

Research Article

An Integrated Bearing Fault Diagnosis Method Based on Multibranch SKNet and Enhanced Inception-ResNet-v2

Baoquan Hu ^{1,2}, Jun Liu ¹, Yue Xu,³ and Tianlong Huo¹

¹School of Mechanical and Electrical Engineering, Lanzhou University of Technology, Lanzhou 730050, China

²School of Engineering, Xi'an International University, Xi'an 710077, China

³School of Automation, Northwestern Polytechnical University, Xi'an 710129, China

Correspondence should be addressed to Jun Liu; liuj@lut.edu.cn

Received 9 September 2023; Revised 12 January 2024; Accepted 29 January 2024; Published 21 February 2024

Academic Editor: Yi Qin

Copyright © 2024 Baoquan Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Deep learning has recently received extensive attention in the field of rolling-bearing fault diagnosis owing to its powerful feature expression capability. With the help of deep learning, we can fully extract the deep features hidden in the data, significantly improving the accuracy and efficiency of fault diagnosis. Despite this progress, deep learning still faces two outstanding problems. (1) Each layer uses the same convolution kernel to extract features, making it difficult to adaptively select convolution kernels based on the features of the input image, which limits the network's adaptability to different input features and leads to weak feature extraction. (2) Large number of parameters and long training time. To solve the above problems, this paper proposes an integrated deep neural network that combines an improved selective kernel network (SKNet) with an enhanced Inception-ResNet-v2, named SIR-CNN. First, based on the SKNet, a new three-branch SKNet was designed. Second, the new SKNet is embedded into a depthwise separable convolution network such that the model can adaptively select convolution kernels of different sizes during training. Furthermore, the convolution structure in the Inception-ResNet-v2 network was replaced by the improved depthwise separable convolution network to achieve effective feature extraction. Finally, the time-frequency maps of the raw vibration signals are obtained through short-time Fourier transform (STFT) and then sent to the proposed SIR-CNN network for experiments. The experimental results show that the proposed SIR-CNN achieves superior performance compared to other methods.

1. Introduction

Rolling bearings are widely used in nuclear energy, wind power, aerospace, petrochemicals, electric power, and other industrial fields [1]. With a series of advantages, such as high precision, good substitutability, and low price, rolling bearings are the core components of mechanical equipment, especially rotating mechanical equipment [2]. However, as the “joint” between the rotating and fixed parts, the operating conditions of rolling bearings inevitably change, owing to the long-term effects of high temperature, high speed, and a variety of alternating loads [3]. In addition, owing to processing errors, poor lubrication, thermal fatigue, work wear, and other factors, it is easy to develop faults in the rolling element, inner race, outer race, and cage inside the

bearing [4]. According to statistics, 45%~55% of mechanical failures are caused by bearing failure [5]. When a rolling bearing fails, it will affect the normal operation of other parts of the equipment, causing a series of chain damage reactions, even leading to serious consequences of machine damage and human death. Therefore, the fault diagnosis of rolling bearings is of great significance [6, 7].

Traditional bearing fault diagnosis methods include noise analysis, acoustic diagnosis, temperature measurements, oil film resistance diagnosis, and vibration signal analysis [8]. Among them, diagnostic technology based on vibration signal analysis is applicable to rolling bearings under various working conditions, which has the advantages of obvious fault characteristics and high diagnostic accuracy [9]. Therefore, it has been widely used for bearing fault

diagnoses. More than 80% of the existing literature on bearing fault diagnosis employs vibration signal-analysis methods. This method usually uses manual approaches such as the fast Fourier transform (FFT) [10], wavelet transform (WT) [11], and empirical mode decomposition (EMD) [12] to extract signal features and then uses a support vector machine (SVM) [13], K-nearest neighbor (KNN) [14], and BP neural network (BPNN) [15] to obtain diagnostic results. However, these feature extraction methods rely on expert experience and knowledge, which can easily introduce artificial errors and have poor generalization ability. In addition, it is difficult to establish complex data-mapping mechanisms for these shallow models; therefore, they have certain limitations.

In recent years, deep learning has been widely used for fault diagnosis [16, 17]. Compared to traditional bearing fault diagnosis methods, deep learning does not require an artificial feature extraction process. Instead, it automatically extracts the representative information and sensitive features from the original data. It has a strong learning representation ability and excellent recognition effect. Therefore, it has several advantages in the field of fault diagnosis. For example, deep discriminative transfer learning network (DDTLN) [18], maximum mean square discrepancy (MMSD) [19], relationship transfer domain generalization network (RTDGN) [20], convolutional neural network (CNN) [21], recurrent neural network (RNN) [22], generative adversarial networks (GAN) [23], autoencoder (AE) [24], ResNet [25], etc. Among them, CNN, as a typical deep learning model, has a strong feature extraction ability due to its unique local connection, weight sharing, pooling operation, and other characteristics; therefore, it has received extensive attention from researchers [9, 16, 26, 27].

Generally, to improve the performance of a CNN, it is necessary to stack convolutional layers continuously to obtain deeper convolutional networks. However, as the number of network layers increases, the number of parameters increases dramatically, and the computing resources required are considerable. What is more terrible is that when the number of layers is too large, the model is prone to the problems of gradient disappearance and overfitting in the process of back propagation, making it difficult for the model to converge, leading to a decline in the identification accuracy [28].

To solve the above problems, Google proposed Inception-ResNet [29], which not only solved the problems of gradient vanishing and loss value increasing but also deepened the network and achieved higher recognition accuracy, receiving widespread attention from experts. For example, Liu et al. [30] achieved transfer learning based on the Inception-ResNet-v2 model by converting raw data into RGB images. Li et al. [31] proposed a bearing fault diagnosis method that combined fault signal spectrum images with the Inception-ResNet-v2 model, achieving good classification accuracy. To address the issue of important feature loss, Jigyasu et al. [32] developed a two-dimensional (2D) image dataset using time-frequency methods and utilized the Inception-ResNet-v2 model for effective feature extraction. Deveci et al. [33] compared commonly used time-frequency

images in bearing fault detection and analyzed which time-frequency methods can more clearly display fault features. Peng et al. [34] replaced the convolution module in Inception-ResNet-v2 with depthwise separable convolution to extract fault features under different receptive fields. Zheng [35] improved the feature extraction layer in the Inception-ResNet-v2 structure, improving the detection accuracy of the network for small-scale targets. Liu et al. [36] proposed a transfer learning method based on Inception-ResNet-v2. By studying different methods for converting one-dimensional (1D) signals into 2D graphics, the best method for structural health monitoring was found. Song et al. [37] reduced the computational complexity of the model by segmenting the input image and replacing the convolution module in Inception-ResNet with depthwise separable convolution. Das et al. [38] introduced a multi-model-integrated network based on Inception-ResNet-v2, which achieved high accuracy. Kasireddy et al. [39] developed a binary classification model based on Inception-ResNet-v2 and a small Inception-ResNet-v2 model. Meel and Kumar Vishwakarma [40] proposed a multimodal fusion model based on Inception-ResNet-v2, which achieved high recognition accuracy through multiple fusions in the early and late stages.

However, the abovementioned Inception-ResNet network still has the following problems: (1) They are unable to adaptively select convolution kernels according to image features. (2) The Inception-ResNet model has a large number of parameters and a long training time. To solve these problems, this study proposes an integrated deep neural network that combines an improved selective kernel network (SKNet) with an enhanced Inception-ResNet-v2. The main contributions of this study are as follows.

- (1) Considering that Inception-ResNet-v2 usually has three different sizes of convolutional kernels, such as 3×3 , 5×5 , and 7×7 , a new three-branch SKNet (NewSKNet) is designed, so that the network can adaptively select the important features extracted by different sizes of convolutional kernels during the training process;
- (2) Embed NewSKNet into a depthwise separable convolution network. Then, the improved depthwise separable convolution network is used to replace the convolutional structure in the Inception-ResNet-v2 network to reduce the parameters and thus shorten the training time of the network;
- (3) A new intelligent bearing fault diagnosis framework built on NewSKNet and enhanced Inception-ResNet-v2 is proposed.

The remainder of this paper is arranged as follows: Section 2 briefly introduces the main theoretical background, including the short-time Fourier transform (STFT), Inception-ResNet-v2 network, and SKNet; Section 3 describes the proposed method in detail; in Section 4, two datasets are used for experimental verification; Section 5 discusses the proposed method. Finally, the results and future work are summarized in Section 6.

2. Theoretical Background

2.1. Short-Time Fourier Transform (STFT). The STFT can obtain both time and frequency domain features of bearing vibration signals, and the transformed 2D matrix is more suitable for CNN processing [41]. The essence of STFT is the Fourier transform with a window, which is calculated by multiplying a window function $h(\tau)$, and assuming that the signal is smooth during the short interval of the analysis window. Subsequently, $h(\tau)$ is shifted on the time axis to obtain the spectrum of the entire time domain. The STFT was calculated as follows:

$$\text{STFT}(\tau, \omega) = \int_{-\infty}^{+\infty} x(\tau)h(\tau - \Delta\tau)e^{-j2\pi\omega\tau} d\tau, \quad (1)$$

where $x(\tau)$ is the input time domain signal and $h(\tau - \Delta\tau)$ is the analysis window function. It can be seen that STFT is the Fourier transform made by multiplying the input signal $x(\tau)$ by a window function $h(\tau - \Delta\tau)$. The variables τ and ω are the time and frequency resolutions, respectively, and the formula is as follows:

$$t = \left\lfloor \frac{N_{sl} - N_{ow}}{N_{ww} - N_{ow}} \right\rfloor, \quad (2)$$

$$\omega = \begin{cases} \frac{N_p}{2} + 1, & \text{if } N_p \text{ is even,} \\ \frac{(N_p + 1)}{2}, & \text{if } N_p \text{ is odd,} \end{cases} \quad (3)$$

where N_{sl} is the sample length, N_{ww} is the window width, N_{ow} is the window overlap width, N_p is the number of points participating in the Fourier transform, and $\lfloor \cdot \rfloor$ is the rounding down function.

As can be seen from equation (2), the sample length and window width affect the resolution in both the time and frequency domains, thus affecting the transformation effect of the STFT. Generally, the sample length is selected based on the first harmonic of the fault characteristic frequency and the window width of the STFT is selected based on the second harmonic. The fault characteristic frequency can be calculated using the following formula:

$$\begin{aligned} f_{IR} &= \frac{N}{2} \left(1 + \frac{d_1}{d_2} \cos \theta \right) f, \\ f_{OR} &= \frac{N}{2} \left(1 - \frac{d_1}{d_2} \cos \theta \right) f, \\ f_{BR} &= \frac{1}{2} \frac{d_2}{d_1} \left(1 - \left(\frac{d_1}{d_2} \right)^2 \cos^2 \theta \right) f, \end{aligned} \quad (4)$$

where f_{IR} and f_{OR} are the frequencies of the rolling element passing through the inner and outer races, respectively; f_{BR} is the rotation frequency of the rolling element; $f = v/60$ is the rotation frequency of the inner race; v is the rotational speed; d_1 and d_2 are the rolling element diameter and pitch

circle diameter, respectively; N is the number of balls; and θ is the contact angle.

2.2. Inception-ResNet-v2. In 2016, Google introduced ResNet to the Inception network, thus proposing the Inception-ResNet-v2. The inception module extracts multi-scale features from different receptive fields using convolution kernels of different sizes in parallel computing. The introduction of the ResNet structure can avoid overfitting and network degradation problems caused by deepening of the model layers.

This study is based on the Inception-ResNet-v2 model, which is mainly composed of Stem, Inception-ResNet, Reduction, Average pooling, Dropout, SoftMax, and other modules, as shown in Figure 1. As can be seen, the original pooling operation inside the inception is replaced with a residual connection, which constitutes a new Inception-ResNet module. The Inception-ResNet module comprises three types: A, B, and C. Their structures are similar; however, the size and number of convolution cores are different. The Stem module was used to preprocess the input data to obtain a deeper network structure. The Reduction module was used to change the size of the feature map to prevent bottlenecks. For the detailed structure of the Inception-ResNet modules, refer to [29].

2.3. Selective Kernel Network (SKNet). In recent years, studies on the mechanism of animal visual nerve action have found that when cats look at objects of different sizes and distances, the size of the receptive field of their visual layer neurons is not fixed but automatically adjusts with the size of the stimulus. Therefore, when constructing a CNN, the size of the convolution kernel should differ for different stimuli. However, existing CNN models generally employ only one type of convolutional kernel in the same layer and rarely consider the role of multiple convolutional kernels.

To solve this problem, in 2019, Li et al. proposed SKNet [42], as shown in Figure 2. The network consisted of three main steps: split, fuse, and selection. In the Split stage, the input image is convolved by two kinds of convolution kernels, 3×3 and 5×5 , respectively; in the Fuse stage, the features calculated in the Split stage are fused by the SoftMax function; finally, in the Select stage, the new feature map is obtained according to the results of the different convolution kernels.

3. Proposed Method

3.1. A New Selective Kernel Network (NewSKNet). The Inception model can adapt to images of different scales by adding multiple convolution kernels; however, the convolution kernels of each layer have the same weight. Correspondingly, the convolution kernel of SKNet differs in size and parameter weight. In addition, they can be easily embedded into other deep learning models. Therefore, based on SKNet, we designed a new three-branch lightweight embedded module called NewSKNet, as shown in Figure 3.

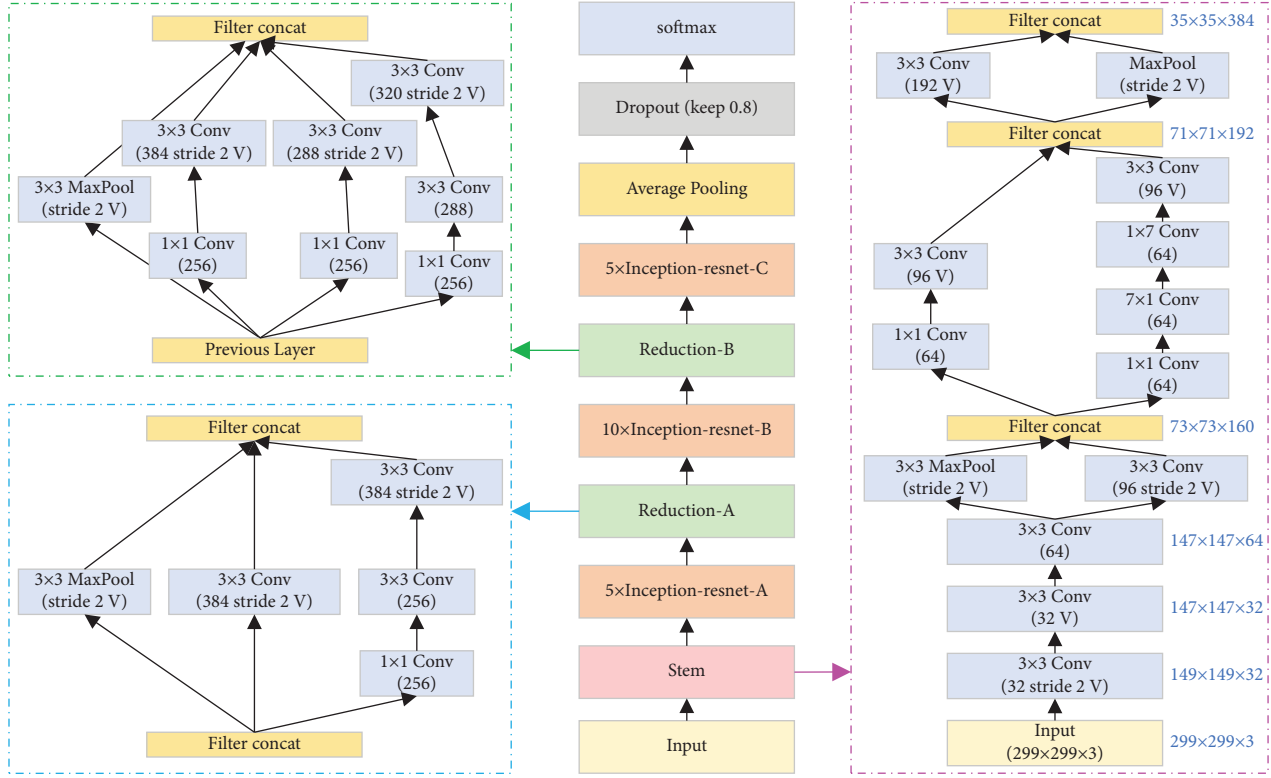


FIGURE 1: Inception-ResNet-v2 model.

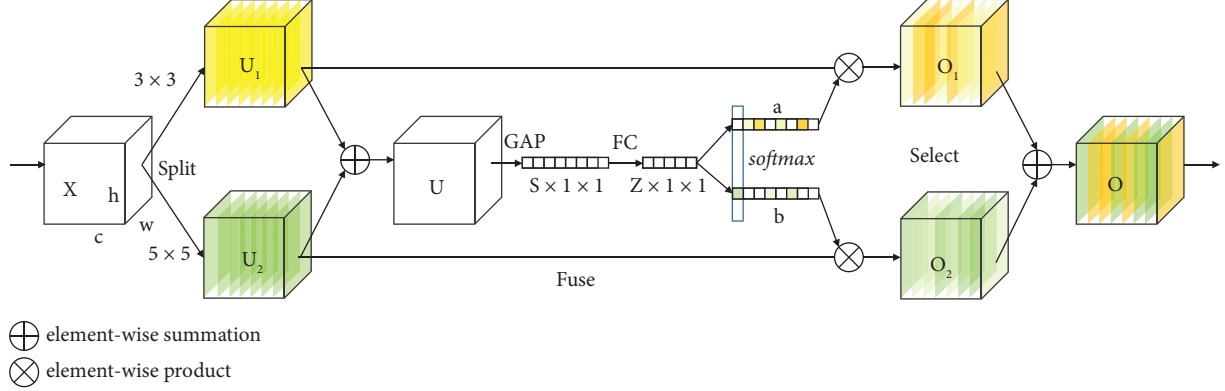


FIGURE 2: SKNet model structure.

For a given characteristic graph $X \in \mathbb{R}^{H \times W \times C}$, three different convolution kernels, 3×3 , 5×5 , and 7×7 , were used for the calculation, and three characteristic graphs, namely, $U_1 \in \mathbb{R}^{H \times W \times C}$, $U_2 \in \mathbb{R}^{H \times W \times C}$, and $U_3 \in \mathbb{R}^{H \times W \times C}$, are obtained. To enable the model to adjust the size of the local receptive field according to the size of the input feature map, the “gate” is used to control the information passing through the three convolution kernels. To achieve this goal, element-wise summation was used to fuse the results of the three convolution kernels.

$$U = U_1 + U_2 + U_3. \quad (5)$$

The statistical information $S \in \mathbb{R}^C$ of different channels is then obtained by global average pooling (GAP), where the statistical information of channel c is

$$\begin{aligned} S_c &= \mathcal{F}_{\text{GAP}}(U_c) \\ &= \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_c(i, j). \end{aligned} \quad (6)$$

Furthermore, a compact feature map $z \in \mathbb{R}^{d \times 1}$ is generated through a fully connected (FC) layer so that the parameters in the network can be reduced, thus improving computational efficiency.

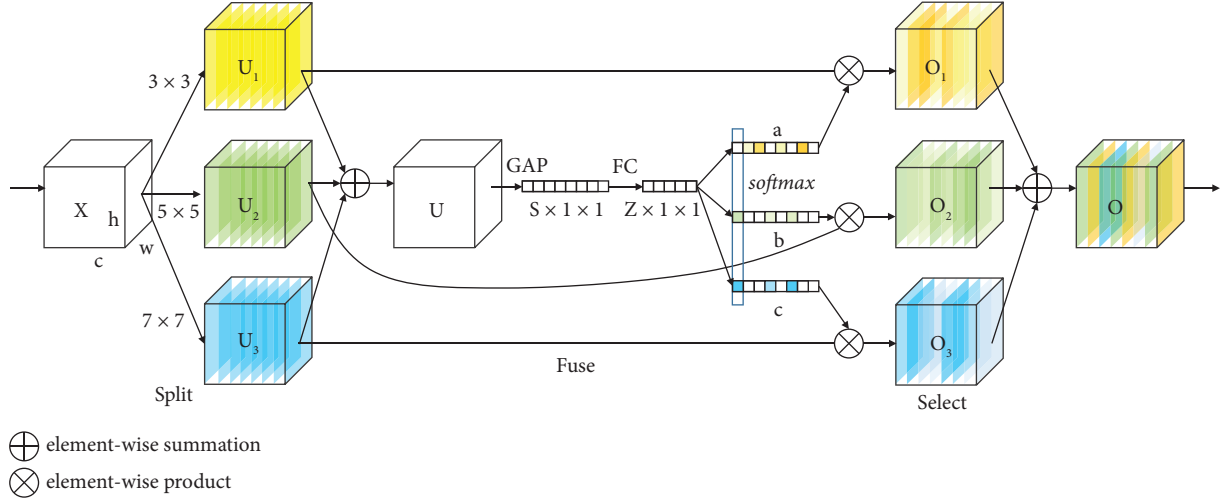


FIGURE 3: NewSKNet model structure.

$$Z = \mathcal{F}_{\text{FC}}(S) = \mathcal{R}(\mathcal{B}(W_S)), \quad (7)$$

where \mathcal{R} is the ReLU activation function and \mathcal{B} denotes the batch normalization. Subsequently, Z is divided into three branches using the SoftMax function.

$$\begin{aligned}
 a &= \frac{e^{AZ}}{e^{AZ} + e^{BZ} + e^{CZ}}, \\
 b &= \frac{e^{BZ}}{e^{AZ} + e^{BZ} + e^{CZ}}, \\
 c &= \frac{e^{CZ}}{e^{AZ} + e^{BZ} + e^{CZ}},
 \end{aligned} \quad (8)$$

where $A, B, C \in \mathbb{R}^{c \times d}$, a, b, c denote the convolution kernel weights of U_1, U_2 , and U_3 , respectively. We then obtain O_1, O_2 , and O_3 through an element-wise product of the convolution kernel weights.

$$\begin{aligned}
 O_1 &= aU_1, \\
 O_2 &= bU_2, \\
 O_3 &= cU_3.
 \end{aligned} \quad (9)$$

Finally, the output of the characteristic graph can be obtained by summation.

$$O = O_1 + O_2 + O_3. \quad (10)$$

3.2. Depth-Wise Separable Convolution Embedded in NewSKNet. Depthwise separable convolution is a lightweight network that includes depthwise convolution and pointwise convolution [43]. In depthwise convolution, each convolution kernel is responsible for one channel; that is, each channel is calculated using only one convolution kernel. Because this operation is an independent convolution operation for each channel, it does not effectively use the feature information of different channels at the same spatial location. Therefore, pointwise convolution is typically used after depthwise convolution. Specifically, it was used to

combine the obtained feature maps again to generate a new feature map.

In this study, a depthwise separable convolutional network embedded in NewSKNet is proposed, as shown in Figure 4. First, the features of the input image are extracted using 1×1 , 3×3 , and 5×5 convolution kernels. Then, three 1×1 convolution kernels are used to merge the extracted features. Finally, the NewSKNet proposed in Section 3.1 is embedded into the model to further extract significant features. Specifically, after the depthwise separable convolution network obtains the feature map and NewSKNet uses three different convolution kernels to perform operations, thereby obtaining three feature maps. Then, the weighted probability of each feature map was calculated using the SoftMax function. Finally, the weighted probability with the highest probability ranking was selected, and the weight was multiplied with the corresponding characteristic graph to obtain the final characteristic graph. Thus, after the NewSKNet operation, feature maps with more features can be selected for improved fault classification.

In a traditional CNN, a certain relationship exists between the number of parameters in the network and the number of feature maps, which can be calculated as follows:

$$G_{(w-x)/l+1, (h-y)/l+1, m} = \sum_{x,y} C_{x,y,z} * G_{w,h,z}, \quad (11)$$

where C is the convolution kernel, x, y , and z represent the size of the convolution kernel and the number of channels; G represents the input characteristic diagram with dimensions as w, h, z ; G represents the output characteristic diagram with dimensions are $(w-x)/l+1, (h-y)/l+1, m$; l represents the step length; and m is the number of output channels. Therefore, the number of parameters in the network is $R_1 = x * y * z * m$.

The parameter formula of depthwise convolution is as follows:

$$G_{(w-x)/l+1, (h-y)/l+1, z} = \sum_{x,y} C_{x,y,z} * G_{w,h,z}. \quad (12)$$

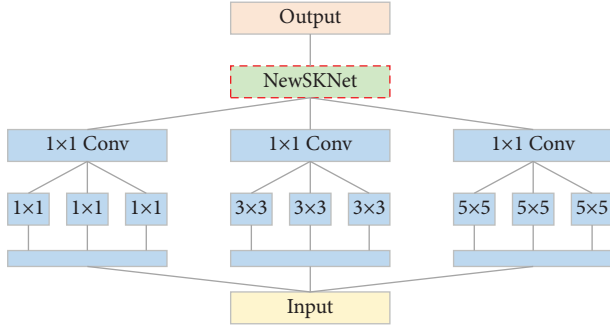


FIGURE 4: Embedding NewSKNet into the depthwise separable convolution model structure.

The parameter formula of pointwise convolution is as follows:

$$G_{w,h,m} = \sum_{z,m} C_{1*1*m} * G_{w,h,z}, \quad (13)$$

where C is the convolution kernel; its size is $1 * 1 * m$; m is the number of output channels. The number of depthwise convolution parameters is $z * x * y$ and the number of pointwise convolution parameters is $m * z$, so the number of depthwise separable convolution parameters is $R_2 = z * x * y + m * z$.

Therefore, the ratio of the number of parameters between depthwise separable convolution and traditional convolution is as follows:

$$\begin{aligned} \frac{R_2}{R_1} &= \frac{z * x * y + m * z}{x * y * z * m} \\ &= \frac{1}{m} + \frac{1}{x * y}. \end{aligned} \quad (14)$$

From equation (14), it can be seen that the parameter calculation of depthwise separable convolution is related to the number of output channels and the size of the convolution kernel used. For example, when the convolutional kernel size is $3 * 3$ and the number of output channels is 64, depthwise separable convolution reduces the computational complexity by about 8-9 times compared to traditional convolution.

3.3. The Proposed Fault Diagnosis Framework Based on SIR-CNN. In fact, when the cat's visual nerve is stimulated, the size of its receptive field is not fixed but automatically adjusts with the change in stimulation. In contrast, in the CNN model constructed by imitating the visual features of cats, the weight of each layer of the convolution kernel is the same, and the convolution kernel cannot be adaptively selected according to the features of the input image. Therefore, this study designed a novel bearing fault diagnosis method based on the proposed SIR-CNN. The core of the diagnosis method is to design a new three-branch SKNet and embed it into a depthwise separable convolution network. Finally, the convolution structure in Inception-ResNet-v2 is replaced by the improved depthwise separable network.

The model can adaptively select a feature map with more features in the training process and extract the multilayer sensitive features in the input image. The established fault diagnosis method and its application process are shown in Figure 5 and Table 1. The implementation steps are summarized as Figure 6 and Table 2.

4. Experimental Validation

To validate the effectiveness of the proposed fault diagnosis method based on SIR-CNN, this section uses a comprehensive fault simulation test bench and the XJTU-SY bearing dataset to conduct experiments. All experiments were implemented on a PC with a Win 10 operating system, Core i7 CPU, 2.9 GHz, 16 GB RAM, and RTX2060 GPU. The software used in the experiments was Python 3.7, TensorFlow 2.3.

4.1. Case 1: Laboratory-Measured Dataset

4.1.1. Experiment Description and Data Acquisition. The HZXT-DS-001 comprehensive fault simulation test bench can be used to simulate various fault types of rolling bearings. As shown in Figure 7, the test bench was mainly composed of a three-phase motor, acceleration sensor, eddy current sensor, shaft, and coupling. In this experiment, EDM technology was used to machine faults in different parts of a rolling bearing (NSK6308). Two radial (X , Y) and one axial (Z) vibration signals of the bearing were measured by three accelerometers. The sampling rate is 8192 Hz and the sampling time is 10 s. In this way, five common health conditions of rolling bearings were simulated: normal (NL), rolling element failure (RF), inner race failure (IF), outer race failure (OF), and cage failure (CF). Figure 8 shows a real picture of the four types of failed bearings.

During the experiment, the motor drives the rotor to run at 2600 r/min, 2800 r/min, 3000 r/min, and 3200 r/min and then uses the data acquisition system (as shown in Figure 9) to collect the vibration signals of the bearing in different states, which are represented by A, B, C, and D, respectively, as shown in Table 3. At each speed, it includes a rolling element fault signal, cage fault signal, inner-race fault signal, normal signal, and outer-race fault signal. Therefore, five different data samples were formed for each dataset, and their time-domain and frequency-domain diagrams are shown in Figure 10.

To fully utilize the powerful ability of the CNN model in image processing, the 1D bearing vibration signal is converted into a 2D time-frequency diagram through STFT, as shown in Figure 11. Then, the obtained time-frequency graph was divided into training, validation, and testing sets. The training and validation sets were used to train the proposed SIR-CNN and the testing set was used to verify the trained model.

4.1.2. Diagnostic Results and Analysis. To explain the classification effect of the rolling bearing fault diagnosis method based on SIR-CNN in more detail, the four datasets in

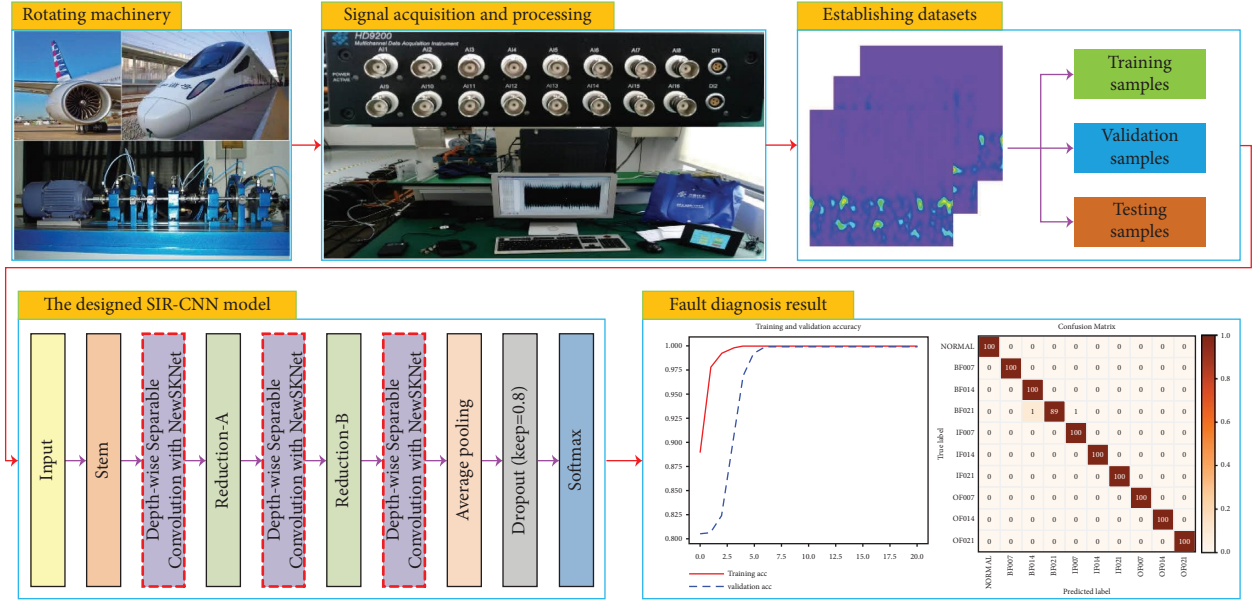


FIGURE 5: Flowchart of the fault diagnosis method based on SIR-CNN.

TABLE 1: Description of SIR-CNN model parameters.

Modules	Blocks	Operations and parameters
Stem	—	As in Figure 1
Depthwise separable convolution with NewSKNet	Branch_0	$C = 1 \times 1 - 3, S = 1 \times 1, P = \text{"same"}; C = 1 \times 1, S = 1 \times 1, P = \text{"valid"}$
	Branch_1	$C = 3 \times 3 - 3, S = 1 \times 1, P = \text{"same"}; C = 1 \times 1, S = 1 \times 1, P = \text{"valid"}$
Reduction A	Branch_2	$C = 5 \times 5 - 3, S = 1 \times 1, P = \text{"same"}; C = 1 \times 1, S = 1 \times 1, P = \text{"valid"}$
	NewSKNet	As in Figure 3
Reduction B	Branch_0	$C = 3 \times 3 - 384, S = 2 \times 2, P = \text{"valid"}$
	Branch_1	$C = 1 \times 1 - 256; 3 \times 3 - 256; 3 \times 3 - 384, S = 2 \times 2, P = \text{"valid"}$
Depthwise separable convolution with NewSKNet	Branch_2	$\text{Max } P = 3 \times 3, S = 2 \times 2, P = \text{"valid"}$
	NewSKNet	As in Figure 3
Reduction B	Branch_0	$C = 1 \times 1 - 256; 3 \times 3 - 384, S = 2 \times 2, P = \text{"valid"}$
	Branch_1	$C = 1 \times 1 - 256; 3 \times 3 - 288, S = 2 \times 2, P = \text{"valid"}$
Depthwise separable convolution with NewSKNet	Branch_2	$C = 1 \times 1 - 256; 3 \times 3 - 288; 3 \times 3 - 320, S = 2 \times 2, P = \text{"valid"}$
	Branch_3	$\text{Max } P = 3 \times 3, S = 2 \times 2, P = \text{"valid"}$
Average pooling	Branch_0	$C = 1 \times 1 - 3, S = 1 \times 1, P = \text{"same"}; C = 1 \times 1, S = 1 \times 1, P = \text{"valid"}$
	Branch_1	$C = 3 \times 3 - 3, S = 1 \times 1, P = \text{"same"}; C = 1 \times 1, S = 1 \times 1, P = \text{"valid"}$
Dropout	Branch_2	$C = 5 \times 5 - 3, S = 1 \times 1, P = \text{"same"}; C = 1 \times 1, S = 1 \times 1, P = \text{"valid"}$
	NewSKNet	As in Figure 3
SoftMax	—	512×1
		Ratios 0.8
		$n_classes \times 1$

Table 3 were used for the experiments, and the confusion matrix was used to display the experimental results. The abscissa of the confusion matrix represents the predicted category label and the ordinate represents the real category label. The number on the diagonal line indicates the number of samples correctly classified for each sample type. The larger the number, the better is the classification effect of the model for this type of sample. The number outside the

diagonal indicates the number of samples of one type incorrectly identified as another. Figure 12 shows the confusion matrix of the proposed method for the four datasets. The diagnostic accuracies were 99.6%, 99.8%, 100%, and 99.2%, with the average diagnostic accuracy of the four datasets at 99.65%. This shows that the proposed method has a strong feature learning ability on the four datasets and achieves a high classification accuracy.

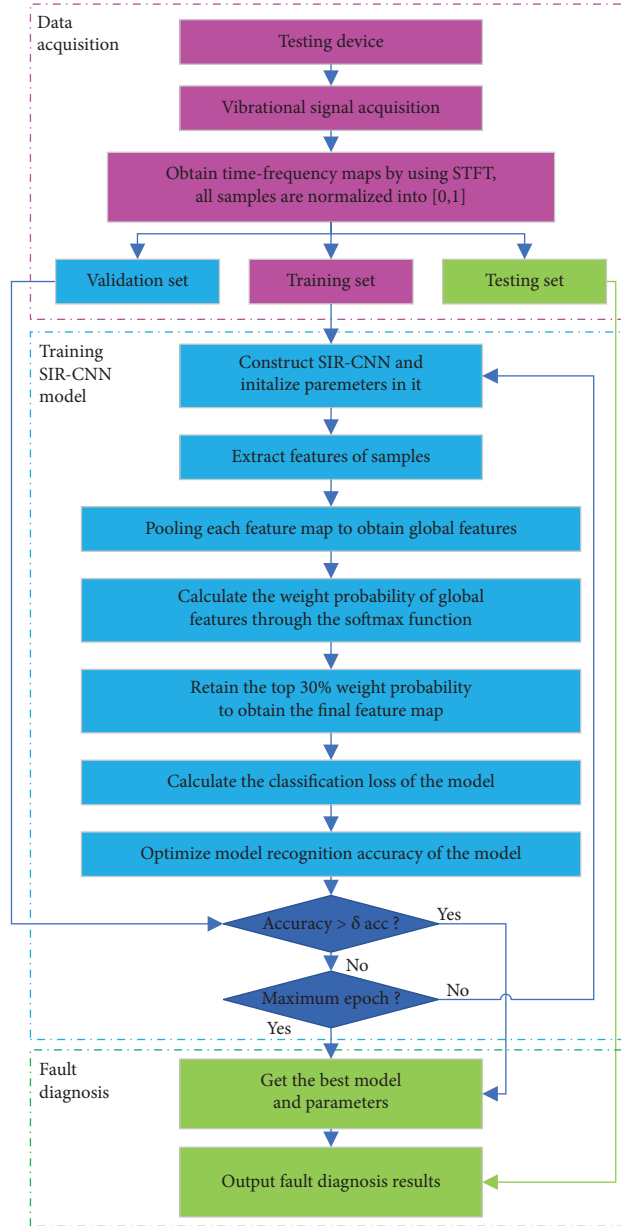


FIGURE 6: Fault diagnosis flowchart of SIR-CNN.

4.1.3. *Performance Comparison of Different Methods.* To further validate the advantages of the proposed method compared with other methods, this section selects shallow models, a conventional CNN model, and deep models based on Inception were used for comparative experiments. (1) Compared with shallow models. First, 15 features of the vibration signal are extracted [44] and then input into the SVM and BPNN for fault identification. Among them, the kernel function of SVM adopts Gaussian radial basis function (RBF), $C = 10$, $\gamma = 0.015$. The hidden layer structure of BPNN is (32, 16), and the activation function is ReLU. (2) Comparison with the conventional CNN. The specific structural parameters were as follows: $\text{Conv}_{16}^{3 \times 3} \rightarrow \text{Pooling}_{16}^{2 \times 2} \rightarrow \text{Conv}_{32}^{3 \times 3} \rightarrow \text{Pooling}_{32}^{2 \times 2} \rightarrow \text{Conv}_{64}^{3 \times 3} \rightarrow \text{Pooling}_{64}^{2 \times 2}$

$\text{FC}_1^{100} \rightarrow \text{FC}_1^5$, where $\text{Conv}_n^{m \times m}$ represents the convolutional layer, with a kernel size of $m \times m$ and a quantity of n ; $\text{Pooling}_k^{c \times c}$ represents the pooling layer, with a pooling size of $c \times c$ and a quantity of k ; FC represents the fully connected layer (3) Compared to the deep model based on inception. In this experiment, Inception-v4, Inception-ResNet-v2, Inception-ResNet-v2+SKNet, and Inception-ResNet-v2+NewSKNet were used. The sample data in Table 3 are input into the model for iterative training. To reduce the impact of randomness, each experiment was conducted ten times.

As shown in Figure 13, the diagnosis accuracy of the shallow networks (BPNN and SVM) is low. This is because the BPNN and SVM need to manually extract features before training. However, the method of manually extracting

TABLE 2: Pseudocode of our proposed method.

 Algorithm: SIR-CNN

Requirement: Samples that have been preprocessed by STFT; learning rate: α ; batch size: n ; total training epoch: N ; optimizer: Adam

While epoch \leq max epoch:
While $i \leq$ the total batch numbers:

- (1) Extract features from the input image using three different convolution kernels: 3×3 , 5×5 , and 7×7 ;
- (2) Three feature maps are obtained: $U_1 \in \mathbb{R}^{H \times W \times C}$, $U_2 \in \mathbb{R}^{H \times W \times C}$, and $U_3 \in \mathbb{R}^{H \times W \times C}$;
- (3) Using equation (7) to fuse the results of three convolutional kernels;
- (4) Calculate the statistical information for each channel using equation (8);
- (5) Generate a compact feature map using equation (9);
- (6) Calculate the weight probability of global features using equation (10);
- (7) Retain the top 30% weight probability to obtain the final feature map;
- (8) obtain the entire loss L ;
- (9) Update parameters in the network;
- (10) $i += 1$

end

epoch += 1

end

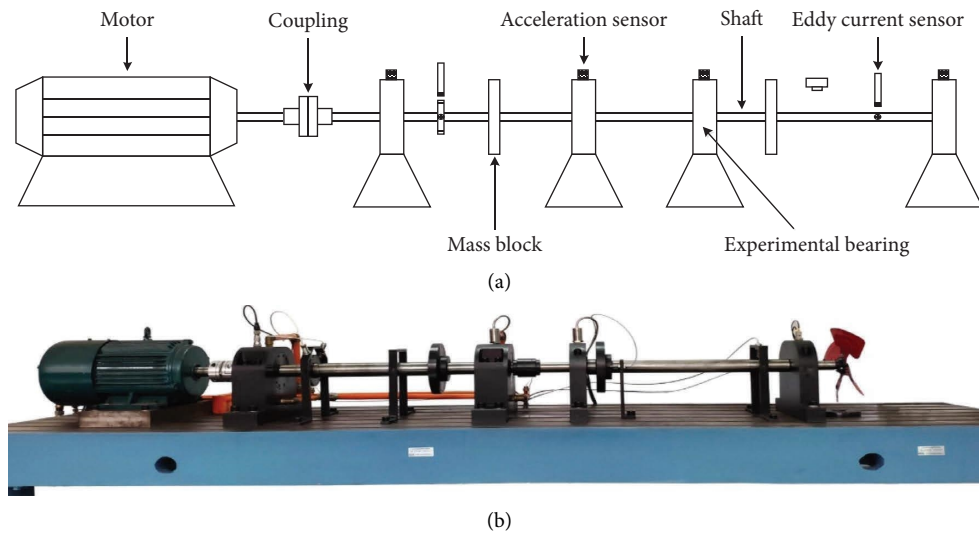


FIGURE 7: HZXT-DS-001 comprehensive fault simulation test bed. (a) Structural diagram of HZXT-DS-001. (b) Real picture of HZXT-DS-001.

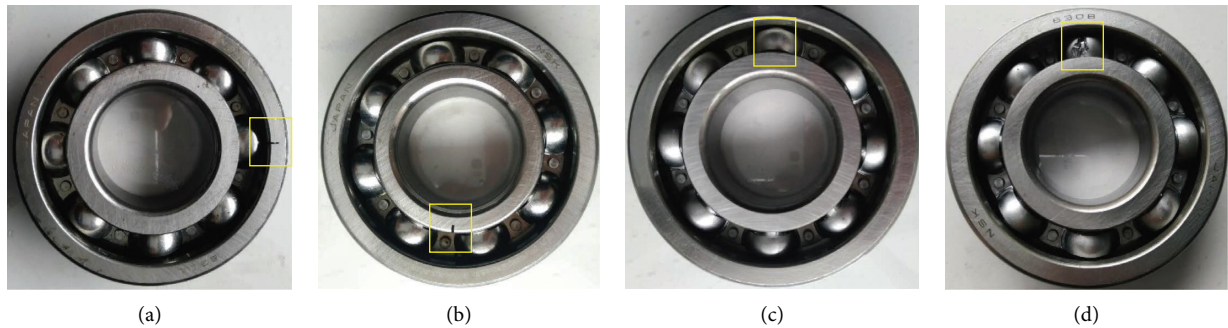


FIGURE 8: Faulty bearing in four states. (a) Outer race failure. (b) Inner race failure. (c) Rolling element failure. (d) Cage failure.

features will miss some important information, which will affect the recognition accuracy. Correspondingly, several other methods have obtained high diagnostic accuracy,

among which the proposed SIR-CNN has the best classification performance. This is owing to the improved SKNet embedded in the SIR-CNN model, which enables the model



FIGURE 9: Operating platform and data acquisition system. (a) PC platform. (b) Data acquisition system.

TABLE 3: Datasets constructed under different speed conditions.

Fault location		Ball	Cage	Inner race	None	Outer race
Class label		0	1	2	3	4
Dataset A (2600 r/m)	Training set	700	700	700	700	700
	Validation set	200	200	200	200	200
	Testing set	100	100	100	100	100
Dataset B (2800 r/m)	Training set	700	700	700	700	700
	Validation set	200	200	200	200	200
	Testing set	100	100	100	100	100
Dataset C (3000 r/m)	Training set	700	700	700	700	700
	Validation set	200	200	200	200	200
	Testing set	100	100	100	100	100
Dataset D (3200 r/m)	Training set	700	700	700	700	700
	Validation set	200	200	200	200	200
	Testing set	100	100	100	100	100

to automatically extract more important features during training.

To verify the antinoise performance of the SIR-CNN, Gaussian white noise with different signal-to-noise ratios (SNR) was added to the bearing vibration signal to simulate the actual working state of the bearing. The SNR is defined as follows:

$$\text{SNR} = 10 \log_{10} \left(\frac{P_{\text{rawsignal}}}{P_{\text{noisesignal}}} \right), \quad (15)$$

where $P_{\text{rawsignal}}$ and $P_{\text{noisesignal}}$ represent the powers of the original vibration and noise signals, respectively. The smaller the SNR value, the stronger the noise interference. 70% of the dataset with added noise was divided into training samples, 20% into validation samples, and the remaining 10% as testing samples.

Figure 14 shows the average diagnostic accuracy of the various methods under different SNR interferences. With a continuous increase in noise, the fault diagnosis accuracy of various methods also decreases correspondingly. The influence of random noise on the SIR-CNN model is relatively small, and its antinoise performance tends to be stable. The shallow models (BPNN and SVM) are more susceptible to noise interference. When $\text{SNR} = -6$ (dB), the diagnostic accuracy was the lowest. Meanwhile, other methods are also affected by random noise. In conclusion, compared with

other diagnostic methods, the SIR-CNN has the best antinoise performance.

4.1.4. Comparison with Other Preprocessing Methods. In order to obtain the optimal STFT, four different types of window functions and window widths were used for comparative analysis, and the experimental results are shown in Figure 15. It can be seen that when the window function is Hamming window and the window width is 64, the recognition accuracy is the highest, reaching 99.91%.

Two data preprocessing methods, Wigner–Ville distribution (WVD) and continuous wavelet transform (CWT), are used to compare with STFT to verify the superiority of STFT-based SIR-CNN. As shown in Figures 11 and 16, the time-frequency map obtained by STFT contains the least amount of information compared with WVD and CWT, and some important fault features may be lost. However, it reduces invalid noise and interference to some extent, and the differences between fault types are more obvious. Correspondingly, the time-frequency map obtained by CWT retains more time-frequency information and therefore is more susceptible to noise [42]. The time-frequency map generated by WVD is of low quality, with low similarity to the original samples, and it is difficult to distinguish between fault types, which increases the difficulty of the model in feature extraction.

Figure 17 shows the test results of three preprocessing methods under different SNR. It can be seen that both CWT-

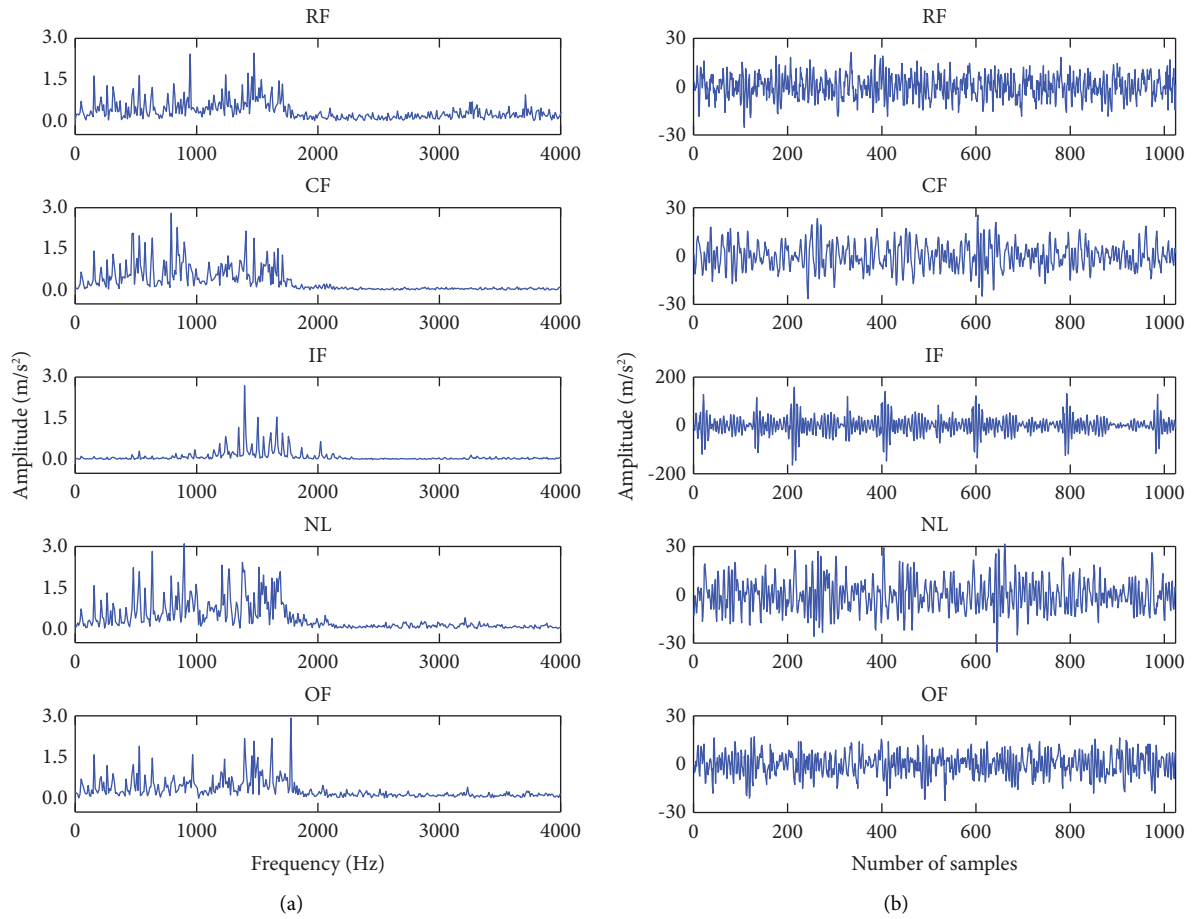


FIGURE 10: Frequency domain and time domain waveforms for five different samples (dataset A). (a) Frequency domain diagram. (b) Time domain diagram.

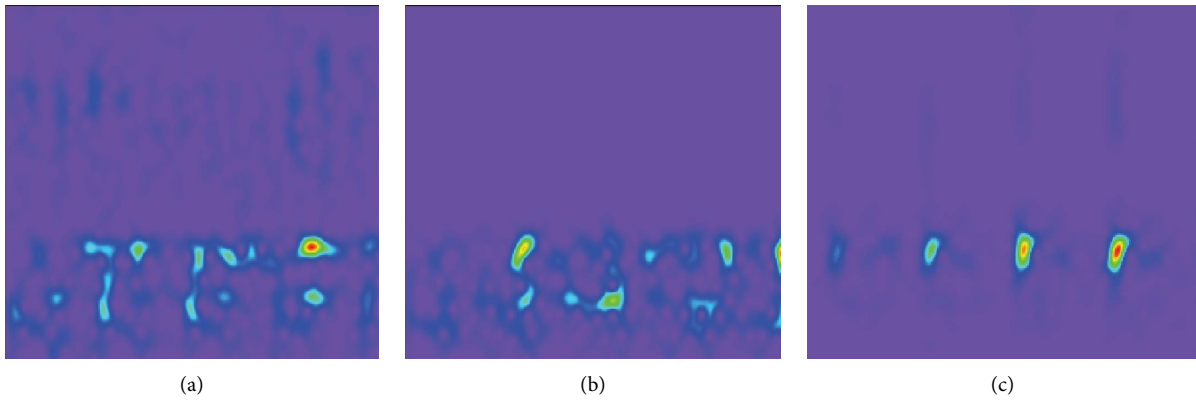


FIGURE 11: Continued.

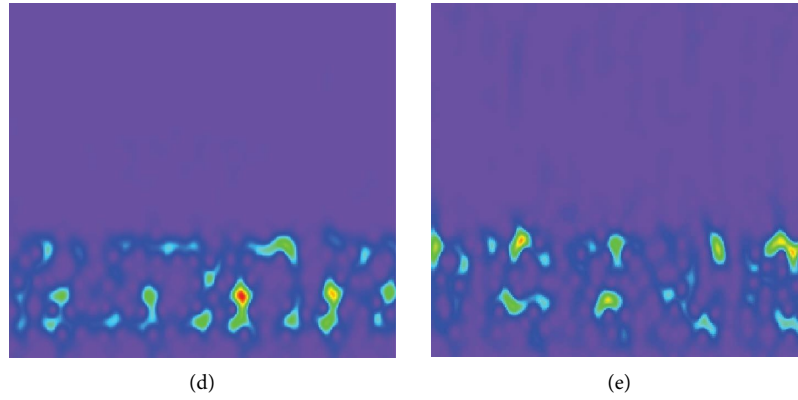


FIGURE 11: Time-frequency diagram obtained by STFT (dataset A). (a) Roll element. (b) Cage. (c) Inner race. (d) Normal. (e) Outer race.

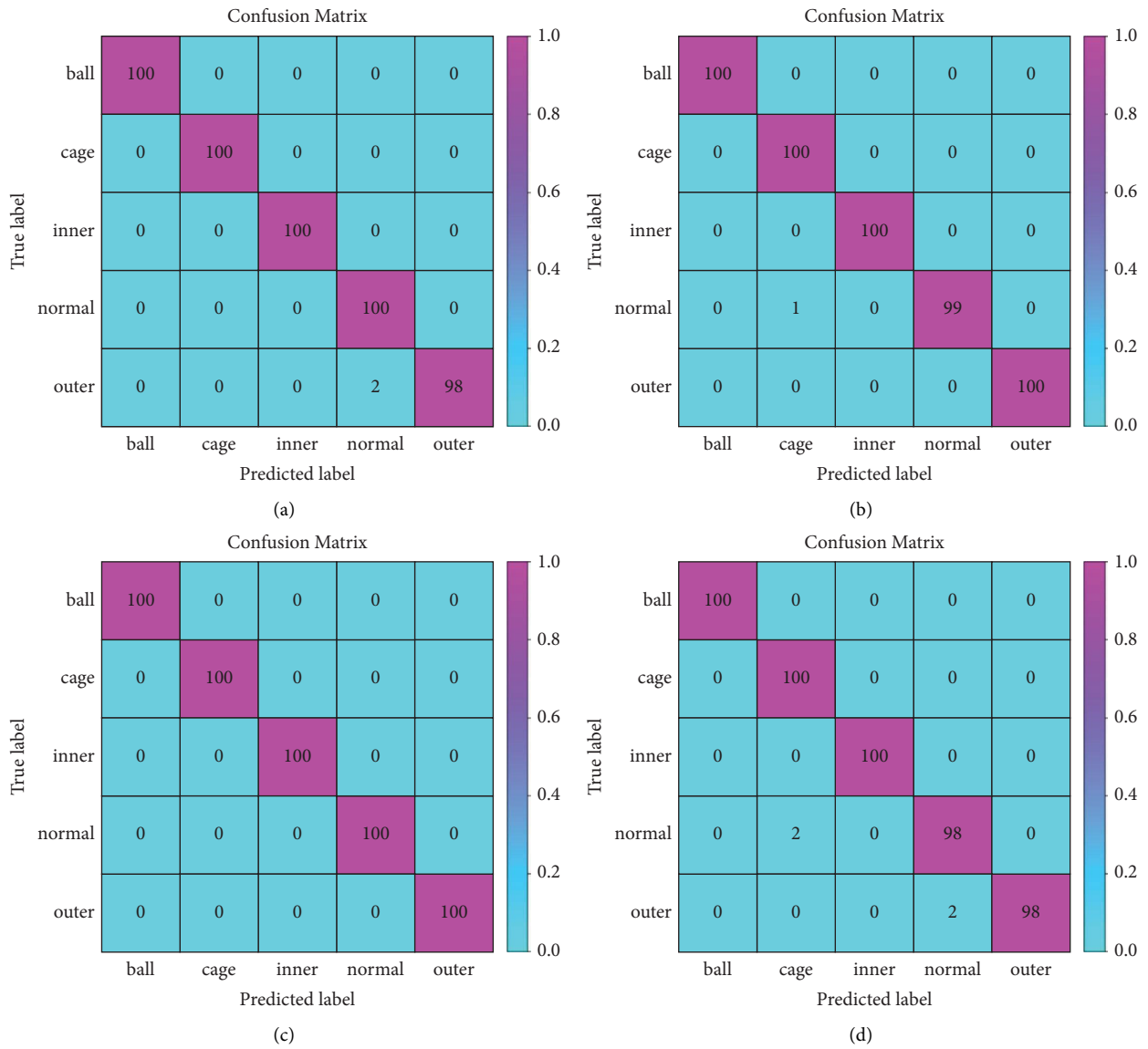


FIGURE 12: Diagnostic results of the model under four datasets composed of different speeds. (a) Results for dataset A. (b) Results for dataset B. (c) Results for dataset C. (d) Results for dataset D.

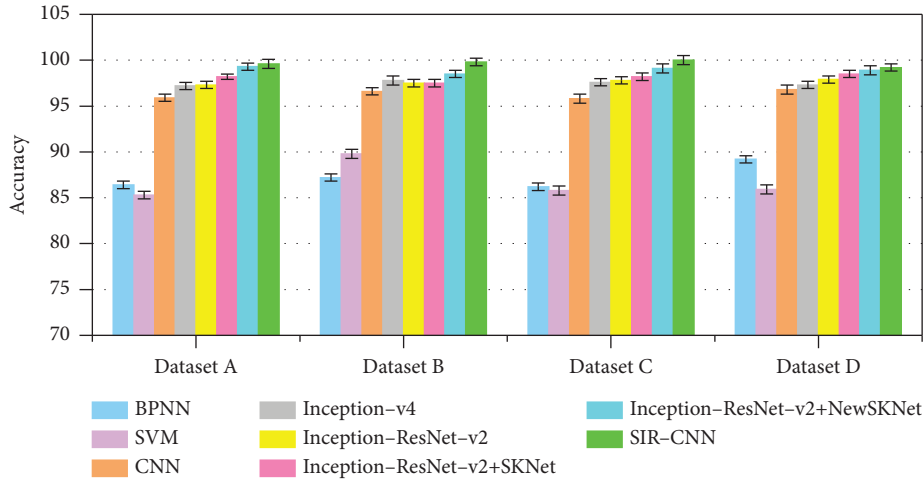


FIGURE 13: Average diagnostic accuracy of various fault diagnosis methods under four datasets.

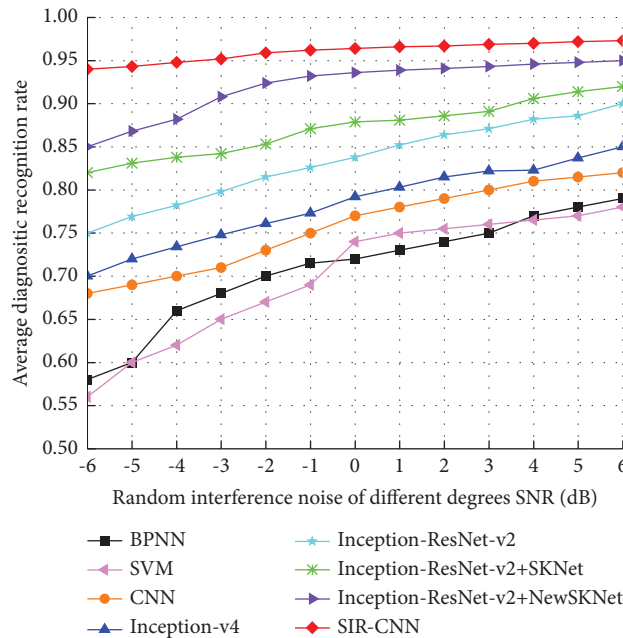


FIGURE 14: Average diagnostic accuracy of various methods under random noise interference.

SIRCNN and STFT-SIRCNN achieved high recognition accuracy when the noise interference was relatively small. However, as the noise interference increased, the recognition accuracy of CWT-SIRCNN decreased significantly. Relatively speaking, the impact of random noise on STFT-SIRCNN was relatively small, and its noise resistance performance tended to stabilize. Although WVD-SIRCNN is not easily affected by noise interference, its overall recognition accuracy is relatively low.

4.2. Case 2: The XJTU-SY Bearing Dataset

4.2.1. Experimental Setup. To further validate the effectiveness of the SIR-CNN algorithm and its fault diagnosis framework proposed in this paper, this section conducted experiments using the XJTU-SY bearing dataset [45]

provided by Xi'an Jiaotong University (XJTU) and the Changxing Sumyoung Technology Co., Ltd. (SY). As shown in Figure 18, the test rig consists of an AC motor, motor speed controller, support bearing, horizontal accelerometer, vertical accelerometer, and tested bearing. A total of three types of working conditions were designed for the test: (1) 2100 rpm with a radial force of 12 kN; (2) 2250 rpm with a radial force of 11 kN; and (3) 2400 rpm with a radial force of 10 kN. Five bearings were used for each working condition for the test, and pictures of the bearings with typical types of failure are given in Figure 19, where it can be seen that the causes of failure of the test bearings include inner race wear, outer race wear, cage fracture, and outer race fracture. In the test, two accelerometers were used to collect the horizontal and vertical vibration signals of the test bearings, respectively, with a sampling frequency of

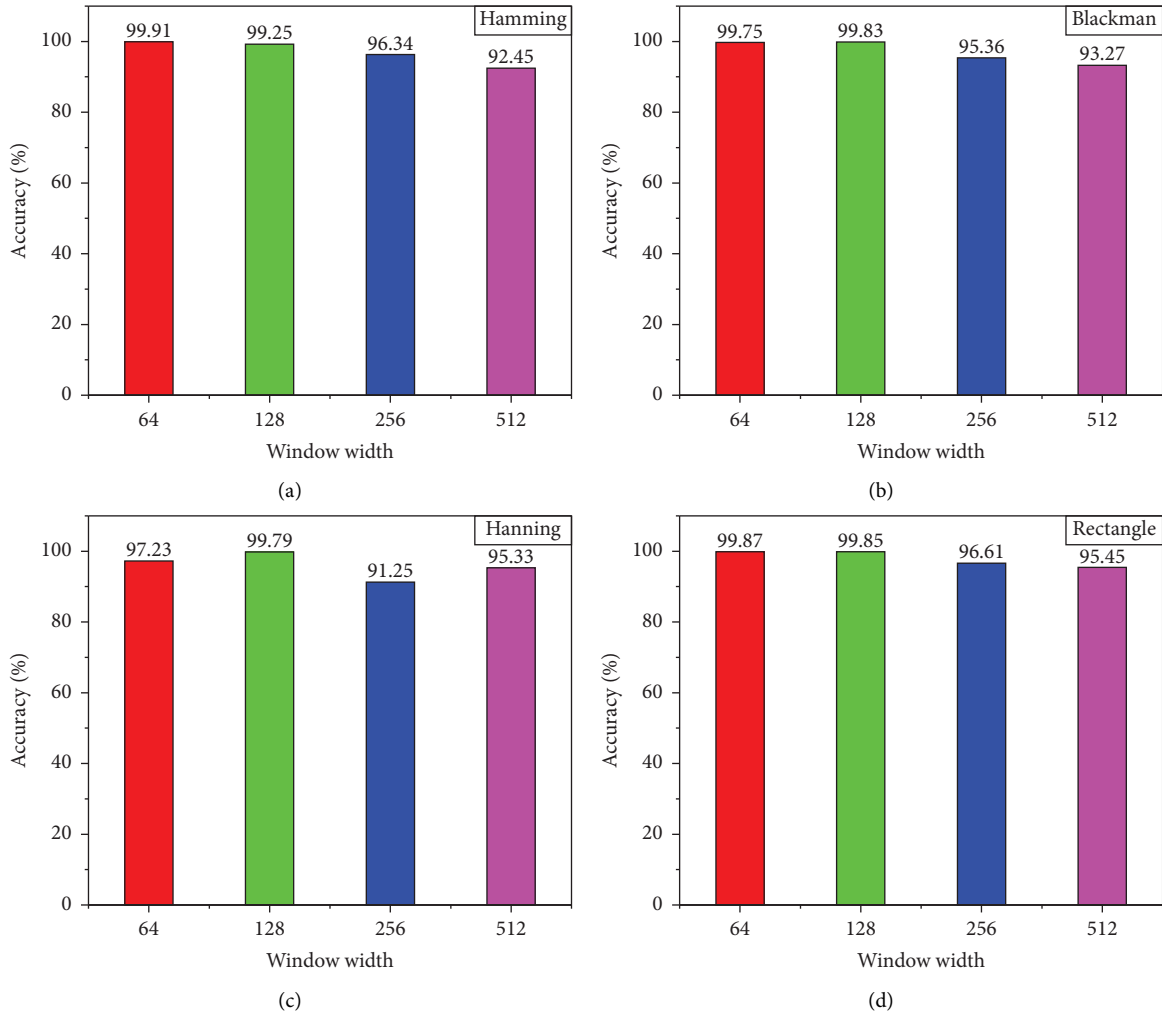


FIGURE 15: The classification accuracies of four window functions. (a) Hamming. (b) Blackman. (c) Hanning. (d) Rectangle.

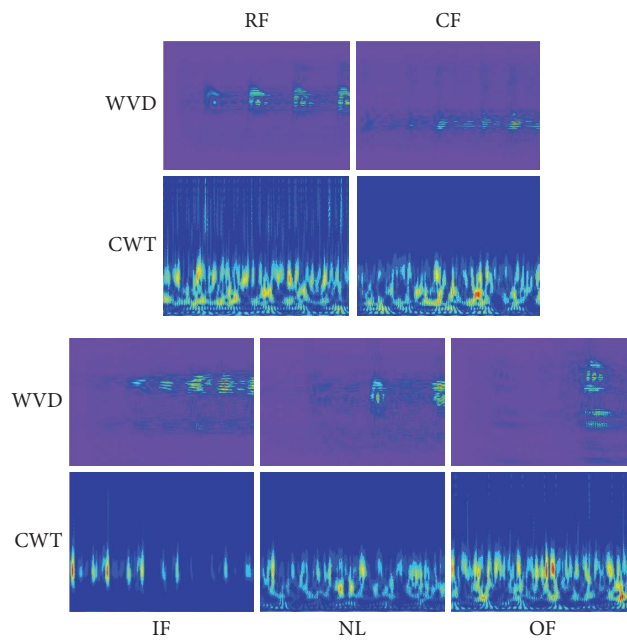


FIGURE 16: Time-frequency maps generated by WVD and CWT (dataset A).

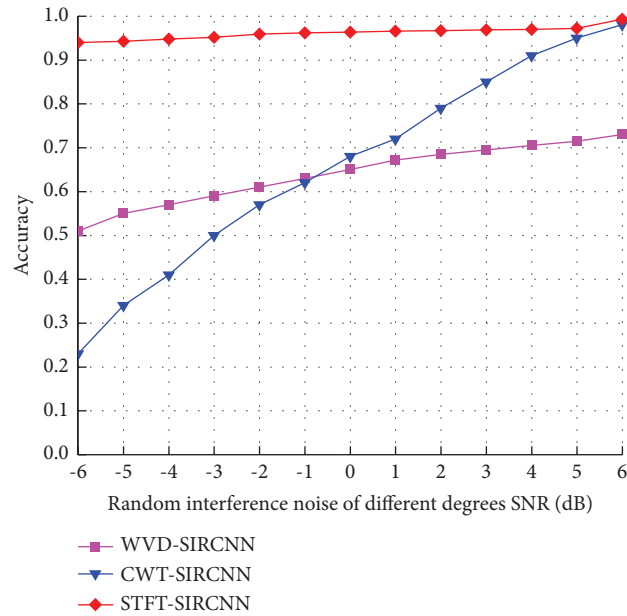


FIGURE 17: Diagnosis results under three preprocessing methods.

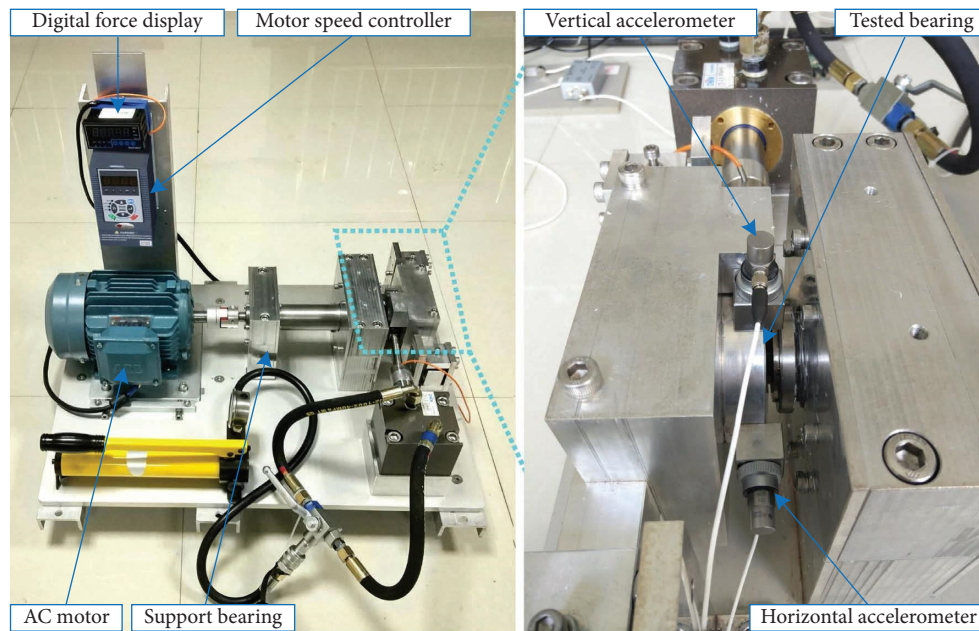


FIGURE 18: XJTU-SY bearing test stand.

25.6 kHz, a sampling interval of 1 min, and a sampling duration of 1.28 s each time.

In this experiment, the outer race dataset is selected in Case 1, the inner race and cage dataset is selected in Case 2, and the outer race, inner race, rolling body, and cage composite fault dataset is selected in Case 3, which constitutes the dataset of four fault types, namely, IR, OR, Cage, and IBCO, as shown in Table 4.

Similarly, in order to fully utilize the powerful image processing capabilities of CNN models, 1D bearing vibration signals were converted into 2D time-frequency maps through STFT, as shown in Figure 20. Then, the obtained

time-frequency map is divided into training set, validation set, and testing set. The training set and validation set are used to train the proposed SIR-CNN, while the testing set is used to validate the trained model.

4.2.2. Diagnostic Results and Comparison. In order to further verify the superiority of the proposed SIR-CNN model, seven other models are selected for comparison experiments, namely, STFT + CNN, STFT + ResNet, enhanced Inception-ResNet-v2, RGB + Inception-ResNet-v2, SKNet + Inception-v4, SKNet + Inception-ResNet-v2, and NewSKNet + Inception-

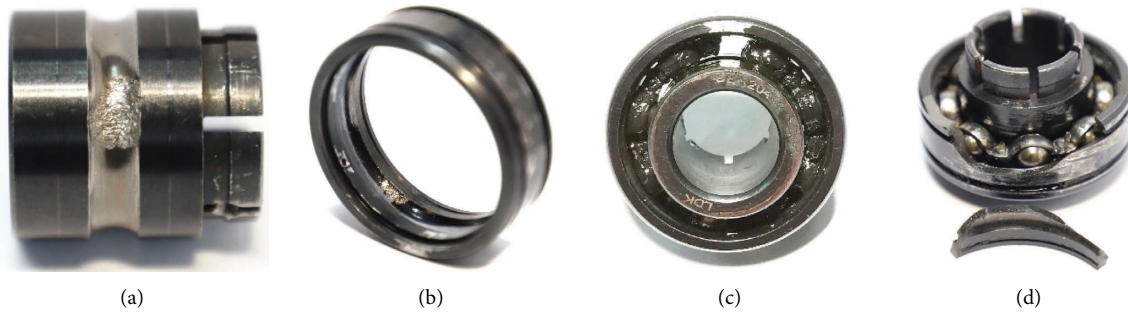


FIGURE 19: Pictures of bearings with typical failure types. (a) Inner race wear. (b) Outer race wear. (c) Cage fracture. (d) Outer race fracture.

TABLE 4: Description of the bearing dataset.

Fault location	Inner race	Outer race	Cage	Inner race, ball, cage, and outer race
Dataset	IR	OR	Cage	IBCO
Training set	700	700	700	700
Validation set	200	200	200	200
Testing set	100	100	100	100

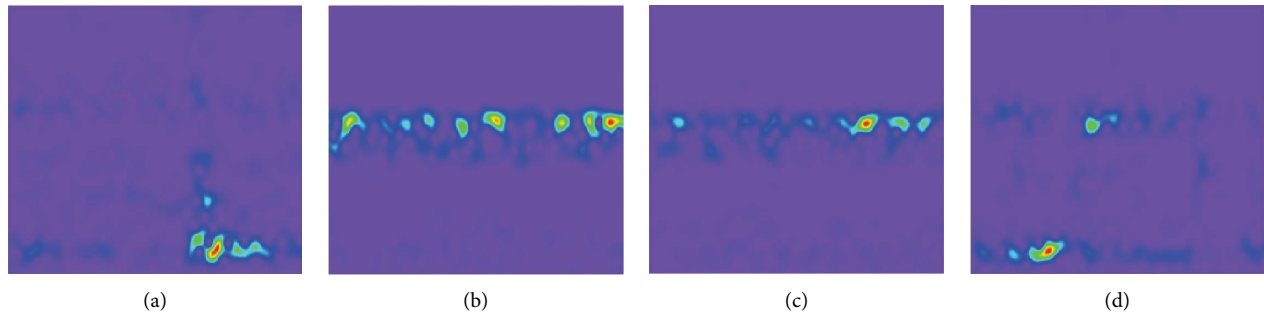


FIGURE 20: Time-frequency diagram obtained by STFT. (a) IR. (b) OR. (c) Cage. (d) IBCO.

ResNet-v2. Each method is experimented 10 times to take the average value, and the results are shown in Table 5. It can be seen that the diagnostic accuracies of the first four models are relatively low, while the two models SKNet + Inception-v4 and SKNet + Inception-ResNet-v2 improve their diagnostic accuracies by at least 3.8% compared to the first four models due to the integration of SKNet. Overall, the proposed SIR-CNN obtains the highest recognition accuracy, with a 1.8% improvement in classification accuracy over NewSKNet + Inception-ResNet-v2 and a 12% improvement over STFT + CNN.

In order to illustrate the classification details of the various methods on different fault category samples in more detail, the confusion matrix is used to show the experimental results. Figure 21 shows the confusion matrices of the proposed method and the other seven methods under the XJTU-SY dataset, and it can be seen that the proposed SIR-CNN method has a strong feature learning capability on each fault category dataset, and a high classification accuracy is obtained.

In order to more intuitively analyze the diagnostic performance of different methods under the XJTU-SY

TABLE 5: Comparative performance of different methods.

Model	Accuracy (%)
STFT + CNN [41]	87.9
STFT + ResNet [46]	89.7
Enhanced Inception-ResNet-v2 [34]	91.0
RGB + Inception-ResNet-v2 [30]	92.5
SKNet + Inception-v4 [47]	96.3
SKNet + Inception-ResNet-v2	96.9
NewSKNet + Inception-ResNet-v2	98.1
SIR-CNN (ours)	99.9

dataset, the features extracted by the model are visualized using t-SNE, and the results are shown in Figure 22. It can be seen that the features extracted by STFT + CNN, STFT + ResNet, enhanced Inception-ResNet-v2, and RGB + Inception-ResNet-v2 are all severely confounded; SKNet + Inception-v4, SKNet + Inception-ResNet-v2, and NewSKNet + Inception-ResNet-v2 extracted features are also slightly aliased. In contrast, in the features extracted by SIR-CNN, the samples of the same category are completely

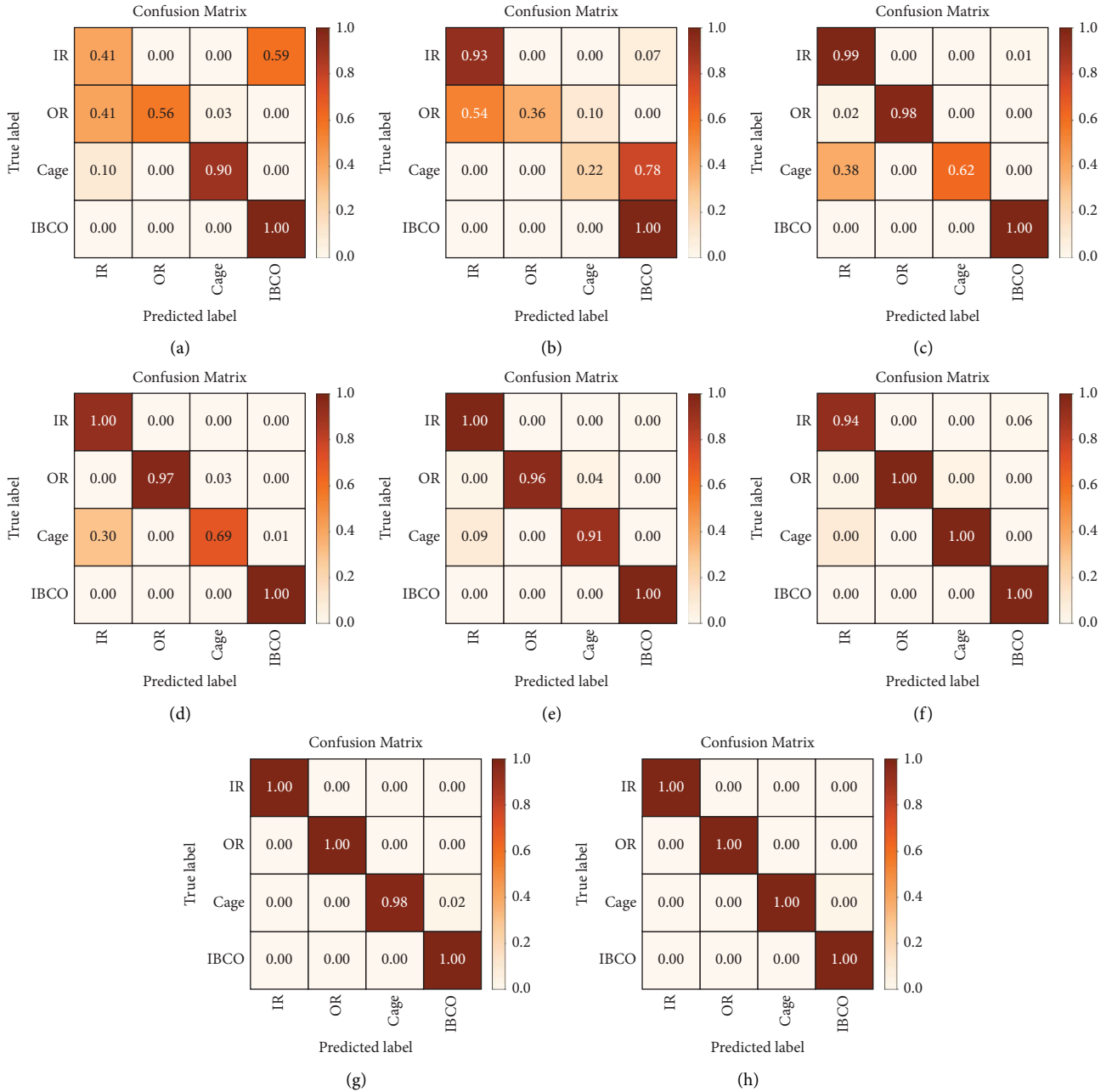


FIGURE 21: The confusion matrix under XJTU-SY dataset. (a) STFT + CNN. (b) STFT + ResNet. (c) Enhanced Inception-ResNet-v2. (d) RGB + Inception-ResNet-v2. (e) SKNet + Inception-v4. (f) SKNet + Inception-ResNet-v2. (g) NewSKNet + Inception-ResNet-v2. (h) SIR-CNN (ours).

aggregated together, and the samples of different categories are completely separated, which again indicates that the SIR-CNN-based fault identification method has a more powerful feature extraction capability and superior classification performance.

4.2.3. The Effectiveness of SIR-CNN. To further validate the effectiveness of model improvement, we compared the number of parameters, network training time, and network testing time between the proposed SIR-CNN and the other four networks. As shown in Table 6, it can be seen that the

proposed SIR-CNN has fewer network parameters, which is consistent with the analysis of equation (14). We also noticed that the number of parameters has an impact on the training and testing times of the network.

Next, we also studied the combination effects of different convolutional kernels. To limit the search space, we only used four different convolutional kernels: 1×1 , 3×3 , 5×5 , and 7×7 . If “NewSKNet” in Table 7 is checked, it means that we use NewSKNet on the corresponding kernels checked in the same row. Otherwise, we simply add the results of these kernels as the output of the model. The results in Table 7 indicate that when the designed NewSKNet is used, lower

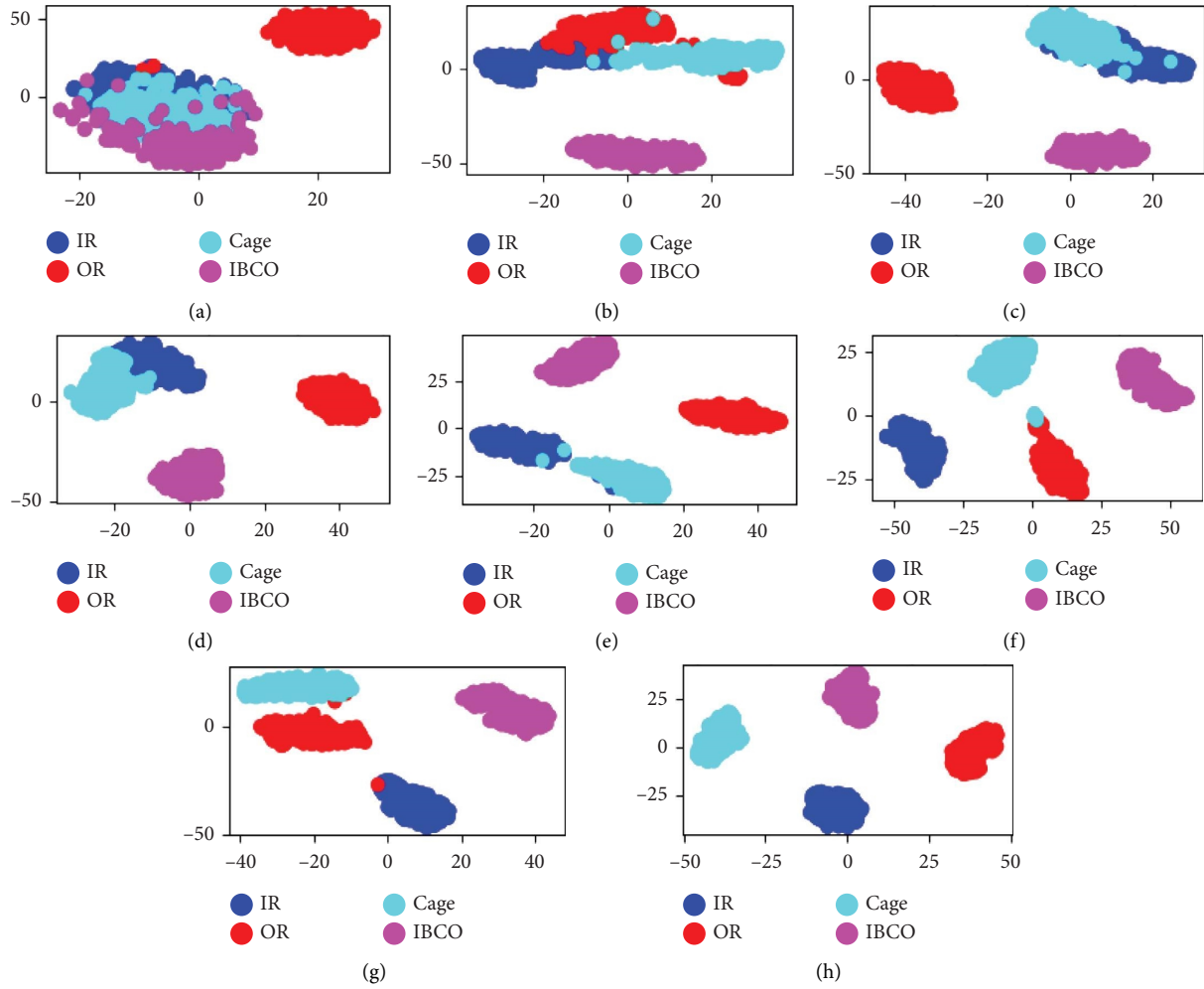


FIGURE 22: Feature visualization. (a) STFT + CNN. (b) STFT + ResNet. (c) Enhanced Inception-ResNet-v2. (d) RGB + Inception-ResNet-v2. (e) SKNet + Inception-v4. (f) SKNet + Inception-ResNet-v2. (g) NewSKNet + Inception-ResNet-v2. (h) SIR-CNN (ours).

TABLE 6: Comparison of parameter quantity and execution time.

Models	Parameters ($\times 10^6$)	Execution time to convergence (h)	Execution time to on test set (s)
Inception-v4 [29]	27.6	8.59	25
Inception-ResNet-v2 [34]	31.3	11.8	31
Inception-ResNet-v2 + SKNet	37.5	13.1	35
Inception-ResNet-v2 + NewSKNet	41.3	15.3	38
SIR-CNN (ours)	5.3	2.56	6

TABLE 7: Combination of different kernels.

1×1	3×3	5×5	7×7	NewSKNet	Loss
✓	✓	✓			23.5
✓	✓	✓		✓	2.4
✓	✓		✓		21.2
✓	✓		✓	✓	2.3
	✓	✓	✓		20.5
	✓	✓	✓	✓	1.4

losses can be achieved, which is attributed to the use of multiple convolutional kernels and the adaptive selection mechanism between them. Due to the fact that lower losses imply higher accuracy, this is consistent with the experimental results mentioned above and further proves the effectiveness of the model improvement.

5. Results and Discussion

As mentioned above, we know that feature extraction is a crucial part in the fault diagnosis process, and its main role is to extract fault-related information from the original signal to match known fault templates to recognize the fault. However, traditional deep learning methods are unable to adaptively select convolutional kernels based on the features of the input image, which results in weak extracted features. In addition, the number of parameters of deep learning models is very large, which usually requires a lot of time to train the models. In order to solve these problems, this paper skillfully fuses the multibranch SKNet, the depthwise separable convolution network and the improved Inception-ResNet-v2 network, and the fused network can extract the features in the input data more effectively and reduce the parameters in the network. Thus, the accuracy of fault diagnosis can be improved and the speed of network convergence can be accelerated. The performance of the method is verified on two datasets. However, there are still some potential problems, which need further improvement and in-depth research.

At high bearing speeds, multiple faults may occur simultaneously or sequentially rather than a single fault. These faults may be correlated or interact with each other, causing the system to exhibit complex fault behavior. However, the present model is tested on two datasets with single faults and has some limitations.

From the empirical values in the literature, the number of convolutional layers, the size of convolutional kernel, the type and size of pooling, and the activation function all have a significant impact on the performance of the model. Therefore, how to choose the parameters of SIR-CNN is a matter to be considered.

It should be noted that the data collected at industrial sites usually contain real noise and are more complex, while the noise data in this paper are modeled by adding Gaussian white noise, which is somewhat different from the real noise data.

The testing of the model is done on two datasets; in the next step of the research work, we need to collect more datasets for further in-depth validation of the effectiveness of the proposed method.

6. Conclusion

To solve the problem that existing intelligent fault diagnosis methods cannot adaptively select the convolution kernel according to different input images, which leads to weak extracted features, this paper proposes an integrated deep neural network, SIR-CNN, which combines proposed NewSKNet and enhanced Inception-ResNet-v2. First, the

1D raw vibration signal is converted into 2D time-frequency diagram using STFT. In this way, more time-frequency features can be obtained, and the powerful image processing capability of the CNN can be fully utilized. Then, based on SKNet, a new three-branch SKNet is designed, and the designed NewSKNet is embedded in the depthwise separable convolution network. Finally, the convolution structure in Inception-ResNet-v2 was replaced by the improved depthwise separable convolution network. The performance of the SIR-CNN was validated on open and measured bearing datasets. Experiments show that the NewSKNet designed in this study can adaptively select convolutional kernels and extract more important signal features, thereby greatly improving the diagnostic accuracy of the model. In addition, as Inception-ResNet-v2 is a complex model with numerous network parameters, we cleverly embed NewSKNet into a depthwise separable convolution and then replace the convolutional module in Inception-ResNet-v2 with a depthwise separable convolution embedded with NewSKNet, which can significantly reduce the parameters in the network and accelerate model fitting.

In the future, we plan to collect more datasets to validate the model and consider introducing transfer learning into the model to further validate the model on real industrial data.

Data Availability

Some or all data, models, and codes used in this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was partially supported by the National Natural Science Foundation of China (Grant no. 71861025), the National Key Research and Development Program of China (Grant no. 2018YFB1703105), the Hongliu First-class Disciplines Development Program of Lanzhou University of Technology, and the Fundamental Ability Enhancement Project for Young and Middle-Aged University Teachers in Guangxi Province (Grant no. 2023KY0810).

References

- [1] W. Fu, X. Jiang, B. Li, C. Tan, B. Chen, and X. Chen, "Rolling bearing fault diagnosis based on 2D time-frequency images and data augmentation technique," *Measurement Science and Technology*, vol. 34, no. 4, 2023.
- [2] X. Zhao, J. Yao, W. Deng, M. Jia, and Z. Liu, "Normalized Conditional Variational Auto-Encoder with adaptive Focal loss for imbalanced fault diagnosis of Bearing-Rotor system," *Mechanical Systems and Signal Processing*, vol. 170, 2022.
- [3] M. Maurya, I. Panigrahi, D. Dash, and C. Malla, "Intelligent fault diagnostic system for rotating machinery based on IoT with cloud computing and artificial intelligence techniques: a review," *Soft Computing*, vol. 28, no. 1, pp. 477–494, 2023.

- [4] Y. Zhang, K. Yu, Z. Lei et al., "Integrated intelligent fault diagnosis approach of offshore wind turbine bearing based on information stream fusion and semi-supervised learning," *Expert Systems with Applications*, vol. 232, 2023.
- [5] N. Kumar and R. Satapathy, "Bearings in aerospace, application, distress, and life: a review," *Journal of Failure Analysis and Prevention*, vol. 23, no. 3, pp. 915–947, 2023.
- [6] H. Zhou, X. Huang, G. Wen et al., "Convolution enabled transformer via random contrastive regularization for rotating machinery diagnosis under time-varying working conditions," *Mechanical Systems and Signal Processing*, vol. 173, 2022.
- [7] J. Zhang, K. Zhang, Y. An, H. Luo, and S. Yin, "An integrated multitasking intelligent bearing fault diagnosis scheme based on representation learning under imbalanced sample condition," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12, 2023.
- [8] H. Tao, J. Qiu, Y. Chen, V. Stojanovic, and L. Cheng, "Un-supervised cross-domain rolling bearing fault diagnosis based on time-frequency information fusion," *Journal of the Franklin Institute*, vol. 360, no. 2, pp. 1454–1477, 2023.
- [9] A. Choudhary, T. Mian, S. Fatima, and B. K. Panigrahi, "Passive thermography based bearing fault diagnosis using transfer learning with varying working conditions," *IEEE Sensors Journal*, vol. 23, no. 5, pp. 4628–4637, 2023.
- [10] I. D. P. Karyatanti, A. Noersena, F. R. Purnomo, R. S. Zulkifli, and A. Wijayanto, "Analysis of outer race bearing damage by calculation of sound signal frequency based on the FFT method," *International Journal of Engineering and Technology Innovation*, vol. 13, no. 1, pp. 28–39, 2023.
- [11] S. Sui, K. Wang, and S. Chen, "Fault detection and quantitative assessment of wheel diameter difference based on ensemble adaptive tunable Q-factor wavelet transform and mixed kernel principal component analysis," *Measurement Science and Technology*, vol. 35, no. 1, 2024.
- [12] B. Qi, Y. Li, W. Yao, and Z. Li, "Application of EMD combined with deep learning and knowledge graph in bearing fault," *Journal of Signal Processing Systems*, vol. 95, no. 8, pp. 935–954, 2023.
- [13] X. Xu, Z. Li, Y. Zhao, and W. Qian, "Fault diagnosis of gas turbine generator bearings using enhanced valuable sample strategy and convolutional neural network," *Measurement Science and Technology*, vol. 35, no. 1, 2024.
- [14] X. Li, Z. Dai, and L. He, "A K-nearest neighbor indoor fingerprint location method based on coarse positioning circular domain and the highest similarity threshold," *Measurement Science and Technology*, vol. 34, no. 1, 2023.
- [15] J. Wang, Z. Lu, G. Wang et al., "Research on fault diagnosis of HMCVT shift hydraulic system based on optimized BPNN and CNN," *Agriculture*, vol. 13, no. 2, p. 461, 2023.
- [16] Z. Zhu, Y. Lei, G. Qi et al., "A review of the application of deep learning in intelligent fault diagnosis of rotating machinery," *Measurement*, vol. 206, 2023.
- [17] J. Yu and Y. Zhang, "Challenges and opportunities of deep learning-based process fault detection and diagnosis: a review," *Neural Computing & Applications*, vol. 35, no. 1, pp. 211–252, 2023.
- [18] W. Zhu, B. Peng, C. Chen, and H. Chen, "Deep discriminative dictionary pair learning for image classification," *Applied Intelligence*, vol. 53, no. 19, pp. 22017–22030, 2023.
- [19] Q. Qian, Y. Wang, T. Zhang, and Y. Qin, "Maximum mean square discrepancy: a new discrepancy representation metric for mechanical fault transfer diagnosis," *Knowledge-Based Systems*, vol. 276, 2023.
- [20] Q. Qian, J. Zhou, and Y. Qin, "Relationship transfer domain generalization network for rotating machinery fault diagnosis under different working conditions," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 9, pp. 9898–9908, 2023.
- [21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [22] Z. Cheng, B. Chen, R. Lu et al., "Recurrent neural networks for snapshot compressive imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 2264–2281, 2023.
- [23] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on generative adversarial networks: algorithms, theory, and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3313–3332, 2023.
- [24] K. Liu, M. Zheng, Y. Liu, J. Yang, and Y. Yao, "Deep autoencoder thermography for defect detection of carbon fiber composites," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 5, pp. 6429–6438, 2023.
- [25] G. Kılıçarslan, C. Koç, F. Özyurt, and Y. Gül, "Breast lesion classification using features fusion and selection of ensemble ResNet method," *International Journal of Imaging Systems and Technology*, vol. 33, no. 5, pp. 1779–1795, 2023.
- [26] H. Wang, Z. Liu, D. Peng, and M. J. Zuo, "Interpretable convolutional neural network with multilayer wavelet for Noise-Robust Machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 195, 2023.
- [27] Z. Zhao and Y. Jiao, "A fault diagnosis method for rotating machinery based on CNN with mixed information," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 8, pp. 9091–9101, 2023.
- [28] J. Zhang, X. Li, J. Tian, H. Luo, and S. Yin, "An integrated multi-head dual sparse self-attention network for remaining useful life prediction," *Reliability Engineering & System Safety*, vol. 233, 2023.
- [29] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," 2016, <https://arxiv.org/abs/1602.07261>.
- [30] Y. Liu, J. Li, Q. Li, and Q. Wang, "Transfer learning with inception ResNet-based model for rolling bearing fault diagnosis," *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, vol. 16, no. 2, 2022.
- [31] Z. Li, H. Wang, J. Chen, Z. Zhou, and W. Chen, "Research on rolling bearing fault diagnosis based on DRS frequency spectrum image and deep learning," *International Journal of Acoustics and Vibration*, vol. 28, no. 2, pp. 211–219, 2023.
- [32] R. Jigyasu, V. Shrivastava, and S. Singh, "Hybrid multi-model feature fusion-based vibration monitoring for rotating machine fault diagnosis," *Journal of Vibration Engineering & Technologies*, 2023.
- [33] B. U. Deveci, M. Celtikoglu, O. Albayrak, P. Unal, and P. Kirci, "Transfer learning enabled bearing fault detection methods based on image representations of single-dimensional signals," *Information Systems Frontiers*, 2023.
- [34] C. Peng, Y. Liu, X. Yuan, and Q. Chen, "Research of image recognition method based on enhanced inception-ResNet-V2," *Multimedia Tools and Applications*, vol. 81, no. 24, pp. 34345–34365, 2022.
- [35] Y. Zheng, "Pipeline multitype artifact recognition method based on Inception_Resnet_V2 structure improving SSD network," *Advances in Multimedia*, vol. 2022, Article ID 6049013, 8 pages, 2022.
- [36] W. Liu, Z. Tang, F. Lv, and X. Chen, "An efficient approach for guided wave structural monitoring of switch rails via deep

- convolutional neural network-based transfer learning,” *Measurement Science and Technology*, vol. 34, no. 2, 2023.
- [37] C. Song, B. Peng, H. Wang et al., “Maize seed appearance quality assessment based on improved Inception-ResNet,” *Frontiers of Plant Science*, vol. 14, 2023.
- [38] B. Das, A. Saha, and S. Mukhopadhyay, “Rain Removal from a Single Image Using Refined Inception ResNet v2,” *Circuits, Systems, and Signal Processing*, vol. 42, no. 6, pp. 3485–3508, 2023.
- [39] H. R. Kasireddy, U. Kallam, S. Mantrala et al., “Deep-learning-based visualization and volumetric analysis of fluid regions in optical coherence tomography scans,” *Diagnostics*, vol. 13, no. 16, p. 2659, 2023.
- [40] P. Meel and D. Kumar Vishwakarma, “Multi-modal fusion using Fine-tuned Self-attention and transfer learning for veracity analysis of web information,” *Expert Systems with Applications*, vol. 229, 2023.
- [41] Q. Zhang and L. Deng, “An intelligent fault diagnosis method of rolling bearings based on short-time fourier transform and convolutional neural network,” *Journal of Failure Analysis and Prevention*, vol. 23, no. 2, pp. 795–811, 2023.
- [42] X. Li, W. Wang, X. Hu, and J. Yang, “Selective kernel networks,” in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 510–519, IEEE, Long Beach, CA, USA, June 2019.
- [43] R. Zhang, F. Zhu, J. Liu, and G. Liu, “Depth-wise separable convolutions and multi-level pooling for an efficient spatial CNN-based steganalysis,” *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1138–1150, 2020.
- [44] B. Hu, J. Liu, R. Zhao, Y. Xu, and T. Huo, “A new fault diagnosis method for unbalanced data based on 1DCNN and L2-SVM,” *Applied Sciences*, vol. 12, no. 19, p. 9880, 2022.
- [45] B. Wang, Y. Lei, N. Li, and N. Li, “A hybrid prognostics approach for estimating remaining useful life of rolling element bearings,” *IEEE Transactions on Reliability*, vol. 69, no. 1, pp. 401–412, 2020.
- [46] M. Wisal and K.-Y. Oh, “A new deep learning framework for imbalance detection of a rotating shaft,” *Sensors*, vol. 23, no. 16, p. 7141, 2023.
- [47] F. Chen, J. Wei, B. Xue, and M. Zhang, “Feature fusion and kernel selective in Inception-v4 network,” *Applied Soft Computing*, vol. 119, 2022.