RESEARCH ARTICLE

Bearing Fault Diagnosis Method Based on Attention Mechanism and Multi-Channel Feature Fusion

HONGFENG GAO[1], JIE MA[2], ZHONGHANG ZHANG[3], AND CHAOZHI CAI[2]

[1]Handan Branch of Hebei Special Equipment Supervision and Inspection Institute, Handan 056000, China
[2]School of Mechanical and Equipment Engineering, Hebei University of Engineering, Handan, Hebei 056038, China
[3]MCC Baosteel Technology Service Company Ltd., Shanghai 200941, China

Corresponding author: Chaozhi Cai (caichaozhi1983@163.com)

ABSTRACT To address the problems of limited identification accuracy and poor generalization ability of bearing fault diagnosis models, a convolutional neural network model for bearing fault diagnosis based on convolutional block attention module and multi-channel feature fusion (CBAM-MFFCNN) is proposed. The method uses signal processing technology to convert one-dimensional vibration signal into three types of two-dimensional time-frequency images, and constructs a network with multi-channel input to learn the three types of images at the same time. To realize the accurate fault diagnosis of bearings in strong noise environment, the structural parameters of the network are optimized. By adding different degrees of Gaussian white noise to the vibration signal, the convolution kernel size and the step of the first layer of the model are optimized. In order to improve the feature extraction ability and generalization performance of the model, the variable load dataset is constructed for training and testing. Experiments are conducted based on the Case Western Reserve University (CWRU) bearing datasets, the experimental results show that compared with the single channel diagnosis model, CBAM-MFFCNN can not only realize accurate identification of bearing fault, but also achieve 100% identification accuracy in fault degree testing.

I. INTRODUCTION

By replacing the sliding friction between parts with rolling friction, rolling bearings not only reduce the energy loss in the internal energy form of the equipment, but also greatly improve the service life of parts, so they are widely used in various fields [1], [2]. However, bearings often operate at high speed in harsh working conditions such as high temperature and pressure, and are prone to failure due to long-term impact alternating loads, mechanical wear and thermal fatigue [3].

Due to the complexity of actual work environment, bearing fault signals have non-stationary characteristics, so it is difficult to extract bearing fault characteristics solely based on time domain or frequency domain signals, and it is more

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

prone to noise interference and reduce the accuracy. Time-frequency analysis method can transform the original bearing vibration signal into frequency domain, and explore their joint distribution information, and clearly display the relationship between signal frequency and time. The typical time-frequency analysis methods include short-time Fourier transform (STFT), wavelet transform (WT), stockwell transform (ST), empirical mode decomposition (EMD) [4], etc.

With the improvement of computer performance and the development of big data technology, data-driven fault diagnosis algorithms have been widely concerned, among which deep learning stands out with its strong feature extraction ability and adaptive ability. In addition, various new technologies are also being developed in the fault detection of power equipment, such as using infrared spectroscopy to distinguish electrical and thermal faults in transformer oil

[5] and using ultraviolet-visible spectroscopy to estimating health index of the transformer [6]. Compared with traditional machine learning methods such as artificial neural networks (ANN) [7], sparse representation [8], support vector machine, etc, research based on deep learning methods has gradually become a hot topic in the field of bearing fault diagnosis. Early deep learning methods such as auto-encoder (AE) [9], deep Boltzmann machine (DBM) [10] and deep belief networks (DBN) [11] have been widely used.

With the continuous development of deep learning, various deep neural network models with high efficiency and complex structure have been proposed. Convolutional neural network (CNN) [12] and recurrent neural network (RNN) [13] are the best developed and most widely used deep learning methods.

Due to the one-dimensional nature of bearing vibration signals, earlier studies are usually carried out by using one-dimensional CNN and RNN. Saghi et al. [14]designed a neural network combined with the feature attention mechanism; this model used three parallel convolutional neural networks with different filter lengths, which can efficiently extract and combine spatial and temporal features of input signals at different frequencies, and solve the problems of vibration information loss and poor anti-noise performance that may exist in single-scale convolution. In order to better extract fault features from bearing vibration signals containing strong noise, Yan et al. [15] proposed a novel anti-noise multi-scale convolutional neural network named AM-CNN. In this model, a residual pre-processing block was proposed according to the principle of noise superposition, and a new loss function was constructed; meanwhile, multi-scale convolution blocks were used to achieve multi-scale feature extraction; in addition, the model used multiple labels to distinguish multiple bearing faults simultaneously, and achieved good recognition results on the constructed noisy composite dataset. Liu et al. [16] proposed a model combined deep convolutional neural networks with wide first-layer kernels (WDCNN) and deep long short-term memory networks (DLSTM), which called WDCNN-DLSTM. This model addressed the issue of traditional fault diagnosis methods cannot fully utilize the temporal nature of data. WDCNN expanded the kernel size of the first layer of traditional CNN and enhanced the model's ability to extract spatial features from one-dimensional vibration signals; DLSTM stacked multiple LSTM modules to enhance the model's ability to extract temporal information from one-dimensional vibration signals; WDCNN-DLSTM utilized their complementary advantages to connect these two models through fusion layers and enhanced the discriminative ability of the model; The experimental results showed that compared with other models, this method has higher accuracy and can still achieve better classification performance even under variable load conditions.

In recent years, in order to better extract the features of vibration signals, the method of extracting the features of vibration signals from high dimensional scales has been gradually adopted [17], and the two-dimensional representation of signals has been widely used in various fault diagnosis research [18]. Zhao et al. [19] proposed a new signal-to-image mapping method, which can convert one-dimensional vibration signals into two-dimensional gray images and establish a convolutional neural network model to extract fault features from gray images and realize fault classification. Yu and Yao [20] proposed an end-to-end lightweight quantitative modeling framework based on ensemble convolutional neural networks. This method has a lower root mean square error (54% and 73%, respectively) for predicting principal elements. They conducted in-depth research on the internal learning mechanism of deep CNN models and validated their proposed method through experiments. Lei et al. [21] proposed a new method based on the combination of Markov Random Field (MRF) and multidimensional CNN in order to deal with the situation of complex actual conditions and few available sample datasets. This model can achieve high fault diagnosis accuracy under variable operating conditions and small sample datasets.

At present, a large number of scholars have applied modern signal processing technology to this field. Among them, the time-frequency analysis method has been widely used because it can express the bearing signal as a two-dimensional function related to time and frequency, and better show the characteristics of the signal in the time domain and frequency domain. However, the traditional method of simply applying time-frequency analysis to extract fault characteristics and designing mathematical models for classification has the following disadvantages.

(1) Overreliance on specific feature extraction without effectively utilizing all information from the original data. When the feature extraction operation is carried out under the definite framework, the extracted target feature usually needs to make a choice among many features, so it is easy to cause the loss of the original information.

(2) Rely on professional background knowledge. The application of time-frequency analysis method to the original signal for noise reduction, filtering and feature extraction puts forward higher requirements for the knowledge level of researchers. The difficulty mainly includes the selection of parameters such as signal characteristics, window function, signal decomposition and layer number of reconstructions according to the bearing type and actual operating condition. These parameters play a crucial role in the effect of fault feature extraction. If the parameters of any step are not selected properly, it is difficult to effectively classify the subsequent signals.

(3) Low processing efficiency. With the advent of the era of big data, on the one hand, due to the explosive growth of monitoring data for electromechanical equipment, it is unrealistic to rely solely on manual feature extraction in the context of massive data; on the other hand, with the complexity of electromechanical equipment, different types and structures

of bearings have emerged, and traditional methods relying on manual feature extraction and classification may not be able to be implemented on new equipment.

To solve the above problems, this paper proposed a convolutional neural network model for bearing fault diagnosis based on convolutional block attention module and multi-channel feature fusion (CBAM-MFFCNN). The model took three types of time-frequency feature images as input and applied the attention mechanism to the model, allowing the model to pay more attention to the information of the fault frequency region corresponding to each time-frequency image. Based on the comprehensive reference of the three types of time-frequency feature images generated by the same signal, the bearing fault category and fault degree can be discriminated, and the generalization ability of the model under complex working conditions can be further improved. The specific contributions and innovations are as follows.

(1) Optimize data preprocessing operations. Multi-class modern signal processing technology is used to preprocess the original signal at the same time, and the time-frequency information of the signal is fully explored from various dimensions.

(2) Research on multi-channel networks. The time-frequency images generated after multi-class signal processing are used as the input of the convolutional neural network in the form of multi-channels. The powerful learning ability of the convolutional neural network is utilized to fuse and extract the features of all kinds of time-frequency signals describing the same vibration signal.

(3) Attention mechanism is used to build efficient feature fusion network. By utilizing attention mechanism, spatial and channel optimization is performed on feature maps, and efficient feature fusion is achieved between input maps.

(4) The comprehensive ability of the model under complex industrial application conditions has been evaluated. All kinds of real conditions are simulated, and the noise resistance experiment, variable load experiment and bearing fault location experiment are set respectively.

II. DATA PREPROCESSING

Data preprocessing refers to the processing of original data before inputting it into the model, including data cleaning, conversion, feature selection, etc. [22]. The selection of preprocessing methods is usually aimed at helping network models achieve optimal performance.

A. EXPERIMENTAL DATA

The experimental data comes from the rolling bearing data center of Case Western Reserve University (CWRU) [23], and the CWRU dataset sampling platform is shown in Fig. 1. All experiments carried out in this paper were based on the CWRU dataset, and the bearing used in this paper was SKF6205. In the training stage of the model, 12 kHz vibration data of the drive end bearing was used.
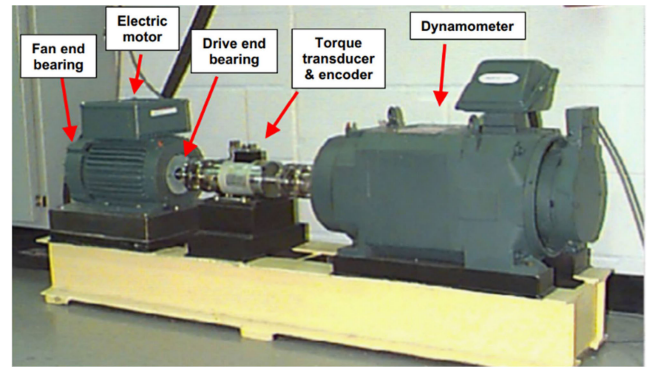

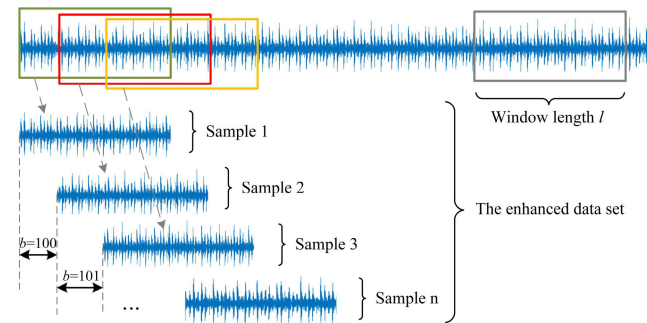**FIGURE 1.** CWRU bearing data sampling platform [23].

FIGURE 2. Signal overlapping sampling method with variable step.

B. DATA ENHANCEMENT AND STANDARDIZATION

Data enhancement generates more isomorphic data through a series of transformations on the training data, which can effectively reduce the overfitting phenomenon of the network and make the trained network have stronger generalization ability [24]. In this paper, a variable step overlapping sampling method was adopted. The step b satisfied the following condition.

$$b = 100 + n \tag{1}$$

where, n represents the remainder obtained after dividing the current number of intercepted samples by 300. The effect of data enhancement with variable step overlapping sampling is shown in Fig. 2 [25].

In the training of fault diagnosis models based on neural networks, it is important to maintain a balance in the total number of different types of samples as much as possible. Otherwise, it will reduce the generalization ability of the network [26]. In this paper, the sample quantity was obtained by the variable step overlapping sampling method. Details are shown in Table 1.

Finally, Z-Score standardization was used to process the original data, and the values were converted to a normal distribution with the mean of 0 and the variance of 1.

C. SPECIALLY CONSTRUCTED TEXTURE IMAGES

This paper proposed a texture image that arranged the time domain signal according to the features of the frequency

TABLE 1. Samples after data enhancement.

Bearing condition	Diameter of damage （inch）	Load （hp）	One-dimensional sample size	Number of samples for single Class condition	Number of single bearing status samples
Normal	0	0,1,2	1×2048	1200	3600
Inner race fault	0.007, 0.014, 0.021	0,1,2	1×2048	400	3600
Outer race fault	0.007, 0.014, 0.021	0,1,2	1×2048	400	3600
Rolling element fault	0.007, 0.014, 0.021	0,1,2	1×2048	400	3600

FIGURE 3. Generation of texture image.

domain, and converted the features of the signal into a texture image that was more suitable for 2D convolution kernel extraction. The size of the generated two-dimensional image was set to 128 × 128. The specific generation process is as follows.

The bearing speed range of the CWRU dataset used was 1750-1797 r/min. The deep groove ball bearing used in this paper contains 9 balls, the pitch diameter D is 1.537 inch, the rolling body diameter d is 0.3126 inch, and the contact angle is 0°. The calculation results show that the frequency f_i ranges of the inner race failure is 157.9-162.2 Hz, the frequency f_o ranges of the outer race failure is 104.6-107.4 Hz, and the frequency f_d ranges of the rolling element failure is 137.5-141.2 Hz.

It is assumed that the calculated value of each fault frequency corresponding to the diagnosed bearing is $f_1, f_2, \ldots, f_n$, where $f_{\max}$ and $f_{\min}$ are their maximum and minimum values respectively, then

$$W = \left[\frac{3}{4}f_{\min}, \frac{4}{3}f_{\max}\right] \quad (2)$$

where, W represents the interval to be extracted. The feature extraction interval of CWRU dataset used in this paper is 78.4-216.2 Hz [27].

According to the obtained feature frequency, the interval W can be extracted and the correlation coefficient K with the longitudinal features of the image can be calculated. If the interval is assumed, the calculation formula of the correlation coefficient K is as follows

$$K = \left\lceil \left(\frac{N}{l} - \frac{N}{h}\right) \div 120 \right\rceil \quad (3)$$

where, K represents the correlation coefficient of texture image in time and frequency; N is sampling frequency (Hz); h represents the upper boundary value of W (Hz). l represents the lower bound value of W (Hz). The smaller the value of K, the higher the accuracy of the longitudinal frequency information in the generated texture image, but the smaller the corresponding feature interval. The construction of the texture image is shown in Fig. 3 [25].
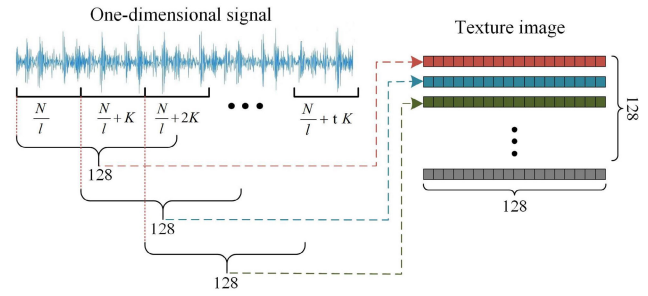
As shown in Fig. 3, the texture image is generated corresponding to a single one-dimensional signal of 2048 length obtained by data enhancement. Multiple 128-length digital signals are intercepted repeatedly on a single one-dimensional signal and longitude-arranged. The initial intercept point interval coefficient of adjacent signal segments is an arithmetic sequence of K.

Assuming that the one-dimensional signal before processing is $X = \{x_1, x_2, \cdots, x_{2048}\}$, the matrix expression of the corresponding generated two-dimensional texture image is shown in formula (4) [25]. The essence of this operation is to split and arrange a one-dimensional signal at a discrete interval frequency f at a selected interval. Assuming that the diagnosed bearing has a certain fault frequency $f \in W$, and the length of a cycle T corresponding to this frequency on the original signal is L, then L must meet formula (5).

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_{128} \\ x_{1+\frac{N}{l}} & x_{2+\frac{N}{l}} & \cdots & x_{128+\frac{N}{l}} \\ x_{1+2\frac{N}{l}+k} & x_{2+2\frac{N}{l}+k} & \cdots & x_{128+2\frac{N}{l}+k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1+t\frac{N}{l}+\frac{t^2k}{2}} & x_{2+t\frac{N}{l}+\frac{t^2k}{2}} & \cdots & x_{128+t\frac{N}{l}+\frac{t^2k}{2}} \\ x_1 & x_2 & \cdots & x_{128} \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix} \quad (4)$$

$$L \in \left[\frac{N}{l}, \frac{N}{l} + t_{\max}K\right] \quad (5)$$

where, $t_{\max}$ represents the coefficient corresponding to the interception of the last 128-length signal, L must be close to a certain interval value [25]. The sequence intercepted by the interval value will have a high degree of similarity with the sequence on both sides, forming stripes, which constructs the fault features that the two-dimensional convolution kernel needs to extract. The time domain, frequency domain (local), and texture images corresponding to each fault category are shown in Fig. 4.

D. TIMT-FREQUENCY FEATURE IMAGE BASED ON DISCRETE BINARY WAVELET TRANSFORM
In this paper, Dobessie wavelet was used, the vanishing moment was 2; the standardized one-dimensional signal with the length of 2048 was decomposed by discrete binary
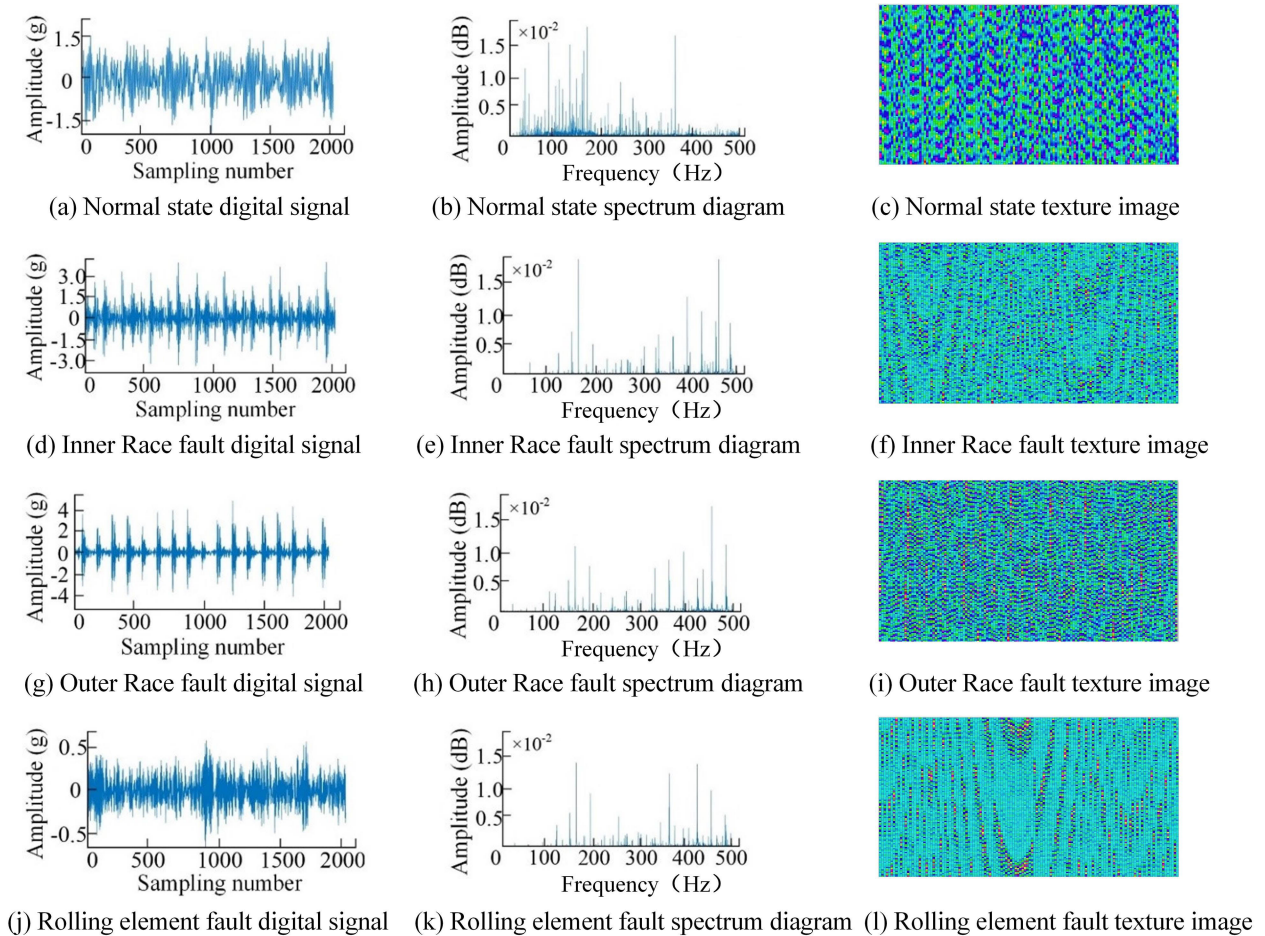
(a) Normal state digital signal

(b) Normal state spectrum diagram

(c) Normal state texture image

(d) Inner Race fault digital signal

(e) Inner Race fault spectrum diagram

(f) Inner Race fault texture image

(g) Outer Race fault digital signal

(h) Outer Race fault spectrum diagram

(i) Outer Race fault texture image

(j) Rolling element fault digital signal

(k) Rolling element fault spectrum diagram

(l) Rolling element fault texture image

FIGURE 4. Time, frequency and texture images of various fault category.

wavelet transform (DWT), and the approximate component (cA) and a cluster of detailed components (cD) were obtained, which correspond to the low frequency component and high frequency component respectively in the frequency domain.

In order to convert the data transformed by wavelet into 128×128, a special arrangement operation was carried out. Firstly, each component was expanded to 2048×1 by copying its value several times. For example, the length of cD1 is 1024, and the length needs to be expanded by two times, so each data point needs to be copied twice and sorted according to the original sequence. Then the generated signal was added to the original signal to form a size of 8×2048, which can be transformed into a size of 128×128 by shape adjustment. The process of converting one-dimensional signals into two-dimensional images through DWT is shown in Fig. 5.

DWT time-frequency images generated by bearing signals with four different health states are shown in Fig. 6. As can be seen from Fig. 6, DWT time-frequency images under four different working conditions have obvious differences in each frequency band, and there is a large degree of differentiation among each image. These images represent the bearing
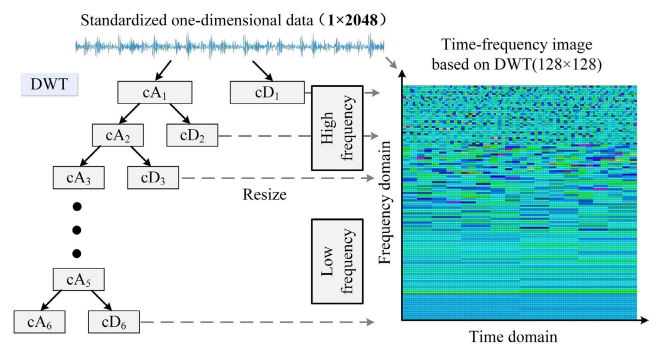

FIGURE 5. The process of converting one-dimensional signals into two-dimensional images through DWT.

vibration signal from a different angle from the texture image.

E. TIMT-FREQUENCY FEATURE IMAGE BASED ON EMPIRICAL MODE DECOMPOSITION

In this paper, the empirical mode decomposition (EMD) method was used to decompose bearing signals. Due to the
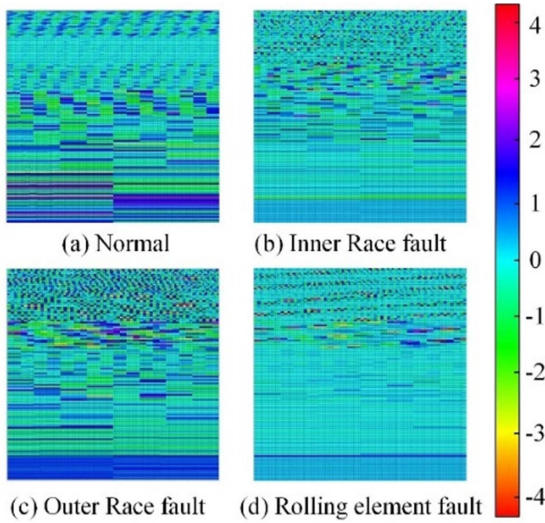
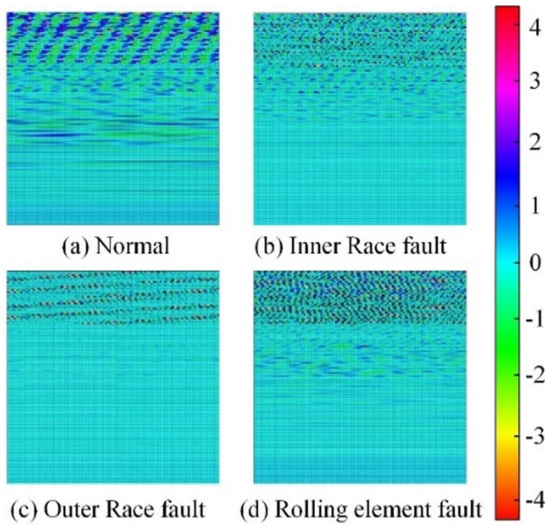**FIGURE 6.** DWT time-frequency images of various vibration signals.

FIGURE 7. EMD time-frequency images of various vibration signals.

number of layers that can be decomposed from CWRU bearing data under various operating conditions was different, but all of them could reach more than 6 layers, so all kinds of signals were screened for 6 times, resulting in 6 Intrinsic mode function (IMF) components and 1 residual component.

In order to convert this cluster of signals into two-dimensional images with size of 128×128, firstly, original signal with size of 8×2048 was added and arranged horizontally in the order of the original signal, IMF1, IMF2,..., IMF6, and residual components, then, the signal was adjusted to size of 128×128 to generate an EMD based time-frequency image. The EMD time-frequency images corresponding to bearing fault signals under various operating conditions are shown in Fig. 7.

Through the visualization of images, the difference between various vibration signals can be effectively observed.
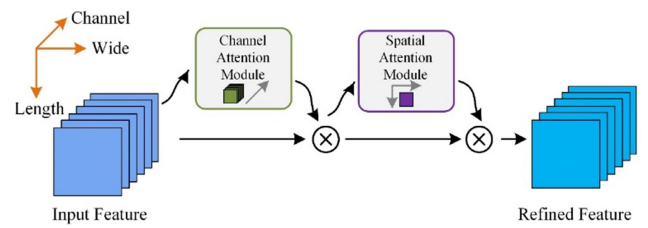

FIGURE 8. CBAM structure.

The main difference lies in the middle and upper part of the image, which corresponds to the original signal and its first few IMF components, indicating that the first few IMF components contain more characteristic information. The latter few IMF components also contain some important information, but the numerical difference is smaller. Limited by the accuracy of the color band, it is difficult to observe the difference of the corresponding image region with the naked eye, but the neural network can capture the subtle difference and then use it as the basis for fault diagnosis. In general, the time-frequency image method based on EMD can effectively convert various vibration signals into time-frequency images with great differences.

III. CONVOLUTIONAL NEURAL NETWORK BASED ON ATTENTION MECHANISM AND MULTI-CHANNEL FEATURE FUSION

A. ATTENTION MECHANISM

The function of the attention mechanism in deep learning is to distinguish the information that is more critical to the current task from numerous data, so it is essentially similar to the selective visual attention mechanism of human beings [28]. Attention mechanism can be divided into soft attention mechanism and hard attention mechanism. This paper used the convolutional block attention module (CBAM) in the soft attention mechanism.

B. CONVOLUTIONAL BLOCK ATTENTION MODULE

CBAM is a soft attention mechanism module proposed by Woo et al in 2018 [29], which mainly consists of channel attention module (CAM) and spatial attention module (SAM). The two submodules adopt series structures, which enable the network to focus the spatial and channel information of image features. The infrastructure is shown in Fig. 8.

1) CHANNEL ATTENTION MODULE

The function of the CAM is to give more weight to the channel where the key information is located, so that the limited computing resources of the computer can concentrate more on the channel containing the key features. The topological structure of CAM is shown in Fig. 9.

The upper input feature map F has three dimensions: length H, width W and channel number C. CAM firstly performs global maximum pooling (GMP) and global average pooling (GAP) on the feature map according to the model
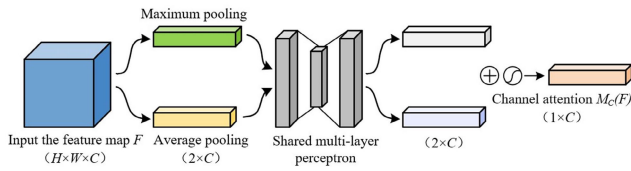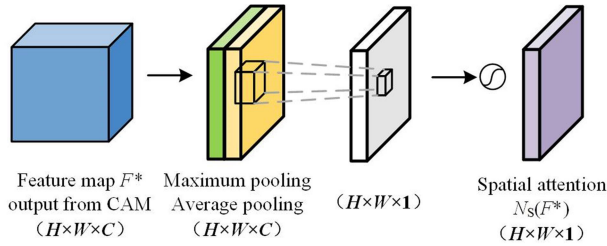
FIGURE 9. Topological structure of CAM.

FIGURE 10. Topological structure of SAM.

channel, thereby aggregating the spatial information of the feature maps and obtaining the corresponding output F_{Avg} and F_{Max}. For a certain channel c, u_c^{Avg} is used to represent the average value of the channel, and u_c^{Max} is used to represent the maximum value of the channel. Then a multilayer perceptron (MLP) with shared weights is used to exchange information between u_c^{Avg} and u_c^{Max}, and the output features of the perceptron are activated by Sigmoid function, so the channel weight coefficient $M_c(F)$ can be obtained [29]. The process can be expressed as

$$M_c(F) = Sigmoid(MLP[Avgpool(F)] + MLP[Maxpool(F)]) \tag{6}$$

The obtained channel weight coefficient is used to express the importance of each channel, which is multiplied with the input features element by element, and the feature map F^* optimized by the channel attention module can be obtained.

$$F^* = M_c(F) \otimes F \tag{7}$$

where, $\otimes$ represents the multiplication of corresponding positions element by element [29].

2) SPATIAL ATTENTION MODULE
The input of SAM is the feature map F^* optimized by the CAM. The function of SAM is to allow the network to learn information about the importance of each region of the feature map, so as to effectively extract features from the specified region. The topological structure of SAM is shown in Fig. 10.

The SAM firstly performs global maximum and global average pooling operations on the feature map F^* in the channel dimension to obtain two layers feature maps with size of $H \times W$. The two layers feature maps are spliced according to the direction of channel dimension and input into the convolutional layer with size of 7×7. The spatial attention weight $N_s(F^*)$ of the feature map can be obtained by applying

the Sigmoid function for activation [29]. The process can be expressed as follows

$$N_s(F^*) = Sigmoid\left(f^{7 \times 7}\left[Avgpool(F^*); Maxpool(F^*)\right]\right) \tag{8}$$

The obtained spatial attention weight $N_s(F^*)$ is used to indicate the importance of each position in the image plane of the feature map. Then this weight is multiplied F^* element by element to obtain the optimized feature map F^{**} through the spatial attention module [29]:

$$F^{**} = N_s(F^*) \otimes F^* \tag{9}$$

C. NETWORK STRUCTUTR OF THE CBAM-MFFCNN
The CBAM-MFFCNN model took three types of time-frequency feature images as input, and used the convolutional neural network which has the powerful feature fusion and feature extraction capabilities to deeply mine the fault features contained in different expression forms among the images. Based on the comprehensive reference of the three types of time-frequency feature images generated by the same signal, the bearing fault category and fault degree were determined. The topological structure of the model is shown in Fig. 11 and Fig. 12.

The whole model includes data preprocessing, feature fusion, feature extraction and classifier. The data preprocessing mainly includes data enhancement, normalization of the original signal and the generation of three types of time-frequency feature images. The data was converted by three different ways and adjusted to the input size required by the network. The CBAM was added to the three channels of the feature fusion part of the network (as shown in Fig. 12). The output of the pooling layer from the three channels was stacked to complete the whole process of feature fusion. After the fusion, the feature map was extracted by multi-layer convolution. In this process, the size of the feature map gradually decreased and the number of corresponding channels increased. The output of the last convolutional layer was processed by GAP and then fed into the classifier composed of the fully connected layer, and the prediction results can be obtained by activating the Softmax function at the end of the classifier.

The size of the first layer of convolutional kernel was not determined at the beginning. Previous studies such as literature [30] have shown that for one-dimensional signals, the use of the first layer of large convolutional kernel can effectively increase the noise resistance of the model. In order to test whether this rule is applicable to image-based networks, relevant experiments as shown in Table 2 were conducted.

In the experiments, the step size of the first convolutional layer of the model was 4, and the number of channels was 32. The size of the convolutional kernel was obtained by comparing the diagnosis accuracy of models with different sizes of convolutional kernel under different levels of noise environment by using the control variable method while keeping other hyperparameters unchanged. The default step of
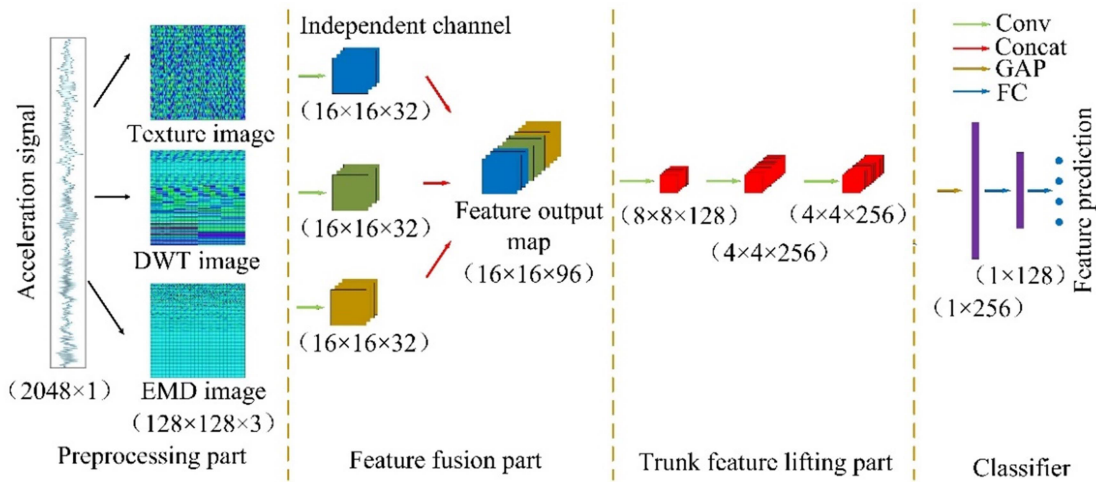
FIGURE 11. Topological structure of multi-channel feature fusion convolutional neural network (MFFCNN).
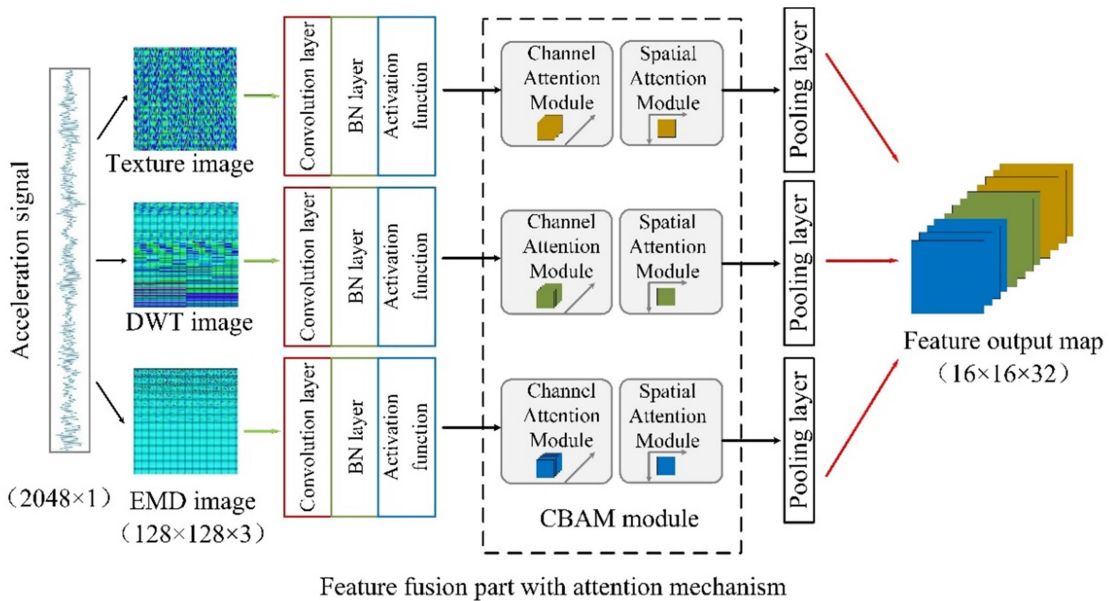

FIGURE 12. Feature fusion network with attention mechanism.

the first convolutional layer was set to 4. Experiments were conducted three times for each group and the mean value was obtained. The test results are shown in Table 2. On the basis of the experimental results in Table 2, the structure parameters of CBAM-MFFCNN were finally determined; it is shown in Table 3.

As shown in Table 2, when the Signal-to-Noise Ratio (SNR) is high, the size of the first-layer convolution kernel has little effect on the anti-noise performance of the model, and models with different convolution kernel sizes can make a good judgment on the fault degree of the bearing. However, with the decrease of SNR, that is, with the proportion of noise increases continuously, the recognition accuracy of models with smaller first layer convolution kernel decreases rapidly.

For example, when the SNR is 0 dB, the accuracy of the model with convolution kernel size of 8 × 8 is only 36.42%, and the prediction results under this accuracy have almost no reference value. The recognition accuracy of the model with convolution kernel size of 128×128 is 99.75% under the same condition. In addition, the larger the convolution kernel size is not the better. When the size of the first-layer convolution kernel exceeds 128 × 128, the anti-noise performance of the model began to decline. Moreover, as the convolution kernel size increases, the parameters of the model also increase rapidly, and the corresponding training time is longer. Therefore, the convolution kernel size of the multi-channel feature fusion network proposed in this paper was set to 128 × 128.

TABLE 2. Effect of different sizes of convolutional on noise immunity of model.

Convolution Kernel Size	Signal-to-Noise Ratio (dB)								
	10	6	4	2	0	-2	-4	-6	-10
8×8	98.61%	91.32%	78.33%	59.35%	36.41%	30.63%	25.37%	17.62%	16.34%
16×16	100%	97.64%	87.82%	70.32%	47.22%	38.43%	29.42%	23.03%	18.33%
24×24	100%	99.44%	90.72%	78.43%	55.20%	42.06%	32.08%	25.64%	21.01%
32×32	100%	99.61%	95.02%	83.54%	68.80%	50.76%	37.22%	27.66%	19.33%
40×40	100%	99.54%	94.35%	81.76%	71.04%	66.88%	42.70%	31.62%	22.11%
48×48	100%	99.54%	97.82%	87.29%	72.45%	62.34%	45.79%	33.00%	25.74%
56×56	100%	99.40%	97.04%	90.23%	70.81%	58.33%	41.62%	33.38%	26.34%
64×64	100%	100%	99.89%	95.07%	88.78%	74.21%	61.04%	40.12%	34.54%
80×80	100%	100%	99.91%	93.42%	87.20%	74.33%	64.33%	49.51%	32.78%
96×96	100%	100%	100%	98.25%	94.32%	93.75%	81.94%	56.39%	37.94%
112×112	100%	100%	100%	98.40%	97.71%	95.21%	84.86%	53.75%	35.90%
128×128	100%	100%	100%	100%	99.75%	95.14%	82.22%	56.03%	35.76%
144×144	100%	100%	98.59%	94.33%	82.80%	70.46%	64.33%	53.10%	33.80%

TABLE 3. structural parameters of CBAM-MFFNN.

Number	Network Layer	Convolution Kernel Size/Number	Stride	Output（Length × Width × Depth）	Dropout	Number of Trainable Parameters
1-a	Convolution layer	128×128 / 32	4	32×32×32	/	524320
1-b	Convolution layer	128×128 / 32	4	32×32×32	/	524320
1-c	Convolution layer	128×128 / 32	4	32×32×32	/	524320
2-a	CBAM module	/	/	32×32×32	/	651
2-b	CBAM module	/	/	32×32×32	/	651
2-c	CBAM module	/	/	32×32×32	/	651
3-a	Pooling layer	2×2 / 32	2	16×16×32	/	0
3-b	Pooling layer	2×2 / 32	2	16×16×32	/	0
3-c	Pooling layer	2×2 / 32	2	16×16×32	/	0
4	Convolution layer	3×3 / 128	1	16×16×128	/	110720
5	Pooling layer	2×2 / 128	2	8×8×128	/	0
6	Convolution layer	3×3 / 256	1	4×4×256	/	295168
7	Pooling layer	2×2 / 256	2	4×4×256	/	0
8	Convolution layer	3×3 / 256	1	4×4×256	/	590080
9	Global average pooling layer	/	/	256	/	0
10	Fully connected layer	/	/	128	0.3	32896
11	Softmax	/	/	4/10	/	516/1290

In addition, in order to verify the feasibility of the data preprocessing method, a single-channel convolutional neural network was first built. The neural network has a total of 10 layers, including a feature extractor composed of
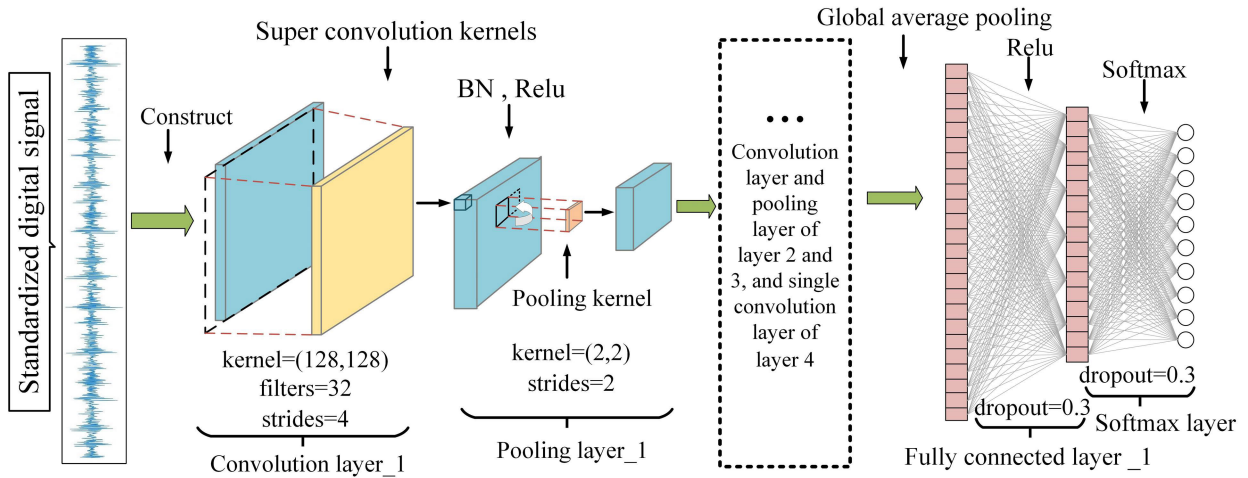
FIGURE 13. Topological structure of single-channel input convolutional neural network.

4 layers of convolutional and pooling layers and a classifier composed of two layers of fully connected layers. The topology of the single channel convolutional neural network is shown in Fig. 13. The performance comparison between the single-channel convolutional neural networks and the multi-channel convolutional neural networks can be seen in section IV.

IV. EXPERIMENT AND ANALYSIS

All experiments performed in this paper were conducted on CWRU datasets. The deep learning framework used for model training was TensorFlow [31] developed by Google, and the model was trained on a computer whose GPU was NVIDIA GeForce RTX3060Ti and CPU was i5-13400F. Python was used to build the above convolutional neural network model under the TensorFlow framework, and then the labeled datasets were randomly scrambled and sent into the neural network in batches for model training. The Mini-batch size was set to 128, the optimization algorithm was set to Adam, and the learning rate was set to 0.001.

A. FAULT CATEGORY AND DEGREE IDENTIFICATION

Firstly, the effectiveness of the preprocessing method was verified based on the single-channel input convolutional neural network, and a 4 categories experiment was set up to verify the fault recognition ability of the model. In the experiment, the training set, validation set and test set were divided into a ratio of 8:1:1; the verification frequency of the verification set in the training process was 1 time/Epoch; each experiment was conducted three times and the mean was taken as the final result. Secondly, considering the actual application requirements, and in order to better understand the performance upper limit of the network based on various datasets, the model was expanded to a 10 classifier, that is, it is required that the model not only identifies the category of bearing faults, but also identifies the degree of bearing faults. Finally,
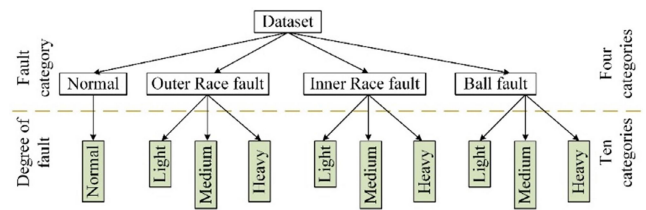

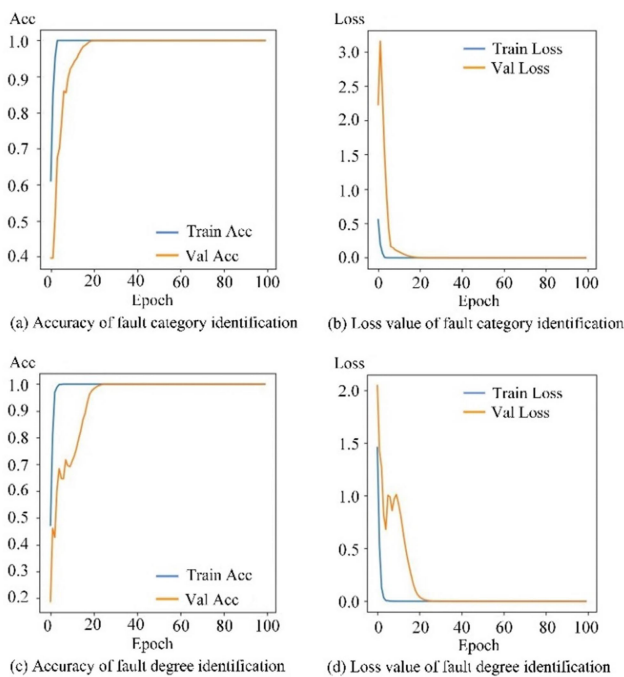**FIGURE 14.** Fault category and degree identification.

while keeping the overall structure of the model unchanged, only the label and network output were adjusted for verification. The difference in the experiment of identifying the category and degree of bearing faults is shown in Fig.14, and the training results of single-channel input convolutional neural network are shown in Table 4.

From Table 4, it can be seen that in the fault category identification experiment, for the texture image, DWT time-frequency image, and EMD time-frequency image datasets, the accuracy of the model on both the training and validation sets is 100%.This indicating that the single-channel input convolutional neural network model can effectively recognize bearing fault categories. However, in the fault degree identification experiment, Although the accuracy of the models for the three types of time-frequency image datasets has reached 100% on the training set, it could not reach 100% on the validation set. Different from the training set, the verification set is not learned by the network and is closer to the unknown situation of the signal in practical application, so it is more referential.

The CBAM-MFFCNN model and the single-channel input network model have the same structure except for the feature fusion part. The parameters used for training and validating the CBAM-MFFCNN model are the same as those of the single-channel input model, and the identification results of fault category and fault degree are shown in Fig. 15.
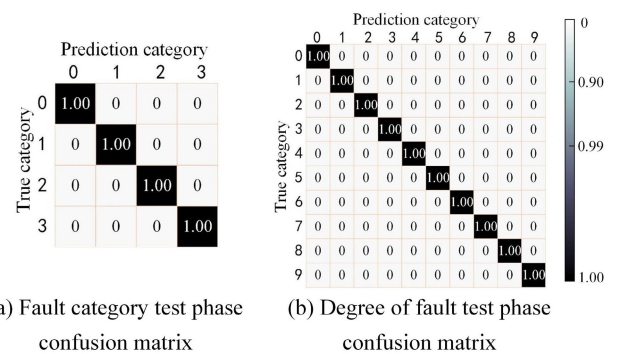
TABLE 4. Identification results of fault category and degree of single-channel input convolutional neural network.

Experiment Category	Number of Label Category	Dataset Category	Training Set Accuracy	Training Set Loss	Validation Set Accuracy	Validation Set Loss	Duration of Each Diagnosis
Fault category identification	4	Texture image	100%	2.96×10^{-6}	100%	1.56×10^{-5}	16.3 ms
		DWT image	100%	2.12×10^{-6}	100%	3.05×10^{-4}	52.7 ms
		EMD image	100%	7.27×10^{-6}	100%	2.20×10^{-4}	46.4 ms
Fault degree identification	10	Texture image	100%	1.21×10^{-5}	99.75%	5.3×10^{-3}	17.2 ms
		DWT image	100%	9.92×10^{-6}	99.42%	7.1×10^{-3}	53.4 ms
		EMD image	100%	5.29×10^{-5}	99.62%	3.8×10^{-3}	47.3 ms

(a) Accuracy of fault category identification

(b) Loss value of fault category identification

(c) Accuracy of fault degree identification

(d) Loss value of fault degree identification

FIGURE 15. Training results of CBAM-MFFCNN.

(a) Fault category test phase confusion matrix

(b) Degree of fault test phase confusion matrix

FIGURE 16. Confusion matrix of CBAM-MFFCNN.

Form Fig 15, it can be seen that in the fault category identification experiment, the accuracy of the CBAM-MFFCNN model has reached 100% on both the training and validation sets, with loss of 4.96×10^{-6} and 9.43×10^{-6} respectively. The model converges after about 20 Epochs training, and the overall convergence is good, which is almost no different from the single-channel input model. However, in the fault degree identification experiment, the single-channel input model can no longer guarantee 100% recognition accuracy; in contrast, the CBAM-MFFCNN model can still achieve 100% recognition accuracy on both the training set and the validation set, with loss of 9.59×10^{-6} and 2.69×10^{-5} respectively.

Although the model training curve has a large fluctuation in the early stage, the model tends to converge after about 20 Epochs training. The total number of trainable parameters of the model is 2605034, and the average single validation time is 137 ms (this time includes the data preprocessing process and the network prediction process). By comparison with the data in Table 4, it can be proved that the CBAM-MFFCNN model has stronger ability to extract the bearing fault degree feature than the single-channel input model and can identify the bearing fault degree more accurately.

The confusion matrix of CBAM-MFFCNN model in fault category and fault degree identification test is shown in Figure 16. In Figure 16, label 0 indicates no fault, labels 1-3 indicate light, medium, and heavy inner race faults, labels 4-6 indicate outer race faults of different degrees, and labels 7-9 indicate rolling element faults of different degrees.

The visualization results of each label dataset after dimensionality reduction are shown in Fig. 17 and Fig. 18.

From Figure 17, it can be found that the overall distribution of each label data input to the network is relatively complex, and it is difficult to distinguish between each data; after passing through each pooling layer, its distribution gradually reaches linear reparability, and after passing through the
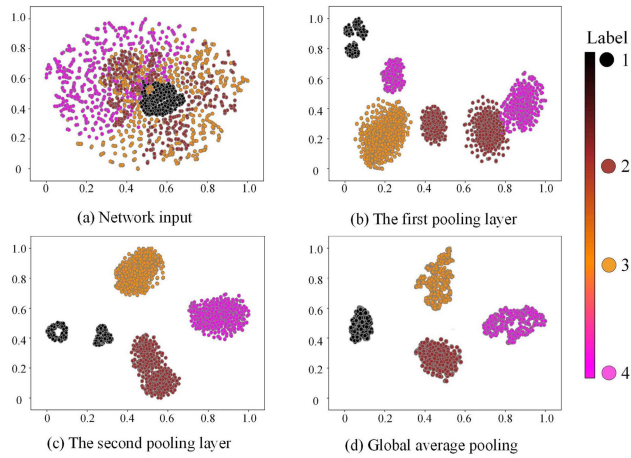
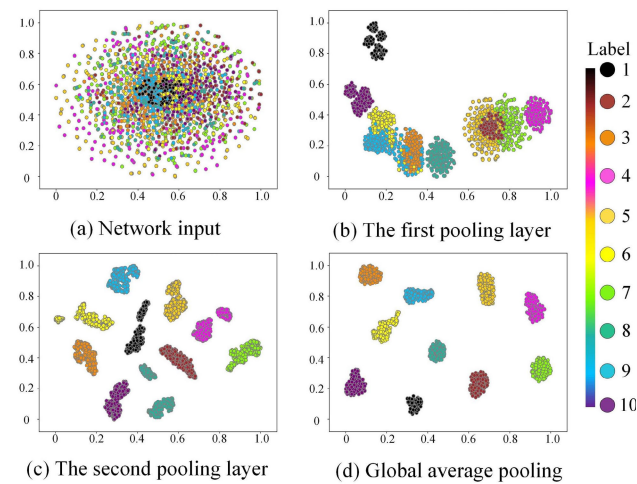**FIGURE 17.** Visualization result of fault category identification.

FIGURE 18. Visualization result of fault degree identification.

TABLE 5. Experimental setup for bearing fault diagnosis under variable load conditions.

Description of Domain	Source Domain	Target Domain
The first type of test	Train set A	Test set B、Test set C
	Train set B	Test set A、Test set C
	Train set C	Test set A、Test set B
The second type of test	Train set A、Train set B	Test set C
	Train set A、Train set C	Test set B
	Train set B、Train set C	Test set A

global average pooling layer, the output clustering effect is obvious.

In addition, it can be seen from Figure 18 that label 1 represents no fault signal, which is the most distinguishable in the figure compared with the signal with fault, indicating that the preprocessing operation proposed in this paper is helpful for the model to quickly diagnose the health status of the bearing.

From the above experimental results, it can be seen that the proposed CBAM-MFFCNN model can achieve better feature extraction for each label category data, and pass the extracted features to subsequent classifiers for identification. It is proved that the model has strong feature extraction and classification ability for CWRU datasets.

B. MODEL PERFORMANCE TESTING UNDER VARIABLE LOAD CONDITIONS

The change of the working load of rotating machinery is a very common phenomenon that cannot be ignored. When the load changes, the vibration response of the bearing will also

change, resulting in a large change in the collected signal. Therefore, it is of great practical significance to construct a highly generalized model trained under a certain load data to realize the effective fault diagnosis of bearing under various variable load conditions.

There are three types of loads in the bearing fault diagnosis experiment, namely 0 horsepower (hp), 1 hp and 2 hp, and the bearing fault datasets obtained under these three loads are A, B, and C. The specific settings of the experiment are shown in Table 5. From Table 5, it can be seen that the experiment is divided into two categories; one uses the dataset of a single load condition as the training set and uses the dataset of the other two load conditions as the testing set; the other uses the dataset of two load conditions as the training set and uses the dataset of the third load condition as the testing set. The two types of experiments were set as 10 categories identification, meaning that the model was required to distinguish the category and degree of bearing fault. The experimental results are shown in Fig. 19.

From Figure 19, it can be found that for the first type of experiment, the average accuracy of SVM based on FFT transform and VGG16 based on time-frequency images is less than 70%, which indicates that these two types of models overfitted the workload information contained in the signal when the training set only had single workload and failed to grasp the inherent common characteristics of the signal; when the distribution of test data in the target domain is different from the training data in the source domain, the performance of the model deteriorates rapidly. Therefore, the recognition ability of these two models may fail to recognize the unlearned load signal in practical application. In contrast, the average accuracy of WDCNN, MFFCNN, and CBAM-MFFCNN is 84.57%, 86.47%, and 91.95% respectively, this indicates that these three models have strong adaptability cross load domain and can better adapt to the situation where the signal load is relatively single due to conditional limitations in the actual signal acquisition process. Additionally, in the first type of experiment, the recognition performance of each model showed significant differences. When using either dataset B or dataset C as the training
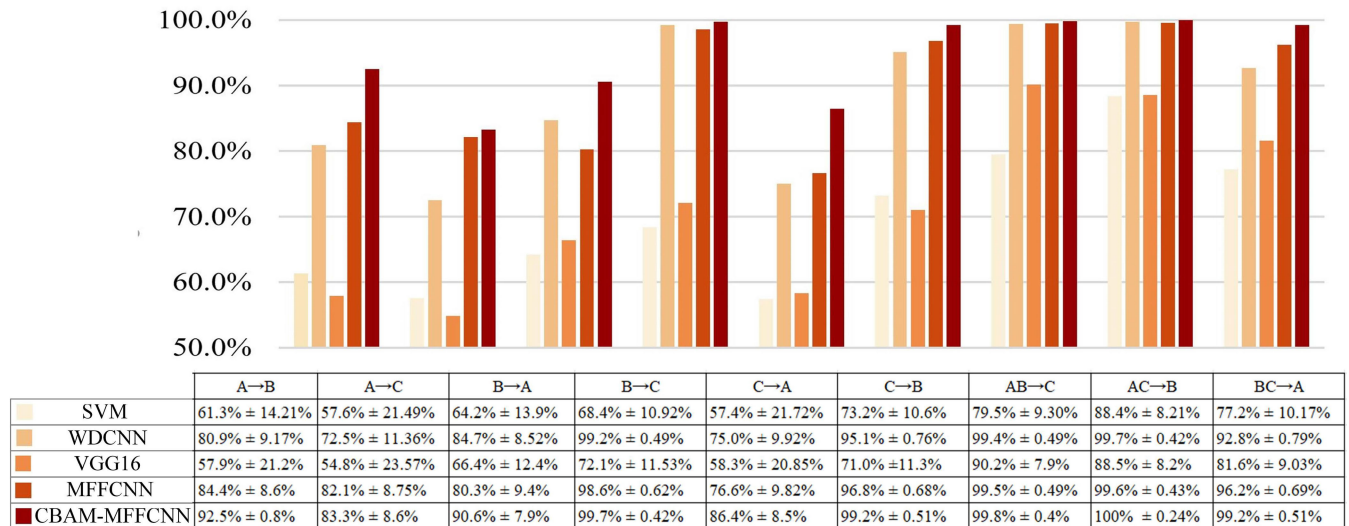
FIGURE 19. Diagnosis accuracy of different models under variable load conditions.

	A→B	A→C	B→A	B→C	C→A	C→B	AB→C	AC→B	BC→A
SVM	61.3% ± 14.21%	57.6% ± 21.49%	64.2% ± 13.9%	68.4% ± 10.92%	57.4% ± 21.72%	73.2% ± 10.6%	79.5% ± 9.30%	88.4% ± 8.21%	77.2% ± 10.17%
WDCNN	80.9% ± 9.17%	72.5% ± 11.36%	84.7% ± 8.52%	99.2% ± 0.49%	75.0% ± 9.92%	95.1% ± 0.76%	99.4% ± 0.49%	99.7% ± 0.42%	92.8% ± 0.79%
VGG16	57.9% ± 21.2%	54.8% ± 23.57%	66.4% ± 12.4%	72.1% ± 11.53%	58.3% ± 20.85%	71.0% ±11.3%	90.2% ± 7.9%	88.5% ± 8.2%	81.6% ± 9.03%
MFFCNN	84.4% ± 8.6%	82.1% ± 8.75%	80.3% ± 9.4%	98.6% ± 0.62%	76.6% ± 9.82%	96.8% ± 0.68%	99.5% ± 0.49%	99.6% ± 0.43%	96.2% ± 0.69%
CBAM-MFFCNN	92.5% ± 0.8%	83.3% ± 8.6%	90.6% ± 7.9%	99.7% ± 0.42%	86.4% ± 8.5%	99.2% ± 0.51%	99.8% ± 0.4%	100% ± 0.24%	99.2% ± 0.51%

set and the other as the testing set, the recognition accuracy obtained is usually higher. However, when dataset A appears in the training or testing set, the recognition rates of each model decrease. The reason may be that the difference between bearing fault data under no-load conditions is greater than that under load conditions. Therefore, in order to carry out efficient fault diagnosis of bearings under a certain load, bearing vibration signals similar to the working load should be collected as far as possible for model training.

By comparing the two types of experiment, it can be seen that the performance of various models trained based on the multi-load condition is higher than that of the model trained under the single-load condition. For SVM and VGG16, the average recognition accuracy has increased by 18.0% and 23.3% respectively based on multi-load training. Meanwhile, for WDCNN, MFFCNN, and CBAM-MFFCNN, the multi-load characteristics of the training set also improve their generalization performance. Among them, CBAM-MFFCNN performs the best, and the average recognition accuracy is 99.67%. Furthermore, when the training set includes all load conditions, the recognition accuracy of the model on testing set under each load can reach 100%. Therefore, when applying a deep learning model for bearing fault diagnosis, the collected training set signals should contain more load conditions, and rich variable load information can enable the model to better grasp the commonality of fault signals, thus avoiding overfitting fault characteristics under specified loads.

V. CONCLUSION
Considering the complexity of practical industrial applications, as well as the common problems of the existing bearing fault diagnosis models such as poor anti-noise performance, weak overall model cohesion and insufficient generalization

ability, this paper started from two aspects of data preprocessing and network structure, and introduced the attention mechanism in computer vision field. A multi-channel feature fusion model with strong noise resistance and generalization was constructed for efficient bearing fault diagnosis. The following conclusions are drawn.

(1) In order to reduce the loss of original signal features caused by data preprocessing and the problem of processed features not being learned by the network, this paper constructed three types of time-frequency feature images to fully explore the time-frequency information of bearing signals from various dimensions. The comparison experimental results between single-channel and multi-channel networks showed that multi-channel input networks had stronger fault feature extraction capability.

(2) In view of the problem that vibration signals collected in industrial environment inevitably included noise, which interfered with the model's extraction of bearing fault features, a feature fusion network with super-large convolution kernel in the first layer and strong noise resistance was proposed. The experimental results of parameter optimization showed that increasing the size of the first layer convolution kernel could effectively enhance the model's noise resistance. The best performance was achieved when the size of the first layer convolution kernel was set to 128×128 and the step was set to 4. When SNR was 0 dB, the fault degree identification accuracy of the model can reach 99.75%.

(3) In order to achieve more efficient feature fusion for the three types of time-frequency feature images, the attention mechanism of deep learning was introduced to optimize, and a network named CBAM-MFFCNN with stronger feature extraction ability and generalization performance was constructed. Variable load experiment results showed that the average recognition accuracy of CBAM-MFFCNN is 99.67%, and when the training set included all loads, the

recognition accuracy of the model on testing set under each load can reach 100%. It showed that this model can better adapt to the subtle changes of testing set data and had stronger generalization ability than the existing similar models.

In addition, although intelligent fault diagnosis of mechanical equipment is a hot research in recent years, through in-depth exploration of it, it is found that there are still many problems, which can be used as the direction of future research.

(1) Data structure. Existing bearing datasets are overly discretized in terms of operating conditions, which deviates from real-world conditions and may lead to model overfitting. Future research can focus on optimizing the data structure and constructing more comprehensive bearing datasets that better represent realistic operating conditions.

(2) Model structure. This includes the connection between data preprocessing methods and networks, as well as the overall optimization of network architecture. Although this study has explored these aspects, there is still significant room for further research. For example, it is worth investigating whether there are more reasonable signal processing methods and the optimal number of channels. Therefore, future research can be conducted in these areas.

(3) Robustness. Training a robust model is crucial. A robust model should be able to handle imperfect situations such as missing data, noise, and outliers without a significant drop in performance. Therefore, robustness should be an important consideration for future research on model stability.

REFERENCES

[1] I. El-Thalji and E. Jantunen, "A summary of fault modelling and predictive health monitoring of rolling element bearings," *Mech. Syst. Signal Process.*, vols. 60–61, pp. 252–272, Aug. 2015.

[2] A. Rai and S. H. Upadhyay, "A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings," *Tribol. Int.*, vol. 96, pp. 289–306, Apr. 2016.

[3] M. Xia, T. Li, L. Xu, L. Liu, and C. W. de Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 101–110, Feb. 2018.

[4] Y. Bai, W. Cheng, W. Wen, and Y. Liu, "Application of time-frequency analysis in rotating machinery fault diagnosis," *Shock Vibrat.*, vol. 2023, pp. 1–16, Jul. 2023.

[5] M. M. F. Darwish, M. H. A. Hassan, N. M. K. Abdel-Gawad, and D. A. Mansour, "Application of infrared spectroscopy for discrimination between electrical and thermal faults in transformer oil," in *Proc. 9th Int. Conf. Condition Monitor. Diagnosis (CMD)*, Nov. 2022, pp. 255–258.

[6] M. M. F. Darwish, M. H. A. Hassan, N. M. K. Abdel-Gawad, and D. A. Mansour, "A new method for estimating transformer health index based on ultraviolet-visible spectroscopy," in *Proc. 23rd Int. Middle East Power Syst. Conf. (MEPCON)*, Cairo, Egypt, Dec. 2022, pp. 1–5.

[7] P. Bangalore and L. B. Tjernberg, "An artificial neural network approach for early fault detection of gearbox bearings," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 980–987, Mar. 2015.

[8] J. Wang, C. Lu, M. Wang, P. Li, S. Yan, and X. Hu, "Robust face recognition via adaptive sparse representation," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2368–2378, Dec. 2014.

[9] C. Xu, Q. Liu, and M. Ye, "Age invariant face recognition and retrieval by coupled auto-encoder networks," *Neurocomputing*, vol. 222, pp. 62–71, Jan. 2017.

[10] J. Wang, K. Wang, Y. Wang, Z. Huang, and R. Xue, "Deep Boltzmann machine based condition prediction for smart manufacturing," *J. Ambient Intell. Humanized Comput.*, vol. 10, no. 3, pp. 851–861, Mar. 2019.

[11] A.-R. Mohamed, G. E. Dahl, and G. Hinton, "Acoustic modeling using deep belief networks," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 1, pp. 14–22, Jan. 2012.

[12] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022.

[13] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Comput. Sci. Rev.*, vol. 3, no. 3, pp. 127–149, Aug. 2009.

[14] T. Saghi, D. Bustan, and S. S. Aphale, "Bearing fault diagnosis based on multi-scale CNN and bidirectional GRU," *Vibration*, vol. 6, no. 1, pp. 11–28, Dec. 2022.

[15] Y. Jin, C. Qin, Z. Zhang, J. Tao, and C. Liu, "A multi-scale convolutional neural network for bearing compound fault diagnosis under various noise conditions," *Sci. China Technolog. Sci.*, vol. 65, no. 11, pp. 2551–2563, Sep. 2022.

[16] W. Liu, Y. Li, and H. Shi, "Fault diagnosis method for rolling bearings based on WDCNN-DLSTM," *Sci. Technol. Eng.*, vol. 23, no. 13, pp. 5522–5529, May 2023.

[17] R. Medina, J. C. Macancela, and P. Lucero, "Gear and bearing fault classification under different load and speed by using Poincaré plot features and SVM," *J. Intell. Manuf.*, vol. 33, pp. 1031–1055, Apr. 2022.

[18] M. Kuncan, K. Kaplan, M. R. Minaz, Y. Kaya, and H. M. Ertunç, "A novel feature extraction method for bearing fault classification with one dimensional ternary patterns," *ISA Trans.*, vol. 100, pp. 346–357, May 2020.

[19] J. Zhao, S. Yang, Q. Li, Y. Liu, X. Gu, and W. Liu, "A new bearing fault diagnosis method based on signal-to-image mapping and convolutional neural network," *Measurement*, vol. 176, May 2021, Art. no. 109088.

[20] Y. Yu and M. Yao, "When convolutional neural networks meet laser-induced breakdown spectroscopy: End-to-end quantitative analysis modeling of ChemCam spectral data for major elements based on ensemble convolutional neural networks," *Remote Sens.*, vol. 15, no. 13, p. 3422, Jul. 2023.

[21] C. Lei, L. Xue, M. Jiao, H. Zhang, and J. Shi, "Rolling bearing fault diagnosis by Markov transition field and multi-dimension convolutional neural network," *Meas. Sci. Technol.*, vol. 33, no. 11, Nov. 2022, Art. no. 114009.

[22] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proc.*, vol. 3, no. 1, pp. 91–99, Jun. 2022.

[23] *Download a Data File | Case School of Engineering | Case Western Reserve University*. Accessed: Dec. 11, 2023. [Online]. Available: https://engineering.case.edu/bearingdatacenter/download-data-file

[24] N. Zhang, L. Wu, J. Yang, and Y. Guan, "Naive Bayes bearing fault diagnosis based on enhanced independence of data," *Sensors*, vol. 18, no. 2, p. 463, Feb. 2018.

[25] T. Luo, Z. Ma, Z.-Q. John Xu, and Y. Zhang, "Theory of the frequency principle for general deep neural networks," 2019, *arXiv:1906.09235*.

[26] Y. Yang, Y. Zhuang, and Y. Pan, "Multiple knowledge representation for big data artificial intelligence: Framework, applications, and case studies," *Frontiers Inf. Technol. Electron. Eng.*, vol. 22, no. 12, pp. 1551–1558, Jan. 2022.

[27] C. Zhu, Y. Hu, and J. Chen, "Research on fault diagnosis technology of rolling bearing (Part 1)," *Railway Vehicles*, vol. 4, pp. 4–8, Apr. 1992.

[28] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, Sep. 2021.

[29] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. ECCV*, Munich, Germany, 2018, pp. 3–19.

[30] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, Feb. 2017.

[31] M. Abadi, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*.

HONGFENG GAO received the B.S. degree in automation from Wuhan University of Science and Technology, in 2007, and the M.S. degree in mechanical engineering from Hebei University of Engineering, in 2013. He is currently with Handan Branch of Hebei Special Equipment Supervision and Inspection Institute. He is a Senior Engineer and specializes in research and development of specialized equipment testing technology.

JIE MA received the B.S. degree from Taizhou Institute of Science and Technology, Nanjing University of Science and Technology. He is currently pursuing the M.S. degree with the School of Mechanical and Equipment Engineering, Hebei University of Engineering, China. His current research interests include bearing fault diagnosis based on signal processing and deep learning.

ZHONGHANG ZHANG received the B.S. degree from Ningbo University of Engineering, and the M.S. degree in mechanical engineering from Hebei University of Engineering. He is currently with MCC Baosteel Technology Service Company Ltd., engaged in the design of metallurgical special vehicles. His research interests include bearing fault diagnosis based on signal processing and deep learning.

CHAOZHI CAI received the B.S. and M.S. degrees in measurement and control technology and instrument from Hebei University of Engineering, Handan, Hebei, China, in 2007 and 2010, respectively, and the Ph.D. degree in mechatronics engineering from Beihang University, Beijing, China, in 2014.

He has been a Professor of mechanical engineering with Hebei University of Engineering, since 2023. He has published more than 30 refereed journal and conference papers in journals sponsored by IEEE, SAGE, ASME, and MDPI. His research interests include process system modeling and control, hydraulic and mechatronic servo control, system identification, robot control, and artificial intelligence based mechanical equipment fault diagnosis and structural health monitoring.

• • •