Experimental Techniques (2023) 47:407–418
https://doi.org/10.1007/s40799-022-00552-x

RESEARCH PAPER

Investigation and Experimental Study on Gearbox Vibration Fault Diagnosis Method Based on Fusion Feature Convolutional Learning Network

X. Zhu[1,2] · X. Ye[1,2] · R. Wang[1,2] · W. Zhao[1,2] · X. Luo[3] · J. Zhao[4] · Z. Han[1,2] · X. Gao[1,2]

Received: 28 November 2020 / Accepted: 26 January 2022 / Published online: 1 March 2022

Abstract

When the gearbox fails, the impulsive fault feature is usually submerged in other frequency component. The recognition accuracy of the network model will be seriously affected when directly using the vibration signal as the sample input of Convolutional Neural Network (CNN). To solve this problem, a state recognition method based on vibration signal fusion feature convolutional learning network (FF-CNN) is proposed to conduct the gear fault diagnosis. This method uses the Resonance-based Sparse Signal Decomposition (RSSD) algorithm to separate the periodic and impulsive components that characterize different fault characteristics of the vibration signal. The impulsive feature is amplified by the Teager energy operator. And then the signal periodic and impulsive components are input into CNN to perform the targeted deep learning. Finally, the efficiency of this method is validated using the gear power transmission failure simulation experimental setup. The research in this paper improves the effectiveness of the feature learning, reduces the complexity of deep learning model, and ensures the accuracy of state recognition. Through experimental research, it is found that the recognition accuracy of the recognition model based on the method in this paper reaches more than 95.6%.

Highlights

- The experiments were carried out in the dynamic engineering laboratory.
- An FF-CNN method is proposed to conduct the fault diagnosis.
- T-SNE tool in the manifold learning method was used to visualize the results.
- FF-CNN can improve the accuracy of fault diagnosis by separating fault features.

Keywords Vibration fault diagnosis · Resonance-based sparse decomposition · Feature fusion · Teager energy operator · Convolutional Neural Network

✉ X. Gao
okspringgao@hotmail.com

1 Department of Power Engineering, North China Electric Power University, Baoding 071003, China

2 Hebei Key Laboratory of Low Carbon and High Efficiency Power Generation Technology, North China Electric Power University, Baoding 071003, Hebei, China

3 Hunan University College of Electrical and Information Engineering, Changsha 410082, Hunan, China

4 North China Municipal Engineering Design and Research Institute, Tianjin 300070, China

Introduction

The gear transmission system has an important position in the engineering field, such as the application in wind power generators, coal mines, petroleum and other fields. Due to its long-term work in a complex and harsh working environment. The gear is prone to various failures, which affects the safety and reliability of the system. Therefore, condition monitoring and fault diagnosis of gears is very necessary.

With the continuous development of data mining and artificial intelligence technology, machine learning-based fault diagnosis models are widely used in condition monitoring of mechanical equipment, such as: Artificial Neural Network

(ANN), support vector machines (SVM), fuzzy recognition [1–3]. However, due to the lack of learning depth, it is difficult to improve the diagnostic accuracy. In recent years, the emergence of Deep Learning has solved the problem of feature adaptive learning [4]. Among them, the application of the Convolutional Neural Network (CNN) in vibration fault diagnosis has become a hot topic. The essence of CNN is to learn a plurality of feature filters capable of extracting input data features, perform layer-by-layer convolution and pooling operations on the input data, and use a back-propagation algorithm to extract features hidden in the data step by step.

In recent years, some scholars have gradually introduced CNN into the field of fault diagnosis. Literature [5] uses CNN to directly perform feature learning on the spectrum data of the vibration signal, and uses the signal's time domain data, spectrum data and combined time–frequency data to test its performance. It is concluded that the method has the highest recognition accuracy of the spectrum data with the accuracy of 84.6%. In [6], the time–frequency images obtained by continuous wavelet transform of vibration signals are recognized based on CNN, which can reach an accuracy of 90.2%. However, when the time–frequency image of the vibration signal is traditionally used as the analysis object, the image's visual characteristics will be seriously affected due to the displacement and scaling of the image, which will seriously affect the intrinsic feature learning of the image. In addition, traditional CNNs must consider the two-dimensional spatial correlation of the image, while the vibration data is only related to time, which is a typical one-dimensional parameter.

Considering the phenomenon of information loss in the process of image transformation of the signal, many studies have used the original signal as the object of deep feature learning. Literature [7] used CNN to learn the characteristics of one-dimensional vibration signals and achieved good results. Literature [8] established a deep neural network (DNN) model to directly identify the signal, and input the vibration signal as a one-dimensional vector to the neural network for feature learning which can retain the true information of the original vibration signal. All of these features make it more suitable for the diagnosis of vibration failure.

In the actual operation of the gearbox. The fault signals of bearings and gears contain complex periodic components and impact components. However, those fault features are often submerged in the external noise generated by the frequency conversion, frequency conversion harmonic components and other vibration sources, accompanied by the vibration phenomenon of the modulation frequency [9], which will greatly increase the complexity of the model and seriously affect the recognition accuracy of the model if directly conducting the characteristics learning. To solve this problem, a more effective method is proposed to demodulate the vibration signal (such as Hilbert-Huang transform [10], Wavelet Transform (WT) [11], Empirical Mode Decomposition (EMD) [12], Ensemble Empirical Mode Decomposition (EEMD) [13], Local Mean Decomposition (LMD) [14], etc.). The demodulated signal components are subjected to special component learning or fusion learning to improve recognition accuracy and reduce the complexity of deep learning models. However, due to the shortcomings of signal demodulation algorithms, such as the selection of wavelet basis functions in WT will be seriously affected by human experience, EMD, LMD and other methods have problems such as endpoint effects and modal aliasing [15], which will lead to distortion of feature extraction and further affects the effect of fault diagnosis.

In order to solve the above problems, this paper proposes a vibration state recognition method based on fusion feature convolutional learning network. Based on the CNN structure This method adds the resonance-based sparse decomposition and the Teager energy operator layer to establish a convolutional learning network of fused features. The model structure is shown in the Fig. 1. Through the resonance-based sparse decomposition and the Teager energy operator layer enhance periodic, impulsive and other effective vibration fault information, the vibration state is finally obtained. Based on this, the influence of model structure and

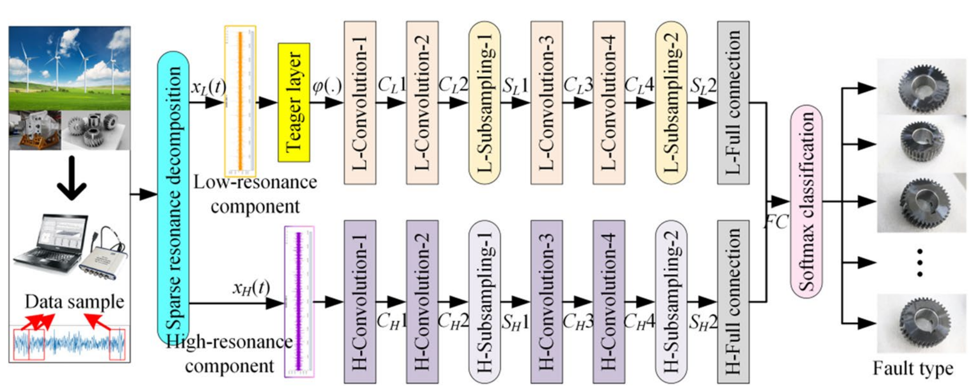

Fig. 1 FF-CNN fault diagnosis system

parameters on accuracy is discussed in this paper, and the model is optimized experimentally.

Vibration Fault Diagnosis Method Based on Fusion Feature Convolutional Learning Network

Gearbox faults include gear fault (such as broken teeth and missing teeth) and bearing fault (such as inner and outer rings fault). They will generate periodic pulses, and the vibration signals generated often include periodic transient pulse components, frequency conversion and its harmonics and noise components. The proposed vibration diagnosis method based on fusion feature convolutional learning in this paper can be divided into the following two parts:

Part 1, uses the resonance-based sparse decomposition algorithm to decompose the original signal in the input layer to obtain the high-resonance and low-resonance components of the signal. The Teager energy operator layer is used to enhance the transient impulsive characteristics of the low-resonance component and then, to achieve the extraction and amplification of the periodic impact on the gear signal;

Part 2, performs convolution, subsampling and other operations on the periodic harmonic components and transient impulsive components (i.e., high-resonance and low-resonance components) obtained by decomposition to complete the deep feature learning of information fusion, and finally identify the vibration state to obtain the fault classification result. Especially in the convolutional neural network, the receptive field of the neural network is increased by the superposition of two convolutional layers, which is more suitable for the feature learning of one-dimensional form data. The structure is shown in Fig. 2.

Signal Resonance-Based Sparse Decomposition

Selesnick [16] proposed the method of signal resonance-based sparse decomposition firstly in 2012. Resonance-based sparse decomposition is to quantify the signal's resonance properties by using the quality factor Q ($Q = f_c / B_W$, where f_c is the frequency of the signal and B_W is the bandwidth of the signal). Signals with more oscillations and high Q-factor are referred as high-frequency-aggregated high-resonance attribute signals. Signals with few oscillations and low Q-factor are referred as low-frequency-aggregated low-resonance-attribute signals, and the original input signal is decomposed into high-resonance and low-resonance based on the magnitude of Q-factor.

Table 1 shows the high-resonance and low-resonance components corresponding to different quality factors Q and the wavelet function library map are drafted in Fig. 3.

It can be seen from Fig. 3 that the resonance-based sparse decomposition of the signal can be characterized according

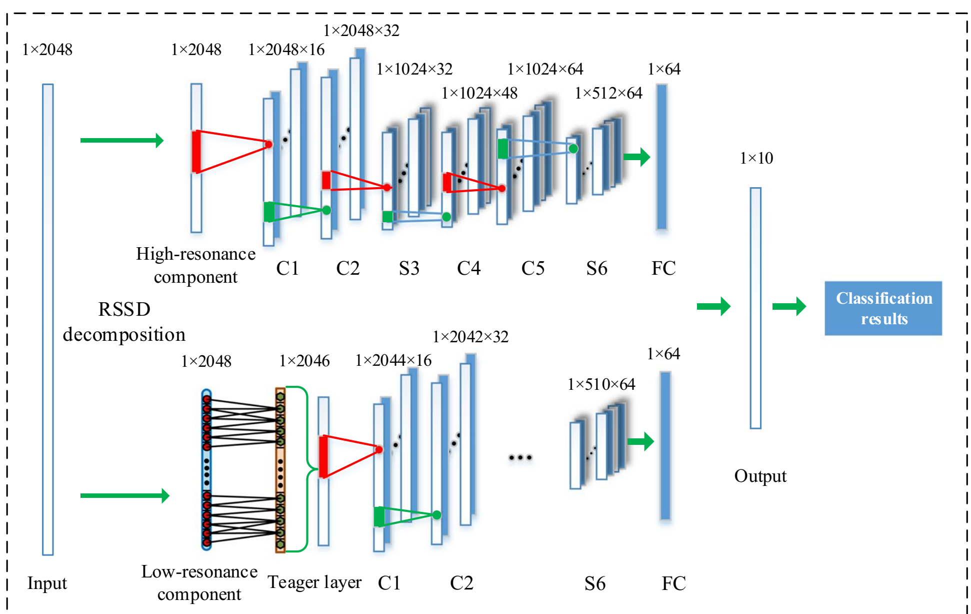

Fig. 2 Structure of vibration signal fusion feature convolutional learning network model

Table 1 High-resonance and low-resonance components corresponding to different quality factors Q

Q-factor Redundancy factor γ	1	3
3	Low-resonance components $x_L(t)$	High-resonance components $x_H(t)$

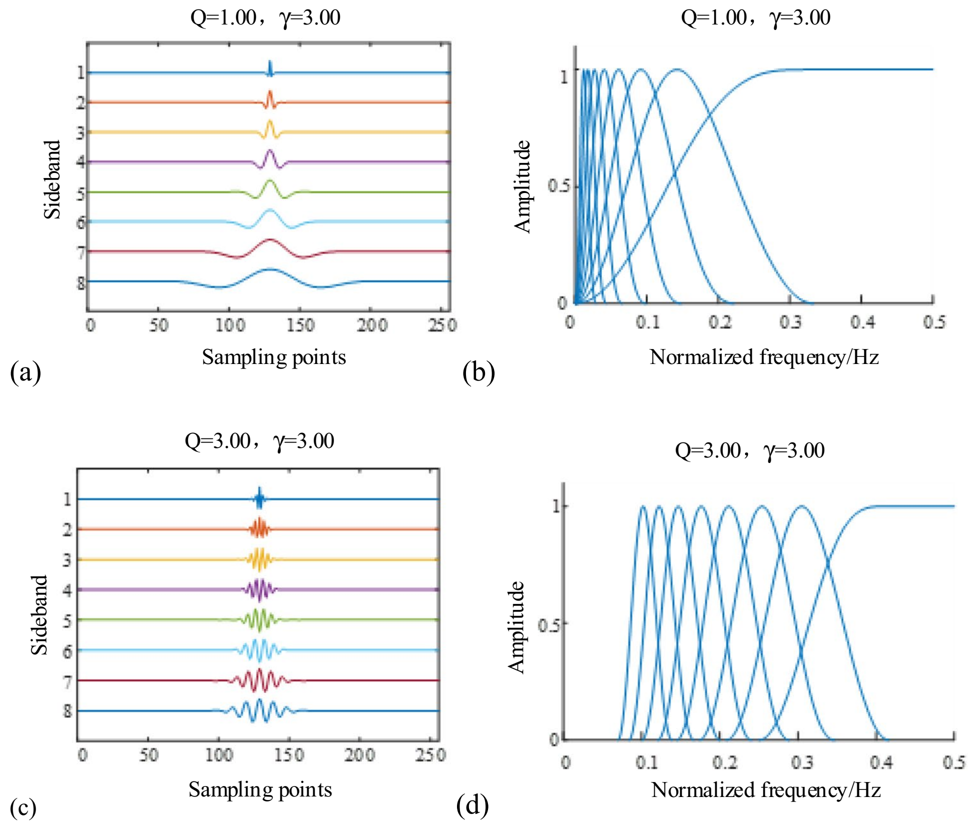

Fig. 3 Adjustable wavelet function library: (**a**) Waveforms of each sub-band (Q = 1, γ = 3, j = 8); (**b**) Spectra waveforms (Q = 1, γ = 3, j = 8); (**c**) Waveforms of each sub-band (Q = 3, γ = 3, j = 8) and (**d**) Spectra waveforms of each sub-band (Q = 3, γ = 3, j = 8)

to the Q-factor: as the Q-factor value increases, the resonance-based intensity of the signal increases, and the signal waveform appears more times on the time domain waveform chart (Fig. 3(a) and (c)); the spectra waveform of the signal shows that the overlap between adjacent two stages decreases as the Q-factor increases, the frequency resolution increases, and the error decreases (Fig. 3(c) and (d)), and when the Q-factor value is as low as 1, it can be seen on the frequency domain waveform graph that the spectra waveform of each level of wavelet starts from the origin and the waveform overlap is high.

After the signal undergoes resonance-based sparse decomposition, the components with different resonance properties contained in the signal can be separated, and the best sparse expressions of the high-resonance and low-resonance components are established [17], completed the signal separation of the high-resonance and low-resonance components.

Energy Operator Layer

Periodic shock is a key feature for judging the local damage. If the periodic shock feature in the original vibration signal can be highlighted, the accuracy of the neural network's classification of gearbox faults will be effectively improved. The Teager energy operator is a non-linear difference operator. Using the instantaneous value of the signal and its non-linear combination to estimate the energy required by the signal source to generate a dynamic signal, the transient feature can be strengthened significantly which is suitable for the impulsive components in the detection signal [18].

The calculation process for the Teager energy operator is as follows:

The general expression of the AM / FM signal *x(t)* with time-varying amplitude *a(t)* and time-varying phase*ω(t)* is

$$x(t) = a(t)cos\varnothing(t) \tag{1}$$

For low resonance components, the signals $x_L(t)$ $(t \le l$, l is the data length of the original signal), Its energy operator is defined as:

$$\varphi\left(x_L(t)\right) = (x_L(t))^2 - x_L(t-1)x_L(t+1)2 \le t \le l-1 \tag{2}$$

For discrete signals, only three samples of data are needed for the Teager energy operator to calculate the value at t. Thus, the calculation amount will not affect the calculation of the convolutional neural network. Meanwhile, it has better resolution for the instantaneous changes of the signal which can highlight the signal transients.

Advantages of the Algorithm

According to analysis, this method has the following advantages:

(1) Preserve the information of the original vibration signal
The traditional CNN model takes a two-dimensional image matrix as input. However, when processing vibration signals, if the vibration signal picture or the extracted feature spectrum is used as the CNN input, there must be differences in image vision due to coordinate axis selection and image stretching which will result in an artificially difference features of the image extracting by CNN, making it difficult to learn the overall difference features of different faults. The object processed by the neural network proposed in this paper is a one-dimensional vector form, which avoids artificial interference caused by image visual factors and retains the information of the original vibration signal.
(2) Impulsive characteristics extracted from gear vibration signals

The traditional signal demodulation method cannot resolve the signal according to the fault impact characteristics of the gear or bearing, and the diagnosis effect of the gear fault is poor. This paper firstly uses the resonance-based sparse decomposition algorithm to process the original one-dimensional vibration signal in the fusion feature convolutional learning network model, decomposes the vibration signal into high-resonance and low-resonance components according to different quality factors Q, then uses the Teager energy operator to further strengthening the shock characteristics contained in the low-resonance component, and finally inputs the high-resonance component and the low-resonance component to the neural network for deep learning to obtain a more accurate gear fault classification result.

Network Training and Optimization

When a gearbox fails, the damaged components will periodically frictionally collide with normal components, which results in the generation of impulse shock. Therefore, the low-resonance and high-resonance components with Deep learning can get effective fault classification results. In order to verify the effectiveness of the diagnostic method based on vibration signal fusion feature convolutional learning in gear fault diagnosis, the experiments of different faults of gears and bearings were carried out in the dynamic engineering laboratory of North China Electric Power University (NCEPU).

The experimental setup fully simulates the transmission system of wind turbine shafting. The motor drives the input shaft of the planetary gearbox, which is fixed to the sun gear. The ring gear is standstill, and the planet carrier is coupled to the output shaft, which drives two-stage parallel shaft gears and loads. The rotation speed of the planetary gearbox input shaft is measured by the tachometer. The physical parameters of the test bench are listed in Table 2. Gear 3 is selected as the fault gear, it is the driving gear of the second stage parallel shaft gears. The different states of parallel gears and rolling bearings are used in the experimental setup. Parallel gears include normal gears, gear wear, cracks, broken teeth and missing teeth. Rolling bearings include normal bearings, outer ring faults, inner ring faults, rolling element faults and compound faults (inner ring, outer ring and rolling element fault). The experimental setup, fault gears kit and fault bearings kit are shown in the Figs. 4, 5 and 6 respectively. In this paper, the vibration data of various fault types are collected at the motor speed of 1200 rpm and the sampling frequency of 8000 Hz.

As shown in Fig. 7(a), (b), (c) and (d), when the gear has gear tooth broke, gear crack, gear wear and tear and gear tooth deficiency failures, the resonance-based sparsity decomposition is used to decompose the original vibration signal. Comparing the high resonance component with the low resonance component, it strongly verifies the effectiveness of the resonance sparse decomposition method to separate the harmonic component and the impact component. The impact characteristic of the low resonance component is obvious, and the impact of the original signal is extracted into the low resonance component signal. Among the high resonance components, the harmonic components are the main ones, and the signal changes relatively smoothly. In the low resonance component spectrum, the fault characteristics and the prominent frequency peaks of the four gear fault types can be observed, indicating that this method can effectively extract gear fault components.

However, for the low-resonance component and the high-resonance component of the gear fault signal, the characteristics of the above-mentioned component cannot be used solely as a fault diagnosis feature, and the diagnosis of the gear fault cannot be achieved. Therefore, it is necessary to comprehensively consider the high-resonance and low-resonance

Table 2 Physical parameters of the test bench

Motor Power (N·m)	Number of sun gear teeth	Number of ring gear teeth	Number of Planet gear teeth	Number of gear one	Number of gear two	Number of gear three	Number of gear four
7	20	40	100	29	95	36	90

Fig. 4 NCEPU gear power transmission failure simulation experimental setup

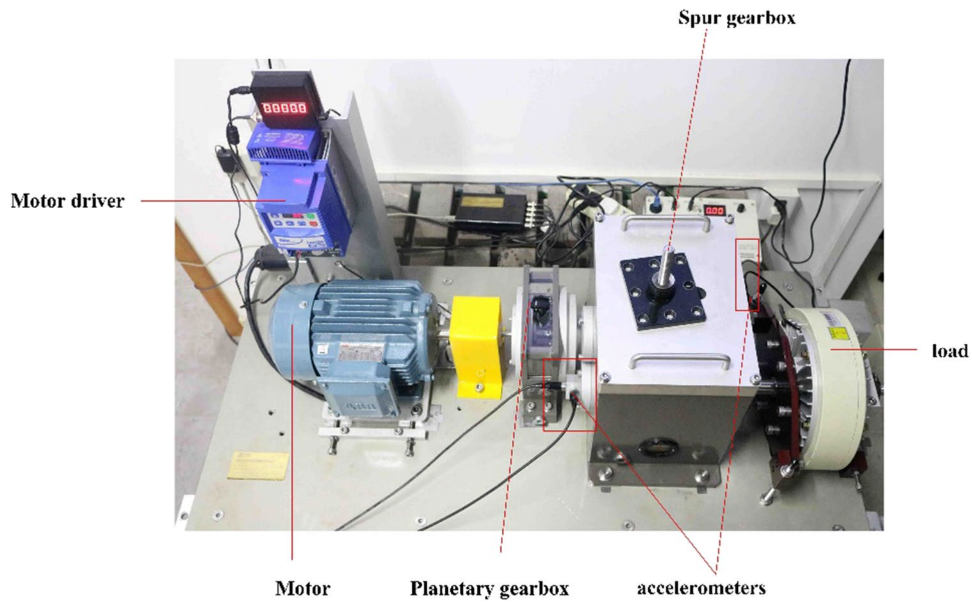

Fig. 5 Fault gear (**a**) gear tooth broke (**b**) gear crack (**c**) gear wear and tear (**d**) gear tooth deficiency

(a) (b) (c) (d)

Fig. 6 Fault bearing (**a**) inner ring fault (**b**) outer ring fault (**c**) rolling element fault (**d**) inner ring, outer ring and rolling element fault

(a) (b) (c) (d)

components of the gear vibration signal to complete the gear fault diagnosis.

In this paper, the vibration data in the NCEPU database is selected as the experimental sample, of which 50% is used as the training sample set and 50% is used as the test sample set. The specific experimental samples are selected as follows (Table 3):

Appropriate structural parameters can improve the recognition accuracy and efficiency of the model in constructing a convolutional neural network model. The hyperparameters are initialized by the network which cannot be learned during the network training process and need to be set firstly. The maximum recognition rate can be obtained by setting the best combination of hyperparameters. In convolutional neural networks, these hyperparameters include: kernel size, learning rate, neural network layers, activation function, loss function and optimizer, etc. In order to find the optimal model structure parameters, this paper firstly designs a basic FF-CNN network structure by screening, and then further determines the structure parameters of the network by trial and error. The network structure is initially selected as it is shown in Fig. 2. The size of the convolutional kernel is 1*5, and the loss function is the squared difference loss function. Based on this, the corresponding optimal parameters are determined through experiments.

Learning Rate and Its Decay Function

For the classification task in this paper, the model is optimized using Stochastic Gradient Descent (SGD). In

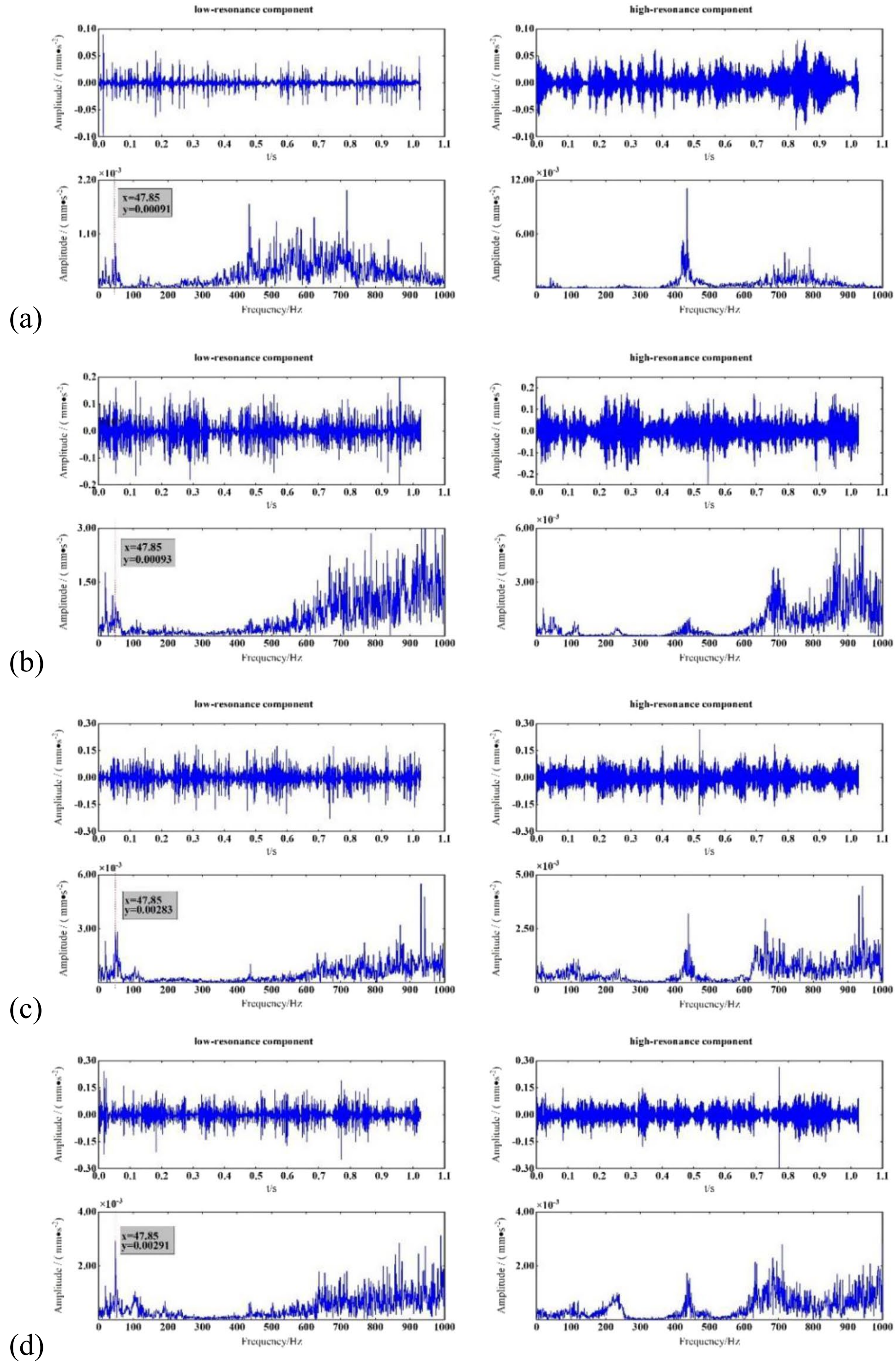

Fig. 7 Waveform and spectrum of high-resonance and low-resonance components under each gear fault type: (**a**) Low-resonance and high-resonance component of gear tooth broke fault signal under 8000 Hz and 1200 rpm; (**b**) Low-resonance and high-resonance component of gear crack fault signal under 8000 Hz and 1200 rpm; (**c**) Low-resonance and high-resonance component of gear wear and tear fault signal under 8000 Hz and 1200 rpm and (**d**) Low-resonance and high-resonance component of gear tooth deficiency fault signal under 8000 Hz and 1200 rpm

order to find the optimal learning rate parameter setting, the classification accuracy rate of the test set is used as the standard. By setting different learning rate parameters, multiple classification experiments are performed to determine the parameter settings. In the experiment, the learning rate is set to 0.001, 0.005, 0.01, and 0.1, and the decay modes are Fixed (fixed), Step (decay with steps), and Exp (exponential decay). The experimental results are shown in Table 4.

From the experimental results in Table 4, it can be seen that the classification accuracy is highest when the learning rate is 0.005. When the learning rate is too high, the training model process will not converge and the classification task cannot be completed. When different learning rate decay

Table 3 Experimental samples of bearing vibration signals

Gear sample types	Sample length	Sample size	bearing sample types	Sample length	Sample size
Normal	2048	200	Normal	2048	200
Gear tooth deficiency	2048	200	Bearing inner ring, outer ring and rolling element fault	2048	200
Gear tooth broke	2048	200	Bearing inner ring failure	2048	200
Gear crack	2048	200	Bearing outer ring fault	2048	200
Gear wear and tear	2048	200	Bearing rolling element fault	2048	200

Table 4 Influence of learning rate parameters on recognition rate

Learning rates Decay models	0.001	0.005	0.01	0.1
Fixed	93.5%	94.6%	92.6%	10.5%
Step	93.7%	95.1%	93.2%	10.2%
Exp	94.8%	95.6%	94.5%	10%

Table 5 Influence of convolutional layer parameters on recognition rate

Size Depth (C1, C2, C3, C4)	1×3	1×4	1×5	1×6
8, 16, 32, 64	92.5%	91.7%	93.1%	92.7%
16, 32, 64, 128	94.8%	95.6%	94.9%	94.6%
32, 48, 64, 94	94.4%	94.7%	94.1%	94.3%
48, 96, 128, 192	93.5%	93.6%	92.6%	93.6%

Table 6 Influence of loss function on recognition rate

Type of loss function	Squared	Absolute difference	Average cross entropy
Recognition rate	92.5%	92.1%	92.2%

functions are used, it can be concluded that the Exp model obtains the highest accuracy. Therefore, the learning rate of the network model in this paper is 0.005, and the decay function is Exp.

Convolutional Layer Parameters

In the process of deep learning, the model learns high-level, abstract features from the classification objects is a prerequisite for completing complex classification tasks. In FF-CNN, the setting of the convolutional layer affects the effect of model feature extraction. In order to effectively extract the feature representations that are most conducive to sample classification from the signal, the optimal network structure can be found by changing the number and depth of the convolutional layers. In this paper, the neural network with the structure of Fig. 2 is determined through preliminary experiments. Based on the structure, the size and the depth of the convolutional layer are determined experimentally.

According to the experimental results in Table 5, the depth of the convolutional layer has a significant impact on the accuracy of the model classification only when the depth of the convolutional layer is insufficient. The training time of the model is greatly increased, and the model convergence is more difficult. Therefore, the convolutional layer depth of 16,32,64,128 is selected to complete the classification task in this paper.

Loss Function

In a multi-class problem where the training sample is N and the total number of categories is C, the classification algorithm usually finds weights and biases that make the network output $f(x)$ of all network inputs x close to Y. The loss function is defined as the gap between the output $f(x)$ of the training input x and Y. The common loss functions in convolutional neural networks are as follows with the experimental results of different loss functions are listed in Table 6.

Loss function of squared:

$$\mathrm{L}(\mathrm{Y},f(x)) = \frac{l}{2}(Y - f(x))^2 \tag{3}$$

Loss function of absolute difference:

$$L(\mathrm{Y},f(x)) = |Y - f(x)| \tag{4}$$

Loss function of average cross entropy:

$$L(\mathrm{Y},f(x)) = -\sum_i Y_i log(f(x_i)) \tag{5}$$

It can be concluded that the classification accuracy of the model is highest when the loss function of squared is used, and the difference between the classification accuracy of the three loss functions is small. However, since the loss function of average cross entropy is no longer affected by saturation compared with the loss function of square. The weight update speed is faster when the error is larger, and the weight update speed is slower when the error is smaller which avoids the decrease of learning efficiency. Therefore, the loss function of average cross entropy is selected as the loss function of the network model in this paper.

Based on the above experimental results, the structural parameters of the FF-CNN model are finally determined and shown in Table 7.

Experimental Study

FF-CNN Test

Using the FF-CNN model parameters in Table 7, the gear vibration data in the NCEPU database is used to train the model. When training the deep learning neural network, the learning rate is set to 10^{-2}, the number of batch samples is set to 32, and the model is used to classify the test set during the model iteration process. The classification results are compared with the marked fault types, and the training model is output. The accuracy of the classification is shown in Fig. 8.

After 25 iterations, the classification accuracy of the model on the test set reached 95.6%. Since the FF-CNN model's learning process for sample features is similar to a black box, in order to get a closer understanding of the processing effect of the deep learning neural network model on different types of failures, the confusion matrix is used to analyse the experimental results. The results are shown in Fig. 9.

According to the test results in Fig. 9, it can be seen that the average recognition accuracy of the five gear states is 95.6% by applying the FF-CNN model. Since the model in this paper introduces the energy operator and resonance sparse decomposition algorithm, in order to study the effect of this front network on the original input signal features, the t-Distributed Stochastic Neighbor Embedding(T-SNE) tool in the manifold learning method is used for model training at different levels to reduce and visualize high-dimensional data. T-SNE manifold learning algorithm is a deep learning non-linear learning algorithm. It can effectively implement high-dimensional data visualization and dimensionality reduction. The results are shown in Fig. 10.

Figure 10(a) is the classification of the original input vibration signal. Because the vibration signal has noise components and contains a variety of feature mixes, the visual distribution of the original vibration signal is chaotic and complex which is difficult to classify it. The classification of the vibration signal after the resonance-based sparse decomposition algorithm and the energy operator layer processing are shown in Fig. 10(b), it can be seen that compared with the original input signal, the classification characteristics are

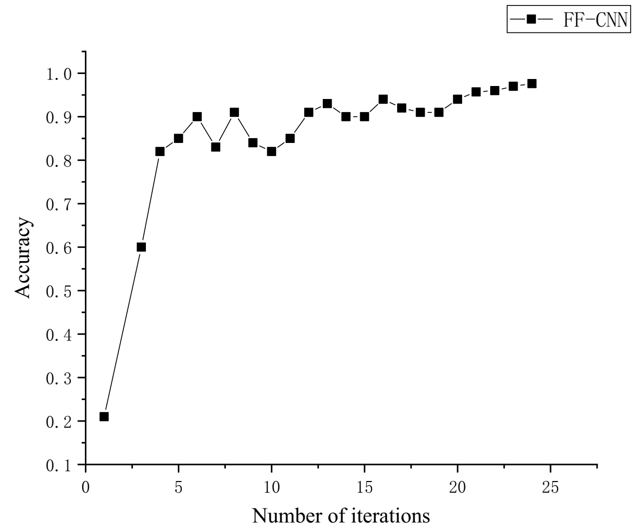

Fig. 8 Test accuracy during model training

Table 7 FF-CNN model parameter settings

Layer	Kernels/ filter (high × width /stride)	Feature maps
Input	–	1 × 2048
Energy Operator	1 × 4/1	1 × 2046
C1	1 × 4/1	1 × 2044 × 16
C2	1 × 4/1	1 × 2042 × 32
S1(max)	1 × 2/2	1 × 1022 × 32
C4	1 × 4/1	1 × 1020 × 48
C5	1 × 4/1	1 × 1018 × 64
S2(max)	1 × 2/2	1 × 510 × 64
FC	–	1 × 64
Output	–	1 × 5

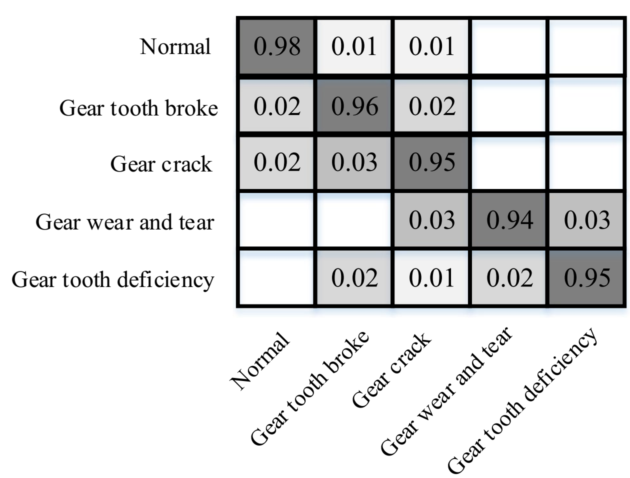

Fig. 9 FF-CNN recognition accuracy of the test set

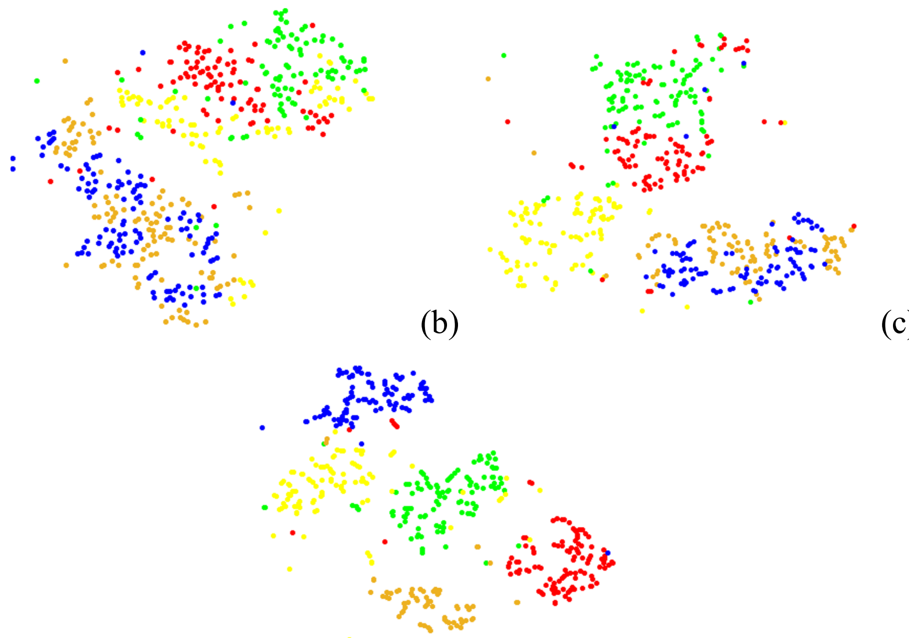

Fig. 10 Classification effect of FF-CNN model: (**a**) Input layer classification diagram; (**b**) Pre-processing network classification effect diagram and (c) FC layer classification effect diagram

obvious. However, the classification accuracy is not enough and in which, still many classifications are mixed together. Figure 10(c) is the final classification of the fully connected layer 1×64 vector after dimensionality reduction. It can be found that the classification characteristics are very obvious at this time. Most of the similar samples are clustered together, and only some unrecognized samples are scattered in different classification regions.

Comparative Study

The conventional intelligent fault diagnosis method usually consists of a feature extraction section and a feature selection section, as it is shown in Table 7. For example, EMD [19] are used to extract the features of the signal, and then the principal component analysis is used to reduce the dimensionality to extract the feature vectors from the high-dimensional matrix. For the selected feature vectors, artificial neural networks such as support vector machines (SVM) are usually used for classification. 1D-CNN [7] is used to filter, select segments, and normalize the original vibration signal to construct training samples and test samples. It uses a one-dimensional convolutional neural network structure to implement feature extraction and failure of the original vibration signal classification.

In this paper, on the basis of traditional CNN, the effective vibration fault information such as signal periodic component and shock component is tracking by adding resonance-based sparse decomposition algorithm and Teager energy operator layer, and then perform deep learning on the characteristic information to obtain higher recognition accuracy. In order to verify the effectiveness of this method, this paper selects some other methods for comparison. (1) The traditional diagnostic method based on SVM which takes the feature extracted by the EMD method as the model input [19]. (2) The 1D-CNN method takes raw vibration data as the model input [7]. (3) Deep learning diagnosis method based on Deep Belief Networks (DBN) [20]. (4) Back Propagation Neural Network (BPNN)takes raw vibration data as the model input. (5) Deep learning diagnosis method CNN takes the feature extracted by the Symmetrized Dot Pattern (SDP) method as the model input [25]. (6) Deep learning diagnosis method CNN takes the time domain waveform as the model input [21]. The results were shown in Table 8.

According to the experimental results in the table, the performance of the diagnosis method is largely determined by the performance of the feature extraction method and the classifier algorithm in the traditional intelligent fault diagnosis method. When the SVM processes the feature samples extracted by the EMD method, limited by the training

Table 8 Classification accuracy of the test set

Classifier	Feature	Gear diagnosis accuracy	Rolling bearing diagnosis accuracy
SVM [19]	EMD	85%	90.3%
1D-CNN [7]	Raw vibration data	90.2%	93.3%
DBN [22]	Raw vibration data	76.5%	90.7%
BPNN [20]	Raw vibration data	77.6%	90.5%
CNN [23]	SDP	85.1%	93.5%
CNN [21]	Time domain waveform	82.9%	91.2%
FF-CNN	Raw vibration data	95.6%	98.7%

method, it need to extract feature vectors from the signals processed by EMD as learning objects of the SVM classification algorithm, the small number of samples during training and verification is also a limitation.

However, using CNN, DBN and BP deep learning model, a comprehensive classification features can be learned from the extracted features and adaptively complete the classification task. Meanwhile, a larger sample size data set can be learned.

When the two-dimensional image matrix is used as the input of the CNN model, the visual differences of the image caused by the selection of the coordinate axis, image stretching, etc. may lead to the artificially distinguished features of the visual difference of the image caused by CNN. Thus, the classification accuracy is also low.

However, the feature extraction and fault classification of the FF-CNN model rely on the deep learning of the neural network on the one hand, and on the resonance-based sparse decomposition signal demodulation algorithm on the other hand. At the same time, it introduces the energy calculation which enhances the characteristics of the original vibration signal. Thus, FF-CNN has higher recognition accuracy than 1D-CNN directly identifying vibration signals, and is more suitable for vibration fault diagnosis.

Conclusion

This paper established and proposed a fault diagnosis method FF-CNN based on deep learning technology which is a fusion feature convolutional learning network designed for gearbox vibration signals. It introduces the resonance-based sparse decomposition algorithm and the Teager energy operator, extracts the shock and periodic components in the vibration signal containing fault features. Refer to the structure of CNN, the strong nonlinear feature learning ability of the convolutional kernel are used to perform targeted learning on the vector containing fault information. Finally, the gearbox fault diagnosis method FF-CNN can automatically learn the fault classification features from the original gearbox diagnosis signals, which can realize intelligent fault diagnosis and does not require manual design feature extraction.

From the experimental results, the advantages of FF-CNN are listed as follows:

(1) The FF-CNN-based fault diagnosis method has a deeper learning depth and a higher accuracy recognition result compared with the traditional signal processing technology EMD, wavelet packet decomposition and shallow neural network combined fault diagnosis method.

(2) The fault diagnosis method with feature learning proposed in this paper has higher recognition accuracy, by adding the front resonance-based sparse decomposition algorithm and energy operator layer to process the longer one-dimensional vibration signal, which strengthens the gearbox vibration fault characteristics.

(3) When the CNN model uses the time–frequency image of the vibration signal as the analysis object, it will affect the visual characteristics of the image due to the displacement and scaling of the image, which severely affects the model recognition accuracy. The research method proposed in this paper avoids this problem. Experimental results prove that the research method proposed in this paper has higher recognition accuracy.

Acknowledgements This paper is supported by the Natural Science Foundation of Hebei Province [No. E2019502080], the National Natural Science Foundation of China (No.52076081), the Natural Science Foundation of Hebei Province (No.E2019502072).

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

Declarations

Conflicts of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Zhang YQ, Zhang PL, Wu DH (2015) Bearing fault diagnosis based on optimal generalized S transform and pulse coupled neural network. J Vibr Shock 34(9):26–31
2. Si JP, Ma JC, Niu JH (2017) An intelligent fault diagnosis expert system based on fuzzy neural network. J Vibr Shock 36(4):164–171
3. Li Y, Xu M, Wei Y et al (2016) A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree. Measurement 77:80–94
4. Zhu X, Hou D, Zhou P et al (2019) Rotor fault diagnosis using a convolutional neural network with symmetrized dot pattern images. Measurement 138:526–535
5. Jing L, Zhao M, Li P et al (2017) A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. Measurement 111:1–10
6. Liao YX, Zeng XQ, Li WH (2017) Wavelet transform based convolutional neural network for gearbox fault classification. In: Liu D, Wang S, Liao H, Zhang B, Miao Q, Peng Y (eds.) 2017 Prognostics and System Health Management Conference. Prognostics and System Health Management Conference, pp. 987–992. Ieee, New York
7. Qu JL, Yu L, Yuan T et al (2018) Adaptive fault diagnosis algorithm for rolling bearings based on one-dimensional convolutional neural network. Chin J Sci Inst 39(07):134–143
8. Jia F, Lei Y, Lin J et al (2016) Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of

rotating machinery with massive data. Mech Syst Signal Process 72:303–315
9. Zhang WY, Yu DJ, Chen XM (2013) Fault diagnosis of rolling bearings based on resonance-based sparse signal decomposition and energy. Proc CSEE 33(20):111-118+19
10. Xia S, Zhang J, Ye S et al (2019) A mechanical fault detection strategy based on the doubly iterative empirical mode decomposition. Appl Acoust 1:346–357
11. Żak G, Wyłomańska A, Zimroz R (2019) Periodically impulsive behavior detection in noisy observation based on generalized fractional order dependency map. Appl Acoust 144:31–39. https://doi.org/10.1016/j.apacoust.2017.05.003
12. Zhao X, Patel TH, Zuo MJ (2012) Multivariate EMD and full spectrum based condition monitoring for rotating machinery[J]. Mech Syst Signal Process 27(1):712–728
13. Zhao HW, Norden EH (2009) Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv Adapt Data Anal 1(1):1–41
14. Smith JS (2005) The local mean decomposition and its application to EEG perception data. J R Soc Interface 2(5):443–454
15. He L, Lin J, Ding J et al (2017) Empirical mode decomposition envelope technique and mode mixing problem in amplitude modulation-frequency modulation signals. J Mech Eng 53(2):1–10
16. Selesnick IW (2011) Resonance-based signal decomposition: A new sparsity-enabled signal analysis method. Signal Process 91(12):2793–2809
17. Zhang WY, Yu DJ, Chen XM (2015) Energy operator demodulating of signal's resonance components for the compound fault diagnosis of gearbox. J Vibr Eng 28(1):148–155
18. Xia JZ, Zhao L, Bai YC et al (2017) Fault feature extraction of rolling element bearings based on Teager energy operator and ZFFT. J Vibr Shock 36(11):106–110
19. Li YB, Meng FL, Lu YJ (2009) Research on Rub impact Fault Diagnosis Method of Rotating Machinery Based on EMD and SVM. 2009 IEEE International Conference on Mechatronics and Automation, Vols 1–7, Conference Proceedings, 4806–4810. https://doi.org/10.1109/Icma.2009.5246424
20. Wang XM, Wang J, Privault M (2018) Artificial intelligent fault diagnosis system of complex electronic equipment. J Intel Fuzzy Syst 35(4):4141–4151
21. Wen L, Li X, Gao L et al (2018) A new convolutional neural network-based data-driven fault diagnosis method. IEEE Trans Industr Electron 65(7):5990–5998
22. Shao H, Jiang H, Zhang H et al (2018) Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing. Mech Syst Signal Process 100:743–765
23. Zhu X, Luo X, Ye X et al (2021) State recognition method of turbine rotor based on depth feature learning. Proc CSEE 41(2):432–442