

Received July 7, 2020, accepted July 14, 2020, date of publication July 20, 2020, date of current version July 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3010272

A Non-Contact Fault Diagnosis Method for Rolling Bearings Based on Acoustic Imaging and Convolutional Neural Networks

RAN WANG¹, FENGKAI LIU¹, FATAO HOU², WEIKANG JIANG², QILIN HOU¹, AND LONGJING YU¹, (Graduate Student Member, IEEE)

¹College of Logistics Engineering, Shanghai Maritime University, Shanghai 201306, China

²State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200240, China

Corresponding author: Ran Wang (ranwang@shmtu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 51505277.

ABSTRACT Rolling bearing fault diagnosis is conventionally performed by vibration-based diagnosis (VBD). However, VBD is restrained in some cases because vibration measurement usually requires the contact with the machine. Acoustical-based fault diagnosis (ABD) has the advantage of non-contact measurement over VBD. However, ABD has received little attention and rarely applied in bearing fault diagnosis. In this paper, a new non-contact ABD method for rolling bearings using acoustic imaging and convolutional neural networks (CNN) is proposed. Firstly, a microphone array is used to acquire the acoustic field radiated by rolling bearings. Then, acoustic imaging is performed with the wave superposition method (WSM). The reconstructed acoustic images can depict the spatial distribution of the acoustic field, which add a new spatial dimension in the acoustic data representation for fault diagnosis and makes it possible to localize the sound sources. Finally, CNN is applied to accomplish bearing fault diagnosis, which can overcome the problems of handcrafted feature extraction in traditional ABD methods. Experimental results verify the effectiveness of the proposed ABD method. Comparisons with peer state-of-the-art ABD methods further validate that the proposed method can mitigate the drawbacks of the existing ABD methods, and obtain more accurate and reliable diagnosis results.


INDEX TERMS Bearing fault diagnosis, acoustic imaging, CNN, wave superposition method, acoustical-based fault diagnosis.

I. INTRODUCTION

Rolling bearings are one of the vital components in rotating machinery, and their failure is one of the most frequent reasons for machine breakdown [1]. Therefore, surveillance and fault diagnosis of rolling bearings have attracted considerable attention over the past several decades. A great deal of research has been done on vibration-based diagnosis (VBD) of rolling bearings, which mainly includes three procedures: vibration measurement, feature extraction, and intelligent fault pattern classification. In order to extract effective fault features, various signal processing methods have been proposed, such as spectral kurtosis, envelop spectrum, empirical mode decomposition, and so on [2]–[6]. Then, some conventional machine learning methods, for example, support vector machine (SVM), and hidden Markov model (HMM), have been adopted for intelligent bearing fault classification

[7], [8]. Although VBD is a well-developed and prevalent rolling bearing fault detection approach, it has some limitations in practical applications. Vibration measurement usually needs to be performed in a contact way. However, in some cases, due to the irregular machinery geometry, or harsh testing environments such as high temperature, high humidity, corrosion, and other harmful situations, it is difficult to mount vibration sensors like accelerometers on the surface of the machine. In these cases, we have to turn to a non-contact method for machinery fault diagnosis.

In this paper, we focus on the non-contact acoustical-based fault diagnosis (ABD) methods. Acoustic signals can be acquired via non-contact microphones, and also contain abundant information of machine health condition. However, compared with the well-studied VBD, ABD methods have not been researched extensively. Baydar and Ball [9], [10] made a comparative study of acoustic signals and vibration signals in gear fault detection using Wigner-Ville distribution and wavelet transform. Yadav *et al.* [11] applied an ABD

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

method in internal combustion engine fault diagnosis using fast Fourier transform (FFT) and correlation analysis. In these studies, acoustic signals were recorded by one or several microphones, and then processed by various traditional signal processing methods in the same way as vibration signals. Although these traditional ABD methods have been proved effective in some applications, they suffer from some shortcomings. Firstly, they are vulnerable to the interference of reflected waves, scattered waves, and noises radiated from other sound sources. Secondly, defective components of the machine are hard to localize with these methods. Moreover, for traditional ABD methods, we need to choose proper measurement positions, which is difficult without prior knowledge.

In recent years, with the development of the acoustic imaging techniques and low-cost MEMS microphone arrays, research interests have been increasing in acoustical-based fault diagnosis. A microphone array can obtain the time-space-frequency information of the sound field simultaneously. Based on acoustic measurement by a microphone array, some researchers have attempted to apply acoustic imaging techniques including nearfield acoustical holography (NAH) and beamforming to machinery fault diagnosis. Lu *et al.* [12], [13] proposed a fault diagnosis scheme based on FFT-based NAH and gray level co-occurrence matrix (GLCM), and applied it in gearbox and rolling element bearing fault diagnosis. Cabada *et al.* [14] combined beamforming and spectral kurtosis to perform fault detection of rotating machines. Wang *et al.* [15] proposed a bearing fault diagnosis method based on acoustic imaging and Gabor wavelet transform (GWT). Darraz *et al.* [16] made a preliminary machinery fault diagnosis study based on acoustic imaging and artificial intelligence. However, in this study, acoustic imaging was just used to select acoustic signals radiated from the direction of the sound source, and hereby to make the fault diagnosis less sensitive to disturbance. In general, as a powerful sound field visualization tool, the application of acoustic imaging in machinery fault diagnosis can overcome the difficulty of choosing measurement positions, and makes it possible to localize the sound sources in the machinery. Also, it can provide a new spatial dimension in acoustic representations for fault diagnosis. However, feature extraction in these methods is conducted manually. It is challenging and time-consuming to choose suitable feature extraction methods because the optimal feature set often varies from case to case in different applications.

Deep learning (DL) can learn multi-scale/hierarchical representation automatically and address large-scale data [17]. In recent years, motivated by its strong ability of automatic feature learning and classification, some researchers have attempted to introduce DL into machinery fault diagnosis. Various DL models have been investigated, such as deep belief network, auto-encoder and so on [18]–[20]. As one of the most popular DL models, convolutional neural networks (CNN) have been widely used for image processing, which can learn identifiable features from high-dimensional

structures like 2-D images. CNNs have also been applied in vibration-based machinery fault diagnosis [21]–[23]. For example, Janssens [24] exploited CNN to learn features from the frequency spectra of bearing vibration signals. Ding and He [25] proposed a deep CNN which took wavelet packet energy image as input for spindle bearing fault diagnosis. Wen *et al.* [26] converted the raw one-dimensional (1-D) vibration time-series into 2-D images and then fed these images into a CNN structure based on LeNet-5 [27], and then tested its effectiveness on a motor bearing dataset. Chen *et al.* [28] proposed a bearing fault diagnosis method based on 2D maps of cyclic spectral coherence and CNN. These studies verified that CNN-based fault diagnosis methods outperformed the traditional methods based on manually designed features and shallow machine learning algorithms.

However, most of the studies on DL-based fault diagnosis were related to VBD. So, their applications in some special circumstances may still be limited due to the requirement of contact measurement. In addition, most of the previous DL models were originally designed for image recognition. To fit the input format of these DL models, sometimes one-dimensional vibration signals had to be converted into 2-D images before they were fed into DL models [26]. So far, there has been few literatures on the application of DL in ABD. Li *et al.* [29] adopted DBM and random forest data fusion of acoustic and vibration signals for gearbox fault diagnosis. Yao *et al.* [30] proposed an end-to-end CNN model for gear fault diagnosis based on acoustic signals. Kumar *et al.* [31] proposed an improved deep CNN model for defect identification of the centrifugal pump using acoustic images. But in this study, acoustic images were obtained by analytical wavelet transform of individual acoustic signals. Generally speaking, in these studies, acoustic signals were fed into DL models just in the same way as vibration signals. Consequently, the disadvantages of traditional ABD methods still exist, such as the difficulty of choosing measurement positions, noise interference, and lack of the capability of noise source localization.

To address the aforementioned issues, inspired by the success of CNN in image recognition, we propose a new ABD method using acoustic imaging and CNN for rolling bearing fault diagnosis. Firstly, a microphone array is used to acquire the acoustic field radiated by rolling bearings. And then, acoustic imaging is conducted to obtain the acoustic images of bearings under different conditions. As a widely used NAH technique, wave superposition method (WSM) [32], [33] can be used for the sound field visualization of arbitrarily shaped sound sources, with high reconstruction accuracy and efficiency. Because of these advantages, WSM is chosen as the acoustic imaging tool in this paper. The reconstructed acoustic images can depict the spatial distribution of the acoustic field which is highly-related to the fault pattern of different bearings. Meanwhile, the sound sources can be intuitively localized through the acoustic images. Finally, a CNN model is designed and introduced to process acoustic images. The reconstructed acoustic images are fed into a

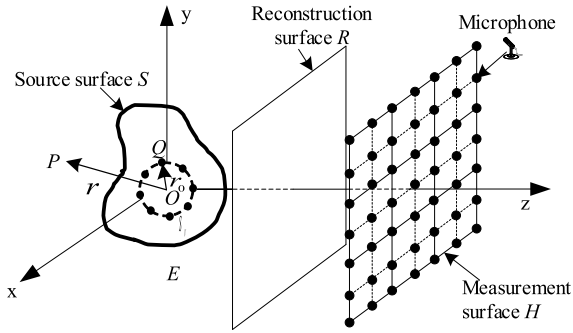


FIGURE 1. Schematic diagram of sound field reconstruction with WSM.

CNN model to automatically learn discriminative and robust features from the acoustic images and accomplish bearing fault diagnosis. Therefore, the main contribution of this paper lies in the combination of WSM-based acoustic imaging and CNN in rolling bearing fault diagnosis. WSM-based acoustic imaging facilitates the usage of the sufficient spatial distribution information of acoustic field in bearing fault diagnosis. It makes the proposed ABD method possible to localize the sound sources and less sensitive to the measurement positions. Furthermore, the combination with CNN can mitigate the problem of handcrafted features, and finally enhance the diagnosis accuracy.

The remainder of this paper is organized as follows. In Section2, the basic principles of WSM-based acoustic imaging are presented. And the theoretical background of CNN is briefly introduced in Section3. Then, Section4 describes the proposed ABD method for rolling bearings using acoustic imaging and a designed CNN model. In Section5, a rolling bearing fault diagnosis experiment is carried out to evaluate the proposed ABD method. At last, some conclusions and discussions are drawn in Section6.

II. THEORY OF WAVE SUPERPOSITION METHOD

Acoustic imaging is usually realized by beamforming or NAH. Beamforming is simple and efficient, which has been widely used for sound source localization in noise control. However, it has a low spatial resolution at low frequencies. By contrast, NAH is a powerful tool for sound field visualization, which can reconstruct the sound field accurately and obtain super-resolution acoustic images. The aim of acoustic imaging in this paper is to obtain acoustic images for fault diagnosis. To this end, acoustic images need to be accurate and have high resolution. On this account, NAH is more suitable for acoustic imaging. Among commonly used NAH techniques, WSM can reconstruct the acoustic radiation from an arbitrarily shaped sound source accurately with far less complexity [34], [35]. It is also very efficient and easy to implement. Because of these advantages, WSM is adopted as the acoustic imaging tool in this paper.

The schematic diagram of sound field reconstruction based on WSM is shown in Fig. 1. An arbitrarily shaped vibrating object with boundary S is immersed in an unbounded homogeneous fluid medium of density ρ and sound speed c . The exterior region outside the source surface S is denoted

as E , and P is a random point in E . The equivalent simple sources (monopoles in this paper) are assumed to be continuously distributed on a virtual surface S_V interior to the source surface S , and Q represents one of them. The location vectors of the field point P and virtual source Q are represented as r and r_0 respectively. O denotes the origin of the coordinate system. The measurement surface H , which is constituted by a microphone array, and the reconstruction surface R are all parallel to the x - y plane.

According to the theory of WSM [36], the acoustic field of an arbitrarily shaped sound source can be equivalently substituted by a series of virtual sources distributed on a virtual surface S_V . So, the acoustic pressure in the exterior region E can be written as

$$p(\mathbf{r}) = i\omega\rho \int_{S_V} q(\mathbf{r}_0)G(\mathbf{r}, \mathbf{r}_0)dS_V(\mathbf{r}_0) \quad (1)$$

where ω is the vibrating angular frequency, $q(\mathbf{r}_0)$ is the source strength, G is the free-space Green's function, which is defined as

$$G(\mathbf{r}, \mathbf{r}_0) = \frac{\exp(ik|\mathbf{r} - \mathbf{r}_0|)}{4\pi|\mathbf{r} - \mathbf{r}_0|} \quad (2)$$

where $k = \omega/c$ is the wave number.

In practical applications, equivalent sources are usually uniformly distributed on a regular virtual surface S_V for simplicity, such as sphere, plane, etc. Meanwhile, In practical applications, equivalent sources are usually uniformly distributed on a regular virtual surface S_V for simplicity, such as sphere, plane, etc. Meanwhile, S_V is discretized into M elements, and the source strength $q(\mathbf{r}_0)$ of each element is assumed to be constant. If the number of the measurement points in the microphone array is N , then according to Eq. (1), the acoustic pressures of the measurement points P_H can be written in a matrix form as

$$\mathbf{P}_H = i\omega\rho\mathbf{G}_{HV}\mathbf{q}, \quad (3)$$

where $\mathbf{P}_H = [p(r_1), p(r_2) \cdots p(r_N)]^T$, \mathbf{q} is the equivalent source strength vector, and \mathbf{G}_{HV} is a propagation matrix relating the equivalent source strengths and the acoustic pressures of the measurement points, which can be calculated according to Eq. (2) as

$$\mathbf{G}_{HV} \mathbf{D} \begin{bmatrix} G(r_1, r_{01}) \cdots G(r_1, r_{0M}) \\ \vdots \\ G(r_N, r_{01}) \cdots G(r_N, r_{0M}) \end{bmatrix} \quad (4)$$

Given the measurement acoustic pressures \mathbf{P}_H , \mathbf{q} can be evaluated by inverting the corresponding matrix \mathbf{G}_{HV} as

$$\mathbf{q} = \frac{1}{i\omega\rho}\mathbf{G}_{HV}^{-1}\mathbf{P}_H \quad (5)$$

where \mathbf{G}_{HV}^{-1} denotes the inverse matrix of \mathbf{G}_{HV} . It should be noted that the solution of \mathbf{q} is a typical inverse problem, and the matrix \mathbf{G}_{HV} is often ill-conditioned. In order to overcome the ill-posed problem, Tikhonov regularization (TR) method

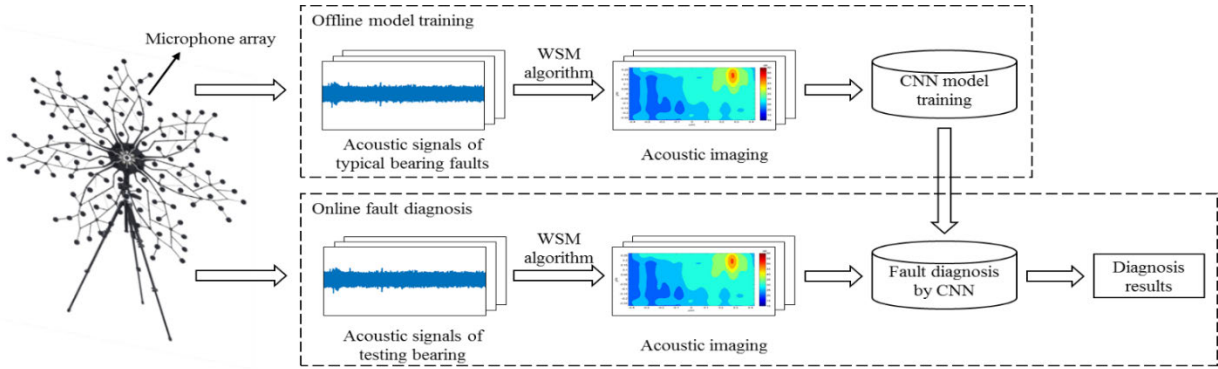


FIGURE 2. Sketch of the proposed ABD method for rolling bearings.

[32], [37] is employed, and its regularization parameter is determined by generalized cross validation (GCV) [38].

Finally, the acoustic pressures P_R of the field points on reconstruction surface R can be reconstructed as

$$P_R = i\omega\rho G_{RV}q \quad (6)$$

where G_{RV} is the propagation matrix relating the equivalent source strengths and the acoustic pressures of the reconstruction points. The elements of G_{RV} can be calculated by substituting the locations of reconstruction point into Eq. (2). Subsequently, the reconstructed acoustic pressures can be represented as 2-D acoustic images, for example, contour maps.

III. THEORETICAL BACKGROUND OF CNNS

CNNs are a special kind of feed-forward neural networks. Inspired by the animal visual cortex, each neuron in a layer is only sparsely connected to a small set of neurons in the previous layer in CNNs. CNNs are mainly characterized by three properties: local connectivity, spatially shared weights, and spatial pooling, which bring the benefits of a faster training phase and less training parameters [24].

Typical CNNs are mainly constituted by three kinds of layers: the convolutional layer, the pooling layer, and the fully connected layer. With multiple combinations of these three layers, CNNs are constructed to realize automatic feature learning and classification.

(1) Convolutional layer. The convolutional layer utilizes a set of kernels/filters to obtain the feature maps of the input images. In this layer, each neuron is connected to a local receptive field of the preceding layer. The kernels, which are actually weights of a grid structure equal to the size of the local receptive field, are convoluted across the input maps of the lower layer. And then, new feature maps are learned through further nonlinear transformation. The convolution operation with nonlinear transformation can be summarized as

$$h_n^{l,r} = \varphi\left(\sum_m x_m^{l-1} * k_n^l + b_n^l\right) \quad (7)$$

where $h_n^{l,r}$ is the n th output feature map in layer l , and r is the number of the kernels. x_m^{l-1} is the m th input in the previous

$l-1$ layer, k_n^l is the kernel, and $*$ denotes the convolution operation. b_n^l is the bias term of the n th output feature map. And $\varphi()$ is the nonlinear activation function. In this article, the rectified linear unit (ReLU) function is adopted as the activation function. It can be seen that in the same layer, m different input maps share the same kernels and bias. Consequently, the number of weights is reduced.

(2) Pooling layer. Pooling layer is used to reduce the dimension of the feature maps after the convolutional layer. Through subsampling, parameters in CNNs will be reduced and the feature maps will be more robust. In this article, max-pooling is adopted as

$$h_m^l(a, b) = \max_{0 \leq p, q < s} \left\{ x_m^{l-1}(a \cdot s + p, b \cdot s + q) \right\} \quad (8)$$

where h_m^l and x_m^{l-1} denote the m th output feature map in layer l and the input feature map in the previous layer, respectively. And s is the pooling size. It can be seen that max-pooling is conducted by sliding a small window over the input feature map while extracting the maximum value of the local region with the size of $s \times s$.

(3) Fully connected layer. The fully connected layer is similar to a traditional multilayer neural network. In this layer, the learned features of the last convolutional layer or pooling layer will be unfolded. And then, a softmax layer is often added as the top layer to perform classification or predictions. Detailed information about CNNs can be seen in reference [27].

IV. ROLLING BEARING FAULT DIAGNOSIS USING ACOUSTIC IMAGING AND CNN

When it comes to rolling bearing fault diagnosis, local pitting is a common cause of bearing failures and has attracted much attention in many studies. During the rotating of the bearing, the local pitting of inner race, outer race, or rolling elements will strike other parts and excite an impulsive vibration. Meanwhile, through propagating in the air, acoustic signals are also radiated from the bearings. In this article, local pitting faults of rolling bearings as well as the fault severity are under consideration. The sketch of the proposed ABD method of rolling bearings is illustrated in Fig. 2.

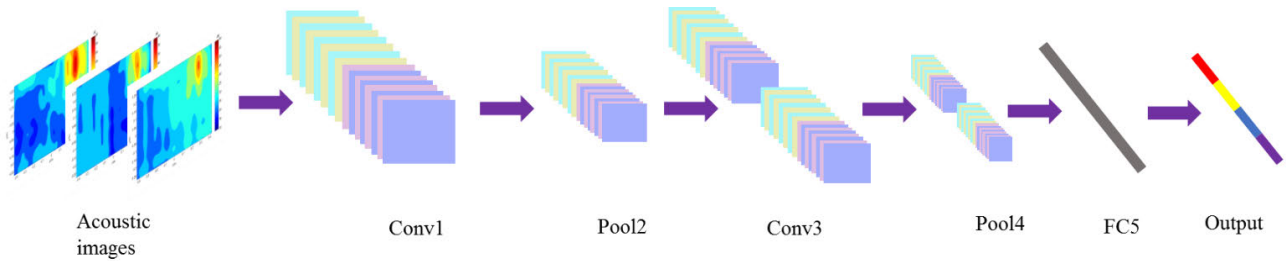


FIGURE 3. The architecture of the modified CNN model.

The proposed method contains two stages: offline training and online fault diagnosis. It can be seen that there are three vital steps in the proposed ABD method, which are described in detail as follows.

A. STEP1: ACOUSTIC SIGNAL MEASUREMENT

In order to record acoustic signals radiated from bearings under different conditions, a microphone array is used. Because WSM is adopted for acoustic imaging, the microphone array needs to be placed in the near-field of the rolling bearings.

B. STEP2: ACOUSTIC IMAGING

When acoustic imaging is performed using WSM, the reconstruction frequency needs to be selected, which is denoted as ω in Eq. (6). As explained above, defective rolling bearings can radiate periodic impulsive acoustic signals, which are typical broadband signals. Anyway, the frequencies of these periodic shocks are related to the locations of the pitting fault, which are widely used in rolling bearing fault diagnosis. These characteristic frequencies include ball pass outer race frequency, ball pass inner race frequency, ball spin frequency, and the fundamental cage frequency, which is denoted as f_o, f_i, f_b and f_c , respectively. Given the shaft rotating frequency and the structural parameters of the rolling bearing, these characteristic frequencies can be calculated by the following equations [1].

$$f_o = \frac{Zf_r}{2} \left(1 - \frac{d}{D} \cos \theta\right) \tag{9}$$

$$f_i = \frac{Zf_r}{2} \left(1 + \frac{d}{D} \cos \theta\right) \tag{10}$$

$$f_b = \frac{f_r D}{d} \left\{ 1 - \left(\frac{d}{D} \cos \theta\right)^2 \right\} \tag{11}$$

$$f_c = \frac{f_r}{2} \left(1 - \frac{d}{D} \cos \theta\right) \tag{12}$$

where f_r is the shaft rotating frequency, Z is the number of rolling elements, d is the ball diameter, D is the pitch diameter, and θ is the contact angle.

Characteristic frequencies in Equations (9)-(12) are usually used in vibration signal analysis. As we know, machinery acoustic signals are caused by vibration. So, it is reasonable to choose these characteristic frequencies as the reconstruction frequencies in acoustic imaging [12]. Subsequently, acoustic images of bearings under different conditions on the reconstruction surface can be obtained

TABLE 1. Detailed parameters of the modified CNN model.

Layer name	Parameter name	Size
Conv1	Kernels	3 x 3 x 64
Pool2	Max Pooling size	2 x 2
Conv3	Kernels	3 x 3 x 64
Pool4	Max Pooling size	2 x 3

with WSM. These reconstructed acoustic images can help the engineers localize and identify the main noise sources of prime concern. Meanwhile, these acoustic images can reflect the spatial distribution of the radiated acoustic field intuitively, which is highly related to the fault patterns of the rolling bearings.

C. STEP3: FAULT DIAGNOSIS BASED ON CNN

Once the acoustic images of rolling bearings are reconstructed with WSM, a CNN model will be established and trained for automatic feature learning and fault classification. As a classical CNN model, LeNet-5 [27], has been successfully applied in handwritten character recognition and other computer vision tasks. Some researchers also have applied LeNet-5 to bearing fault diagnosis [26], [28]. Inspired by the typical LeNet-5, in this section, a CNN model is designed to automatically learn features from the acoustic images for bearing fault diagnosis. Since the task of acoustic image classification is not very complex, the adopted CNN model can be simplified to enhance the computation efficiency. The architecture of the proposed CNN model is illustrated in Fig. 3.

As shown in Fig. 3, the proposed CNN model consists of two alternating convolutional and pooling layers, one fully connected layer, and one output layer. In this paper, the convolutional layer, pooling layer, and fully connected layer are abbreviated as Conv, Pool, and FC, respectively. Detailed parameters of the CNN model are summarized in Table 1.

The inputs of the CNN model in both the offline model training stage and fault diagnosis stage are all acoustic images of rolling bearings, which are obtained by following steps 1 and 2. In the offline training stage, acoustic image features are firstly learned by Conv 1 and then down-sampled by Pool 2. In the convolution process, zero-padding is adopted to keep the spatial dimension of the input. Then, high-level features are extracted by repeating the feature learning and subsampling processes in Conv 3 and Pool 4. Next, features in Pool 4 are flattened into a vector and fed into FC 5.

The dropout strategy is adopted in FC 5 to reduce the overfitting risk. In the last output layer, softmax regression is exploited as the classifier to diagnosis the bearing health conditions based on the learned features from acoustic images. The output number of the softmax classifier is determined by the number of bearing health states. Softmax regression can provide a probability for every possible class, and the one with the highest probability will be predicted as the target fault type.

In the offline training stage, weights of the whole CNN model are firstly initialized randomly, and then back propagation (BP) is performed to update the network weights through minimizing the cross-entropy loss between the softmax output and the true labels. Eventually, in the online fault diagnosis stage, acoustic images of the testing bearings are obtained by acoustic imaging, and then fed into the trained CNN model to gain the final diagnosis results. Practical issues of applying the proposed method will be discussed in the following experimental research.

V. EXPERIMENTAL VERIFICATION

A. EXPERIMENTAL SETUP

To verify the effectiveness of the proposed ABD method, an experiment of rolling bearing fault diagnosis was carried out in the semi-anechoic chamber of the state key laboratory of mechanical system and vibration, Shanghai Jiao Tong University. The experimental setup is shown in Fig. 4. The test rig consists of a machine set, test bearings, the microphone array, three reference microphones marked as 1, 2, 3 in Fig. 4, and the data acquisition system. The machine set mainly consists of a motor, shafts, belt drive unit, a coupling, a clamp, the lubrication system, and the chassis. The test bearing is fixed by the clamp.

The linear microphone array in Fig. 4(a) is constituted by 12 microphones with a uniform 5cm spacing. The measurement surface is 5cm above the upper surface of the bearing test bench. The distance between the measurement surface and the test bearing is about 10cm. The measurement is taken by scanning on the measurement surface step by step with the linear microphone array. The scanning step distance is 5cm, and the number of steps is 18. By this means, an 85 × 55cm measurement hologram is obtained. It should be noted that acoustic signals recorded in different steps are not temporal synchronous. So, a reference microphone, which is labeled as no. 1 in Fig. 4(a), is placed near the bearing for phase compensation. Moreover, two additional microphones, labeled as no. 2 and 3, are also mounted near the machine set for comparative study. A 32-channel high-speed data acquisition system is used for acoustic signal acquisition, which is shown in Fig. 4(b).

The type of test bearings is GB6203 rolling element bearings, whose structural parameters are displayed in Table 2. In this study, three typical bearing fault types are under consideration, including outer race fault (ORF), inner race fault (IRF), and rolling ball fault (RBF). And the local pitting faults are processed in the inner races, outer races,

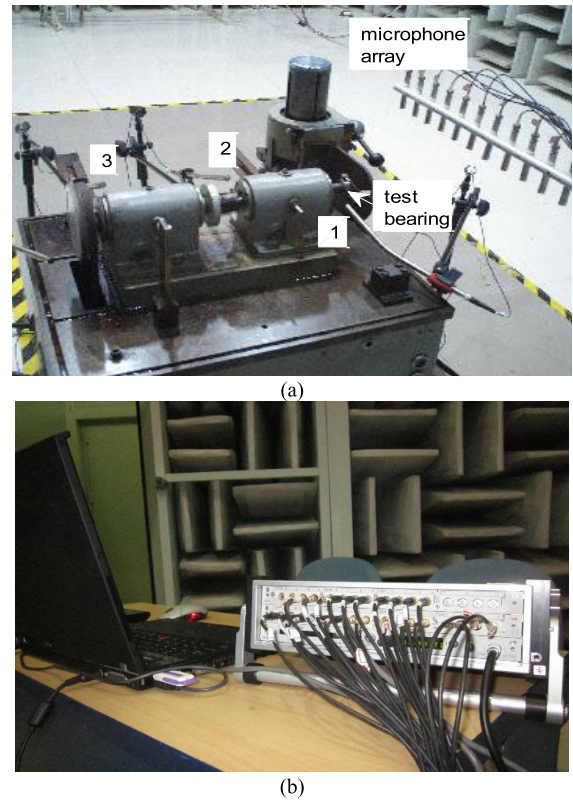


FIGURE 4. Experimental setup of bearing fault diagnosis (a) bearing test rig, and (b) data acquisition system.

TABLE 2. Parameters of GB6203 rolling bearings.

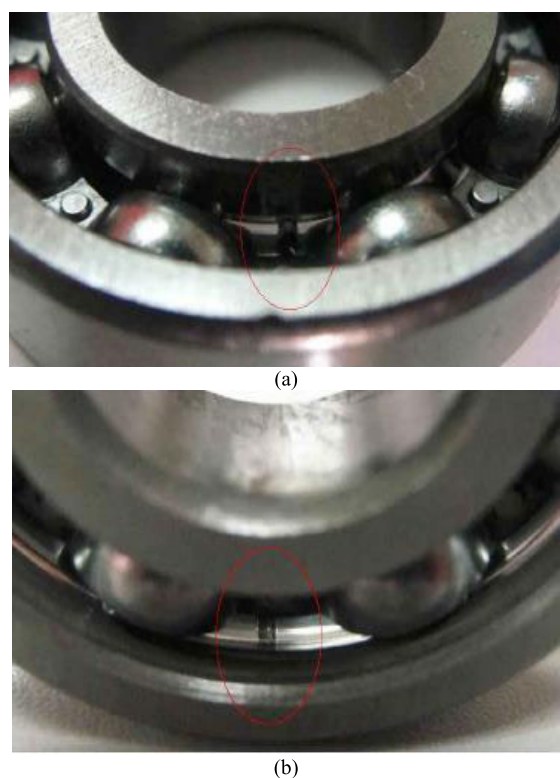
Type	Ball number	Ball diameter (mm)	Pitch diameter (mm)	Contact angle(rad)
GB6203	8	6.747	28.5	0

and rolling balls of bearings by electrical discharge machining technique. In order to simulate different fault severities, for ORF and IRF, there are three different damage sizes, which are 3.5×0.5×0.5mm, 3.5×1.0×0.5mm, 3.5×1.5×0.5 mm. For RBF, the damage size is 3.5×0.5×0.5mm. To sum up, there are totally eight health conditions in this study, including seven faulty conditions and one normal condition (NC). The health conditions of the rolling bearings and their labels are summarized in Table 3, where three different damage sizes are abbreviated as S, M, and L, respectively. Test bearings with outer race fault of small size (abbreviated as ORF-S) and inner race fault of small size (IRF-S) are shown in Fig. 5.

Acoustic pressures of rolling bearings under eight conditions are recorded by the data acquisition system. The data sampling frequency is 4096Hz. For every health condition, the acoustic signals are gathered under five different loads, which introduces variability to the dataset. Hence, 8 conditions × 5 loads = 40 recordings are acquired. There are 45 samples for every health condition under each load, with the data length of 16384 points. So, there are 1800 samples in the whole dataset.

TABLE 3. Eight health conditions of the rolling bearings.

Fault type	Damage size	Condition label
IRF	S	1
IRF	M	2
IRF	L	3
RBF	S	4
ORF	S	5
ORF	M	6
ORF	L	7
NC	/	8

**FIGURE 5.** Test bearings in condition (a) IRF-S, and (b) ORF-S.

B. ACOUSTIC IMAGING

For every sample in the dataset, acoustic imaging is conducted with WSM through sound field reconstruction. The reconstruction surface is chosen at the upper surface of the bearing test bench, which is parallel and identical to the measurement surface. The distance between the reconstruction surface and the measurement surface is 5cm. Just like the measurement hologram, the reconstruction surface is divided into a rectangular grid of 18×12 reconstruction points. According to the guidelines in reference [39], 18×12 virtual monopoles serving as equivalent sources are collocated uniformly on a fictitious planar surface, which is also parallel and

identical to the reconstruction surface with a 0.05m retreat distance below the reconstruction surface.

In the experiment, the rotating frequency of the shaft f_r is about 12Hz. According to Equations (9)-(11), the characteristic frequencies of ORF, IRF and RBF can be determined. Here, $f_o=36.6\text{Hz}$, $f_b=47.8\text{Hz}$, $f_i=59.4\text{Hz}$. As explained above, these three characteristic frequencies are chosen as the reconstruction frequencies in WSM. In practical application, the accuracy of the characteristic frequencies is probably not exactly in accordance with the resolution of the frequency analysis. To deal with this problem, in this experiment, three frequency components around each characteristic frequency are used as the reconstruction frequencies, and their reconstructed acoustic pressures are summed up to represent the reconstruction result at this characteristic frequency. As a result, for every acoustic signal sample of the test bearings, acoustic pressures on the reconstruction surface can be reconstructed at f_o , f_i , f_b with WSM, and three acoustic images can be obtained consequently. In total, 5400 acoustic images of rolling bearings under eight health conditions are finally obtained. Some of the acoustic images reconstructed at three characteristic frequencies under different bearing conditions are displayed in Fig. 6, where the sound pressure levels in the colormaps are limited from 10dB to 60dB, and the color bars are omitted for convenience.

In Fig. 6, in order to demonstrate the sound source localization results more intuitively, the reconstructed acoustic images are overlapped with the sketch map of the bearing test rig (white lines) from the top view. Firstly, it can be seen that the hot spots in the acoustic images are almost at the position (0.275m, 0.125m), which is coincident with the location of the test bearings. Besides that, the sound pressure levels around the coordinate (0.275m, 0.225m) are also high. That is mainly because the radiated sound waves are superposed with the reflected sound waves in this region. On the whole, the test bearings can be identified correctly in the acoustic images.

C. BEARING FAULT DIAGNOSIS USING CNN

As shown in Fig. 6, there are some differences in the reconstructed acoustic images at different frequencies under different bearing conditions. For example, the sound pressure levels at the hot spot region, textures in the acoustic images which can reflect the spatial distribution of acoustic pressures. However, it is still impossible to identify the bearing conditions directly from the acoustic images. Various image processing methods can indeed be utilized to extract image features for bearing fault diagnosis. However, it is still a hard task to choose an optimal image processing method because it is difficult to find which features are effective for bearing fault diagnosis. In order to overcome this difficulty, by virtue of its advantages in image processing and recognition, CNN is adopted to realize bearing fault diagnosis though automatically feature learning and classification from the reconstructed acoustic images.

As stated above, there are 1800 samples for all the eight health conditions in the whole dataset. Then, 1500 samples

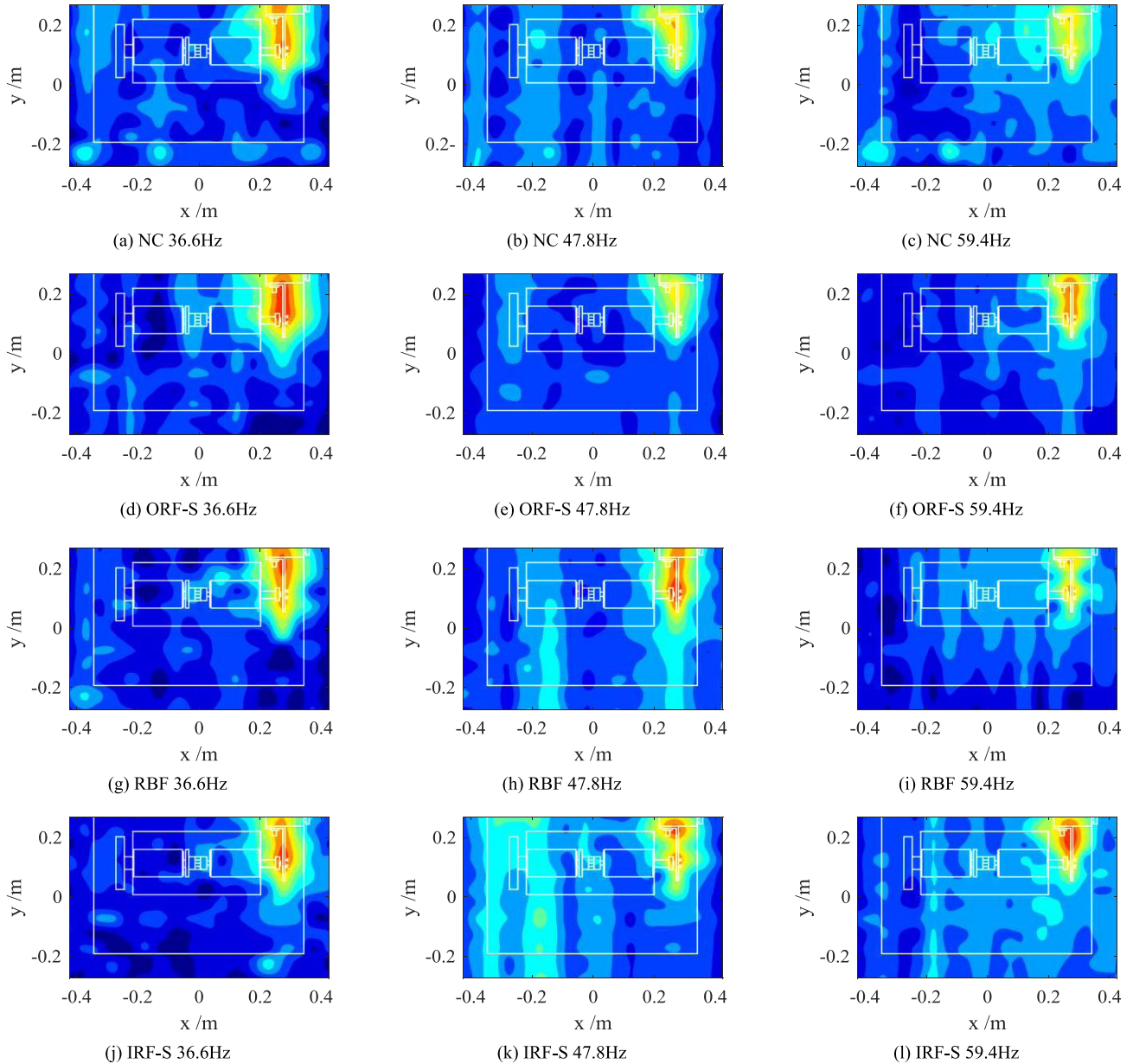


FIGURE 6. Reconstructed acoustic images at three characteristic frequencies of rolling bearings under different conditions: (a) NC at 36.6Hz, (b) NC at 47.8Hz, (c) NC at 59.4Hz, (d) ORF-S at 36.6Hz (e) ORF-S at 47.8Hz, (f) ORF-S at 59.4Hz, (g) RBF at 36.6Hz, (h) RBF at 47.8Hz, (i) RBF at 59.4Hz, (j) IRF-S at 36.6Hz, (k) IRF-S at 47.8Hz, (l) IRF-S at 59.4Hz.

are randomly selected from the whole dataset as the training dataset. Meanwhile, the number of samples for each health condition is almost the same. And the remaining 300 samples are used as testing dataset. After acoustic imaging, for every acoustic sample, three acoustic images are reconstructed at three characteristic frequencies 36.6 Hz, 47.8 Hz, and 59.4Hz. These three acoustic images for the designed CNN model are just corresponding to three RGB channels for the original LeNet-5 model. It is obvious that this data format can meet the input format requirement of the proposed CNN model.

In the CNN training stage, the preprocessed acoustic images in the training dataset are fed into the proposed

CNN model, whose parameters on each layer are presented in Table 1. Specifically, in this experiment, the number of outputs in the last layer is eight, which is equal to the number of bearing health conditions. Adam gradient optimization algorithm is adopted to adjust the weights of the network, and its learning rate is 3×10^{-4} . The total number of epochs is set as 1000, and the mini-batch size is 150. The dropout rate is chosen as 0.5, which is a commonly-used value. Training and testing losses in every epoch are recorded. Fig. 7 shows the training and testing loss curves. It can be seen that as the increase of the number of epochs, the training and testing loss curves decrease gradually and remains stable when the epoch number reaches to a certain value. Both of the loss curves are

TABLE 4. Testing accuracies of 10 trials using the proposed ABD method.

Trial	1	2	3	4	5	6	7	8	9	10	Average	Std
Accuracy (%)	99.67	99.33	99	99.33	99.67	98.67	99.67	98.33	98.67	99	99.13	0.0048

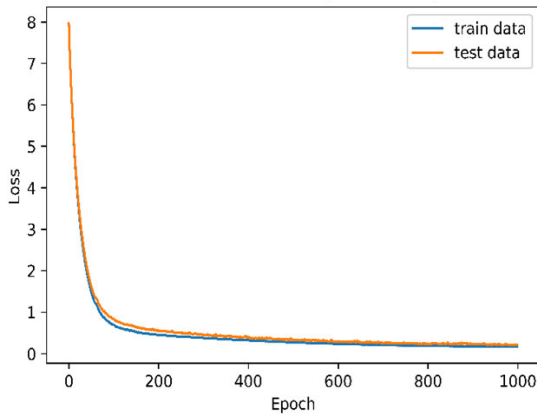


FIGURE 7. Loss curves of proposed CNN.

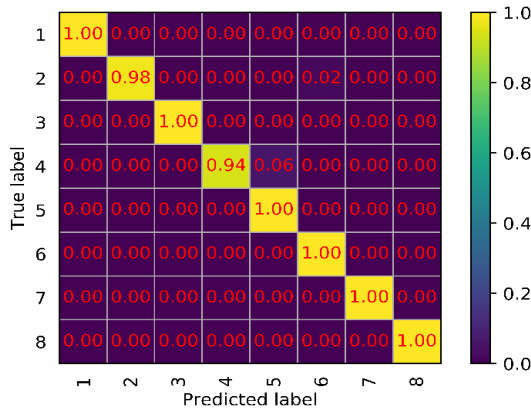


FIGURE 8. The confusion matrix of the diagnosis result.

smooth and convergent. Moreover, the differences between the training and testing errors are relatively small, which indicates that the trained CNN model has good generalization performance.

In the test stage, the preprocessed acoustic images in the test dataset are fed into the trained CNN model. It should be noted that 10 trials are carried out to reduce the influence of randomness. The testing accuracies of these 10 trials are presented in Table 4, where ‘Std’ denotes the standard deviation. It can be seen that the proposed ABD method achieves good diagnosis results in all the 10 trials. The maximum accuracy is 99.67%, and the minimum accuracy is 98.33%. Overall, the proposed ABD method obtains a mean accuracy of 99.13% with a 0.0048 standard deviation. The diagnosis accuracy is high and stable.

To show the diagnosis results of each health condition in detail, the confusion matrix of testing accuracies in the first trial is presented in Fig. 8. The corresponding health conditions of the labels can be found in Table 3. In Fig. 8, it can be seen that just a few samples in condition 2 (IRF-M)

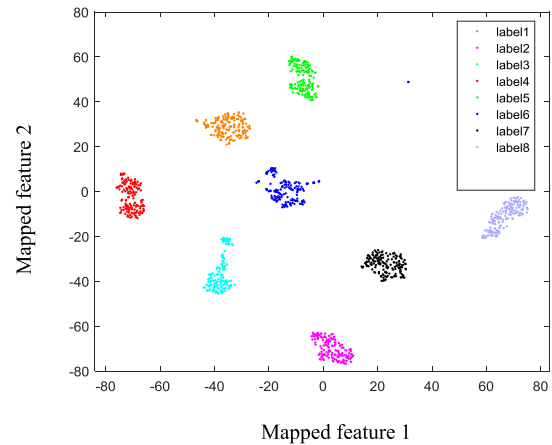


FIGURE 9. The visualization result of learned features via t-SNE.

and condition 4 (RBF) are misclassified, and most of the test samples are classified correctly. These results indicate that the adopted CNN model can learn fault features from the acoustic images of different bearings adaptively and detect the bearing health conditions effectively.

In order to further verify the feature learning ability of the proposed ABD method, feature visualization analysis is performed by t-distributed Stochastic Neighbor Embedding (t-SNE) [40]. High-level learned features of the fully connected layer FC5 are mapped into two-dimensional features. Mapped features of the testing samples are shown in Fig. 9, where the legends label 1-8 are in line with eight bearing health conditions in Table 3. It can be seen that features under the same bearing condition gather closely, and features under different bearing health conditions separate well in Fig.9. This result shows that the designed CNN model can learn discriminative features from the reconstructed bearing acoustic images.

D. COMPARISON WITH OTHER STATE-OF-THE-ART ABD METHODS

In order to verify the superiority of the proposed method based on acoustic imaging and CNN, the proposed ABD method is compared with some existing ABD methods. As reviewed briefly in section 1, there are generally three kinds of ABD methods. The first kind is the so-called traditional ABD methods, which are based on manual feature extraction from single or multi-channel acoustic signals and conventional shallow machine learning algorithms. The second kind is the end-to-end ABD method, which feeds single or multi-channel acoustic signals into a deep learning model for machinery fault diagnosis without feature engineering [30]. The third kind of ABD methods are based on acoustic imaging, image feature extraction, and shallow machine learning [12], [15]. Some traditional ABD methods have been proven to be vulnerable to noise interference

TABLE 5. Layer configurations of the 1-D CNN model.

Layer name	Parameter name	Parameter size
Conv1	Kernels	1 x 32 x 1
Pool2	Max Pooling size	2 x 2
Conv3	Kernels	1 x 16 x 32
Pool4	Max Pooling size	2 x 2
Conv5	Kernels	1 x 4 x 64
Pool6	Max Pooling size	2 x 2
Conv7	Kernels	1 x 4 x 64
Pool8	Max Pooling size	1 x 2

and have low diagnosis accuracy in previous studies [12], [30]. In this paper, some representative methods in the state-of-the-art ABD methods are performed for comparison. The prediction accuracy is used as the main measure term for the comparison. These methods and their diagnosis results are detailed as follows.

1) END-TO END ABD METHOD

According to the end-to-end ABD method [30], in this experiment, acoustic signals of the three reference microphones are respectively fed into a 1-D CNN model for bearing fault diagnosis. The 1-D CNN model consists of four alternating convolutional and pooling layers, one fully connected layer, and the last softmax output layer. The layer configurations of the 1-D CNN model are listed in Table 5. The output number in the softmax classifier is also eight.

Acoustic signals of each reference microphone are segmented into 15000 samples with a data length of 1024 points. Then, 10000 samples are randomly selected as the training data. And the rest of 5000 samples construct the test dataset. The training settings are similar to the CNN model used in the proposed method. Adam optimization algorithm is utilized, and the learning rate is 3×10^{-4} . The dropout ratio is 0.5, and the mini-batch size is 100. In order to provide a reliable comparison, other parameters are also adjusted to be optimal values. Acoustic samples of three reference microphones are fed into the CNN model respectively to train the model. The training and testing loss curves using acoustic signals of reference mic. 1 are plotted in Fig. 10. It can be seen that the training tends to converge around 30 epochs.

After 100 epochs' training, the testing accuracies of the end-to-end ABD method using three different reference microphones are obtained. Similarly, 10 trials are conducted for acoustic signals of each microphone to reduce randomness. Testing accuracies of three different reference microphones in 10 trials, as well as the average accuracies and standard deviations, are shown in Fig. 11.

It can be seen that the end-to-end ABD method achieves the best diagnosis average accuracy 97.73% at the nearest reference microphone mic. 1, but it's still lower than the mean accuracy 99.13% of the proposed ABD method. And the standard deviation of mic. 1 is the lowest one. However, the diagnosis accuracy decreases rapidly as the microphone

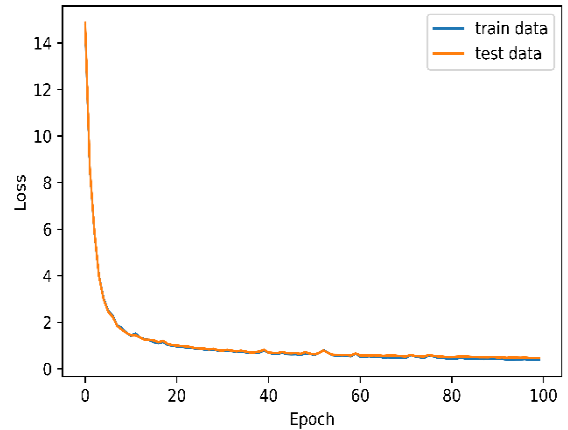


FIGURE 10. Loss curves of 1-D CNN in end-to-end ABD method.

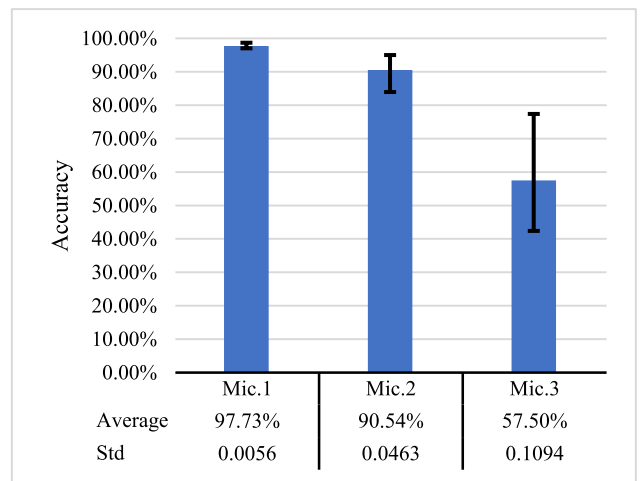


FIGURE 11. Testing accuracies of three reference microphones.

becomes far away from the tested bearing, and the standard deviation increases from mic.1 to mic.3. The diagnosis accuracy is very low at mic. 3, and the standard deviation of diagnosis accuracies in 10 trials is high, which indicates that the diagnosis results are not stable. Meanwhile, it costs a long time to train the CNN model using acoustic signals of mic.3, because it is difficult for the model to be convergent. So, it is obvious that the performance of the end-to-end ABD method is sensitive to the measurement positions.

To further explain this phenomenon, acoustic signals recorded by the three reference microphones under bearing condition 1 (IRF-S) are shown in Fig. 12.

It can be seen that, there are obvious periodic impacts in Fig. 12(a), which can reflect fault information. By contrast, impacts become weaker in Fig.12 (b) and even cannot be identified in Fig. 12 (c). In other words, the signal to noise ratio (SNR) of mic. 2 and mic. 3 are much lower than that of mic. 1. That is why the diagnosis accuracy varies at different measurement positions. However, in practical application, due to the complicated structure of modern machinery, it is difficult to choose proper measurement positions without prior knowledge.

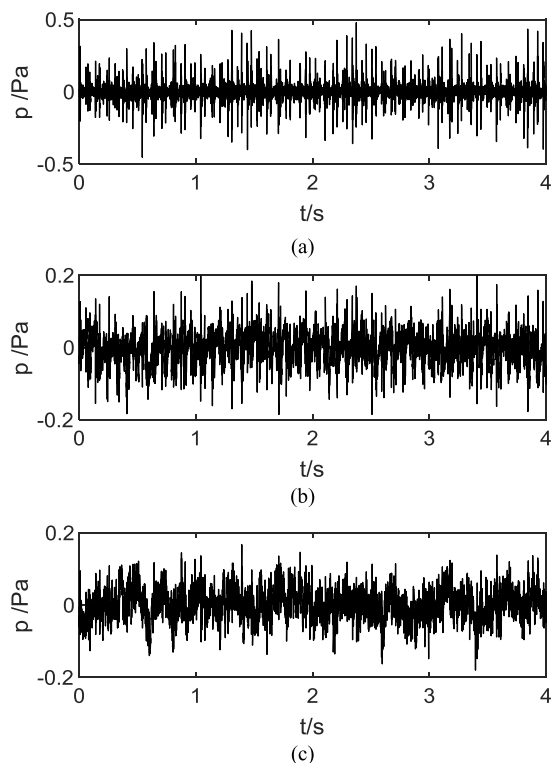


FIGURE 12. Acoustic signals of different reference microphones: (a) mic. 1, (b) mic. 2, (c) mic. 3.

In conclusion, even though the end-to-end ABD method can take advantage of deep learning to overcome the problem of feature engineering in traditional ABD methods, it is still based on single-channel acoustic measurement. As a result, it cannot overcome the difficulty of choosing proper measurement positions without prior knowledge. Moreover, it cannot identify the sound source locations, which are important for machinery fault diagnosis and further noise control.

2) ABD METHOD BASED ON ACOUSTIC IMAGING AND IMAGE FEATURE EXTRACTION

In this section, two different state-of-the-art ABD methods based on acoustic imaging and image feature extraction are performed for bearing fault diagnosis. In order to compare with the proposed ABD method based on WSM and CNN, the acoustic imaging tools adopted in these two methods are also WSM. The only difference between these two methods lies in the image feature extraction method. The first method is designed according to Lu *et al.* [12], where gray-level co-occurrence matrix (GLCM) based texture features from four directions (0°, 45°, 90°, 135°) are extracted from the reconstructed acoustic images. And the other feature extraction method is designed according to Wang *et al.* [15], where Gabor wavelet transform (GWT) is adopted. After feature extraction, SVM is also adopted for bearing fault diagnosis. For convenience, these two methods can be briefly abbreviated as WSM+GLCM+SVM, and WSM+GWT+SVM.

Since WSM is employed for acoustic imaging, the acoustic images used for fault diagnosis are the same as those

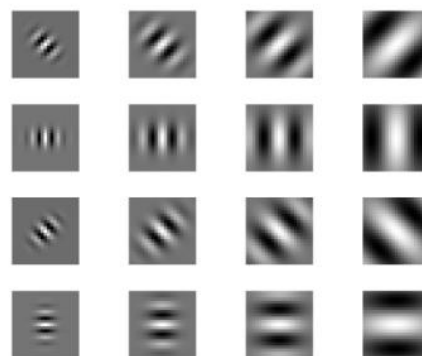


FIGURE 13. Real part of the Gabor-wavelet filters in the spatial domain.

used in the proposed method, as shown in Fig. 6. Then, in WSM+GLCM+SVM method, the quantized gray level L is set as 32 according to reference [12]. While calculating GLCMs in four directions, the inter-pixel distance is chosen as 1. For every acoustic image, 12 statistical features are extracted from each GLCM in one direction, such as angular second moment, contrast, correlation, entropy, and so on. Detailed feature extraction procedures of GLCM can be seen in reference [12]. Since the feature dimensions extracted by GLCM is very high (12 statistical features \times 4 directions \times 3 characteristic frequencies = 144), principal component analysis (PCA) is conducted. The cumulative explained variance ratio is chosen to be 0.95. After PCA, the dimension of GLCM features is reduced from 144 to 20.

In the WSM+GWT+SVM method, parameters of Gabor filters are designed according to reference [15]. While designing the Gabor filters, the number of scales and orientations of the Gabor wavelet function are both set as 4. As a result, 16 Gabor filters are generated, whose real parts in the spatial domain are shown in Fig.13, where the x and y axes are both pixels.

Then, all the acoustic images are convolved with these 16 designed Gabor filters, generating 16 Gabor images at different scales and directions for every acoustic image. Then, 32 mean and standard deviation features can be extracted from every acoustic image. For every acoustic signal sample, 96 Gabor features at three characteristic frequencies are extracted and then reduced to 24 by PCA. The division of the training dataset and the test dataset is the same as the proposed method. Using the trained SVM models, the diagnosis accuracies of every bearing condition, and the overall accuracy of the whole dataset obtained by these two methods are listed in Table 6. Furthermore, 10 trials are conducted to reduce the effects of randomness. Overall diagnosis accuracies in 10 trials using WSM+GLCM+SVM and WSM+GWT+SVM are shown in Fig. 14.

In Fig. 14, the average accuracy of WSM+GLCM+SVM is 65.45%, and the standard deviation is 0.028. The average accuracy of WSM+GWT+SVM is 96.26%, and the standard deviation is 0.01. It can be seen in Table 6 and Fig.14 that the fault diagnosis accuracies of WSM+GLCM+SVM are

TABLE 6. Testing accuracies (%) of two different acoustic imaging-based fault diagnosis methods.

ABD method	IRF-S	IRF-M	IRF-L	RBF	ORF-S	ORF-M	ORF-L	NC	Overall accuracy
WSM+GLCM+SVM	51.35	75	73.33	69.05	66.67	66.67	64.1	92.86	70
WSM+GWT+SVM	97.22	91.67	88.89	100	100	95	97.14	97.37	96

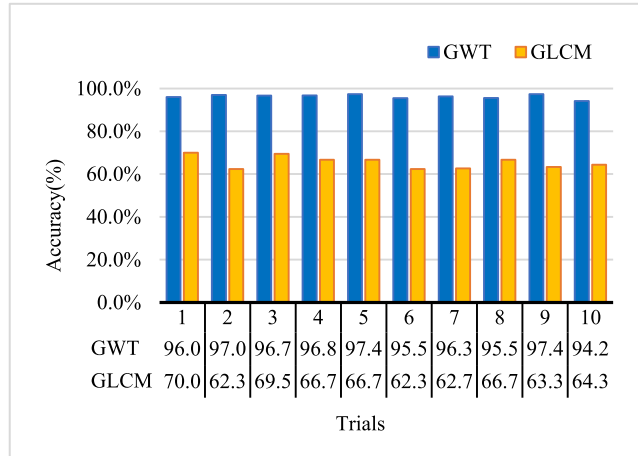


FIGURE 14. Diagnosis accuracies in 10 trials using WSM+GLCM+SVM and WSM+GWT+SVM.

much lower than the accuracies of WSM+GWT+SVM, and only the testing accuracy of the normal condition (NC) is acceptable. This means that the GLCM features of the acoustic images are not effective for bearing fault diagnosis, while Gabor features obtained through GWT are more effective. It should be mentioned that the diagnosis results of the WSM+GLCM+SVM method are worse than those in reference [12]. That is mainly because in reference [12], there are only four bearing conditions with constant load and the same defect size. However, in this article, both the fault type and fault severity are considered. There are eight bearing health conditions with different defect sizes, and there are five different loads under each health condition. So, the ABD method needs to have better generalization performance to tackle this task.

Compared with Fig. 14, the average diagnosis accuracy of the proposed method (99.13%) in Table 4 is still higher than WSM+GWT+SVM and WSM+GLCM+SVM, and the standard deviation of the proposed method is also the lowest one. This comparison shows that although the state-of-the-art methods based on acoustic imaging and image feature extraction can overcome the difficulty of measurement position selection and realize sound source identification, their diagnosis results rely seriously on the image processing techniques. In comparison, the proposed ABD method based on acoustic imaging and CNN can automatically learn effective features from the acoustic images and avoid the influences of image feature engineering on diagnosis results. Meanwhile, better diagnosis results can be obtained with the proposed method.

In conclusion, compared with the above three state-of-the-art ABD methods, by virtue of acoustic imaging

and deep learning, the proposed ABD method can mitigate the difficulty of measurement position selection, identify the sound source location, eliminate the influences of manual feature extraction, and obtain higher diagnosis accuracies in bearing fault diagnosis.

VI. CONCLUSION

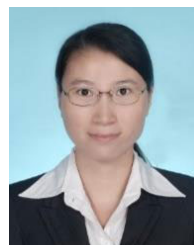
ABD method can overcome the limitation of VBD methods by virtue of its non-contact measurement. In this paper, a non-contact ABD method using acoustic imaging and CNN is proposed for rolling bearing fault diagnosis. Based on acoustic measurement by a microphone array, acoustic images of bearings under different conditions can be reconstructed with WSM. Acoustic imaging can add a new spatial dimension in the acoustic data representation for fault diagnosis. Moreover, sound source identification can be easily realized using acoustic images. Then, a CNN model is designed and applied to learn features automatically from the acoustic images and diagnose the bearing condition. Experimental results demonstrate that the proposed method can detect various bearing conditions with different damage sizes effectively, and obtain a mean accuracy of 99.13%. Comparisons with peer state-of-the-art ABD methods further verify the superiority of the proposed method. It can eliminate the drawbacks of the existing ABD methods, such as the difficulty of measurement position selection, the dependence on feature extraction methods, and so on.

In conclusion, the diagnosis result of the proposed method is more accurate and reliable. It may promote the development of bearing ABD, and has the potential to be extended to the fault diagnosis of other rotating machinery. Nevertheless, it also has some disadvantages, i.e., the need for expensive microphone array and relatively long computation time. In the future, the rapid development of MEMS microphone array, embedded acquisition devices, edge computing techniques, and high-speed computing devices will significantly facilitate the application of this ABD method.

REFERENCES

- [1] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics-A tutorial," *Mech. Syst. Signal Process.*, vol. 25, no. 2, pp. 485–520, 2011.
- [2] J. Antoni and R. B. Randall, "The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines," *Mech. Syst. Signal Process.*, vol. 20, no. 2, pp. 308–331, Feb. 2006.
- [3] Z. K. Peng, P. W. Tse, and F. L. Chu, "A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing," *Mech. Syst. Signal Process.*, vol. 19, no. 5, pp. 974–988, 2005.
- [4] H. Li, T. Liu, X. Wu, and Q. Chen, "Application of EEMD and improved frequency band entropy in bearing fault feature extraction," *ISA Trans.*, vol. 88, pp. 170–185, May 2019.

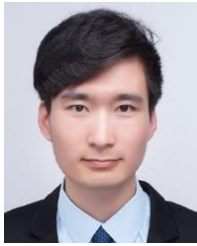
- [5] D. Wang, X. Zhao, L.-L. Kou, Y. Qin, Y. Zhao, and K.-L. Tsui, "A simple and fast guideline for generating enhanced/squared envelope spectra from spectral coherence for bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 122, pp. 754–768, May 2019.
- [6] D. Wang, Z. Peng, and L. Xi, "Theoretical and experimental investigations on spectral Lp/Lq norm ratio and spectral gini index for rotating machine health monitoring," *IEEE Trans. Autom. Sci. Eng.*, early access, May 29, 2020, doi: 10.1109/TASE.2020.2994741.
- [7] H. Jiang, J. Chen, and G. Dong, "Hidden Markov model and nuisance attribute projection based bearing performance degradation assessment," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 184–205, May 2016.
- [8] H. Jiang, J. Yuan, Q. Zhao, H. Yan, S. Wang, and Y. Shao, "A robust performance degradation modeling approach based on student's t-HMM and nuisance attribute projection," *IEEE Access*, vol. 8, pp. 49629–49644, 2020.
- [9] N. Baydar and A. Ball, "A comparative study of acoustic and vibration signals in detection of gear failures using Wigner–Ville distribution," *Mech. Syst. Signal Process.*, vol. 15, no. 6, pp. 1091–1107, Nov. 2001.
- [10] N. Baydar and A. Ball, "Detection of gear failures via vibration and acoustic signals using wavelet transform," *Mech. Syst. Signal Process.*, vol. 17, no. 4, pp. 787–804, Jul. 2003.
- [11] S. K. Yadav, K. Tyagi, B. Shah, and P. K. Kalra, "Audio signature-based condition monitoring of internal combustion engine using FFT and correlation approach," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 4, pp. 1217–1226, Apr. 2011.
- [12] W. Lu, W. Jiang, H. Wu, and J. Hou, "A fault diagnosis scheme of rolling element bearing based on near-field acoustic holography and gray level co-occurrence matrix," *J. Sound Vib.*, vol. 331, no. 15, pp. 3663–3674, Jul. 2012.
- [13] W. Lu, W. Jiang, G. Yuan, and L. Yan, "A gearbox fault diagnosis scheme based on near-field acoustic holography and spatial distribution features of sound field," *J. Sound Vib.*, vol. 332, no. 10, pp. 2593–2610, May 2013.
- [14] E. Cardenas Cabada, Q. Leclere, J. Antoni, and N. Hamzaoui, "Fault detection in rotating machines with beamforming: Spatial visualization of diagnosis features," *Mech. Syst. Signal Process.*, vol. 97, pp. 33–43, Dec. 2017.
- [15] R. Wang, C. Wang, J. Li, and W. Lu, "An intelligent fault diagnosis method of rolling element bearing based on acoustic imaging and Gabor wavelet transform," in *Proc. 25th Int. Congr. Sound Vib.*, vol. 5, Hiroshima, Japan, Jul. 2018, pp. 2684–2691.
- [16] A. Darraz, J. Antoni, P. Mollon, G. Kirie, J.-H. Thomas, and A. Larcher, "Rotating machine diagnosis using acoustic imaging and artificial intelligence," presented at the *Surveill. Vishno AVE Conf.*, Lyon, France, 2019.
- [17] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.
- [18] H. Zhiyi, S. Haidong, J. Lin, C. Junsheng, and Y. Yu, "Transfer fault diagnosis of bearing installed in different machines using enhanced deep auto-encoder," *Measurement*, vol. 152, Feb. 2020, Art. no. 107393.
- [19] Z. He, H. Shao, X. Zhang, J. Cheng, and Y. Yang, "Improved deep transfer auto-encoder for fault diagnosis of gearbox under variable working conditions with small training samples," *IEEE Access*, vol. 7, pp. 115368–115377, 2019.
- [20] S. Tang, C. Shen, D. Wang, S. Li, W. Huang, and Z. Zhu, "Adaptive deep feature learning network with nesterov momentum and its application to rotating machinery fault diagnosis," *Neurocomputing*, vol. 305, pp. 1–14, Aug. 2018.
- [21] R. Huang, Y. Liao, S. Zhang, and W. Li, "Deep decoupling convolutional neural network for intelligent compound fault diagnosis," *IEEE Access*, vol. 7, pp. 1848–1858, 2019.
- [22] S. Wang and J. Xiang, "A minimum entropy deconvolution-enhanced convolutional neural networks for fault diagnosis of axial piston pumps," *Soft Comput.*, vol. 24, no. 4, pp. 2983–2997, Feb. 2020.
- [23] S. Wang, J. Xiang, Y. Zhong, and Y. Zhou, "Convolutional neural network-based hidden Markov models for rolling element bearing fault identification," *Knowl.-Based Syst.*, vol. 144, pp. 65–76, Mar. 2018.
- [24] O. Janssens, R. Van de Walle, M. Loccufier, and S. Van Hoecke, "Deep learning for infrared thermal image based machine health monitoring," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 151–159, Feb. 2018.
- [25] X. Ding and Q. He, "Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 8, pp. 1926–1935, Aug. 2017.
- [26] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.
- [27] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [28] Z. Chen, A. Mauricio, W. Li, and K. Gryllias, "A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks," *Mech. Syst. Signal Process.*, vol. 140, Jun. 2020, Art. no. 106683.
- [29] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez, "Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals," *Mech. Syst. Signal Process.*, vols. 76–77, pp. 283–293, Aug. 2016.
- [30] Y. Yao, H. Wang, S. Li, Z. Liu, G. Gui, Y. Dan, and J. Hu, "End-to-end convolutional neural network model for gear fault diagnosis based on sound signals," (in English), *Appl. Sci.*, vol. 8, no. 9, p. 1584, Sep. 2018.
- [31] A. Kumar, C. P. Gandhi, Y. Q. Zhou, R. Kumar, and J. W. Xiang, "Improved deep convolution neural network (CNN) for the identification of defects in the centrifugal pump using acoustic images," (in English), *Appl. Acoust.*, vol. 167, Oct. 2020, Art. no. 107399.
- [32] A. Sarkissian, "Method of superposition applied to patch near-field acoustic holography," *J. Acoust. Soc. Amer.*, vol. 118, no. 2, pp. 671–678, Aug. 2005.
- [33] R. Wang, J. Chen, and G. Dong, "Data interpolation method based on the wave superposition algorithm," *J. Vib. Control*, vol. 20, no. 3, pp. 421–435, Feb. 2014.
- [34] L. Song, G. H. Koopmann, and J. B. Fahline, "Numerical errors associated with the method of superposition for computing acoustic fields," *J. Acoust. Soc. Amer.*, vol. 89, no. 6, pp. 2625–2633, Jun. 1991.
- [35] R. Jeans and I. C. Mathews, "The wave superposition method as a robust technique for computing acoustic fields," *J. Acoust. Soc. Amer.*, vol. 92, no. 2, pp. 1156–1166, Aug. 1992.
- [36] G. H. Koopmann, L. Song, and J. B. Fahline, "A method for computing acoustic fields based on the principle of wave superposition," *J. Acoust. Soc. Amer.*, vol. 86, no. 6, pp. 2433–2438, Dec. 1989.
- [37] E. G. Williams, "Regularization methods for near-field acoustical holography," *J. Acoust. Soc. Amer.*, vol. 110, no. 4, pp. 1976–1988, Oct. 2001.
- [38] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter" *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.
- [39] M. R. Bai, C.-C. Chen, and J.-H. Lin, "On optimal retreat distance for the equivalent source method-based nearfield acoustical holography," *J. Acoust. Soc. Amer.*, vol. 129, no. 3, pp. 1407–1416, Mar. 2011.
- [40] L. van der Maaten and G. Hinton, "Visualizing high-dimensional data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 2, pp. 2579–2605, Nov. 2008.



RAN WANG received the B.E. degree in vehicle information engineering from Southwest Jiaotong University, in 2008, and the Ph.D. degree in mechanical engineering from Shanghai Jiao Tong University, in 2015. She is currently a Lecturer with the School of Logistics Engineering, Shanghai Maritime University, China. Her current research interests include machinery fault diagnosis, performance degradation assessment, noise source identification, and deep learning.



FENGKAI LIU received the B.E. degree from Shandong Agricultural University, in 2017. He is currently pursuing the M.E. degree with the School of Logistics Engineering, Shanghai Maritime University, China. His current research interests include machinery fault diagnosis and deep learning.



FATAO HOU received the B.E. degree in mechanical design, manufacturing and automation from Shandong University, Jinan, China, in 2015. He is currently pursuing the Ph.D. degree in mechanical engineering with Shanghai Jiao Tong University, Shanghai, China. His current research interests include rotating machinery fault diagnosis and digital signal processing.



QILIN HOU received the B.E. degree from Shanghai Polytechnic University, in 2018. He is currently pursuing the M.E. degree with the School of Logistics Engineering, Shanghai Maritime University, China. His current research interests include machinery fault diagnosis, lithium battery remaining useful life prediction, and deep learning.



WEIKANG JIANG received the B.E. and Ph.D. degrees from Shanghai Jiao Tong University (SJTU), in 1982 and 1990, respectively. He is currently a Professor with the School of Mechanical Engineering, SJTU, where he is also the Director of the Institute of Vibration, Shock and Noise and a Deputy Director with the State Key Laboratory of Mechanical Systems and Vibration. His research interests include near field acoustic holography, beamforming, noise and vibration control, NVH of various vehicles, and environmental noise. He is a Councillor with the Chinese Society of Acoustics and a Vice-Director with the Committee of Environmental Acoustics of CSA.



LONGJING YU (Graduate Student Member, IEEE) received the B.E. degree in vehicle engineering from Hunan University, in 2019. He is currently pursuing the M.E. degree with the School of Logistics Engineering, Shanghai Maritime University, China. His current research interests include machinery fault diagnosis and acoustic signal processing.

...