Multibranch and Multiscale CNN for Fault Diagnosis of Wheelset Bearings Under Strong Noise and Variable Load Condition

Dandan Peng, Huan Wang, Zhiliang Liu, *Member, IEEE*, Wei Zhang, Ming J. Zuo, *Senior Member, IEEE*, and Jian Chen

***Abstract*—The critical issue for fault diagnosis of wheelset bearings in high-speed trains is to extract fault features from vibration signals. To handle high complexity, strong coupling, and low signal-to-noise ratio of the vibration signals, this article proposes a novel multibranch and multiscale convolutional neural network that can automatically learn and fuse abundant and complementary fault information from the multiple signal components and time scales of the vibration signals. The proposed method combines the conventional filtering methods and the idea of the multiscale learning, which can extend the breadth and depth of the feature learning process. Consequently, the proposed network can perform better. The experimental results on the wheelset bearing dataset demonstrate that the proposed method has better antinoise ability and load domain adaptability and can diagnose 12 fault types more accurately when compared with the five state-of-the-art networks.**

***Index Terms*—Convolutional neural network (CNN), fault diagnosis, multibranch learning, multiscale learning, wheelset bearing.**

I. Introduction

WHEELSET bearings are core parts of high-speed train (HST) bogies. Their health status has a significant impact on the HST safety. Because the wheelset bearings work for a long time under the harsh conditions of strong impact, high speed, and large workload, they are vulnerable to suffering various faults, which may lead to catastrophic accidents [1]. Therefore, it is critically important to accurately diagnose various faults of the wheelset bearings under strong noise and variable load condition.

Traditional intelligent fault diagnosis methods need to manually extract features from vibration signals [2], such as local mean decomposition [3], empirical mode decomposition [4], Hilbert–Huang transform [5], wavelet transform [6], etc. The manually extracted features are fed into machine learning methods to obtain diagnosis results. Common examples of these machine learning methods include k-nearest neighbor [7], random forest [8], support vector machine [9], etc. However, the conventional intelligent fault diagnosis methods have the following disadvantages.

1) Manual feature extraction cannot fully generate the complex dynamic characteristics of wheelset bearings, as the vibration signals of wheelset bearings are highly complicated due to the influence of the uneven rail surface and other excitation parts of bogies.
2) Manual feature extraction is hard to truly extract the intrinsic fault features, this is because the fault features are completely submerged in strong noises and distributed in different feature intervals under variable load condition.
3) The abovementioned machine learning methods belong to shallow learning networks, and they are difficult to learn complex nonlinear relations and thus are prone to making misjudgment.

Overall, it is worth noting that the final diagnostic performance is highly related to the quality of feature extraction. Therefore, it is very important to propose an appropriate feature extraction method so that the wheelset bearing faults can be efficiently diagnosed.

Recently, many impressive achievements of implementing convolutional neural network (CNN) have been reported in the fields of computer vision [10], speech recognition [11], and natural language processing [12]. In particular, in the field of fault diagnosis of rotating machinery, vibration signals are often highly nonlinear and complex, which contain a lot of information irrelevant to fault diagnosis. Fortunately, the CNN is specifically designed for highly nonlinear and complex signals. Most

Manuscript received November 20, 2019; accepted January 7, 2020. Date of publication January 17, 2020; date of current version March 17, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61833002, in part by the State Key Laboratory of Traction Power, Southwest Jiaotong University under Grant TPL1608, and in part by the Natural Sciences and Engineering Research Council of Canada under Grant #RGPIN-2015-04897. Paper no. TII-19-5054. *(Dandan Peng and Huan Wang contributed equally to this work.) (Corresponding author: Zhiliang Liu.)*

D. Peng, H. Wang, and Z. Liu are with the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: dandanpeng2@gmail.com; wh.huanwang@gmail.com; zhiliang_liu@uestc.edu.cn).

W. Zhang is with the Glasgow College, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: 2016200106019@std.uestc.edu.cn).

M. J. Zuo is with the Department of Mechanical Engineering, University of Alberta, Edmonton AB T6G 1H9, Canada (e-mail: ming.zuo@ualberta.ca).

J. Chen is with the School of Engineering, University of Leicester, LE1 7RH Leicester, U.K. (e-mail: jc734@leicester.ac.uk).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2020.2967557

importantly, the convolution operation is the most special feature of the CNN. It behaves like a filter (e.g., lowpass, bandpass, or highpass) and can help the CNN extract sensitive frequency bands of bearing faults and to remove noise in the vibration signals. According to Chen *et al*. [13], most of the features extracted by convolutional kernels are medium-frequency features or low-frequency features, and these kernels only extract features in certain frequency band. Therefore, we think that the CNN is mostly suitable for rotating machinery fault diagnosis in a comparison with other machine learning methods. Because of its special characteristics, the CNN has been widely used in the fault diagnosis of gearboxes [14]–[17], bearings [13], [18]–[24], and other rotating machinery [25], [26]. The CNN-based methods can be generally categorized into two groups: one-dimensional CNN (1DCNN) and two-dimensional CNN (2DCNN). The inputs of 1DCNN mainly can be the raw signal [13], [15], [17], [20], [23]–[25], manually extracted statistical features [14], filtering signal [26], Fourier amplitude spectrum [19], etc. The inputs of 2DCNN mainly can be the direct reorganization of 1-D vibration signal into 2-D image [19], [21] and time–frequency spectrum [22]. The abovementioned studies prove that the CNN performed better than the conventional machine learning methods in rotating machinery fault diagnosis. As mentioned above, due to the complicated and time-varying working conditions of the HSTs, the vibration signal of the wheelset bearings is prone to being contaminated by noises, which makes the feature extraction extremely challenging. Therefore, it is difficult and is improper to apply the existing CNN-based methods to the fault diagnosis of wheelset bearings. Therefore, this article attempts to solve the difficulty from the following two aspects.

In the first aspect, the energy of the low-frequency impulse excitations generated by the faulty wheelset bearings is relatively low. Additionally, the noises from uneven rail surface and other parts of bogie seriously contaminate the vibration signals for fault diagnosis. Therefore, learning the fault features of the standard CNN based on vibration signals is extremely difficult. By combining conventional filtering technology with the CNN, a multibranch network that can learn abundant features from different signal components is proposed to solve this difficulty. In particular, the proposed multibranch network uses three parallel branches to learn feature information from the raw vibration signal, the signal component with high signal-to-noise ratio (SNR), and the low-frequency component. Then, multifeature fusion is used to adaptively fuse and optimize features extracted by the multibranch network. Because the extracted features from multiple branches are different, the CNN can make the best judgment by synthesizing the features of multiple degrees.

In the second aspect, because load changes dramatically, the vibration signals measured from the wheelset bearings is typically nonlinear with high complexity, coupling, and uncertainty. Moreover, there may be a variety of fault types and fault severity levels for the wheelset bearings, and the feature frequencies caused by these faults are varying a lot, resulting in the distribution of fault features in different scales. Therefore, the vibration signals exhibit multiscale properties and contain complicated feature information on multiple time scales. To address this issue, this article introduces the multiscale learning into the CNN. The multiscale learning is implemented by multiscale convolution module. Multiple convolutional layers with different convolution kernels are used to simultaneously learn the signal features at different time scales, and the serial feature fusion layers are used to fuse the multiscale features. Based on this module, the multiscale network is proposed to excavate deeper and richer multiscale features from the vibration signals and consequently enhance the CNN robustness on the fault feature learning.

Combining the abovementioned two innovative aspects, a novel multibranch and multiscale CNN (MBSCNN) is proposed to automatically learn and fuse the abundant and complementary features from multiple signal components and time scales of the vibration signals. Based on the real vibration datasets recorded from the wheelset bearing test rig, the experimental results show that the MBSCNN can accurately diagnose various fault types of the wheelset bearings, and it also have been proved that the MBSCNN can achieve better performance than the five state-of-the-art networks.

The contribution points of this article are summarized as follows.

1) This article combines the conventional filtering technology with the CNN and proposes a multibranch network to extract abundant features from multiple signal components.
2) This article uses the idea of the multiscale learning and develops a multiscale convolution module, which can learn both long-term and short-term features from vibration signals and, thus, improving the learning ability of the CNN on multiscale features.
3) A multifeature fusion layer is used to adaptively fuse and optimize the abundant features learnt by the multibranch network and the multiscale network.
4) Based on the abovementioned improvements, the MBSCNN-based framework is finally generated, and it can automatically learn and fuse the abundant and complementary features and consequently accurately diagnose various fault types of the wheelset bearings under strong noise and variable load condition.

The rest of this article is organized as follows. Section II elaborates in detail the proposed method. In Section III, the proposed method is evaluated by the real experimental wheelset bearing datasets. In Section IV, the necessity of the multibranch learning and the multiscale learning is verified and followed by two main discussions in Section V. Finally, Section VI concludes this article.

II. Proposed Method

This article aims to automatically learn and fuse abundant and complementary features from various components and various time scales of the vibration signal, so as to solve the fault diagnosis task of the wheelset bearings under strong noise and variable load condition. A robust intelligent fault diagnosis method should allow for in-depth and comprehensive learning of signal features, such as the ability to simultaneously learn

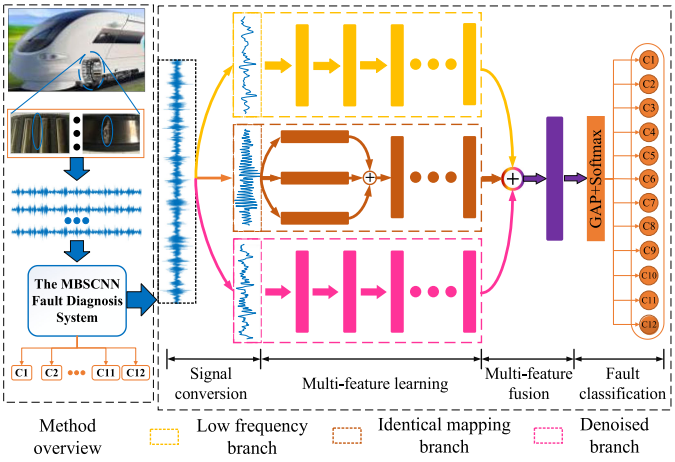


Fig. 1. MBSCNN network structure.

long-term and short-term features or to extract low-frequency signal features associated with faults. Therefore, this article constructs an end-to-end network, which integrates the signal conversion, the multifeature learning, the multifeature fusion, and the fault classification. The proposed method combines the conventional filtering technology and the multiscale learning, and its core contribution is to propose a deep and comprehensive feature extraction and feature fusion framework. The overall architecture of the MBSCNN of wheelset bearing fault diagnosis is shown in Fig. 1, among which, the MBSCNN is composed of four main parts—multibranch feature learning, multiscale feature learning, multifeature fusion, and fault classification.

A. Multibranch Learning

1) Identical Mapping Branch: The vibration signal contains the most abundant signal features, and it can reflect the health status of bearings in the most comprehensive way. In this way, the identical mapping branch (IMB) aims to learn the global fault feature information from the raw signals. The IMB is a wide spectral range analysis. Given a vibration signal $x = [x_1, x_2, \ldots, x_N]$, where N is the size of x, and the input of the IMB is x. Additionally, due to the inherent multiscale properties of vibration signal, this article constructs an architecture for the multiscale learning to obtain more comprehensive signal features. This will be introduced in detail in Section II-B.

2) Denoising Branch: The CNN has powerful feature learning ability which could learn the complex nonlinear relation. However, the weak fault features of wheelset bearings are often submerged in strong noise. Although the CNN extracts the fault features, it inevitably learns the noise features, which may lead to the overfitting problem. Therefore, the denoising branch (DB) aims to obtain signal components with high SNR from raw signals, so as to enhance the CNN's learning and discrimination ability for fault features. In the real HST operation, noise often comes from multiple sources. We assume that the noise obeys different probability distributions and are independent of each other. According to the central limit theorem, these signals' normalized sum tends to be a Gaussian distribution as the number of noise sources increases. Therefore, for the first time,

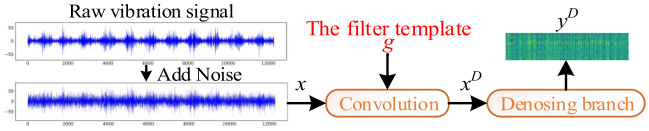


Fig. 2. Data flow of the DB.

the DB introduces the Gaussian filter into the vibration signals and combines it with the CNN to improve the antinoise ability of the networks.

The 1-D discrete Gaussian function $f[j]$ is given as

$$f[j] = \frac{1}{\delta\sqrt{2\pi}} \exp\left(-\frac{j^2}{2\delta^2}\right) \tag{1}$$

where j is integer, and δ is the standard deviation of j ($\delta = 1$).

Then, take the center of the template as the coordinate origin, and select the length of the template g so as to obtain the template $g = [f[-2], f[-1], f[0], f[1], f[2]]$. The length is set to 5, which is selected from four candidates (3, 5, 7, and 9) by empirical testing of CNN performance.

The data flow of the DB is shown in Fig. 2, where the raw vibration signal means recording data directly from accelerometers, x is the vibration signal after adding Gaussian white noise, and x^D is the convolution between x and g. The DB adopts the network branch constructed by multiple convolutional layers to learn the fault features of x^D. In this article, the DB adopts five convolutional layers, so the output y^D of the DB can be represented as $y^D = \text{Conv}(\text{Conv}\ldots\text{Conv}(x^D))$.

3) Low-Frequency Branch: The low-frequency branch (LFB) is designed to extract more discriminant fault features from the low-frequency components of the raw signals. The LFB is a low-frequency spectral range analysis. More specifically, relative to the carrier signal, the fault signal is often characterized by a low-frequency component, which contains more discriminant fault features. First, the LFB obtains the low-frequency component of x through the low-pass filter and then uses multiple convolutional layers to learn the features of the low-frequency component.

In signal processing, there are many low-pass filters, such as limited-amplitude filter, median filter, and moving average filter. The LFB is implemented hereby by the moving average filter, which can be directly implemented on x and generate a low-frequency signal. The moving average works by converting x into a new time series $x^L = \{x_1^l, x_2^l, \ldots x_N^l\}$, where x_i^L is the value at the i-node of x^L and represented as

$$x_i^l = \frac{1}{m} \sum_{i}^{i+m-1} x_i, \quad 1 \le i + m \le N \tag{2}$$

where m is the window size. The LFB adopts the same network as the DB to learn the more essential fault features from x^L. The output of the LFB is $y^L = \text{Conv}(\text{Conv}\ldots\text{Conv}(x^L))$.

B. Multiscale Learning

1) Multiscale Convolutional Module: The core idea of the multiscale learning is to learn the complementary long-term and short-term features in the raw signals from different time scales.

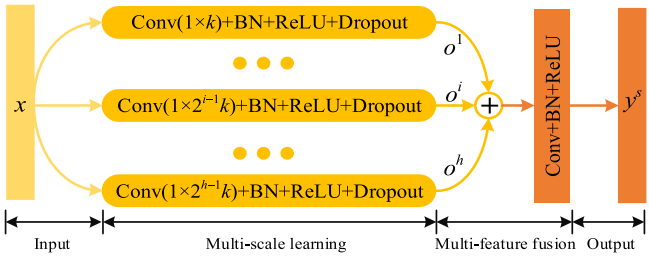


Fig. 3. Structure of the multiscale convolution module.

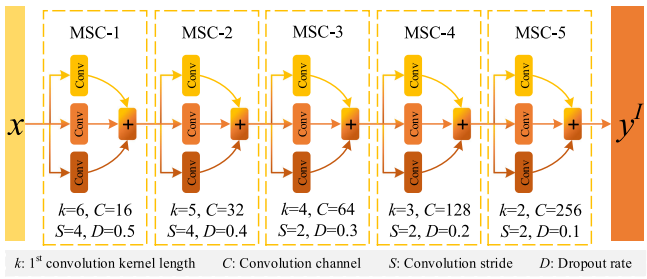


Fig. 4. Structure of the DMSN.

The multiscale convolutional module (MSC) skillfully adjusts the size of convolution kernels to enable convolutional layers to learn features of different time scales and, thus, effectively combining the multiscale learning with the CNN. Specifically, the MSC uses several parallel convolutional layers with different sizes of convolution kernels to simultaneously learn rich features at different scales and adopts convolution operations to fuse multiscale features across channels.

According to Fig. 3, first, the MSC uses h parallel convolutional layers to simultaneously learn multiscale features of the input signal x. The output of each convolutional layer is expressed as

$$o^i = \beta_i \left(w_i \otimes x + b_i \right) \tag{3}$$

where $\otimes$ represents convolution operation, i is the layer index, w_i and b_i are convolution kernel and bias of the ith convolutional layer, respectively, the size of w_i is $1 \times 2^{i-1}k$, k is an integer, and $\beta(\cdot)$ means the function transformation of the operation rectified linear unit (ReLU) [27], batch normalization (BN) [28], and dropout [29].

For a different convolution kernel, the convolution operation is capable of extracting information at a different scale from the raw signals. Next, the feature o^i is concatenated into the feature vector $O = [o^1, o^2, \ldots, o^h]$ through channels. Then, O is fed into the feature fusion layer C_s to effectively fuse the complementary features of different time scales. Finally, the output of the MSC is $y^s = \mathrm{MSC}(x) = C_s(O)$.

2) Deep Multiscale Network: According to Lin *et al.* [30], the shallow layers of the network learns the low-level features of signals, whereas the deep layers of the network learns the high-level features of signals. The CNN can learn the low-level/middle-level/high-level features of input signals, and the level of features can be enriched by simply stacking convolution layers. To learn high-level fault features, this article constructs the deep multiscale network (DMSN) that consists of multiple stacked MSCs. The final output of the IMB is $y^I = \mathrm{MSC}(\mathrm{MSC} \ldots \mathrm{MSC}(x))$.

The DMSN is shown in Fig. 4, where dropout rate (D) refers to the probability of randomly dropping neurons in the process of network training, and convolution stride (S) refers to the distance of each move of the convolution kernel. If $S = 2$, the convolution kernel slides by 2 at a time. The DMSN replaces five standard convolution layers in the DB and the LFB by the five MSC modules, where the convolution stride and the dropout rate of the convolution layer in the DB and the LFB are consistent with the DMSN. In addition, the DMSN follows several design rules.

1) The number of MSCs determines the depth of the DMSN.
2) The size of the convolution kernel and the dropout rate in the DMSN decrease with the increase in depth.
3) The number of channels in the DMSN increases with the increase in depth.

Parameter settings in the DMSN can be adjusted according to the real situation. For example, more MSC modules can be stacked to extract high-level features of more complex signals, and the number of channels and the size of convolution kernel can be adjusted as well.

C. Multifeature Fusion

The features of multiple branches or scales that the MBSCNN learns are the different levels of understanding of raw signals. These features do not always promote each other and may even contradict and cause misjudgment. Therefore, an effective multifeature fusion mechanism is necessary for the CNN. The MBSCNN adopts the multifeature fusion technology in two aspects. One is to use the feature fusion layer C_s to fuse multiscale features in the MSC, and the other is to use the feature fusion layer C_b to fuse the multiple signal features learnt by three branches. Both C_s and C_b use the feature learning and abstraction ability of convolutional layers to adaptively fuse multifeature information. The main difference lies in the number of the convolutional layers and the convolution kernels. According to Fig. 1, the MBSCNN concatenates y^I, y^D, and y^L of the multifeature learning into the feature vector $y^C = [y^I, y^D, y^L]$. Then, the multifeature fusion layer C_b adaptively fuses the rich feature information in y^C to obtain the final feature $y = C_b(y^C) = C_b([y^I, y^D, y^L])$ that the MBSCNN learns from x.

D. Fault Classification

In the MBSCNN, global average pooling [31] is adopted to replace the fully connected layers in the CNN. It can effectively avoid the overfitting problem that easily occurs in the full-connection layer and, thus, improve the generalization ability of the CNN. For the wheelset bearing fault diagnosis task, the Softmax function is used in the classification layer to give the results. Assuming that there are n classes for the input samples, the output probability Q_j for the class j is calculated as shown

in (4). The diagnosis output is the fault label corresponding to the largest Q_j

$$Q_j = \frac{\exp(\theta^j \mathrm{GAP}(y))}{\sum_{j=1}^{n} \exp(\theta^j \mathrm{GAP}(y))}, \quad j = 1, 2, \ldots, n \tag{4}$$

where θ^j is the classification layer parameter and $\sum_{j=1}^{n} Q_j = 1$.

III. Validation of Overall Performance

When the HST is in operation, the working condition of the wheelset bearings changes greatly. First, the weak fault signal features are easily submerged in the background noise. Therefore, it is very important and challenging to have high-precision fault diagnosis ability in noisy environment. Second, the HST's workload changes from time to time, and the related signal features change accordingly. Because it is not realistic to collect and label enough training data, it greatly improves the efficiency and applicability of the diagnostic method if the fault diagnosis method has the ability to diagnose the health status of samples under other loads by learning the training data under the existing loads. Therefore, from the abovementioned two aspects, this section verifies the effectiveness and superiority of the proposed MBSCNN on the wheelset bearing dataset.

A. Wheelset Bearing Data Description

The training data can be roughly grouped into two categories: normal data and fault data. The normal data are usually very easy to obtain because the real HST works under the normal state in most time; however, obtaining real fault data is a challenge. One solution for the fault data collection uses the experimental HST and its on-board data acquisition system. The inspection train is an option of the experimental HSTs. It is designed to detect potential abnormalities/faults of the whole rail system. Because the inspection train is fully controllable and carries no passengers, it can be used as the experimental HST, and faults embedded into the inspection train are allowed. The data acquisition system is then used to collect the data corresponding to the faults. Although the inspection train is still different from the real operating HST in terms of, for example, speed and load, this solution can generate the most real and relevant fault data mimicking real HST operation in a short period. The training data collection solution used in this article is similar to the first solution. The only difference is that the wheelset bearing is not mounted on the experimental HST.

The wheelset bearing test rig is mainly composed of a drive motor, a belt transmission system, a real wheel and axle, two real axle boxes and wheelset bearings, a vertical loading set, a lateral loading set, two fan motors, and a control system, as shown in Fig. 5. Vibration signals sampled at 5120 Hz are collected via four accelerometers mounted on the axle boxes of the HSTs. The installation of sensors should consider the following three general principles [32], [33].

1) Sensors should be placed as close to load zone of bearings as possible.
2) Transfer path between the sensors and the target bearing should be as short as possible.
3) Installation place should be considered for ease of installation, and installation surface should be as clean and flat as possible.

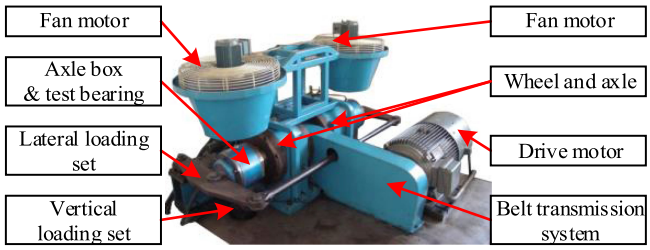


Fig. 5. Wheelset bearing test rig.

TABLE I
Description of 12 Health Conditions

Fault Description	Class Label
Normal	C1
Inner race pitting	C2
Rolling element pitting	C3
Rolling element flaking with a size of 3 mm×35 mm	C4
Inner race flaking with a size of 3 mm×45 mm	C5
Rolling element cracking	C6
Mixed fault with outer race flaking (10 mm×45 mm) and rolling element pitting	C7
Inner race flaking with a size of 10 mm×45 mm	C8
Outer race flaking with a size of 10 mm×30 mm	C9
Rolling element flaking with a size of 1 mm×1 mm	C10
Cage cracking	C11
Outer race flaking with a size of 10 mm×45 mm	C12

For the wheelset bearing test rig, the vibration sensors are placed on the housing surface of the axle box in the 12 o'clock and the 3 o'clock. The vertical load zone for the wheelset bearing test rig is in the 6 o'clock direction; however, this direction is hard for sensor installation. The first and the third principles are conflicting with each other. For the HST, according to the abovementioned three principles, the vibration sensors are usually placed on the top housing of the axle box in the 12 o'clock direction that is also the vertical load zone of wheelset bearings. By comparing the two sensor installation strategies, the recorded data become more effective in the strategy for the HST than that used in the wheelset bearing test rig. In this sense, our strategy is more rigorous.

Considering that various failure patterns may occur in the real operation of wheelset bearings, 12 typical health status of wheelset bearings are set. The fault information with respect to the experimental bearings is listed in Table I. To simulate the complex and changing working conditions of the HSTs during their operation as much as possible, under each health condition, five running speeds are designed, 60, 90, 120, 150, and 180 km/h, four vertical load conditions are conducted: 56, 146, 236 and 272 kN, and two lateral load conditions are set, 0 and 20 kN. Therefore, each health condition includes 40 kinds of working conditions, which is a challenge for fault diagnosis methods.

A large amount of data are required for the development of deep-learning-based fault diagnosis methods, so the data augmentation is a practical technique to increase the number of samples. This article adopts the same sliding segmentation

method reported in [24]. In our experiments, the length of each sample is 2048, and the step size of sliding segmentation is set to 256. Finally, 188 088 samples are obtained.

B. Comparison Settings

The MBSCNN is implemented by the Keras library under Python 3.5. Training and testing of the networks are performed on a workstation with Ubuntu 16.04 operating system, an Intel Core i7-6850K CPU, and a GTX 1080Ti GPU. During the training process, we adopt the cross-entropy loss function and the Adam optimization algorithm with a learning rate of 0.0001 and a batch size of 96.

We compare the proposed MBSCNN with the five state-of-the-art networks. The five networks are a new CNN based on LeNet-5 proposed by Wen *et al.* (Wen-CNN) [21], adaptive deep convolution neural network (ADCNN) [19] based on 2DCNN, deep convolutional neural networks with wide first-layer kernels (WDCNN) [20] based on 1DCNN, multiscale convolutional neural network (MSCNN) [15] based on the multiscale learning, and four-layer back propagation neural network (BPNN) with the same structure as presented in [34]. The training strategies of the six methods are the same in the comparison experiments. Each MSC module in the MBSCNN adopts four scales, namely $h = 4$.

The MSCNN proposed by Jiang *et al.* [15] uses the multiscale learning to solve the problem of wind turbine gearbox fault diagnosis and achieved a satisfactory performance. The MSCNN and the proposed MBSCNN have the following differences.

1) Multiscale operation: The MBSCNN uses multiple parallel convolutional layers to learn multiscale features. The MSCNN adopts the multiscale coarse-grained operation, which downsamples the raw signals and is prone to losing some features of the raw signals.
2) Multibranch operation: There are three branches in the MSCNN, and these three branches act only to learn the multiscale features. The MBSCNN, however, combines multibranch learning with the conventional filtering methods to extract more essential fault signal features from different signal components.
3) Multifeature fusion: The MSCNN does not use any feature fusion operation, and the learning features are directly fed into the classifier to get the result. The MBSCNN adopts multiple feature fusion layers to fuse and optimize the feature information learned by the multibranch learning and the multiscale learning, to promote the CNN to make the best judgment.

We adopt the indicator of accuracy that is a commonly comprehensive performance metric defined in (5). The accuracy is estimated with fourfold cross-validation for each set of experiments

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \tag{5}$$

where TP, FP, TN, and FN refer to the number of true positive samples, false positive samples, true negative samples, and false negative samples, respectively. The value of *Accuracy* ranges from 0 to 1, and a larger value means better classification performance.

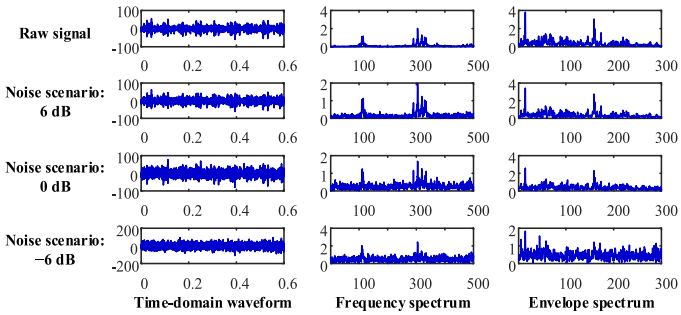


Fig. 6. Time-domain waveforms, frequency spectra, and envelope spectra of the raw signal and noise signals under the three SNR scenarios.

C. Experimental Results

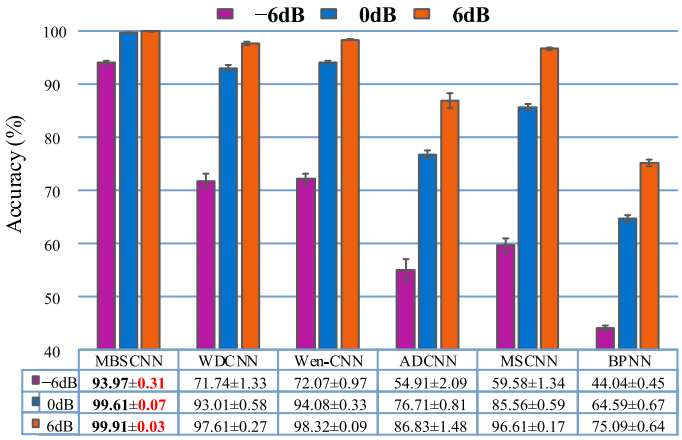
1) Robustness Against Noise: To better simulate the strong noise interference faced by the HSTs, we add additional white Gaussian noises into the raw signals. The SNR is defined as

$$\text{SNR}_{\text{dB}} = 10\log_{10}\left(\frac{P_{\text{signal}}}{P_{\text{noise}}}\right) \tag{6}$$

where P_{signal} and P_{noise} are the power of the raw signal and the added noise, respectively.

In this section, we verify the diagnostic performance of the MBSCNN in different noise environments. We set up three SNR scenarios (i.e., −6, 0, and 6 dB), respectively, simulating the wheelset bearings under the strong, the medium, and the weak noise conditions. Fig. 6 shows an example of time-domain waveforms, frequency spectra and envelope spectra of the raw C5 signals, and the noise signals under the three SNR scenarios. It can be found that with the increase in noise, the time-domain vibration signal becomes complex, and the noise in frequency spectrum and envelope spectrum gradually increases, drowning out some impulse signals.

The experimental results are shown in Fig. 7. Obviously, the MBSCNN achieves the best diagnostic performance under each SNR scenario. In particular, even when SNR $= -6$ dB (that is, the power of noise is about 3.98 times of that of the raw signal), the MBSCNN still obtains 93.97% diagnostic accuracy. Furthermore, the MBSCNN has almost 22% improvement comparing with the Wen-CNN. It indicates that the MBSCNN has strong antinoise ability without requiring any additional denoising operations. The BPNN has the lowest accuracy among the six comparison networks, which again indicates that the CNN-based methods can be better applied to the fault diagnosis of wheelset bearings. The MSCNN performs poorly in the noise environment, indicating that the MBSCNN can extract richer and more comprehensive discriminant features and adaptively fuse and optimize multiple features to enable the CNN to obtain more accurate judgment. Moreover, according to the variance results presented in Fig. 7, the MBSCNN has the smallest



	MBSCNN	WDCNN	Wen-CNN	ADCNN	MSCNN	BPNN
−6dB	**93.97±0.31**	71.74±1.33	72.07±0.97	54.91±2.09	59.58±1.34	44.04±0.45
0dB	**99.61±0.07**	93.01±0.58	94.08±0.33	76.71±0.81	85.56±0.59	64.59±0.67
6dB	**99.91±0.03**	97.61±0.27	98.32±0.09	86.83±1.48	96.61±0.17	75.09±0.64

Fig. 7. Performance of the six comparison networks under the three SNR scenarios. The black bold font represents the maximum accuracy; and the red bold font indicates the minimum variance of the accuracy.

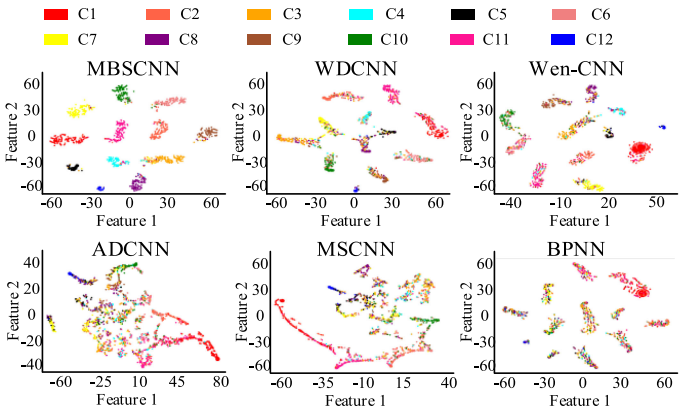
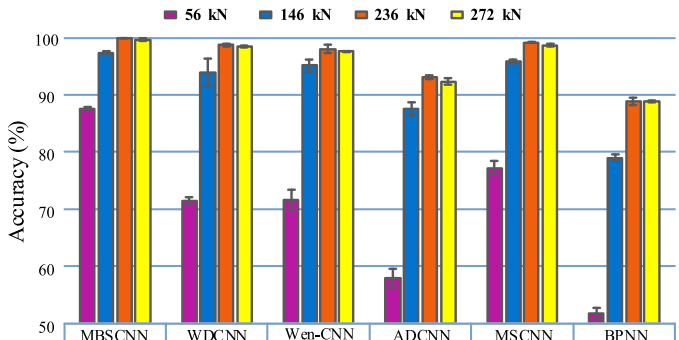


Fig. 8. Visualization of features learned by the six comparison networks under strong noise condition (SNR = −6 dB).

variance under the three noise scenarios, so its network stability is the best among the six comparison networks.

We use the t-distributed stochastic neighbor embedding (t-SNE) [35] to visualize the distribution of the 2000 testing samples in a 2-D space. In Fig. 8, colors represent health conditions of wheelset bearings. Obviously, the feature information learned by the MBSCNN has the best discrimination, indicating that the MBSCNN can learn more discriminant fault features from complex vibration signals.

2) Domain Adaptability of Variable Loads: Domain adaptability is an important task in the wheelset bearing fault diagnosis under variable loads. We cannot collect training data from all load conditions; therefore, diagnostic models trained under existing loads should be able to work under a new load condition. In such application, the feature space and the class space are identical between the training and the test data, and only the feature distribution is inconsistent. It is a typical domain adaptability problem. It is also one of the research contents in transfer learning that studies not only the inconsistency of the feature distribution, but also the inconsistency of the feature space or the class space. In this section, we validate the load domain adaptability of the MBSCNN. The vibration signals of



	MBSCNN	WDCNN	Wen-CNN	ADCNN	MSCNN	BPNN
56kN	**87.61±0.19**	71.51±0.64	71.59±1.77	58.01±1.63	77.14±1.22	51.79±0.96
146kN	**97.39±0.23**	93.92±2.43	95.13±1.12	87.55±1.07	95.83±0.34	78.98±0.53
236kN	**99.92±0.03**	98.75±0.27	98.05±0.76	93.05±0.37	99.14±0.09	88.91±0.63
272kN	**99.66**±0.33	98.42±0.19	97.61±**0.12**	92.35±0.53	98.68±0.19	88.85±0.20

Fig. 9. Performance of the six comparison networks in load domain adaptation. The black bold font represents the maximum accuracy; and the red bold font indicates the minimum variance of the accuracy.

wheelset bearings are collected under various load conditions. We select the four kinds of vibration signals under the vertical load condition as datasets, which includes 56, 146, 236, and 272 kN. Four sets of experimental data were obtained by taking one load data as the testing set and the other three load data as the training set.

Each set of experiment is repeated three times. The experimental results are shown in Fig. 9. Obviously, the MBSCNN obtained the best mean diagnostic results in the four load domain adaptability tasks, indicating that when the working load changes, the MBSCNN has good load domain adaptability without requiring specific domain adaptation methods. We have two observations from the trend of accuracy. 1) The maximum accuracy is presented in the case of load 236 kN, and standing at this point, the accuracy decreases along both load sides. This may be explained by the nonlinear data distribution, and extrapolation usually has a bigger error than interpolation. 2) The left-hand side using training data from a bigger load to predict testing data from a smaller load is more difficult than the right-hand side using training data from a smaller load to predict testing data from a bigger load. According to forced vibration theory in rotating machinery dynamics, the vibration amplitude is positively proportional to the external excitation force whose value is indicated by the load. The smaller the load is, the weaker the vibration response of wheelset bearing is, and thus the corresponding fault features are weaker too. Therefore, discriminant fault features learning from a bigger load most likely be blur in a smaller load, and in contrast, discriminant fault features learning from a smaller load are usually still reserved in a bigger load. Thus, accuracy decreases faster along the left-hand side to small load than that along the right-hand side to big load.

The accuracy of the MBSCNN under the smallest load (56 kN) is 87.61%. Although this accuracy is not high as other load cases, our performance is still the best among the six comparison networks. It shows that the fault learning features have good generalization. The MSCNN has the best domain adaptability among the five comparison networks, which indicates that

TABLE II
TESTING TIME OF THE SIX COMPARISON NETWORKS

Method	MBSCNN	WDCNN	Wen-CNN	ADCNN	MSCNN	BPNN
Testing time / s	1.882	0.694	0.622	0.613	0.793	0.671

multiscale learning can effectively dig out more discriminant fault features from the raw signals. However, under the 56 kN load, the accuracy of the MSCNN is nearly 10% lower than that of the MBSCNN, which once again indicates that the MBSCNN has stronger multiscale learning ability and more effective multifeature fusion ability. Moreover, according to the variance results presented in Fig. 9, the stability of the MBSCNN is generally better than the five comparison networks. To sum up, through domain adaptability, the generalization capabilities of the MBSCNN has been tested.

3) Computational Burden of the Networks: Computational burden is an important feature of the CNN-based fault diagnosis methods. This section calculates testing time of the six comparison networks. Table II records the testing time for a batch size (96 samples). According to Table II, the MBSCNN consumes more time than other methods. This is not a surprise because the MBSCNN has the multibranch and the multiscale modules, which brings more parameters and also more complex structure than other methods. It is worth pointing out that the testing time of the MBSCNN for a batch size is 1.882 s, which is acceptable in engineering practice.

IV. VALIDATION OF THE THREE IMPROVEMENTS

The performance of intelligent fault diagnosis methods is closely related to the quality of feature extraction. The core of the MBSCNN is to learn and fuse abundant and complementary features from different perspectives through the multibranch learning, the multibranch learning and the multifeature fusion layers, so as to obtain better performance. Therefore, the following three aspects are used to explore the impact of different feature information on diagnostic performance. Antinoise experiments and load domain adaptability experiments are carried out under the SNR scenario of −6 dB and the load of 56 kN respectively, and the testing accuracy of each epoch is recorded in the whole process.

A. Necessity of the Multibranch Learning

To quantitatively evaluate the effectiveness of the multibranch learning, three CNN structures, including MBCNN-B1 (only the IMB), MBCNN-B2 (with the IMB and the LFB) and MBCNN-B3 (with the IMB, the LFB, and the DB), are set in this experiment. Note that the abovementioned three CNNs do not include the multiscale module, so we call them multibranch CNN (MBCNN). In other words, the IMB does not use multiscale convolution modules, it only uses the same multilayer convolutional network as the LFB and the DB.

The experimental results are shown in Table III. Fig. 10(a) represents the testing results of each epoch under the strong noise condition (−6 dB), and Fig. 10(b) represents the testing results of each epoch under the variable load condition. Obviously, the multibranch learning has better antinoise ability and load domain adaptability. This is because the multibranch learning can learn more fault-related features from multiple signal components and complement each other with features learned from the raw signals, which effectively improves the robustness of feature learning. In the experiments of load domain adaptability, the feature distributions of the testing samples are different from those of the training samples, but the accuracy of the MBCNN-B3 is still improved nearly 20% compared with the MBCNN-B1, which indicates that the multibranch learning can extract more domain invariant features from different signal components. In the antinoise experiments, the diagnostic performance of the MBCNN-B3 is similar to that of the MBCNN-B2, but is 3.6% higher than that of the MBCNN-B1. This is because the moving average filter method used in the LFB has certain denoised ability, which is repeated with the Gaussian filter of the DB, and more branches do not show better antinoise performance. Therefore, in the future work, advanced filtering methods can be developed to improve the MBSCNN's performance for more complex tasks.

TABLE III
EXPERIMENTAL RESULTS OF DIFFERENT BRANCHES

Accuracy (%)	MBCNN-B1	MBCNN-B2	MBCNN-B3
Strong Noise	83.71	87.54	87.33
Variable Loads	64.30	74.60	82.01

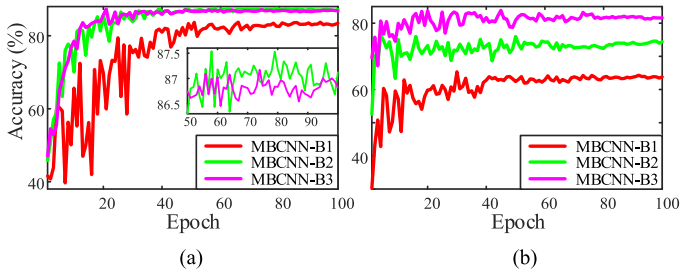


Fig. 10. Performance of different branches with respect to epoch. (a) Antinoise experiments. (b) Load domain adaptation experiments.

TABLE IV
EXPERIMENTAL RESULTS OF THE FOUR SCALES

Accuracy (%)	IMB-S1	IMB-S2	IMB-S3	IMB-S4
Strong Noise	71.25	88.47	92.17	93.89
Variable Loads	64.67	74.40	75.76	82.32

B. Necessity of the Multiscale Learning

To quantitatively evaluate the effectiveness of the multiscale learning, four CNN structures, including the IMB-S1 ($h = 1$), the IMB-S2 ($h = 2$), the IMB-S3 ($h = 3$), and the IMB-S4 ($h = 4$), are set in this section. The four networks do not contain the LFB and the DB.

The experimental results are shown in Table IV. Fig. 11(a) represents the testing results of each epoch under the strong noise condition, and Fig. 11(b) represents the testing results of each epoch under the variable load condition. Obviously, under the strong noise condition, the diagnostic performance increases

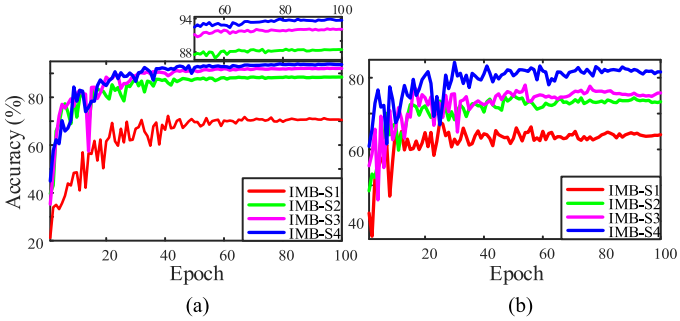


Fig. 11. Performance of different scales with respect to epoch. (a) Antinoise experiments. (b) Load domain adaptation experiments.

TABLE V
EXPERIMENTAL RESULTS OF THE FOUR FEATURE FUSION CASES

Accuracy (%)	MBSCNN	MBSCNN -NSF	MBSCNN -NBF	MBSCNN -NSF-BF
Strong Noise	93.97	93.10	93.19	92.93
Variable Loads	87.61	74.75	78.92	73.06

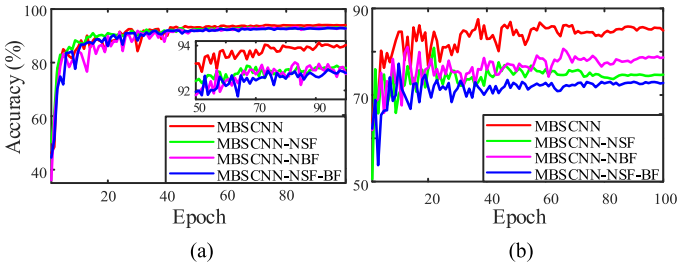


Fig. 12. Performance of different feature fusion cases with respect to epoch. (a) Antinoise experiments. (b) Load domain adaptation experiments.

with the increase in the number of scales. Moreover, the IMB-S4 has roughly a 22.5% improvement over the IMB-S1, which shows that multiscale network can extract more comprehensive fault features from multiscale features under the strong noise condition, so as to obtain better diagnosis results.

According to Fig. 11(b), with the increase in the number of scales, the CNN's load domain adaptability increases. The accuracy of the IMB-S4 has nearly 18% improvement over that of the IMB-S1. This further indicates that the MBSCNN with multiscale feature learning ability has huge advantages over the conventional CNN. It also proves that the proposed multiscale convolution module has excellent multiscale feature learning ability.

C. Necessity of the Multifeature Fusion

To quantitatively evaluate the effectiveness of multifeature fusion, the MBSCNN, the MBSCNN-NSF (no multiscale feature fusion layer in the MSC), the MBSCNN-NBF (no multi-branch feature fusion layer), and the MBSCNN-NSF-BF (no multiscale and multibranch feature fusion layers) are set in this section. The four networks contain the multibranch and the multiscale structures.

The experimental results are shown in Table V. Fig. 12(a) represents the testing results of each epoch under the strong noise condition, and Fig. 12(b) represents the testing results of each epoch under the variable load condition. For load domain adaptive tasks, the CNN needs to learn more discriminant fault features to adapt to fault diagnosis tasks under variable loads. The proposed multifeature fusion layer can fuse and optimize features learned from different sources, so as to provide consistent discriminant information for the classification layer. Therefore, the multifeature fusion layer can significantly improve the CNN's load domain adaptability. In particular, under the 56 kN load, the diagnostic performance of the MBSCNN is improved by 12.62% compared with that of the MBSCNN-NSF-BF. Additionally, under the strong noise condition, because the multibranch learning and the multiscale learning can help the CNN to learn discriminant fault features well, the effect of multifeature fusion layer is not particularly prominent. However, when SNR $= -6$ dB, the MBSCNN also has a 1.2% increase compared with the MBSCNN-NSF-BF, indicating that the multifeature fusion layer can further improve the antinoise ability of the MBSCNN.

V. DISCUSSIONS

A. Discussion on the Multiscale and Multibranch Learning

The core of this article is to get more features that can reflect the intrinsic nature of the faults from the input raw signals. Aiming at this problem, we introduce two mechanisms. One is to build a more complex network to obtain more features from the raw signals, namely the multiscale learning module in the IMB. The other is to try to filter the signal to remove the redundant information in the raw signals and make fault-related features easier to be acquired, that is, multibranch learning. According to the experimental results, the two mechanisms complement each other in diagnostic performance but also overlap to some extent.

For the multiscale learning, as shown in Table IV, from the IMB-S1 to the IMB-S4, the diagnostic performance of the CNN under noise and load domain adaptivity increased from 71.25% and 64.67% to 93.89% and 82.32%, respectively. This is a surprising achievement demonstrating that by capturing and integrating information from different time scales in vibration signals, the CNN can understand and characterize signal features from both local and global perspectives, thus achieving better performance. It is worth noting that when dealing with other datasets with different sampling frequencies, the convolution kernel size should be adjusted accordingly, so as to maintain the MBSCNN's good performance. The higher the sampling frequency is, the wider the convolution kernel should be.

For the multibranch learning, we design three branches: the IMB, the DB, and the LFB. The IMB focuses on obtaining richer signal features from the raw signals containing complex information. The DB and the LFB, respectively, focus on removing redundant information of the raw signals from different aspects and, thus, extract signal components that can better reflect fault features. According to Table III, from the MBCNN-B1 to the MBCNN-B3, the diagnostic performance of

the CNN under noise and load domain adaptation increased from 83.71% and 64.30% to 87.33% and 82.01%, respectively. This proves that the DB and the LFB can provide signal features from different aspects and enhance the diagnostic performance of the CNN. Specifically, for the load domain adaptability task, the performance is greatly improved with each additional branch, ranging from 64.30%, 74.60% to 82.01%. This shows that the model learns complementary features from multiple branches, which improves the model performance and proves the validity of different branches. Therefore, the DB and the LFB are both necessary. However, under the noise condition, the performance of the MBCNN-B3 is not improved compared with that of the MBCNN-B2. This is because Gaussian noise is essentially a high-frequency signal, so the contribution of the DB on the CNN performance under the noise condition is covered by the LFB. Similarly, the performance of the MBSCNN in Table V is only 0.08% higher than that of IMB with multiscale learning module, so the contribution of the multibranch learning on the CNN performance is also covered by the multiscale learning. This proves that the multibranch learning can be further improved. It inspires us to explore other advanced signal processing methods (including adaptive signal decomposition methods, such as empirical mode decomposition and variational mode decomposition) to further improve the proposed MBSCNN in the future.

B. Discussion on Potential Applications of the MBSCNN

In this section, we discuss the potential of applying the MBSCNN in real applications from the perspectives of experimental setting and accuracy performance.

1) Experimental setting. First, the test rig has a real wheel and axle, two real axle boxes, and wheelset bearings. The test rig also has the vertical and lateral loading sets to mimic 2-D loads in real HST operation. In addition, a fan device is used to simulate the wind generated in the HST movement. Second, the faulty wheelset bearings are collected from the HSTs in the real running line, that is, the faults were naturally generated and were 100% real. Thus, the test rig can effectively simulate the vibration response caused by the real faulty wheelset bearings. Third, we have also considered 40 different operating conditions, which cover a wide range of the real HST operation condition. Fourth, the test rig is unable to simulate the random excitation of the track and the disturbance of the axle movement unbalance during the operation of the HST, thus we use the artificial noise scenarios to simulate these noise conditions. For real HSTs, the SNR of vibration signals is difficult to quantify. To the best of our knowledge, we do not find studies reporting the noise range in the considered application field. We subjectively select the three SNR scenarios, but this selection covers a wide range of noises from weak to strong.
2) Accuracy performance. On the one hand, under the SNR scenarios of 6 and 0 dB, the MBSCNN can achieve high accuracy of 99.91% and 99.61%, respectively. This indicates that it can provide an acceptable diagnostic performance under regular practical conditions. Under the SNR scenario of −6 dB (that is, the adding noise power is 3.98 times the signal power), the accuracy decreases to 93.97%, but the noise power far exceeds the real signal power, which rarely happens in the real HST operation. In this article, the SNR scenario of −6 dB is set to verify the detection capability of the MBSCNN for extremely weak faults. On the other hand, the processing of domain adaptivity task is a challenge for all networks. Nevertheless, the MBSCNN still achieves acceptable accuracy of 99.66% and 99.92% under the load condition of 272 kN and 236 kN, respectively. Under the load condition of 56 kN, its accuracy decreases to 87.61%. The reasons have been explained in Section III-C.2. In the practical application, the performance degeneration in the 56 kN load condition can be avoided by simply including data from small loads in the training dataset, so as to guarantee a satisfactory performance of domain adaptability in a wide load range.

To sum up, various efforts have been made to make the experiments simulate the real HST as much as possible, so the MBSCNN verified on this dataset have considerable practical application potential. In addition, the MBSCNN can provide an acceptable fault diagnosis performance under regular practical conditions, but it needs to be further improved under extreme noise and load domain adaptivity conditions.

VI. Conclusion

This article proposed the MBSCNN-based fault diagnosis method to intelligently diagnose the wheelset bearing's faults from the vibration signals under strong noise and variable load condition. The MBSCNN combines the conventional filtering technology and the multiscale learning and can automatically learn and fuse rich and complementary features from the multiple signal components and time scales, so as to accurately judge the health status of the wheelset bearings.

The experimental results on the wheelset bearing dataset demonstrated that the MBSCNN is superior to the five state-of-the-art networks in the antinoise ability and load domain adaptability. The experimental results proved that the multibranch network, multiscale network, and the multifeature fusion can significantly improve the CNN performance. Therefore, it is convinced that the MBSCNN can be applied in intelligent fault diagnosis on the wheelset bearings of the HSTs. Moreover, there is a high potential for the MBSCNN to be applied to other industrial systems with similar strong noise and variable operation conditions.

For the future work, first, we are going to introduce the transfer learning to deal with the problem of speed domain adaptability. Under the variable speed condition, the rotational speed of wheelset bearings changes, which caused a huge change in the distribution of sample feature. Therefore, it is a better choice to use the transfer-learning-based method to learn the domain invariant features of samples under variable speeds. Second, for the HSTs in operation, the number of normal data is far more than that of fault data; therefore, the study on the class imbalance problem is also an important direction.

REFERENCES

[1] H. Cao, F. Fan, K. Zhou, and Z. He, "Wheel-bearing fault diagnosis of trains using empirical wavelet transform," *Measurement*, vol. 82, pp. 439–449, 2016.

[2] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.,* vol. 108, pp. 33–47, 2018.

[3] Z. Liu, M. J. Zuo, Y. Jin, D. Pan, and Y. Qin, "Improved local mean decomposition for modulation information mining and its application to machinery fault diagnosis," *J. Sound Vib.*, vol. 397, pp. 266–281, 2017.

[4] J. Faiz, V. Ghorbanian and B.M. Ebrahimi, "EMD-based analysis of industrial induction motors with broken rotor bars for identification of operating point at different supply modes," *IEEE Trans. Ind. Inf.*, vol. 10, no. 2, pp. 957–966, 2014.

[5] 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. Sig. Process.*, vol. 19. no. 5, pp. 974–988, 2005.

[6] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: A review with applications," *Signal Process.*, vol. 96, no. 5, pp. 1–15, 2014.

[7] D. Wang, "K-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: Revisited," *Mech. Syst. Sig. Process.*, vol. 70/71, pp. 201–208, 2016.

[8] S. Shevchik, F. Saeidi, B. Meylan, and K. Wasmer, "Prediction of failure in lubricated surfaces using acoustic time-frequency features and random forest algorithm," *IEEE Trans. Ind. Inform.*, vol. 13, no. 4, pp. 1541–1553, Aug. 2017.

[9] D. You, X. Gao and S. Katayama, "Multisensor fusion system for monitoring high-power disk laser welding using support vector machine," *IEEE Trans. Ind. Inform.*, vol. 10, no. 2, pp. 1285–1295, May 2014.

[10] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Syst.*, Nevada, USA, 2012, pp. 1–9.

[11] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[12] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. Conf. Empirical Methods Natural Lang.*, Doha, Qatar, 2014, pp. 1746–1751.

[13] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," in *Mech. Syst. Signal Pr.*, vol. 100, pp. 439–453, 2017.

[14] Z. Chen, C. Li and R. Sanchez, "Gearbox fault identification and classification with convolutional neural networks," *Shock Vib.*, vol. 2015, no. 2, pp. 1–10, 2015.

[15] G. Jiang, H. He, J. Yan, and P. Xie, "Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3196–3207, Apr. 2019.

[16] L. Jing, M. Zhao, and P. Li, "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox," *Measurement*, vol. 111, pp. 1–10, 2017.

[17] J. Senanayaka, K. Huynh and K. Robbersmyr, "Multiple classifiers and data fusion for robust diagnosis of gearbox mixed faults," *IEEE Trans. Ind. Inform*., vol. 15, no. 8, pp. 4569–4579, Aug. 2019.

[18] O. Janssens *et al.*, "Convolutional neural network based fault detection for rotating machinery," *J. Sound Vib.*, vol. 377, pp. 331–345, 2016.

[19] X. Guo, L. Chen and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, 2016.

[20] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, pp. 425–446, 2017.

[21] 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.

[22] S. Li, G. Liu and X. Tang, "An ensemble deep convolutional neural network model with improved D-S evidence fusion for bearing fault diagnosis," *Sensors*, vol. 17, no. 8, p. 1729, 2017.

[23] H. Shao, H. Jiang, H. Zhang and T. Liang, "Electric locomotive bearing fault diagnosis using novel convolutional deep belief network," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2727–2736, Mar. 2018.

[24] D. Peng, Z. Liu, H. Wang, Y. Qin, and L. Jia, "A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains," *IEEE Access*, vol. 1, pp. 10278–10293, 2019.

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

[26] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 63, no. 11, pp. 7067–7075, Nov. 2016.

[27] V. Nair and G. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, Haifa, Israel, 2010, pp. 807–814.

[28] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, Lille, France, PMLR, vol. 37, pp. 448–456, 2015.

[29] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res*., vol. 15, no. 1, pp. 1929–1958, 2014.

[30] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. 13th Eur. Conf. Comput. Vis.*, Zurich, Switzerland, 2014, pp. 818–833.

[31] M. Lin, Q. Chen, and S. Yan, "Network in network," in *Proc. Int. Conf. Learn. Representations*, Banff, Canada, 2014, pp. 1–10.

[32] P. D. Mcfadden and M.M. Toozhy, "Application of synchronous averaging to vibration monitoring of rolling element bearings," *Mech. Syst. Signal Process*, vol. 14, no. 6, pp. 891–906, 2000.

[33] M. Dumont, A. Cook and N. Kinsley, "Acceleration measurement optimization: Mounting considerations and sensor mass effect," in *Topics in Modal Analysis & Testing*, Mains M. Eds. Cham, Switzerland: Springer, 2016.

[34] G. F. Bin, J. J. Gao, X. J. Li, and B. S. Dhillon, "Early fault diagnosis of rotating machinery based on wavelet packets—Empirical mode decomposition feature extraction and neural network," *Mech. Syst. Sig. Process.*. vol. 27, no. 1, pp. 696–711, 2012.

[35] L. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res*., vol. 9, pp. 2579–2605, 2008.

Dandan Peng was born in Shaanxi, China, in 1992. She received the B.S. and M.S. degrees in mechanical engineering from the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2016 and 2019, respectively. She is currently working toward the Ph.D. degree in mechanical engineering from Catholic University of Leuven, Leuven, Belgium.

Her research interests include Hilbert–Huang transform, convolutional neural network, machinery condition monitoring, and fault diagnosis.

Huan Wang was born in Hunan, China. He received the B.S. degree in mechanical engineering in 2016 from the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China, where he is currently working toward the M.S. degree.

His research interests include mechanical fault diagnosis, image recognition, deep learning, and machine learning.

Zhiliang Liu (Member, IEEE) was born in Rizhao, China, in 1984. He received the Ph.D. degree in detection technology and automatic equipment from the School of Automation Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2013.

From 2009 to 2011, he was a Visiting Scholar with the University of Alberta. From 2013 to 2015, he was an Assistant Professor with the School of Mechanical and Electrical Engineering, UESTC, where he has been an Associate Professor since 2015. He has authored/coauthored more than 60 papers, including 20+ SCI-Indexed journal papers. His research interests include fault diagnosis and prognostics of rotating machinery by using advanced signal processing and data mining methods.

Dr. Liu currently holds more than ten research grants from the National Natural Science Foundation of China, Open Grants of National Key Laboratory, China Postdoctoral Science Foundation, etc.

Wei Zhang was born in Nanjing, China. He is currently working toward the B.S. degree in electrical and electronic engineering from the University of Electronic Science and Technology of China, Chengdu, China.

His research interests include deep learning and high-accuracy positioning system based on ultra wide band devices.

Ming J. Zuo (Senior Member, IEEE) received the B.S. degree in agricultural engineering from the Shandong Institute of Technology, Zibo, China, in 1982, and the M.S. and Ph.D. degrees from Iowa State University, Ames, IA, USA, in 1986 and 1989, respectively both in industrial engineering.

He is currently a Full Professor with the Department of Mechanical Engineering, University of Alberta, Edmonton, AB, Canada. His research interests include system reliability analysis, maintenance modeling and optimization, signal processing, and fault diagnosis.

Dr. Zuo is the Department Editor for the *IISE Transactions*, an Associate Editor for the IEEE TRANSACTIONS ON RELIABILITY and *Journal of Risk and Reliability*, a Regional Editor for the *International Journal of Strategic Engineering Asset Management*, and an Editorial Board Member of the *Reliability Engineering and System Safety*, *Journal of Traffic and Transportation Engineering*, *International Journal of Quality, Reliability and Safety Engineering*, and *International Journal of Performability Engineering*. He is a Fellow of the Institute of Industrial and Systems Engineers, a Fellow of the Engineering Institute of Canada, and a Founding Fellow of the International Society of Engineering Asset Management.

Jian Chen, photograph and biography not available at the time of publication.