Research on a Rolling Bearing Fault Detection Method With Wavelet Convolutional Deep Transfer Learning

MENGLIANG LIAO, CHANG LIU, CONG WANG, AND JIANWEI YANG

Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming 650093, China

Corresponding author: Chang Liu (liuchang3385@gmail.com)

This work was supported in part by the National Major Science and Technology Projects of China under Grant 2018YFB1306100, and in part by the Scientific Research Foundation for Introduced Talent Kunming University of Science and Technology under Grant KK23201801048.

ABSTRACT

Many deep learning models for fault diagnosis have not considered the prior diagnosis knowledge of the rolling bearing. Moreover, some measuring locations cannot collect adequate data to diagnose due to equipment size or installation space problems. This paper proposes a wavelet convolutional deep transfer learning model for rolling bearing fault detection on cross-measurement points. A new convolution layer includes a redesigned convolution kernel and a new energy pooling layer. The convolution kernel is designed based on wavelet construction for mining time-frequency characteristics. The energy pooling layer is proposed to extract the energy of different frequency bands. The associated location’s fault information has been transferred to domain knowledge to enhance the target features. Different domain features have adaptive matches based on multiple kernel variants of maximum mean discrepancy. The experimental results demonstrate that the accuracy of fault detection can reach 99.73%, and the robustness of the proposed method is also verified.

INDEX TERMS

Backpropagation, ball bearings, convolutional neural networks, deep learning, fault detection, feature extraction, frequency-domain analysis, kernel, knowledge transfer, wavelet transforms.

I. INTRODUCTION

Rolling bearings are essential components of rotating machinery [1]. Bearing failure diagnosis is crucial for equipment health [2]. The sensor cannot be installed in the most sensitive position due to equipment size and installation conditions problems. These issues will restrict the bearing status data acquisition and affect the accuracy of the equipment status assessment and fault diagnosis [3], [4]. Therefore, solving the problem of bearing cross-measurement transfer diagnosis has excellent engineering significance.

Transfer learning theory provides new methods for solving these problems. Transfer learning is a method that uses existing knowledge to solve problems in different but related fields. Transfer learning methods have successfully been applied in image recognition [5], speech recognition [6], text recognition [7], and other fields. Research on the combination of transfer learning and fault diagnosis has gradually become a research hotspot. Considerable research has been focused on fault diagnosis under variable working conditions by using transfer learning. Li et al. proposed a novel transfer learning method for diagnostics based on deep learning. The diagnostic knowledge learned from multiple rotating machines was transferred to the target equipment with domain adversarial training [8]. Qian et al. constructed a three-stage deep fault diagnosis network utilizing Adaptive Batch Normalization (AdaBN), which was highly efficient without the target datasets during training [9]. Xiao et al. proposed a method based on a Convolutional Neural Network (CNN) architecture embedded with the Maximum Mean Discrepancy (MMD). This method realized the recognition of the motor’s health status under variable operating conditions [10].

Recently, deep learning was introduced into transfer learning to improve learning performance and diagnosis accuracy. Shao et al. proposed an adversarial domain adaptation method based on deep transfer learning. This approach used a deep residual network to process a time-frequency image to achieve cross-domain diagnosis [11].
Wen et al. proposed a three-layer Sparse Auto-Encoder (SAE) to extract the features from raw data, and then the MMD term was applied to minimize the discrepancy penalty of two domains. This encoder was proven effective when transferred between multiple locations [12]. Zheng et al. elaborated a diagnosis scheme combining a priori diagnosis knowledge and a Deep Domain Generalization network for Fault Diagnosis (DDGFD). The superiority of DDGFD was validated on cross-domain tasks organized using broad bearing datasets [13]. Jalayer et al. proposed a new feature engineering model which combined Fast Fourier Transform (FFT), Continuous Wavelet Transform (CWT), and statistical features of raw signals [14]. Grezmai et al. proposed an explainable CNN for gearbox fault diagnosis based on Layer-wise Relevance Propagation (LRP), which obtained a high classification accuracy [15].

However, there are some problems in the above research. 1) Many models for fault diagnosis have not considered the prior diagnosis knowledge of the rolling bearing. Feature extraction in the early stage can only rely on the learning ability of the network. The lack of consideration of extracting domain invariant features from different domains limits the generalization ability of the model to a certain level. 2) Much research has been focused on fault diagnosis under variable working conditions. However, there have been few studies regarding diagnosis information transfer between different domains based on cross locations coupling information. 3) The cross-domain diagnosis method based on a CNN converts one-dimensional (1D) signals into image data and uses the model in image recognition for classification. Such methods require a preprocessing stage and the insufficient consideration of the original vibration signal.

This paper proposes a Wavelet Convolution Deep Transfer Learning (WDTL) method for rolling bearing fault detection, which integrates a multiscale wavelet structure and deep transfer learning theory. The proposed neural network is improved based on AlexNet. The convolution layer is redesigned by a new convolution kernel and a new energy pooling layer. We use the fine-tuning and domain adaptive match method of the improved model and apply it to the target domain. The main contributions are as follows:

1) The model improves the ability of the CNNs to mine vibration signal characteristics. A wavelet convolution kernel is proposed to improve the network framework. The improved network can effectively dig time-frequency local characteristics of the raw vibration data. Energy information of different frequency bands is extracted by energy pooling. The purpose is to explore the domain features of signals in different domains and improve the ability of the CNNs to explore signal energy features.

2) The Multiple Kernel Variants of Maximum Mean Discrepancy (MK-MMD), which integrates into the deep transfer learning framework, can reduce the difference distribution between the source domain and target domain. The fault information of the associated location is converted into domain knowledge to realize cross-domain fault detection.

3) The effectiveness of the proposed method is verified by experiment.

The remainder of this paper is organized as follows. Section II briefly introduces the CNNs structure and basic theory. Section III is dedicated to describing the proposed WDTL. Two transfer learning cases and corresponding analyses are conducted in Section IV. Finally, conclusions and discussions are summarized in Section V.

II. THEORETICAL BACKGROUND

The main structures of a CNN are the convolutional layer, pooling layer, and fully connected layer [16]. AlexNet is one of the leading CNN architectures. This model performed well in the ImageNet contest and has proven to have higher performance than previous methods in image classification tasks [17]. This architecture consists of 5 convolutional layers and 3 Fully Connected Layers (FC). The time-frequency image fault diagnosis based on the AlexNet framework has achieved good results [18], [19], but has difficulty processing 1D signals, and feature extraction depends on the learning ability of the network, which lacks prior knowledge of the rolling bearing diagnosis. To directly apply the two-dimensional (2D) CNN model, signal processing methods such as a wavelet packet and continuous wavelet transform are used in the field of fault diagnosis to upgrade the 1D vibration signal [20]. Another method is to adjust 1D data to a 2D structure by constructing different data structures [21]. Unknown false information may be introduced due to an increasing signal dimension is designed. Therefore, in this paper, an end-to-end CNN model. Considering the basic architecture, convolution kernel, and pooling method, the AlexNet-based model is improved to adapt to 1D vibration signals to avoid the unknown false information that may be introduced due to a signal upgrade.

A. CONVOLUTION LAYER

First, a 1D convolutional layer is used as the basis of the CNN model in this paper. The convolutional layer of the CNN convolves and pools the input data layer by layer, and extracts the topological structure features contained in the input data layer by layer. The essence of the convolution kernel of the convolution operation is to extract features from the input data. In signal processing, a convolution in the time domain corresponds to a multiplication in the frequency domain. The convolution between the original vibration data and the convolution kernel is essentially the selection of frequency domain information. Therefore, the convolution layer of 1D-CNN extracts different frequency domain vibration input information through convolution operations and extracts higher-dimensional features through layered superposition. The output calculation of the neuron in the convolutional layer is shown in Equation (1)

$$g(i) = \sum_{x=1}^{m} \sum_{y=1}^{n} \sum_{z=1}^{p} a_{x,y,z} \times w'_{x,y,z} + b'$$

(1)
where $i$ represents the $i$-th convolution kernel (layer $l$), $g(i)$ is the feature map learned by the $i$ convolution kernel, $a_{k,y,z}$ is the value of node $(x, y, z)$ in the convolution kernel, and $b^i_l$ is the bias of the convolution kernel. $m, n, p$ represents the dimension of the input data, $m$ and $n$ are the length and width of the receiving field, respectively, and $p$ is the number of feature maps in the upper layer. The latter two dimensions need to be simplified for 1D time-domain signals.

Since the convolution kernel is mostly linear, it is more suitable for learning the potential features of linear separability. For the diagnosis of rolling bearing vibration signals, the frequency component of the signal in the frequency domain is more important than the performance in the time domain [22]. Therefore, an improvement in the traditional convolution kernel can make the CNN model more suitable for vibration signals. The time-frequency method based on a wavelet transform can effectively extract the features of the bearing vibration signals. This method is widely used in rotating mechanical bearing fault diagnosis [23].

The essence of the wavelet analysis method is to choose a different scale basis function $\psi(f, t)$ and convolve it with $f(t)$, where convolution is the translational invariance in time. By constructing a suitable basis function, a more ideal time-frequency analysis result can be obtained.

$$S(a, \tau) = f(t) \ast \psi(f, t) = \int f(\tau)\psi(\tau − t, a) dt$$

In digital signal processing, for the discrete time series $x[n]$ and $h[n]$, the convolution can be expressed as

$$y[n] = x[n] \ast h[n] = \sum_{k=-\infty}^{\infty} x[k] \cdot h[n−k]$$

It can be seen in Equations (1) to (3) that when the convolution kernel of the first layer is taken as different scale basis functions, its convolution with the original signal can achieve the effect of wavelet analysis. In this way, the time-domain signals are converted to a frequency domain analysis, and different convolution kernels obtain signals with different frequency bands. The entire model will excvate the bearing fault information from the frequency domain to identify the bearing state more efficiently. Based on the classic CNN, the author uses the wavelet multiscale structure to initialize the first layer of the convolution kernel. We use the wavelet's ability to excvate signal time-frequency domain features to provide a basis for extracting associated fault features based on wavelet frequency band energy information. Finally, the model is combined with prior knowledge in the field of signal processing.

In this paper, a convolution kernel with a series of wavelet structures is constructed based on the Daubechies wavelet to obtain multiple analysis results. We use the ability of wavelet analysis to mine the time-frequency domain features of the signal to obtain the effective associated fault features [24]. At the same time, gradient update learning is allowed. Compared with the general features of extracting data from a random convolution kernel, the wavelet convolution kernel proposed in this paper is more targeted for the extraction of device features. It can effectively improve the accuracy and robustness of the diagnostic model.

Daubechies wavelet is a wavelet function constructed by Inrid Daubechies, which has the characteristic of orthogonality:

$$Hypothesis \ P(y) = \sum_{k=0}^{N−1} c_k^{N−1+k} y^k, \ \text{where } c_k^{N−1+k} \ \text{is the binomial coefficient,}$$

$$|m_0(\omega)|^2 = (\cos \frac{\omega}{2})^N P(\sin \frac{\omega}{2}) (4)$$

where $m_0(\omega) = \frac{1}{\sqrt{2}} \sum_{k=0}^{2N−1} h_k e^{-jk\omega}$.

**B. POOLING LAYER**

For the CNN model in this article, 1D pooling is used for feature extraction and data dimension reduction. The purpose of the pooling layer is to reduce the dimensionality of the feature map so that the model focuses on the existence of certain features rather than the specific location of the feature, and tolerates slight displacement of the feature. Maximum pooling and average pooling are commonly used pooling methods for models. The former takes the largest element in the receiving field.

$$P^{l+1}_i(j) = \max_{(j−1)W+1 \leq j \leq jW} \{g_i(t)\}, \quad t \in [(j−1)W + 1, jW]$$

where $w$ is the width of the convolution kernel. $g_i(t)$ denotes the value of the $t$-th feature in the $i$-th feature of the $l$-th layer, and $P^{l+1}_i(j)$ is the value corresponding to the $(l+1)$-th neuron.

The latter uses the arithmetic mean of elements in the receiving field. Maximum pooling has certain advantages in capturing the invariance of the image data and has achieved good results in image classification tasks. However, this advantage comes at the cost of losing spatial information, which may overfit the training data. As a result, good generalization ability cannot be guaranteed in the test data. Average pooling considers all the elements in the receiving field, which causes some low-order elements to reduce the weight of high-order elements, resulting in a smaller response [25].

Based on bearing vibration signal and frequency domain analysis, the frequency band energy information extracted by a wavelet convolution kernel is used to characterize equipment fault characteristics. An energy pooling layer method based on frequency band energy information is proposed to extract the energy characteristics of different frequency bands in the signal. However, the Root Mean Square (RMS) value, which is used to describe the statistical characteristic index of the vibration signal energy, has the characteristics of stability and good repeatability in the diagnosis index. It is an important indicator for judging the operating status of equipment and diagnosing component failures. When the index exceeds the normal value by a large amount, the equipment is certainly at fault or has a hidden danger. For signal $x$, the RMS value
calculation formula is as follows

\[ X_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} x_i^2} \]  \hspace{1cm} (6)

The energy pooling method of the signal channel is calculated as follows

\[ f(X) = \sqrt{\frac{1}{N} \sum_{x \in X} x^p} \]  \hspace{1cm} (7)

where \( p \) is the coefficient. When \( p \) is infinite, it is equivalent to the maximum pooling operation, and when \( p = 1 \), it is equivalent to an average pooling operation. In this article, \( p = 2 \) is used.

When \( p = 2 \), compared with Equations (6) and (7), the energy pooling method is approximate in extracting the RMS value of the input channel. This method can be used as a frequency band energy index to describe and better identify the abnormal situation in the vibration signal. When the pooling kernel is 2 × 2 and the stride is 2, a comparison of the maximum pooling, average pooling, and energy pooling results are shown in Fig. 1. First, energy pooling is compared to maximum pooling. Compared with the pooling results in the upper left corner of the figure, the energy pooling not only retains the maximum value of 9 in the receiving field but also retains the influence of other elements overall. The results of pooling in the other receiving fields are also better than the maximum pooling, with less information lost. Second, energy pooling and average pooling are compared. For the results of pooling in the lower right corner of the figure, 3.87 is a more reasonable result. This is because the energy pooling eliminates the offset effect of the pooling results in the average pooling, that is, the offsetting of a low-order to a high-order.

C. FULLY CONNECTED LAYER

The fully connected layer acts as a classifier in the CNN. The convolution, pooling, and other layers map the original data to the hidden layer feature space, while the fully connected layer maps the learned feature representation to the sample label space. For multi-class problems, the Softmax function is commonly used in the fully connected layer to predict the probability distribution of the input data.

The Softmax function is calculated as follows

\[ y_k = \frac{e^{d_{k}^u}}{\sum_{j=1}^{n} e^{d_{j}^u}} \]  \hspace{1cm} (8)

where there are \( n \) neurons in the output layer, \( y_k \) denotes the output of the \( k \)-th neuron, and \( d_{k}^u \) is the input data.

D. FEATURE DOMAIN ADAPTATION BASED ON MK-MMD

\( D_k \) represents the data obtained in the laboratory environment representing the source domain. Its sample space is \( \{ x_i^t \}_{i=1}^{n_t} \), and the tag space is \( \{ y_i^t \}_{i=1}^{n_t} \). \( n_t \) is the number of samples in the source domain, and \( x_i^t \) follows the distribution \( p \). We also denote the target domain \( D_s \), which represents the data in the engineering environment, which contains the sample space \( \{ x_i^s \}_{i=1}^{n_s} \). The number of samples in the target domain is \( n_t \) and \( x_i^s \) obeys distribution \( q \). Deep transfer learning aims to build a deep neural network that uses the knowledge learned from the source domain \( D_k \) to improve the performance of the target task classifier \( y = \theta(x) \).

This paper uses MMD [26] to evaluate the distribution difference between the two samples. Assuming that for all functions \( f(\bullet) \), if the mean values of the images obtained by the \( f(\bullet) \) mapping of the two distributions are equal, then the two distributions can be considered to be the same distribution, which is defined as follows

\[ MMD(X^s, X^t) = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(x_i^s) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi(x_j^t) \right\|_H \]  \hspace{1cm} (9)

where \( H \) represents the Reproducing Kernel Hilbert Space (RKHS) and \( \phi(\bullet) \) is the mapping function. The key to MMD is to find the appropriate \( \phi(\bullet) \) as a mapping function to map the source domain and target domain to \( H \). Then, the mean difference between the two parts of the data after the mapping is calculated as their difference. The most important concept is kernel \( k \). The Gaussian kernel function \( k(u, v) = e^{-\frac{\|u-v\|^2}{\sigma^2}} \) is generally used as the kernel function in the MMD algorithm, which can map data to infinite-dimensional space.

In this paper, we use the MK-MMD proposed by Gretton et al. [27]. The total kernel \( K \) is constructed with multi-kernel \( k \). The square of MK-MMD is defined as follows

\[ d_k^2(p, q) \overset{\Delta}{=} \left\| E_p \left[ \phi(x^s) \right] - E_q \left[ \phi(x^t) \right] \right\|_{H_k}^2 \]  \hspace{1cm} (10)

where \( H_k \) is the RKHS endowed with a characteristic kernel \( k \). If \( p = q \), then \( d_k^2(p, q) = 0 \).

The total kernel \( K \) is defined as

\[ \kappa \overset{\Delta}{=} \left\{ k = \sum_{u=1}^{m} \beta_u k_u : \sum_{u=1}^{m} \beta_u = 1, \beta_u \geq 0, \forall u \right\} \]  \hspace{1cm} (11)

where \( \{ \beta_u \} \) is the coefficient, which is the weight of \( K \), it is the parameter that needs to be learned. The aim is to ensure that
the variance in the MMD distance generated by each kernel is the minimum to guarantee that the derived multi-kernel $k$ is characteristic.

The optimization objective of the WDTL model consists of two parts: the loss function and distribution distance. The loss function is used to measure the difference between the predicted value and the true value. The distribution distance is the MK-MMD distance, which is used to evaluate the maximum difference between the two samples. The optimization objectives are described as follows

$$\min_{\theta} \frac{1}{n_a} \sum_{i=1}^{n_a} J(\theta(x_i^a), y_i^a) + \lambda \sum_{\ell=l_1}^{l_2} d^2_{\ell}(D^s_\ell, D^t_\ell)$$

(12)

where $\theta$ represents the weight and bias parameters of the network, which is the goal of learning. $l_1, l_2$ denote the number of layers that the network needs to adapt, from $l_1$ to $l_2$. $\lambda$ is a penalty parameter and $J(\bullet)$ is the loss function. The cross-entropy loss function is used in this paper.

III. PROPOSED DETECTION METHOD

A. PROPOSED ARCHITECTURE

The structure of the WDTL deep migration diagnosis model is shown in Fig. 2. The architecture is based on the AlexNet model, which is an effective CNN architecture. The improved method proposed is used in the design of the WDTL model. The transfer learning algorithm is integrated into the WDTL model to realize the knowledge transfer. The WDTL model is composed of five 1D convolutional layers and three fully connected layers. The input of the model is the 1D raw data of the rolling bearing vibration signal and the output is the fault label corresponding to the signal. The output size of the convolutional layer is shown at the top of the figure. The first convolutional layer in the figure is the wavelet convolutional layer mentioned in this article. During the pretraining, the input is the source domain data and the model parameters are adjusted, as shown in the upper part of the figure. When transferring the model, we take the source and target domain data as input and fine-tune the original model, which is shown at the bottom of the figure. The source domain data is used to participate in the MK-MMD domain distribution difference calculation.

The specifications of the proposed architecture are shown in Table 1. The wavelet convolution proposed in this paper is applied to the first convolutional layer and convolves with the original signal to obtain the signals of different frequency bands. This method can provide frequency information for subsequent layers, which is beneficial for enhancing the performance of WDTL feature extraction. After the wavelet convolution, the proposed energy pooling is used to extract the energy value characteristics of the multi-band signal. Several subsequent 1D convolutional layers with random kernels are used to extract abstract fault features from the band energy. The dropout technique is used after Conv 2 and Conv 4 to prevent overfitting. The following 3 fully connected layers classify the features, and after the last fully connected layer, Softmax is used to calculate the weighted score of

| NO. | Layers       | Kernel Size | Activation | Output size |
|-----|--------------|-------------|------------|-------------|
| 1   | Wavelet Conv1| 1×27×55     | ReLU       | 2048×27     |
| 2   | Pooling 1    | 16×1        |            |             |
| 3   | Conv 2       | 27×27×55    | ReLU       | 128×27      |
| 4   | Dropout      |             |            |             |
| 5   | Conv 3       | 27×27×55    | ReLU       | 128×27      |
| 6   | Pooling 2    | 16×27       |            |             |
| 7   | Conv 4       | 27×27×55    | ReLU       | 8×27        |
| 8   | Dropout      |             |            |             |
| 9   | Conv 5       | 27×27×55    | ReLU       | 8×27        |
| 10  | Flatten      |             |            | 216×1       |
| 11  | FC 1         | 216×216     | ReLU       | 216×1       |
| 12  | FC 2         | 216×64      | ReLU       | 64×1        |
| 13  | FC 3         | 64×4        | Softmax    | 4×1         |
each category. The advantage of the proposed WDTL model is that the wavelet convolution kernel is trainable, which means the wavelet basis function can be adaptive. The model fine-tunes the wavelet convolution kernel by using the gradient descent backpropagation algorithm to obtain better fitting results. This approach ensures a more reliable fault detection model. The WDTL method aims to still obtain valid diagnostic results at the target location, even with physical limitations. The principle of knowledge transfer fault diagnosis across different points is that models trained using the data in one point can be applied to process the data on other points. Signal coupling always exists in different locations in the same system, so it provides the possibility of knowledge transfer. Bearing failures at different locations have the same failure feature index, but the feature space distribution is different. Therefore, we improve the feature extraction ability of the WDTL model for vibration signals so that the diagnostic knowledge in the source domain can identify the health status of the target domain equipment.

According to transfer learning theory, the features extracted by the model are distributed differently among different tasks, and the features change from conventional features to specific features along with the network. With the increase in domain differences, the transferability of high-level features will be greatly reduced. Therefore, we fix the layers of Conv 1-Conv 3 and fine-tune the two layers of Conv 4 and Conv 5. For different diagnostic tasks, FC1-FC3 is adjusted in combination with the MK-MMD method. Finally, the fault detection of rolling bearing is realized.

B. TRAINING AND TRANSFER STRATEGY OF WDTL

The WDTL method training is carried out in the source domain and the wavelet convolution kernel is fine-tuned and adapted. Then, the transfer learning strategy is used to apply the knowledge learned from the source domain to the detection task of the target domain and to improve the performance of the new task. When the wavelet convolution kernel is fixed, the wavelet Conv 1 layer can be regarded as a constant wavelet transform data processing module, and the remaining conventional play feature extraction and classification roles.

If the wavelet convolution kernel is unfrozen to be trainable, then the model is trained through the gradient descent backpropagation algorithm, which can be thought of as the process of model learning in the source domain. The main steps of the proposed method are as follows.

1) Dataset division: The datasets of the source domain and the target domain are divided by the fault size, measuring point, and load as the variables.

2) Based on the 5-layer AlexNet model, the WDTL model is constructed using a 1D structure. The first convolutional layer is initialized with the proposed wavelet function, and the other convolutional layers are initialized with random values. Energy pooling is used to construct the pooling layer.

3) Pretraining of the model: We use the data in the source domain to train the source CNN model. The length of the input data is 2,048, the size of the first-layer wavelet convolution kernel is $55 \times 1$, and the number is 27. The pooling window size is set to $16 \times 1$. All parameters are updated through backpropagation and the Adam optimization algorithm. The initial learning rate is set to 0.001 and the decay rate is 0.99. The training process uses mini-batch learning, the size is set to 32, and Batch Normalization (BN) [28] is used to solve the vanishing or exploding gradient problem.

4) Step (3) is repeated until the pretraining is completed and the best parameters are obtained. At this point, the source model can effectively complete the fault detection task in the source domain.

5) Domain adaptation transfer: We embed the MK-MMD algorithm into the WDTL model based on the source model, and set the model objective optimization function based on Equation (10). The parameters in the source model are transferred to the target WDTL model. We use the source domain and target domain data as the model input, reduce the domain differences through backpropagation and fine-tune the model using the fine-tuning method mentioned in the previous section.

6) Step (5) is repeated until the domain adaptation migration is completed. Finally, knowledge is transferred from the source domain to the target domain and a better fault detection model is obtained.

C. PREPROCESSING OF THE INPUT SIGNAL

To expand the size of the training data sample set and reduce the demand for field status data, the input signal is processed by redundant data segmentation based on the sliding window mechanism. If the total length of the vibration signal is $L$, then the total number of training samples $n$ is described as follows

$$n = \left( \frac{L - l}{s} \right) + 1 \quad (13)$$

where $l$ is the sample length, and $s$ is the stride. In this case, $l = 2048$, $s = 500$. The sample length should contain sufficient equipment health information, and the selection of stride should be greater than the number of data points required for one revolution of the bearing. The schematic diagram of the sliding window segmentation mechanism of the dataset is shown in Fig. 3 [29]. This article uses the method to expand the quantity data by nearly 5 times, making full use of the existing data.

IV. EXPERIMENTAL VERIFICATION

To verify the effectiveness of the proposed WDTL model, 27 migration tasks were performed on the Case Western Reserve University (CWRU) bearings dataset under complex conditions. The proposed method was implemented in Python 3.7 with the PyTorch library and was trained using an NVIDIA GT720 GPU and 64G RAM.

A. DATASETS DESCRIPTION

In this paper, the WDTL performance model is verified using the CWRU bearing failure laboratory experimental data [30].
The experimental platform is shown in Fig. 4. A single-point fault is machined on the bearing using electro-discharge machining technology. Acceleration sensors are used to collect vibration signals.

The data used in this experiment is the data collected from the motor Drive End (DE) and the Fan End (FE) at a sampling frequency of 12kHz. The data contains four types: Normal (NO), Inner race Fault (IF), Outer race Fault (OF), and Ball Fault (BF). The fault diameter of each type is 0.021 inches (Class A), 0.014 inches (Class B), and 0.007 inches (Class C) under loads of 0, 1, 2, and 3 hp. In this paper, the selected data is divided into 24 categories according to the three variables of the measuring point location, load, and fault size. There are 12 categories on the DE and the same on the FE. Each category contains four health states of the bearing. The available data length for each health state is approximately 240k. The length of our sample is 2,048 data points, and the method proposed in Section III.C is used to expand the data sample. Therefore, the total number of samples is 11, 280 × 4, and the number of samples in each category is 470 × 4. The dataset of each category is divided into a training sample set and test sample set according to the proportion of 80% and 20%. The training dataset of each category contains 1,504 samples, and the test dataset contains 376 samples. The details of bearing datasets are depicted in Table 2.

To verify the accuracy of the WDTL model proposed in this paper, the CWRU bearing data is used for verification. In this paper, two types of transfer detection tasks are designed according to the degree of failure between the two measurement points, specifically as follows.

1) SAME DEGREE OF FAILURE
When the degree of failure between two measuring points is the same or similar, there are two situations: the same load and different loads. There are 4 transfer tasks under the same load, i.e., A1→A2, A3→A4, A5→A6, and A7→A8, where A1→A2 denotes the scenario that the DE bearing is the source domain with 0.021 inches and a 0 hp load operating condition, and the FE bearing is the target domain with 0.021 inches and a 0 hp load operating condition. Three transmission tasks are designed for different loads, i.e., A1→A4, A3→A6, and A5→A8. The source domain data comes from the DE, and the target domain is the FE. Feature visualization is performed for each layer of the WDTL model to verify the improvement in the model feature learning ability by the proposed method. We compare with the classic CNN model by observing the learning process of the model features. White noise is added to the Class A dataset and verified in two types of loads to test the robustness of the model.

2) DIFFERENT DEGREE OF FAILURE
In the actual situation, the fault degree between the two measuring points is not uniform, and the data distribution is quite different. Therefore, we design cross-measurement point transfer with different failure degrees under multiple load conditions. The source domain data comes from the DE, and the target domain is the FE. There are 16 transfer tasks, i.e., A1→B2, A3→B4, A5→B6, A7→B8, B1→C2, B3→C4, B5→C6, B7→C8, A1→B1→B2, A3→B3→B4, A5→B5→B6, A7→B7→B8, B1→C1→C2, B3→C3→C4, B5→C5→C6, and B7→C7→C8, including direct transfer and secondary transfer. The secondary transfer eliminates the influence of working conditions on the accuracy of model diagnosis. Then, the state detection of the bearing at the target position is realized through cross-domain transfer. To compare the ability of the pooling method to the wavelet band information extraction, we test two pooling methods.

Comparative studies are carried out from the aspects of the shallow learning methods, CNN, and other deep transfer learning methods for testing the effects of the methods. First, the proposed method is compared with typical shallow learning methods. The Support Vector Machine (SVM) [31]
classifier using the radial basis function is selected. The fault characteristics used by the SVM algorithm are the effective value, skewness, kurtosis, peak value, form factor, impulse factor, crest factor, and margin. Second, the proposed model is compared with the deep transfer learning method and the CNN method. The network structure of the classic CNN used for comparison in this article is consistent with the proposed WDTL. The CNN contains five convolutional layers and three fully connected layers, and the structural parameters of each layer are the same as those in Table 1. However, the first layer of the CNN is a random convolution kernel and uses the maximum pooling method. The training and transfer methods are consistent with the proposed method. Two deep transfer learning methods are selected: the Transfer Component Analysis (TCA) [32] and the Joint Distribution Adaptation (JDA) [33]. To further demonstrate the effectiveness of classification, the dataset is used to test all models. In addition, considering the influence of field noise on the signal, the situation of whether the signal contains noise is compared to verify the robustness of the proposed method.

**B. CASE 1: SAME DEGREE OF FAILURE**

1) RESULTS AND ANALYSIS

In this section, the author verifies the performance of the WDTL model in the diagnosis transfer between different measurement points on the same failure dataset degree. Seven transfer tasks were performed according to the proposed method. Each task was compared to the other methods and the corresponding results are displayed in Table 3.

Table 3 shows the comparative diagnosis results of the bearing health detection of the target measuring point under complex working conditions. The data in the table are the average diagnostic accuracy of the four rolling bearing types. It can be seen in the table that the average accuracy of all transfer tasks of the proposed model exceeds that of other methods by more than 20%, which shows that the proposed model can obtain good performance improvement from the improved wavelet convolution kernel.

When the source domain and target domain data are from the same load situation, the CNN method performs better than the other three methods in the diagnosis task. However, due to the improvement in the convolutional layer and the addition of the transfer learning mechanism, the method proposed in this paper can achieve a 99.73% classification accuracy on the target domain. Under different load conditions, the CNN classification accuracy in the comparison test increased by more than 15%, while the diagnostic performance of the proposed method in this paper is still maintained at a good level. The accuracy of the proposed methods is more than 15% higher than that of the other four types of methods. These results verify the effectiveness and superiority of the proposed method for feature extraction and feature domain adaptation. Furthermore, the average time of this method for one fine-tuning epoch is 9.30 seconds, and the testing time for the testing datasets is 0.45 seconds. Compared with the training time required for transfer learning methods in other references [19], [34], [35], the method in this paper takes less time, with higher computational efficiency. This method has a lower fine-tuning time cost and a higher test efficiency, which is acceptable for real-time bearing fault detection in industrial applications.

Fig. 5 exhibits the confusion matrix of the best results in the test datasets of the two working conditions. It can be seen that the IF and OF are the most easily detected fault types. From Fig. 5(a), 99% of NO is correctly classified, and approximately 0.01% of NO is incorrectly classified as the BF. The precision of the BF is 98.95%. In Fig. 5(b), 100% prediction accuracies were achieved under all health conditions except for the BF. The overall accuracy is 99.73% and 99.20%. We perform sensitivity and specificity analysis on the abovementioned confusion matrix. The analysis results are shown in Table 4. Observing Table 4, except that the sensitivities of NO and BF are 98.94% and 96.81%, respectively, the sensitivity of other bearing states can reach 99.9%. However, the specificity of BF and IF is lower than other health conditions, 99.65% and 98.94%, respectively. The results show that even if the datasets come from different measuring points and different motor loads, the proposed method can well detect bearing conditions. This strongly proves the reliability of the proposed method.

In a real environment, vibration signals are easily contaminated by noise. Therefore, the robustness of the model under noisy conditions is very important. The Class A dataset is added with Gaussian white noise to verify the diagnostic performance and robustness of the WDTL model under noisy conditions. The Signal-Noise Ratio (SNR) is taken as four. Under the same load and different load conditions, the proposed method is compared with the standard deep transfer model CNN. The comparison results are shown in Table 5. In this test comparison, the diagnosis of the proposed model is still very good in the presence of noise, with an accuracy rate of 98%. However, in the model without constructing wavelet convolution, the diagnostic accuracy is only approximately 80%. The diagnostic accuracy has dropped by more than 15%. It is proved that the proposed deep transfer model has a strong anti-noise ability and robustness.

---

**TABLE 3. Detection accuracy of different methods, including seven transfer tasks on Class A dataset (%).**

| Source-Target DE→FE | WDTL | SVM | CNN | TCA | JDA |
|---------------------|------|-----|-----|-----|-----|
| A1→A2               | 99.73| 52.13| 84.31| 77.66| 68.88|
| A3→A4               | 98.40| 39.63| 81.65| 65.43| 62.23|
| A5→A6               | 95.74| 45.48| 71.81| 65.69| 61.17|
| A7→A8               | 94.41| 47.34| 71.81| 55.05| 57.45|
| A1→A4               | 85.64| 50.80| 63.30| 71.54| 51.06|
| A3→A6               | 89.20| 44.68| 76.86| 60.64| 61.70|
| A5→A8               | 97.07| 46.54| 62.50| 52.13| 60.64|
| Avg                 | 95.74| 46.66| 73.18| 64.02| 60.47|

---

45182 VOLUME 9, 2021
2) FEATURE VISUALIZATION

The Class A dataset case obtained from the CWRU division is used to verify the ability of the wavelet convolution kernel to mine signal fault features. The author uses the t-distributed Stochastic Neighbor Embedding (t-SNE) method proposed by Laurens and Hinton [36] to visualize the convolutional learning features in the WDTL model. The t-SNE method is used to check the feature extraction of each convolutional layer in the network structure proposed by this method to verify the enhancement of the feature learning ability of the proposed wavelet convolution. The feature visualization of each convolutional layer is shown in Fig. 6.

Fig. 6 shows that the features become increasingly divisible as the layer goes deeper. The features are not divisible in the early layers, while in the fully connected layer, the features are very easy to divide. Second, from the Conv layer 2, the features of each category begin to gather together, and the features of different categories start to separate. For the row signal features, the distribution is messy, which indicates that the proposed model adapts to differences in the domain distribution. Finally, the classification effect of the fully connected layer shows that the method proposed in this paper has a good clustering effect on the target domain. In the model with the MK-MMD algorithm added, we can easily distinguish different categories, which intuitively demonstrates the effectiveness of the algorithm and the proposed model in cross-measurement fault detection.

Fig. 7 displays the three-dimensional representations of high-dimensional feature maps from the Conv layer 2 to the Conv layer 5 of the WDTL and CNN model. This representation verifies in more detail that the proposed method improves the feature extraction capability of the model and also shows the complexity of the features with the three-dimensional map. As shown in the feature map in Fig. 7(a), the samples of different categories have a good degree of discrimination in Conv layer 3. The samples of the same category are well gathered in Conv layer 4. Compared with the features in Fig. 7(b) obtained by the CNN, the features of different categories in Conv layer 5 can be more easily distinguished. It can be observed visually that the CNN method is inefficient in feature learning, which is manifested in the scattered distribution among the same categories and the tight distribution among different categories. This finding renders the final classification effect of the model unsatisfactory.

We can intuitively see that the proposed model is faster than the CNN with respect to feature learning efficiency, which verifies that the proposed method has unique advantages for feature extraction.

The visualization results are compared with the classical CNN model to verify the effectiveness of the improved wavelet convolution kernel in the WDTL model. As shown in Fig. 7, the improved convolution kernel method proposed in this paper can make full use of the prior knowledge of signal time-frequency characteristics. On the output features of the second convolutional layer, the clustering effect has
been relatively obvious, and the classification effect continues to improve with the deepening of the convolutional layer. In contrast, although the CNN model has a better classification in the third layer, the discrimination of some categories is low. It is still not valid to distinguish all categories in layer 5, and it is proven that the wavelet convolution kernel proposed in this paper has advantages in feature mining.

C. CASE 2: DIFFERENT DEGREE OF FAILURE

1) DIRECT TRANSFER METHOD

Table 6 shows the results of cross-detection under different fault degrees. The diagnostic tasks are (DE)An→(FE)Bn and (DE)Bn→(FE)Cn. It can be seen that the average accuracy of all transmission tasks in this method is 96.43%, which is higher than the other methods. SVM, TCA, and JDA perform poorly in cross-domain detection, with an average accuracy of 38.83%, 37.40%, and 46.71%, respectively. In comparison, the best diagnostic accuracy of the proposed WDTL method in detection tasks is 99.47% and 99.93%. The data show that the proposed model has a good generalization ability. However, the average accuracy of the CNN is 70.05%, indicating that our method improves by 26%. The results indicate the effectiveness of the proposed method in the sharing and transmission of fault diagnosis knowledge in different fields. The proposed model can extract better domain invariant features based on wavelet convolution and energy pooling, and the high cross-domain detection accuracy proves the superiority of this method.

The confusion matrix of the transfer results under different failure degrees is shown in Fig. 8. The abscissa represents the predicted result, and the ordinate represents the true result. In Fig. 8(a), only a small number of BFs are not correctly classified. In Fig. 8(b), only 3% of OFs are incorrectly classified as BFs, and the remainder is correctly classified. As shown in Table 7, 100% of the prediction accuracies were achieved under all health conditions except for the BF and OF. For the specificity, the BF under the Bn→Cn task is 98.94%, while the specificity is higher than 99.50% in the other cases. This result further proves the effectiveness of the proposed method in cross-domain fault detection.
FIGURE 7. Feature visualization comparison of the two methods under the cross-measurement point for the transfer detection task. Visualization from the Conv layer 2 to the Conv layer 5: (a) WDTL model and (b) CNN model.

The maximum pooling method was used as a comparison, and the classification accuracy of the model was tested in the An→Bn task. The comparison results are shown in Table 8. The diagnostic effect of using energy pooling was increased by 8% compared to the maximum pooling. It can be seen that the diagnostic accuracy of the optimized method is higher than 95%, which proves that the proposed energy pooling can effectively improve the diagnostic accuracy of the WDTL method.

2) SECONDARY TRANSFER METHOD
The direct transfer method does not consider the influence of different load conditions on WDTL modeling. Therefore, on the basis of the direct transfer method, the secondary transfer method is adopted. The first transfer is carried out between different measuring points under the same working condition, and the second transfer is carried out under different working conditions on the same side. The diagnostic tasks are (DE)An→(DE)Bn→(FE)Bn and (DE)Bn→(DE)Cn→(FE)Cn. The purpose of the secondary transfer is to eliminate the influence of the working conditions on the model diagnosis task. The experimental results are shown in Table 9.

From the results, SVM, TCA, and JDA performed poorly in the target location diagnosis, with the accuracy of all three less than 65%. These methods demonstrate poor adaptability in the transfer of diagnostic tasks. In comparison, the diagnostic accuracy of the CNN increased to 80% but is still lower than that of WDTL, with an accuracy rate of 99.2%. It can be seen that the average accuracy of the proposed method is 94.98%, which is the best performance. This result clearly shows the effectiveness of the proposed wavelet convolution in improving the time-frequency information extraction capability of the model and the advantages of an adaptive

| Source-Target | WDTL | SVM  | CNN  | TCA  | JDA  |
|---------------|------|------|------|------|------|
| An→Bn         |      |      |      |      |      |
| A1→B2         | 96.28| 42.55| 73.14| 41.22| 49.73|
| A3→B4         | 97.07| 25.80| 68.35| 24.73| 53.99|
| A5→B6         | 97.07| 38.56| 66.49| 38.30| 41.49|
| A7→B8         | 99.47| 44.68| 79.52| 38.03| 41.22|
| B1→C2         | 87.50| 40.96| 82.71| 41.22| 30.85|
| B3→C4         | 99.20| 43.35| 70.74| 32.98| 54.79|
| B5→C6         | 94.95| 36.17| 68.62| 41.76| 52.66|
| B7→C8         | 99.93| 38.56| 50.80| 40.96| 48.94|
| Avg           | 96.43| 38.83| 70.05| 37.40| 46.71|

TABLE 7. Analysis result of the sensitivity and specificity of the proposed confusion matrix in Case 2 (%).

| Fault type  | NO  | IF  | BF  | OF  | NO  | IF  | BF  | OF  |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|
| Sensitivity | 100 | 100 | 97.87| 100 | 100 | 96.81|     |     |
| Specificity | 100 | 99.65| 100 | 99.65| 100 | 98.94| 100 |     |

TABLE 8. Comparison results of the proposed model with energy pooling and with maximum pooling (%).

| Source-Target | WDTL with Energy Pool | WDTL with Max Pool |
|---------------|-----------------------|--------------------|
| DE→FE        |                       |                    |
| A1→B2        | 96.28                 | 85.64              |
| A3→B4        | 97.07                 | 87.77              |
| A5→B6        | 97.07                 | 87.50              |
| A7→B8        | 99.47                 | 97.87              |
| Avg          | 97.47                 | 89.70              |
TABLE 9. Detection accuracy of all comparison methods in eight secondary transfer tasks (%).

| Source-target | WDTL | SVM  | CNN  | TCA  | JDA  |
|---------------|------|------|------|------|------|
| Transfer from |      |      |      |      |      |
| fault A to fault |      |      |      |      |      |
| A1→B1→B2 | 90.96 | 42.29 | 75.80 | 45.21 | 51.33 |
| A3→B3→B4 | 97.87 | 27.66 | 60.11 | 37.23 | 41.49 |
| A5→B5→B6 | 99.20 | 39.89 | 84.31 | 59.04 | 39.36 |
| B A7→B7→B8 | 94.95 | 38.03 | 82.71 | 54.78 | 44.95 |
| Transfer from |      |      |      |      |      |
| fault B to fault |      |      |      |      |      |
| B1→C1→C2 | 92.02 | 42.02 | 83.78 | 63.03 | 53.99 |
| B3→C3→C4 | 94.41 | 45.74 | 84.04 | 53.19 | 52.66 |
| B5→C5→C6 | 95.48 | 47.61 | 82.18 | 49.47 | 52.39 |
| C B7→C7→C8 | 94.95 | 44.15 | 76.60 | 58.78 | 44.14 |
| Avg | 94.98 | 40.92 | 78.69 | 52.59 | 47.54 |

FIGURE 8. Confusion matrix of the best transfer results of each fault type: (a) the classification accuracy in transfer task A1 → B2 and (b) the classification accuracy in transfer task B3 → C4.

FIGURE 9. Accuracy of different diagnostic tasks in Case 2. The proposal achieves the highest diagnostic accuracy: (a) direct transfer and (b) secondary transfer.

domain in improving the generalization capability of the model. Notably, the proposed WDTL classification accuracy is higher for diagnostic tasks (DE)An→(DE)Bn→(FE)Bn than for (DE)An→(FE)Bn when compared to Tables 5 and 7. This result is consistent with the actual situation because the latter is closer to the characteristics of the target position after the elimination of the working condition, which proves that the proposed method has better diagnostic performance on the target measuring points.

To illustrate the classification results under the two diagnostic tasks in detail, the accuracy comparison graphs are plotted. It can be seen in Fig. 9 that the proposed methods achieved the highest diagnostic accuracy for the different detection tasks. Specifically, Fig. 9(a) shows a direct transfer diagnostic task under different working conditions. The accuracy of the other four diagnostic methods is lower than that of the method mentioned in this paper. This change is because the data distribution of the source domain and the target domain are different under variable conditions. The contrast method does not utilize the signal processing prior knowledge, so it is difficult to reduce the difference. In Fig. 9(b), the performance of the classic CNN improved, which benefits from the secondary transfer. It eliminates the influence of the changes in the working conditions and reduces the difference between the source domain and the target domain but it is still inferior to the proposed method. This finding proves the advantages of the proposed model in the utilization of prior knowledge and cross-test point transfer diagnosis. The model can still maintain good diagnostic performance and good robustness under variable conditions.
V. CONCLUSION

This paper proposes a fault detection method for rolling bearings to solve the problem of bearing fault detection on mechanical equipment that cannot directly obtain status data. The proposed method directly uses the original vibration data as the input of the model, and mines the time-frequency characteristic information of the associated signal through the wavelet convolution kernel. Energy pooling is used to obtain the energy characteristics of the signal channel, and the MK-MMD algorithm is used to enhance the characteristic energy adaptively. Finally, the feature transferability in a specific layer of the deep transfer model is improved. The experimental results show that the combination of the proposed improved wavelet convolution and energy pooling with the domain adaptive algorithm can improve the classification accuracy by more than 15% in different diagnostic tasks. It can achieve a high accuracy rate of 99.73%. The proposed method can realize the fault diagnosis of the target measuring point based on the data of a similar measuring point, and it also has good robustness under strong background noise and variable working conditions. The research results of this paper can provide a reference for the related research of deep transfer learning in the field of fault diagnosis.

REFERENCES

[1] M. Sadoughi and C. Hu, “Physics-based convolutional neural network for fault diagnosis of rolling element bearings,” IEEE Sensors J., vol. 19, no. 11, pp. 4181–4192, Jun. 2019.
[2] X. Li and W. Zhang, “Deep learning-based partial domain adaptation method on intelligent machinery fault diagnostics,” IEEE Trans. Ind. Electron., vol. 68, no. 5, pp. 4351–4361, May 2021.
[3] J. Zhu, N. Chen, and C. Shen, “A new multiple source domain adaptation fault diagnosis method between different rotating machines,” IEEE Trans. Ind. Informat., early access, Sep. 3 2020, doi: 10.1109/TII.2020.3021406.
[4] S. Tang, S. Yuan, and Y. Zhu, “Data preprocessing techniques in convolutional neural network based on fault diagnosis towards rotating machinery,” IEEE Access, vol. 8, pp. 373–382, Jul. 2020.
[5] J. Kim, Y. Ko, and J. Seo, “Construction of machine-labeled data for improving named entity recognition by transfer learning,” IEEE Access, vol. 8, pp. 59684–59693, 2020.
[6] X. Li, W. Zhang, Q. Ding, and X. Li, “Diagnosing rotating machines with weakly supervised data using deep transfer learning,” IEEE Trans. Ind. Informat., vol. 16, no. 3, pp. 1688–1697, Mar. 2020.
[7] W. Qian, S. Li, J. Wang, Y. Xin, and H. Ma, “A new deep transfer learning algorithm combining a priori diagnosis knowledge toward cross-domain fault diagnosis of rolling bearing,” IEEE Trans. Instrum. Meas., vol. 70, pp. 1–11, 2021.
[8] J. Gereznik, P. Wang, C. Sun, and R. X. Gao, “Explainable convolutional neural network for gearbox fault diagnosis,” Proc. CIRP, vol. 80, pp. 476–481, Jan. 2019.
[9] D. Zhang, Y. Chen, F. Guo, H. R. Karimi, H. Dong, and Q. Xuan, “A new interpretable learning method for fault diagnosis of rolling bearings,” IEEE Trans. Instrum. Meas., vol. 70, pp. 1–10, 2021.
[10] A. Khan, A. Sohail, U. Zahoor, and A. S. Qureshi, “A survey of the recent architectures of deep convolutional neural networks,” Artif. Intell. Rev., vol. 53, no. 8, pp. 5855–58561, Dec. 2020.
[11] X. Shi, Y. Cheng, B. Zhang, and H. Zhang, “Intelligent fault diagnosis of bearings based on feature model and alexnet neural network,” in Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM), Jun. 2020, pp. 1–6.
[12] T. Lu, F. Yu, B. Han, and J. Wang, “A generic intelligent bearing fault diagnosis system using convolutional neural networks with transfer learning,” IEEE Access, vol. 8, pp. 164807–164814, 2020.
[13] M. M. Islam and J.-M. Kim, “Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network,” Comput. Ind., vol. 106, pp. 142–153, Apr. 2019.
[14] A. Kumar, Y. Zhou, C. P. Gandhi, R. Kumar, and J. Xiang, “Bearing defect size assessment using wavelet transform based deep convolutional neural network (DCNN),” Alexandria Eng. J., vol. 59, no. 2, pp. 999–1012, Apr. 2020.
[15] K. Khalil, O. Eldash, A. Kumar, and M. Bayoumi, “Intelligent fault prediction assisted self-healing for embryonic hardware,” IEEE Trans. Biomed. Circuits Syst., vol. 14, no. 4, pp. 852–866, Aug. 2020.
[16] B. Chen, B. Shen, F. Chen, H. Tian, W. Xiao, F. Zhang, and C. Zhao, “Fault diagnosis method based on integration of RSSD and wavelet transform to rolling bearing,” Measurement, vol. 131, pp. 400–411, Jan. 2019.
[17] K. Khalil, O. Eldash, A. Kumar, and M. Bayoumi, “Machine learning-based approach for hardware faults prediction,” IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 67, no. 11, pp. 3880–3892, Nov. 2020.
[18] C. Zhang, J. Feng, C. Hu, Z. Liu, L. Cheng, and Y. Zhou, “An intelligent fault diagnosis method of rolling bearing under variable working loads using 1-D stacked dilated convolutional neural network,” IEEE Access, vol. 8, pp. 63027–63042, 2020.
[19] M. Long, Y. Cao, J. Wang, and M. I. J. Ordan, “Learning transferable features with deep adaptation networks,” in Proc. 22nd Int. Conf. Mach. Learn. (ICML), vol. 1, Jun. 2015, pp. 97–105.
[20] A. Grettov, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, “A kernel two-sample test,” J. Mach. Learn. Res., vol. 13, pp. 723–773, Mar. 2012.
[21] J. Zheng, H. Sun, X. Wang, J. Liu, and C. Zhu, “A batch-normalized deep neural networks and its application in bearing fault diagnosis,” in Proc. 11th Int. Conf. Intell. Hum.-Mach. Syst. Cybern. (HIMSC), Aug. 2019, pp. 121–124.
[22] M. J. Hasan, M. Sohail, and J.-M. Kim, “1D CNN-based transfer learning model for bearing fault diagnosis under variable working conditions,” in Proc. Int. Conf. Comput. Intell. Inf. Syst., Nov. 2018, pp. 13–23.
[23] D. Neupane and J. Seok, “Bearing fault detection and diagnosis using case-based reserve university dataset with deep learning approaches: A review,” IEEE Access, vol. 8, pp. 93155–93178, 2020.
[24] Y.-K. Gu, X.-Q. Zhou, D.-P. Yu, and Y.-J. Shen, “Fault diagnosis method of rolling bearing using principal component analysis and support vector machine,” J. Mech. Sci. Technol., vol. 32, no. 11, pp. 5079–5088, Nov. 2018.
[25] W. Mao, J. He, and M. J. Zuo, “Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning,” IEEE Trans. Instrum. Meas., vol. 69, no. 4, pp. 1594–1608, Apr. 2020.
[26] Z. Xu, H. Darong, G. Guan, and W. Yongchao, “A fault diagnosis method based on improved adaptive filtering and joint distribution adaptation,” IEEE Access, vol. 8, pp. 159683–159695, 2020.
[27] S. Xiong, H. Zhou, S. He, L. Zhang, Q. Xia, J. Xuan, and T. Shi, “A novel end-to-end fault diagnosis approach for rolling bearings by integrating wavelet packet transform into convolutional neural network structures,” Sensors, vol. 20, no. 17, pp. 4965, Sep. 2020.
[35] F. Li, T. Tang, B. Tang, and Q. He, “Deep convolution domain-adversarial transfer learning for fault diagnosis of rolling bearings,” Measurement, vol. 169, Feb. 2021, Art. no. 108339.

[36] L. van der Maaten and G. Hinton, “Visualizing data using t-SNE,” J. Mach. Learn. Res., vol. 9, pp. 2579–2605, Nov. 2008.

MENGLIANG LIAO was born in Hunan, China, in 1996. He received the bachelor’s degree in mechanical engineering from the Kunming University of Science and Technology, in 2018, where he is currently pursuing the master’s degree with the Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology. His research interests include deep learning and fault diagnosis.

CHANG LIU received the Ph.D. degree in engineering from the Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, in 2017. In the past five years, he has hosted or participated in several scientific research projects. He published more than ten articles. His current research interests include signal processing theory and methods, compressed sensing and sparse representation, and machinery fault diagnosis. He is a member of the Fault Diagnosis Professional Committee and the Chinese Society of Vibration Engineering.

MENGLIANG LIAO was born in Hunan, China, in 1996. He received the bachelor’s degree in mechanical engineering from the Kunming University of Science and Technology, in 2018, where he is currently pursuing the master’s degree with the Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology. His research interests include deep learning and fault diagnosis.

CONG WANG was born in Shaanxi, China, in 1992. She received the bachelor’s degree in mechanical engineering from the Xi’an University of Architecture and Technology, in 2015. She is currently pursuing the master’s degree with the Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology. Her research interests include compressed sensing and fault diagnosis.

JIANWEI YANG was born in Yunnan, China, in 1998. She received the bachelor’s degree in mechanical engineering from the Kunming University of Science and Technology, in 2019, where she is currently pursuing the master’s degree with the Faculty of Mechanical and Electrical Engineering. Her research interests include multi-source heterogeneous and fault diagnosis.

***