IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 10, 2020, accepted October 3, 2020, date of publication October 12, 2020, date of current version October 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3030058

Imbalanced Fault Diagnosis of Rolling Bearing Using Enhanced Generative Adversarial Networks

HONGLIANG ZHANG[1], RUI WANG[1], RUILIN PAN[1,2], AND HAIYANG PAN[3]
[1]School of Management Science and Engineering, Anhui University of Technology, Ma'anshan 243032, China
[2]The Key Laboratory of Multidisciplinary Management and Control of Complex Systems, Department of Education Anhui Province, Ma'anshan 243032, China
[3]School of Mechanical Engineering, Anhui University of Technology, Ma'anshan 243032, China

Corresponding author: Ruilin Pan (alltimefight@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 71772002 and Grant 61702006.

ABSTRACT Machinery fault diagnosis tasks have been well addressed when sufficient and abundant data are available. However, the data imbalance problem widely exists in real-world scenarios, which leads to the performance deterioration of fault diagnosis markedly. To solve this problem, we present a novel imbalanced fault diagnosis method based on the enhanced generative adversarial networks (GAN). By artificially generating fake samples, the proposed method can mitigate the loss caused by the lack of real fault data. Specifically, in order to improve the quality of generated samples, a new discriminator is designed using spectrum normalization (SN) strategy and a two time-scale update rule (TTUR) method is used to stabilize the training process of GAN. Then, an enhanced Wasserstein GAN with gradient penalty is developed to generate high-quality synthetic samples for the fault samples set. Finally, a deep convolutional classifier is constructed to carry out fault classification. The performance and effectiveness of the proposed method are validated on the Case Western Reserve University bearing dataset and rolling bearing dataset acquired from our laboratory. The simulation results show that the proposed method has a superior performance than other methods for imbalanced fault diagnosis tasks.

INDEX TERMS Fault diagnosis, rolling bearing, generative adversarial networks, imbalanced data, convolutional neural networks.

I. INTRODUCTION

As one of the most important components for rotating machinery, rolling bearing is widely used in manufacturing system, electric system, and other mechanical equipment. Unfortunately, rolling bearing is subject to unexpected failures under complex operating conditions, which causes huge economic loss and casualties in engineering practice [1], [2]. Therefore, it is of great significance to study rolling bearing fault diagnosis to ensure the safety and reliability of facilities.

Traditional fault diagnosis algorithms can extract fault features from raw vibrational signal to recognize fault types. Feature extraction methods in the time and frequency domains, such as wavelet transform [3], variational mode decomposition [4], permutation entropy [5], have been widely used to improve fault diagnosis performance in existing research. Besides, intelligent fault diagnosis methods have been successfully developed in recent years. Especially, data-driven fault diagnosis approaches, such as deep learning (DL) [6], have shown its superiority in the task of fault classification since the prior knowledge of automatic system is not required. For DL-based methods, abundant balanced data are the primary requirement to fully train deep neural networks. However, in most time of industrial processes, rolling bearing usually works in normal state and the fault data are hard to collect. The amount of data acquired in normal state is far more than that in faulty state. Therefore, there is an imbalance between normal samples and fault samples, which leads to the performance deterioration of fault diagnosis markedly.

Existing works on the data imbalance problem can be classified into two classes. The first class focuses on utilizing data preprocessing techniques, such as over-sampling and under-sampling, to balance the data distribution. The Synthetic Minority Oversampling Technique (SMOTE) [7] is one of the well-established methods to address the data imbalance problem. Elreedy and Atiya [8] conducted a comprehensive

The associate editor coordinating the review of this manuscript and approving it for publication was Farid Boussaid.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

analysis of SMOTE from both theory and experiment aspects. The experiment results demonstrated that SMOTE was an effective method to generate extra samples for the fault samples set. Douzas *et al.* [9] combined SMOTE with K-means algorithm to avoid the generation of noise and overcome the imbalance between and within classes. Maldonado *et al.* [10] proposed an improved SMOTE over-sampling approach to deal with the class-imbalanced problem for high-dimensional datasets. Zhang *et al.* [11] proposed a novel synthetic over-sampling approach called weighted minority over-sampling (WMO) and set up a three-stage fault diagnosis model, which achieved better results than that of SMOTE-based data learning methods. The second class focuses on improving the traditional diagnosis algorithm structure or designing new algorithms by means of the distribution characteristics of fault data. Mao *et al.* [12] presented an online sequential prediction method based on extreme learning machine, in which the principal curve and granular division strategy were introduced to conduct over-sampling and under-sampling processes. A new support vector classifier integrating support vector data description was proposed by Duan *et al.* [13] to deal with fault classification problems for imbalanced datasets. Jia *et al.* [14] addressed the imbalanced fault diagnosis problem from data distribution perspective. A deep normalized convolutional neural network and a neuron activation maximization algorithm were used. Although these methods have made improvements on the imbalanced fault classification tasks, one of their drawbacks is hard to generate high-quality data samples.

As a prospective deep learning tool, GAN was firstly introduced by Good-fellow *et al.* [15] in 2014, which is composed of two models: generator and discriminator, and has been successfully applied in the fields of speech recognition [16], image generation [17], sentiment analysis [18], etc. Since GAN is able to learn the potential distribution of original samples, and then generate new samples which are different from original samples but have similar distribution, it has attracted attention of many scholars on fault diagnosis field in recent years. Based on deep generative neural networks, Li *et al.* [19] proposed a cross-domain fault diagnosis method. When the testing data under machine fault conditions were not available for training, the proposed method can provide effective diagnosis results. Mao *et al.* [20] presented an imbalanced fault diagnosis method based on GAN and Stacking Denoising Auto Encoder (SDAE), in which GAN was used to generate the synthetic minority samples and SDAE was regarded as a classifier to diagnose fault types. By using Wasserstein distance to construct Wasserstein generative adversarial networks (WGAN), Wang *et al.* [21] presented a deep learning model to study the imbalanced fault diagnosis for rotating machinery.

In order to enhance model stability and improve the quality of generated samples, Gao *et al.* [22] proposed a data augmentation approach based on Wasserstein generative adversarial network with gradient penalty (WGAN-GP), which redesigned the loss function of WGAN [23]. Shao *et al.* [24] employed one-dimensional convolutional neural network (1D-CNN) to construct an auxiliary classifier GAN (ACGAN) for data augmentation, where additional label information was conducive to generating the corresponding fault samples. Zhou *et al.* [25] designed a new generator and discriminator using global optimization mechanism to generate discriminant fault samples. Li *et al.* [26] utilized an adaptive training ratio strategy to enhance the convergence and stability of GAN.

Through literature review, it can be found that the fault diagnosis performance with imbalanced data can be improved by utilizing GAN architecture to generate synthetic samples. However, due to the existing of model collapse and vanishing gradient problem, GAN tends to generate some meaningless samples in the training process, which leads to the performance deterioration of fault diagnosis dramatically. Thus, the quality of generated samples is the key factor to improve the performance of imbalanced fault diagnosis. Although the aforementioned works enhance the quality of generated samples to some extent, the performance control of GAN model is still a persisting challenge. To enhance the stability of GAN model and obtain high-quality synthetic samples, this paper proposes a novel fault diagnosis method based on WGAN-GP method to address the data imbalance problem. Utilizing GAN architecture, the proposed method can learn the potential distribution of each health condition and generate new synthetic samples to expand and balance the fault samples set. Especially, with the goal of learning the high-dimensional manifold of real distribution accurately, a new discriminator is designed using spectrum normalization strategy. Meanwhile, in order to make GAN have a better convergence, a two time-scale update rule (TTUR) method is used to stabilize the training process of WGAN-GP. Furthermore, based on the intrinsic characteristics of time series, the deep 1D-CNN classifier is constructed to extract hierarchical features and carry out fault classification. The main contributions of this paper are summarized as follows.

1) Based on the intrinsic characteristics of time series, this paper redesigns the structure of discriminator using spectrum normalization strategy. The new discriminator has excellent learning ability for multimodal data structure since the spectrum normalization layer imposes a mild constraint on network parameters. Moreover, compared with other normalization layers, spectral normalization layers are computationally light and easy to incorporate into WGAN-GP.

2) To make the generator have low-rank perturbations during the training process, this paper takes advantage of a two time-scale update rule to enhance the training stability of WGAN-GP. Compared with the equal time-scale rule, TTUR can prevent the discriminator from overtraining on the current generator. Meanwhile, WGAN-GP trained with TTUR converges to a locally stationary Nash equilibrium.

3) The proposed method can learn the data manifold of fault samples set with more stable process and generate samples with higher quality. The high-quality generated samples can be used to train fault classifier together with original

samples to improve the generalization ability of the diagnostic model. This will promote the practical application of our method in intelligent fault diagnosis under small sample size conditions.

The rest of paper is organized as follows. Theoretical background about convolutional neural networks, generative adversarial networks, and Wasserstein GAN with gradient penalty is introduced in Section II. Section III illustrates the proposed fault diagnosis method in detail. Experimental results and analysis are given in Section IV. The final section concludes this paper with some discussions on the future work.

II. THEORETICAL BACKGROUND

A. CONVOLUTIONAL NEURAL NETWORKS

Based on hierarchical network structure, the CNN can extract discriminative features from a great deal of redundant information. In general, the CNN is composed of three main basic units: convolutional layer, pooling layer and fully-connected layer.

The input signal of CNN is denoted as $X = [x_1, x_2, \ldots, x_L]$, where L denotes the length of raw vibrational signal. When it is fed into the convolutional layer, the corresponding convolutional operation can be conducted as follows:

$$z_j^{(l)} = \varphi(\sum_{i \in K_j} x_i^{(l-1)} * w_{ij}^{(l)} + b_j^{(l)}) \tag{1}$$

In order to reduce the size of convolution features and filter redundant information, the output features in convolution layer will be transferred to the pooling layer to conduct a pool operation. The max pooling function is used to extract the maximum value of the input features. Specifically, the max pool process is described as:

$$p_j^{(l)} = \max \left\{ z_{j \times W:(j+1) \times W}^{(l)} \right\} \tag{2}$$

After convolution and pooling operations, the input features extracted by the previous layers are mapped into one-dimensional vector through the fully-connected layer without information loss. Finally, the health conditions of machines can be recognized.

B. GENERATIVE ADVERSARIAL NETWORKS

Inspired by binomial zero-sum game theory, GAN consists of two parts: the generator G and the discriminator D, which are trained in opposition to each other. The generator aims to generate realistic synthetic samples to fool the discriminator, while the discriminator is used to distinguish real samples from fake samples produced by the generator. The training process can be briefly described as follows. Firstly, the generator G takes a random noise vector z as input and produces synthetic samples $Xg = G(z)$, then synthetic samples $G(z)$ and real samples X are mixed as the input of discriminator D, which is trained to identify fake samples and real data by outputting the true or false probability. During the training process of GAN, the generator G and the discriminator D are trained alternately until reaching the Nash equilibrium. Concretely, the objective function of GAN model is expressed as:

$$\begin{aligned} &\min_G \max_D L(D, G) \\ &\quad = E_{x \sim P_r}[\log D(x)] + E_{z \sim P_g}[\log(1 - D(G(z)))] \end{aligned} \tag{3}$$

By the adversarial training mechanism, GAN shows a powerful ability to generate synthetic samples which are different from original samples but with similar distribution. Under ideal conditions, the process of zero-sum game can achieve a global optimum at $P_g = P_r$, i.e., the data distributions of generated samples and real samples are coincident.

C. WASSERSTEIN GAN WITH GRADIENT PENALTY

In practice, the training process of GAN is always unstable or even difficult to train, mainly because the Jensen-Shannon (JS) divergence is discontinuous, which tends to cause the discriminator saturates and gradients vanish in the training process. In order to solve the above-mentioned problems effectively, Wasserstein GAN (WGAN) [23] is proposed, which uses Wasserstein distance *W(Pg,Pr)* to measure the difference between generated distribution and real distribution. Specifically, the Wasserstein distance *W(Pg,Pr)* represents the minimum cost of mass to move the generative distribution Pg to real distribution Pr. The objective function of WGAN is expressed as:

$$\min_G \max_{D \in R1} E_{x \sim P_r}[D(x)] - E_{z \sim Pg}[D(z)] \tag{4}$$

where $R1$ denotes the set of 1-Lipschitz functions.

The discriminator of WGAN essentially becomes a critic of distance measurement rather than merely the classifier of real and fake samples. Meanwhile, as the Wasserstein distance is continuous, WGAN has some inherent advantages, i.e., the training process is more stable and the quality of generated samples is higher.

To make the critic satisfy the Lipschitz constraint, WGAN clips the weights in each layer of critic into the range [-c, c]. Unfortunately, this strategy not only restricts the ability of the critic to fit complex data distribution, but may cause the vanishing gradient problem. To improve the performance of WGAN, Wasserstein GAN with gradient penalty (WGAN-GP) was proposed [27], which used an alternative method named gradient penalty to improve the value function of WGAN. The objective function of WGAN-GP is represented as follows:

$$E_{z \sim Pg}[D(z)] - E_{x \sim P_r}[D(x)] + \chi E_{\widehat{x} \sim P_{\widehat{x}}}[(\|\nabla_{\widehat{x}} D(\widehat{x})\|_2 - 1)^2] \tag{5}$$

Compared with the objective function of WGAN, WGAN-GP adds a new gradient penalty term, in which λ is the penalty coefficient, and $\widehat{x}$ represents the random sampling along the straight lines between real distribution Pr and generative distribution Pg. It is noticed that the gradient penalty term makes all gradient norms of the critic be close

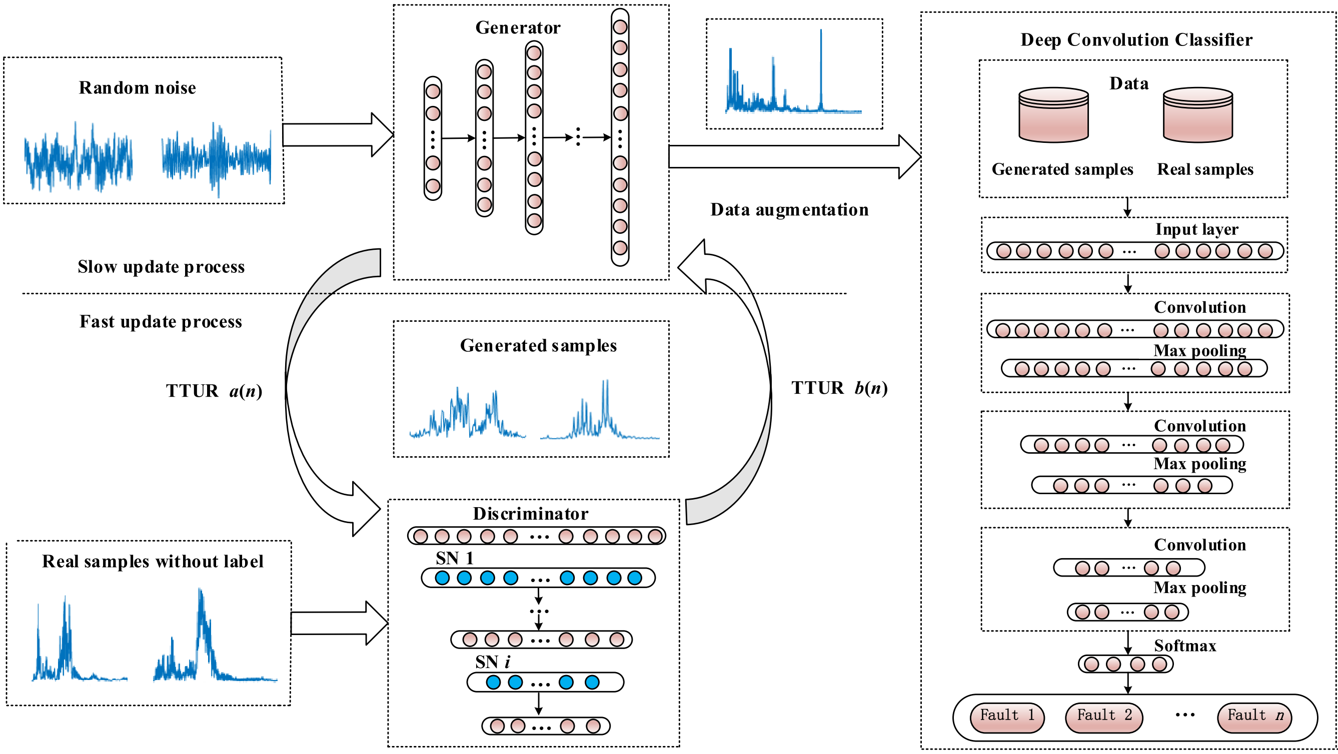

FIGURE 1. The flow diagram of enhanced model.

to 1 rather than restrict network weights to a compact space. In this way, the critic will get a more stable gradient and a faster convergence process. As a result, WGAN-GP can generate data samples with higher quality.

III. IMBALANCED FAULT DIAGNOSIS BASED ON THE ENHANCED WGAN-GP

As described in II-C, WGAN-GP can effectively mine the intrinsic distribution of a dataset using only a few samples. Its generator is able to generate samples from random noise of latent space, and its discriminator determines whether the input samples come from the real data distribution or not. Because the CNN is capable of fitting complex data distribution and approximating the density ratio of real distribution, it can be used to build the generator and discriminator. The training procedures of WGAN-GP are briefly presented as follows.

1) The generator produces synthetic samples from random noise of latent space.

2) Generated samples and real data samples are mixed together and fed into the discriminator. Based on the equation (5), the discriminator is trained and its parameters are updated.

3) After training the discriminator, the parameters of it are frozen. During this step, only the parameters in generator need to be updated, and the generator is trained to produce more realistic synthetic samples.

4) Repeating the processes above until reaching the Nash equilibrium.

Based on WGAN-GP method, this paper proposes a novel fault diagnosis method to solve the data imbalance problem. The framework of the proposed method is shown in Fig. 1, and the whole framework consists of three components: sample preprocessing, the enhanced WGAN-GP method and deep convolutional classifier. Each component is described in detail as follows.

Sample Preprocessing: As rolling bearing generally operates under complex conditions, the vibration signal sampled from transducer often contains complex and strong noise, which greatly influences the extraction of signal feature. To eliminate the interference of noise signal, this paper uses fast Fourier transform to obtain the frequency spectrum of raw vibration signal. Compared with the feature extracted from original vibration signal, the frequency spectrum samples contain more significant features while eliminating noise interference. Therefore, this paper selects frequency spectrum samples as the inputs of GAN to generate synthetic samples.

Enhanced WGAN-GP: After sample preprocessing, an enhanced model based on WGAN-GP method is presented to generate the spectrum synthetic samples whose distribution is similar to original samples. In order to make the training process much more stable and generate high-quality samples, this paper applies two crucial strategies to improve WGAN-GP method, i.e., a two time-scale update rule (TTUR) [28] and spectral normalization [29]. The enhanced model is described in III-A in detail. When the enhanced model reaches the Nash equilibrium after fully training, the generator is capable of generating realistic synthetic samples for the fault samples set, and the high-quality generated samples are mixed with real samples to expand and balance the training dataset.

Deep Convolutional Classifier: In the fault diagnosis stage, the generated samples along with real samples are mixed together and fed into fault classifier to build its diagnostic ability. Then, the test dataset is inputted to fault classifier to obtain diagnostic results. The detailed information about deep convolutional classifier is described in III-C.

A. THE ENHANCED WGAN-GP METHOD

Although WGAN-GP employs gradient penalty mechanism to avoid the occurrence of vanishing gradient problem, its training process may suffer from some unexpected situations. For instance, when the discriminator is saturated enough in the initial training stages, the generator generates some meaningless samples, which leads to the performance deterioration of fault diagnosis. To further improve the quality of generated samples and achieve better fault diagnosis performance, this paper applies two crucial strategies to enhance the training stability of WGAN-GP, including spectral normalization and TTUR strategy.

For mechanical vibration signals, their data distribute in a high-dimensional space, and the density ratio estimation of discriminator is often inaccurate and unstable during the training process. Once the discriminator cannot perfectly distinguish the real distribution from generated distribution, the generator will fail to learn the multimodal structure of the real distribution. Even worse, when the derivative of the discriminator with respect to the input turns out to be 0, the training of the generator comes to complete stop. To avoid the above-mentioned problems, many training tricks have been developed to enhance the training stability of discriminator, including batch normalization (BN), layer normalization (LN), and so on. Ref. [29] has proved that spectral normalization is a better weight normalization technique than other regularization terms. Spectral normalization controls the Lipschitz constant of discriminator function f by literally constraining the spectrum norm of each layer, which makes the discriminator not require intensive tuning while Lipschitz constant is the only hyper-parameter. By contrast, other normalization terms impose stronger constraint on weight matrix W than intended, which restricts the ability of discriminator to identify generated distribution and real distribution. Therefore, this paper redesigns the network structure of discriminator using spectral normalization strategy.

Typically, the slow update rule corresponds to the generator and the fast update rule is equivalent to the discriminator during the adversarial training process. Once the generator changes slowly enough, the discriminator will go into steady region without capturing the discriminant features of original samples. To ensure the generator has low-rank perturbations, this paper applies TTUR strategy to adjust the training process of WGAN-GP so that the generator does not affect discriminator learning in an undesired way. With regard to TTUR strategy, it uses different learning rates $a(n)$ and $b(n)$ to conduct adversarial training process. As shown in Ref. [28], TTUR improves learning performance of WGAN-GP, making it converge to the locally stationary Nash equilibrium. Detailed network architecture about enhanced WGAN-GP method is shown in Table 1.

TABLE 1. The network structure of enhanced WGAN-GP method.

	Layer Type	Channel	Kernal	Stride
Generator	Upsample 1	-	-	-
	Convolution 1	128	8	1
	Upsample 2	-	-	-
	Convolution 2	128	8	1
	Upsample 3	-	-	-
	Convolution 3	1	8	1
	Fully-connected	1	1200	-
Discriminator	Convolution 1	32	16	2
	SN 1	-	-	-
	Convolution 2	8	16	2
	SN 2	-	-	-
	Convolution 3	4	16	2
	SN 3	-	-	-
	Fully-connected	1	564	-

B. ADVERSARIAL TRAINING FOR THE ENHANCED WGAN-GP METHOD

The enhanced WGAN-GP method contains three main parts: the generating process of generator, the training process of discriminator and adversarial training mechanism. The detailed process is described as follows.

Step 1: the random noise $\{z^{(i)}\}_{i=1}^{m}$ sampled from Gaussian distribution Pg is input into the generator, where m denotes the size of a mini-batch. Then, the synthetic samples $G(z^{(i)})_{i=1}^{m}$ with the same size of real samples are generated. Specifically, there are three upsampling layers and three layers of 1D-convolution operation followed by a batch of normalization respectively in generator.

Step 2: After synthetic samples generation, the real batch samples $\{x^{(i)}\}_{i=1}^{m}$ sampled from the real distribution along with the synthetic samples are fed into the enhanced discriminator for authenticity discrimination, which obtains the predicted values d_{real} and d_{fake}. In detail, the enhanced discriminator consists of three 1D-CNN layers followed by a spectral normalization layer respectively, three dropout layers, a flatten layer and a fully-connected layer. Based on the gradient penalty term, the discriminator is trained by maximizing the loss function, as shown in Eq. (5). Furthermore, the updated values of the parameters of D are obtained by implementing an Adam optimization method.

Step 3: After the discriminator is trained, the synthetic samples $G(z^{(i)})_{i=1}^{m}$ are input into the discriminator for authenticity discrimination. It is noticed that the generator is optimized by means of discriminator. The corresponding loss function is denoted as:

$$L_G = -\frac{1}{m}\sum_{i=1}^{m} D(G(z^{(i)})) \tag{6}$$

Similarly, the updated values of the parameters of G are obtained by implementing an Adam optimization method, while keeping the parameters in D constant.

***Step 4*:** During the training process, the discriminator and the generator are optimized alternatively until reaching the Nash equilibrium. At the end of adversarial training process, the generator is capable of generating synthetic samples that are similar to real samples, and the discriminator cannot distinguish the fake samples from the real samples any more.

C. IMPLEMENTATION OF FAULT DIAGNOSIS MODEL

Considering the intrinsic characteristics of time series and the advantages of 1D-CNN in analyzing sensor signal, this paper constructs the fault classifier based on deep 1D-CNN, as shown in Fig. 1. Firstly, the enhanced WGAN-GP method is used to generate the fault samples set, and the real samples $\{x_i, y_i\}_{i=1}^{N}$ and the generated samples $\{G(z^{(i)})\}_{i=1}^{M}$ are mixed together to expand and balance the training set. Then, they are fed together into 1D-CNN network to extract hierarchical representations. Finally, a SoftMax classifier is utilized to carry out fault classification and outputs specific labels.

IV. EXPERIMENTS VERIFICATION AND ANALYSIS

To verify the performance and effectiveness of the proposed method, we first testify it on rolling bearing benchmark dataset from Case Western Reserve University (CWRU) [30], and then further validate it by using HVC bearing fault dataset obtained from a fault simulation testbed of our laboratory.

A. COMPARISON METHODS AND EVALUATION METRICS

We first use the CNN model trained with the imbalanced datasets as baseline comparison model, which determines the lower bound of diagnosis accuracy. Then, SMOTE [7] and its evolution ADASYN [31], well-established methods addressing the data imbalance problem, are compared with our method. SMOTE creates new samples for each health condition by nearest neighbor algorithm. ADASYN improves the performance of SMOTE by adding noise to the new samples, which avoids linear correlation with the parent instances. The new samples and the original samples are mixed together to expand and balance the training dataset. After the training dataset is expanded, the same classifier with CNN is built to conduct fault classification tasks. Furthermore, we compare two other generative models with our approach to evaluate the incidence of GAN model, including WGAN and WGAN-GP.

To quantitatively analyze fault diagnosis performance of the proposed method, we introduce several performance metrics tools: accuracy, recall and F1 score, which are widely used in classification problems. As to our fault diagnosis model, accuracy, recall, and F1 score can be calculated as follows:

$$accuracy = \frac{TP + TN}{N} \tag{7}$$

$$recall = \frac{TP}{TP + FN} \tag{8}$$

TABLE 2. Confusion matrix.

Confusion matrix		True label	
		Positive	Negative
Predicted label	Positive	TP	FP
	Negative	FN	TN

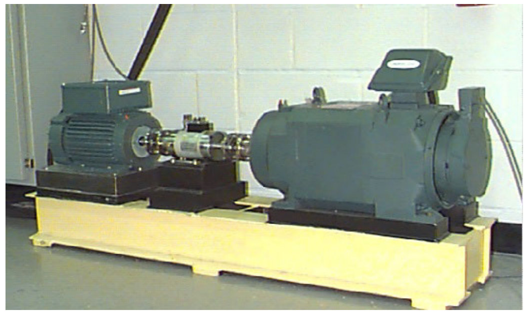

FIGURE 2. The CWRU bearing testbed.

$$F1 = \frac{2 * precision * recall}{precision + recall} \tag{9}$$

where N denotes the total number of samples, TP is the number of the current fault sample S correctly classified as S, FP is the number of other categories samples incorrectly predicted to be sample S, and FN is the number of the current fault sample S incorrectly classified as other categories. The above three values can be obtained from the confusion matrix, shown in Table 2.

B. CASE STUDY ON CWRU BEARING DATASET

1) DATA DESCRIPTION AND PREPROCESSING

As shown in Fig. 2, the test stand consists of a two-hp motor, a torque transducer/encoder, a dynamometer, and control electronics. The test bearings support the motor shaft. Subjected to electro-sparking, four different fault types are introduced, including normal condition (NC), inner-race fault (IF), outer-race fault (OF) and ball fault (BF). The vibration signal is sampled by accelerometers attached to the housing with magnetic bases. Besides, the vibration signal is collected from the drive end of motor under 3 different loads (1 hp, 2 hp, 3 hp) with a sampling frequency of 48 kHz. By using electro-discharge machining with fault diameters of 0.007 inches, 0.014 inches, 0.021 inches and 0.028 inches, the vibration signal is processed.

Raw vibration signal is collected with a 48 kHz sampling frequency for 10 seconds, containing 480000 data points. A sample with 2400 points is successively selected from raw vibration signal. Then, FFT is implemented on each signal and the 2400 Fourier coefficients are obtained. Since the coefficients are symmetric, the first 1200 coefficients are used in each sample. The working conditions vary along with three motor loads, corresponding to three datasets A/B/C. Each dataset consists of ten classes, including nine kinds of faults and a normal condition, and each class contains

TABLE 3. Description of the datasets.

Dataset	Working load (hp)	Numbers of samples	Fault types	Fault diameter (inches)	Label
A/B/C	1/2/3	200/200/200	NC	-	1
		200/200/200	BF	0.007	2
		200/200/200	IF	0.007	3
		200/200/200	OF	0.007	4
		200/200/200	BF	0.014	5
		200/200/200	IF	0.014	6
		200/200/200	OF	0.014	7
		200/200/200	BF	0.021	8
		200/200/200	IF	0.021	9
		200/200/200	OF	0.021	10

TABLE 4. Parameters for enhanced WGAN-GP method.

Notation	Description	Value
r	Random vector dimension	100
s	A mini-batch	64
$a(n)$	The learning rate of generator	5×10^{-4}
$b(n)$	The learning rate of discriminator	10^{-4}
R	The ratio of training times of D and G	2
N_1	Maximum adversarial training epoch	1000
λ	Gradient penalty coefficient	1
IR	Initial learning rate	10^{-2}
N_2	Maximum diagnosis epoch	50

200 samples. The detailed information of the datasets is presented in Table 3.

Datasets A/B/C are processed according to the size of imbalance ratio, which is defined as the ratio of the number of normal samples to that of samples of each fault type. Taking the imbalance ratio 5:1 as an example, each normal condition contains 200 samples, so the number of samples of each fault type is 40. Then, each class are further divided into a training set and a testing set at a ratio of 9:1. Therefore, there are 180 normal samples that can be used to train the diagnostic model. Similarly, the number of samples of each fault type in the training set is 36, and the total number of fault samples of nine fault classes in the training set is 324. The rest of all samples are used for testing. As a result, 504 samples in each dataset are used for training and the diagnosis performance is testified on 1496 samples.

2) IMPLEMENTATION DETAILS

In order to achieve the optimal fault diagnosis performance of the proposed method, we conduct appropriate fine-tuning for the parameters of model, and report the best parameter settings, which are listed in Table 4. As for the parameter settings of GAN model, the Gaussian distribution is used as the prior noise distribution and the random noise vector dimension r is set as 100. The size of a mini-batch s is 64, the learning rate of the generator $a(n)$ and the discriminator $b(n)$ is 5×10^{-4} and 10^{-4}, respectively. The ratio R of the training times of D and G in one adversarial training epoch is 2. The maximum adversarial training epoch N_1 is 1000.

TABLE 5. The various imbalance ratios on the training set.

Imbalance ratio	NC	BF	IF	OF
2:1	200	100	100	100
5:1	200	40	40	40
10:1	200	20	20	20
20:1	200	10	10	10
50:1	200	4	4	4

The gradient penalty coefficient λ is 1. We implement our approach by using PyTorch framework, and an Adam optimization method is used to update the weight sets of all neural networks with momentums $\beta_1 = 0.5$ and $\beta_2 = 0.9$. The detailed network parameter settings of our approach can be found in Table 1.

In the fault diagnosis stage, we adjust the learning rate with a decay rate 0.1, i.e., for each 10 epochs of training, the learning rate decreases to one tenth of the current value, and initial learning rate IR is 10^{-2}. Similarly, the size of mini-batch is 64 and an Adam optimization method is utilized. The training epoch N_2 is 50. In addition, in order to evaluate the efficiency of our model under various imbalance degrees, five comparison experiments are carried out by setting imbalance ratio as 2:1, 5:1, 10:1, 20:1 and 50:1 respectively, shown in Table 5. The larger the imbalance ratio, the smaller the number of fault samples in training sets is. All above experiments are repeated 10 times in each case and the average value is taken as the final fault diagnosis result.

3) EXPERIMENT RESULTS

As shown in Table 6, the diagnosis results for the testing sets of datasets A, B and C are obtained under the above-mentioned parameter settings. It is observed that the CNN model can only achieve lower accuracy of testing with imbalanced data, meaning that the CNN model is not able to distinguish each health conditions effectively. Whereas, our approach obtains the higher accuracy of testing with a lower standard deviation, and the scores of recalls and F1 are also higher. The testing accuracy of our approach for each dataset is greater than 99%, which suggests that the proposed method

TABLE 6. Diagnosis accuracy using the CNN and enhanced model for datasets A/B/C.

Accuracy	Dataset A		Dataset B		Dataset C	
Label	CNN	Enhanced	CNN	Enhanced	CNN	Enhanced
1	100.0	100.0	100.0	100.0	100.0	100.0
2	83.51	100.0	78.19	98.40	84.57	99.47
3	97.34	100.0	95.74	100.0	100.0	100.0
4	100.0	100.0	98.45	100.0	94.68	100.0
5	71.81	99.47	89.89	100.0	96.27	97.87
6	99.47	99.47	93.62	100.0	84.04	97.87
7	94.14	100.0	92.55	97.87	93.65	99.47
8	76.59	100.0	75.53	100.0	54.79	99.47
9	77.13	100.0	93.62	100.0	90.42	99.47
10	94.14	100.0	88.29	96.28	97.87	100.0
Average	89.41	**99.31**	90.59	**99.26**	89.63	**99.36**
Recall	88.72	**99.23**	89.53	**98.94**	88.06	**99.30**
F1	89.15	**99.23**	90.34	**98.94**	88.21	**99.30**
Std	1.06	**0.13**	1.38	**0.27**	0.99	**0.16**

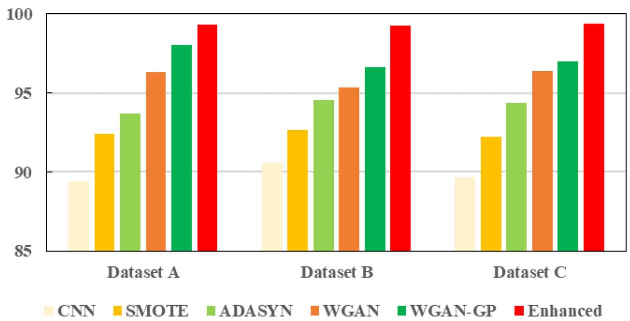

FIGURE 3. The increments of diagnosis accuracy on three datasets under imbalance ratio 5:1.

can significantly improve the performance of fault diagnosis with imbalanced data. The standard deviations under ten independent trials are below 0.3%, indicating that our approach has stable diagnostic ability.

In order to show the diagnosis result obtained from each dataset clearly, Fig. 3 is given to show the increment of the diagnosis accuracy of our approach relative to other comparison methods when the imbalance ratio is 5:1. It can be seen that SMOTE and ADASYN can improve fault diagnosis performance with imbalanced data, and the testing accuracy of SMOTE and ADASYN is 92.68% and 94.67%, respectively. Compared with SMOTE and ADASYN, WGAN and WGAN-GP achieve relatively higher accuracy of testing. The average testing accuracy is 96.32% and 98.02%, respectively, which illustrates the synthetic samples generated by GAN model can contribute to fault diagnosis. Remarkably, it is clearly seen from Fig. 3 that the enhanced method can obtain the highest accuracy of testing. Concretely, the average testing accuracy of our approach reaches 99.31%, 99.26% and 99.36%, respectively, which shows that the quality of generated samples is significantly improved, thus facilitating the final fault classification.

To further validate the performance of our approach, the comparisons of diagnostic results under different imbalance ratios corresponding to datasets A, B and C, are listed in Table 7. It can be seen that as the imbalance ratio increases, the diagnosis results of the CNN model gradually become worse, which means that the diagnosis results based on CNN model depend on the training sample size. Although SMOTE and ADASYN algorithms improve the fault diagnosis performance to some extent, the diagnosis accuracy under extremely imbalanced data conditions is poor. When the imbalance ratio is 50:1, the testing accuracy of SMOTE corresponding to dataset A/B/C is only 53.24%, 52.71% and 45.97%. Under the same conditions, the accuracy of WGAN and WGAN-GP rises about 16%. It is obvious that the diagnosis results of our approach reach 73.61%, 77.64% and 63.22%, which is superior to other comparison methods. It is noticed that under different imbalance ratios, our approach maintains a significant superiority. Especially under extremely imbalanced data conditions, the diagnostic results of our approach have obvious advantages.

4) ANALYSIS AND DISCUSSION

In order to explain why our approach can achieve excellent performance on rolling bearing fault diagnosis with imbalanced data, we provide a visual insight for the training process of enhanced WGAN-GP method, including the similarity between generated samples and real samples, and the high-level representations extracted by the discriminator.

The purpose of using GAN architecture is to generate realistic synthetic samples. Therefore, it is of significance to evaluate the similarity between generated samples and real samples. For image generation tasks, there are several critical evaluation metrics for quantitative analysis, such as inception score (IS) [32] and the Fréchet inception distance (FID) [28]. However, for mechanical vibration signals, it is unreasonable to calculate the scores of the above metrics directly. As the frequency spectrum can reflect inherent characteristics of raw vibration signal, we utilize the frequency spectrum of generated samples and real samples to visualize the adversarial training process of the enhanced model. Taking the imbalance ratio 5:1 as an example, the frequency spectrum

TABLE 7. Diagnosis accuracy for the dataset A/B/C under different imbalance ratios.

Dataset	Imbalance ratio	CNN	SMOTE (CNN)	ADASYN (CNN)	WGAN (CNN)	WGAN-GP (CNN)	Enhanced
A	2:1	93.50	94.89	95.21	98.56	99.23	100.0
	5:1	89.41	92.43	93.68	96.32	98.02	99.31
	10:1	80.01	86.54	88.26	90.04	91.20	95.58
	20:1	68.15	75.12	78.23	83.31	85.45	93.87
	50:1	45.33	53.24	56.36	61.52	63.08	73.61
B	2:1	93.65	95.04	96.18	98.24	99.12	100.0
	5:1	90.59	92.68	94.57	95.34	96.64	99.26
	10:1	81.75	84.81	87.36	90.41	92.25	97.60
	20:1	68.88	73.45	76.17	80.05	83.06	89.24
	50:1	43.35	52.71	56.69	62.24	65.32	77.64
C	2:1	92.90	94.43	95.36	98.15	98.56	100.0
	5:1	89.63	92.21	94.38	96.37	97.02	99.36
	10:1	80.12	85.02	87.96	91.25	93.54	96.06
	20:1	64.75	72.84	75.54	79.23	81.05	87.40
	50:1	40.31	45.97	47.07	52.64	54.16	63.22

of generated samples and the corresponding real samples for dataset A is presented in Fig. 4. It can be seen from this figure that the generated samples have the same trend with the real samples, which means that the generator is able to learn the potential distribution of the real samples. In particular, the generator has captured the inherent characteristics of frequency spectrum samples, which greatly facilitate the final fault classification. Thus, our approach obtains a higher diagnosis accuracy.

To demonstrate how the discriminator distinguishes the samples in each condition, we visualize the features extracted from the fully-connected layer before the output layer. t-Distributed Stochastic Neighbor Embedding (t-SNE) is a technique for dimensionality reduction, which is widely used to visualize deep neural networks. In this paper, we take advantage of t-SNE to visualize the high-dimensional data representation by mapping the samples from the original feature space into a 3-dimensional space map.

When the imbalance ratio is 5:1, the visualization results for the datasets A, B and C are displayed in Fig. 5. For convenience, EGAN_A denotes the visualization result of our approach corresponding to dataset A. The others follow the similar meanings. It can be observed that the sample points with the same label cluster together, but different types of fault samples cannot be separated completely in the CNN model. When the training set is expanded by the proposed method, it is clearly noticed that the same label samples cluster well, and meanwhile all the data samples of different health conditions are completely separated. That is the basis for accurate fault diagnosis. Furthermore, the preferable divisibility for different health conditions indicates that the generated samples are able to provide extra discriminant information for the final fault classification. Although the enhanced WGAN-GP method is trained by means of unsupervised learning, the quality of generated samples is prominent. That is why our approach outperforms comparison methods with imbalanced data.

TABLE 8. The detailed description about experiment data.

Dataset	Load (KN)	Fault type	Fault size (mm)	Label
A/B	0/2	NC	-	1
		BF	0.2	2
		IF	0.2	3
		IF	0.4	4
		OF	0.2	5
		OF	0.3	6

C. CASE STUDY ON HVC BEARING DATASET

1) DATA DESCRIPTION AND PREPROCESSING

To testify the generalization ability of the proposed method, we use it to study the measured signal of the rolling bearing simulation fault testbed. As presented in Fig. 6, the testbed is composed of a control equipment, a load motor, a test bearing, torque sensors, an idler and three accelerometer sensors, in which sensor 1 is used to collect the radial vibration signal, the horizontal vibration signal are sampled by sensor 2, and sensor 3 is utilized to acquire the axial vibration signal. The used test bearings are 6206-2RS deep groove ball bearings and the electrical discharge matching technology is used to set the single point failure to rolling bearings. The faulty locations are produced in the inner race, the outer race and rolling ball element with different sizes (0.2mm, 0.3mm, 0.4mm). The faulty types corresponding to four fault modes can be seen in Fig. 7. The vibration signal is collected under different motor loads (0 KN, 2 KN) with the sampling frequency 51200 Hz. The detailed description about the experimental data is shown in Table 8.

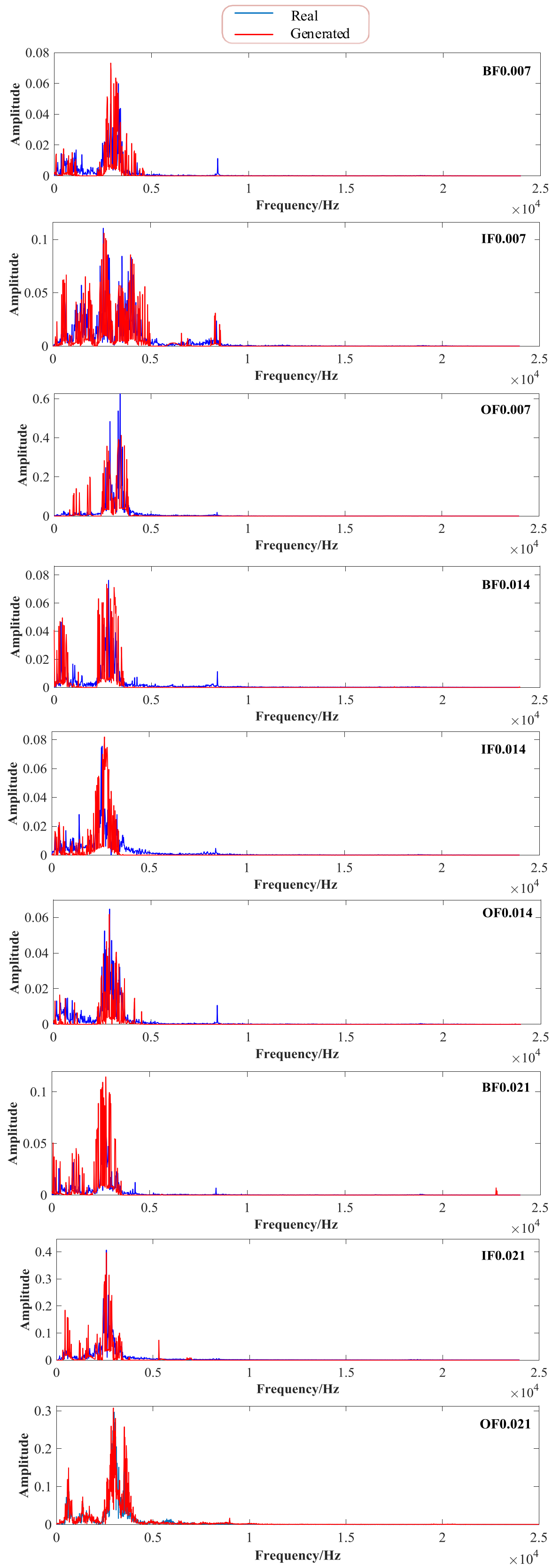

FIGURE 4. The generated samples and the corresponding real samples of the dataset A.

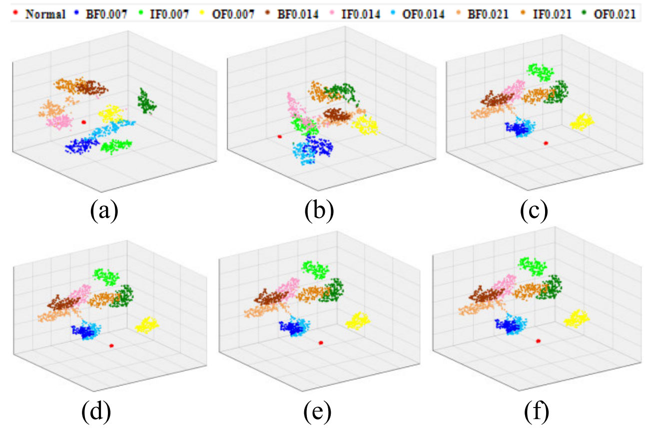

FIGURE 5. Scatters diagrams of the high-level representations extracted by the discriminator for (a) CNN_A, (b) CNN_B, (c) CNN_C, (d) EGAN_A, (e) EGAN_B, (f) EGAN_C.

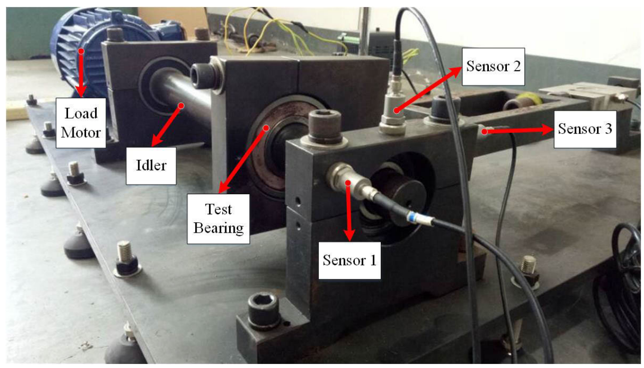

FIGURE 6. The testbed of our laboratory.

In this paper, datasets A and B collected from different working loads (0 KN, 2 KN) are respectively used. For each dataset, the vibration signal is sampled at a sampling frequency of 51200 Hz for 10 seconds, so there are 512000 data points for each health condition. Similarly, the vibration signal with length 1200 is successively selected from the raw vibration signal, and each health condition contains 200 subsamples. Then, FFT is used to obtain 2400 Fourier coefficients and the first 1200 coefficients are used in each sample. In addition, there are six classes under each dataset, including five kinds of faults and one normal state. Therefore, each dataset contains 1200 samples of six classes collected from corresponding working condition.

2) IMPLEMENTATION DETAILS

Based on the experimental setup in IV-B, the network parameters of the enhanced generative model are fine-tuned to achieve optimal fault diagnosis performance. As for the parameter settings of GAN model, we adjust the kernel size of each convolutional layer in discriminator to 8 and the number of channels for each layer is set as 128, 64, and 4, respectively. Meanwhile, the number of convolution layer of discriminator and the parameter settings of generator are unchanged. In addition, the penalty coefficient χ is set as 3,

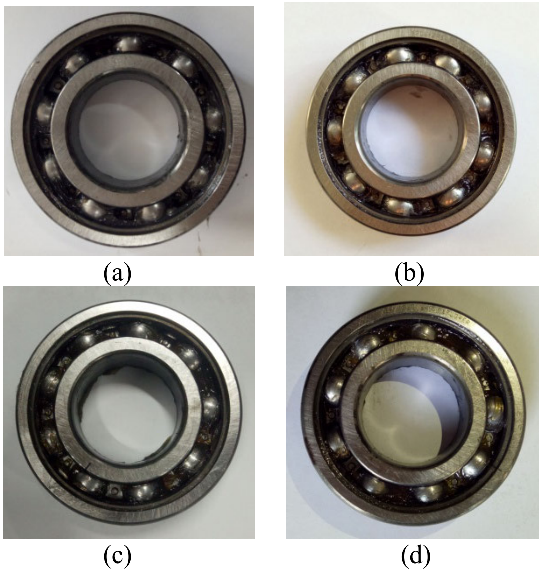

FIGURE 7. The bearings corresponding to the four fault modes: (a) normal state, (b) ball fault, (c) inner race fault, (d) outer race fault.

TABLE 9. Diagnosis accuracy using the CNN and enhanced model of datasets A/B.

Accuracy	Dataset A		Dataset B	
Label	CNN	Enhanced	CNN	Enhanced
1	100.0	100.0	100.0	100.0
2	85.11	100.0	62.23	95.74
3	87.77	100.0	90.42	100.0
4	90.42	100.0	95.74	100.0
5	100.0	100.0	97.87	100.0
6	100.0	100.0	87.77	96.27
Average	93.88	**100.0**	89.01	**98.69**
Recall	92.64	**100.0**	88.20	**98.02**
F1	93.02	**100.0**	88.75	**98.40**
Std	0.70	**0**	1.46	**0.18**

and the ratio R of training times of D and G in one adversarial training epoch is set as 3:1. The learning rates of the generator $a(n)$ and the discriminator $b(n)$ is set as 5×10^{-4} and 2×10^{-4}, respectively. The maximum adversarial training epoch N_1 is set as 500. All neural networks are trained using an Adam optimization method with momentums $\beta_1 = 0.5$ and $\beta_2 = 0.9$ to update weight sets.

In the fault diagnosis stage, each dataset is divided into training sets and testing sets according to the ratio of 9:1. The number of training sets and testing sets depends on the size of imbalance ratio. When the imbalance ratio is 5:1, there are 360 samples for training and 840 samples for testing. To eliminate the effect of randomness, ten independent experiments are conducted for each dataset and the average value is taken as the final fault diagnosis result.

3) EXPERIMENT RESULTS

Table 9 displays the accuracy, recalls and F1 scores of each dataset under different fault conditions. We take the

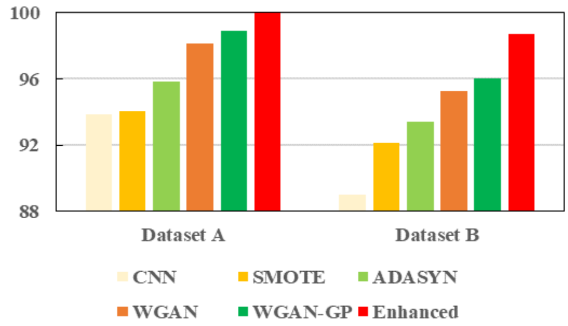

FIGURE 8. The increments of diagnosis accuracy on two datasets under imbalance ratio 5:1.

imbalance ratio 5:1 as representative. The average testing accuracy of CNN model are close to 90% for the two different datasets. By contrast, the accuracy of our approach reaches 99%, which suggests that the proposed method can improve the fault diagnosis performance significantly with imbalanced data in actual scenario. In particular, the proposed method obtains a completely diagnosis result for dataset A. The detailed experimental results are shown in Table 9.

Comparative experiments are conducted for each dataset under various imbalance ratios. Fig. 8 shows the increments of diagnosis accuracy our approach relative to other comparison methods when the imbalance ratio is 5:1. It can be seen from this figure clearly that the proposed method can achieve better diagnosis results compared with other methods. In particular, the testing accuracy of our approach for datasets A/B reaches 100% and 98.69%, respectively, which indicates that the proposed method is able to distinguish each health condition effectively. In addition, comparing enhanced WGAN-GP with WGAN-GP, it can be seen that the performance of enhanced WGAN-GP model is superior to WGAN-GP for each dataset, meaning that the quality of generated sample can be further improved using spectrum normalization and TTUR strategies.

Table 10 shows the fault diagnosis results of several comparison methods which were tested under different imbalance ratios corresponding to datasets A and B. It can be seen from Table 10 that although SMOTE and ADASYN algorithms improve the fault diagnosis performance in actual scenario, the diagnosis accuracy are still poor. Compared with the CNN model, the testing accuracy of SMOTE and ADASYN rises up about 19% when the imbalance ratio is 20:1. Under the same conditions, the accuracy of WGAN and WGAN-GP rises up about 31%. Strikingly, the accuracy of our approach rises about 38%, and the testing accuracy reaches 94.04%.

4) ANALYSIS AND DISCUSSION

In order to explain why our approach can outperform other comparison methods with imbalanced data, the generated samples acquired by the generator of the enhanced

TABLE 10. Diagnosis accuracy for the dataset A/B under different imbalance ratios.

Dataset	Imbalance ratio	CNN	SMOTE (CNN)	ADASYN (CNN)	WGAN (CNN)	WGAN-GP (CNN)	Enhanced
A	2:1	95.34	96.01	96.45	98.51	99.26	100.0
	5:1	93.88	94.02	95.87	98.12	98.93	100.0
	10:1	90.16	92.78	94.02	96.04	97.23	99.47
	20:1	85.05	88.32	88.25	92.26	95.48	98.62
	50:1	80.32	84.04	86.28	88.03	91.20	95.87
B	2:1	95.06	96.42	97.53	98.24	99.06	100.0
	5:1	89.01	92.13	93.41	95.24	96.04	98.69
	10:1	88.28	90.56	91.29	94.46	96.01	97.72
	20:1	67.36	80.20	82.27	86.12	90.36	94.04
	50:1	46.80	74.84	78.52	82.16	85.21	90.54

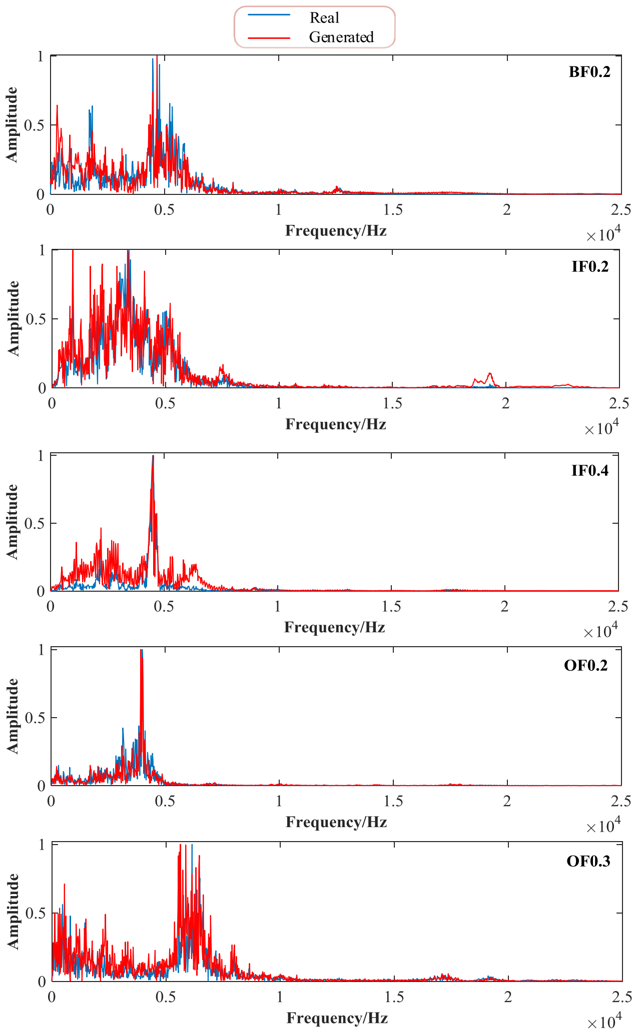

FIGURE 9. The generated samples and the corresponding real samples of rolling bearing.

WGAN-GP and the high-level representations extracted by the discriminator are visualized. Taking the imbalance ratio 5:1 as an example, the frequency spectrum of the generated samples and the corresponding real samples of dataset A are presented in Fig. 9. It can be seen clearly from this figure that the frequency spectrum of generated samples has

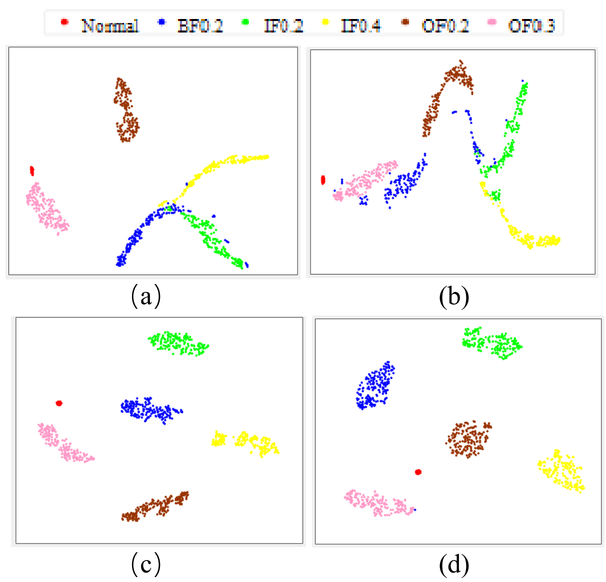

FIGURE 10. Scatters diagrams of the high-level representations extracted by the discriminator for (a) CNN_A, (b) CNN_B, (c) EGAN_A, (d) EGAN_B.

the same trend with that of the real samples. Although there is a slight difference between them, the generator has captured the inherent frequency characteristics from real frequency spectrum, which reflects the generated samples can provide extra discriminative features to boost the final fault diagnosis.

To reveal the discrimination ability of discriminator under each condition, we visualize the features extracted by the fully-connected layer before the output layer. The visualization results obtained by t-SNE under no load condition are shown in Fig. 10. EGAN_A denotes the visualization result of our approach corresponding to dataset A. The others follow the similar meanings. It can be observed that the CNN model cannot identify each type of fault samples completely. When the dataset is expanded by the proposed method, it is noticed clearly that the feature points of the same label cluster well, and all the data samples of different health conditions are completely separated, which reflects the enhanced WGAN-GP can generate data samples with higher quality.

V. CONCLUSION

GAN have been successfully introduced into the field of imbalanced fault diagnosis. Based on the WGAN-GP, this paper proposes a novel fault diagnosis method to address the data imbalance problem which often encounters in the real scenario. The proposed method has strong generalization ability and provides an effective way to solve the intelligent fault diagnosis problem under small sample size conditions. In order to improve the quality of the generated samples, a new discriminator is designed using spectrum normalization strategy. Meanwhile, TTUR strategy is applied to make WGAN-GP converge a locally stationary Nash equilibrium. In this way, the generator is capable of learning the data manifold in real sense with more stable process. During the adversarial training process, the enhanced WGAN-GP can generate new samples that have similar distribution with the original samples to expand the fault samples set. Generated samples of high quality can be used to train fault classifier together with original samples, which improves the generalization ability of diagnostic model. To evaluate the performance and effectiveness of the proposed method, two bearing datasets, including CWRU and HVC, are investigated in the testing experiments. The classification results show that the diagnosis accuracy of the proposed method can reach above 99% for both two bearing datasets. In particular, when we only use 5% of all fault samples to train GAN model, the diagnosis accuracy of the proposed method can still achieve about 90%, which indicates that the proposed method maintains a competitive superiority performance under extremely imbalanced data conditions. Therefore, when the fault data are difficult to obtain in practice, the proposed method has large potential value for industrial application. Furthermore, because the implementation of the proposed method requires less prior knowledge and experience in fault diagnosis fields, this method can be further applied to diagnosis tasks involving other objects.

Although the proposed method can promote the practical application of intelligent fault diagnosis under small sample size conditions, it is assumed that the training and testing data are drawn from the same distribution. However, due to the variation of operating condition, data distribution discrepancy may exist between the training and testing data. Therefore, the next challenge is to validate the effectiveness of our proposed method across domains. Besides, since the training process of GAN requires a lot of parameters adjustment, we will attempt to train GAN by adaptive adjustment mechanism in the future work.

REFERENCES

[1] C. Li, L. Ledo, M. Delgado, M. Cerrada, F. Pacheco, D. Cabrera, R.-V. Sanchez, and J. V. de Oliveiraae, "A Bayesian approach to consequent parameter estimation in probabilistic fuzzy systems and its application to bearing fault classification," *Knowl.-Based Syst.*, vol. 129, pp. 39–60, Aug. 2017.

[2] Z. Wei, Y. Wang, S. He, and J. Bao, "A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection," *Knowl.-Based Syst.*, vol. 116, pp. 1–12, Jan. 2017.

[3] B. J. Chen, B. M. Shen, F. F. Chen, H. L. Tian, W. R. Xiao, F. J. Zhang, and C. H. Zhao, "Fault diagnosis method based on integration of RSSD and wavelet transform to rolling bearing," *Measurement*, vol. 131, pp. 400–411, Jan. 2019.

[4] F. Li, R. Li, L. Tian, L. Chen, and J. Liu, "Data-driven time-frequency analysis method based on variational mode decomposition and its application to gear fault diagnosis in variable working conditions," *Mech. Syst. Signal Process.*, vol. 116, pp. 462–479, Feb. 2019.

[5] Y. Tian, Z. Wang, and C. Lu, "Self-adaptive bearing fault diagnosis based on permutation entropy and manifold-based dynamic time warping," *Mech. Syst. Signal Process.*, vol. 114, pp. 658–673, Jan. 2019.

[6] J. R. Wang, S. M. Li, Z. H. An, X. X. Jiang, W. W. Qian, and S. S. Ji, "Batch-normalized deep neural networks for achieving fast intelligent fault diagnosis of machines, *Neurocomputing*, vol. 329, pp. 53–65, Feb. 2019.

[7] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, Jun. 2002.

[8] D. Elreedy and A. F. Atiya, "A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance," *Inf. Sci.*, vol. 505, pp. 32–64, Dec. 2019.

[9] G. Douzas, F. Bacao, and F. Last, "Improving imbalanced learning through a heuristic oversampling method based on K-means and SMOTE," *Inf. Sci.*, vol. 465, pp. 1–20, Oct. 2018.

[10] S. Maldonado, J. López, and C. Vairetti, "An alternative SMOTE oversampling strategy for high-dimensional datasets," *Appl. Soft Comput.*, vol. 76, pp. 380–389, Mar. 2019.

[11] Y. Zhang, X. Li, L. Gao, L. Wang, and L. Wen, "Imbalanced data fault diagnosis of rotating machinery using synthetic oversampling and feature learning," *J. Manuf. Syst.*, vol. 48, pp. 34–50, Jul. 2018.

[12] W. Mao, L. He, Y. Yan, and J. Wang, "Online sequential prediction of bearings imbalanced fault diagnosis by extreme learning machine," *Mech. Syst. Signal Process.*, vol. 83, pp. 450–473, Jan. 2017.

[13] L. Duan, M. Xie, T. Bai, and J. Wang, "A new support vector data description method for machinery fault diagnosis with unbalanced datasets," *Expert Syst. Appl.*, vol. 64, pp. 239–246, Dec. 2016.

[14] F. Jia, Y. Lei, N. Lu, and S. Xing, "Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization," *Mech. Syst. Signal Process.*, vol. 110, pp. 349–367, Sep. 2018.

[15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[16] Y. Yuan, C. Tian, and X. Lu, "Auxiliary loss multimodal GRU model in audio-visual speech recognition," *IEEE Access*, vol. 6, pp. 5573–5583, 2018.

[17] R. Togo, T. Ogawa, and M. Haseyama, "Synthetic gastritis image generation via loss function-based conditional PGGAN," *IEEE Access*, vol. 7, pp. 87448–87457, 2019.

[18] Y. Li, Q. Pan, S. Wang, T. Yang, and E. Cambria, "A generative model for category text generation," *Inf. Sci.*, vol. 450, pp. 301–315, Jun. 2018.

[19] X. Li, W. Zhang, and Q. Ding, "Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks," *IEEE Trans. Ind. Electron.*, vol. 66, no. 7, pp. 5525–5534, Jul. 2019.

[20] W. Mao, Y. Liu, L. Ding, and Y. Li, "Imbalanced fault diagnosis of rolling bearing based on generative adversarial network: A comparative study," *IEEE Access*, vol. 7, pp. 9515–9530, 2019.

[21] J. Wang, S. Li, B. Han, Z. An, H. Bao, and S. Ji, "Generalization of deep neural networks for imbalanced fault classification of machinery using generative adversarial networks," *IEEE Access*, vol. 7, pp. 111168–111180, 2019.

[22] X. Gao, F. Deng, and X. Yue, "Data augmentation in fault diagnosis based on the Wasserstein generative adversarial network with gradient penalty," *Neurocomputing*, vol. 396, pp. 487–494, Jul. 2020.

[23] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*. [Online]. Available: http://arxiv.org/abs/1701.07875

[24] S. Shao, P. Wang, and R. Yan, "Generative adversarial networks for data augmentation in machine fault diagnosis," *Comput. Ind.*, vol. 106, pp. 85–93, Apr. 2019.

[25] F. Zhou, S. Yang, H. Fujita, D. Chen, and C. Wen, "Deep learning fault diagnosis method based on global optimization GAN for unbalanced data," *Knowl.-Based Syst.*, vol. 187, Jan. 2020, Art. no. 104837.

[26] Q. Li, L. Chen, C. Shen, B. Yang, and Z. Zhu, "Enhanced generative adversarial networks for fault diagnosis of rotating machinery with imbalanced data," *Meas. Sci. Technol.*, vol. 30, no. 11, Nov. 2019, Art. no. 115005.
[27] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," Mar. 2017, *arXiv:1704.00028*. [Online]. Available: http://arxiv.org/abs/1704.00028
[28] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," Jun. 2017, *arXiv:1706.08500*. [Online]. Available: http://arxiv.org/abs/1706.08500
[29] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," Feb. 2018, *arXiv:1802.05957*. [Online]. Available: http://arxiv.org/abs/1802.05957
[30] Y. Ding, L. Ma, J. Ma, C. Wang, and C. Lu, "A generative adversarial network-based intelligent fault diagnosis method for rotating machinery under small sample size conditions," *IEEE Access*, vol. 7, pp. 149736–149749, 2019.
[31] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 1322–1328.
[32] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," Jun. 2016, *arXiv:1606.03498*. [Online]. Available: http://arxiv.org/abs/1606.03498

HONGLIANG ZHANG received the Ph.D. degree in industrial engineering from Tianjin University, Tianjin, China, in 2010. He is currently an Associate Professor with the Anhui University of Technology, China. His research interests include fault diagnosis, intelligent scheduling, and machine learning.

RUI WANG received the B.S. degree in information management and information system from the Anhui university of Technology, China, in 2018. He is currently pursuing the M.S. degree. His research interests include intelligent fault diagnosis, signal processing, and deep learning.

RUILIN PAN received the Ph.D. degree from the CIMS Center, Dalian University of Technology, Dalian, China, in 2010. He is currently a Professor with the Anhui University of Technology, China. His research interests include fault diagnosis, pattern recognition, and scheduling optimization.

HAIYANG PAN received the Ph.D. degree in mechanical engineering from Hunan University, Changsha, China, in 2019. He is currently a Teacher with the School of Mechanical Engineering, Anhui University of Technology. He has authored more than 30 articles. His research interests include intelligent fault diagnosis, signal processing, and pattern recognition.