IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 15, 2020, accepted August 2, 2020, date of publication August 5, 2020, date of current version August 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3014340

A Novel Data Augmentation Method for Intelligent Fault Diagnosis Under Speed Fluctuation Condition

XIAOYU WANG[1], ZHENYUN CHU[1], BAOKUN HAN[1], JINRUI WANG[1], GUOWEI ZHANG[1], AND XINGXING JIANG[2]

[1]College of Mechanical and Electronic Engineering, Shandong University of Science and Technology, Qingdao 266590, China
[2]School of Rail Transportation, Soochow University, Suzhou 215006, China

Corresponding author: Zhenyun Chu (chuzhenyun@sdust.edu.cn)

This work was supported by the China Postdoctoral Science Foundation under Grant 2019M662399.

ABSTRACT The problem of insufficient datasets has long been a hot topic in the field of prognosis and health management of rotary machines. Generative adversarial network (GAN) and other data augmentation algorithms can solve the problem of insufficient samples. However, the premise of the above method is the signal collected at a constant speed rather than at large speed fluctuation. To deal with data augmentation under large speed fluctuation, this article proposes an effective deep learning method, namely, domain adaptive efficient sub-pixel network (DAESPN). The core idea of DAESPN is to enhance the resolution of the original sample for data augmentation. The DAESPN framework is implemented as follows: after the data passes through the fully connected neural network, the multi-feature maps of the four channels are outputted. A group of high resolution (HR) features is obtained through the sub-pixel fully connected layer. In addition, maximum mean discrepancy (MMD) and mean square error (MSE) are used to construct the loss function of the model. Experimental results of gearbox and bearing datasets show that the DAESPN model has strong feasibility to carry out data augmentation for fault diagnosis of rotating machines under speed fluctuation condition. In addition, the feature learning process of DAESPN is visually displayed and analyzed.

INDEX TERMS Data augmentation, fault diagnosis, large speed fluctuation, signal resolution enhancement, maximum mean discrepancy, domain adaptation.

I. INTRODUCTION

Rotating parts, which are the core components of machines, directly affect their operation and have a huge effect on the performance, stability, and life of the mechanism [1]. The traditional machinery industry is rapidly changing toward automation and intelligence, and the application of intelligent fault diagnosis methods has increased and matured [2], [3]. In general, sensors are used to collect the vibration signals of a constant-speed rotating mechanism and apply intelligent methods to identify fault types [4], [5]. However, the speed of rotating parts of mechanical transmission fluctuates due to the influence of working conditions and loads [6]. For example, in the cutting process of computer numerical control machine tools, a variable speed operation mode is used to prevent the occurrence of chatter [7]. The speed of the gearbox in the wind turbine will change continuously with changing wind speed and wind direction [8], [9]. The speed of the top-drive transmission system in an oil drilling machine fluctuates considerably during the working process [10]. In summary, speed fluctuation widely occurs in mechanical equipment in different fields.

Intelligent fault diagnosis of mechanical transmission rotating parts under high-speed fluctuation has gained increasing attention. Xue *et al.* [11] proposed an incipient fault detection method for bearings under varying speed conditions; a novel alarm trigger mechanism is designed by constructing selected negative log-likelihood probability (SNLLP) health indicator and used to detect incipient fault. An *et al.* [12] built a three-dimensional numerical simulation

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

model by using advantage finite element method (FEM) for accurate diagnosis under rotating-speed fluctuations. Han *et al.* [13] set up a regularized sparse filtering model for intelligent fault diagnosis under large speed fluctuation. Some researchers [14]–[16] effectively used sequential tracking to eliminate the effects of speed fluctuations.

Speed diversity and change rate under speed fluctuations condition result in a huge amount of data to train a better model. When the training data set is small, changes in the intrinsic features of the sample are limited and the performance of the neural network will decline in most cases due to the sparse nature of the data [17]. At present, few studies were conducted on intelligent fault diagnosis of large speed fluctuation under insufficient samples, and this topic needs further attention. Resolution enhancement technology is widely used in the field of image processing. Dong *et al.* [18] proposed super-resolution (SR) method using deep convolutional neural network (CNN) to take the low resolution (LR) image as the input and the high resolution (HR) one as the output. Kim *et al.* [19] proposed an image super-resolution (SR) method by using a deeply recursive convolutional network (DRCN). Ledig *et al.* [20] presented super-resolution generative adversarial network (SRGAN), a generative adversarial network (GAN) for image SR; this framework is the first that can infer photo-realistic natural images for 4× upscaling factors. Shi *et al.* [21]proposed efficient sub-pixel convolutional neural network (ESPCN), which learned an array of upscaling filters to upscale the final LR feature maps into the HR output. Based on the ESPCN framework, this article proposes an effective neural network framework, namely, domain adaptive efficient sub-pixel network (DAESPN), for data augmentation under large speed fluctuation. The proposed method was verified using the stacked autoencoder (SAE) [22] classification network.

The main contributions of this article are summarized as follows:

(1) This article attempts to use deep learning to improve the resolution of samples for data augmentation of insufficient samples under large speed fluctuation.

(2) The maximum mean discrepancy (MMD) [23], which is widely used in domain adaptation, is used in the sub-pixel fully connected layer of the proposed framework to ensure the similarity and stability of the generated features.

(3) At the end of the proposed framework, a sub-pixel fully connected layer is used for upsampling, and SAE is used to verify the performance of the reconstructed data obtained by this method.

The rest of this article is organized as follows. In Section II, the theoretical backgrounds of MMD and ESPCN are introduced. In Section III, the proposed method is introduced in detail. In Section IV, the two diagnosis cases of gear and bearing datasets are investigated using the proposed method. Conclusions are drawn in the final section.

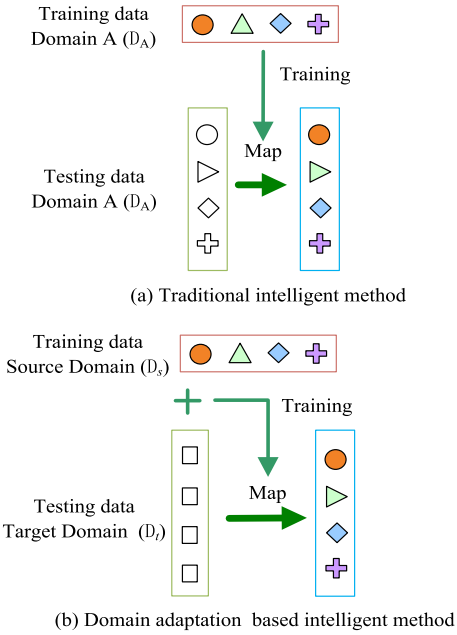


FIGURE 1. Intelligent learning systems.

II. THEORETICAL BACKGROUNDS

A. DOMAIN ADAPTATION

As shown in Fig. 1, traditional machine learning often requires training and testing data from the same domain. Domain adaptation, however, allows the distributions of the training and testing data to be different and is a representative method in transfer learning. This method uses information-rich source domain samples to improve the performance of the target domain model and is widely applied to intelligent fault diagnosis of mechanical equipment. The problem of domain adaptation has two crucial concepts: source domain and target domain. The source domain has a wealth of labeled information but has a different domain from the test sample. The target domain indicates the domain where the test sample is located, with no or only few tags. The source and target domains often belong to the same type of task but has different distribution.

The background and purpose of domain adaptation is converted into the following mathematical expressions to facilitate problem description.

(1) $\boldsymbol{X}$ is a data space, and $P(\boldsymbol{X})$ is a marginal probability distribution. Therefore, $\{\boldsymbol{X}, P(\boldsymbol{X})\}$ represents that the dataset $\boldsymbol{X}$ belongs to the data space $\boldsymbol{X}$ and follows the distribution $P(\boldsymbol{X})$. The problem to be solved in domain adaptation is that source domain $\boldsymbol{Ds}$ and target domain $\boldsymbol{Dt}$ have different data spaces and marginal distributions, that is, $\boldsymbol{Xs} \neq \boldsymbol{Xt}$ and $\mathrm{P}(\boldsymbol{Xs}) \neq \mathrm{P}(\boldsymbol{Xt})$.

(2) The label space for the source and target domains is the same, that is, $\boldsymbol{Ys} = \boldsymbol{Yt}$.

(3) The method of marginal probability distribution adaptation in domain adaptation is to find a map F only with labeled samples from $\boldsymbol{Ds}$ and unlabeled samples from $\boldsymbol{Dt}$ for satisfying $\mathrm{P}(\mathrm{F}(\boldsymbol{Xs})) = \mathrm{P}(\mathrm{F}(\boldsymbol{Xt}))$ and $\mathrm{P}(\boldsymbol{Ys}|\mathrm{F}(\boldsymbol{Xs})) = \mathrm{P}(\boldsymbol{Yt}|\mathrm{F}(\boldsymbol{Xt}))$, where $\mathrm{P}(\boldsymbol{Y}|F(\boldsymbol{X}))$ is the conditional probability distribution.

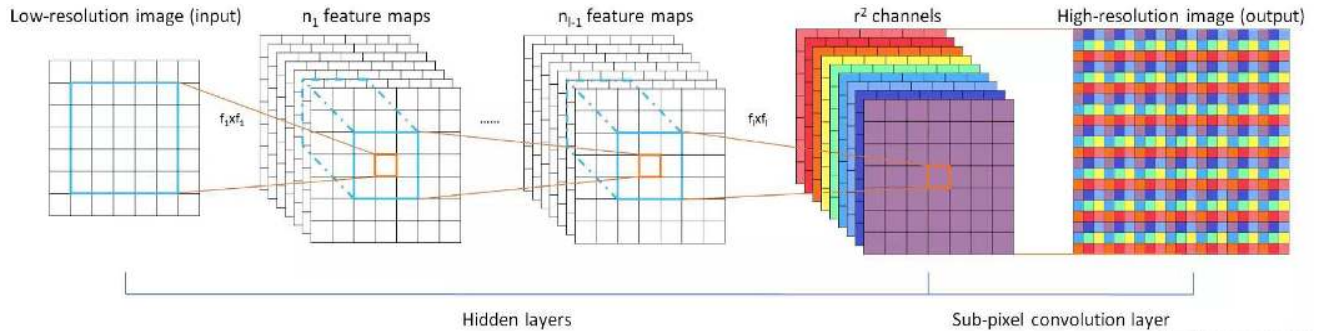


FIGURE 2. Structure of ESPCN.

B. MAXIMUM MEAN DISCREPANCY

MMD is a loss function widely used in transfer learning, especially in domain adaptation. As an effective criterion for comparing distributions, MMD can measure the difference in probability distribution between two samples. Given the two probability distributions p and q in $\boldsymbol{X}$, MMD is defined as:

$$\mathrm{MMD}\left(\Phi, p, q\right) = \sup_{f\in \mathrm{F}} \left(\mathrm{E}_{x_1\sim p}\left[f\left(x_1\right)\right] - \mathrm{E}_{x_2\sim p}\left[f\left(x_2\right)\right]\right) \quad (1)$$

where Φ is a class of functions $f: \boldsymbol{X} \to \boldsymbol{H}$. $\boldsymbol{H}$ means reproducing kernel Hilbert space (RKHS) [24].

$\{\boldsymbol{x}(i)\, s\}_{i=1,\ldots,ns}$ and $\{\boldsymbol{x}(j)\, t\}_{j=1,\ldots,nt}$ are data vectors drawn from distributions $\boldsymbol{D}s$ and $\boldsymbol{D}t$ on the data space $\boldsymbol{X}$, respectively. Based on the fact that f is in the unit ball in a universal RKHS, the empirical estimate of MMD can be rewritten as follows:

$$\mathrm{MMD}_e\left(x_s, x_t\right) = \left\| \frac{1}{n_s}\sum_{i=1}^{n_s}\phi\left(x_s^{(i)}\right) - \frac{1}{n_t}\sum_{j=1}^{n_t}\phi\left(x_t^{(j)}\right)\right\|_{\mathrm{H}} \quad (2)$$

where $\phi(\cdot): \boldsymbol{X} \to \boldsymbol{H}$ is referred to as the feature space map.

The kernel method is used to calculate the distribution distance of high-level learned features in different domains [25]. Furthermore, the practical computation of MMD is written as:

$$\begin{aligned} &D\left(x_s, x_t\right) \\ &= \left(\frac{1}{n_s^2}\sum_{i=1}^{n_s}\sum_{j=1}^{n_s}\left\langle\phi\left(x_s^{(i)}\right), \phi\left(x_s^{(j)}\right)\right\rangle + \frac{1}{n_t^2}\sum_{i=1}^{n_t}\sum_{j=1}^{n_t}\left\langle\phi\left(x_t^{(i)}\right),\right.\right. \\ &\quad \left.\left.\phi\left(x_t^{(j)}\right)\right\rangle - \frac{2}{n_s n_t}\sum_{i=1}^{n_s}\sum_{j=1}^{n_t}\left\langle\phi\left(x_s^{(i)}\right), \phi\left(x_t^{(j)}\right)\right\rangle\right)^{\frac{1}{2}} \\ &= \left(\frac{1}{n_s^2}\sum_{i=1}^{n_s}\sum_{j=1}^{n_s}k\left(x_s^{(i)}, x_s^{(j)}\right) + \frac{1}{n_t^2}\sum_{i=1}^{n_t}\sum_{j=1}^{n_t}k\left(x_t^{(i)}, x_t^{(j)}\right)\right. \\ &\quad \left. - \frac{2}{n_s n_t}\sum_{i=1}^{n_s}\sum_{j=1}^{n_t}k\left(x_s^{(i)}, x_t^{(j)}\right)\right)^{\frac{1}{2}} \end{aligned} \quad (3)$$

where $\boldsymbol{D}(x_s, x_t)$ is the unbiased estimation of $\mathrm{MMDe}(x_s, x_t)$. $k(\cdot, \cdot)$ is a kernel function that could compute the inner product in a higher dimensional space, i.e., $k(x, y) = <\phi(x), \phi(y)>$.

C. EFFICIENT SUB-PIXEL CONVOLUTIONAL NEURAL NETWORK

The network structure of efficient sub-pixel convolutional neural network (ESPCN) is shown in Fig. 2. ESPCN is mainly composed of two convolution layers for feature map extraction and a sub - pixel convolution layer. L layer convolutional neural network directly to the LR image, and a sub-pixel convolution layer is applied to upscale the LR feature maps to produce $\boldsymbol{I}_{SR}$. For a network composed of L layers, the first l-1 layers can be described as follows:

$$f_1\left(\boldsymbol{I}_{LR}; \boldsymbol{W}_1; b_1\right) = \varphi\left(\boldsymbol{W}_1 * \boldsymbol{I}_{LR} + b_1\right) \quad (4)$$

$$\boldsymbol{f}_l\left(\boldsymbol{I}_{LR}; \boldsymbol{W}_{1:l}; b_{1:l}\right) = \varphi\left(\boldsymbol{W}_l * \boldsymbol{f}_{l-1}\left(\boldsymbol{I}_{LR}\right) + b_l\right) \quad (5)$$

where $\boldsymbol{W}_l, b_l, l \in (1, L-1)$ are learnable weights and offsets, $\boldsymbol{W}_l$ is a 2D convolution tensor with the size of $n_{l-1} \times n_l \times k_l \times k_l$, where n_l is the feature numbers at layer l, k_l is the convolution kernel number at layer l, offset b_l is the vector with the length of n_l, and activation function φ is applied element-wise and is fixed.

The sub-pixel convolution layer consists of two parts, a convolution layer and subsequent arrangement of pixels. The convolution layer partially outputs r^2 channel feature maps, where r is the upscaling ratio. The last layer f_L has to convert the LR feature maps to HR image $\boldsymbol{I}_{SR}$ in the following form:

$$\boldsymbol{I}_{SR} = f_L\left(\boldsymbol{I}_{LR}\right) = PS\left(\boldsymbol{W}_L * f_{L-1}\left(\boldsymbol{I}_{LR}\right) + b_L\right) \quad (6)$$

where PS is a periodic shuffling operator that rearranges the elements with shape $H \times W \times C \cdot r^2$ to shape $rH \times rW \times C$, $\boldsymbol{I}_{LR}$ and $\boldsymbol{I}_{SR}$ have C color channels. PS is important to insert LR features into the HR image periodically, and the mathematical description is as follows:

$$PS(T)_{x,y,c} = T_{\lfloor x/r\rfloor, \lfloor y/r\rfloor, c.r.\mathrm{mod}(y/r)+c.\mathrm{mod}(x/r)} \quad (7)$$

where x, y are the output pixel coordinates in the HR space.

Pixel-wise mean squared error (MSE) is employed as an objective function to train the network, and the formula is as follows:

$$\xi = \left(\boldsymbol{W}_{1:L}, b_{1:L}\right) = \frac{1}{r^2 HW}\sum_{x=1}^{rH}\sum_{y=1}^{rW}\left(\boldsymbol{I}_{HR}^{x,y} - f_L^{x,y}\left(\boldsymbol{I}_{LR}\right)\right)^2 \quad (8)$$

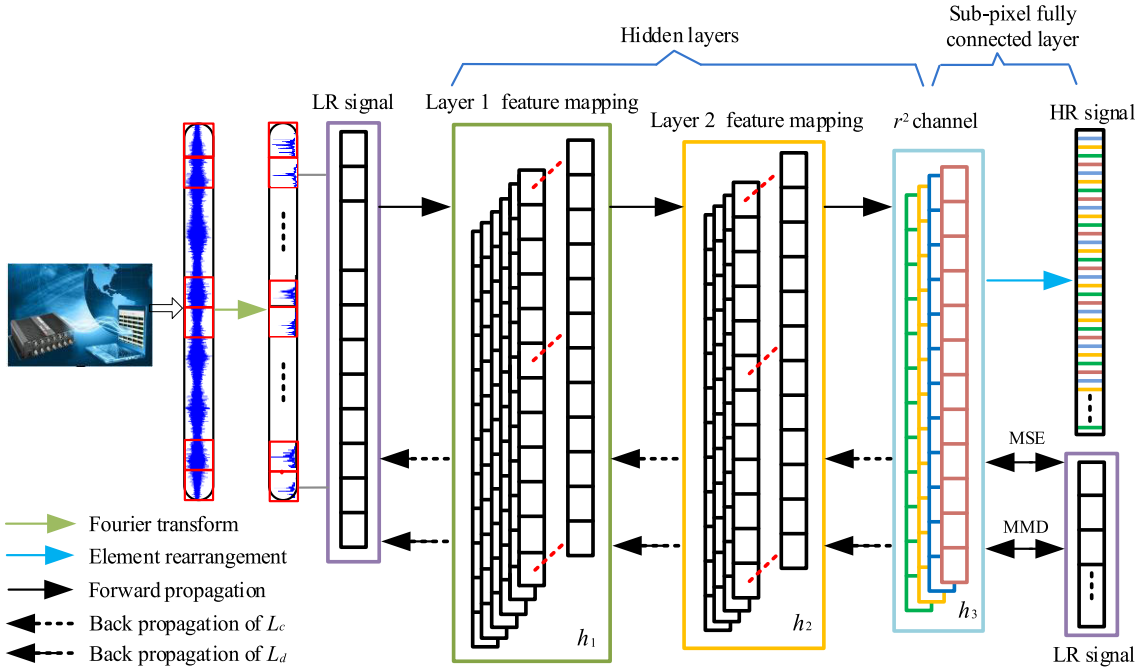


FIGURE 3. Illustration of the proposed framework.

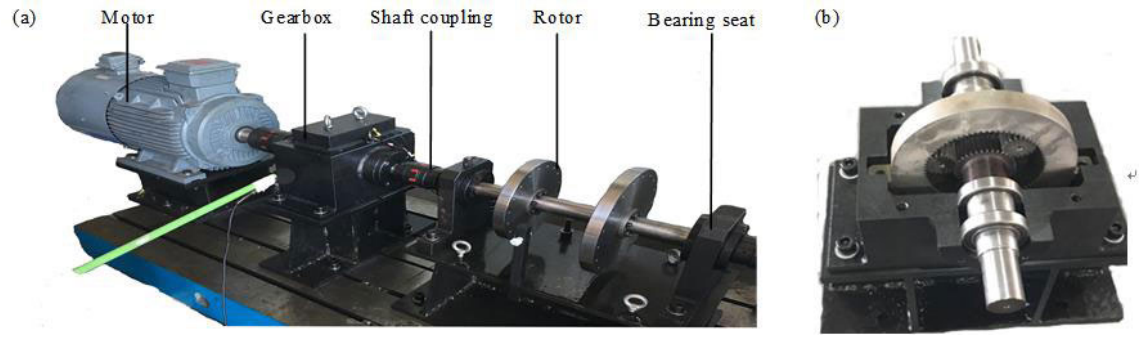


FIGURE 4. (a) Bench of the fault planetary gearbox and (b) inner structure of the planetary gearbox.

where $\boldsymbol{I}^n_{HR}(n = 1.... N)$ represents the HR image examples, and $\boldsymbol{I}^n_{\boldsymbol{LR}}(n = 1.... N)$ represents the resulting LR image examples.

III. PROPOSED FRAMEWORK

As shown in Fig. 3, data augment domain adaptive efficient sub-pixel network (DAESPN) is mainly composed of the fully connected neural network layer and sub-pixel fully connected layer. The fully connected neural network contains two fully connected layers and can be described as follows:

$$\boldsymbol{h}_1 = f_1(\boldsymbol{X}_{LR}; \boldsymbol{W}_1) = \phi(\boldsymbol{W}_1 * \boldsymbol{X}_{LR}) \quad (9)$$

$$\boldsymbol{h}_2 = f_2(\boldsymbol{X}_{LR}; \boldsymbol{W}_2) = \phi(\boldsymbol{W}_2 * f_1(\boldsymbol{X}_{LR})) \quad (10)$$

where $\boldsymbol{h}_i$ represents the hidden layer feature of layer i, $\boldsymbol{W}_i$ represents the weight of the i-layer fully connected layer. For each fully connected layer, 64 and 32 are set as the number of channels. The LR data X_{LR} is output to the sub-pixel fully connected layer through the fully connected neural layer for resolution enhancement.

The sub-pixel fully connected layer consists of a fully connected layer and a subsequent arrangement of pixels. The fully connected layer $f_3(\cdot)$ outputs r^2 $(r = 2)$ channel data of the same dimension as the input data.

$$\boldsymbol{h}_3 = f_3(\boldsymbol{X}_{LR}; \boldsymbol{W}_3) = \phi(\boldsymbol{W}_3 * f_2(\boldsymbol{X}_{LR})) \quad (11)$$

In the last layer $f_4(\cdot)$ of the network, the SR data $\boldsymbol{X}_{SR}$ are obtained through the PS function:

$$\boldsymbol{X}_{SR} = f_4(\boldsymbol{X}_{\boldsymbol{LR}}) = PS(\boldsymbol{W}_3 * f_2(\boldsymbol{X}_{\boldsymbol{LR}})) \quad (12)$$

The form of pixel-wise MSE is as follows:

$$L_c = \xi(\boldsymbol{W}_{1:L}, b_{1:L}) = \frac{1}{r^2 HW} \sum_{x=1}^{rH} \sum_{y=1}^{rW} (\boldsymbol{h}_3^{x,y} - \boldsymbol{X}_{\mathrm{LR}})^2 \quad (13)$$

where $\boldsymbol{h}_3$ represents the feature of the third hidden layer, that is, the simulated signal generated by fully connected neural network. $\boldsymbol{X}_{LR}$ represents the original LR signal.

The purpose of domain loss L_d is to let the deep features of the source and target domains into the same feature space.

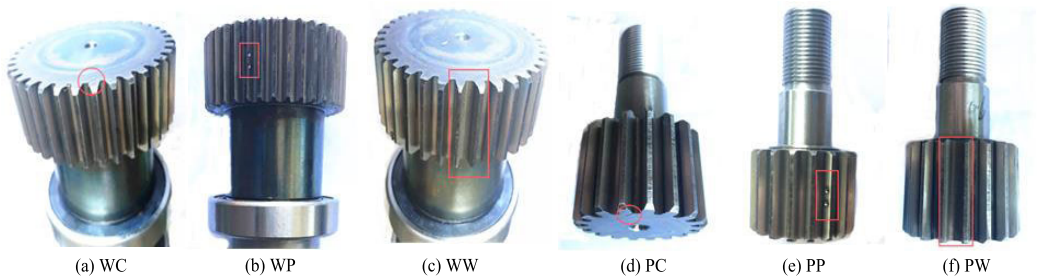


FIGURE 5. Fault feature diagrams of sun and planet wheels.

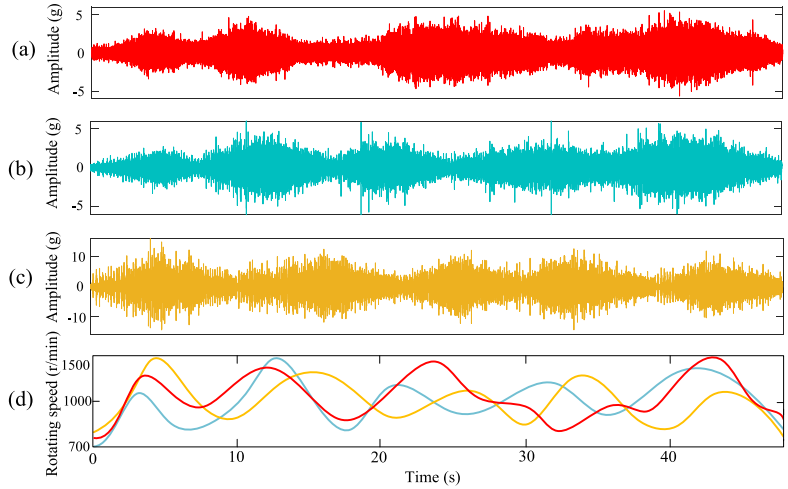


FIGURE 6. Speed fluctuation information of the three gear fault samples: (a) NC, (b) WW, and (c) WWPW. (d) Speed curves.

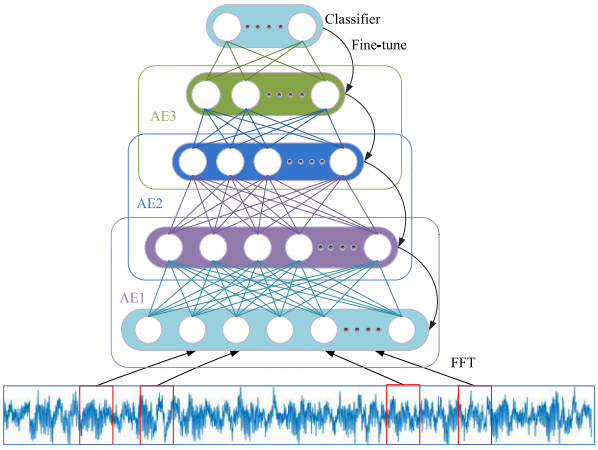


FIGURE 7. Architecture of SAE.

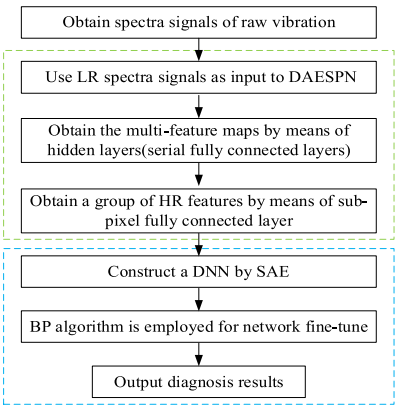


FIGURE 8. Flow chart of the proposed fault diagnosis method.

In this article, $\boldsymbol{X}_{LR}$ is the source domain data, and $\boldsymbol{h}_3$ is the target domain data. Multiple-kernel MMD (MK-MMD) [26] is employed to reduce the distribution distance between two domains to ensure quality of the generated simulation sample, and the form is as follows:

$$L_d = \mathrm{MK}-\mathrm{MMD}^2(\boldsymbol{h}_3, \boldsymbol{X}_{LR}) = \sum_{i=1}^{K} D_i^2(\boldsymbol{h}_3, \boldsymbol{X}_{LR}) \quad (14)$$

where K represents the number of different kernels.

Combining the two optimization objects, the final optimization object can be written as:

$$L = L_c + L_d \quad (15)$$

IV. FAULT DIAGNOSIS USING THE PROPOSED METHOD

A. CASE 1: FAULT DIAGNOSIS OF A PLANETARY GEARBOX

1) EXPERIMENT DEVICE AND DATA INTRODUCTION

A large speed fluctuation experiment of the planetary gearbox fault test bench is carried out to verify the ability of the proposed method. The test bench shown in Fig. 4 includes

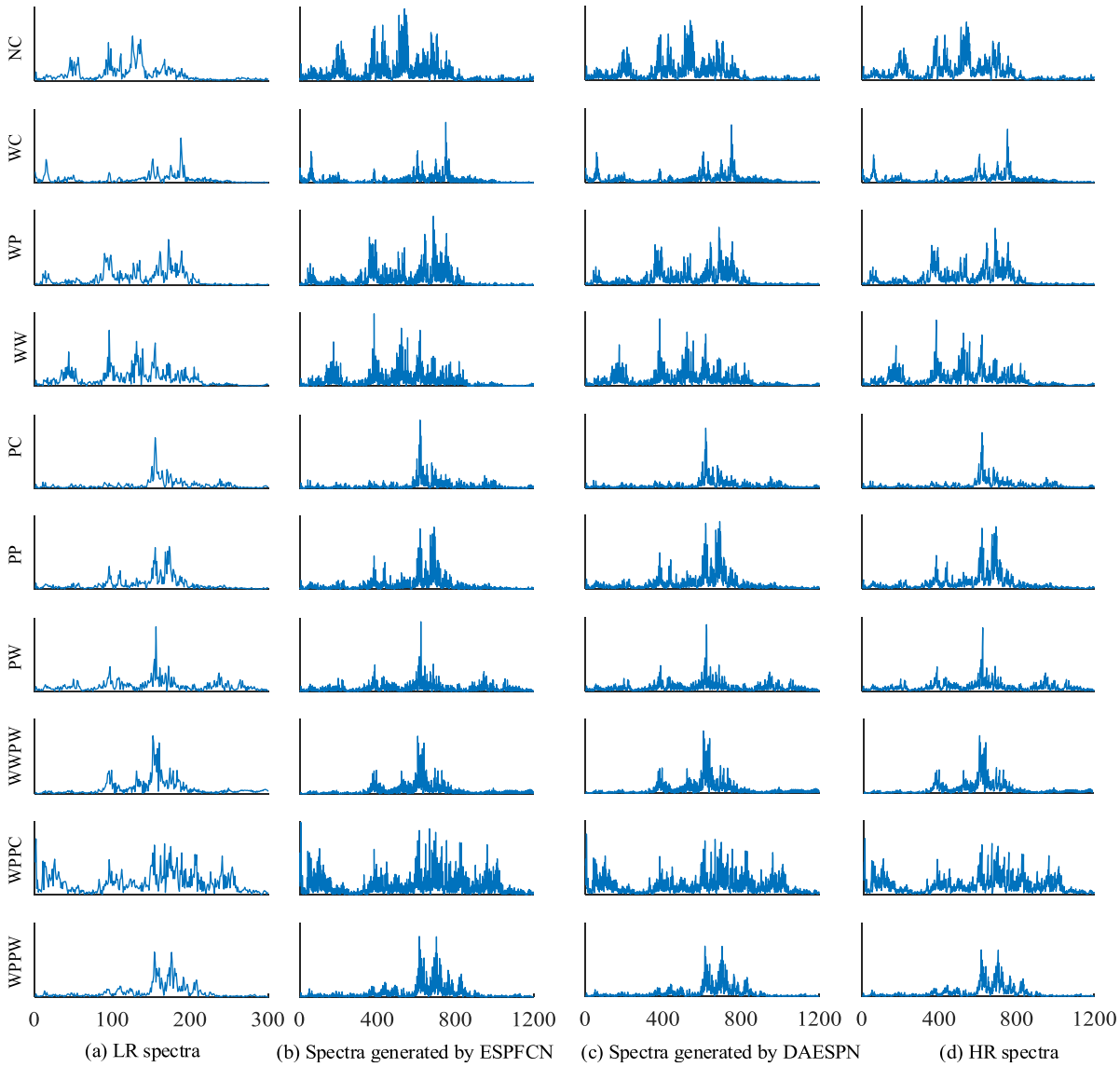


FIGURE 9. Spectra comparison of four data types.

a motor, a planetary gearbox, shaft couplings, and bearing seats. Fig. 5 shows gears in different health conditions: normal condition (NC), sun wheel crack (WC), sun wheel pit (WP), sun wheel worn (WW), pinion crack (PC), pinion pit (PP), pinion worn (PW), wheel worn and pinion worn (WWPW), wheel pit and pinion crack (WPPC), and wheel pit and pinion worn (WPPW). The sampling frequency used in the experiment is 12.8 kHz., and the rotating speed ranges from 700 r/min to 1500 r/min. According to the formula [27]:

$$\Delta f = \frac{f_s}{N} \tag{16}$$

where f_s represents the sampling frequency, N represents the sampling points, and Δf represents the frequency resolution. Assuming that the sampling frequency does not change, the value of the frequency resolution decreases with increasing number of sampling points, that is, the frequency resolution is improved. Therefore, the essence of DAESPN is to generate four times of sampling points under the premise of unchanged sampling frequency, thereby enhancing the frequency resolution. In addition to the HR dataset generated by DAESPN, three sets of comparative experiments are set up, namely, LR dataset, efficient sub-pixel fully connected network (ESPFCN) generated dataset, and HR dataset. Except that there is no MMD loss function, the network structure of ESPFCN is the same as that of DAESPN. For the HR dataset, each health condition contains 200 samples, and each sample contains 2400 data points. For the LR dataset, each health condition contains 200 samples, and each sample contains 600 data points. The input of DAESPN and ESPFCN are LR data, and the output is the simulated HR data. Fast Fourier transform (FFT) analysis is performed to each sample to convert the time domain signal into the frequency domain signal. Given that the data points of the LR dataset

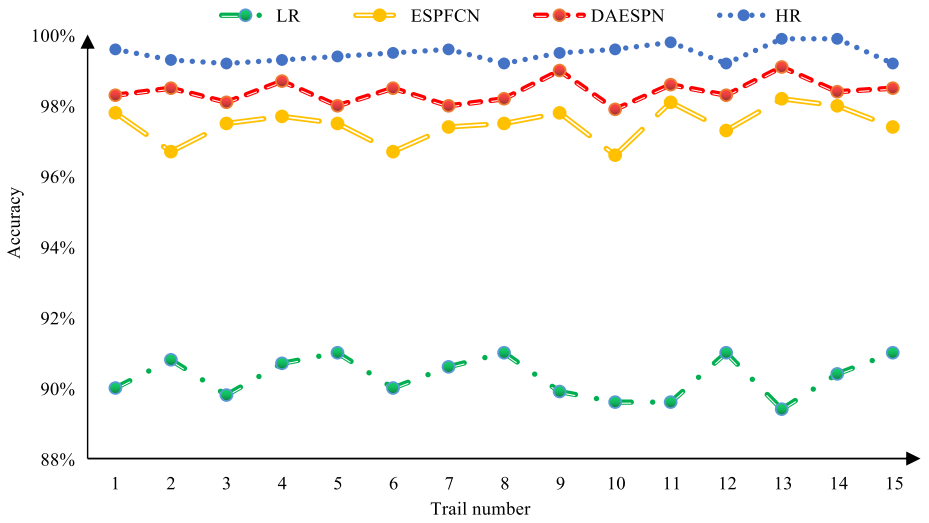


FIGURE 10. Testing accuracies of four data types.

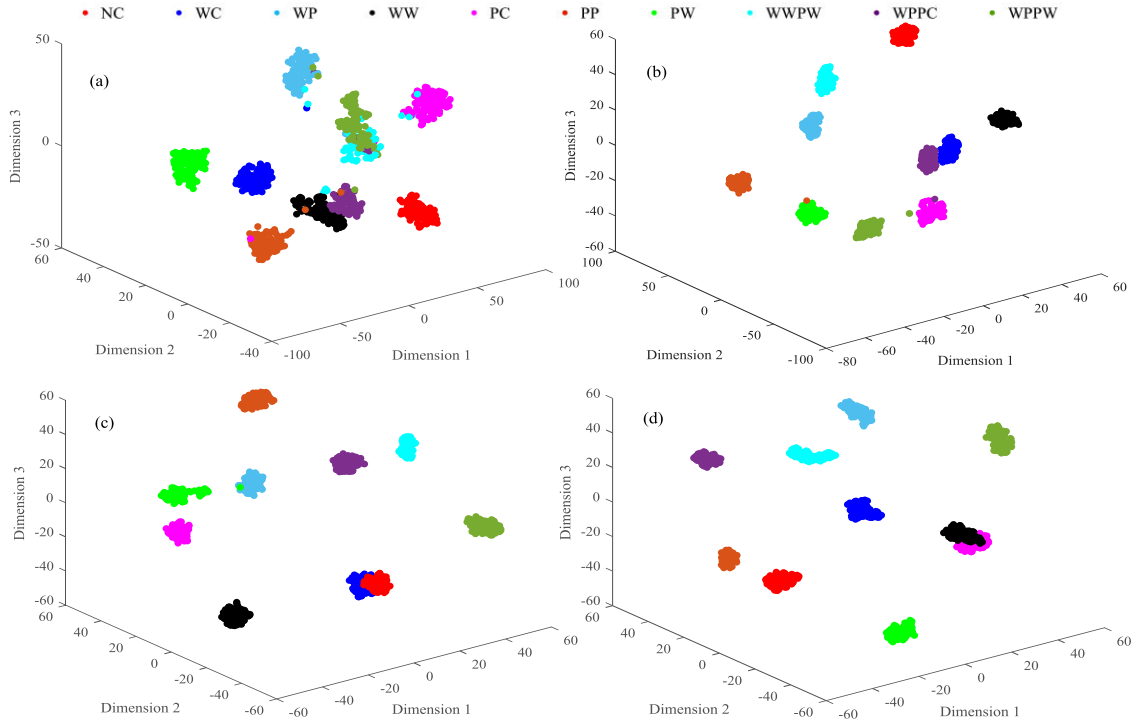


FIGURE 11. Feature visualization map of the t-SNE result of four data types: (a) LR, (b) ESPFCN, (c) DAESPN, and (d) HR.

are 1/4 of the data points of the HR dataset, the former can be regarded as an insufficient dataset. The process of generating HR dataset by DAESPN can be regarded as data augmentation.

As shown in Figs. 6(a)-(c), three gear examples (NC, WW, and WWPW) are randomly selected under the speed fluctuation condition. The vibration signals of different gear fault types show different speed fluctuations in the time domain. The irregular rotation rate fluctuation curves of three gear fault types between 700 and 1000 r/min are shown in Fig. 6 (d).

2) DIAGNOSIS RESULTS

The parameters of DAESPN are set as follows. The training epoch is set to 200, and the activation function is Leaky ReLU. Adam is employed to train the model with learning rate $0.01/(1+10 \times q)0.75$, and q is the training progress that changes from 0 to 1. As shown in Fig. 7, SAE is employed to test the validity of the data generated by DAESPN. The number of hidden layer neurons is arranged as follows: when the input dimension is 300, the number of neurons in the hidden layer is 200-150-100; when the input dimension is 1200, the number of neurons in the hidden layer is 600-200-100.

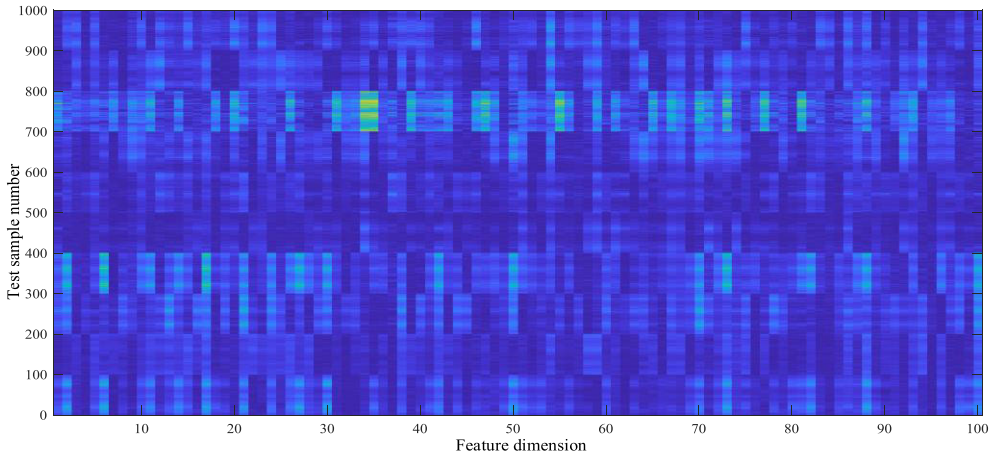


FIGURE 12. Feature visualization of SAE output layer.

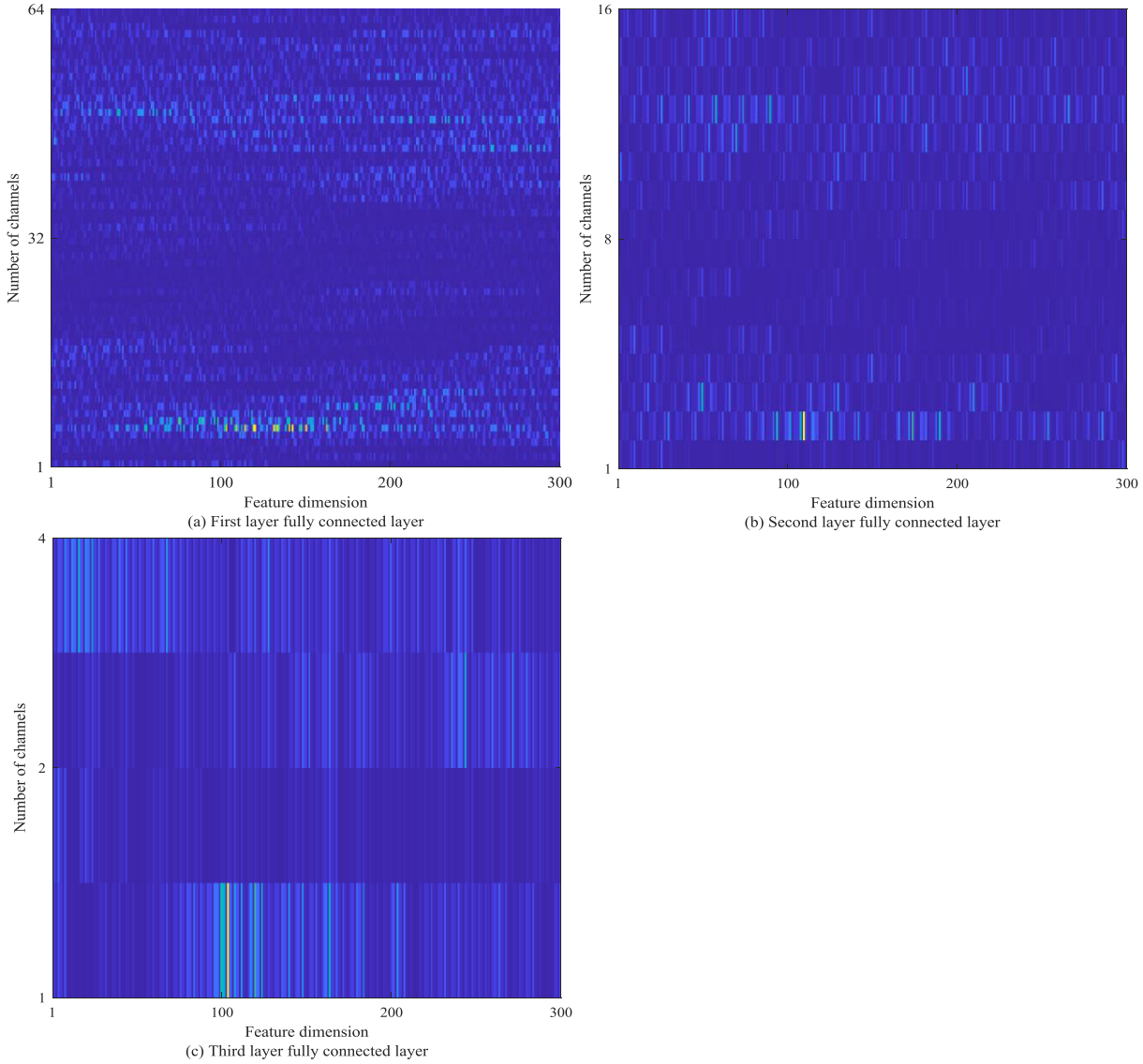


FIGURE 13. Feature visualization of different layer for gear signal in DAESPN.

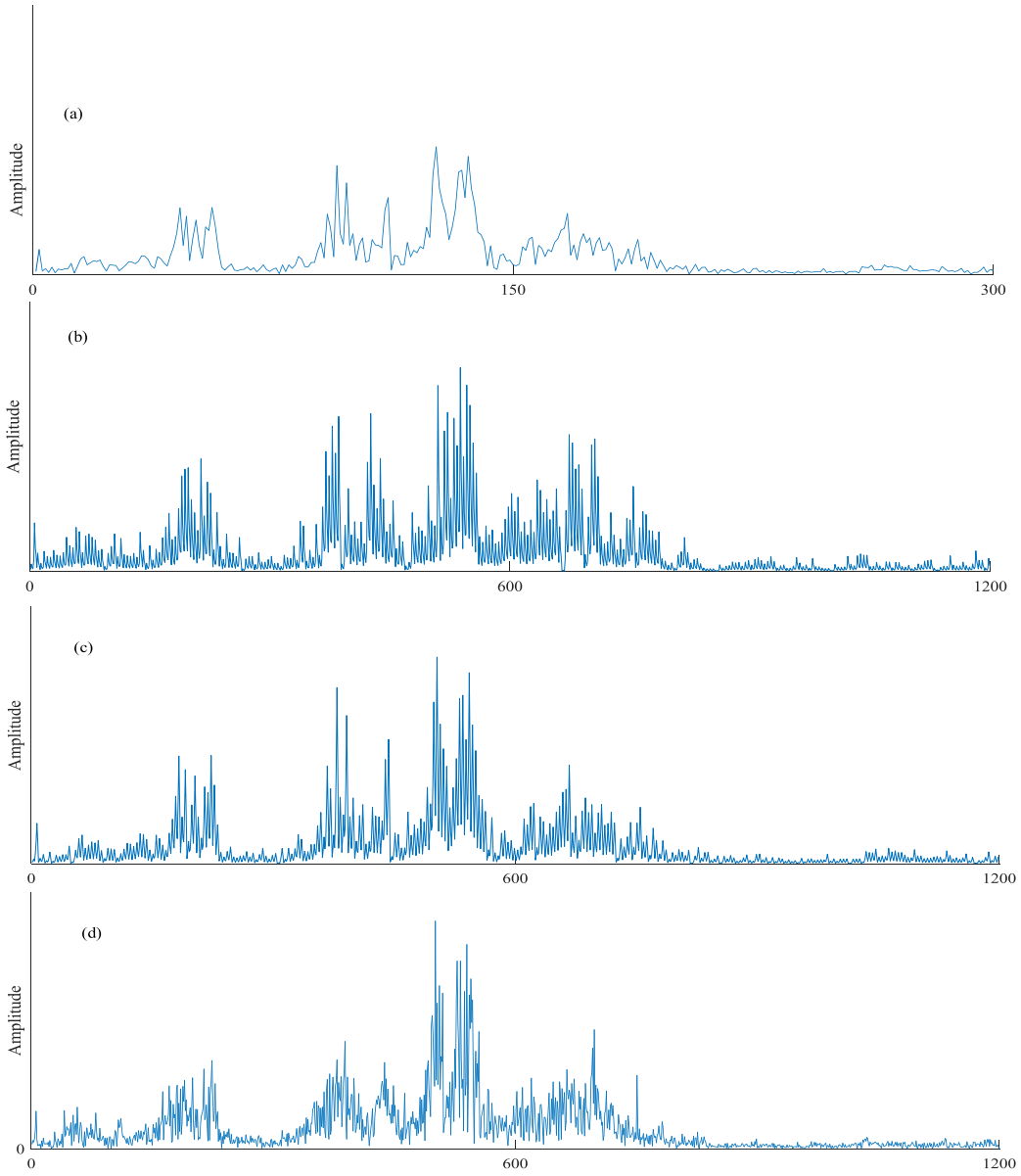


FIGURE 14. Spectrum visualization of four data types: (a) LR, (b) ESPFCN, (c) DAESPN, and (d) HR.

The activation function of the network is Sigmoid, the number of iterative training is 20, the learning rate is 1E-4, and the batch size is 20. The last layer uses the softmax classifier, and the back propagation (BP) algorithm [28] is used for network fine-tuning. Meanwhile, the batch normalization (BN) algorithm [29] is used before each activation layer of the SAE. The dataset generated by DAESPN is the training sample, and the original HR dataset is the testing sample. The original HR sample and the LR sample are used for comparison. About 50% of the samples are randomly selected from different health conditions as the training set, and the remaining 50% of the samples are the testing set. It should be noted that due to the large variation range of the LR samples of different fault types, the range of the spectral amplitude generated by ESPFCN and DAESPN has changed, so the datasets need to be standardized and normalized. In Fig. 8, this model includes three parts: signal collection, data augmentation, and fault diagnosis. Specially, the raw vibration signal are collected using experimental device, and then DAESPN is employed to generate simulated HR samples, finally SAE is adopted for fault diagnosis.

Fig. 9 shows the spectra of four different data types. Compared with the spectra of the three other sample types, each LR sample contains only 300 data points, and the characteristics of different fault types are not evident. Fig. 9 (b) (c) shows the frequency spectra generated by ESPFCN and DAESPN, respectively. In both spectra, more features are generated on the basis of the LR samples and the discrimination of samples with different fault types becomes larger and has a characteristic trend consistent with that of the LR samples.

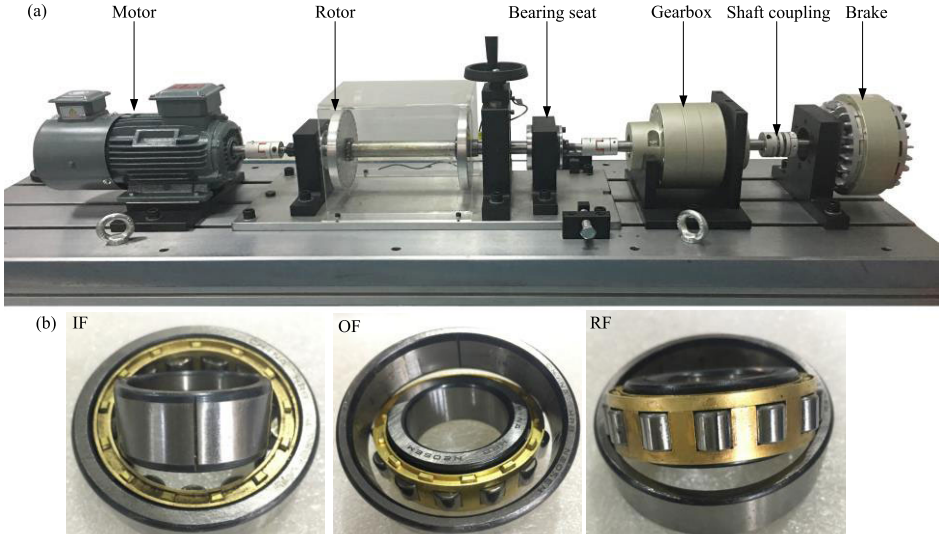


FIGURE 15. (a) Bearing fault test rig and (b) three fault bearings.

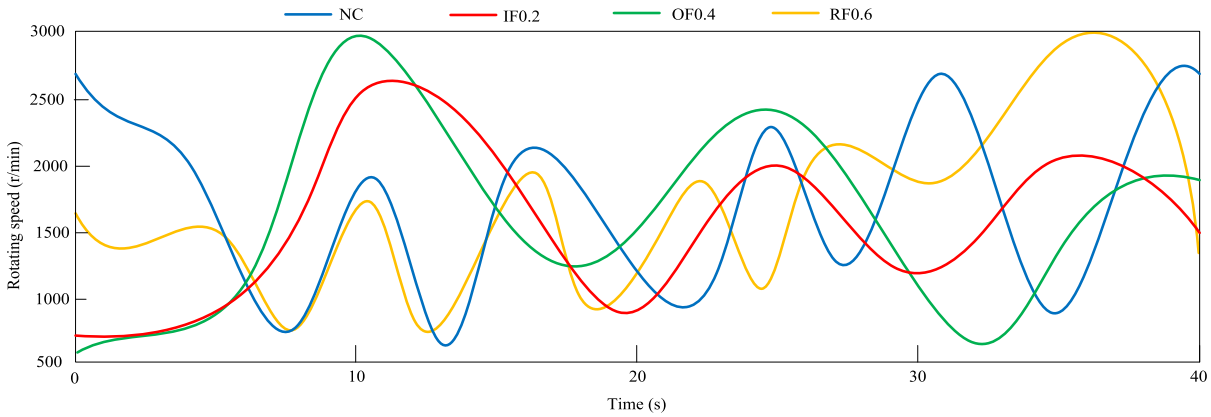


FIGURE 16. Large speed fluctuation information of the four bearing fault types.

The spectrum generated by ESPFCN generates redundant feature points, and the spectrum generated by DAESPN is more similar to the HR spectrum in Fig. 9 (d).

The classification results of the three data types obtained through SAE are shown in Fig. 10. Fifteen experiments are carried out to reduce the influence of randomness. The results from the HR dataset achieve the highest average testing accuracy (99.48%) and the lowest standard deviation (0.24%). The testing accuracies of the LR dataset is 90.32% with 0.57% standard deviation, which is lower than the other experimental results. The performance of ESPFCN is better than the LR dataset, and the average accuracy is 97.48% with 0.48% standard deviation. The testing accuracies of DAESPN are close to that of the HR dataset, and the average accuracy is 98.41% with 0.34% standard deviation. T-distributed stochastic neighbor embedding (t-SNE) [30] is employed to show the visual classification results of dimension reduction. Fig. 11(a) shows that the classification effect of the LR dataset is the worst, and samples of various fault types have different degrees of misclassification. As shown in Fig. 11 (b - d), the dimension reduction result of DAESPN is only second to the HR dataset and is superior to ESPFCN. Only the sample PW is misclassified into the sample WP, and the samples of other fault types are all correctly classified. Furthermore, the effect of feature clustering is good for the same failure type sample, and the separation between features is obvious for samples of different fault types. Fig. 12 shows the output feature of SAE in the third hidden layer, the samples under the same health condition have similar feature trend, the samples under the different health conditions are obviously different. In summary, after the verification of the SAE discriminant network, the performance of the samples generated by DAESPN is excellent.

3) FEATURE LEARNING PROCESS

Fig. 13 (a-c) shows the feature mapping of a sample through three fully connected layers to understand the data augmentation of DAESPN. As the number of fully connected layers increases, the feature dimension of the sample does not change, the number of channels of the sample is determined by 64 reduced to 4, and the characteristics of different channels are more obvious. In Fig. 14 (b) (c), the spectrum

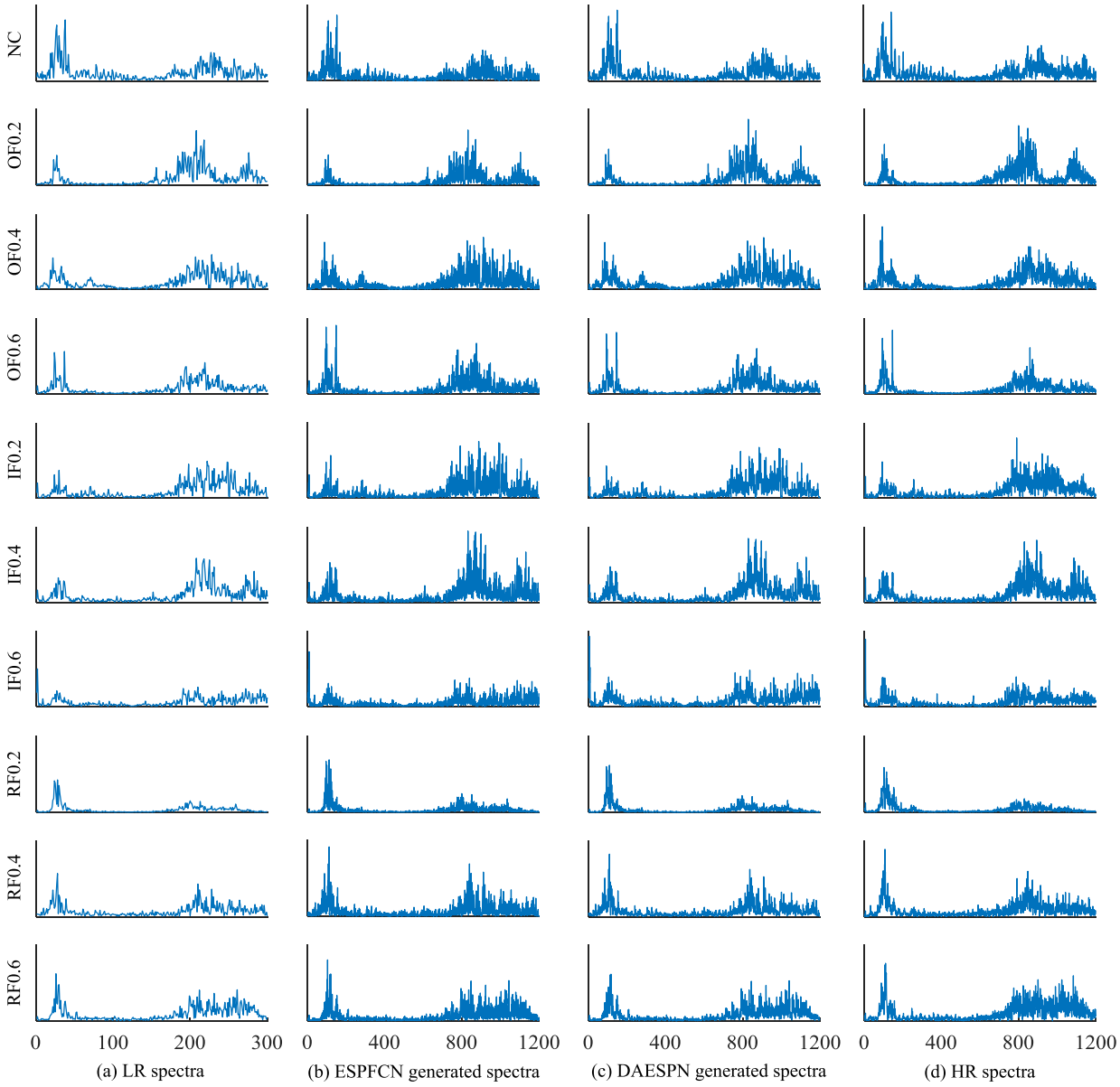


FIGURE 17. Spectra comparison of four data types.

generated by ESPFCN and DAESPN generates more feature points according to the characteristic trend of the spectrum in Fig. 14(a). The similarity between the spectrum in Fig. 14(c) and Fig. 14(d) is higher, and the effect of the spectrum generated by the DAESPCN method is better than that generated by ESPFCN. Therefore, DAESPN can generate reliable HR samples and achieve data augmentation of insufficient samples.

B. CASE STUDY 2: FAULT DIAGNOSIS OF A MOTOR BEARING

1) EXPERIMENT DEVICE AND DATA INTRODUCTION

As is shown in Fig. 15 (a), the platform is mainly composed of motor, shaft couplings, bearing seat, and brake. The bearing dataset contains four different fault types: normal condition (NC), outer race (OF), inner race (IF) and roller faults (RF). There are three different degrees of damages for each fault type: 0.2mm, 0.4mm and 0.6mm. Therefore, there are 10 health conditions of the bearing dataset. The three fault bearings are depicted in Fig. 15 (b). The accelerometer is mounted on the bearing box with a sampling frequency of 25.6 kHz. Fig. 16 shows the irregular speed fluctuation information of the four different bearing fault types between 500 and 3000 r/min and the four different curves corresponding to the four different gear speeds. The network structure and parameter set are similar to those in Case 1.

2) DIAGNOSIS RESULTS

The network structure and parameter set are similar to those in Case 1. Fig. 17 (a) shows few feature points of the sample.

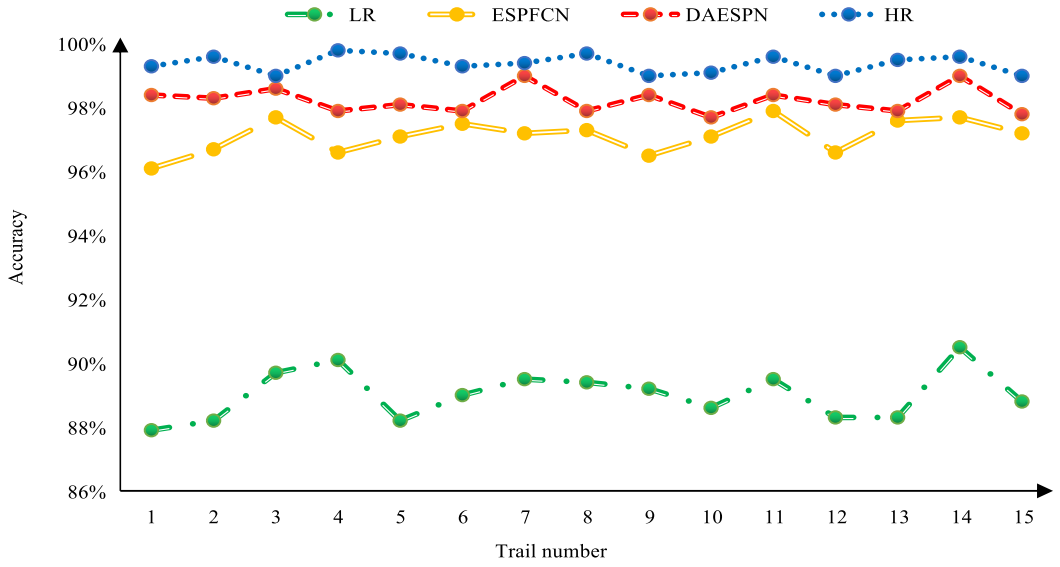


FIGURE 18. Testing accuracies of four datasets.

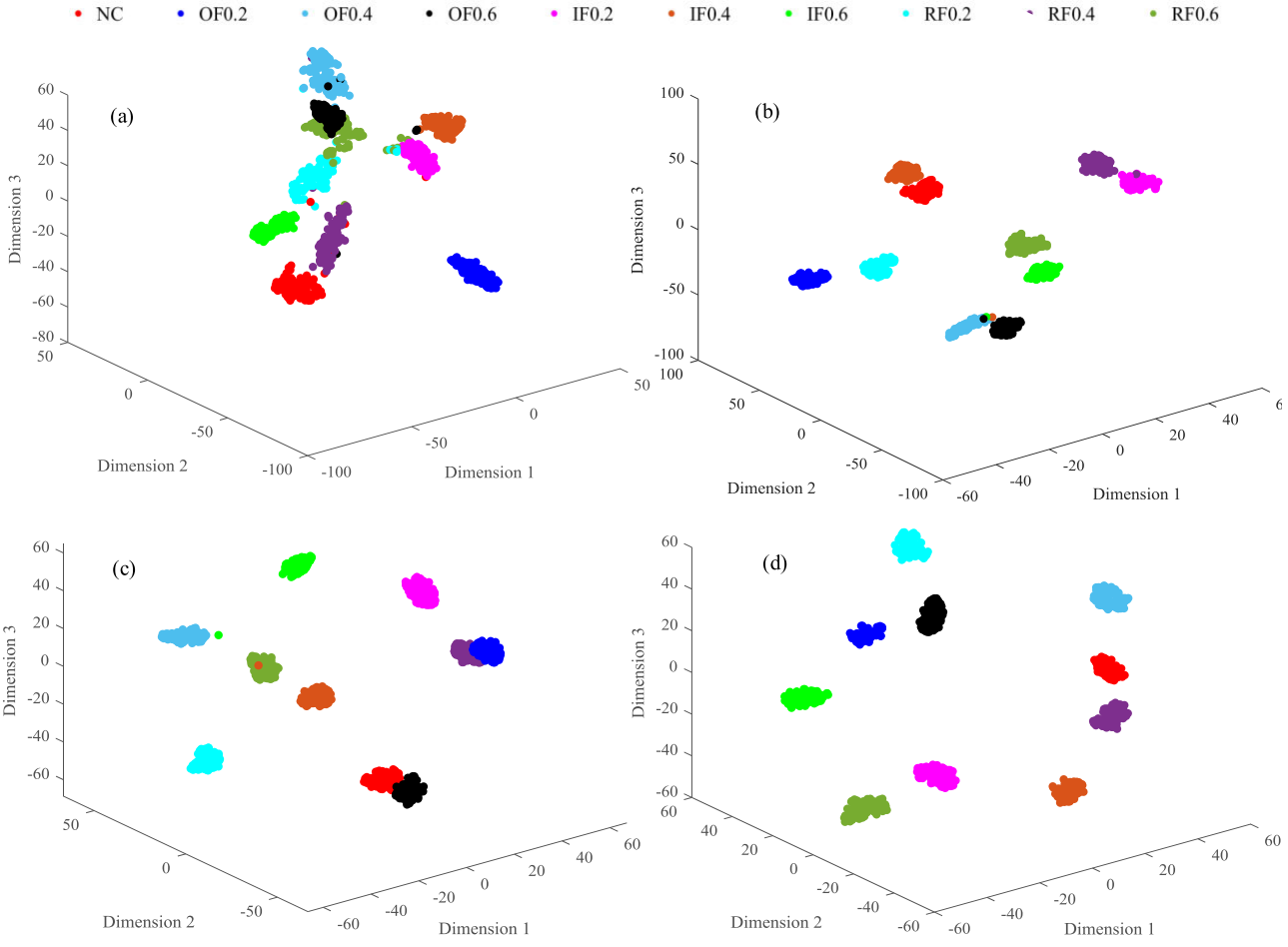


FIGURE 19. Feature visualization map of the t-SNE result of four data types: (a) LR, (b) ESPFCN, (c) DAESPN, and (d) HR.

The features of different fault types are not obvious. The HR spectrum generated by ESPFCN and DAESPN have more data points. As can be seen from the spectrum in Fig. 17(b) and Fig. 17(c), the generation effect of Fig. 17(c) is better,

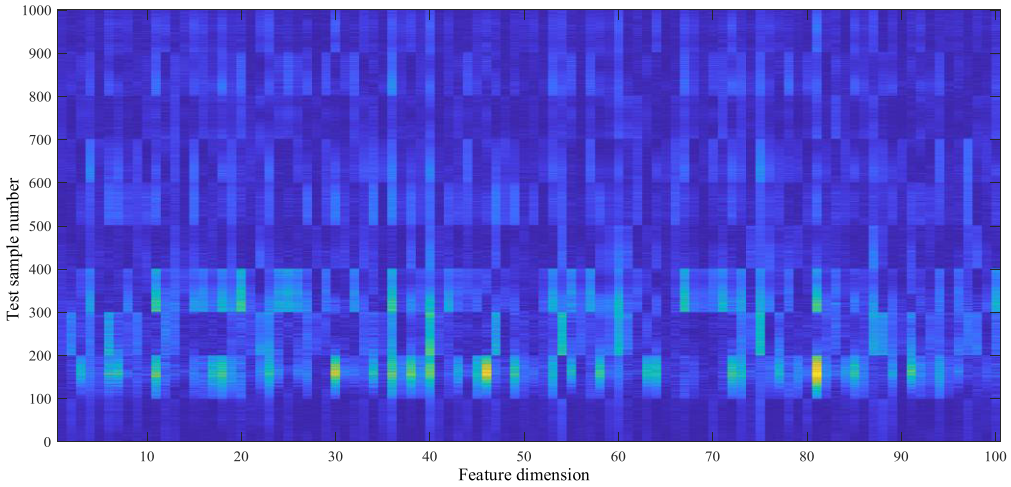


FIGURE 20. Feature visualization of SAE output layer.

the difference among different fault types is obvious and is basically consistent with the characteristic trend of the HR spectrum in Fig. 17 (d).

The test accuracy of the three data types after fifteen experiments is shown in Fig. 18. The test accuracy of the LR dataset has the lowest accuracy (89.01%) with the highest standard deviation (0.77%). The performance of ESPFCN is better than that of the LR dataset; the average accuracy is 97.12%, and the standard deviation is 0.51%. The best results are obtained from the HR dataset, and the average accuracy is 99.37% with 0.28% standard deviation. The performance of DAESPN is second only to the test results of the HR dataset, and the average accuracy is 98.23% with 0.40% standard deviation. Finally, t-SNE is employed to investigate the classification results of the three datasets. As shown in Fig. 19(a), the scattered points of the gear samples under different health conditions are interlaced. Fig. 19 (b)(c) illustrates that the dimension reduction features of DAESPN is better than those of ESPFCN. Moreover, few samples are misclassified in the proposed method, and the dimension reduction results are close to that of the HR dataset in Fig. 19(d). As shown in Fig. 20, the sample features under different health conditions have a high degree of discrimination, and the sample features under the same health status are similarity.

V. CONCLUSION

A framework called DAESPN is proposed for insufficient fault classification under large speed fluctuation of rotating machineries. DAESPN is mainly composed of serial fully connected layers and a fully connected sub-pixel layer. The framework uses MSE and MMD as loss function. Through the sub-pixel fully connected layer, the sample resolution is changed from low to high, and the data points of the sample are augmented 4 times. The gearbox dataset and bearing dataset are employed to verify the proposed DAESPN. The experimental results show that the spectrum generated by DAESPN is close to that of the HR dataset, and both have the same characteristic trend. Moreover, the performance of DAESPN is better than that of ESPFCN without MMD. The comprehensive performance of DAESPN is close to that of the HR dataset.

REFERENCES

[1] X. Jiang, J. Wang, J. Shi, C. Shen, W. Huang, and Z. Zhu, "A coarse-to-fine decomposing strategy of VMD for extraction of weak repetitive transients in fault diagnosis of rotating machines," *Mech. Syst. Signal Process.*, vol. 116, pp. 668–692, Feb. 2019.

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

[3] B. Han, X. Wang, S. Ji, G. Zhang, S. Jia, and J. He, "Data-enhanced stacked autoencoders for insufficient fault classification of machinery and its understanding via visualization," *IEEE Access*, vol. 8, pp. 67790–67798, 2020.

[4] Z. Wang, L. Zheng, W. Du, W. Cai, J. Zhou, J. Wang, X. Han, and G. He, "A novel method for intelligent fault diagnosis of bearing based on capsule neural network," *Complexity*, vol. 2019, pp. 1–17, Jun. 2019.

[5] J. Wang, S. Li, B. Han, Z. An, Y. Xin, W. Qian, and Q. Wu, "Construction of a batch-normalized autoencoder network and its application in mechanical intelligent fault diagnosis," *Meas. Sci. Technol.*, vol. 30, no. 1, pp. 1–17, Jan. 2019.

[6] W. Huang, G. Gao, N. Li, X. Jiang, and Z. Zhu, "Time-frequency squeezing and generalized demodulation combined for variable speed bearing fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 8, pp. 2819–2829, Aug. 2019.

[7] E. Al-Regib, J. Ni, and S.-H. Lee, "Programming spindle speed variation for machine tool chatter suppression," *Int. J. Mach. Tools Manuf.*, vol. 43, no. 12, pp. 1229–1240, Sep. 2003.

[8] J. Urbanek, T. Barszcz, N. Sawalhi, and R. Randall, "Comparison of amplitude-based and phase-based methods for speed tracking in application to wind turbines," *Metrol. Meas. Syst.*, vol. 18, no. 2, pp. 295–304, Jan. 2011.

[9] L. F. Villa, A. Reñones, J. R. Perán, and L. J. de Miguel, "Angular resampling for vibration analysis in wind turbines under non-linear speed fluctuation," *Mech. Syst. Signal Process.*, vol. 25, no. 6, pp. 2157–2168, Aug. 2011.

[10] J. Urbanek, T. Barszcz, R. Zimroz, and J. Antoni, "Application of averaged instantaneous power spectrum for diagnostics of machinery operating under non-stationary operational conditions," *Measurement*, vol. 45, no. 7, pp. 1782–1791, Aug. 2012.

[11] L. Xue, N. Li, Y. Lei, and N. Li, "Incipient fault detection for rolling element bearings under varying speed conditions," *Materials*, vol. 10, no. 6, p. 675, Jun. 2017.

[12] Z. An, S. Li, W. Qian, and J. Wang, "An intelligent fault diagnosis method in the case of rotating speed fluctuations," in *Proc. Prognostics Syst. Health Manage. Conf.*, Jul. 2017, pp. 1–6.
[13] B. Han, G. Zhang, J. Wang, X. Wang, S. Jia, and J. He, "Research and application of regularized sparse filtering model for intelligent fault diagnosis under large speed fluctuation," *IEEE Access*, vol. 8, pp. 39809–39818, 2020.
[14] R. Randall and J. Antoni, "Rolling element bearing diagnostics—A tutorial," *Mech. Syst. Signal Process.*, vol. 25, pp. 485–520, Feb. 2011.
[15] P. Borghesani, R. Ricci, S. Chatterton, and P. Pennacchi, "A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions," *Mech. Syst. Signal Process.*, vol. 38, no. 1, pp. 23–35, Jul. 2013.
[16] P. D. Mcfadden and M. M. Toozhy, "Application of synchronous averaging to vibration monitoring of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 14, no. 6, pp. 891–906, Nov. 2000.
[17] X. Li, W. Zhang, Q. Ding, and J.-Q. Sun, "Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation," *J. Intell. Manuf.*, vol. 31, no. 2, pp. 433–452, Feb. 2020.
[18] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
[19] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1637–1645.
[20] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. Comput. Vis. Pattern Recognit.*, 2016, pp. 4681–4690. [Online]. Available: http://arxiv.org/abs/1609.04802
[21] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1874–1883.
[22] Y. Bengio, *Learning Deep Architectures for AI*. Boston, MA, USA: Now, 2009, pp. 1–127.
[23] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *J. Mach. Learn. Res.*, vol. 13, pp: 723-773, 2012.
[24] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Scholkopf, and A. J. Smola, "Integrating structured biological data by kernel maximum mean discrepancy," *Bioinformatics*, vol. 22, no. 14, pp. e49–e57, Jul. 2006.
[25] Z. An, S. Li, J. Wang, Y. Xin, and K. Xu, "Generalization of deep neural network for bearing fault diagnosis under different working conditions using multiple kernel method," *Neurocomputing*, vol. 352, pp. 42–53, Aug. 2019.
[26] A. Gretton, K. Borgwardt, M. J. Rasch, B. Scholkopf, and A. J. Smola, "A kernel method for the two-sample problem," in *Proc. Neural Inf. Process. Syst.*, 2006, pp. 513–520.
[27] C. Yuan, D. He, and G. Yu, *The Analysis for the Frequency Resolution of Vibration Data Acquisition Device*. Sichuan, China: Dongfang Turbine, 2013, pp. 19–22.
[28] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 2, 1997, pp. 396–404.
[29] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[30] V. Laurens and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

XIAOYU WANG received the B.S. degree from the Liren College, Yanshan University, Qinhuangdao, China, in 2017. He is currently pursuing the master's degree with the Shandong University of Science and Technology, Qingdao, China.

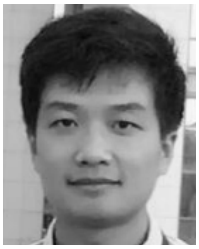
His current research interests include intelligent fault diagnosis of machines, noise and vibration analysis, and control.

ZHENYUN CHU received the M.S. degree from the Shandong University of Science and Technology, Qingdao, China, in 2005.

He is currently working as a Teacher with the Shandong University of Science and Technology. His current research interests include signal processing and fault diagnosis, vehicle dynamics, noise and vibration analysis, and control.

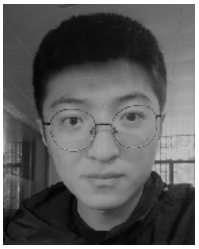
BAOKUN HAN received the Ph.D. degree from the School of Mechanical Engineering, Beijing Institute of Technology, Beijing, China, in 2003.

He is currently a Professor with the Shandong University of Science and Technology. His current research interests include signal processing and fault diagnosis, vehicle dynamics, noise and vibration analysis, and control.



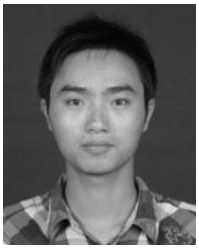
JINRUI WANG received the Ph.D. degree from the Nanjing University of Aeronautics and Astronautics, China, in 2019.

He is currently working as a Postdoctoral Researcher with the Shandong University of Science and Technology, Qingdao, China. His current research interest focuses on intelligent fault diagnosis of machines.



GUOWEI ZHANG received the B.S. degree from Zaozhuang University, Zaozhuang, China, in 2018. He is currently pursuing the master's degree with the Shandong University of Science and Technology, Qingdao, China.

His current research interests include intelligent fault diagnosis of machines, noise and vibration analysis, and control.



XINGXING JIANG received the Ph.D. degree from the Nanjing University of Aeronautics and Astronautics, China, in 2016.

He is currently working as a Postdoctoral Researcher with Soochow University, Suzhou, China. His current research interest focuses on intelligent fault diagnosis of machines.