IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 13, 2019, accepted August 15, 2019, date of publication August 21, 2019, date of current version September 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2936625

An Adaptive Weighted Multiscale Convolutional Neural Network for Rotating Machinery Fault Diagnosis Under Variable Operating Conditions

HUIHUI QIAO[1,2], TAIYONG WANG[1,3], PENG WANG[1,2], (Member, IEEE), LAN ZHANG[1], AND MINGDA XU[1]

[1]School of Mechanical Engineering, Tianjin University, Tianjin 300350, China
[2]Key Laboratory of Mechanism Theory and Equipment Design of Ministry of Education, Tianjin University, Tianjin 300350, China
[3]Renai College, Tianjin University, Tianjin 301636, China

Corresponding author: Peng Wang (pengwang@tju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51975402.

ABSTRACT Extracting robust fault sensitive features of vibration signals remains a challenge for rotating machinery fault diagnosis under variable operating conditions. Most existing fault diagnosis methods based on the convolutional neural network (CNN) can only extract single-scale features, which not only loss fault sensitive information on other scales, but also suffer from the domain shift problem. In this work, a novel end-to-end deep learning network named adaptive weighted multiscale convolutional neural network (AWMSCNN) is proposed to adaptively extract robust and discriminative multiscale fusion features from raw vibration signals. The AWMSCNN consists of three main components: the denoising layer, the adaptive weighted multiscale convolutional (AWMSC) block, and the multiscale feature fusion layer. The AWMSC block can learn rich and complementary features on multiple scales in parallel. Then, an adaptive weight vector is introduced to modulate multiscale features to emphasize fault sensitive features and suppress features that are sensitive to operating conditions. The train wheelset bearing dataset and the bearing dataset provided by Case Western Reserve University (CWRU) are used to verify the superiority of the proposed model over the basic CNN and other multiscale CNN models. The experiment results show that the proposed model has strong fault discriminative ability and domain adaptive ability against variable operating conditions.

INDEX TERMS Adaptive weighted multiscale feature learning, convolutional neural network, deep learning, fault diagnosis, rotating machinery, variable operating conditions.

I. INTRODUCTION

Rotating machinery is widely used in transportation, electric equipment, and manufacturing equipment [1]. Rotating machinery often operates under complex conditions such as variable speed, variable load, and strong noise [2], [3]. Under the influence of alternating stress and various random factors, faults will inevitably occur. Any small fault of the rotating machinery may evolve into a major safety accident. Therefore, effective fault diagnosis of rotating machinery under variable operating conditions is crucial to guarantee the system safety and reliability and to reduce maintenance costs [4].

In traditional data-driven based fault diagnosis methods, handcrafted feature extraction is a key step affecting the final diagnosis accuracy. The purpose of feature extraction is to extract fault sensitive information from sensor signals. Then, the extracted features are sent into a shallow machine learning model, such as support vector machines (SVM), neural networks, and decision trees, to implement fault detection. However, it is difficult and time-consuming to determine which features should be extracted [5]. What is worse, rotating machinery typically operates under variable operating conditions in practice, especially under variable speed and

The associate editor coordinating the review of this article and approving it for publication was Sungroh Yoon.

variable load, which will directly lead to the failure of traditional feature extraction methods based on the assumption of stationary rotating speed. Therefore, it is still a challenge to extract robust fault sensitive features under variable operating conditions.

To eliminate the influence of variable operating conditions on feature extraction, several methods including order tracking [6], [7], tacholess order tracking [8]–[10], high-order synchrosqueezing transform [11], variational mode decomposition [12], and Vold-Kalman filter with the multiscale sample entropy [13] have been proposed for fault diagnosis under variable operating conditions. These methods have achieved good fault diagnosis performances under variable operating conditions, but the feature extraction processes of these methods are all depend on prior domain knowledge and expert experience, so they have poor generalization ability. Extracting these hand-crafted features is time-consuming and labor-intensive with low efficiency, so it is difficult to meet the online monitoring requirement. Moreover, the feature extraction process and shallow machine learning models are independent of each other and can't be jointly optimized [14].

Deep learning provides a powerful solution to above weaknesses and has achieved breakthrough performance in the machinery fault diagnosis field taking advantage of its excellent automatic feature learning ability [15]. The convolutional neural network (CNN) is a commonly known deep learning model and has attracted more and more attention benefitting from its local connection mechanism, weight sharing strategy, and spatial pooling layers. In recent years, some CNN-based deep learning models have been developed for mechanical fault diagnosis. The excellent performance of existing CNN-based fault diagnosis models is based on the assumption that the training dataset and the testing dataset are drawn from the same distribution [16]. This assumption requires that the training dataset and the testing dataset are collected from the same operating condition. However, in the case of fault diagnosis under variable operating conditions, signals under only one or a few operating conditions are available for model training, and the testing signals are often collected from other different working conditions, which inevitably cause the domain shift problem [17]. Therefore, fault diagnosis models based on CNN have poor fault diagnosis performance under variable operating conditions.

A fault occurring in rotating machinery will introduce shock components in the vibration signal. The frequency of these shock components varies with rotating speed and the response on the temporal vibration signal is that the fault sensitive features appear on different time scales. Since the operating condition of rotating machinery often changes, we cannot judge which time scale features are sensitive to faults. Moreover, even under a constant rotating speed, the features that sensitive to different faults also appear in different frequency bands in the frequency domain. In other words, the features that sensitive to different faults appear on different time scales in the time domain. To extract fault sensitive features comprehensively, multiscale feature learning is a promising solution. However, the classical CNN is not designed specifically for vibration signals and lack of multiscale feature extraction ability [18]. Convolutional kernels with different widths have different local reception fields to learn features in different observation scales. All kernels in a convolutional layer of traditional CNN model typically have the same shape, so each convolutional layer can only extract features on a single scale and will lose useful information on other scales. The single scale features that extracted by CNN model are not only sensitive to faults but also sensitive to the change of operating condition. Therefore, the traditional CNN-based fault diagnosis models cannot capture effective features under variable operating conditions. This motivates us to use different shapes of kernels to learn multiscale features. Although several other works have achieved multiscale feature learning, but it is unclear which scales features are more sensitive to machinery faults. In general, different channels of a multiscale feature map may have different sensitivity levels to different types of faults under different operating conditions. So it motivates us to adaptively apply different weights to each channel of a multiscale feature map.

To extract robust fault sensitive features to withstand the change of operating condition, a new end-to-end deep learning model named adaptive weighted multiscale convolutional neural network (AWMSCNN) is proposed in this paper. The AWMSCNN integrates two innovations into the traditional CNN, which are the multiscale feature learning and the adaptive multiscale feature weighting. The innovations enable the AWMSCNN to have both fault discriminative ability and domain adaptive ability to effectively detect faults under variable operating conditions. The main contributions of this work are summarized as follows:

1) To boost the feature learning ability of the fault diagnosis model, several convolutional kernels with different widths are adopted to learn rich and complementary fault sensitive features on multiple time scales in parallel.
2) To extract robust fault sensitive features and improve the domain adaptive ability of the fault diagnosis model under variable operating conditions, an adaptive weight vector is introduced into the adaptive weighted multiscale convolutional (AWMSC) block to emphasize fault sensitive features and suppress features that are sensitive to operating conditions.
3) The AWMSCNN is an end-to-end deep learning model and directly takes raw vibration signals as input without any time-consuming and domain-variant handcrafted features. The AWMSCNN has a high efficiency to meet the online monitoring requirement.
4) The AWMSCNN is evaluated through comprehensive experiments on the wheelset bearing dataset and the CWRU dataset. The experiment results show that the AWMSCNN is a robust method for rotating machinery fault diagnosis under variable operating conditions.

The remainder of the paper is organized as follows: In section II, some related works are reviewed. In Section III, the proposed AWMSCNN model for rotating machinery fault diagnosis under variable operating conditions is described in detail. In Section IV, some experiments are implemented to validate the strong fault discriminative ability and domain adaptive ability of the proposed fault diagnosis model. Finally, conclusions are drawn in Section V.

II. RELATED WORKS

A. DFAULT DIAGNOSIS MODELS BASED ON CNN

CNN has many successful applications in the field of mechanical fault diagnosis. In some works, the raw sensor time series data has been preprocessed by some methods such as frequency transformation and time-frequency transformation before being input to the 2D CNN. Ding and He [19] proposed a deep CNN where wavelet packet energy images were used as input for spindle bearing fault diagnosis. A method of planetary gear fault diagnosis via feature image extraction based on multi central frequencies and vibration signal frequency spectrum is proposed in [20]. CNN can also directly address raw temporal signals without any time-consuming preliminary transformation. For instance, Jing *et al.* [21] proposed an adaptive multi-sensor data fusion method based on deep CNN for fault diagnosis with multivariate time-series data as the input. Qian *et al.* [22] construct a fault diagnosis framework called adaptive overlapping CNN to deal with one dimension (1D) raw vibration signals directly. Peng *et al.* [23] proposed a novel deeper 1D convolutional neural network (Der-1DCNN) with residual learning for fault diagnosis of wheelset bearings in high-speed trains. But, above general CNN based fault diagnosis methods only consider the single scale feature, which is not enough to capture effective features under variable operation conditions.

B. MULTISCALE FEATURE LEARNING MODELS BASED ON CNN

To capture complementary fault information from vibration signals at different time-scales, some multiscale feature extraction methods are proposed. Jiang *et al.* [18] proposed the MSCNN architecture, in which, a multiscale coarse-grained layer is introduced to represent the raw vibration signal over a range of scales, and results in multiple coarse-grained signals. Then, a traditional CNN is used to learn more useful and robust feature representations from each coarse-grained signal in parallel. In [24], the outputs of different feed-forward convolutional layers in CNN are integrated to further explore multi-scale feature information. Different from the traditional CNN structure that only utilizes features in the last convolutional layer, another MSCNN structure is proposed in [25], which integrates the last convolutional layer with the pooling layer before to form a multiscale layer. By the multiscale layer, the global and local features are maintained to enhance the network capacity. Though above works can extract multiscale features to some extent and can improve the model performance, but they do not take into account the different sensitivities of each scale feature to faults.

III. PROPOSED METHOD

The proposed AWMSCNN is implemented under the assumption that labeled training datasets under at least two operating conditions are available for model training. Then, the well trained model can be used for fault diagnosis under some unknown operating conditions. The key to the effectiveness of the model is that it can extract multiscale features and can adaptively modulate multiscale features using an adaptive weight vector. The extracted rich features are only sensitive to faults and insensitive to operating conditions.

The framework of the proposed AWMSCNN is shown in Fig. 1. The input of AWMSCNN is a segment of raw temporal vibration signal without any transformation, and the output is the predicted label indicating the health condition. The input sample $X \in R^l$ is denoted as $X = \{x_1, x_2, \ldots, x_l\}$, where $l$is the length of the input sample. The output Y is a one-hot vector. The fault diagnosis task is defined to obtain the label Y based on raw temporal vibration signal X using the AWMSCNN model. The AWMSCNN model has four parts: the denoising layer, the AWMSC block, the multiscale feature fusion layer, and the classification layers. The details of each part are elaborated in the following subsections.

A. DENOISING LAYER

In real industries, raw vibration signals measured from rotating machinery often contain various noise that produced by the complex working environment. Therefore, a denoising layer is applied to suppress background noise before extracting multiscale features. The denoising layer is actually a 1D convolutional (Conv1D) layer.

A Conv1D layer usually contains multiple channels of 1D convolution kernel, each channel is used to extract one type of feature. Assuming that the number of channels is c, the convolutional operation can be expressed by

$$F_i^d = f\left(K_i * X^{d-1} + b_i\right). \tag{1}$$

where $F_i^d, (i = 1, 2, \ldots c)$ is the ith channel feature in the feature map of the dth layer, K_i is the convolution kernel of the ith channel, X^{d-1} is the feature map of the $(d-1)$th layer, $\mathrm{b_i}$ is the bias, f is the ReLU [26] activation function. c channels can output a feature map F, which be expressed as

$$\boldsymbol{F} = [\boldsymbol{F_1^d}; \boldsymbol{F_2^d}; \ldots; \boldsymbol{F_c^d}]. \tag{2}$$

In the AWMSCNN, the parameter of the denoising layer is $c_1@k_1$. The width k_1 of the 1D kernel is wider than kernels in other layers. In addition, we choose a larger convolution stride s_1 than 1 used in the image recognition field to reduce the length of the extracted features. A max-pooling layer with a pooling length of 2 is adopted to capture more concise and local invariant features. Finally, the denoising layer returns a feature F_d with shape of $(c_1 \times l_1)$, where $l_1 = (l - k_1)/2s_1 + 1$.

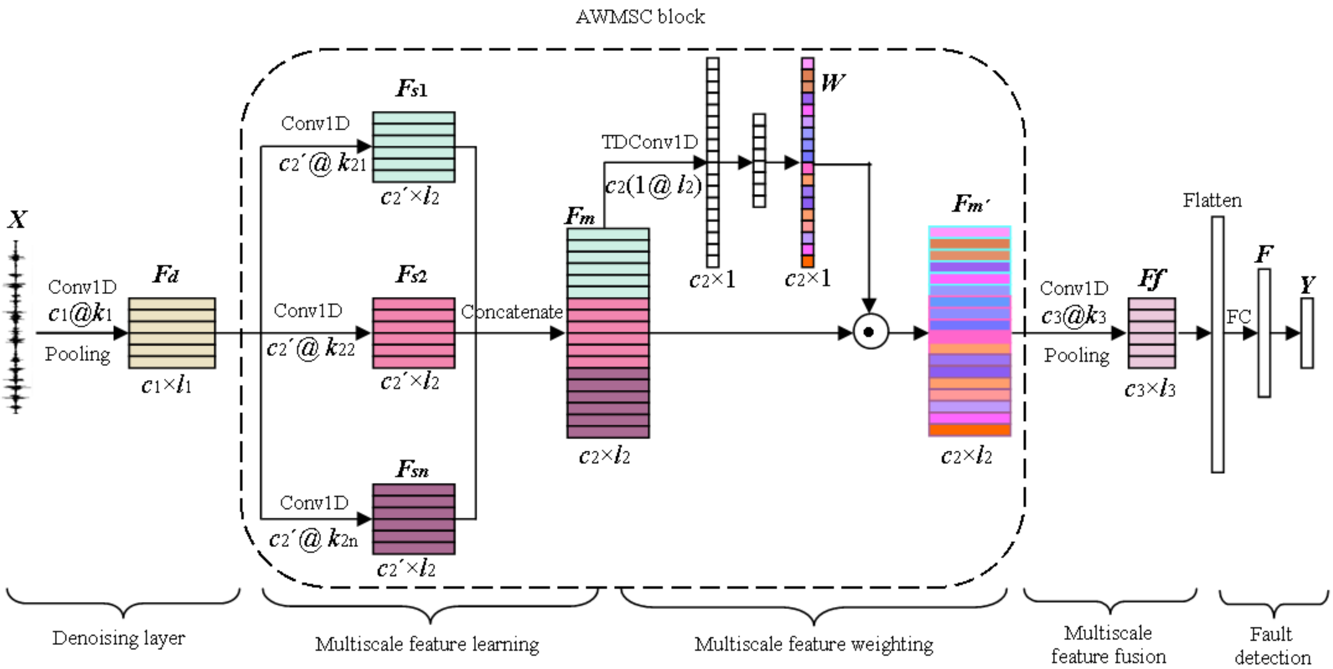


FIGURE 1. Framework of the proposed AWMSCNN model. The expression $c@k$ denotes the parameters in Conv1D layer. The channel is c and the kernel width is k. In the expression $t(c@k)$, t denotes the time steps of the TDConv1D layer. $c' \times l$ denotes the shape of a feature map, c' is the channel and l is the length of 1D feature in each channel. '·' denote the multiplication.

A wide kernel has a larger reception field than a narrow kernel and can capture lower frequency features. The wide kernel acts as a low-pass filter, so the convolutional layer with wide kernels can better suppress high frequency noise, which has been verified in [27] and [28]. Moreover, the convolutional layer with wide kernels and the pooling layer can turn the raw input vibration signal into features with shorter length, which can speed up the calculation of subsequent layers.

B. AWMSC BLOCK

The AWMSC block contains two processes: the multiscale feature learning process and the multiscale feature weighting process, which is shown in Fig. 1.

In the AWMSC block, the feature $\boldsymbol{F_d}$ extracted by the denoising layer is fed into n parallel Conv1D layers. The kernels of the n Conv1D layers have different widths and denoted as k_{2i}, $(i = 1, 2, \ldots n)$. The kernels with different widths act as filters with different scales of frequency domain resolution to simultaneously extract features of different frequency bands of the input signal. It is worth noting that there are c_2' channels in each of the n Conv1D layers. All Conv1D layers have the same convolution stride, which is set to 1. Each Conv1D layer in the AWMSC block can extract a feature map on one scale and the extracted feature map is denoted as $\boldsymbol{F_{Si}}$. In order to make the features of each scale have the same length, padding strategy is adopted here. The shape of $\boldsymbol{F_{Si}}$ is $(c_2' \times l_2)$, where $l_2 = l_1$.

Furthermore, considering that features of different channels on different scales have different fault sensitivity levels and the fault sensitivity level of each channel may also change under different operating conditions and different health conditions, an adaptive weight vector $\boldsymbol{W}$ is introduced to the AWMSC block to dynamically modulate multiscale features.

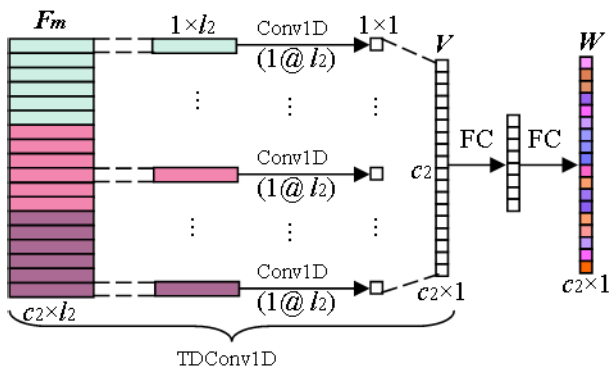


FIGURE 2. The process of adaptive calculation of the weight vector W.

As shown in Fig. 1, a concatenation layer is used to combine features of multiple scales along the channel dimension to form a multiscale feature map $\boldsymbol{F_m}$, which is expressed as

$$\boldsymbol{F_m} = [\boldsymbol{F_{S1}}; \boldsymbol{F_{S2}}; \cdots ; \boldsymbol{F_{Sn}}]. \tag{3}$$

The shape of $\boldsymbol{F_m}$ is $(c_2 \times l_2)$, where $c_2 = nc_2'$. The concatenation layer act a collector, which can aggregate features of all scales to form a multiscale feature set, so the concatenation layer can retain all the sensitive features coming from the convolutonal layer of different scales of kernel. Whereafter, a time-distributed Conv1D layer (TDConv1D) is applied to the $\boldsymbol{F_m}$ and returns a feature $\boldsymbol{V}$ with shape of $(c_2 \times 1)$. The structure of the TDConv1D is shown in Fig. 2. In the TDConv1D, a Conv1D layer is applied to each feature channel of the $\boldsymbol{F_m}$ simultaneously to learn the global information of each channel. The channel and the kernel width of the TDConv1D layer are set to 1 and l_2 respectively. Ultimately, the$c$th channel in the $\boldsymbol{F_m}$ is compressed into the cth element in the feature $\boldsymbol{V}$. Then, the weight vector $\boldsymbol{W}$ is calculated by the following equation referring to the excitation process in the SENet [29]:

$$\boldsymbol{W} = \sigma \left(W_2 f \left(W_1 \boldsymbol{V} + b_1 \right) + b_2 \right). \tag{4}$$

where σ refers to the sigmoid activation, which can implement the gating mechanism. The weight vector $\boldsymbol{W}$ is a 1D vector with shape of $(c_2 \times 1)$ and the cth element in $\boldsymbol{W}$ represents the fault sensitivity level of the cth channel feature of the $\boldsymbol{F_m}$.

The weight vector $\boldsymbol{W}$ is determined by model parameters and $\boldsymbol{F}_m$. After the proposed AWMSCNN model is well trained, model parameters have been fixed, so the weight vector $\boldsymbol{W}$ only changes with $\boldsymbol{F_m}$, i.e., the weight vector $\boldsymbol{W}$ changes with the model input $\boldsymbol{X}$. Therefore, the weight vector $\boldsymbol{W}$ has better adaptivity than a fixed vector. Finally, we modulate the multiscale feature $\boldsymbol{F_m}$ by

$$\boldsymbol{F}'_{\boldsymbol{m}} = \boldsymbol{F_m} \cdot \boldsymbol{W}. \tag{5}$$

where '$\cdot$ ' denotes channel-wise multiplication operation.

Since $\boldsymbol{W}$ can be dynamically adjusted by the model according the input signal, the multiscale feature $\boldsymbol{F_m}$ can be adaptively modulated to get optimized multiscale feature $\boldsymbol{F}'_{\boldsymbol{m}}$. $\boldsymbol{F_m}'$ and $\boldsymbol{F_m}$ have the same shape. Each element in $\boldsymbol{W}$ is a value between 0 and 1. The weight vector $\boldsymbol{W}$ acts as a gate, which can emphasize fault sensitive features and suppress features that are sensitive to operating conditions. Therefore, $\boldsymbol{F_m}'$is more sensitive to faults and insensitive to changes of operating condition.

C. MULTISCALE FEATURE FUSION LAYER

After the AWMSC block, a multiscale feature fusion layer is used to further extract multiscale fusion features from the optimized multiscale feature $\boldsymbol{F_m}'$. The goal of this layer is learn the dependency between features of different scales and different channels. The multiscale feature fusion layer is in fact a Conv1D layer with parameters of $c_3@k_3$, the stride is s_3. A max-pooling layer with pooling length of 2 is adopted to reduce the dimension of features. A feature $\boldsymbol{F_f}$ with shape of $(c_3 \times l_3)$ is returned by this layer, where $l_3 = (l_2 - k_3) / 2s_3 + 1$.

D. CLASSIFICATION LAYERS

At last, the feature $\boldsymbol{F_f}$ is flattened into a 1D feature vector and fed into a FC layer and a softmax layer. The softmax layer is defined as

$$P(y = j) = e^{\theta_j^T \boldsymbol{v}} \Big/ \sum\nolimits_{m=1}^{M} e^{\theta_j^T \boldsymbol{v}}. \tag{6}$$

where m is the label and M is the total number of labels. θdenotes parameters of the softmax layer.

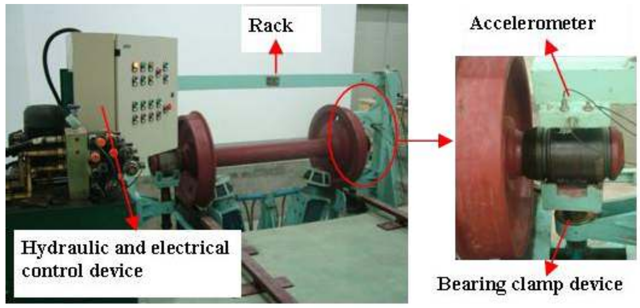


FIGURE 3. Train wheelset bearing test rig.

IV. EXPERIMENTS AND DISCUSSION

A. COMPARED MODELS

To prove the advantages of the proposed AWMSCNN model, the following deep models are implemented as comparisons in this study:

1) CNN: In the CNN, there are three pairs of Conv1D layers and pooling layers.
2) MSCNN: A multiscale coarse-grained layer is introduced to represent the raw vibration signal over a range of scales. Then two pairs of convolutional layers and pooling layers are used to extract features of different scales in parallel. More details of the MSCNN can be found in [18].
3) AWMSCNN-II: We remove the multiscale feature weighting process in the proposed AWMSCNN model and reserve other layers to form a new model. The new model is named as AWMSCNN-II.

All the experiments are carried out on a PC with an NVIDIA GeForce 1060Ti GPU and 8GB of RAM. The fault diagnosis accuracy is used for model performance evaluation and comparison.

B. CASE 1: EXPERIMENT RESULTS ON THE TRAIN WHEELSET BEARING DATASET AND PERFORMANCE ANALYSIS

1) DATA DESCRIPTION

The train wheelset bearing dataset was collected from a freight train wheelset rolling bearing fault diagnosis test rig. The test rig is shown in Fig. 3. The type of bearing is 197726. This study considers five bearing health conditions, including cage fracture (CF), roller crack (RC), roller indentation (RI), outer race peeling (OR), and normal (N). All the fault bearings used in this test rig were from the actual operating freight train. Bearings with each health condition were operated at four bearing rotating speeds (175 rpm, 270 rpm, 365 rpm, and 460 rpm). Vibration signals were collected using an accelerometer mounted vertically on the bearing housing with sampling frequency of 5.12 KHz. Every 10240 data points in the vibration signal were cut out with overlap to form a sample. There are 320 samples for each health condition under each rotating speed. The detailed description of the dataset is summarized in Table 1.

TABLE 1. Description of the train wheelset bearing dataset.

Health condition	Samples				Class label
	175 rpm	270 rpm	365 rpm	460 rpm	
CF	320	320	320	320	0
RC	320	320	320	320	1
RI	320	320	320	320	2
OR	320	320	320	320	3
N	320	320	320	320	4

TABLE 2. The information of each fault diagnosis task on the train wheelset bearing dataset under variable rotating speed. Unit: rpm.

Task name	A1	A2	A3	A4
Train	270, 365, 460	175, 365, 460	175, 270, 460	175, 270, 365
Test	175	270	365	460

2) DOMAIN ADAPTIVE ABILITY AGAINST VARIABLE ROTATING SPEED

The train wheelset bearing dataset is used to verify the performance of the AWMSCNN model under variable rotating speed. Four fault diagnosis tasks i.e. A1, A2, A3, and A4 are organized in this experiment for comprehensive verification. The task information is presented in Table 2. For instance, the task A1 denotes the scenario that the samples under rotating speeds of 270rpm, 365rpm, and 460 rpm are used for model training, and the samples under another rotating speed of 175rpm are used for testing. The other tasks follow the similar pattern. Different with the 10-fold cross-validation method used in [18], the training samples and testing samples in this work belong to different working conditions, and there exists the domain shift problem. This experiment method is more able to prove the generalization ability and the domain adaptive ability of the proposed model under variable operating conditions.

The task A2 is used for model parameter selection through cross-validated experiments. In our proposed model, a 4@64 Conv1D layer with stride size 8 is used in the denoising layer. The multiscale feature learning process in the AWMSC block consists of 4@2, 4@4, 4@6, and 4@8 Conv1D layers with stride size 1. The multiscale feature weighting process consists of a 16(1@636) TDConv1D layer and two FC layer with neuron number of 6 and 16. A 4@8 Conv1D layer with stride size 4 is used in the multiscale feature fusion layer. The dropout rate in the flatten layer is 0.2. To provide a fair comparison, all the compared models have the same model depth with the proposed AWMSCNN. In the CNN model, the parameters of three Conv1D layer are 4@8 with stride size 4, 4@4 with stride size 1 and 4@8 with stride size 4, respectively. The pooling size of each pooling layer is 2. In the MSCNN, four scales are used in the multiscale coarse-grained layer. The parameters of the two convolutional layers are 4@64 and 4@8, respectively. The parameters of the shared layers in the AWMSCNN-II and the AWMSCNN

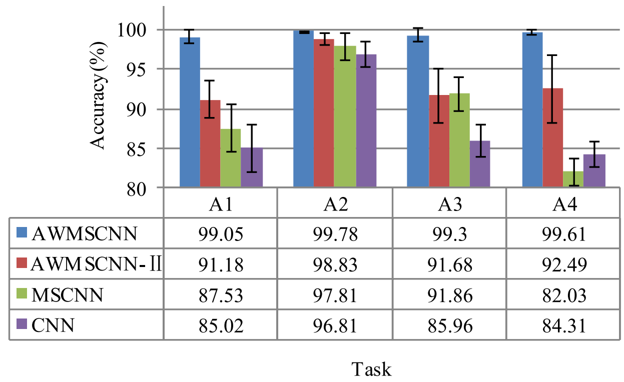


FIGURE 4. Performance of the proposed AWMSCNN in four tasks (A1-A4) compared with other models.

are same. The last two layers in each model are a FC layer with 50 neurons and a softmax layer with 5 neurons. Batch Normalization (BN) [30] is applied after each layer to accelerate the training process. The same input size, batch size, epoch number and gradient descent optimization algorithm parameters are used in all models. The batch size is 50, the epoch number is 50, and the optimization algorithm is Adadelta [31].

The testing fault diagnosis accuracies of four tasks using different models are shown in Fig. 4. The reported experiment results are averaged by 10 replicate experiments to reduce the effect of randomness. The error bar represents the standard deviation of 10 replicate experiments and shows the stability of the model.

It can be seen in Fig. 4 that the proposed AWMSCNN achieves the best performance in each task and the results in four tasks are all above 99%, which shows that the AWMSCNN has good fault discriminative ability and domain adaptive ability against variation of rotating speed. The standard deviations of the AWMSCNN in all tasks are small, which indicates that the AWMSCNN has good stability. So the AWMSCNN can be used for fault diagnosis in the scenario that the rotating speed of the testing data is not included in the training dataset. In addition, AWMSCNN and AWMSCNN-II both perform better than CNN in all four tasks. The MSCNN performs better than CNN only in task A1, A2, and A3, but performs worse than CNN in task A4. Based on the above results, we can draw two conclusions. The first conclusion is that multiscale feature learning models perform better that CNN. The other one is that the proposed adaptive weighted multiscale feature learning method using different shapes of kernels can extract more robust and fault sensitive multiscale features than MSCNN using the coarse-grained layer. Although AWMSCNN-II, MSCNN, and CNN can achieve high diagnosis accuracy in task A2, they perform poorly in other tasks indicating that these three models have poor generalization ability and domain adaptive ability. Comparing the results of AWMSCNN and AWMSCNN-II, it can be concluded that the adaptive multiscale feature weighting process plays an important role in improving fault diagnosis

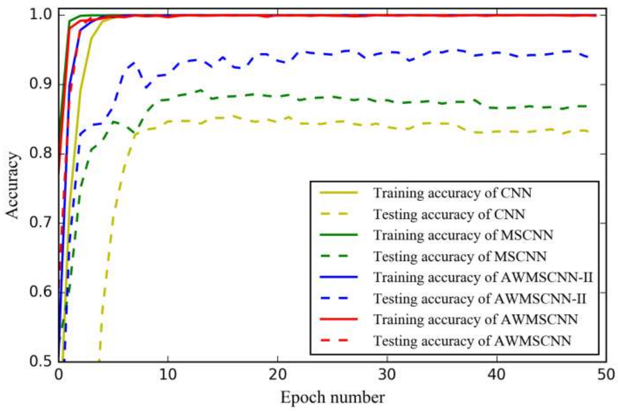


FIGURE 5. The convergence performances of 4 different models in task A1.

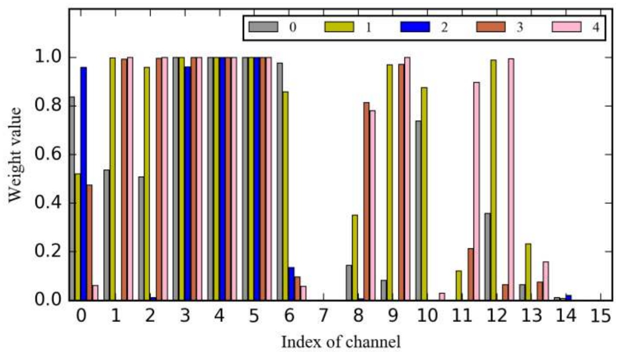


FIGURE 6. Weight vector visualization for five health conditions under the rotating speed of 460rpm.

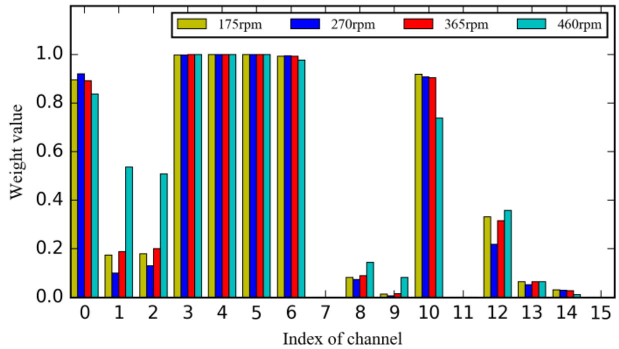


FIGURE 7. Weight vector visualization for the CF condition under four rotating speeds.

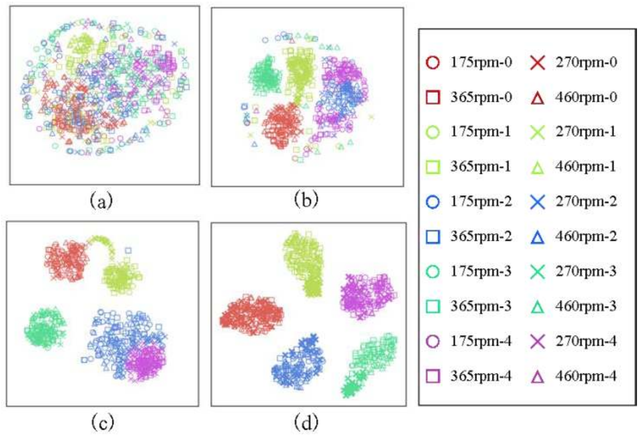


FIGURE 8. Feature visualization of the AWMSCNN for different health conditions under different rotating speeds: (a) feature F_m, (b) feature $F_m{}'$, (c) feature F_f , (d) finally feature F . 25% data are randomly selected to be presented. *a* rpm-*b* denotes the class label of the feature is *b* and the rotating speed is *a* rpm.

accuracy and improving the domain adaptive ability of the model against variation of rotating speed. The convergence performances of 4 different models are shown in Fig. 5. From Fig. 5, we can see that the proposed AWMSCNN converges faster than other models.

3) VISUALIZATION OF THE WEIGHT VECTOR AND FEATURES

To further verify the adaptability of the weight vector in the proposed model, the learned weight vectors are visualized. The weight vector includes weight values of 16 channels. First, the weight vectors of five health conditions under the rotating speed of 460rpm are presented in Fig. 6, where different colors represent different health conditions described in Table 1. From Fig. 6, we can see that the weight vectors calculated from samples of different health conditions at the same rotating speed are different. Specifically, the weight values of most channels vary with health condition. So, we can conclude that the weight vector can be adaptively adjusted by the model when the health condition changes under the same rotating speed. Then, the weight vectors of the CF condition under four rotating speeds are presented in Fig. 7, where different colors represent different rotating speeds described in Table 1. It can be observed in Fig. 7 that the weight values of most channels in the weight vector vary with rotating speed. So, we can also conclude that the weight vector can be adaptively adjusted by the model when the rotating speed changes.

In order to further prove that the weight vector can improve the domain adaptability and the robustness of the model under variable speed conditions, the t-SNE method [32] is used to show the feature maps learned by the proposed model. The two-dimensional feature maps are shown in Fig. 8, where, features of different fault types are distinguished by different colors and features of different rotating speeds are distinguished by different marks.

It can be seen in Fig. 8(a) that although after the multiscale feature extraction process, $\boldsymbol{F_m}$ of the same fault type are still not clustered together, and $\boldsymbol{F_m}$ of different fault types are still mixed together. It is shown in Fig. 8(b) that features $\boldsymbol{F_m}'$ of the same fault type under different rotating speeds are clustered to some extent after the adaptive multiscale feature weighting process. Specifically, except for fault types 2 and 4, other fault types are basically separated. It can be explained that the weight vector can be adaptively adjusted to emphasize fault-discriminative features and suppress features

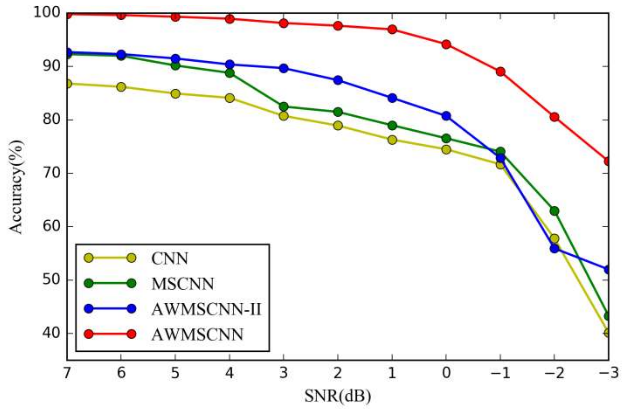


FIGURE 9. Performance comparisons of different models in noisy environments with different SNRs.

that sensitive to operating conditions, so that the features of the same fault type at different speeds can be modulated by the weight vector to have the same distribution. Then, the multiscale feature fusion layer can further extract more abstract features $\boldsymbol{F_f}$ from $\boldsymbol{F_m}'$. We can see in Fig. 8(c) that the multiscale feature fusion layer further clusters the features $\boldsymbol{F_f}$ of the same fault type and further separates the features of different fault types. At last, after the FC layer, the features of different fault types are completely separated, which is shown in Fig. 8(d). The proposed AWMSCNN model can finally learn robust and fault discriminative features, and has good domain adaptive ability and fault discriminative ability under variable operating conditions.

4) ROBUSTNESS AGAINST NOISE

In order to simulate the noisy working environment in real industries, we inject additive Gaussian white noise to the raw vibration signals to construct noisy signals with different signal-to-noise ratios (SNRs). The SNR is defined as

$$SNR_{dB} = 10log_{10}\left(\frac{P_{signal}}{P_{noise}}\right). \tag{7}$$

where P_{signal} and P_{noise} denote the power of the raw signal and the injected noise, respectively.

The task A3 is implemented using noisy signals with different SNRs to evaluate the robustness of the AWMSCNN against noise. The evaluation results for different models are shown in Fig. 9. Obviously, the proposed AWMSCNN outperforms other models and has more than 70% accuracy with each SNR. When the SNR is over 0dB, the AWMSCNN can achieve over 94% accuracy. Comparing the performance of AWMSCNN and AWMSCNN-II, we can conclude that the adaptive multiscale feature weighting process in the AWMSCNN can improve the anti-noise ability of the model. In addition, all multiscale models are outperform the CNN, which indicates that multiscale feature leaning can extract more robust features form noisy input signal than single scale CNN. The above results mean that the proposed AWMSCNN can be used for fault diagnosis under variable rotating speed in real noisy industrial environment.

TABLE 3. Description of the CWRU dataset.

Health condition	fault diameters (in.)	Samples				Class label
		0 hp	1 hp	2 hp	3 hp	
N	-	600	600	600	600	0
OF	0.007	600	600	600	600	1
	0.014	600	600	600	600	2
	0.021	600	600	600	600	3
IF	0.007	600	600	600	600	4
	0.014	600	600	600	600	5
	0.021	600	600	600	600	6
BF	0.007	600	600	600	600	7
	0.014	600	600	600	600	8
	0.021	600	600	600	600	9

TABLE 4. The information of each fault diagnosis task on the CWRU dataset under variable load. Unit: hp.

Task name	B1	B2	B3	B4
Train	1, 2, 3	0, 2, 3	0, 1, 3	0, 1, 2
Test	0	1	2	3

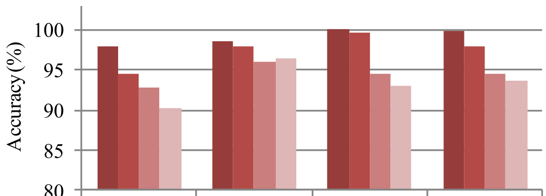
C. CASE 2: EXPERIMENT RESULTS ON THE CWRU DATASET AND PERFORMANCE ANALYSIS

1) DATA DESCRIPTION

The CWRU motor bearing dataset is provided by the Bearing Data Center of Case Western Reserve University [33]. There are four different health conditions: normal condition (N), outer race fault (OF), inner race fault (IF), and ball fault (BF). Each fault type contains three fault diameters: 0.007 inch, 0.014 inch, and 0.021 inch. Therefore, the CWRU dataset contains 10 bearing health conditions. The vibration signals of each health condition are collected from the drive end of the motor under four loads (0 hp, 1 hp, 2 hp, and 3 hp) with sampling frequency of 12 kHz, where the same health condition under different loads is treated as one class. Every 1000 data points in the vibration signal are cut out with overlap to form a sample. There are 600 samples for each health condition under each load. The detailed description of the CWRU dataset is summarized in Table 3.

2) DOMAIN ADAPYIVE ABILITY AGAINST VARIABLE LOAD

The CWRU dataset is used to verify the domain adaptive ability of the AWMSCNN against variation of load. Four fault diagnosis tasks i.e. B1, B2, B3, and B4 are organized in this experiment for comprehensive verification. The task information is presented in Table 4. For instance, the task B1 denotes the scenario that the samples under loads of 1hp, 2hp, and 3hp are used for model training, and the samples under another load of 0hp are used for testing. The other tasks follow the similar pattern. Each task contains 18000 samples for training and 6000 samples for testing.



	B1	B2	B3	B4
AWMSCNN	97.97	98.57	99.98	99.75
AWMSCNN-II	94.55	97.82	99.72	97.98
MSCNN	92.75	96.03	94.5	94.55
CNN	90.33	96.43	92.97	93.58

Task

FIGURE 10. **Performance of the proposed AWMSCNN in four tasks (B1-B4) compared with other models.**

TABLE 5. **The cost time of 4 different models on two dataset.**

Methods	Train wheelset bearing dataset		CWRU dataset	
	Training time (s)	Testing time (ms)	Training time (s)	Testing time (ms)
AWMSCNN	1.7289	2.7551	3.3501	1.7302
AWMSCNN-II	1.5274	2.1243	2.7294	1.2710
MSCNN	2.6889	3.6440	4.5313	2.1177
CNN	1.4151	2.0607	2.4961	1.2348

As Fig. 10 shows, the proposed AWMSCNN achieves the best performance compared with other models in each task. Fault diagnosis accuracies of AWMSCNN in four tasks are all above 97.97, which further validates its superior fault discriminative ability and domain adaptive ability against variation of load. CNN performs poorly than other multiscale models in each task, which indicates that extracting multiscale features is essential for the model to extract more robust features in variable load scenarios. Comparing results of AWMSCNN and AWMSCNN-II, we can conclude that the adaptive multiscale feature weighting process can further improve the domain adaptive ability of the AWMSCNN. In this experiment, the time required by the AWMSCNN to calculate each test sample is just 2ms, which proves that the AWMSCNN can be used for real-time bearing fault diagnosis.

D. TIME CONSUMPTION

The training time of one epoch and the testing time of one testing sample spent by different models on two dataset are listed in Table 5. It should be declared again that to provide a fair comparison, all the compared models have the same model depth with the proposed AWMSCNN. From Table 5, we can find that the AWMSCNN, AWMSCNN-II, and MSCNN are all need more training time and testing time than CNN, which can be explained that multiscale feature learning models will bring more parameters to be trained and therefore require more computing time. Since the model is trained offline, the training time is not a critical aspect of evaluating model performance. But, the testing time is a key factor affecting the performance of the well-trained model used for online fault diagnosis. In term of the testing time, although the AWMSCNN need more time than AWMSCNN-II and CNN, the testing time spent by AWMSCNN is just 2.7551ms on the train wheelset bearing dataset and just 1.7302ms on the CWRU dataset, which proves that the AWMSCNN can be used for online fault diagnosis.

V. CONCLUSION

This paper focus on intelligent fault diagnosis of rotating machinery under variable operating conditions and proposes a novel end-to-end AWMSCNN model. The proposed AWMSCNN directly takes raw vibration signals as input without any handcrafted features and can automatically extract fault sensitive features. The main innovations of the proposed model are multiscale feature learning and adaptive multiscale feature weighting, which can help the model extract robust features and improve the domain adaptive ability of the model against noise, variable rotating speed and variable load. In the experiments, the AWMSCNN outperforms the basic single-scale CNN model and other multiscale models. In addition, the weight vector and features are visualized. We found that when the rotating speed or the health condition changes, the weight vector can be adaptively adjusted by the model, so that the model can finally extract robust and discriminative features. The experiment results show that the proposed AWMSCNN can be used for fault diagnosis under variable operating conditions in real industrial scenario.

Considering that it is often difficult to obtain labeled data for model training in real industries, in our further work, we will further combine the AWMSCNN with transfer learning strategy to achieve good performance when labeled data is rarely available for model training.

REFERENCES

[1] Z. Zhao, B. Qiao, S. Wang, Z. Shen, and X. Chen, "A weighted multi-scale dictionary learning model and its applications on bearing fault diagnosis," *J. Sound Vib.*, vol. 446, pp. 429–452, Apr. 2019.

[2] S. Pang and X. Yang, "A cross-domain stacked denoising autoencoders for rotating machinery fault diagnosis under different working conditions," *IEEE Access*, vol. 7, pp. 77277–77292, 2019. doi: 10.1109/ACCESS.2019.2919535.

[3] J. Chen, Z. Li, J. Pan, G. Chen, and Y. Zi, J. Yuan, B. Chen, and Z. He, "Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vols. 70–71, pp. 1–35, Mar. 2016.

[4] Z. Meng, X. Guo, Z. Pan, D. Sun, and S. Liu, "Data segmentation and augmentation methods based on raw data using deep neural networks approach for rotating machinery fault diagnosis," *IEEE Access*, vol. 7, pp. 79510–79522, 2019. doi: 10.1109/ACCESS.2019.2923417.

[5] M. Zhao, M. Kang, B. Tang, and M. Pecht, "Multiple wavelet coefficients fusion in deep residual networks for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 66, no. 6, pp. 4696–4706, Jun. 2019. doi: 10.1109/TIE.2018.2866050.

[6] D. Liu, W. Cheng, and W. Wen, "Rolling bearing fault diagnosis via ConceFT-based time-frequency reconfiguration order spectrum analysis," *IEEE Access*, vol. 6, pp. 67131–67143, 2018. doi: 10.1109/ACCESS.2018.2877711.

[7] Y. Guo, T.-W. Liu, J. Na, and R.-F. Fung, "Envelope order tracking for fault detection in rolling element bearings," *J. Sound Vib.*, vol. 331, no. 25, pp. 5644–5654, Dec. 2012. doi: 10.1016/j.jsv.2012.07.026.

[8] Y. Wang, B. Tang, L. Meng, and B. Hou, "Adaptive estimation of instantaneous angular speed for wind turbine planetary gearbox fault detection," *IEEE Access*, vol. 7, pp. 49974–49984, 2019. doi: 10.1109/ACCESS.2019.2908192.

[9] S. Lu, Y. Qin, J. Hang, B. Zhang, and Q. Wang, "Adaptively estimating rotation speed from DC motor current ripple for order tracking and fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 3, pp. 741–753, Mar. 2019. doi: 10.1109/TIM.2018.2852978.
[10] J. Niu, S. Lu, Y. Liu, J. Zhao, and Q. Wang, "Intelligent bearing fault diagnosis based on tacholess order tracking for a variable-speed AC electric machine," *IEEE Sensors J.*, vol. 19, no. 5, pp. 1850–1861, Mar. 2019. doi: 10.1109/JSEN.2018.2883955.
[11] L. Wang and J. Xiang, "A two-stage method using spline-kernelled chirplet transform and angle synchronous averaging to detect faults at variable speed," *IEEE Access*, vol. 7, pp. 22471–22485, 2019. doi: 10.1109/ACCESS.2019.2898884.
[12] F. Li, R. Li, L. Tian, L. Chen, and J. Liu, "Data-driven time-frequency analysis method based on variational mode decomposition and its application to gear fault diagnosis in variable working conditions," *Mech. Syst. Signal Process.*, vol. 116, pp. 462–479, Feb. 2019. doi: 10.1016/j.ymssp.2018.06.055.
[13] Y. Li, K. Feng, X. Liang, and M. J. Zuo, "A fault diagnosis method for planetary gearboxes under non-stationary working conditions using improved Vold-Kalman filter and multi-scale sample entropy," *J. Sound Vib.*, vol. 439, pp. 271–286, Jan. 2019. doi: 10.1016/j.jsv.2018.09.054.
[14] H. Qiao, T. Wang, P. Wang, S. Qiao, and L. Zhang, "A time-distributed spatiotemporal feature learning method for machine health monitoring with multi-sensor time series," *Sensors*, vol. 18, no. 10, p. 2932, Sep. 2018. doi: 10.3390/s18092932.
[15] 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. doi: 10.1016/j.ymssp.2018.05.050.
[16] X. Li, W. Zhang, and Q. Ding, "Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks," *IEEE Trans. Ind. Electron.*, vol. 66, no. 7, pp. 5525–5534, Jul. 2019. doi: 10.1109/TIE.2018.2868023.
[17] X. Li, W. Zhang, and Q. Ding, "A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning," *Neurocomputing*, vol. 310, pp. 77–95, Oct. 2018.
[18] G. Q. 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. 2018. doi: 10.1109/TIE.2018.2844805.
[19] 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. doi: 10.1109/TIM.2017.2674738.
[20] Y. Li, G. Cheng, Y. Pang, and M. Kuai, "Planetary gear fault diagnosis via feature image extraction based on multi central frequencies and vibration signal frequency spectrum," *Sensors*, vol. 18, no. 6, p. 1735, Jun. 2018. doi: 10.3390/s18061735.
[21] L. Jing, T. Wang, M. Zhao, and P. Wang, "An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox," *Sensors*, vol. 17, no. 2, p. 414, Feb. 2017. doi: 10.3390/s17020414.
[22] W. Qian, S. Li, J. Wang, Z. An, and X. Jiang, "An intelligent fault diagnosis framework for raw vibration signals: Adaptive overlapping convolutional neural network," *Meas. Sci. Technol.*, vol. 29, no. 9, Sep. 2018, Art. no. 095009. doi: 10.1088/1361-6501/aad101.
[23] 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. 7, pp. 10278–10293, 2019. doi: 10.1109/ACCESS.2018.2888842.
[24] X. Li, W. Zhang, and Q. Ding, "Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction," *Rel. Eng. Syst. Saf.*, vol. 182, pp. 208–218, Feb. 2019. doi: 10.1016/j.ress.2018.11.011.
[25] J. Zhu, N. Chen, and W. Peng, "Estimation of bearing remaining useful life based on multiscale convolutional neural network," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3208–3216, Apr. 2019.
[26] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, Fort Lauderdale, FL, USA, 2011, pp. 315–323.
[27] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, Feb. 2017. doi: 10.3390/s17020425.
[28] 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," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–453, Feb. 2018. doi: 10.1016/j.ymssp.2017.06.022.
[29] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published. doi: 10.1109/TPAMI.2019.2913372.
[30] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Feb. 2015, *arXiv:1502.03167*. [Online]. Available: https://arxiv.org/abs/1502.03167
[31] M. D. Zeiler, "ADADELTA: An adaptive learning rate method," Dec. 2012, *arXiv:1212.5701*. [Online]. Available: https://arxiv.org/abs/1212.5701
[32] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[33] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mech. Syst. Signal Process.*, vols. 64–65, pp. 100–131, Dec. 2015. doi: 10.1016/j.ymssp.2015.04.021.

HUIHUI QIAO was born in Shijiazhuang, Hebei, China, in 1991. She received the B.S. degree in measurement and control technology and instrumentation from East China Jiaotong University, Nanchang, Jiangxi, China, in 2013, and the M.S. degree in measurement technology and automation equipment from Shijiazhuang Tiedao University, Shijiazhuang, in 2016. She is currently pursuing the Ph.D. degree in mechanical engineering with Tianjin University, Tianjin, China.

Her research interests include the advanced signal processing algorithms, intelligent fault diagnosis and prognostic, data mining, cloud computing, and deep learning for machine health monitoring.

TAIYONG WANG was born in Changzhi, Shanxi, China, in 1962. He received the B.S. degree in mechanical engineering from the Shenyang University of Technology, Shenyang, Liaoning, China, in 1983, and the M.S. and Ph.D. degrees in mechanical engineering from Tianjin University, Tianjin, China, in 1986 and 1995, respectively.

Since 1995, he has been a Professor with the School of Mechanical Engineering, Tianjin University, where he is currently the Chairman of the International Conference on Functional Manufacturing and Mechanical Dynamics (ICFMT&MD), the Director of the Tianjin Engineering Research Center of NC Technology, and the Director of the Institute of Digital Manufacturing and Measurement and Control Technology. His research interests include intelligent fault diagnosis and dynamic measurement and control technology, big data and deep learning, reconfigurable CNC system, and intelligent manufacturing system.

PENG WANG was born in Tianjin, China, in 1987. He received the B.S. degree in bioinformatics from the Huazhong University of Science and Technology, Wuhan, Hubei, China, in 2010, the M.S. degree in measurement technology and automation equipment from Shijiazhuang Tiedao University, Shijiazhuang, Hebei, in 2014, and the Ph.D. degree in mechanical engineering from Tianjin University, Tianjin, China, in 2018.

Since 2018, he has been a Lecturer with the School of Mechanical Engineering, Tianjin University. His research interests include time–frequency signal processing method, intelligent fault diagnosis, and development of intelligent diagnosis instruments.

LAN ZHANG was born in Shijiazhuang, Hebei, China, in 1995. She received the B.S. degree in mechanical and electrical engineering from the North University of China, Taiyuan, China, in 2017. She is currently pursuing the master's degree in mechanical engineering with Tianjin University, Tianjin, China.

Her research interests include the advanced signal processing algorithms, intelligent fault diagnosis, and deep learning for machine health monitoring.

MINGDA XU was born in Sanming, Fujian, China, in 1995. He received the B.S. degree in mechanical design and manufacturing and automation from the Shenyang University of Technology, Shenyang, Liaoning, China, in 2017. He is currently pursuing the M.S. degree in mechanical engineering with Tianjin University, Tianjin, China.

His research interests include the advanced signal processing algorithms, data mining, cloud computing, and deep learning for machine health monitoring.