Hindawi
Journal of Sensors
Volume 2022, Article ID 5420478, 11 pages
https://doi.org/10.1155/2022/5420478

Research Article

A Novel Generative Method for Machine Fault Diagnosis

Zhipeng Dong,[1] Yucheng Liu,[2,3] Jianshe Kang,[1] and Shaohui Zhang[2]

[1]*Army Engineering University of PLA, Shijiazhuang, China*
[2]*School of Mechanical Engineering, Dongguan University of Technology, Dongguan 523808, China*
[3]*College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China*

Correspondence should be addressed to Shaohui Zhang; zhangsh@dgut.edu.cn

Received 15 November 2021; Accepted 17 December 2021; Published 11 January 2022

Academic Editor: Haidong Shao

Deep learning is widely used in fault diagnosis of mechanical equipment and has achieved good results. However, these deep learning models require a large number of labeled samples for training, which is difficult to obtain enough labeled samples in the actual production process. However, it is easier to obtain unlabeled samples in the industrial environment. To overcome this problem, this paper proposes a novel method to generative enough label samples for training deep learning models. Unlike the generative adversarial networks, which required complex computing time, the calculation of the proposed novel generative method is simple and effective. First, we calculate the Euclidean distance between the training sample and the test sample; then, the weight coefficient between the training sample and the test sample is settled to generate pseudosamples; finally, combine with the pseudosamples, the deep learning method is training for machine fault diagnosis. In order to verify the effectiveness of the proposed method, two experiment datasets with planetary gearboxes and wind gearboxes are carried out with different activation functions. Experimental results show that the proposed method is effective for most activation function models.

1. Introduction

With the continuous development of industrial intelligence, people are focusing on equipment health monitoring and fault diagnosis. Gearboxes are widely used in mechanical equipment, especially in large and complex equipment, and gearboxes are the main transmission device. Therefore, condition monitoring and fault diagnosis of gearboxes are very important [1, 2]. In the manufacturing industry, machine fault will directly affect machining accuracy and machining quality, reducing production efficiency [3]. Equipment status monitoring and fault diagnosis are essential to ensure the machine's normal operation, reduce maintenance costs, and improve production efficiency. Therefore, it is of great significance to conduct health monitoring and fault diagnosis of mechanical engineering equipment.

The data-driven fault diagnosis method usually includes three stages: (1) using sensors to obtain sample data; (2) denoising the sample data and extracting features; (3) inputting the extracted features into the classification algorithm for fault identification [4]. Deep learning is an effective method of fault diagnosis. In order to improve the accuracy of fault diagnosis, researchers have developed several deep learning networks. Such as recurrent neural networks (RNN) [5], autoencoder(AE) [6, 7], long short-term memory (LSTM) [8], deep belief network (DBN) [9], and convolutional neural network (CNN) [10], the advantage of these deep learning algorithms is to reduce feature redundancy and extract more information features for predictive models. Deep learning technology has shown good performance in fault detection and diagnosis with its powerful feature extraction ability and excellent classification performance, thus, it has become a research hotspot. Guo et al. [11] proposed an intelligent method based on deep belief network (DBN) and hyperparameter optimization for fault diagnosis of rolling bearings. Li et al. [12] proposed a deep autoencoder network for cross-machine fault diagnosis. Zhou et al. [13] designed a new generative confrontation network generator and discriminator, using a global optimization scheme to generate more discriminable fault samples. Ma

and Mao [14] developed a novel deep neural network, a convolution-based long short-term memory (CLSTM) network, to predict the remaining service life (RUL) of rotating machinery in field vibration data. Wang et al. [15] formed a new fault diagnosis method by integrating the deformable convolutional neural network (CNN), deep long short-term memory (DLSTM), and transfer learning strategies. Singh et al. [16] proposed a deep learning-based domain adaptation (DA) method for fault diagnosis of gearboxes when the speed changes significantly. Xue et al. [17] proposed an enhanced deep sparse autoencoder (ADSAE) for diagnosing gear pitting faults. He et al. [18] proposed an improved deep autoencoder (MDAE) driven by multisource parameters for cross-domain fault prediction.

Deep learning has a good effect on fault diagnosis; however, it requires a large number of condition samples in the training process to achieve satisfactory accuracy. The limited number of samples will directly lead to a reduction in model performance. Due to different working conditions, it is difficult and expensive to collect enough fault samples for training models in actual industrial production. Therefore, it is important to research the deep learning method for machine fault diagnosis in small samples situation. Li et al. [19] proposed a fusion framework based on the confidence weight support matrix machine (CWSMM) for strong interference and unbalanced data sets. Ti et al. [20] proposed a weighted extended neural network (W-ENN) model for fault diagnosis of a small piece of steam turbine generator sets. He et al. [21] suggested using depth transfer multiwavelet autoencoder to diagnose the gearbox fault with a few training samples. Zhang et al. [22] used generative adversarial networks to learn the mapping between noise distribution and actual mechanical time vibration data and then generate similar fake samples to expand further the available data set for fault diagnosis. Xiao et al. [23] proposed a fault diagnosis framework using an improved TrAdaBoost algorithm and a convolutional neural network for a small amount of target data for transfer learning. Meng et al. [24] proposed a data enhancement method that divides a single sample into multiple cells and then reorganizes the cells to increase the number of data samples. Li et al. [25] proposed an enhanced generative adversarial network (EGAN), which uses a generator to generate specified samples and automatically enrich small sample data sets for fault diagnosis of rotating machinery. In order to solve the problem of unbalanced sample allocation, Zhang et al. [26] designed a weighted minority oversampling (WMO) comprehensive oversampling method.

Although the deep learning model has high diagnostic accuracy and prediction accuracy, it requires a large number of label samples for training. Otherwise, the diagnostic accuracy will be greatly reduced. Since several novel deep learning methods have been proposed to overcome the problem of the small sample, most of these methods are based on deep learning network architecture and require a lot of computing time. Therefore, a novel method with litter computing time has been proposed for the small sample problem. The main contributions of the paper contains (1) a novel, low-computing, and effective intelligent diagnosis method is proposed for small samples problem; (2) the proposed method calculates the Euclidean distance between a small label samples and a large number of unlabeled samples and generates pseudo samples with labels by a weight; (3) the proposed method is used for fault detection of planetary gearboxes, and the accuracy has been greatly improved.

The other sections of this paper are arranged as follows. The second section mainly introduces the theory of the proposed method. Sections 3 and 4 discuss the arrangement of experiments on planetary gearboxes and industrial robots and the analysis of the corresponding results. Finally, Section 5 introduces the relevant conclusions.

2. Methodology

For the small sample problem of fault diagnosis, we use unknown as training samples X_{train} and then use a number of label samples as test samples X_{test}. The training sample matrix is composed of $\boldsymbol{m} \times \boldsymbol{n}$ dimensional vectors, expressed as

$$X_{\text{train}} = [x_{11}, x_{12}, \cdots, x_{1n}; \cdots; x_{m1}, x_{m2}, \cdots, x_{mn}] X_{train} \in R^{\text{m}\times n}. \tag{1}$$

Similarly, the test sample matrix is composed of $\boldsymbol{f} \times \boldsymbol{n}$ dimensional vectors, expressed as

$$X_{\text{test}} = \left[x_{11}, x_{12}, \cdots, x_{1n}; \cdots; x_{f1}, x_{f2}, \cdots, x_{fn}\right] X_{test} \in R^{f\times n}. \tag{2}$$

By calculating the Euclidean distance between the test sample and the training sample and comparing the distance, find the training sample point closest to the test sample.

2.1. Novel Generative Method. Euclidean distance is the actual distance between two points in n-dimensional space or the distance from the point to the origin (that is, the length of the vector). The Euclidean distance obtained in two and three dimensions is equal to the actual distance between two points. It is the simplest and most direct method to calculate the distance between data sample points. Suppose two n-dimensional vectors $A = [a_1, a_2, \cdots, a_n]$ and $B = [b_1, b_2, \cdots, b_n]$, then, the Euclidean distance formula in n-dimensional space is

$$\text{Dist}_{(A,B)} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2}. \tag{3}$$

The above formula can calculate the Euclidean distance between the vectors, extended from the Euclidean distance between the vectors to the Euclidean distance between the matrices. Let $i \times n$ dimensional matrix $\boldsymbol{C}$ as $\boldsymbol{C} = [x_{11}, x_{12}, \cdots, x_{1n}; \cdots; x_{i1}, x_{i2}, \cdots, x_{\text{in}}]$, $j \times n$ dimensional matrix D as $\boldsymbol{D} = [y_{11}, y_{12}, \cdots, y_{1n}; \cdots; y_{j1}, y_{j2}, \cdots, y_{jn}]$. Calculating the Euclidean distance of matrix $\boldsymbol{C}$ and matrix $\boldsymbol{D}$ will get distance matrix E. The formula for calculating Euclidean distance between matrices is

$$\text{Dist}_{c_{ij}} = \sqrt{\left(x_{i1} - y_{j1}\right)^2 + \left(x_{i2} - y_{j2}\right)^2 + \cdots + \left(x_{\text{in}} - y_{jn}\right)^2}. \tag{4}$$

The above formula can be found that each row vector of matrix $\boldsymbol{C}$ has been calculated j times (the number of rows in D). Therefore, the calculation of the Euclidean distance between the matrices relies on the measure of the Euclidean distance between the vectors, so the dimension of the matrix E is $i \times j$. After calculating the Euclidean distance of the data sample points, compare the distance between the sample points. Find the test sample point $\text{Dist}_{\min}$ with the smallest Euclidean distance from the training sample in the test sample and then generate the label sample by calculating the following formula.

$$\boldsymbol{X}_{\text{new}} = (\text{dist}_{\min}, :) \times \boldsymbol{k} + \boldsymbol{X}_{\text{test}}. \tag{5}$$

In the formula, X_{new} represents a new sample, and k represents a scale factor. We automatically classify sample points with sufficiently small distance into one category, generate a unique sample point at the distance k between the two sample points with the shortest distance, and expand the data sample by this method in Figure 1.

2.2. SAE-Based Network Model

2.2.1. Autoencoder (AE). Autoencoders were used for dimensionality reduction processing and feature learning of high-dimensional complex data, which has a positive effect on the development of deep learning neural networks. The autoencoders use unsupervised neural network learning methods to learn unlabeled raw data and extract low-dimensional data features of high-dimensional complex data. The network structure of the autoencoder is shown in Figure 2. It is composed of three layers of neural networks, namely, the input layer, the hidden layer, and the output layer. The hidden layer means that the high-dimensional data is processed to obtain the low-dimensional data features. The output layer has the same number of nodes as the input layer, which means that the input and output data dimensions are the same. The autoencoder aims to reconstruct its input; that is, it uses the backpropagation algorithm to make the output equal to the input as much as possible. A function that the self-encoding neural network tries to learn is

$$y_{w,b}(x) \approx x. \tag{6}$$

The autoencoder tries to approximate an identity function so that the output $y(i)$ approximates the input $x(i)$. By minimizing the reconstruction error, the input data can be reconstructed as much as possible in the output layer, thereby exerting the unsupervised learning effect and effectively extracting low-dimensional data features. The autoencoder method has been widely used in fault diagnosis.

The autoencoders are composed of a three-layer neural network, which can also be seen as composed of two parts: encoder and decoder. The encoder consists of an input layer and a hidden layer. Through training, the input layer data x is encoded and converted into a deterministic mapping of the feature form h of the hidden layer. It can perform affine mapping and nonlinear mapping. The coding network is defined as

$$\boldsymbol{h} = f_\theta(\boldsymbol{w}_1 x + \boldsymbol{b}_1), \tag{7}$$

where f_θ is the activation function in the coding network, w_1 is the weight vector of the coding stage, b_1 is the offset vector of the coding stage, and $\theta = \{w_1, b_1\}$ is the trainable parameter set of the encoder and decoder. Then, in the decoding stage, the decoder consists of a hidden layer and an output layer. The decoding network maps the feature h of the hidden layer to the input layer, reconstructs the input data x, and obtains the output layer data with the same dimensions. Similarly, the decoding process can be defined as

$$x' = y_{w,b}(x) = f'_\theta(\boldsymbol{w}_2 h + \boldsymbol{b}_2). \tag{8}$$

The vector of approximate input data is reconstructed by the output layer, the activation function in the decoding network, w_2 is the weight vector of the decoding stage, b_2 is the

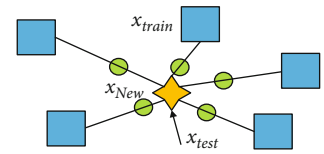

Figure 1: Pseudosample generation.

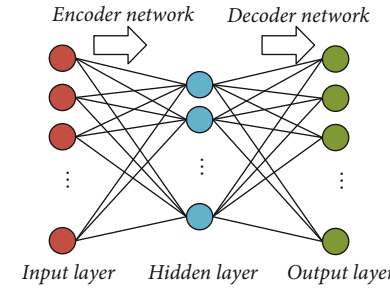

Figure 2: The structure of an AE neural network.

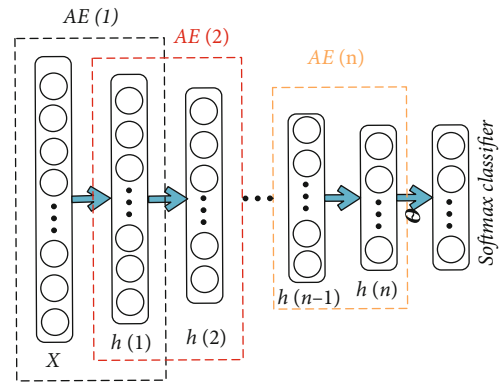

Figure 3: The structure of the SAE network.

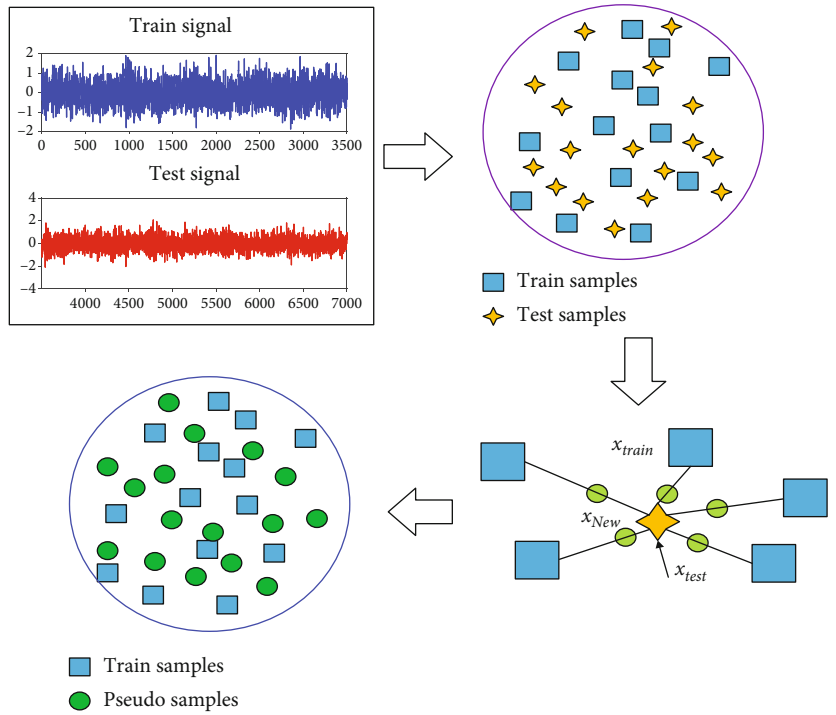

Figure 4: The overall proposed fast generative samples structure.

offset vector of the decoding stage, and $\theta = \{w_2, b_2\}$ is the trainable parameter set of the encoder and decoder. The reconstruction error $J(w, b; x, x')$ between the output x' and the input x is expressed as

$$J\left(\boldsymbol{w}, \boldsymbol{b}; \boldsymbol{x}, \boldsymbol{x}'\right) = \frac{1}{2}\left\|\boldsymbol{x} - \boldsymbol{x}'\right\|^2. \tag{9}$$

In the training process, given a training set of m samples, we define the total cost function as

$$J(\boldsymbol{w}, \boldsymbol{b}) = \left[\frac{1}{m}\sum_{i=1}^{m} J\left(\boldsymbol{w}, \boldsymbol{b}; \boldsymbol{x}, \boldsymbol{x}'\right)\right] + \lambda T, \tag{10}$$

where $J(w, b)$ represents the total cost function of the entire data set, the last term is the average sum of square error, λT is the regularization term (also called the weight penalty term), and the weight penalty parameter λ is used to limit the weight in order to achieve the purpose of preventing overfitting.

2.2.2. SAE-Based Network. The SAE network structure is composed of multiple autoencoder structure hierarchically, and there is a classifier in the output layer. The SAE network structure is shown in Figure 3.

The AE network has the problem of extracting feature redundancy. In order to solve this problem, regularization constraints are introduced in the SAE network, and constraints are added to the hidden layer neurons. In the encoding and decoding process, in order to make the hidden layer sparse, we need to add constraints to each hidden layer.

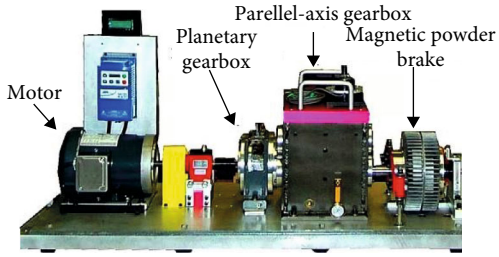

Figure 5: The drivetrain diagnostics simulator.

Only a few neuron nodes are active. The average activation degree of the hidden layer neural unit node j can be described as

$$\boldsymbol{\rho_j} = \frac{1}{m}\sum_{i=1}^{m}\left[h_j(x(i))\right]. \tag{11}$$

The average activation degree of the hidden layer neuron nodes generally approaches 0; that is, most neuron nodes are disabled. Therefore, in order to ensure that the activation degree of each node is close to 0, additional penalty terms need to be added to the cost function, which is described by the mathematical formula as follows:

$$KL\left(\rho \| \rho_j\right) = \rho \log \frac{\rho}{\rho_j} + (1-\rho) \log \frac{1-\rho}{1-\rho_j}, \tag{12}$$

where ρ is the sparsity parameter, and $KL(.)$ is the Kulback-Leibler divergence used as a penalty metric between the

Table 1: Faulty patterns set in the experiments.

Faulty pattern	Fault type	Load condition	Faulty pattern	Fault type	Load condition
F1	Surface wear	0 nm	F11	Chipped tooth	2.8 nm
F2	Surface wear	1.4 nm	F12	Chipped tooth	5.2 nm
F3	Surface wear	2.8 nm	F13	Missing tooth	0 nm
F4	Surface wear	5.2 nm	F14	Missing tooth	1.4 nm
F5	Crack tooth	0 nm	F15	Missing tooth	2.8 nm
F6	Crack tooth	1.4 nm	F16	Missing tooth	5.2 nm
F7	Crack tooth	2.8 nm	F17	Normal	0 nm
F8	Crack tooth	5.2 nm	F18	Normal	1.4 nm
F9	Chipped tooth	0 nm	F19	Normal	2.8 nm
F10	Chipped tooth	1.4 nm	F20	Normal	5.2 nm

expected distribution and the actual distribution. The penalty term has the property that when $\rho_{j=}\rho, KL(\rho \| \rho_j) = 0$. In the SAE network, the sparse penalty is added to the cost function, which can be expressed by the following formula:

$$\boldsymbol{J}_{\text{sparse}}(w, b) = J(w, b) + \beta \sum_{j=1}^{s} KL\left(\rho \| \rho_j\right), \tag{13}$$

where β is the weight of the sparsity parameter. The parameters $\{w, b\}$ can be updated using a stochastic gradient descent algorithm.

2.3. Classical Activation Functions. The input datasets of the neural network are weighted and, applied to the activation function, which can enhance the nonlinearity of the entire network model. Since the activation function can introduce nonlinearity to neurons, the expressive ability of the network model becomes stronger. The Sigmoid function is a commonly used activation function, but it has the problem of vanishing gradient. In order to solve the problem of vanishing gradient, many activation functions have been developed. In order to verify the effectiveness of the proposed method, the following activation functions are used in the SAE model.

(1) According to the literature [27], the expression of Tanh function is

$$\text{Tanh}(\boldsymbol{x}) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}, \tag{14}$$

$$\text{Tanh}'(\boldsymbol{x}) = \frac{4}{\left(e^{x} + e^{-x}\right)^{2}}. \tag{15}$$

Tanh is a common activation function, and its value range is [-1, 1]. Compared with Sigmoid, its output mean value is 0, and the number of iterations is less. Moreover, the convergence speed is faster. Note that the Tanh has soft saturation and will cause the problem of gradient disappearance.

(2) According to the literature [28], the expression of RelTanh function is

$$\text{RelTanh}(\boldsymbol{x}) = \begin{cases} \text{Tanh}'\left(\lambda^{+}\right)\left(x - \lambda^{+}\right) + \text{Tanh}\left(\lambda^{+}\right) & x \geq \lambda^{+}, \\ \text{Tanh}(x) & \lambda^{-} < x < \lambda^{+}, \\ \text{Tanh}'(\lambda^{-})(x - \lambda^{-}) + \text{Tanh}(\lambda^{-}) & x \leq \lambda^{-}. \end{cases} \tag{16}$$

The RelTanh activation function is composed of a linear function at both ends and a nonlinear Tanh function at the middle. Therefore, it can solve the problems of the Tanh function and ReLU function. λ^{+} and λ^{-} are mainly used to constrain the learnable range of the slope and avoid the disappearance of the gradient. The initial threshold will be set to $\lambda^{+} = 0$ and $\lambda^{-} = -1$.

(3) According to the literature [29], the expression of ELU function is

$$\text{ELU}(\boldsymbol{x}) = \begin{cases} x & x \geq 0, \\ \alpha(\exp\,(x) - 1) & x < 0. \end{cases} \tag{17}$$

ELU is an improved version of the ReLU function. Through the parameter α, the output of the negative interval is no longer to zero. The output has a certain degree of anti-interference ability and enhances the robustness to noise; however, it still has gradient disappearance.

2.4. The Proposed Algorithm. This paper proposes a novel generative method to provide enough label samples for training deep learning models. By calculating the relationship between the label samples and their nearest unlabeled samples, the proposed method can generate enough pseudo-samples with labels. Combine with the pseudosamples, the deep learning method is training for machine fault diagnosis. The structure of proposed in this paper is shown in Figure 4 and summarized as follows.

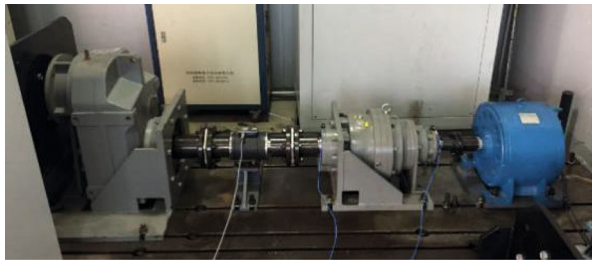

Figure 6: Wind gearboxes test rig.

Table 2: Faulty patterns set in the experiments.

Faulty pattern	Fault location	Degree of failure
A	Normal	Normal
B	Planet gear	Broken
C	Planet gear	Moderate
D	Ring left	Moderate
E	Sun gear	Broken

Table 3: The average accuracy of 100_train samples.

Samples Function	100	500	700	900
Reltanh	0.3827	0.3600	0.4984	0.5637
Tanh	0.3535	0.3942	0.3705	0.4228
ELU	0.3476	0.4142	0.3700	0.4561

(1) Collected the condition data set from machine experiments

(2) Divide data into training and testing data set

(3) Calculating the Euclidean distance between the test sample and the training sample, and comparing the distance

(4) Finding the training sample point closest to the test sample and generative the pseudosamples with labels

(5) Train SAE model with different activation functions

Fault classification using the trained SAE model and evaluate the accuracy.

3. Experiments

3.1. Fault Experiment of Planetary Gearbox. Design experiments to verify the validity of the proposed method. The experimental device used the power transmission fault diagnosis experiment platform (DDS) designed by SpectraQuest, as shown in Figure 5.

The failure of the transmission system is mainly caused by the wear of the tooth surface in the spur gear and the helical gear, the crack of the tooth surface, the pitting of the tooth surface, and the lack of teeth. Therefore, this experiment sets these four types of failures. Since the planetary gearbox's secondary sun gear has a relatively high probability of failure, this experiment's focus is to test the secondary sun gear of the planetary gearbox. In order to obtain a variety of vibration signals, by controlling the magnetic brake under four different load conditions (0 Nm, 1.4 Nm, 2.8 Nm, and 5.2 Nm), the vibration signals of normal conditions and four fault conditions are collected. A total of 20 groups of vibration signals, as shown in Table 1.

Table 4: The average accuracy of 200_train samples.

Samples Function	200	1000	1400	1800
Reltanh	0.4893	0.6313	0.61660	0.6691
Tanh	0.4662	0.6416	0.5314	0.6267
ELU	0.4779	0.4640	0.5371	0.6084

Table 5: The average accuracy of 300_train samples.

Samples Function	300	1500	2100	2700
Reltanh	0.5767	0.7334	0.7435	0.7834
Tanh	0.5082	0.6410	0.7318	0.7547
ELU	0.5671	0.6450	0.6907	0.8212

Table 6: The average accuracy of 400_train samples.

Samples Function	400	2000	2800	3600
Reltanh	0.7036	0.7355	0.7667	0.8500
Tanh	0.5928	0.5696	0.6685	0.8304
ELU	0.6548	0.5851	0.7455	0.8058

Table 7: The average accuracy of 500_train samples.

Samples Function	500	2500	3500	4500
Reltanh	0.7725	0.8266	0.8674	0.8675
Tanh	0.6695	0.7410	0.7589	0.8276
ELU	0.7809	0.7910	0.8004	0.8776

3.2. Fault Experiment of Wind Gearboxes. In order to further verify the effectiveness and versatility of the method, we use the wind gearboxes failure data set to test the method again. The experiment was carried out on the industrial robot experimental platform, as shown in Figure 6. The test bench consists of a 3.7 kW electric motor, a two-stage parallel gearbox with a speed increase ratio of 20, a 3 kW permanent magnet synchronous motor, and a 3 kW load box. In order to ensure the rationality of the collected signals, the sensor is installed on the bearing chock of the gearbox, and the sensor is connected to the computer to store the signal data. In this case, it can better reflect the advantages of the proposed method in adapting to different types of data sets under small sample conditions.

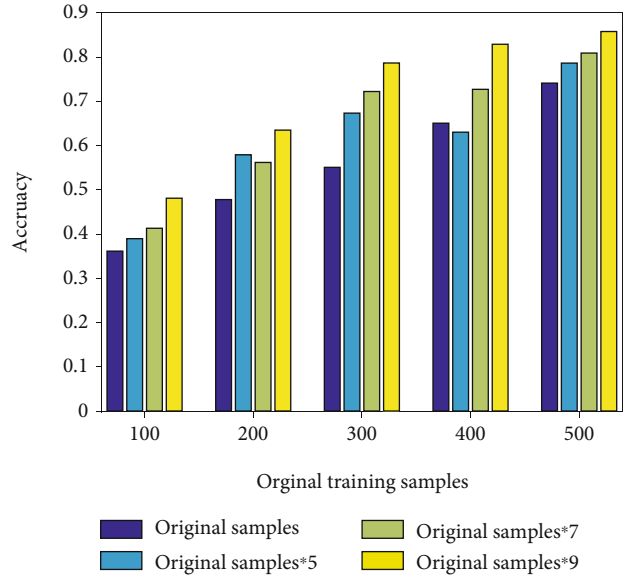

Figure 7: Average accuracies of all activation functions.

The gearbox is a key component to ensure the normal operation of the wind. The gearbox is mainly composed of the two-stage parallel gearbox. As the internal gear is prone to pitting and cracking failures, it will lead to reduced work efficiency. Therefore, in this experiment, the sun gear and planetary gears are tested, and a set of normal modes and four sets of failure modes are set. Table 2 shows the detailed description of the gear fault location and fault degree of the gear simulated in this experiment. The experiment is carried out under the condition that the motor speed is 300 r/min and no load.

4. Fault Diagnosis and Result Analysis

4.1. Fault Diagnosis and Result Analysis for Planetary Gearbox. By analyzing the fault diagnosis accuracy of each activation function in the experiment, the effectiveness of the method under the condition of small samples is verified. In order to avoid the problem of the disappearance of the gradient of the activation function, a 4-layer SAE network structure was constructed, and different activation functions were applied when experimenting with the gearbox. In order to ensure the universality and reliability of the experimental results of this method under small sample conditions, we conducted experiments based on five groups of different sample numbers. In the experiment, the basic numbers of samples are 100, 200, 300, 400, and 500. To ensure the validity of the experimental results, in the experiment, each group of basic samples used the method proposed in this paper to carry out three generate sample, and four independent experiments were carried out after each generate. The average diagnostic accuracy is a key measurable indicator that reflects the functional differences between activation functions and compares the performance of activation functions.

The proposed method is used to generate the label samples, and then deep learning models with different activation functions are used for fault diagnosis. The average accuracy of each activation function under different sample sizes is listed in Tables 3 to 7.

In the case of 100 train samples, the activation functions, as RelTanh, Tanh, and ELU, have improved diagnostic accuracy. When the train samples extended to 900, the highest accuracy of activation function is RelTanh, which exceeds 50% (as shown in Table 3). For the basic samples number is 200, the classification rate of the activation function is ranging from 46.62% to 48.93%. When the training samples are increasing to 1,800 by the proposed method, the accuracy of the activation functions is above 60%. As the number of train samples increases, the accuracy of overall function is significantly improved. When the train samples is 300 and without samples generation, the accuracy of the activation function range between 50.82% and 57.67%. Since the number of train samples becomes 2700 with generated method, the average diagnostic accuracy of the ELU function reaches 82.12%. In the case of the original train samples is 400, the diagnostic accuracy of the activation function is significantly improved, and the diagnostic accuracy of RelTanh and Tanh are exceeded 80%. For the original train samples is 500, the accuracy of the activation function is ranging from 66.95% to 78.09%. The accuracy of RelTanh, Tanh, and ELU are still the highest. From Tables 3 to 7, it can be shown that the average accuracy of RelTanh and Tanh functions is higher in these cases, since they overcome the vanishing gradient problem and the diagnostic accuracy will increase as the number of samples increases.

The average accuracies of each activation function with the proposed generate method are compared with the original method, as shown in Figure 7. Form the Figure 7, it shows that the fault diagnosis accuracy of activation function is significantly improved as increasing the training samples, which proves the effectiveness of the proposed method.

The ELU function has an average diagnostic accuracy rate of 87.76% when the sample base is expanded from 500 to 4500, which has an absolute advantage over the activation

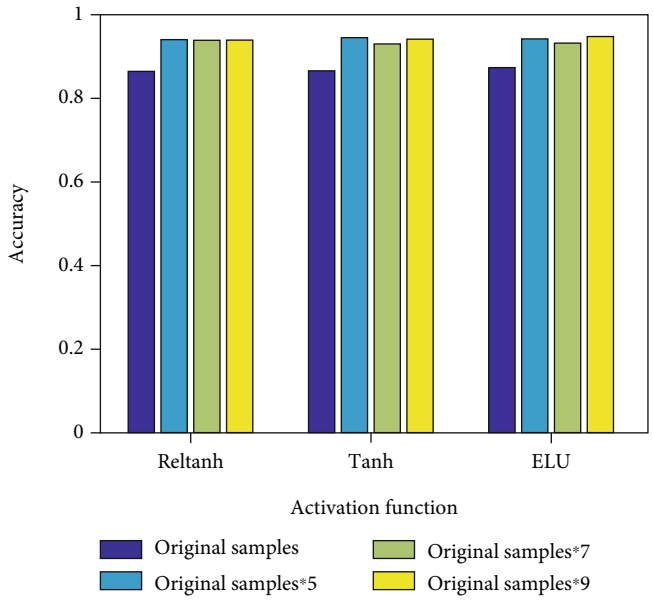

Figure 8: Average accuracies of all activation functions based on 100 train samples.

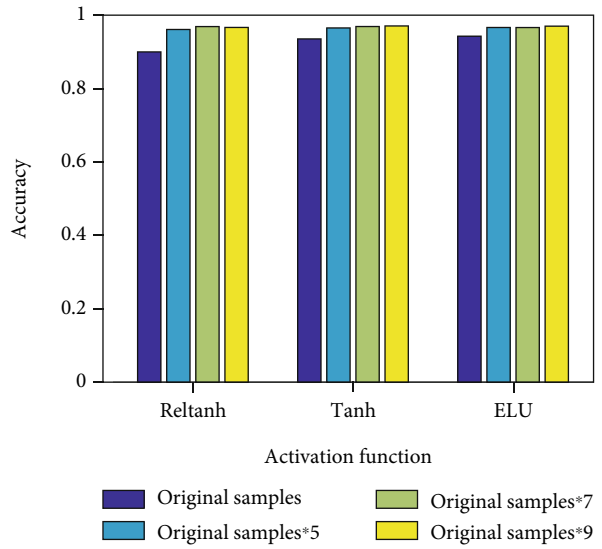

Figure 9: Average accuracies of all activation functions based on 200 train samples.

function of other functions. The ELU function negative interval is saturated, allowing negative abnormal input and avoiding a certain degree of gradient disappearance while having stronger noise robustness. In summary, the ELU, RelTanh, and Tanh function of our proposed sample expansion method perform best. It is worthy of further development and application in the case of small samples.

4.2. Fault Diagnosis and Result Analysis for Wind Gearboxes. Like the experiment introduced in Section A, we use the SAE network model in the wind gearboxes fault diagnosis experiment. In the experiment, this method is used to expand and train five groups of different numbers of samples. In order to ensure the validity and rationality of the results, three generate sample experiments were carried out for each group of samples, and four independent experiments were carried out for each group of generated the sample.

In the wind gearboxes fault diagnosis experiment, each activation function average diagnosis accuracy under different sample bases is shown in Figures 8 to 12. Experimental results show that the proposed method has certain effects in the case of small samples, and the activation function (Reltanh) has highest diagnostic accuracy. The diagnostic accuracy is increasing as generates training samples increases. For the original train samples is 100 (as shown in Figure 8), the accuracy of Reltanh is ranging from 0.8743 to 0.9511. When the sample base is 200, the diagnosis accuracy rates are ranging from 0.9009 to 0.9700, 0.9364 to

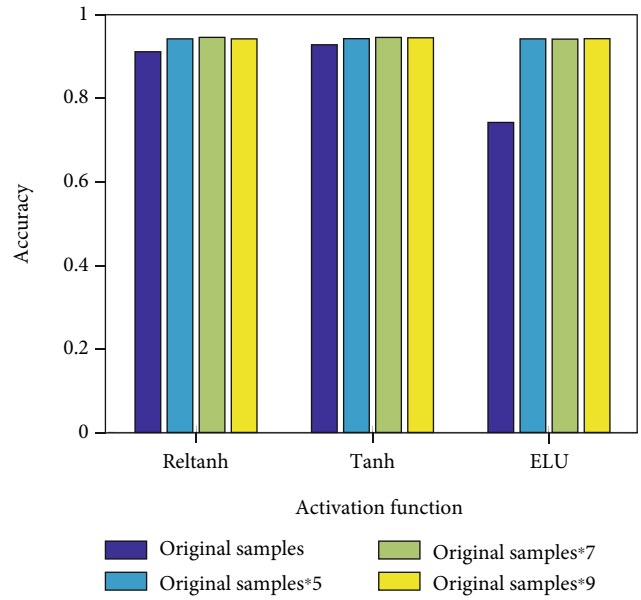

Figure 10: Average accuracies of all activation functions based on 300 train samples.

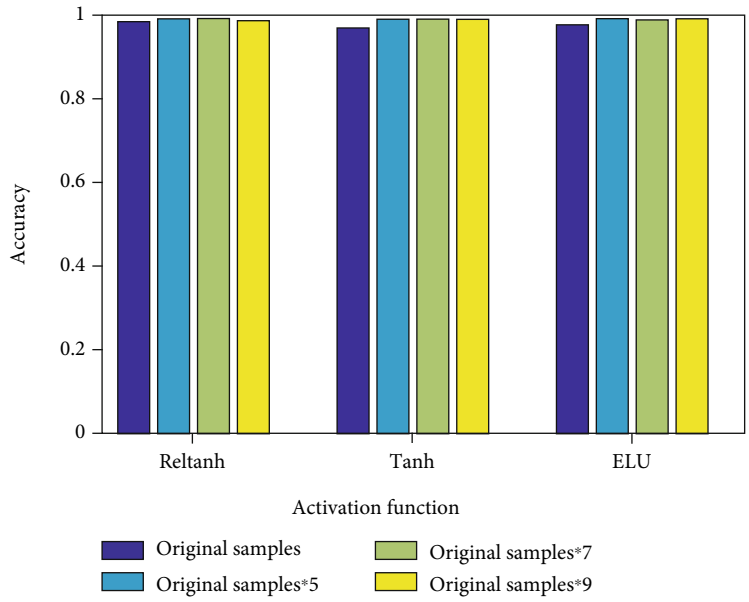

Figure 11: Average accuracies of all activation functions based on 400 train samples.

0.9717, and 0.9438 to 0.9711 for Reltanh, Tanh, and ELU, respectively. For the original samples is 300 (as shown in Figure 10), the accuracy rate is ranging between 0.7697 and 0.9626 without samples generated. However, the accuracy rates increase to more than 0.9766 under the proposed method.

In the case of the sample base is 400, the diagnosis accuracy rate of Reltanh is 97.56 ~ 98.30%. The Tanh function accuracy is increased by 2.09%, when the samples are increased to 2,800. When the basic sample number is 500, the diagnostic accuracy rate of the Reltanh function reaches to 98.50% (as shown in Figure 12).

From Figures 8–12, it can be seen that the Reltanh function has the highest fault diagnosis accuracy. As the sample size increases, the diagnostic accuracy of all activation functions has raised with different degrees. This result proved the effectiveness of the proposed method. In general, our method has strong stability in the case of small samples and has good results on different data sets with small training samples.

The average accuracies of each activation function with the proposed generate method are compared with the original method, as shown in Table 8. It should be noted that nine-time of the original train samples is generated by the

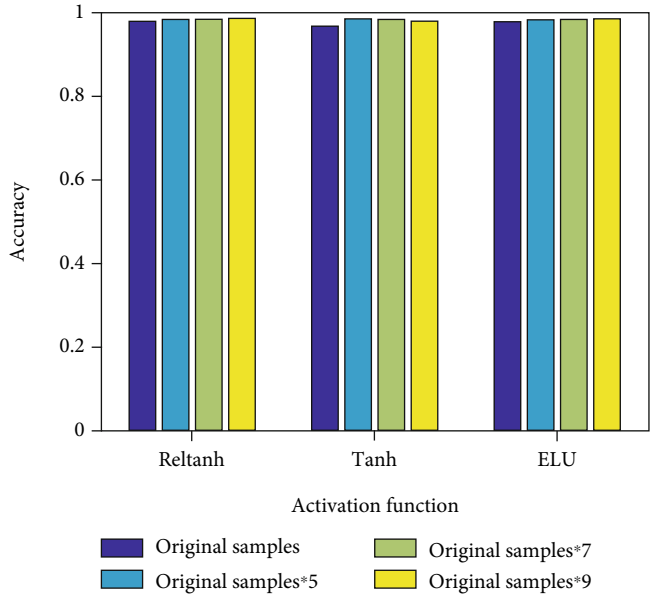

Figure 12: Average accuracies of all activation functions based on 500 train samples.

Table 8: Average accuraciesof all activation functions.

Training samples	The SAE-based DNNs	The proposed method
100	87.78%	95.04%
200	92.70%	96.82%
300	89.25%	97.84%
400	96.82%	98.14%
500	97.38%	98.26%

proposed method, and the average accuracies are calculated. The average accuracy of the proposed method is better than that of the SAE-based DNNs method for different training samples. The accuracy is increased by 0.88% to 8.59% for different training samples, when the samples are increased by the proposed method.

5. Conclusions

Deep learning has been widely used in fault diagnosis of mechanical equipment and has achieved ideal results. To overcome the problem of the small sample, this paper proposed a novel generative method to provide enough label samples for deep learning model. First, the Euclidean distance between a label sample and unlabeled samples is calculated, and the nearest samples are selected to generate new label samples. Then, the pseudosamples with label samples are altogether for training deep model. The results of fault diagnosis on planetary gearboxes and wind gearboxes show that the proposed method can greatly improve accuracy. It is worth noting that the effect of this method is related to the activation function. In the gearbox experiment, as the number of original samples increases, the proposed method has a good effect on most activation functions, and "ReLTanh," "Tanh," or "ELU" is the best activation function.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported in part by the Special projects in key fields of ordinary colleges and universities in Guangdong Province (new generation of information technology) (2020ZDZX3029) and Dongketong [2021] No. 55 Dongguan Science and Technology Commissioner Project (20201800500212, 20201800500282).

References

[1] Z. Feng, A. Gao, K. Li, and H. Ma, "Planetary gearbox fault diagnosis via rotary encoder signal analysis," *Mechanical Systems and Signal Processing*, vol. 149, article 107325, 2021.

[2] M. S. Raghav and R. B. Sharma, "A review on fault diagnosis and condition monitoring of gearboxes by using AE technique," *Archives of Computational Methods in Engineering*, vol. 28, no. 4, pp. 2845–2859, 2021.

[3] A. Kumar, C. Gandhi, Y. Zhou, R. Kumar, and J. Xiang, "Improved deep convolution neural network (CNN) for the identification of defects in the centrifugal pump using acoustic images," *Applied Acoustics*, vol. 167, article 107399, 2020.

[4] Z. Hu, Y. Wang, M. Ge, and J. Liu, "Data-driven fault diagnosis method based on compressed sensing and improved multiscale network," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 4, pp. 3216–3225, 2020.

[5] J. Wang and C. Zhang, "Software reliability prediction using a deep learning model based on the RNN encoder-decoder,"

Reliability Engineering & System Safety, vol. 170, pp. 73–82, 2018.

[6] H. Zhu, J. Cheng, C. Zhang, J. Wu, and X. Shao, "Stacked pruning sparse denoising autoencoder based intelligent fault diagnosis of rolling bearings," *Applied Soft Computing*, vol. 88, article 106060, 2020.

[7] J. Liu, K. Zhou, C. Yang, and G. Lu, "Imbalanced fault diagnosis of rotating machinery using autoencoder-based supergraph feature learning," *Frontiers of Mechanical Engineering*, pp. 1–11, 2021.

[8] J. Lei, C. Liu, and D. Jiang, "Fault diagnosis of wind turbine based on long short-term memory networks," *Renewable Energy*, vol. 133, pp. 422–432, 2019.

[9] C. Shen, J. Xie, D. Wang, X. Jiang, J. Shi, and Z. Zhu, "Improved hierarchical adaptive deep belief network for bearing fault diagnosis," *Applied Sciences*, vol. 9, no. 16, p. 3374, 2019.

[10] Z. Chen, A. Mauricio, W. Li, and K. Gryllias, "A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks," *Mechanical Systems and Signal Processing*, vol. 140, article 106683, 2020.

[11] C. Guo, L. Li, Y. Hu, and J. Yan, "A deep learning based fault diagnosis method with hyperparameter optimization by using parallel computing," *IEEE Access*, vol. 8, pp. 131248–131256, 2020.

[12] X. Li, X.-D. Jia, W. Zhang, H. Ma, Z. Luo, and X. Li, "Intelligent cross-machine fault diagnosis approach with deep autoencoder and domain adaptation," *Neurocomputing*, vol. 383, pp. 235–247, 2020.

[13] F. Zhou, S. Yang, H. Fujita, D. Chen, and C. Wen, "Deep learning fault diagnosis method based on global optimization GAN for unbalanced data," *Knowledge-Based Systems*, vol. 187, article 104837, 2020.

[14] M. Ma and Z. Mao, "Deep convolution-based LSTM network for remaining useful life prediction," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 1658–1667, 2021.

[15] Z. Wang, Q. Liu, H. Chen, and X. Chu, "A deformable CNN-DLSTM based transfer learning method for fault diagnosis of rolling bearing under multiple working conditions," *International Journal of Production Research*, vol. 59, pp. 1–15, 2021.

[16] J. Singh, M. Azamfar, A. Ainapure, and J. Lee, "Deep learning-based cross-domain adaptation for gearbox fault diagnosis under variable speed conditions," *Measurement Science and Technology*, vol. 31, no. 5, article 055601, 2020.

[17] L. Xueyi, L. Jialin, Q. Yongzhi, and H. David, "Semi-supervised gear fault diagnosis using raw vibration signal based on deep learning," *Chinese Journal of Aeronautics*, vol. 33, no. 2, pp. 418–426, 2020.

[18] Z. He, H. Shao, Z. Ding, H. Jiang, and J. Cheng, "Modified deep autoencoder driven by multisource parameters for fault transfer prognosis of aeroengine," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 1, pp. 845–855, 2022.

[19] X. Li, J. Cheng, H. Shao, K. Liu, and B. Cai, "A fusion CWSMM-based framework for rotating machinery fault diagnosis under strong interference and imbalanced Case," *IEEE Transactions on Industrial Informatics*, p. 1, 2021.

[20] W. Tichun, W. Jiayun, W. Yong, and X. Sheng, "A fault diagnosis model based on weighted extension neural network for turbo- generator sets on small samples with noise," *Chinese Journal of Aeronautics*, vol. 33, no. 10, pp. 2757–2769, 2020.

[21] Z. He, H. Shao, P. Wang, J. J. Lin, J. Cheng, and Y. Yang, "Deep transfer multi-wavelet auto-encoder for intelligent fault diagnosis of gearbox with few target training samples," *Knowledge-Based Systems*, vol. 191, article 105313, 2020.

[22] W. Zhang, X. Li, X.-D. Jia, H. Ma, Z. Luo, and X. Li, "Machinery fault diagnosis with imbalanced data using deep generative adversarial networks," *Measurement*, vol. 152, article ???, 2020.

[23] D. Xiao, Y. Huang, C. Qin, Z. Liu, Y. Li, and C. Liu, "Transfer learning with convolutional neural networks for small sample size problem in machinery fault diagnosis," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 233, no. 14, pp. 5131–5143, 2019.

[24] Z. Meng, X. Guo, Z. Pan, D. Sun, and S. Liu, "Data segmentation and augmentation methods based on raw data using deep neural networks approach for rotating machinery fault diagnosis," *IEEE Access*, vol. 7, pp. 79510–79522, 2019.

[25] Q. Li, L. Chen, C. Shen, B. Yang, and Z. Zhu, "Enhanced generative adversarial networks for fault diagnosis of rotating machinery with imbalanced data," *Measurement Science and Technology*, vol. 30, no. 11, article 115005, 2019.

[26] Y. Zhang, X. Li, L. Gao, L. Wang, and L. Wen, "Imbalanced data fault diagnosis of rotating machinery using synthetic oversampling and feature learning," *Journal of Manufacturing Systems*, vol. 48, pp. 34–50, 2018.

[27] E. Fan, "Extended tanh-function method and its applications to nonlinear equations," *Physics Letters A*, vol. 277, no. 4-5, pp. 212–218, 2000.

[28] X. Wang, Y. Qin, Y. Wang, S. Xiang, and H. Chen, "ReLTanh: an activation function with vanishing gradient resistance for SAE- based DNNs and its application to rotating machinery fault diagnosis," *Neurocomputing*, vol. 363, pp. 88–98, 2019.

[29] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," https://arxiv.org/abs/1511.07289, 2015.