

## Research Article

# Intelligent Deep Adversarial Network Fault Diagnosis Method Using Semisupervised Learning

Juan Xu <sup>1,2</sup>, Yongfang Shi,<sup>1</sup> Lei Shi,<sup>1</sup> Zihui Ren,<sup>3</sup> and Yang Lu <sup>1</sup>

<sup>1</sup>School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230009, China

<sup>2</sup>Anhui Fuhuagang Steel Structure Co., Ltd., Hefei 238076, China

<sup>3</sup>Anhui Fuhuagang Technology Co., Ltd., Hefei 230088, China

Correspondence should be addressed to Juan Xu; xujuan@hfut.edu.cn

Received 17 December 2019; Revised 3 March 2020; Accepted 23 April 2020; Published 22 May 2020

Academic Editor: Kai Zhang

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

In recent years, deep learning has become a popular issue in the intelligent fault diagnosis of industrial equipment. Under practical working conditions, although the collected vibration data are of large capacity, most of the vibration data are not labeled. Collecting and labeling sufficient fault data for each condition are unrealistic. Therefore, constructing a reliable fault diagnosis model with a small amount of labeled vibration data is a significant problem. In this paper, the vibration time-domain signal of the fault bearing is transformed into a 2-dimensional image by wavelet transform to obtain the time-frequency domain information of the original data. A deep adversarial convolutional neural network based on semisupervised learning is proposed. A large amount of fake data generated by the generator and unlabeled true vibration data are used in the discriminator to learn the overall distribution of data by judging the authenticity of the input. Three regular terms for different loss functions are designed to constrain the parameters of the discriminator to improve the learning ability of the model. The proposed method is validated by two bearing fault diagnosis cases. The experiment results show that the proposed method has higher diagnostic accuracy than traditional deep models on multigroup small datasets of different labels capacities. The proposed method provides a new solution to the fault diagnosis problem with large vibration data but few labels.

## 1. Introduction

In complex industrial systems, the study of advanced methods of mechanical fault diagnosis is an important part of ensuring the safety of equipment [1, 2]. Deep learning (DL) theory has become a popular issue in the field of data-driven intelligent faults diagnosis with its powerful modeling and characterization capabilities [3, 4]. DL-based fault diagnosis method has a strong ability to learn the fault feature of raw data. However, a large amount of labeled data is required for the training of DL model [5, 6]. Under practical working conditions, although the collected vibration data are of large capacity, most of the vibration data are not labeled. Collecting and labeling sufficient fault data for each condition are unrealistic [7, 8].

In order to solve the fault diagnosis problem of insufficient labeled data, there are some research studies on fault

diagnosis of limited data. Hang et al. proposed principal component analysis (PCA) and applied it to the field of high-dimensional imbalance fault diagnosis data [9]. Duan et al. applied a new support vector data description method for mechanical fault diagnosis of unbalanced datasets [10]. Many scholars introduced the Generative Adversarial Networks (GANs) [11] idea to generate the required enough vibration data for classifiers. Wang et al. proposed a novel fault diagnosis method that combined the GAN and Stacked Automatic Encoder (SAE) [12]. The samples produced by the generator were used to expand the sample size on the input along with the original samples. Xie and Zhang proposed a deep convolutional GAN model that simulated the original distribution of several fault categories and solved the imbalance of training data by generating new data [13]. Li et al. proposed a cross-domain fault diagnosis method based on deep generation neural network, which provided reliable

cross-domain diagnosis results by artificially generating fake samples for domain adaptation [14]. Mao et al. used FFT to preprocess the original vibration signal to obtain the spectrum data of fault samples as the input of GAN and generated fake samples for a few fault classes [15]. Then, a stacked denoising auto encoder model for fault diagnosis was established and verified on a series of comparison experiments.

From the above, GAN is used to generate the required target data for establishing effective mapping relationships of fault feature [16, 17]. However, in the practical industrial environment, the fault features in the raw vibration signal are complicated [18, 19]. The data generated by the GAN can only fit the significant features of the fault signal, which finds difficulty in fully meeting the original signal in actual situation. Moreover, the fault mode of vibration data is variable under different working conditions. If we used the vibration data under one condition to establish the mapping relationship of the different fault features, the robustness of the GAN model is unsatisfactory. Therefore, it is difficult to obtain high fault classification accuracy under variable working conditions.

Considering that labeled fault data under variable working conditions are difficult to collect, this paper proposes a semisupervised learning deep adversarial convolutional neural network model (SACNN). We no longer use GANs to generate target data but directly train high-precision diagnostic models with few labeled data and large amounts of unlabeled data.

The main insights and contributions of this study are summarized as follows:

- (1) A semisupervised learning deep convolutional network is proposed for fault diagnosis with insufficient labeled data. The semisupervised GAN is composed of a generator and a variant discriminator, which are constructed by multiple groups of convolutional network modules. By respectively defining true and fake labels for the unlabeled data and the generated data to train the discriminator, the model can obtain the fault features of the overall sample distribution.
- (2) Three regular terms, one-sided label smoothing, L2 weight penalty, and feature-matching regular terms, are, respectively, defined into the loss function of the model to constrain the parameters of the discriminator, which can improve the classification accuracy and stability of the model.
- (3) The series experiments are conducted on the CWRU bearing dataset and lab-built bearing experimental platform. Compared with three other popular methods, the advantages of the model in a few labeled datasets are verified.

The rest of the paper is organized as follows. Section 2 introduces preliminaries of SACNN. Section 3 details the proposed SACNN model, including problem description, data preprocessing, model structure, and optimization objectives. In Section 4, experimental verification and corresponding analysis are conducted. The conclusions are drawn in Section 5.

## 2. Preliminaries

**2.1. Generative Adversarial Networks.** The generation adversarial network (GAN) is a generic model that can randomly generate observation data. GAN is consisted of two subnetworks: a generator and a discriminator, as shown in Figure 1. The noise from the prior distribution  $z_{\text{noise}}$  as the input of the generator hopes to generate the data  $G(z_{\text{noise}})$ , which is expected to be true. The discriminator is used to determine whether the input data are generated data or a real sample. If the input is the real data, then the output of the discriminator is expected to be 1, where  $D(x) = 1$ ; otherwise,  $D(x) = 0$ . Through this training process, the discriminator learns the characteristics of the real data. The loss function of the discriminator is expressed as

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

where the first part is used to identify real data and the second part is used to identify fake data.

The loss function of generator is

$$\min_G V(G) = E_{z \sim p_z} [\log(1 - D(G(z)))]. \quad (2)$$

GAN is usually expressed in the form of a zero-sum game, using the loss function  $V(D, G)$  to represent the loss of the GAN, where  $G$  wants to minimize  $V$ , while  $D$  wants to maximize  $V$ :

$$\min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}} [\log D(x)] + E_{z \sim p_z} [\log(1 - D(G(z)))]. \quad (3)$$

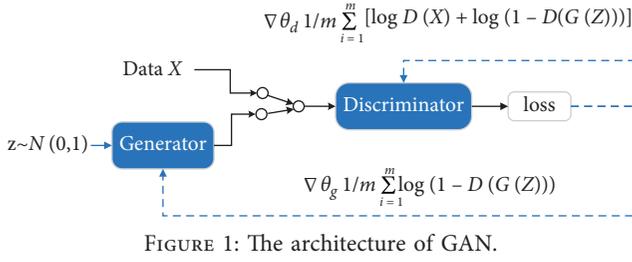
The parameters of the two networks are updated by alternating gradient descent. The parameters of the generator are fixed firstly; the discriminator is trained by using the generated and the true data. Then the parameters of discriminator are fixed and the generator's parameters are updated. The generator can finally generate realistic data by repeated alternating training.

## 3. Proposed Method

**3.1. Problem Description.** GAN can realize the expansion of small samples, but the generated data are difficult to meet the actual situation. The semisupervised learning can add unlabeled data to the supervised classification algorithm. Although unlabeled data do not provide labels, it provides information about the overall distribution of datasets, which provide a better classification surface for the model.

Therefore, according to the idea of semisupervised learning, the intelligent fault diagnosis problem based on semisupervised learning deep adversarial network is formally defined:

- (1) There is a fault bearing dataset, including two parts of data. One has  $n^L$  labeled vibration fault dataset  $X^L = \{x_i^L, y_i^L\}_{i=1}^{n^L}$ , and the other has  $n^U$  unlabeled vibration fault dataset  $X^U = \{x_i^U\}_{i=1}^{n^U}$  and  $n^L \ll n^U$ .



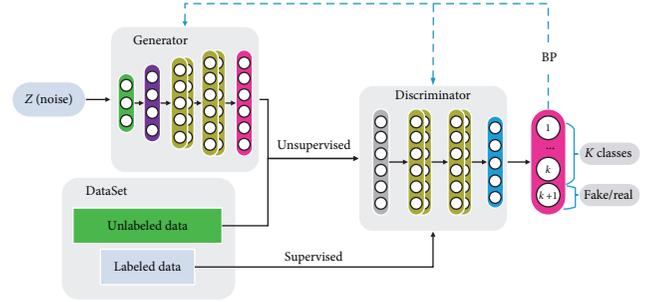
- (2)  $X^L$  is used to supervise training and  $X^U$  is used for assisted training to improve the fault classification accuracy of the model.
- (3) GAN is used in the semisupervised learning field; it is still responsible for generating images from the input noise data. So we define the data generated by the generator as  $X^z = \{x_i^z\}_{i=1}^z$ . Define all data labels in  $X^U$  as true and in  $X^z$  as fake.
- (4) The discriminator  $D$  is not a simple true-fake classification (binary classifier). Assuming that the input data have  $k$  classes,  $D$  is a classifier of  $k + 1$ , and the  $k + 1$ th class is the discriminator of whether the input is true or fake. The overview of the problem description is shown in Figure 2.

The output of discriminator network  $D$  extracts the estimated probability of input data from the data generation distribution. Traditionally, this is a feedforward network, which ends with a single sigmoid unit. The output based on the sigmoid unit is [True, Fake]. But, in our paper, we modify the discriminator network, ending with the Softmax output layer. Thus, the discriminator network has  $n + 1$  output units, corresponding to [1-class, 2-class, ...,  $k$ -class, True/Fake]. In this case, the labeled data can be input into the discriminator for training. The discriminator can not only distinguish between true and fake but also output fault category.

**3.2. Data Preprocessing.** To obtain the time-frequency domain information of the original vibration signal, we used the wavelet transform (WT) [21] to transform the one-dimensional time-domain vibration signal into 2-dimensional image, as shown in Figure 3. This transformation enables localized analysis of the time-frequency domain and progressive multiscale refinement of the signal. Finally, the time domain is subdivided at the high frequency of the signal and the frequency domain is subdivided at the low frequency. Thus, WT can focus on any detail of the signal, so as to automatically adapt to the analysis requirements of time-frequency signal. The formula of continuous WT is given by

$$\text{WT}(a, t) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) * \psi\left(\frac{t - \tau}{a}\right) dt, \quad (4)$$

where  $a$  is the scaling factor of the wavelet function;  $\tau$  is the translation factor of the wavelet function;  $f(t)$  is the raw vibration signal; and  $\psi(t)$  is the wavelet basis function used in this paper, expressed as follows:



$$\psi(t) = \exp^{-t^2/2} \cos(5t). \quad (5)$$

In this paper, we selected 512 time-domain data points of the original signal for wavelet transformation each time. The wavelet function center rate was 0.8125, and the transformation results were converted into black and white images with a size of  $64 * 64$ , which were input into the model.

**3.3. Model Structure of SACNN.** This proposed model consists of two parts: a generator and a discriminator. Generator  $G$  receives an  $n$ -dimensional vector as input,  $x \in R^n$ , which is mapped to  $m$ -dimensional data space,  $G: R^n \rightarrow R^m$ . The output is  $Z = G(x; \theta_g)$ ,  $G(\cdot)$  represents the data space mapping the noise to real images,  $\theta_g$  represents the parameters of generator, and  $Z$  is the output data.

The network structure of the generator is first a fully connected layer that maps the original input to 1024 dimensional vectors, followed by a Batch Normalization (BN) layer, a ReLU activation function layer, and finally a deconvolution layer. A deconvolution module is shown as follows:

$$h^g = W^g \otimes (\text{Relu}(BN(x))) + b^g, \quad (6)$$

where  $\otimes$  is the deconvolution operation,  $W^g$  and  $b^g$  are the deconvolution kernel and the bias, and  $h^g$  is the result of the deconvolution operation. The generator is composed of four such deconvolution modules. The size of the four deconvolution cores is  $5 \times 5$ , and the number of deconvolution cores is 8, 16, 32, and 64, successively. Then, BN layer is used to speed up convergence and improve generalization performance.

The structure of the discriminator consists of three convolution modules. Each convolution module is firstly a 2D convolution operation and then a Max\_Pooling layer, followed by a BN layer and then a Leaky ReLU activation function layer. The form of the convolution block is expressed as follows:

$$\begin{aligned} m^d &= \text{Max\_pool}(W^d * x + b^d), \\ h^d &= \text{LRelu}(BN(m^d)), \end{aligned} \quad (7)$$

where  $*$  is the convolution operation,  $W^d$  and  $b^d$  are the convolution kernel and the bias, and  $h^d$  is the result of the convolution operation. The sizes of the three convolution kernels are  $5 \times 5$ ,  $3 \times 3$ , and  $3 \times 3$ , respectively, and the numbers of convolution kernels are 64, 128, and 256, respectively.

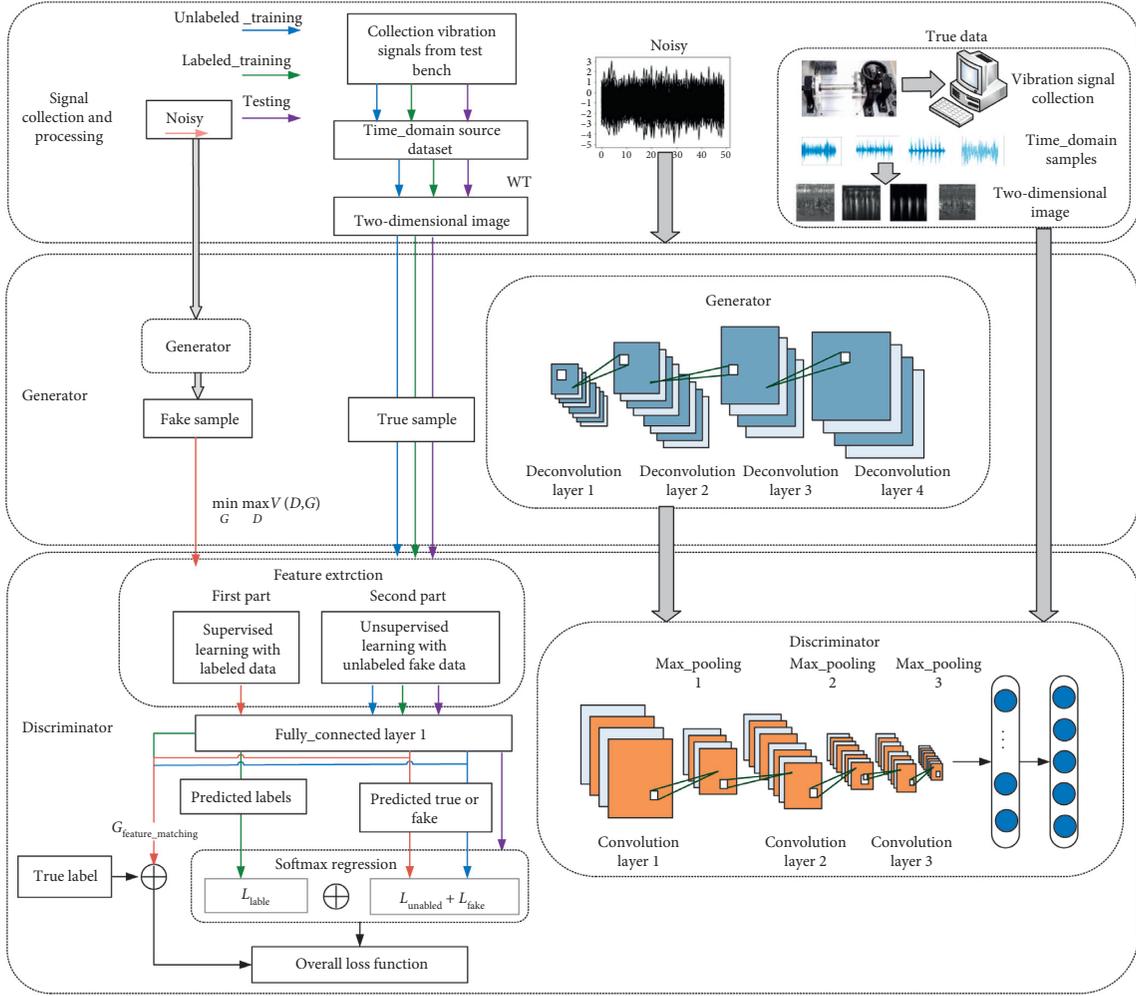


FIGURE 3: Flow chart of the proposed model.

For identifying multiple fault classes, we enhanced the discriminator as a standard classifier based on previous work proposed by Salimans et al. [22]. We use the Softmax functions to replace the output of the discriminator. Suppose that the random vector  $z$  has a uniform noise distribution  $P_z(z)$ , and  $G(z)$  maps it to the data space of the real image. Assuming that there is a distribution  $P_{\text{data}}(x, y)$  of training dataset, the input  $x$  of the discriminator is a real or fake image with the label  $y$ . Discriminator output is a  $k + 1$  dimensional vector  $l = \{l_1, l_2, \dots, l_{k+1}\}$ . Finally, it is converted into a  $k + 1$  dimensional vector of class probability  $P = \{p_1, p_2, \dots, p_{k+1}\}$  by Softmax function. The probability  $p_j$  for the  $j$ th class is expressed as

$$p_j = \frac{e^{l_j}}{\sum_{i=1}^{k+1} e^{l_i}}, \quad j \in \{1, 2, \dots, k + 1\}. \quad (8)$$

Then, the real data are divided into one of the former  $k$  classes, and the fake data are divided into the  $k + 1$ th class. The flow chart of the proposed model is shown is Figure 3.

**3.4. Optimization Objective.** In the process of training, we input labeled data and unlabeled data into the model

training at the same time. We divide the input data into two parts according to semisupervised learning. The first part is labeled data, and the second part is data generated by the generator and unlabeled data. When the input is labeled data, it is supervised training and the discriminator needs to identify the specific fault category. When inputting the true unlabeled data and the fake data generated by the generator, it is unsupervised training, and the discriminator only needs to complete true-fake discrimination.

The training of the whole model needs to optimize three kinds of loss function. For the labeled data in the training set, the model needs to calculate the corresponding probability of fault classification; the classification loss  $L_{\text{label}}$  is as follows:

$$L_{\text{label}} = -E_{x, y \sim P_{\text{data}}} [\log p_{\text{model}}(y | x, y < k + 1)]. \quad (9)$$

For the unlabeled data in the training set, by calculating the probability of not being estimated as the  $k + 1$ th class, the loss  $L_{\text{unlabel}}$  discriminating the truth and fakeness of the input data is as follows:

$$L_{\text{unlabel}} = -E_{x \sim P_{\text{data}}} [\log 1 - p_{\text{model}}(y | x, y = k + 1)]. \quad (10)$$

For the fake data generated by the generator, by calculating the probability of being estimated as the  $k + 1$ th

class, the loss  $L_{\text{fake}}$  discriminating the truth and fakeness of the input data is as follows:

$$L_{\text{fake}} = -E_{x \sim G} [\log p_{\text{model}}(y | x, y = k + 1)], \quad (11)$$

where  $P_{\text{model}}(y | x)$  represents the probability that the discriminator output is fake,  $G$  represents the probability distribution of fake data produced by the generator, and  $P_{\text{data}}$  represents the probability distribution of the real data.

When we calculate all the loss functions, one-sided label smoothing method is used for each label of the data in order to encourage the discriminator to estimate a flexible probability rather than predicting the classification of extreme confidence. If the discriminator relies too heavily on a certain set of features to detect the real data, the generator can quickly mimic these features to fool the discriminator in the GAN. Then the discriminators will be easily exploited by generators to seriously damage classification accuracy. To avoid this problem, we optimize the discriminator when the prediction value of any real data exceed 0.9 ( $D(\text{real\_data}) > 0.9$ ). The target label value is set to 0.9 instead of 1.0.

When the model input is real unlabeled data, it is sufficient to maximize the probability that the output is real data, without requiring specific classification, that is, maximize  $\log(p_{\text{model}}(y|x, y < k + 1))$ . As the training data of generator consist of two parts, generated fake data and unlabeled real vibration data, the loss function of the discriminator  $D$  consists of two parts: one is the supervised learning loss and the other is the unsupervised learning loss. The formula is as follows:

$$L_D = L_{\text{label}} + L_{\text{unlabel}} + L_{\text{fake}} + D_{\text{regular}}, \quad (12)$$

where  $D_{\text{regular}}$  represents the L2 regularization of all weights in the discriminator.

For the generator, we hope the output data can fool the classifier. Thus, the loss function of the generator  $L_G$  is

$$L_G = -L_{\text{fake}} + 0.01 * G_{\text{feature\_matching}}, \quad (13)$$

where  $G_{\text{feature\_matching}}$  is a regularization term, which is proposed by Goodfellow et al. [9] to improve the stability of the model during the training process. The fake data and real data are input to the discriminator to match the fault characteristics of the middle layer, so that the results of each layer are as similar as possible. Suppose that  $f(x)$  represents the activation function of the middle layer of discriminator, and  $Z$  is the noise output of the generator.  $G_{\text{feature\_matching}}$  is defined to match the fault feature between the training data distribution and the generated fake data, expressed as

$$G_{\text{feature\_matching}} = \left\| E_{x \sim p_{\text{data}}} - E_{z \sim P_z(Z)} f(G(Z)) \right\|_2^2. \quad (14)$$

After the design of the two objective functions, the backpropagation algorithm can be used to train the model parameters. In this paper, the adaptive moment estimation (Adam) algorithm is used to update the parameters. The algorithm of the gradient provides independent adaptive learning rates for different parameters by calculating the first-order moment estimation and the second-order

moment estimation of the gradient. Based on the aforementioned equation, the parameters  $\theta_d$  and  $\theta_g$  are updated as follows:

$$\begin{aligned} L(\theta_g^*, \theta_d^*) &= \min_{\theta_g, \theta_d} L_G(\theta_g) + L_D(\theta_d), \\ \theta_d &\leftarrow \text{Adam}(\nabla_{\theta_d} L_D, \theta_d, \alpha, \beta_1, \beta_2), \\ \theta_g &\leftarrow \text{Adam}(\nabla_{\theta_g} L_G, \theta_g, \alpha, \beta_1, \beta_2), \end{aligned} \quad (15)$$

where  $\alpha, \beta_1$ , and  $\beta_2$  are the Adam optimizer parameters and  $\theta_d$  and  $\theta_g$  are the determiner and generator network parameters. The algorithm pseudocode is shown in Algorithm 1.

## 4. Case Study

In this section, the performance of our proposed SACNN model is evaluated on the CWRU bearing datasets [20] and datasets of the lab-built experimental platform.

### 4.1. Case Study I: Fault Diagnosis of CWRU Bearing

#### 4.1.1. Data Description

*CWRU Bearing Dataset.* Fault bearing datasets from the Case Western Bearing Data Center are one of the most commonly used benchmark datasets in the field of fault diagnosis. The test bench consists of a 2 HP motor (left), a torque sensor (middle), a dynamometer (right), and control electronics, as shown in Figure 4.

Single point pitting fault bearings with EDM have three fault diameters of 0.007, 0.014, and 0.021 inches, respectively. The load is 0 HP and the sampling frequency is 12,000 Hz. Each fault diameter contains three faults: Ball fault (BF), Inner ring fault (IF), and Outer ring fault (OF), as well as Normal signals (N) under this condition. Figure 5 shows the conversion of four different fault vibration signals to 2-dimensional image. It can be seen from Figure 5 that the vibration signals of the different fault types have distinctly different distribution.

Based on the above fault data, we set the training and testing datasets with different capacity to verify the proposed model, as shown in Table 1. Each of the datasets contained four faults with different fault diameters. The training set contained a small amount of labeled data and a large amount of unlabeled data. The labeled data are randomly selected from vibration data of different fault types.

*4.1.2. Experimental Results and Analysis.* The fault diagnosis results of the proposed method on four different training datasets are shown in Figure 6, compared with three different popular methods, CNN, CNN\_SVM, and SAE; the details of three methods are shown in Table 2. Convolutional neural network (CNN) is widely used in fault diagnosis due to its advantages of local perception and parameter sharing. In this paper, we constructed the CNN model with a three-layer two-dimensional convolutional network, and Softmax

Algorithm: SACNN Training Algorithm  
**Demand:** Number of total iteration  $I$ ; The number of step to apply to the discriminator  $t$ ; Batch size  $m$ ; Adam optimizer parameters  $\alpha, \beta_1, \beta_2$ ; Number of fault categories  $k$ ; Amount of labeled data control parameters flag.  
**For**  $I$  epochs **do**  
  **For**  $t$  steps **do**  
    **If**  $t < \text{flag}$  **then**  
      Draw  $m$  real samples  $\{(x^{(1)}, y^{(1)}) \dots (x^{(m)}, y^{(m)})\}$  from  $P_{\text{data}}(X)$   
      **Else**  
        Draw  $m$  real samples  $\{x^{(1)} \dots x^{(m)}\}$  from  $P_{\text{data}}(X)$   
        Draw  $m$  noise samples  $\{z^{(1)}, \dots, z^{(m)}\}$  from noise  $P_g(Z)$   
         $x_{\text{noise}} \leftarrow G(Z)$   
      **End If**  
      The discriminator is updated by ascending its stochastic gradient:  
       $\nabla_{\theta_d} L_D \leftarrow \nabla_{\theta_d} - (1/m) \sum_{i=1}^m [\sum_{j=1}^k y \log(p_j(y | x^i, y < k + 1)) + \log(1 - p_j(y | x^i, y < k + 1)) + \log p_j(y | x_{\text{noise}}^i, y < k + 1)]$   
       $\theta_d \leftarrow \text{Adam}(\nabla_{\theta_d} L_D, \theta_d, \alpha, \beta_1, \beta_2)$   
      **End for**  
      Draw  $m$  noise samples  $\{z^{(1)}, \dots, z^{(m)}\}$  from noise  $P_g(Z)$   
       $x_{\text{noise}} \leftarrow G(Z)$   
      The generator is updated by descending its stochastic gradient:  
       $\nabla_{\theta_g} L_G \leftarrow \nabla_{\theta_g} (1/m) \sum_{i=1}^m [\log(p_j(y | x_{\text{noise}}^i, y < k + 1))]$   
       $\theta_g \leftarrow \text{Adam}(\nabla_{\theta_g} L_G, \theta_g, \alpha, \beta_1, \beta_2)$   
    **End for**

ALGORITHM 1: Algorithm pseudocode.

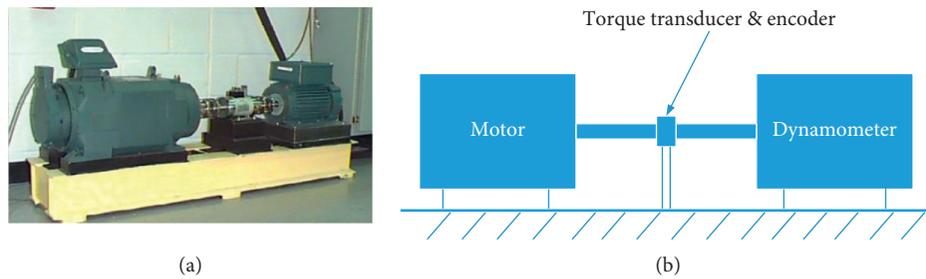


FIGURE 4: Bearing test bench: (a) picture of bearing test bench; (b) schematic diagram of bearing test bench.

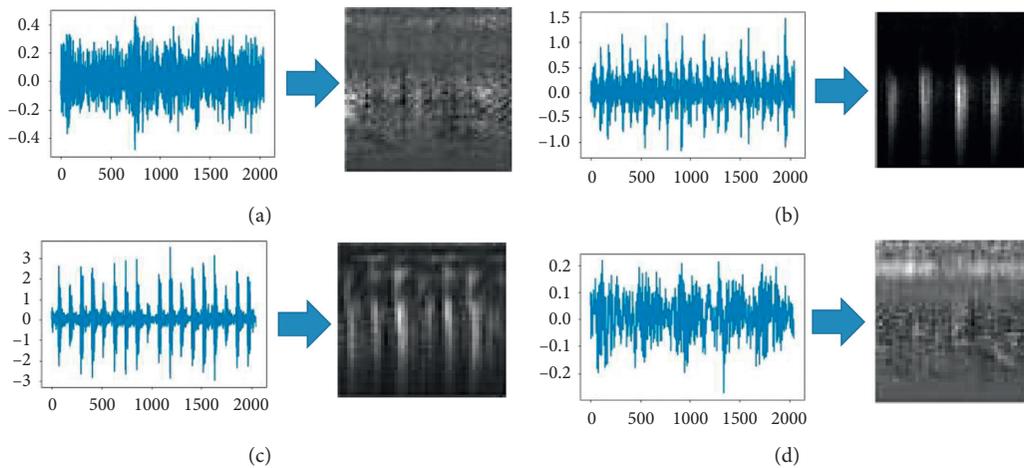


FIGURE 5: Transformations of original vibration signal to the 2-dimensional image: (a) Ball; (b) Inner; (c) Outer; (d) Normal.

is used as the classifier. Support vector machine (SVM) classifier has good performance and generalization with few samples. Thus, we chose CNN based on Support Vector

Machine (CNN\_SVM) method as another contrast method, which uses a three-layer two-dimensional convolutional network as the feature extractor and support vector machine

TABLE 1: Four dataset settings.

No.	Fault type	Fault size	Training dataset Labeled/unlabeled	Testing dataset Labeled
1	BF, OF, IF, N	0.007/0.014/0.021	50/4645	2005
2	BF, OF, IF, N	0.007/0.014/0.021	100/4595	2005
3	BF, OF, IF, N	0.007/0.014/0.021	150/4545	2005
4	BF, OF, IF, N	0.007/0.014/0.021	200/4495	2005

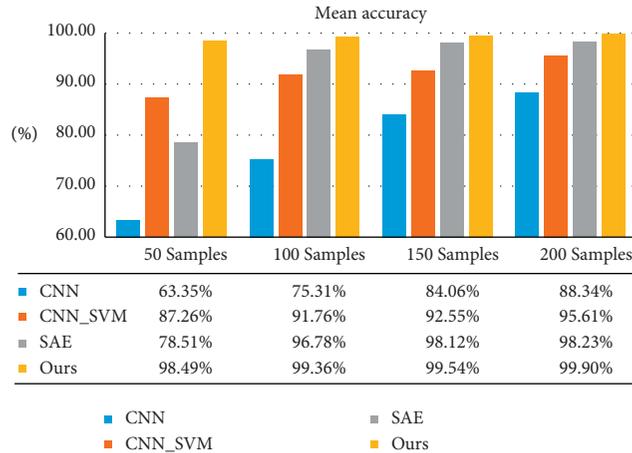


FIGURE 6: Comparison of classification accuracy of different methods.

TABLE 2: Details of comparison methods.

Methods	Operator	Number of layers	Size of kernel/AE	Classification	Optimizer
CNN	Convolution	3	$5 \times 5$ , $3 \times 3$ , $3 \times 3$	Softmax	Adam
CNN_SVM	Convolution	3	$5 \times 5$ $3 \times 3$ , $3 \times 3$	SVM	Adam
SAE	AutoEncoder	$2 \times 2$	$400 \times 100$ , $400 \times 100$	Softmax	Adam

as a classifier. Stacked Automatic Encoder (SAE) initializes the network parameters through unsupervised layer-wise pretraining, so as to improve the convergence speed of deep network and slow down the influence of gradient disappearance. We constructed the SAE network with two self-encoders and uses Softmax as the classifier. The experimental results are the average results of multiple experiments. The DL framework TensorFlow is used for the implementation of each method, and all methods are solved using the Adam algorithm.

Figure 6 shows the classification accuracy of the four methods on datasets nos. 1, 2, 3, and 4, where 50 samples represent 50 labeled data in the training dataset, the same as 50 samples, 150 samples, and 200 samples. The CNN model has the same 3 convolutional layers as the discriminator of SACNN model. Because only a limited amount of labeled data is used for supervised training and no unlabeled data is used for auxiliary training, the accuracy of the CNN model is the lowest in the four models. The diagnosis accuracy on 50 labeled samples is only 63.35%, while the diagnosis accuracy on 200 labeled samples is 88.34%.

Similarly, the CNN\_SVM model also contains three convolutional layers. Due to the enhanced constraints of the network parameters, the classification accuracy of CNN\_SVM on four datasets is higher than that of CNN model. The diagnosis accuracy of CNN\_SVM on 50 labeled samples was 87.26%, while the diagnosis accuracy on 200 labeled samples was 95.61%.

The SAE method only uses labeled data for supervised training. The diagnosis accuracy of SAE on four datasets is higher than that of CNN and CNN\_SVM. Because each layer of SAE network is trained separately, which is equivalent to initializing a reasonable value, such a network will be easier to train and have faster convergence and higher accuracy.

As we know that the proposed SACNN model is trained on not only small labeled samples but also a large number of unlabeled samples, besides the fact that model parameters are constrained by the loss of three regularization terms, the diagnostic accuracy of the proposed model on four datasets is the highest among the four models. For example, the

accuracy of the SACNN model trained on 50 labeled samples is 98.49%, which is about 20% higher than that of SAE model. Therefore, the results prove that the SACNN model has obvious advantage over the traditional deep model in the limited dataset.

Since the SACNN method performs well on all four datasets, we only chose one case to discuss the details of the model performance on the dataset No. 3, as shown in Figure 7. Figure 7(a) is the relationship between test accuracy and number of iterations during training, and Figures 7(b) and 7(c) are training loss of the generator and discriminator in the training process. From the three figures, we can see the fast convergence and high precision of the SACNN model.

Furthermore, in order to observe the impact of different labeled samples on the SACNN model, the diagnosis accuracy on four datasets with different capacities is shown in Figure 8. It can be concluded that the fault diagnosis results will be better with the increase of the numbers of training labeled samples. These results are reasonable. Even if there are only 50 labeled samples in the training set, the diagnosis accuracy of the model can still reach 98.49%, which strongly proved the effectiveness of the model with large vibration data but few labels.

## 4.2. Case Study II: Fault Diagnosis of Lab-Built Experimental Platform

### 4.2.1. Data Description

*Bearing Dataset of Lab-Built Experimental Platform.* In order to further study the performance of the proposed method in practice working conditions, we built the experimental platform in our lab, as shown in Figure 9. The three-phase motor through a flexible coupling controls the speed of bearing. Radial load is applied to the bearing block to simulate real working conditions, and vibration signals are collected by the acceleration sensor on the bearing block. The bearing contains four faults: Ball fault (BF), Inner ring fault (IF), and Outer ring fault (OF), as well as Normal bearing (N). The sampling frequency is 12.8 kHz and the rotating speed is 1500 r/min. We selected the original vibration signal from four types of fault bearings, as shown in Figure 10.

Comparing the time-domain vibration signals from Figures 5 and 10, it can be seen that the data distributions in the two datasets are quite different. Because the speed, load, and other conditions we set in the experimental platform are consistent with the actual working conditions, the collected vibration data can be considered as close to the actual working conditions, which can be used to test the performance of the method under actual complex working conditions.

We set up training and testing datasets with different capacities to validate the proposed model, as shown in Table 3. The training set contains a small amount of labeled data and a large amount of unlabeled data. The labeled data are randomly selected from vibration data of four different fault types. Each type of fault sample contains 1400 pieces of

unlabeled training data and 600 pieces of testing data. All data have been transformed into a 2-dimensional image by wavelet transform from 512 sample points.

*4.2.2. Experimental Results and Analysis.* The diagnosis accuracy on the different datasets is shown in Table 4, compared with the three other models. The results are the average accuracy of ten experiments on each dataset. It can be seen that the diagnostic accuracy of the proposed model on four datasets is the highest among the four models. For instance, the accuracy of our proposed model can achieve 96.91% on 50 labeled samples, while that of CNN model is 57.36%, that of CNN\_SVM model is 82.33%, and that of SAE model is 78.39%.

Moreover, we also find that the diagnosis accuracy of the four models on the datasets of lab-built experimental platform is lower than that on the CWRU bearing dataset. We analyze that the reason is that, compared with the standard data set of CWRU, the operation condition of lab-built experimental platform is closer to the actual working condition, so the vibration signals collected from the fault bearings are more complex, and then the fault characteristics are more difficult to learn in the model. Thereby the diagnosis accuracy of the model is lower than that of the CWRU dataset.

In order to illustrate the diagnosis results of each fault type in detail, we draw a confusion matrix for the test accuracy on the CWRU dataset No. 3. Figure 11 shows the classification accuracy of the model for each fault at different training stages. The model has different recognition accuracy for different fault classes. In Figure 11(a), after training 20 epochs, some Inner ring faults are misclassified as Outer ring faults, and some Outer ring faults are misclassified as Normal and Inner ring faults. Therefore, the accuracy of Outer and Inner ring faults using SACNN model is 62% and 64%, respectively.

Besides, as the number of iterations increases, the classification accuracy of each type of fault is improved. It can be seen from Figure 11(c) that the diagnosis accuracy of Outer ring faults is nearly 100% after training 300 epochs, while it is 88% after training 100 epochs and 62% after training 20 epochs.

In order to intuitively judge the quality of the features learned by the proposed method, we visualize the fault features trained on dataset No. 3, as shown in Figure 12. For different epochs, we use the T-SNE [23] method to reduce the high-dimensional fault features obtained from the previous layer of the classifier to 2 dimensions. Figure 12(a) is the result of direct dimensionality reduction of the original signal, and Figures 12(b)–12(f) are 2-dimensional images after feature dimensionality reduction after 50 epoch trainings. The colors of the dot indicate the types of fault: purple indicates Ball fault (BF), yellow indicates Normal signal (N), blue indicates Inner ring fault (IF), and green indicates Outer ring fault (OF).

We describe its characteristics qualitatively: the feature distribution points of different categories should be separated as much as possible, which indicates that the learned

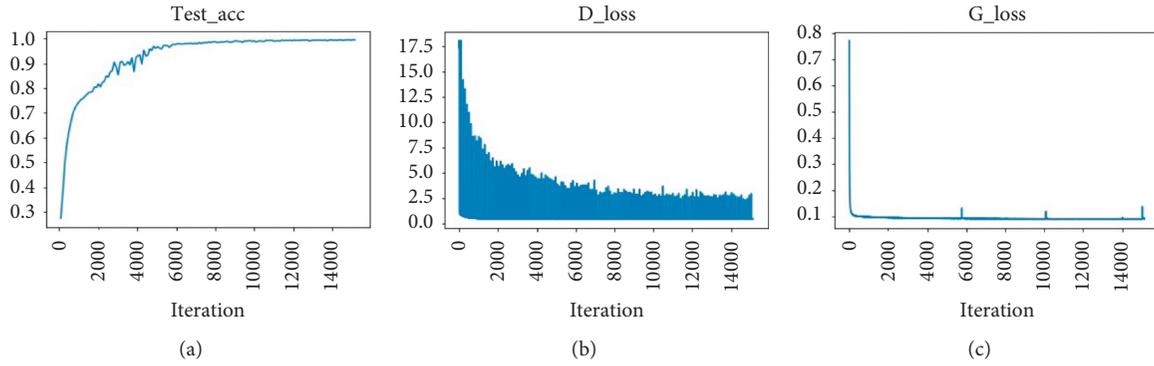


FIGURE 7: SACNN model performance on dataset No. 3: (a) testing accuracy of the classifier; (b) convergence performance of generator during training; (c) convergence performance of discriminator during training.

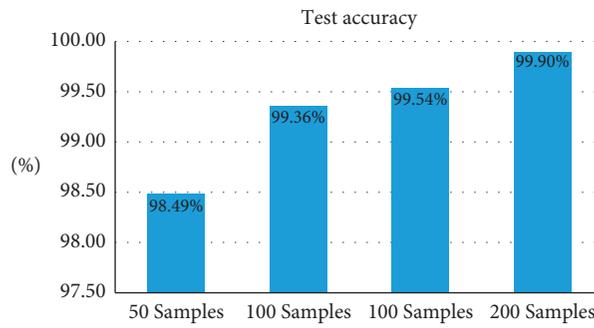


FIGURE 8: Comparison of diagnosis accuracy with different label dataset sizes.

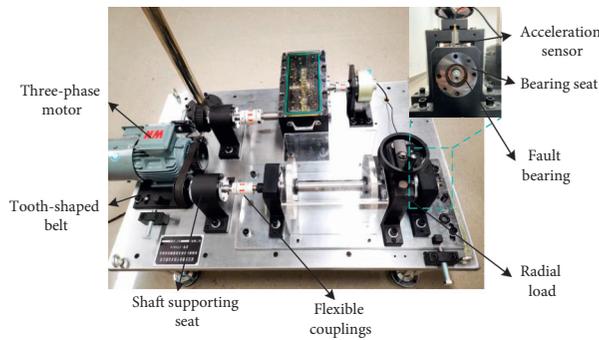


FIGURE 9: Lab-built experimental platform.

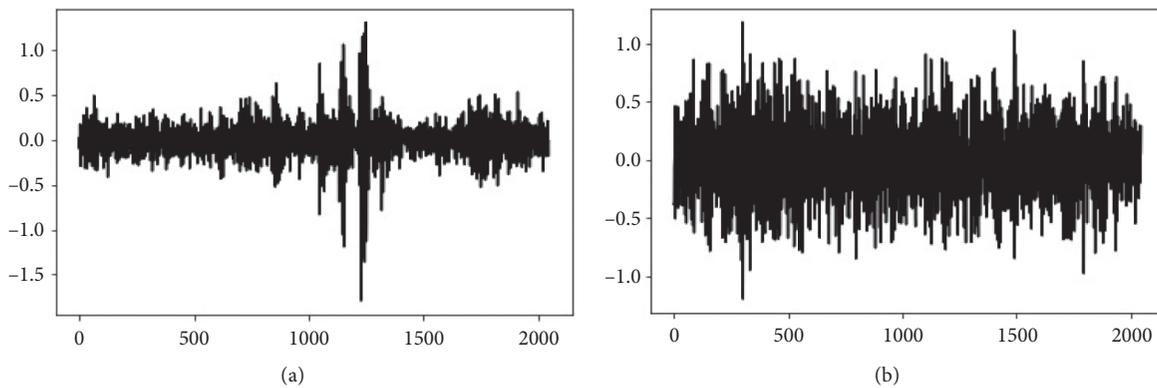


FIGURE 10: Continued.

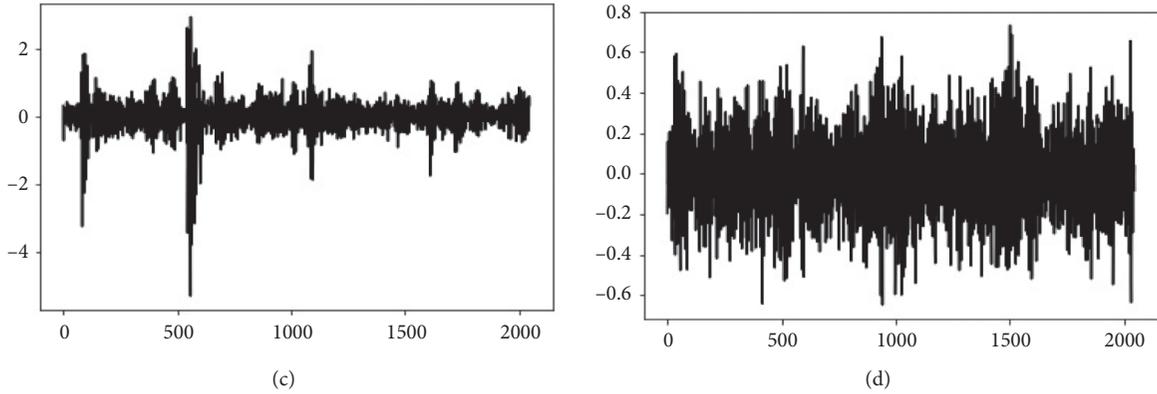


FIGURE 10: Time-domain vibration signal collected by the lab-built experimental platform: (a) Ball; (b) Outer; (c) Inner; (d) Normal.

TABLE 3: Dataset settings of the experimental platform.

No.	Fault type	Training dataset Labeled/unlabeled	Testing dataset Labeled
5	BF, OF, IF, N	50/5600	2400
6	BF, OF, IF, N	100/5600	2400
7	BF, OF, IF, N	150/5600	2400
8	BF, OF, IF, N	200/5600	2400

TABLE 4: Verification results of lab-built experimental platform datasets.

Method	50 samples	100 samples	150 samples	200 samples
CNN	0.5736	0.7728	0.7868	0.8139
CNN_SVM	0.8233	0.8879	0.9399	0.9516
SAE	0.7839	0.8433	0.8728	0.8922
Ours	0.9691	0.9776	0.9882	0.9928

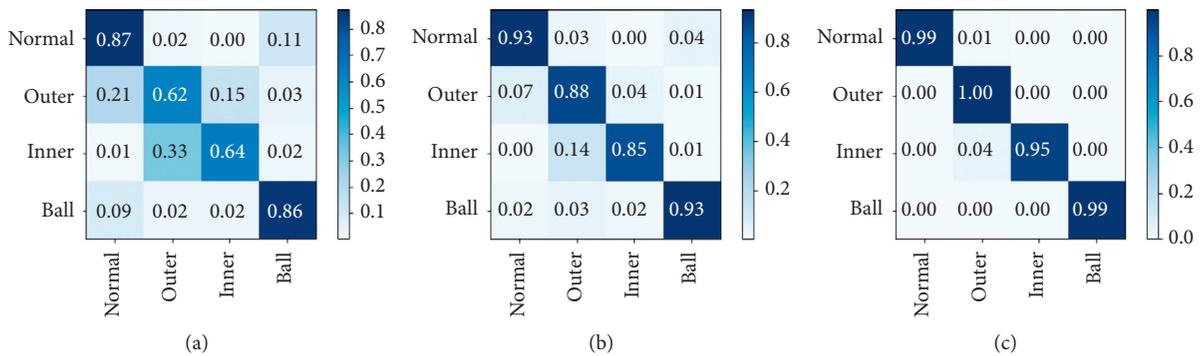


FIGURE 11: Confusion matrix of SACNN model on dataset No. 3: (a) training after 20 epochs; (b) training after 100 epochs; (c) training after 300 epochs.

fault features are more separable and the classification accuracy is higher. As can be seen from Figure 12(a), the feature points of the original vibration signal are completely interleaved and cannot be directly used for classification. After training 50 epochs, Figure 12(b) shows that

different types of fault data can be roughly separated. However, a small number of fault feature data points are overlapped between OF, IF, and N fault data. From Figures 12(c)–12(f), we can obviously find that as the number of epochs increases, the overlapping data points

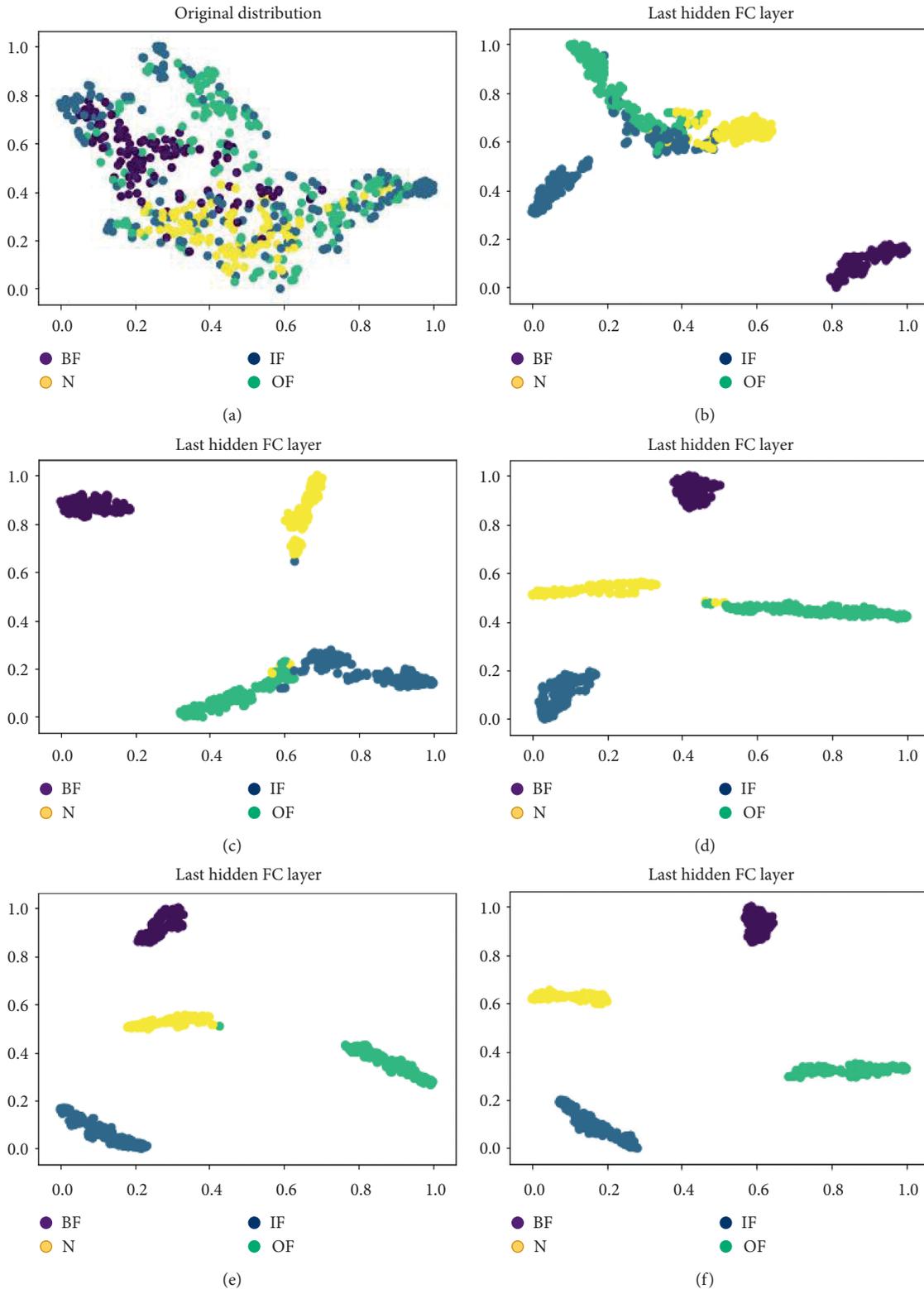


FIGURE 12: Feature visualization.

are decreasing, and the distinction between different fault feature data points is becoming increasingly apparent. Therefore, the features learned during the training process are highly discriminative, and the model has satisfactory diagnosis performance.

### 5. Conclusion

In this study, a fault diagnosis method based on semi-supervised learning deep adversarial network model is proposed to solve the problem of insufficient fault label data

under practical working conditions. Based on semi-supervised learning, a small amount of labeled data can be trained in supervised learning. In unsupervised learning combined with GAN, we added a large amount of unlabeled data to help model training. The network structure and the optimization loss of the model are designed. We demonstrate the performance of our proposed method on two different datasets, compared with three other common models. The change process of the characteristics learned during the whole training process is further analyzed by visualization technology to verify the superiority of the method.

Despite obtaining the promising results, the method in this paper can effectively identify known faults, but the algorithm cannot be generalized to new fault categories to achieve cross-category fault diagnosis. In future work, we will attempt to combine few-shot learning to study cross-category fault diagnosis.

### Data Availability

The data used in this paper include two parts. The first part is the public bearing dataset from the Western Reserve University Data Center, and the second part is the bearing data collected by the experimental platform set up by our laboratory.

### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

### Acknowledgments

This research was funded by the National Key Research & Development Plan of China (no. 2018YFB2000505), National Natural Science Foundation of China (no. 61806067), and the Natural Science Foundation of Anhui Province (no. 1908085ME132).

### References

- [1] Y. Lei, "Remaining useful life prediction," *Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery*, vol. 6, pp. 281–358, 2017.
- [2] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp. 98–109, 2017.
- [3] F. Jia, Y. Lei, L. Guo, J. Lin, and S. Xing, "A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines," *Neurocomputing*, vol. 272, pp. 619–628, 2018.
- [4] 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," *Knowledge-Based Systems*, vol. 116, no. C, pp. 1–12, 2017.
- [5] T. Sorsa, H. N. Koivo, and H. Koivisto, "Neural networks in process fault diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 4, pp. 815–825, 1991.
- [6] P. K. Kankar, S. C. Sharma, and S. P. Harsha, "Fault diagnosis of ball bearings using machine learning methods," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1876–1886, 2011.
- [7] M. Gan, C. Wang, and C. A. Zhu, "Construction of hierarchical diagnosis network based on Deep learning and its application in the fault pattern recognition of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 72–73, pp. 92–104, 2016.
- [8] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," *Mechanical Systems and Signal Processing*, vol. 107, no. 1, pp. 241–265, 2018.
- [9] Q. Hang, J. Yang, and L. Xing, "Diagnosis of rolling bearing based on classification for high dimensional unbalanced data," *IEEE Access*, vol. 7, pp. 79159–79172, 2019.
- [10] L. Duan, M. Xie, T. Bai, and J. Wang, "A new support vector data description method for machinery fault diagnosis with unbalanced datasets," *Expert Systems with Applications*, vol. 64, pp. 239–246, 2016.
- [11] I. J. Goodfellow, J. P. Abadie, M. Mirza, and B. Xu, "David warde-farley, sherjil ozair, aaron courville, yoshua bengio. Generative adversarial nets," in *Proceedings of the 28th Conference on Neural Information Processing Systems (NIPS)*, Montreal, Canada, 2014.
- [12] 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, no. 310, pp. 213–222, 2018.
- [13] Y. Xie and T. Zhang, "Imbalanced learning for fault diagnosis problem of rotating machinery based on generative adversarial networks," in *Proceedings of the 28th 37th Chinese Control Conference (CCC)*, Wuhan, China, 2018.
- [14] X. Li, W. Zhang, and Q. Ding, "Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 7, pp. 5525–5534, 2019.
- [15] 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.
- [16] S. Shao, P. Wang, and R. Yan, "Generative adversarial networks for data augmentation in machine fault diagnosis," *Computers in Industry*, vol. 106, pp. 85–93, 2019.
- [17] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, and X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, pp. 171–178, 2016.
- [18] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mechanical Systems and Signal Processing*, vol. 100, pp. 439–453, 2018.
- [19] Z. Wei, P. Gaoliang, and L. Chuanhao, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, pp. 425–446, 2017.
- [20] Y. Lei, Z. He, Y. Zi, and X. Chen, "New clustering algorithm-based fault diagnosis using compensation distance evaluation technique," *Mechanical Systems and Signal Processing*, vol. 22, no. 2, pp. 419–435, 2008.
- [21] P. Konar and P. Chattopadhyay, "Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs)," *Applied Soft Computing*, vol. 11, no. 6, pp. 4203–4211, 2011.

- [22] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved Techniques for Training GANs," in *Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS)*, Barcelona, Spain, 2016.
- [23] V. D. M. Laurens, "Accelerating t-SNE using tree-based algorithms," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 3221–3245, 2014.