IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 21, 2021, accepted January 31, 2021, date of publication February 10, 2021, date of current version February 19, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3058334

An Imbalanced Fault Diagnosis Method for Rolling Bearing Based on Semi-Supervised Conditional Generative Adversarial Network With Spectral Normalization

MINQIU XU AND YOUQING WANG, (Senior Member, IEEE)
College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China

Corresponding author: Youqing Wang (wang.youqing@ieee.org)

This work was supported in part by the National Natural Science Foundation of China under Grant 61822308, in part by the Shandong Province Natural Science Foundation under Grant JQ201812, and in part by the Program for Entrepreneurial and Innovative Leading Talents of Qingdao under Grant 19-3-2-4-zhc.

ABSTRACT In actual industrial applications, rolling bearings are under normal working conditions most of the time, and the fault data that can be collected are insufficient, so they are prone to data imbalance. Due to the high cost of labeling all fault data, most fault data are unlabeled. In this study, the Semi-supervised Conditional Generative Adversarial Network with Spectral Normalization (SN-SSCGAN) is proposed to solve these problems. Its core idea is to generate new samples with similar distribution by using partially labeled minority fault samples to balance the dataset. First, this method first applies wavelet transform to preprocess a vibration signal and obtain a time-frequency matrix. Second, the partially labeled time-frequency fault data are taken as the input of SN-SSCGAN, Nash equilibrium is achieved through adversarial training, and then data with similar distribution are generated. Lastly, the generated fault data are added to the dataset for balancing, and a convolutional neural network is used for fault diagnosis. The effectiveness of the proposed method is verified with comparative experiments in the CWRU bearing dataset. Results show that this method can generate high-quality samples and determine satisfactory results in bearing fault diagnosis when only a small number of labeled samples and the remaining unlabeled samples are used.

INDEX TERMS Generative adversarial network, imbalanced fault, fault diagnosis, wavelet transform, rolling bearing.

I. INTRODUCTION

Rolling bearing is one of the most frequently used parts in automation equipment. In industrial production, it is prone to damage because of its continuous operation in complex working environments, such as high load, variable working conditions, high temperature, and high pressure. Relevant data have shown that about 30% of mechanical faults are caused by rolling bearing faults [1]. The consequences caused by the fault of the rolling bearing destroy the automation system and may result in serious or even catastrophic accidents. For this reason, the fault diagnosis of rolling bearing is of great significance to improve the reliability of equipment and reduce economic losses.

In many studies, some data-driven methods for bearing fault diagnosis assume that the dataset is balanced, that is, the data collected and labeled under different working conditions are the same in quantity. However, rolling bearing mostly works in normal operating conditions, and the fault data that can be collected are insufficient. When the training data are imbalanced, the accuracy of the majority samples is relatively high, and the accuracy of the minority samples is low; it even has an ''undo'' effect on the minority [2]. Moreover, labeling all fault data is more expensive, and most of the fault data are unlabeled [3]. Through the semi-supervised learning method, only a small number of labeled samples and the remaining unlabeled samples are needed to train the model,

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

which reduces the cost and is more in line with the actual situation.

In recent decades, various machine learning methods have been extensively studied. When the given data are limited, some methods, such as support vector machine (SVM), have a low classification accuracy [4].

Studies have been implemented to solve the problem of imbalanced data and insufficient labeled samples. To solve the imbalanced data problem, the synthetic minority oversampling technique (SMOTE) was proposed through which new samples between two minority samples are synthesized through linear interpolation. Han *et al.* [5] proposed an improved SMOTE method named Borderline-SMOTE, which only considers samples close to the classification boundary, and oversamples them to increase the minority samples at the boundary. He *et al.* [6] developed an adaptive synthetic sampling approach (ADASYN), which adaptively changes its weight in accordance with the distribution of minority samples. Douzas *et al.* [7] designed the K-means SMOTE algorithm, which involves K-means to perform SMOTE oversampling on minority samples with a high degree of clustering, avoid noise generation, and effectively improve data imbalance. However, most of these methods based on synthetic oversampling are generated through data interpolation, which is still insufficient in reflecting the potential distribution of actual data and cannot learn the features of data distribution. Many semi-supervised algorithms have emerged to address the problem of insufficient labeled data. Monroy *et al.* [8] proposed a semi-supervised method by integrating independent component analysis, Gaussian mixed model, and SVM. Jiang *et al.* [9] proposed semi-supervised fisher analysis named SSKMFA for bearing fault diagnosis. Li and Zhou [10] put forward a novel semi-supervised method called SWKC-GS for clustering the labeled and unlabeled data. Jiang *et al.* [11] proposed a semi-supervised hierarchical sparse neural network for fault diagnosis. Therefore, these methods have good performance in fault diagnosis, but they cannot automatically learn the distribution features of samples.

With the continuous improvement of deep learning algorithms, it may become the most potential method to deal with the problem of data imbalance [12]. Goodfellow *et al.* [13] first proposed Generative Adversarial Network (GAN) in 2014; since then, GAN has one of the most popular research directions in the field of deep learning. As a generative model, it can generate new samples with similar distributions to original data and has been widely used in various fields, such as emotional speech [14], computer vision [15], super-resolution image [16]. On the basis of this idea, scholars proposed various improved algorithms for GAN. In order to solve the unstable factors of GAN in the training process, such as gradient explosion and gradient vanishing problem, the following improvement methods are proposed. Arjovsky *et al.* [17] proposed Wasserstein GAN (WGAN), which uses smooth Wasserstein distance instead of discontinuous JS divergence as the loss function of GAN, and forced the discriminator to be restricted to a 1-Lipschitz function space through weight clipping. Gulrajani *et al.* [18] proposed Wasserstein GANs with gradient penalty (WGAN-GP)and replaced weight clipping by directly limiting the gradient norm of the discriminator's output relative to its input, making the training progress more stable. Miyato *et al.* [19] proposed the Spectrally Normalized GAN (SNGAN), which transforms the training instability problem of GAN into a process of finding the maximum singular value of the weight matrix in the discriminator. In semi-supervised learning, the following improvements are proposed. Salimans *et al.* [20] proposed semi-supervised GAN (SSGAN), which improves the effectiveness of GAN for semi-supervised learning and introduced some training techniques to stabilize training. Odena *et al.* [21] proposed the Auxiliary Classifier GAN (ACGAN), which is improved on the basis of SSGAN. Sricharan *et al.* [22] proposed semi-supervised conditional GAN (SSCGAN) to further improve the performance of semi-supervised GAN. In recent studies, some GAN-based methods have been gradually applied to the field of fault diagnosis because of the ability to generate minority samples. Wang *et al.* [23] used a stacked denoising autoencoder as the structure of the discriminator to solve the noise problem and realize small samples of planetary gearbox fault diagnosis. Han *et al.* [24] introduced the adversarial ideas in GAN into convolutional neural networks to improve its generalization performance. Mao *et al.* [25] proposed a GAN-based fault diagnosis method for imbalanced data and conducted a detailed comparative study. Liang *et al.* [26] combined the wavelet transform and semi-supervised GAN for the fault diagnosis. Pan *et al.* [27] proposed a semi-supervised multiscale convolutional GAN framework, which can effectively identify bearing faults. Relevant research has been conducted on data imbalance and semi-supervised learning respectively. They have mainly focused on the following points: (1) labeled minority samples are used for training, and (2) a balanced dataset is used while some of the samples are labeled for training. However, in practical applications, the number of minority samples that can be collected is limited, labeled samples are expensive, and unlabeled data are insufficient. Therefore, studies should be performed to address the problems of data imbalance and insufficient labeled samples simultaneously.

In this study, a new fault diagnosis method SN-SSCGAN is proposed to solve the problem of data imbalance and insufficient labeled samples. Spectral normalization is combined with the improved semi-supervised conditional GAN for the fault diagnosis of bearing imbalanced data. Wavelet transform is adopted to extract the time-frequency features of vibration signals, and labeled and unlabeled samples are simultaneously used for training. The contributions of this study are summarized as follows:

(1) Considering the problems of data imbalance and insufficient labeled samples in practical applications, an improved semi-supervised conditional GAN is proposed to make full use of labeled data and unlabeled data to balance the dataset.

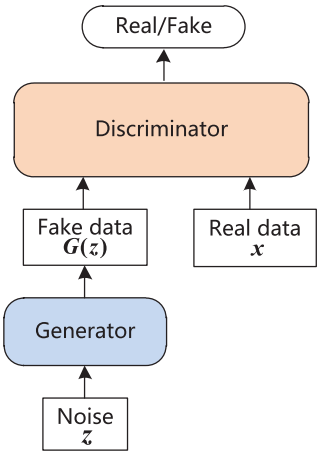

FIGURE 1. Structure of GAN.

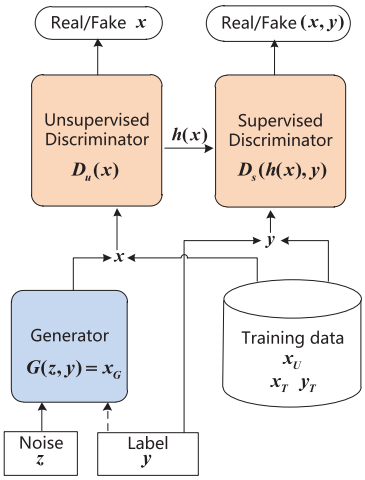

FIGURE 2. Structure of SSCGAN [22].

(2) The minority class fault samples are used to train the proposed SN-SSCGAN, and the balanced minority and majority classes are classified by the fault diagnosis classifier.

(3) Spectral normalization is used in the discriminator to normalize the weight matrix to fundamentally solve the problem of instability during training.

(4) The effectiveness of the proposed method is tested on the CWRU bearing dataset. The results show that the performance of this method is better than that of the three other comparison methods under the condition of limited labeled and unlabeled samples.

The remaining parts are organized as follows. Section II briefly presents the basic theories of GAN, SSCGAN, spectral normalization, and wavelet transform. Section III introduces the proposed method in detail. Section IV discusses the application of this method in the bearing dataset and its comparison with other methods. Section V summarizes the conclusions of this study.

II. RELATED WORKS

A. GAN

Inspired by binomial zero-sum game theory, Goodfellow *et al.* [13] proposed GAN, which is mainly composed of a discriminator D and a generator G, as shown in Fig. 1. The goal of the generator is to capture the potential distribution of real data $P_{data}(x)$ as much as possible to confuse the discriminator. The generator takes random noise z as an input to generate fake data $G(z)$. The goal of the discriminator is to determine the source (real data x or fake data $G(z)$) of the input data and output the corresponding probability, where 0 means completely fake data and 1 means completely real data. The objective function of GAN is expressed in Eq. (1).

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

Through adversarial learning, the parameters of the discriminator and the generator are updated. When the GAN is trained well, the distribution of the fake data generated by the generator is almost exactly the same as the distribution of the real data. The discriminator cannot distinguish the source of the data, and the output probability is 0.5.

B. SEMI-SUPERVISED CONDITIONAL GAN

Semi-supervised learning is widely used in GAN. Salimans *et al.* [20] proposed SSGAN, which involves the use of a discriminator to reconstruct label information, improve the quality of the generated data, and obtain better results on computer vision datasets. Odena *et al.* [21] proposed an ACGAN, which uses label information for training, reconstructs label information through the discriminator, and generates new samples according to the specified label. Sricharan *et al.* [22] proposed SSCGAN, which has better results than other semi-supervised GAN models. The structure of SSCGAN is shown as Fig. 2, where x_G represents the data generated by the generator, x_U represents the unlabeled data, x_T represents the labeled data, and y_T represents the label corresponding to x_T.

In SSCGAN, a stacked discriminator architecture composed of unsupervised part D_u and supervised part D_s is considered, where D_u is responsible for distinguishing the true and false of unlabeled data x, and D_s is responsible for distinguishing the true and false of labeled data (x, y). In the model, the unsupervised discriminator D_u yields the intermediate layers output $h(x)$, and the label y is subsequently appended to $h(x)$ and fed to the supervised discriminator D_s. By providing the feature $h(x)$ learned by D_u to D_s, the advantage over directly providing data x to D_s is that D_s cannot overfit to the minority labeled samples. Instead, it must rely on the features of the population to uncover the dependency between x and y.

During the training process, whether the generator adds label y is determined according to whether the training data have a label, and the label y input to the generator is consistent with the label y_T of the input D_s. When the training is completed, the label y in $G(z, y)$ is arbitrarily specified. The stacked discriminator and the generator apply batch normalization [28] to stabilize training, and input concat is used to add labels.

The loss function of the generator and discriminator pair is shown in the following equations. When the generator is fixed, the discriminator loss function, including the supervised part L_D^s and the unsupervised part L_D^u, should be minimized. When the discriminator is fixed, the parameters of the generator are updated by minimizing L_G. The α controls the effect of the labeled data on the unlabeled part. The α is set to 0 when the input data are unlabeled; otherwise, it is 1.

$$L_D = L_D^s + L_D^u \tag{2}$$

$$L_D^s = -\{E_{x,y\sim P_{data}(x,y)}[\log D_s(x,y)] + E_{z,y\sim P_z(z,y)}[\log(1 - D_s(G(z,y)))]\} \tag{3}$$

$$L_D^u = -\{E_{x\sim P_{data}(x)}[\log D_u(x)] + E_{z\sim P_z(z)}[\log(1 - D_u(G(z)))]\} \tag{4}$$

$$L_G = -\{E_{z\sim P_z(z)}[\log(D_u(G(z)))] + \alpha E_{z,y\sim P_z(z,y)}[\log(D_s(G(z,y)))]\} \tag{5}$$

C. SPECTRAL NORMALIZATION

Miyato *et al.* [19] proposed a new weight normalization method called spectral normalization to solve the problem of instability of the discriminator during training. It controls the Lipschitz constant by limiting the spectral norm (L2 matrix norm) of the weight matrix of each layer f in the discriminator D.

For the linear layer $f(x) = Wx$, its Lipschitz norm is shown in Eq. (6) according to the definition, where $\sigma(W)$ represents the L2 matrix norm of matrix W, which is also equal to the maximum singular value of W.

$$\|f\|_{Lip} = \sup_x \sigma(\nabla f(x)) = \sup_x \sigma(W) = \sigma(W) \tag{6}$$

$$\sigma(W) = \max_{x\neq 0} \frac{\|Wx\|_2}{\|x\|_2} = \max_{\|x\|_2\leq 1} \|Wx\|_2 \tag{7}$$

If the Lipschitz norm of the activation function a selected for each layer is 1 (such as ReLU), according to norm compatibility, the boundary of the Lipschitz norm in discriminator D can be obtained, as presented in inequality (8), where L is the layer's number of D.

$$\|D\|_{Lip} \leq \prod_{l=1}^{L+1} \left\|W^l x_{l-1}\right\|_{Lip} = \prod_{l=1}^{L+1} \sigma(W^l) \tag{8}$$

Therefore, a spectral normalization method is needed to ensure that $\sigma(W^l)$ is always equal to 1. Spectral normalization is shown in Eq. (9).

$$\bar{W}_{SN}(W) = \frac{W}{\sigma(W)} \tag{9}$$

Eq. (9) is used to normalize the weight matrix W^l of each layer, thereby obtaining $\sigma(\bar{W}_{SN}(W^l)) = 1$ so that D can satisfy the 1-Lipschitz constraint. The training instability problem of the discriminator is transformed into the problem of obtaining the maximum singular value $\sigma(W^l)$. $\sigma(W^l)$ can be determined by applying power iteration method [29]. The specific process is as follows, where $\tilde{u}_l$ represents the random initialized vector of each layer of weight.

$$\tilde{v}_l \leftarrow (W^l)^T \tilde{u}_l / \left\|(W^l)^T \tilde{u}_l\right\|_2 \tag{10}$$

$$\tilde{u}_l \leftarrow W^l \tilde{v}_l / \left\|W^l \tilde{v}_l\right\|_2 \tag{11}$$

$$\sigma(W^l) = \tilde{u}_l^T W^l \tilde{v}_l \tag{12}$$

D. CONTINUOUS WAVELET TRANSFORM

Wavelet transform, as a recognized time-frequency domain analysis method, has a powerful time-frequency feature extraction ability and has been widely used in the fault diagnosis of rolling bearings [30]–[32]. In comparison with the one-dimensional frequency domain signal obtained by Fourier transform, wavelet transform can obtain a two-dimensional time-frequency image, which contains more usable information. Wavelet transform has a higher frequency resolution and a lower time resolution in the low frequency range, and vice versa [26]. In our study, continuous wavelet transform (CWT) is implemented to preprocess the vibration signal as the input data of the discriminator. The CWT provides continuous translating and scaling of the wavelet basis function, as defined in Eq. (13),

$$W_\psi(a,\tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t-\tau}{a}\right)dt \tag{13}$$

where $x(t)$ is the input vibration signal, ψ is the wavelet basis function (WBF), ψ^* represents the complex conjugate of ψ, a is the scaling factor, which control the expansion and contraction of WBF, and τ is the translation factor, which determines the position of WBF [33].

Morlet WBF is chosen as the basis function of CWT with an explicit analytic equation. It is symmetric, has good smoothness, and has a limited support length. Therefore, compared with Db WBF, it is more reasonable to use Morlet WBF to extract bearing fault features in wavelet transform [34].

III. PROPOSED METHOD

A noval fault diagnosis method called SN-SSCGAN is proposed to solve the problems of data imbalance and insufficient labeled samples. In this method, spectral normalization is combined with an improved semi-supervised conditional GAN, and limited labeled minority fault samples are used to generate new samples with a similar distribution. The balanced minority fault samples and majority normal samples are applied to the CNN classifier for fault diagnosis. Fig. 3 shows the flow chart of this method, which includes three parts: data preprocessing, data generation and fault diagnosis.

A. DATA PREPROCESSING

In this method, the CWT is used to transform vibration signals into time-frequency images and normalize them to the interval [-1,1]. The minority data are extracted to train SN-SSCGAN, and the majority data are used for fault diagnosis classifier.

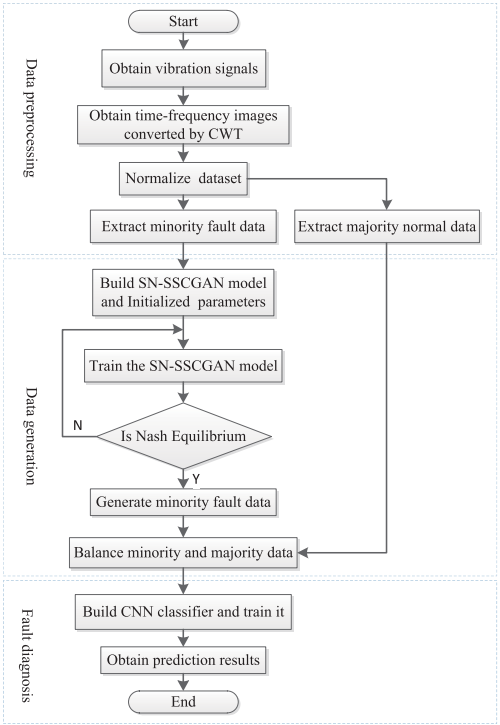

FIGURE 3. Flow chart of the proposed method.

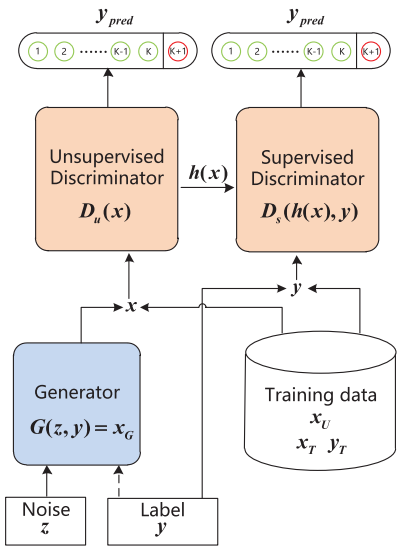

FIGURE 4. SN-SSCGAN framework.

B. DATA GENERATION

1) SN-SSCGAN FRAMEWORK

The method proposed in this study is improved on the basis of SSCGAN by adding label nodes to the output of D_u and D_s in the discriminator and applying spectral normalization in the stacked discriminator. The generator generates new samples by inputting random noise z. In terms of the presence of the label of the sample input to the stack discriminator, this study determines whether the generator adds a label y. The label y input to the generator is consistent with the label y_T input to D_s. The framework of SN-SSCGAN is shown in Fig. 4, where x_G represents the data generated by the generator, x_U represents the unlabeled data, x_T represents the labeled data, and y_T represents the label corresponding to x_T.

2) GENERATOR AND DISCRIMINATOR DESIGN

Convolution and deconvolution have an excellent performance in image feature extraction and reconstruction, so these two structures are used to build the discriminator and the generator.

In the generator, 128-dimensional noise data are input to a fully connected (FC) layer with 16384 nodes. The output of the FC layer is converted into $4 \times 4 \times 1024$ (width×height×channel number) images through a simple reshape function.

Then, six deconvolution layers is used for processing, where the size of filters is 5×5, and the stride is 2. The first five deconvolution layers have 1024, 512, 256, 128, and 64 filters respectively, and the activation function is ReLU. The number of filters in the last deconvolution layer is 1, the activation function uses tanh, and the final output is matrices with a value range of [-1,1] and a size of 64×64.

When the discriminator is input with labeled data, each layer in the generator adopts conditional batch normalization (CBN) [35] to insert label y to generate a new sample of a specific label. On the contrary, the generator inputs unlabeled data and uses batch normalization (BN) instead of CBN. The generator architecture is shown in Fig. 5.

The stacked discriminator is divided into unsupervised discriminator D_u and supervised discriminator D_s. The input data of the stacked discriminator are matrices of 64×64 dimensions, and features are extracted through five convolutional layers. The size of the filters of each layer is 3×3, the stride is 2, the activation function is Leaky ReLU (abbreviated as LReLU), and the size of the output data is $4 \times 4 \times 1024$. Afterward, the middle layer output feature $h(x)$ of D_u is obtained through the average pooling and the FC layer with 512 neuron nodes.

In inputting the unlabeled data, the true/false and classification information is output through the FC layer with $K+1$ neuron nodes and the softmax layer. In the output $K+1$ type data $y_{pred} = \{y^1_{pred}, y^2_{pred}, \cdots, y^{K+1}_{pred}\}$, the first K type data $\{y^1_{pred}, y^2_{pred}, \cdots, y^K_{pred}\}$ represents the predicted label, and the $K+1$ type y^{K+1}_{pred} represents the true or false prediction. The proposed method only uses the true/false information at y^{K+1}_{pred} to update D_u and the generator G.

When the label data are used as input, the feature $h(x)$ is input to D_s, and the label y is inserted using the inner product method. Then the FC layer with $K+1$ neuron nodes and the softmax layer are used to output the true/false and classification information. The true/false information at y^{K+1}_{pred} and the label information at $\{y^1_{pred}, y^2_{pred}, \cdots, y^K_{pred}\}$ are applied to update the stacked discriminator D (including D_u and D_s) and the generator G.

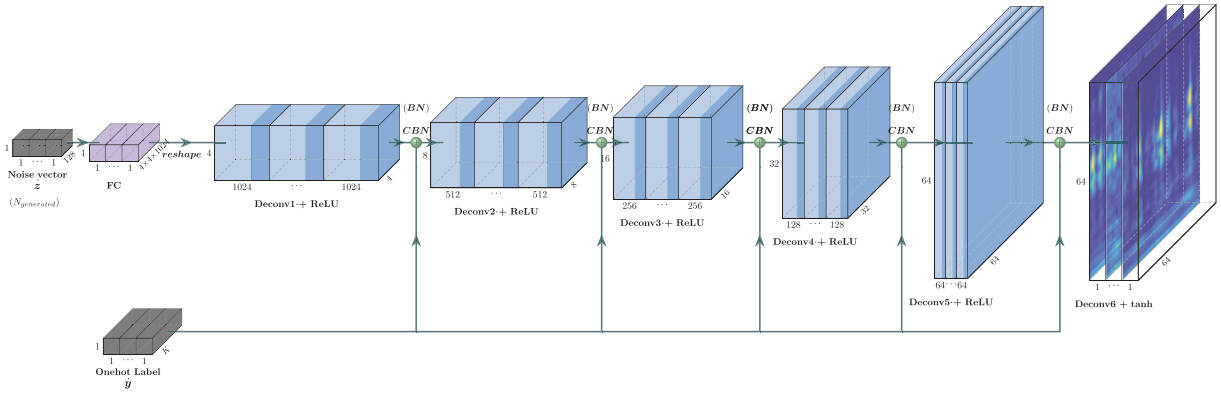

FIGURE 5. Architecture of the proposed generator (*CBN* indicates Conditional Batch Normalization).

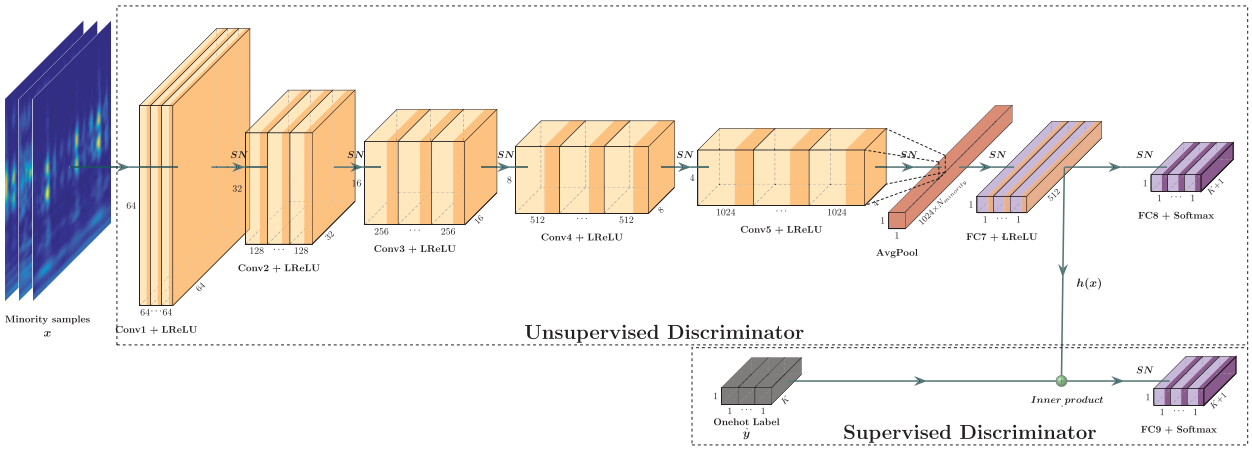

FIGURE 6. Architecture of the proposed discriminator (*SN* indicates Spectral Normalization).

Spectral normalization is applied to stabilize the training process between the layers of the stacked discriminator, s. The discriminator architecture is illustrated in Fig. 6.

3) MODEL TRAINING PROCESS

Model training is divided into a supervised part and an unsupervised part. When the labeled data are input, the random noise z is linked with the label y, and the fake data $G(z, y)$ generated by the generator are input to D_s through D_u with the real data (x, y). The classified data y_{pred} are output. In inputting the unlabeled data, the fake data $G(z)$ and the real data x are inputted into D_u, and the classified data y_{pred} are directly outputted. The generator and the discriminator are alternately trained with the mini-batch gradient descent until the Nash equilibrium is reached.

In the unsupervised part of this study, the loss function proposed in the geometric GAN [36] is adopted to replace the commonly used cross-entropy loss function. Geometric GAN uses support vector machine to separate hyperplane to maximize the margin between real and generated data. The geometric GAN is derived on the basis of geometric intuition similar to the derivation of the SVM compared with most of the existing approaches based on a statistical design criterion, which has more advantages. A large number of numerical experiments show that geometric GAN has less mode collapse and a more stable training process. This method is not described in detail here because of space limitation. The loss functions of the discriminator and the generator are shown in the following equations,

$$L_D = L_D^s + L_D^u \tag{14}$$

$$L_D^s = -E_{x,y\sim P_{data}(x,y)}[\log D_s(h(x), y)] \tag{15}$$

$$L_D^u = -\{E_{x\sim P_{data}(x)}[\max(0, 1 - D_u(x)] + E_{z\sim P_z(z)}[\max(0, 1 + D_u(G(z)))]\} \tag{16}$$

$$L_G = -\{\alpha E_{z\sim P_z(z)}[D_u(G(z))] + (1-\alpha)E_{z,y\sim P_z(z,y)}[D_s(G(z, y)]\} \tag{17}$$

where $D_s(h(x), y)$ corresponds to the first K label information of y_{pred}, $D_u(x)$ and $D_u(G(z))$ represent the $(K+1)$-th true/false information of y_{pred}. α controls the generator input data type. In inputting the unlabeled data, α is set to 1; otherwise, it is set to 0.

The training parameters are as follows. In one training epoch, the ratio of the training times of the stacked discriminator D and the generator G is 1. The size of the mini-batch

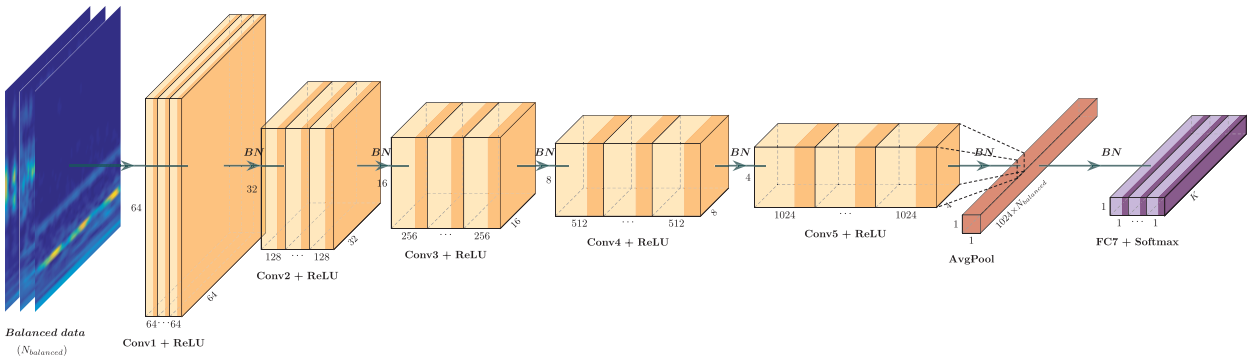

FIGURE 7. Architecture of the proposed fault diagnosis classifier (*BN* indicates Batch Normalization).

is 7. Adam is used as the optimizer with learning rates of G and D of 0.0001, and the model is trained for 300 epochs.

C. FAULT DIAGNOSIS

The fault diagnosis classifier used in this study is shown in Fig. 7. The input of the classifier consists of the balanced minority fault samples and majority normal samples. First, five convolutional layers are used, with 64, 128, 256, 512, and 1024 filters, with a filter size of 3 × 3 and a stride of 2. Then, an average pooling layer and a FC layer with K neuron nodes are used. Lastly, the softmax function is used for classification. ReLU is selected as the activation function between convolutional layers, and the mini-batch size is 50. The Adam optimizer is utilized with the learning rate of 0.0002 to train 100 epochs.

IV. EXPERIMENT

A. BEARING DATASET FROM CWRU

1) DATASET DESCRIPTION

In this study, the bearing dataset is provided by the Case Western Reserve University (CWRU) Bearing Data Center [37], which is a recognized benchmark dataset in the field of fault diagnosis. The test equipment consists of a 2 hp motor, a load motor, a torque transducer/encoder, a power meter and an electronic controller. An electro-discharge machining is used to create different degrees of single point faults on the ball, inner race and outer race of the bearing. The fault diameters are 0.007, 0.014, 0.021, 0.028 and 0.04 inches. The acceleration sensor is used to collect vibration signals along the vertical direction from the housing of the fan end bearing (FE), the drive end bearing (DE), and the base plate (BA). These vibration signals are used as a fault diagnosis dataset.

In this study, the DE vibration signals of different fault diameters are adopted to perform fault diagnosis on normal condition (NC), ball fault (B), inner race fault (IR) and outer race fault (OR). The sampling frequency is 12 kHz, the load is 0 hp, and the selected fault diameters are 0.007, 0.014, and 0.021 inches. Therefore, 10 types of data can be obtained, and they are labeled as NC, B007, B014, B021, IR007, IR014, IR021, OR007, OR014, and OR021.

2) DATA PREPROCESSING

Among the 10 types of the data collected, the majority class (NC) has about 244000 data points, and the minority class (B, IR, and OR) has about 121000 data points. The data points are divided with a sampling method similarly to that in a previous study [26], [38]. In this study, a sliding window method is used to divide the one-dimensional vibration signal. If the length of the sliding window is set too small, it will not be able to cover a sufficient length of the vibration signal, which will cause samples confusion and reduce diagnostic accuracy [39]. When the sampling frequency is 12 kHz and the motor speed is 1797 r/min, the number of data points corresponding to one revolution of the bearing is $N = 60 \div 1797$ (rpm)$\times 12000$ (frequency)≈ 400. Then, 1200 data points (three bearing rotation cycles) are used as the sliding window, and the sliding step length is 600 data points. The first 120600 data points of the minority class and the first 240600 data points of the majority class are intercepted. Through the sliding window, 200 samples of the minority class and 400 samples of the majority class can be obtained. The vibration signal in each sample is converted into a time-frequency image through CWT. Then, the logarithm of all data is taken to reduce the gap between different types of data and normalized to [-1,1]. The proposed method involves $64 \times 64 \times 1$ images as input, so all wavelet transform image sizes are converted to $64 \times 64 \times 1$. The time-frequency images obtained with wavelet transform are shown in Fig. 8, and the details of the division of training set and testing set are presented in Table 1.

According to Table 1, four training sets with different imbalanced ratios can be obtained, and new samples are generated by training the minority class through GAN. The balanced minority samples are combined with the majority samples and inputted into the fault diagnosis classifier for classification. A total of 600 test samples are used to test the classifier performance.

Some samples are randomly labeled from the minority training set, and the dataset is divided into labeled and unlabeled data to further study the performance difference between different semi-supervised GAN methods.

TABLE 1. The details of dataset.

Fault type	Label	Training set				Testing set
		20:1	10:1	5:1	2:1	
NC	1	280	280	280	280	60
B007	2	14	28	56	140	60
B014	3	14	28	56	140	60
B021	4	14	28	56	140	60
IR007	5	14	28	56	140	60
IR014	6	14	28	56	140	60
IR021	7	14	28	56	140	60
OR007	8	14	28	56	140	60
OR014	9	14	28	56	140	60
OR021	10	14	28	56	140	60

TABLE 2. Labeled data under different imbalance ratios.

Minority labeled rate	imbalanced ratio			
	20:1	10:1	5:1	2:1
0.25	28	63	126	315
0.5	63	126	252	630
0.75	91	189	378	945
1	126	252	504	1260

The labeling rates are 0.25, 0.5, 0.75, and 1. The details of the labeled data are shown in Table 2.

3) DATA GENERATION AND VISUALIZED FEATURE DISTRIBUTION

In this section, nine kinds of minority fault samples are used to train SN-SSCGAN, and new samples are generated to balance the dataset. The number of the generated samples is $N^i_{generated} = N_{Major}\text{-}N^i_{labeled}$, where i represents the i-th minority fault type. For example, at an imbalanced ratio of 2:1 and a labeled rate of 0.5, Fig. 9 illustrates the comparison between original and the generated samples. It can be seen that the generated samples learn almost the same features from the original samples. Furthermore, t-distributed stochastic neighbor embedding (t-SNE) [40] is adopted to visualize the feature distribution of these samples and intuitively understand original and generated samples. The dimension-reduced visualization results of original and generated samples are shown in Fig. 10.

B. EXPERIMENTAL RESULTS

The proposed SN-SSCGAN is compared with the improved semi-supervised GAN methods, namely, ACGAN [21] and SSCGAN [22], to demonstrate the advantages of the proposed method in generating data. The network architecture of these two methods is shown in Table 3 and Table 4. The training parameters are consistent with those of the proposed SN-SSCGAN. These methods are also compared with the traditional oversampling method SMOTE. The main idea of this study is to use SN-SSCGAN, SSCGAN, ACGAN, and SMOTE to generate new samples to balance the data, and then input them into the same fault diagnosis classifier for training, and test the quality of the generated data through the evaluation indices of the testing set.

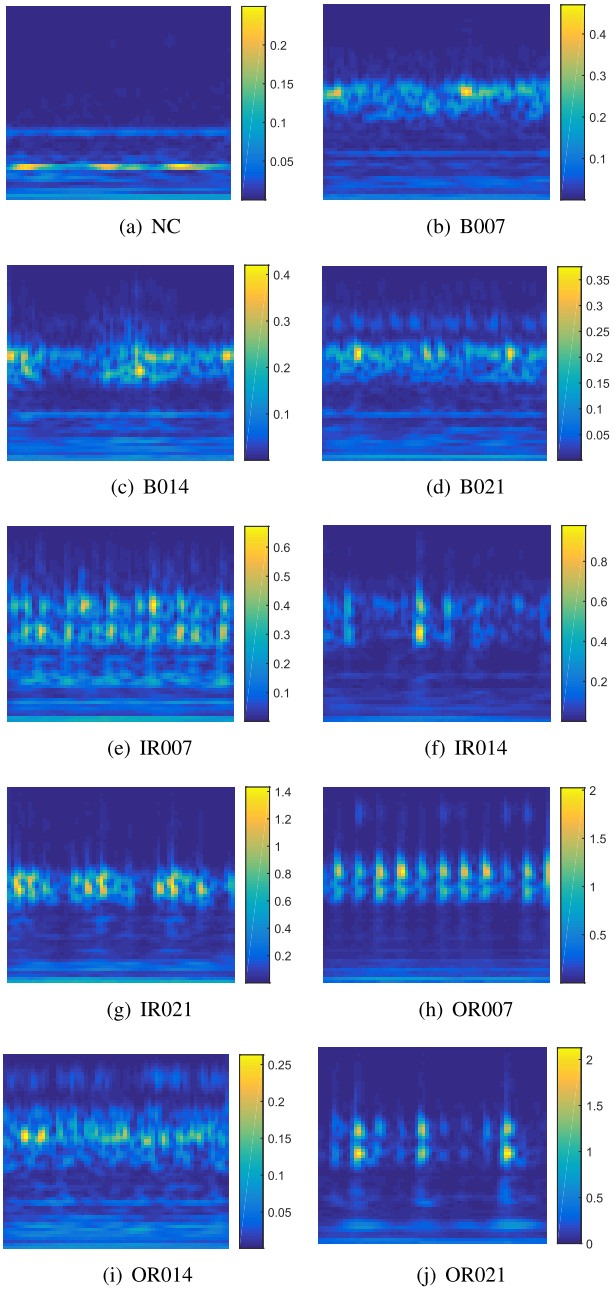

FIGURE 8. Continuous wavelet transform images of the 10 types of samples in the bearing dataset.

For a comprehensive comparison, the four commonly used evaluation indices for imbalanced data are as follows: F1-measure [41], G-mean [42], AUC value, and receiver operating characteristic (ROC) curve.

The definitions of F1-measure and G-mean are shown in Eqs. (18) and (19), where T/F represents the actual true/false sample, and P/N represents the predicted positive/negative sample, which can be combined into 4 situations: TP, TN, FP, and FN. The larger these two values are, the better the performance of the fault diagnosis classifier is, that is, the more accurate the generated data are.

$$F1 = \frac{2 \times TP}{2TP + FN + FP} \tag{18}$$

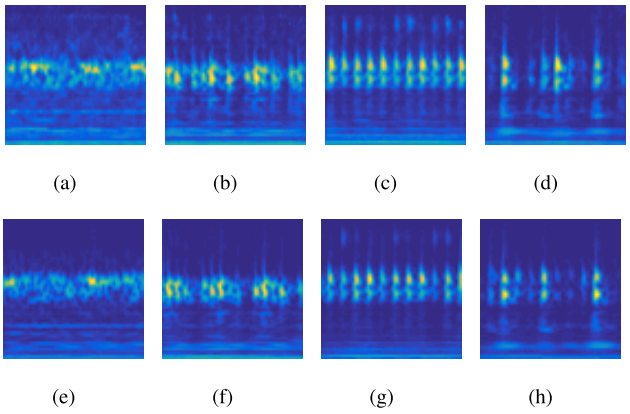

FIGURE 9. Comparison of wavelet transform images of generated samples and original samples: (a) generated B007; (b) generated IR021; (c) generated OR007; (d) generated OR021; (e) original B007; (f) original IR021; (g) original OR007; (h) original OR021.

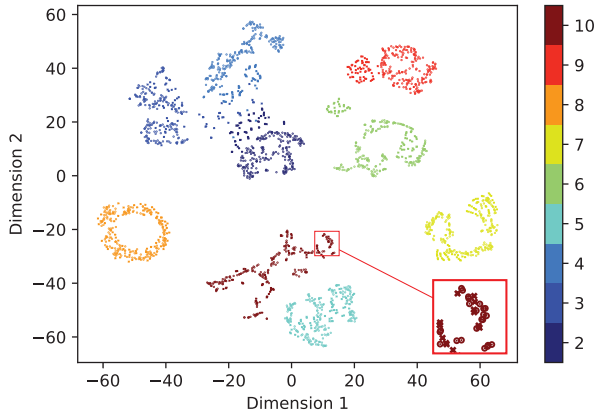

FIGURE 10. Visualized feature distribution (crosses indicate original testing samples, dots indicate generated samples).

$$G-mean=\sqrt{\frac{TP}{TP+FN}\times\frac{TN}{TN+FP}} \tag{19}$$

The ROC is a curve with a false positive rate (*FPR*) and true positive rate (*TPR*) as the abscissa and the ordinate, respectively. *FPR* and *TPR* are defined as Eqs. (20) and (21), and the AUC value is the area under the ROC curve. The closer the ROC curve is to the upper left corner, that is, the closer the AUC value is to 1, indicating that the higher the quality of the generated samples is higher.

$$FPR=\frac{FP}{FP+TN} \tag{20}$$

$$TPR=\frac{TP}{TP+FN} \tag{21}$$

A total of 16 sets of data with different imbalanced ratios and labeled rates are used for training, and the final comparison results are obtained. All these methods are repeated 10 trials and the mean value is taken to obtain the final comparison result to make the results more convincing and reduce the influence of randomness. With the space limitation, the imbalance ratio is 20:1, and the labeling rate is 0.25 as a representative. The experiment is repeated 10 times to obtain the comparison results of different methods, as shown in Fig. 11. The F1-measure index of the method proposed in this study is superior to other methods. In other cases, the comparison results of the F1-measure index, G-mean index, and the AUC index are shown in Fig. 12, 13, and 14, respectively. The imbalanced ratio of 20:1 and the labeled rate of 0.25 are taken as an example to show the comparison results of the ROC curves and make an obvious comparison (Fig. 15).

TABLE 3. Network architecture of SSCGAN.

Layer	Out channels	Filters/Strides	Output
Discriminator			
Input	-	-	(64,64,1)
Conv(CBN/LReLU)	64	(3, 3) / 1	(64, 64, 64)
Conv(CBN/LReLU)	128	(3, 3) / 1	(32, 32, 128)
Conv(CBN/LReLU)	256	(3, 3) / 1	(16, 16, 256)
Conv(CBN/LReLU)	512	(3, 3) / 1	(8, 8, 512)
Conv(CBN/LReLU)	1024	(3, 3) / 1	(4, 4, 1024)
Average Pooling	-	-	1024
FC(LReLU)	-	-	512
FC(Sigmoid)	-	-	1
Generator			
Input	-	-	128
FC	-	-	4×4×1024
Deconv(CBN/ReLU)	1024	(5, 5) / 1	(4, 4, 1024)
Deconv(CBN/ReLU)	512	(5, 5) / 1	(8, 8, 512)
Deconv(CBN/ReLU)	256	(5, 5) / 1	(16, 16, 256)
Deconv(CBN/ReLU)	128	(5, 5) / 1	(32, 32, 128)
Deconv(CBN/ReLU)	64	(5, 5) / 1	(64, 64, 64)
Deconv(CBN/tanh)	1	(5, 5) / 1	(64, 64, 1)

TABLE 4. Network architecture of ACGAN.

Layer	Out channels	Filters/Strides	Output
Discriminator			
Input	-	-	(64,64,1)
Conv(CBN/LReLU)	64	(3, 3) / 1	(64, 64, 64)
Conv(CBN/LReLU)	128	(3, 3) / 1	(32, 32, 128)
Conv(CBN/LReLU)	256	(3, 3) / 1	(16, 16, 256)
Conv(CBN/LReLU)	512	(3, 3) / 1	(8, 8, 512)
Conv(CBN/LReLU)	1024	(3, 3) / 1	(4, 4, 1024)
Average Pooling	-	-	1024
FC(LReLU)	-	-	512
FC(Sigmoid)	-	-	1, 9
Generator			
Input	-	-	128
FC	-	-	4×4×1024
Deconv(CBN/ReLU)	1024	(5, 5) / 1	(4, 4, 1024)
Deconv(CBN/ReLU)	512	(5, 5) / 1	(8, 8, 512)
Deconv(CBN/ReLU)	256	(5, 5) / 1	(16, 16, 256)
Deconv(CBN/ReLU)	128	(5, 5) / 1	(32, 32, 128)
Deconv(CBN/ReLU)	64	(5, 5) / 1	(64, 64, 64)
Deconv(CBN/ReLU)	1	(5, 5) / 1	(64, 64, 1)

C. RESULTS AND DISCUSSIONS

In Fig. 12, when the imbalanced ratios are 2:1 and 5:1, the proposed method is basically the same as the results of SSC-GAN, ACGAN, and SMOTE because the training samples are sufficient, and the F1-measure is 1. The features of the samples generated by the four methods are consistent with those of the original samples. When the imbalanced ratio is 10:1, as the number of the decrease of labeled samples decreases, the F1-measure index gradually decreases. In com-

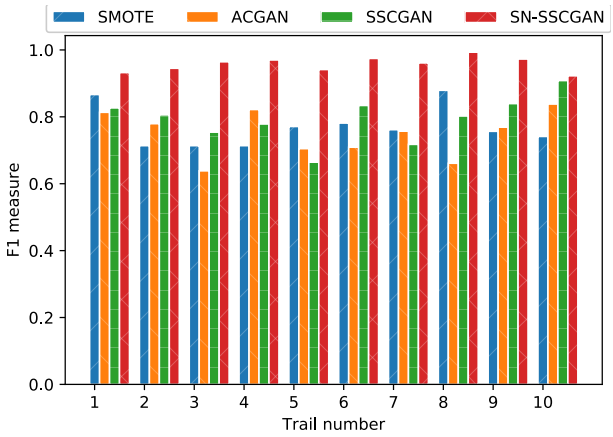

FIGURE 11. F1 measure of the ten trials in testing set.

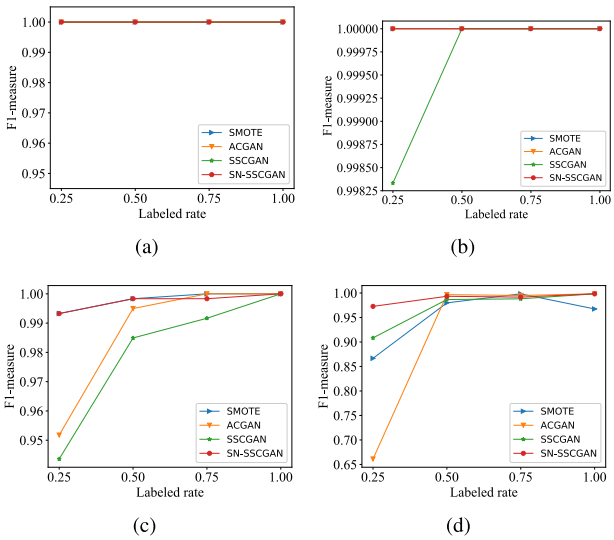

FIGURE 12. F1-measure in different methods with different imbalanced ratios and labeled rates: (a) 2:1; (b) 5:1; (c) 10:1; (d) 20:1.

parison with SSCGAN and ACGAN, the proposed method is significantly improved and not much different from the SMOTE method. When the imbalanced ratio is 20:1, the data are more imbalanced. As the number of the labeled samples decreases, the F1-measure indices of the four methods decrease more significantly. SMOTE is no longer advantageous for synthesizing data through linear interpolation, whereas the semi-supervised GAN can make full use of unlabeled samples to generate new samples. The F1-measure of SN-SSCGAN is higher than that of SSCGAN and ACGAN. The proposed SN-SSCGAN can generate samples that are closer to the original sample features than to the other methods under the condition of the less labeled and unlabeled data. Similarly, in Figs. 13 and 14, the proposed method is more effective than that other methods in terms of G-mean and AUC indices.

In Fig. 15, the ROC curve of the proposed method is closer to the upper left part, indicating that the feature distribution of the generated sample is closer to the real sample and consistent with the results of other evaluation indices.

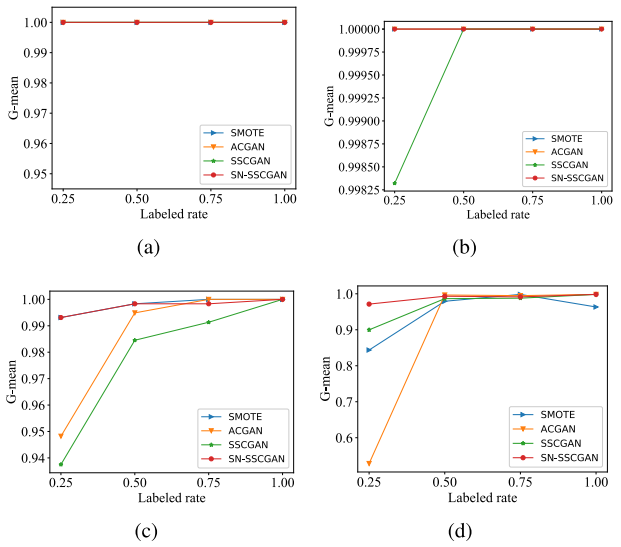

FIGURE 13. G-mean in different methods with different imbalanced ratios and labeled rates: (a) 2:1; (b) 5:1; (c) 10:1; (d) 20:1.

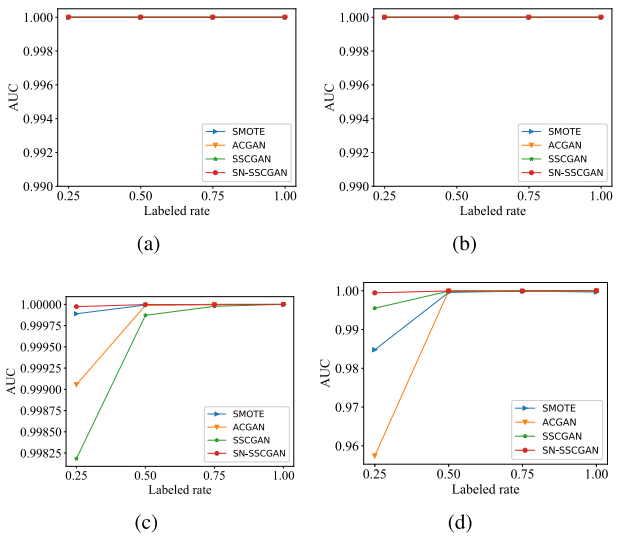

FIGURE 14. AUC value in different methods with different imbalanced ratios and labeled rates: (a) 2:1; (b) 5:1; (c) 10:1; (d) 20:1.

The confusion matrix with an imbalanced ratio of 20:1 and a labeled rate of 0.25 is shown in Fig. 16 to illustrate the classification of the four methods in detail. The abscissa of the confusion matrix is the predicted fault type, the ordinate is the real fault type, diagonal elements are the percentage of correct predictions, and the off-diagonal elements are the percentage of prediction errors. In the case of the imbalanced data, the confusion matrix can intuitively display the accuracy of the classification model corresponding to each category and indirectly reflect the quality of the generated data. In Fig. 16, the performance of the proposed SN-SSCGAN is the best among the four methods. The accuracy rates of other fault types are above 95%, but the accuracy rate of the B014 (label 3) fault is 80%. In Fig. 16(a), SMOTE misclassifies 1.7% samples of the NC working condition, 15% samples of B014, 63.3% samples of IR014, and 38.3% samples of OR021. In Fig. 16(b),

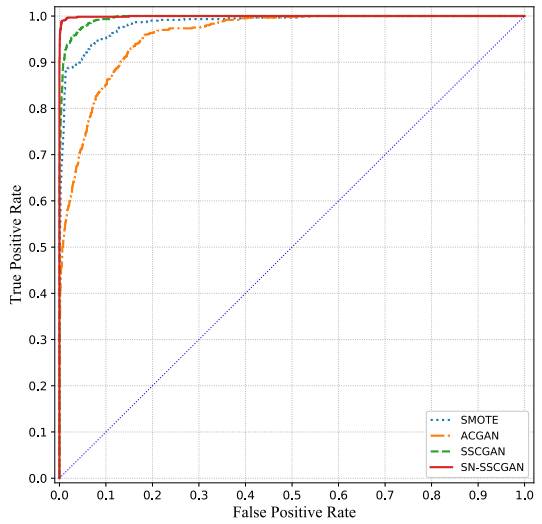

FIGURE 15. ROC curve with imbalanced ratio of 20:1 and labeled rate of 0.25.

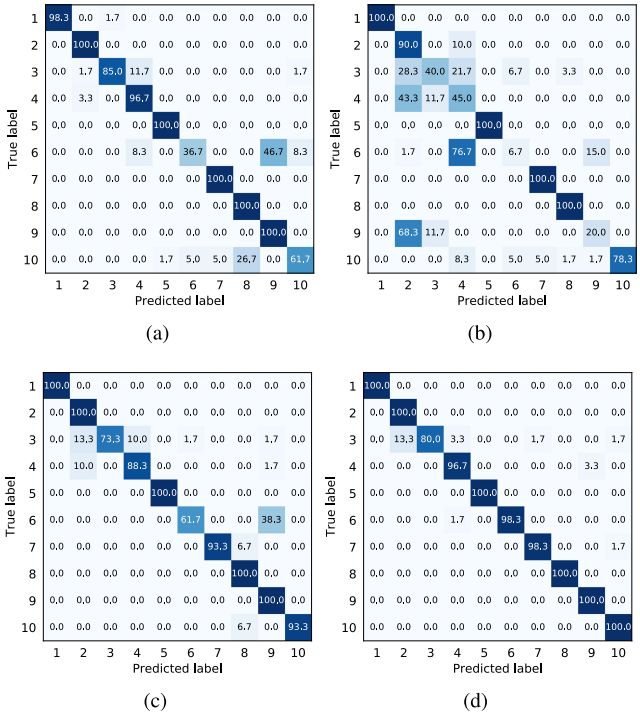

FIGURE 16. The confusion matrixes with imbalanced ratio of 20:1 and labeled rate of 0.25: (a) SMOTE; (b) ACGAN; (c) SSCGAN; (d) SN-SSCGAN.

ACGAN misclassifies 35% samples of B014, 30% samples of B021, 15% samples of IR007, 68.3% samples of IR014, 18.3% samples of IR021, 41.7% samples of OR007, and 3.3% samples of OR021. In Fig. 16(c), SSCGAN misclassifies 26.7% samples of B014, 11.7% samples of B021, 38.3% samples of IR014, 6.7% samples of IR021, and 6.7% samples of OR021. By contrast, the proposed SN-SSCGAN only misclassifies 20% of the B014 faults, 3.3% of the B021 faults, 1.7% of the IR014 faults, and 1.7% of the IR021 faults. The comparison result of confusion matrices shows that the proposed SN-SSCGAN method is superior to other methods in terms of the fault diagnosis of imbalanced data when labeled and unlabeled samples are fewer.

V. CONCLUSION

In practical applications, bearing fault diagnosis suffers from data imbalance. The collected fault data are limited, and the cost of labeling all fault data is high. The accuracy and cost of fault diagnosis must be considered simultaneously. In this study, the proposed imbalanced fault diagnosis method called SN-SSCGAN has a combination of spectral normalization with an improved semi-supervised conditional GAN. Partially labeled minority fault samples and sufficient unlabeled samples are utilized to generate new samples with similar distributions, balance the dataset, and perform fault diagnosis. Spectral normalization is applied to stabilize the adversarial training process. The analysis of the experimental results with other methods in the CWRU bearing dataset confirm that the proposed method has an excellent performance in the imbalanced fault diagnosis of rolling bearing. The obtained conclusions are as follows:

(1) The spectral normalization method can be used to normalize the weight matrix, limit the gradient of the discriminator to a certain range, and fundamentally solve the interference factors in the training process.

(2) The proposed method can be utilized to generate high-quality new samples, expand the imbalanced dataset, and further improve the performance of fault diagnosis.

(3) SN-SSCGAN has advantages over other methods (SMOTE, ACGAN, and SSCGAN) in terms of evaluation indices when few labeled and unlabeled samples are used.

The proposed method achieves excellent results in single fault diagnosis. In the following work, we will consider the problem of imbalanced composite fault diagnosis by using fewer labeled samples, combining the method with the latest deep neural network, and designing a new semi-supervised GAN to enhance the methods performance.

REFERENCES

[1] W. Xi, L. Bai, M. Hui, and Q. Wu, "A novel rolling bearing fault detect method based on empirical wavelet transform," in *Proc. 13th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, May 2018, pp. 2764–2768.

[2] Y. Tang, Y.-Q. Zhang, N. V. Chawla, and S. Krasser, "SVMs modeling for highly imbalanced classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 1, pp. 281–288, Feb. 2009.

[3] B. Yang, Y. Lei, F. Jia, and S. Xing, "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," *Mech. Syst. Signal Process.*, vol. 122, pp. 692–706, May 2019.

[4] 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.

[5] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-smote: A new over-sampling method in imbalanced data sets learning," in *Proc. Int. Conf. Intell. Comput.* Berlin, Germany: Springer, 2005, pp. 878–887.

[6] 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.

[7] 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.

[8] I. Monroy, R. Benitez, G. Escudero, and M. Graells, "A semi-supervised approach to fault diagnosis for chemical processes," *Comput. Chem. Eng.*, vol. 34, no. 5, pp. 631–642, May 2010.
[9] L. Jiang, J. Xuan, and T. Shi, "Feature extraction based on semi-supervised kernel marginal Fisher analysis and its application in bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 41, nos. 1–2, pp. 113–126, Dec. 2013.
[10] C. Li and J. Zhou, "Semi-supervised weighted kernel clustering based on gravitational search for fault diagnosis," *ISA Trans.*, vol. 53, no. 5, pp. 1534–1543, Sep. 2014.
[11] L. Jiang, Z. Ge, and Z. Song, "Semi-supervised fault classification based on dynamic sparse stacked auto-encoders model," *Chemometric Intell. Lab. Syst.*, vol. 168, pp. 72–83, Sep. 2017.
[12] H. Salehinejad, S. Valaee, T. Dowdell, E. Colak, and J. Barfett, "Generalization of deep neural networks for chest pathology classification in X-Rays using generative adversarial networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 990–994.
[13] 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.
[14] J. Chang and S. Scherer, "Learning representations of emotional speech with deep convolutional generative adversarial networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 2746–2750.
[15] X. Liang, Z. Hu, H. Zhang, C. Gan, and E. P. Xing, "Recurrent topic-transition GAN for visual paragraph generation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3362–3371.
[16] V. Das, S. Dandapat, and P. K. Bora, "Unsupervised super-resolution of OCT images using generative adversarial network for improved age-related macular degeneration diagnosis," *IEEE Sensors J.*, vol. 20, no. 15, pp. 8746–8756, Aug. 2020.
[17] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*. [Online]. Available: http://arxiv.org/abs/1701.07875
[18] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.
[19] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," 2018, *arXiv:1802.05957*. [Online]. Available: http://arxiv.org/abs/1802.05957
[20] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2234–2242.
[21] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2642–2651.
[22] K. Sricharan, R. Bala, M. Shreve, H. Ding, K. Saketh, and J. Sun, "Semi-supervised conditional GANs," 2017, *arXiv:1708.05789*. [Online]. Available: http://arxiv.org/abs/1708.05789
[23] 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, Oct. 2018.
[24] T. Han, C. Liu, W. Yang, and D. Jiang, "A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults," *Knowl.-Based Syst.*, vol. 165, pp. 474–487, Feb. 2019.
[25] 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.
[26] P. Liang, C. Deng, J. Wu, G. Li, Z. Yang, and Y. Wang, "Intelligent fault diagnosis via semisupervised generative adversarial nets and wavelet transform," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4659–4671, Jul. 2020.
[27] T. Pan, J. Chen, J. Xie, Y. Chang, and Z. Zhou, "Intelligent fault identification for industrial automation system via multi-scale convolutional generative adversarial network with partially labeled samples," *ISA Transactions*, vol. 101, pp. 379–389, Jun. 2020.
[28] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167
[29] G. H. Golub and H. A. van der Vorst, "Eigenvalue computation in the 20th century," *J. Comput. Appl. Math.*, vol. 123, nos. 1–2, pp. 35–65, Nov. 2000.
[30] X. Lou and K. A. Loparo, "Bearing fault diagnosis based on wavelet transform and fuzzy inference," *Mech. Syst. Signal Process.*, vol. 18, no. 5, pp. 1077–1095, Sep. 2004.
[31] Z. K. Peng, P. W. Tse, and F. L. Chu, "A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing," *Mech. Syst. Signal Process.*, vol. 19, pp. 974–988, Sep. 2005.
[32] P. K. Kankar, S. C. Sharma, and S. P. Harsha, "Rolling element bearing fault diagnosis using wavelet transform," *Neurocomputing*, vol. 74, no. 10, pp. 1638–1645, May 2011.
[33] L. Huang and J. Wang, "Forecasting energy fluctuation model by wavelet decomposition and stochastic recurrent wavelet neural network," *Neurocomputing*, vol. 309, pp. 70–82, Oct. 2018.
[34] P. Liang, C. Deng, J. Wu, and Z. Yang, "Intelligent fault diagnosis of rotating machinery via wavelet transform, generative adversarial nets and convolutional neural network," *Measurement*, vol. 159, Jul. 2020, Art. no. 107768.
[35] T. Miyato and M. Koyama, "CGANs with projection discriminator," 2018, *arXiv:1802.05637*. [Online]. Available: http://arxiv.org/abs/1802.05637
[36] J. H. Lim and J. C. Ye, "Geometric GAN," 2017, *arXiv:1705.02894*. [Online]. Available: http://arxiv.org/abs/1705.02894
[37] *The Case Western Reserve University Bearing Data Center Website*. Accessed: Jun. 10, 2020. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/home
[38] Y. Gao, F. Villecco, M. Li, and W. Song, "Multi-scale permutation entropy based on improved LMD and HMM for rolling bearing diagnosis," *Entropy*, vol. 19, no. 4, p. 176, Apr. 2017.
[39] 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.
[40] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[41] Z. C. Lipton, C. Elkan, and B. Naryanaswamy, "Optimal thresholding of classifiers to maximize F1 measure," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Berlin, Germany: Springer, 2014, pp. 225–239.
[42] 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.

MINQIU XU received the B.S. degree in automation from the Weifang University of Science and Technology, Weifang, China, in 2018. He is currently pursuing the master's degree in control engineering with the Shandong University of Science and Technology. His research interests include deep learning and its application in the field of bearing fault diagnosis.

YOUQING WANG (Senior Member, IEEE) received the B.S. degree from Shandong University, Jinan, China, in 2003, and the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2008.

From 2006 to 2007, he was a Research Assistant with the Hong Kong University of Science and Technology, Hong Kong. From 2008 to 2010, he was a Senior Investigator with the University of California, Santa Barbara, CA, USA. He worked as a Visiting Professor with the University of Alberta, Edmonton, AB, Canada, in 2015. He is currently a Professor with the Shandong University of Science and Technology. His research interests include fault-tolerant control, state monitoring, modeling and control of biomedical processes, such as artificial pancreas systems, and iterative learning controls. He was a recipient of several research awards, including the NSFC Excellent Young Scientists Fund, the Journal of Process Control Survey Paper Prize, and the ADCHEM2015 Young Author Prize.