Measurement 201 (2022) 111597

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

A wavelet packet transform-based deep feature transfer learning method for bearing fault diagnosis under different working conditions

Xiao Yu [a,b], Zhongting Liang [a,b], Youjie Wang [a,b], Hongshen Yin [a,b], Xiaowen Liu [b,c,*], Wanli Yu [d], Yanqiu Huang [e]

[a] IOT Perception Mine Research Center, China University of Mining and Technology, Xuzhou, 221000, China
[b] School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, 221000, China
[c] School of Electrical and Power Engineering, China University of Mining and Technology, Xuzhou, 221000, China
[d] Institute of Electrodynamics and Microelectronics, University of Bremen, Bremen, 28359, Germany
[e] Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, 7522NH, The Netherlands

ARTICLE INFO

Keywords:
Wavelet packet transform
Fault diagnosis
Residual network
Multi-kernel maximum mean discrepancy
Domain adaptation method

ABSTRACT

Deep learning has achieved significant advances in the fault diagnosis of rotating machinery. However, it still suffers many challenges such as various working conditions, large environmental noise interference and insufficient effective data samples. Signal time–frequency analysis and feature transfer learning methods can help solve these problems. Combining wavelet packet transform (WPT) and multi-kernel maximum mean discrepancy (MK-MMD), this paper proposes a novel residual network (ResNet)-based deep transfer diagnosis model for bearing faults. Firstly, this paper devises a distinctive WPT time–frequency feature map (WPT-TFFM) construction method using WPT for time–frequency analysis on nonlinear and non-stationary vibration signals. Then, a modified multi-group parallel ResNet network is structured to extract the depth features of WPT-TFFM for the characteristics of small size and feature dispersion. Then, MK-MMD is further applied to evaluate the distribution difference between the depth features of the source and target domain data. Combining with the classification loss of the sample set with the source domain, the depth features extraction network is optimized to achieve better cross-domain invariance and fault state differentiation capability of the depth features. To evaluate the proposed method, this work conducts comparative experiments on two test rigs under different working loads and speeds. The results reveal that the proposed method offers excellent fault diagnosis and noise prevention capability for working condition transfer tasks.

1. Introduction

Rotating machines are indispensable key equipment for both civilian and industry field. As one of the key components of rotating machines, rolling bearings are very prone to failure under the working conditions of high speed, heavy load and strong impact. Once the failure occurs, it will affect the efficiency of machinery and equipment, and even cause casualties. Therefore, it is of great significance to explore the intelligent fault diagnosis method for rolling bearings to ensure the safe and efficient operation of equipment [1–3].

When the key components such as bearings and gears are in an abnormal state, vibration signals will show different time–frequency characteristics. The analysis method based on the vibration signal is currently the most commonly used method for diagnosis of bearing fault condition [4,5]. Time–frequency analysis methods have unique advantages in analyzing nonlinear and non-stationary rotating mechanical vibration signals, and the commonly used time–frequency analysis methods include the short time Fourier transform (STFT) [6], wavelet and wavelet packet analysis [7], Wigner-Ville distribution (WVD) [8], empirical mode decomposition (EMD) [9] and stochastic resonance demodulation. In the past years, combined with time–frequency analysis methods, machine learning methods such as Back Propagation (BP) neural network [10], support vector machine [11,12], fuzzy c-means clustering [13], etc., have been widely used in bearing fault diagnosis. However, these methods need effective signal analysis means and prior knowledge as guidance, which cannot meet the development needs of big data in the industrial field [14–16].

Compared with traditional machine learning, deep learning can adaptively extract deep semantic features from vibration signals by using end-to-end learning ideas, which can reduce the participation

* Corresponding author at: School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, 221000, China.
E-mail address: xwliucumt@126.com (X. Liu).

https://doi.org/10.1016/j.measurement.2022.111597
Received 13 February 2022; Received in revised form 9 June 2022; Accepted 2 July 2022
Available online 8 July 2022

of empirical knowledge and is more suitable for processing complex bearing vibration signals [17–20]. As a result, it has been widely used in fault diagnosis [21–24]. The deep Convolutional Neural Networks (CNN) is one of the most used deep learning method in fault diagnosis [25,26]. However, with the increase of network depth, CNN network is prone to gradient dispersion or explosion during training. To solve this problem, He et al. [27] proposed a depth residual ResNet model in 2015, which can solve the problem of gradient disappearance, which has become one of the most widely used depth features extraction networks at present [14,28].

Moreover, in real application scenarios, it is very often to face the problem of changes in equipment working conditions, and different test targets from the training samples, which leads to differences in the distribution of data in the source and target domains. To discover general diagnosis knowledge from multiple source domains and apply the knowledge to facilitate new tasks in target domain, Zheng et al. [29] constructed multiscale transfer symbolic dynamic entropy (MTSDE). Li et al. [30] used local Fisher discriminant analysis (LFDA) and Grassmann manifold to reduce the risk of negative transfer considering multiple source domains. Shen et al. [31] proposed a new penalty domain selection machine (PDSM) transfer learning model for gearbox fault recognition. Tong et al. [32] proposed Fault Sensitivity Assessment Model (FSAM) based on the Maximum Mean Difference (MMD) to transmit the most sensitive data stream, and then analyze the fault.

Combining the feature adaptation extraction ability of deep learning, the scene adaptability of feature transfer method is further improved. To achieve sufficient generalization performance, Zheng et al. [33] proposed a cross-domain diagnosis scheme combining the priori diagnosis knowledge and deep domain generalization network. Ragab et al. [34] designed a deep learning architecture for adversarial unsupervised domain adaptation to address the single-source multiple-target (1SmT) problem. Li et al. [35] introduced a domain adversarial network to project the data collected from different sensors into a shared subspace, which can provide better generalization of fault diagnostic knowledge in different feature spaces. Guo et al. [36] constructed a deep convolution transfer learning network (DCTLN) by combining the deep convolution network and the MMD, and used unlabeled samples from the target domain for intelligent identification of bearing faults.

Considering that the time–frequency characteristics of the vibration signal contain the mechanism of bearing damage, on the basis of the above researches, this paper combines the time–frequency analysis method with the deep transfer learning method to improve the performance of the diagnostic model. As mentioned above, the Maximum Mean Difference (MMD) method is a commonly used method to evaluate the distribution difference between two domains in transfer learning [37–39]. However, as the choice of kernel function is crucial to the metric distance of MMD, the parameter selection of each kernel drastically affects the final performance of domain adaptation. To solve this problem, this paper chooses the Multi-Kernel Maximum Mean Difference (MK-MMD) to provide an effective estimation of the mapping multi-kernel space.

Different from previous studies, this paper devises a distinctive WPT-based time–frequency feature map construction method to provide more effective fault time–frequency feature information for deep feature extraction networks. Meanwhile, a lightweight ResNet is structure and a deep transfer learning method for fault diagnosis is established with MK-MMD. The primary contributions of this paper are as follows.

(1) It devises a distinctive WPT time–frequency feature map (WPT-TFFM) construction method to extract preliminary features for state expression and noise immunity.

(2) It establishes a lightweight ResNet and structures a modified multi group Resnet (MGRN) framework to extract the depth features of WPT-TFFM in multi-scale and multi-resolution.

(3) It then proposes a novel roller bearing fault diagnosis model termed as WMGRNMM. MK-MMD is employed to evaluate the distribution difference between the source and target domain data depth features.

(4) The experiment results demonstrate that the proposed WMGRNMM model offers excellent transfer diagnosis capability in variable working mode, including fault diagnosis accuracy, scene adaptation and noise resistance capability.

The remainder of this paper is organized as follows. Section 2 introduces the fundamentals in WPT, ResNet and MK-MMD. Section 3 focuses on the proposed deep transfer diagnosis model WMGRNMM. In Section 4, multiple sets of experiments are carried out through two bearing failure test rigs and the experimental results are analyzed Finally, the conclusion is drawn in the fifth section.

2. Fundamentals

This section briefly states the preliminaries about WPT, ResNet and MK-MMD, which will be used to construct the proposed WMGRNMM model.

2.1. WPT

Wavelet analysis is a multi-scale time–frequency analysis method, which is a powerful tool for nonlinear non-smooth vibration signal analysis. However, wavelet transform has several limitations: it only decomposes the low-frequency components of the signal and does not decompose the high-frequency component, and is insensitive to the features of the high-frequency components of the signal. The WPT, which is proposed based on wavelet transform, is able to decompose the high-frequency part of the signal in detail, has better local characteristics, and extracts the high-frequency features of the signal [40].

In wavelet packet analysis, the Hilbert space $L^2(R)$ can be divided into an orthogonal sum of all subspaces $W_j (j \in Z)$ according to different scale factors j, i.e., $L^2(R) = \bigoplus_{j\in Z} W_j$, where W_j is the wavelet subspace of the wavelet function $\Psi(t)$. To improve the frequency resolution, the wavelet packet further subdivides the wavelet subspace W_j according to the binary system. The orthogonal basis of this space is $u^n_{j,k}(t) = 2^{-j/2}u^n(2^{-j}t - k)$, where U^n_j denotes the j first wavelet n subspace of the scale ($n = 0, 1, 2, 3, \ldots, 2^j - 1$, called the frequency factor) and k is the translation factor. The orthogonal basis$U^n_{j.k}$ satisfies the following two-scale equation:

$$u^n_{j,0} = \sum_k g_0(k)\, u^i_{j-1,k} \qquad \text{(n is even)} \tag{1}$$

$$u^n_{j,0} = \sum_k g_1(k)\, u^i_{j-1,k} \qquad \text{(n is odd)} \tag{2}$$

where, $j, k \in Z$, $n = 1, 2, 3, \ldots, 2^j - 1$, $g_0(k)$, $g_1(k)$ are a pair of orthogonal filters that satisfy $g_1(k) = (-1)^{1-k} g_0(1-k)$.

After performing the layer WPT decomposition j, the Unj frequency band corresponding to each subspace of the first j layer is:
$\left\{\left[0, \frac{f_s}{2^{j+1}}\right]; \left[\frac{f_s}{2^{j+1}}, \frac{2f_s}{2^{j+1}}\right]; \left[\frac{2f_s}{2^{j+1}}, \frac{3f_s}{2^{j+1}}\right]; \ldots; \left[\frac{(2^j-1)f_s}{2^{j+1}}, \frac{f_s}{2}\right]\right\}$, where f_s is the sampling frequency.

2.2. Deep residual network (ResNet)

As the depth of the network increases, CNN model training becomes more and more difficult, thus causing gradient dispersion or explosion, leading to a decrease in the classification accuracy of the model classification. To solve this problem, K.He et al. proposed the deep residual ResNet model in 2015. The basic unit of ResNet is the residual block, as shown in Fig. 1, which uses a shortcut connection method to connect the input x with the $F(x)$ obtained after stacking weight layers across layers to obtain the output $H(x) = F(x)+x$. At this point $F(x) = H(x)-x$ is the residual. The formula is as follows:

$$F = W_2\sigma\left(W_1 x\right) \tag{3}$$

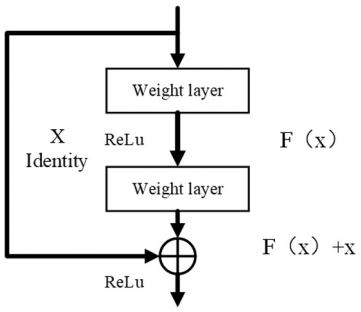


Fig. 1. Residual learning module.

Where F is the residual function, σ is the Relu activation function, $W1$, $W2$ is the weight layer.

After the cross-layer connection operation of ResNet, the gradient does not disappear with the increase of network depth, which can effectively solve the problem of model degradation caused by too deep CNN network.

2.3. Multi-kernel maximum mean discrepancy (MK-MMD)

In the intelligent diagnostic models for rotating machine components, changes in operating conditions such as speed and load can easily lead to distribution differences between the source domain data and the target domain data, which in turn affects the classification accuracy of the diagnostic model. For the known working conditions with labels (source domain data set D_s) and the unknown working conditions without labels (target domain data set D_t), the goal of feature transfer learning is to make the feature space distribution of D_s and with the D_t as consistent as possible through the distance measurement criterion, so as to improve the adaptability of the diagnostic model to the variable working condition scenarios.

Maximum Mean discrepancy (MMD) is the most used distance measure in transfer learning, which measures the distance between two distributions in regenerated Hilbert space. The choice of kernel function is crucial to the metric distance of MMD, and the choice of parameters of each kernel affects the final performance of the mapping. To address the influence of kernel functions on the final performance of the mapping, this paper used the MK-MMD [39] method proposed by Gretton [4], which maximizes the two-sample testing capability and minimizes the type II error. Reconstructing the MK-MMD distance defined in the regenerated Hilbert space, i.e., the Euclidean distance between the edge distribution $P(Xs)$ and the kernel $Q(Xt)$ embedding, the Euclidean distance formula of MK-MMD is defined as follows.

$$d_k^2(P,Q) \triangleq \left\| \mathrm{E}_P\left[\phi\left(\mathrm{X}^{\mathrm{s}}\right)\right] - \mathrm{E}_Q\left[\phi\left(\mathrm{X}^{\mathrm{t}}\right)\right] \right\|_{H_k}^2 \tag{4}$$

where E represents the mathematical expectation, ϕ represents the mapping of the regenerated Hilbert space, and H_k represents the regenerated kernel Hilbert space with characteristic kernel k.

3. The proposed fault diagnosis algorithm

In this section, it presents the details of the proposed WMGRNMM model, a bearing fault depth features transfer diagnosis model based on WPT, ResNet and MK-MMD methods. It first extract the time–frequency features of vibration signals at different scales using WPT. Then, it constructs the time–frequency maps based on the leaflet node coefficients of WPT in layers 4, 5 and 6 of vibration signals to obtain the time–frequency features of vibration signals at different scales and resolutions. Moreover, a parallel ResNet depth features extraction network, denoted as multi group Resnet (MGRN), is established to separately extract the depth features of the 4, 5 and 6-layer wavelet packet time–frequency maps and generate the combined depth features for the final state identification. To solve the transfer diagnosis problem under different operating conditions (the experiments contain different speeds or loads), MK-MMD is further applied in WMGRNMM to evaluate the distribution differences between the source and target domain depth features. The domain adaptation capability and state classification capability of the depth features will be optimized to improve the generalization capability of the diagnostic model to transfer working conditions.

For the illustration purpose, this section will first present the time–frequency map construction based on WPT, then give the details of ResNet-based deep feature extraction network and the MK-MMD-based transfer diagnosis model, and finally present the detailed steps of the proposed WMGRNMM model.

3.1. WPT time–frequency feature map construction

Typically, the vibration signal is nonlinear and non-stationary, and if the vibration signal is directly converted into a 2D matrix form, the fault time–frequency information of the signal will be scattered throughout the 2D matrix, which will increase the difficulty of fault feature extraction of vibration signal by 2D convolutional kernels in ResNet. WPT is one of the most commonly used time–frequency analysis methods for non-linear and non-stationary signals [7]. The wavelet packet space is complete and orthogonal, and the vibration signal can be decomposed into frequency information under different subbands that represents the time–frequency features of the vibration signal without losing information. Therefore, this paper uses WPT to process the vibration signal and construct the WPT-TFFM to improve the fault feature extraction ability of the 2D convolutional kernels.

WPT is performed on the vibration signal, and the obtained wavelet packet node coefficients denote the time–frequency characteristics of the vibration signal at different scales. Since different wavelet packet node coefficients have different physical meanings, we propose to construct the coefficient matrix using the coefficients of individual wavelet packet leaf nodes. To facilitate the subsequent 2D convolution kernel to extract the depth features of wavelet coefficients, the coefficient matrices of all leaf nodes are stitched together in a zigzag order to build a WPT time–frequency feature map of the vibration signal, here we referred it as WPT-TFFM.

Here, we take 4096 samples of vibration signals as an example and introduce the construction process of WPT-TFFM, as shown in Fig. 2.

(1) The vibration signal samples are decomposed by 4 layers of wavelet packets to obtain 16 wavelet packet leaflet nodes with 256 wavelet packet coefficients per wavelet leaflet node.

(2) The wavelet packet coefficients of each leaf node are constructed into a 16*16 coefficient matrix.

(3) Zigzag stitching of the coefficient matrices of these 16 wavelet packet leaf nodes by numbers, which gives the WPT-TFFM of the vibration signal samples.

In existing deep learning fault diagnosis methods, the vibration signals or their time–frequency spectrograms are often directly used as input to the deep feature extraction network, while the influence of the form of signal input on the network is rarely considered. In this paper, a new input construction method is proposed to obtain the multi-scale time–frequency information of the vibration signals by utilizing the multi-scale time–frequency analysis capability of WPT. To enable the depth feature extraction network to better extract time–frequency features for fault classification, the coefficient matrix of wavelet packet leaf node is reordered by Zigzag stitching. As a result, the perceptual field of components with similar frequency intervals is more concentrated, so that the depth features can more effectively reflect the time–frequency response characteristics of different fault states of rolling bearing.

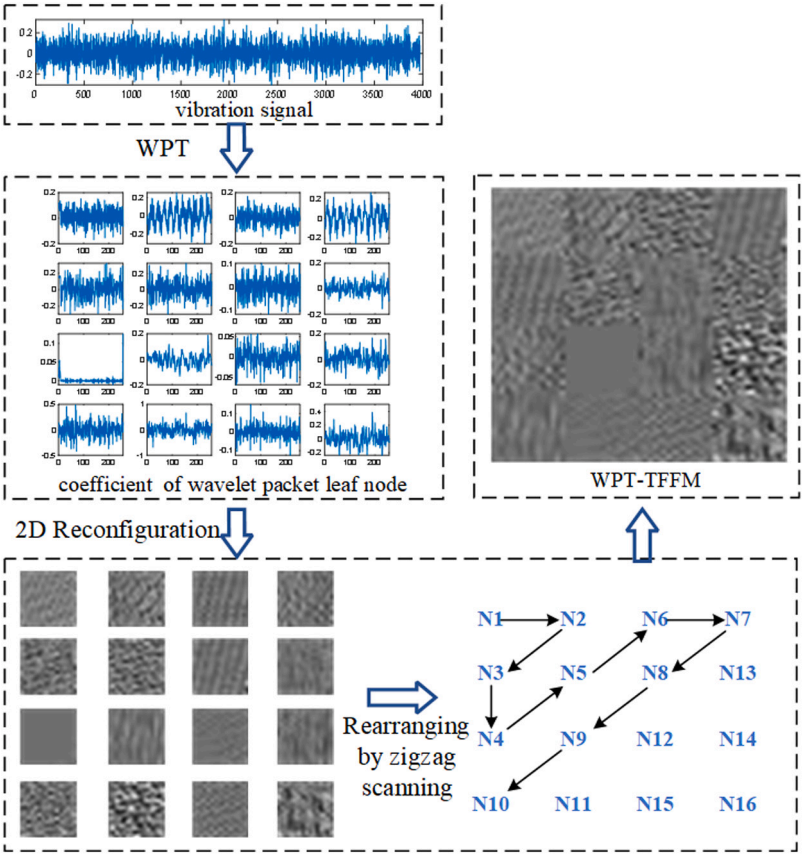


Fig. 2. WPT-TFFM construction workflow.

3.2. ResNet-based depth features extraction network

To better analyze the time–frequency characteristics of the vibration signals at different scales and resolutions, 4, 5 and 6-layer WPT decomposition are used to decompose the vibration signals, the corresponding time–frequency feature maps WPT-TFFM (4), WPT-TFFM (5) and WPT-TFFM (6) are constructed, and the vibration signal WPT-TFFM under different fault states are shown in Fig. 1. A ResNet-based depth features extraction network is established to extract the depth features of WPT-TFFM (4), WPT-TFFM (5), and WPT-TFFM (6) simultaneously. By analyzing Fig. 3, the ResNet depth features extraction network has the following problems:

(1) The WPT-TFFM size of the vibration signal is small and the state features are dispersed in the time–frequency diagram.

(2) The time–frequency feature maps obtained from WPT decomposition of different layers have their own structural characteristics.

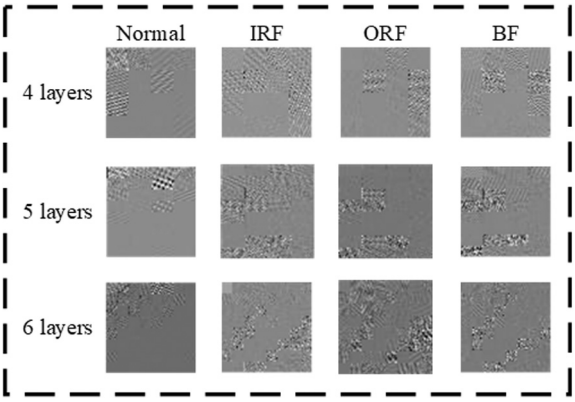


Fig. 3. WPT-TFFMs of different number of WPT layers. (IRF: inner race fault. ORF: outer race fault. BF: ball fault).

For problem (1), for the small size WPT-TFFM, here we first use 3*3 convolution kernels in the ResNet network for feature extraction, the convolution step size is set to 1, and the pooling operation in the middle layer is eliminated. As shown in Fig. 4, unlike the conventional Resnet by increasing the network depth to improve the feature extraction ability, the depth of the Resnet network is limited here, To reduce the impact of small size images and feature scattering on the generalization ability of the model, we design a 7-layers Resnet Block (here referred to as RNB-7).

For problem (2), three RNB-7s are designed to form a parallel network MGRN to extract the feature information of WPT-TFFM of 4, 5, and 6 layers respectively. As shown in Fig. 5 three RNB-7s with the same structure and independent parameters can extract personalized structural features of different types of WPT-TFFMs more effectively, and the depth features extracted by the three RNB-7s are spliced into combined depth features that are fed into the fully connected layer and finally into the Softmax classifier for classification.

3.3. The MK-MMD-based transfer diagnosis model

The fault diagnosis model in Section 3.2 is trained using source domain labeled data, which can obtain depth features expressing the fault states of the source domain data. However, the classification accuracy of this model in the target domain decreases significantly when there is a difference in the distribution of depth features between the source and target domain data. To improve the similarity of the probability distributions of the depth features of the source domain data and the target domain data, MK-MMD is used here to perform the

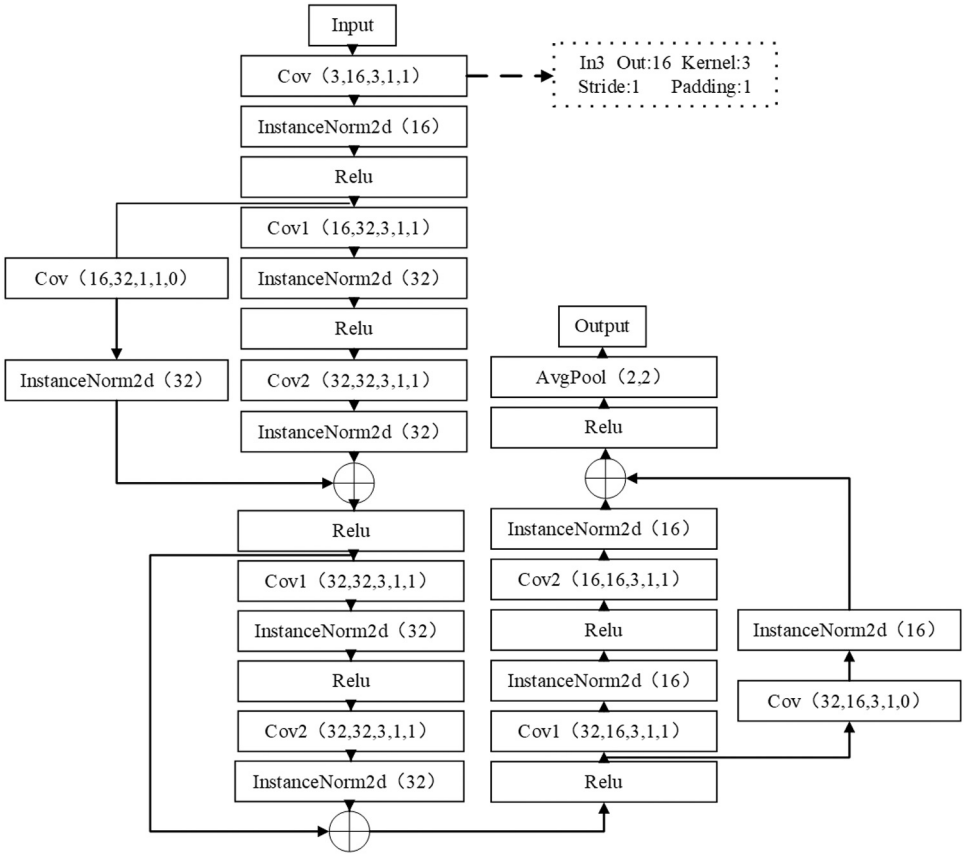


Fig. 4. Flowchart of the designed 7-layers ResNet Block.

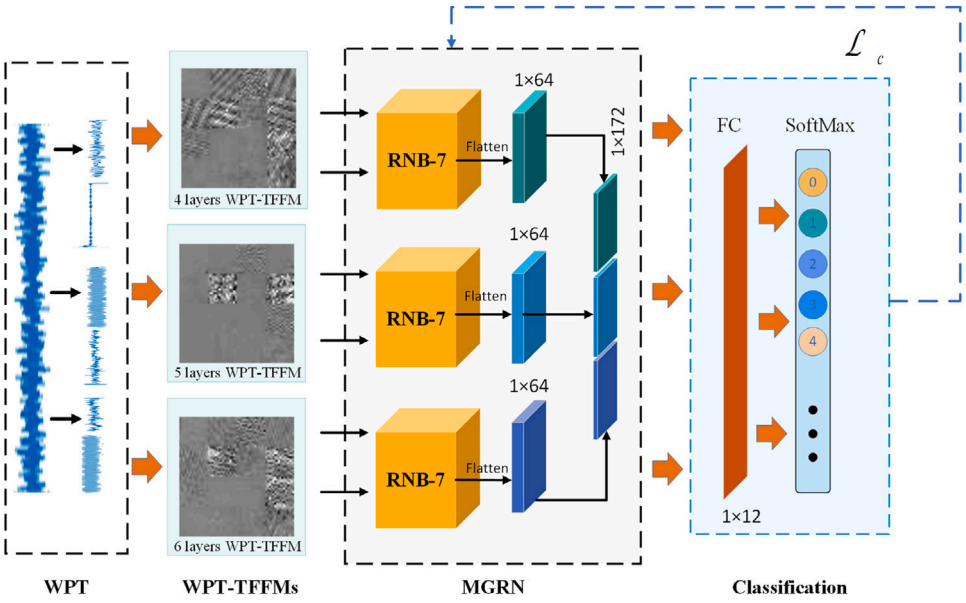


Fig. 5. Flowchart of the parallel network MGRN.

domain fitness metric on the depth features of the source domain data and the target domain data. The labeled source domain data and the unlabeled target domain data are fed into the MGRN in Section 3.2, respectively, to obtain the depth features of the source and target domain datasets, denoted as F_s and F_t, and calculate the MK-MMD loss between F_t and F_s, denoted as L_{MK-MMD}, as shown in Eq. (5).

$$L_{MK-MMD}(F_s, F_t) \triangleq \left\| \mathrm{E}\left[\phi\left(F_s\right)\right] - \mathrm{E}\left[\phi\left(F_t\right)\right] \right\|_{H_k}^2 \tag{5}$$

where E is the mathematical expectation, ϕ represents the mapping of the regenerated Hilbert space, H_k denotes the regenerated kernel Hilbert space with characteristic kernel k.

In addition, after going through the fully connected layer and Softmax classifier, we obtain the predicted category information, combined with the label information, the cross-entropy loss can be calculated,

denoted as L_C, as shown in Eq. (6).

$$L_c = \frac{1}{N}\sum_{i}\sum_{c=1}^{M} F_{s_{ic}} \log\left(p_{ic}\right) \tag{6}$$

where M is the number of categories, $F_{s_{ic}}$ is the sign function (0 or 1), and P_{ic} samples the predicted probability of i belonging to category c.

Then, we calculate the overall loss of the transfer classification model L_{MGRN}, as in Eq. (7):

$$L_{MGRN} = L_c + \lambda_{MK-MMD} L_{MK-MMD}\left(F_s, F_t\right) \tag{7}$$

Where λ_{MK-MMD} is the trade-off parameter of the total loss.

In this paper, the source and target domains share depth features to extract network parameters. The goal is to minimize the total loss, using back propagation and gradient descent methods to optimize the depth network as well as fully connected network parameters. The depth features extracted by the MK-MMD-based transfer diagnosis model can express the fault state of the source domain data. At the same time, reducing the difference between the depth features of the source and target domain data improves the generalization ability of the model.

3.4. Diagnostic steps for bearing faults in transfer tasks

The training process of WMGRNMM model is shown in Fig. 6, and the main process of bearing fault diagnosis in WMGRNMM-based work transfer task is described as follows:

Step 1: We collect the vibration signals of various state bearings under different working conditions, the sample working condition vibration signals with labels are used as the training set, and the sample working condition vibration signals without labels are used as the test set.

Step 2: The 4, 5, and 6-layer WPT decomposition is performed on the vibration signal samples from the training set (source domain) and the test set (target domain) to obtain the WPT-TFFM (4), WPT-TFFM (5), and WPT-TFFM (6) for each sample.

Step 3: The WPT-TFFMs of the source and target domains are fed to MGRN and the combined depth features of WPT-TFFM(4), WPT-TFFM(5), and WPT-TFFM(6) are extracted using MGRN.

Step 4: The depth features of the source domain samples are fed into the fully connected layer FC, which combines the Softmax classifier and the source domain label information to calculate the cross-entropy loss L_C.

Step 5: The MK-MMD distance of the source and target domain sample depth features is calculated to obtain the source and target domain depth features space of L_{MK-MMD}.

Step 6: The fully connected layer FC is optimized by using L_C back propagation, and the L_{MGRN} is calculated and back-propagated to optimize the MGRN.

Step 7: Iterate steps (3)–(6) until L_{MGRN} is less than the set value or the number of iterations reaches the target requirement to obtain the trained MGRN and FC.

Step 8: The depth features of the target domain sample WPT-TFFM are extracted using the trained MGRN network, and the depth features are fed into the trained FC to obtain the category labels of the samples.

4. Experimental results

This section conducts extensive experiments on the Case Western Reserve University (CWRU) bearing fault experimental platform and Machinery Fault Simulator and Rotor Dynamics Simulator (MFS-RDS) experimental platform to validate the performance of the proposed WMGRNMM model. To demonstrate the superiority of WMGRNMM, we compare it with several other fault diagnosis models.

4.1. Results of the CWRU bearing fault experimental platform

4.1.1. Platform description

We first validate the model using the CWRU bearing experimental platform [41], as shown in Fig. 7. It consists of a 2 HP motor (left), torque sensor/encoder (center), dynamometer (right), and control electronics. Vibration data was collected from the drive side and fan side by accelerometers placed in the motor housing. In total, four experiments were conducted to collect vibration data for different power levels of 0 HP, 1 HP, 2 HP, and 3 HP.

In our experiment, vibration data from the drive end with a sampling frequency of 12 kHz were analyzed for a total of 12 different states of the bearing: (0) normal, (1) inner race damage with a diameter of 0.007 inches, (2) outer race damage with a diameter of 0.007 inches, (3) rolling element damage with a diameter 0.007 inches, (4) inner race with a diameter of 0.014 inches, (5) outer race damage with a diameter of 0.014 inches, (6) rolling body damage with a diameter of 0.014 inches, (7) inner race damage with a diameter of 0.021 inches, (8) outer race damage with a diameter of 0.021 inches, (9) rolling body damage with a diameter of 0.021 inches, (10) inner race damage with a diameter of 0.028 inches, and (11) rolling body damage with a diameter of 0.028 inches.

We take 4096 sampling points of vibration signal sequence as a sample and obtain 30 vibration signal samples for each state of the bearing under a single power, and the vibration data under four different powers are used to establish four datasets for four different working conditions, which are dataset 0, dataset 1, dataset 2, and dataset 3, each of which includes 360 vibration signal samples, as shown in Table 1. Here we performed mutual transfer diagnostic tests for the data sets in the four working conditions states. There are twelve transfer tasks (0 -> 1, 0 -> 2, 0 -> 3, 1 -> 0, 1 -> 2, 1 -> 3, 2 -> 0, 2 -> 1, 2 -> 3, 3 -> 0, 3 -> 1, 3 -> 2), where 0 -> 1 indicates transfer from the training set (source domain) dataset 0 to the test set (target domain) dataset 1.

4.1.2. Diagnostic results under different transfer tasks

In this section, we evaluate the WMGRNMM model for transfer diagnosis between different operating states. To verify the effectiveness of our theoretical analysis and WMGRNMM model in Section 3, we introduce some models for comparison experiments. The settings of various comparison models are shown below:

(1) WMGRN model. It has the same network structure as the WMGRNMM model, but does not use MK-MMD for domain adaptation optimization.

(2) WRNMM model. Compared with the WMGRNMM, it changes the depth features extraction network by using a single RNB-7 network instead of MGRN, with WP-TFF(4), WP-TFF(5), and WP-TFF(6) as three channels input into a single RNB-7.

(3) WRN18MM and WRN34MM models. Compared with WRNMM, it increases the depth of ResNet using classical ResNet18 and ResNet34 networks instead of the proposed RNB-7 network.

Based on the results in Table 2, the following conclusion can be drawn:

(1) The WMGRN model performs unstably in various transfer task, reaching a maximum of 97% and a minimum of only 81%, with excessive differences in accuracy rates. This indicates that the depth features of the source and target domain data extracted by WPT and MGRN have certain distribution differences under different working conditions, resulting in the diagnostic models trained by the source domain data failing to achieve better classification accuracy in the target domain. After adding the MK-MMD domain transfer method, the classification results of WMGRNMM model in various modes are significantly improved, which achieves an average accuracy of 98%.

(2) The average accuracy of the WRNMM model is lower than that of the WMGRNMM model, and the classification accuracy under various transfer tasks is also lower than that of WMGRNMM, which is consistent with our analysis in Section 3.2. For time–frequency features

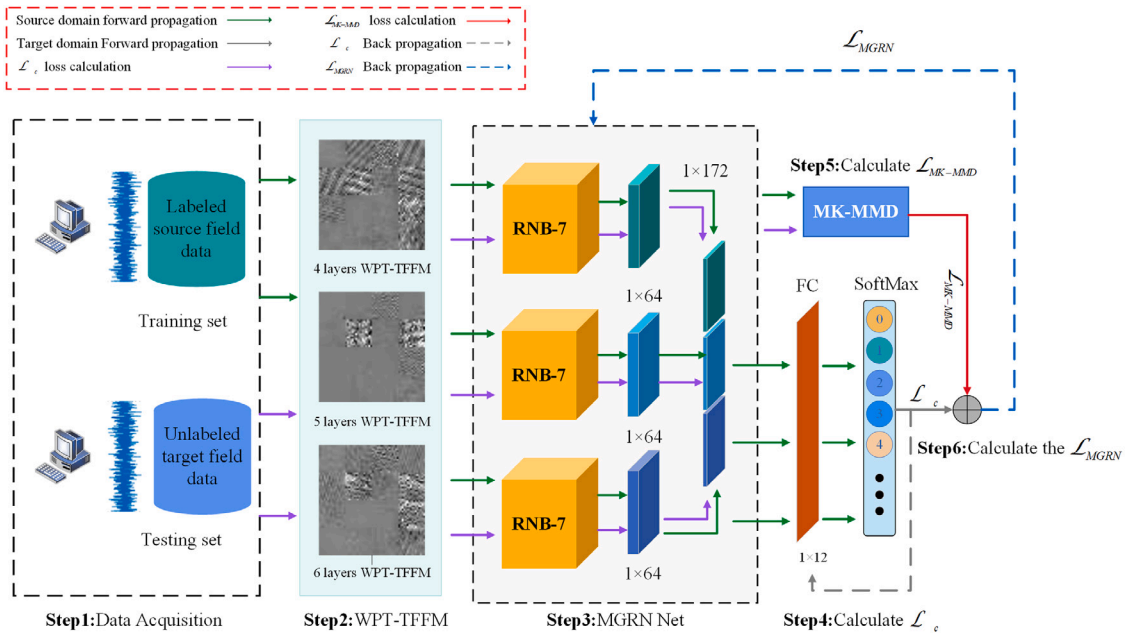


Fig. 6. Flowchart of WMGRNMM for bearing fault diagnosis in working condition transfer task.

Table 1
Experimental datasets of CWRU experimental platform.

	HP	Normal	IRF				ORF			BF			
Damage (in)	–	–	0.007	0.014	0.021	0.028	0.007	0.014	0.021	0.007	0.014	0.021	0.028
label	–	0	1	2	3	4	5	6	7	8	9	10	11
Dataset 0	0 HP	30	30	30	30	30	30	30	30	30	30	30	30
Dataset 1	1 HP	30	30	30	30	30	30	30	30	30	30	30	30
Dataset 2	2 HP	30	30	30	30	30	30	30	30	30	30	30	30
Dataset 3	3 HP	30	30	30	30	30	30	30	30	30	30	30	30

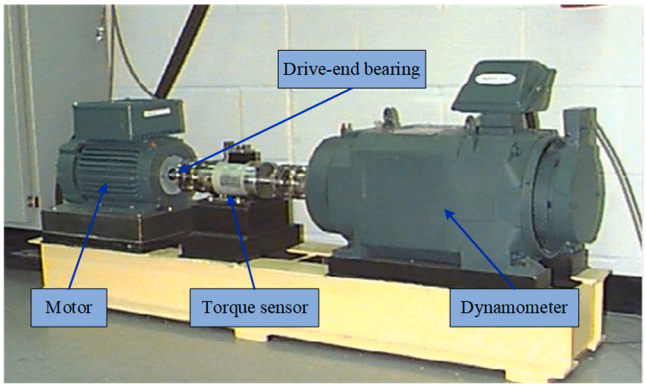


Fig. 7. The CWRU bearing fault experimental setup.

Table 2
Classification accuracy of transfer fault diagnosis based on various models.

Transfer task	WRN18MM	WRN34MM	WRNMM	WMGRN	WMGRNMM
0-> 1	93.61%	94.44%	90.83%	89.72%	**97.78%**
0-> 2	96.67%	93.06%	94.17%	86.39%	**100.00%**
0-> 3	93.33%	93.61%	98.33%	81.11%	**99.72%**
1-> 0	91.11%	88.33%	97.22%	95.00%	**97.50%**
1-> 2	93.61%	92.78%	99.44%	97.50%	**100.00%**
1-> 3	94.72%	92.50%	93.89%	84.44%	**97.50%**
2-> 0	90.83%	93.89%	94.17%	91.67%	**97.22%**
2-> 1	94.44%	89.17%	96.11%	92.78%	**97.50%**
2-> 3	97.78%	97.50%	97.22%	95.28%	**100.00%**
3-> 0	91.39%	88.06%	92.78%	85.00%	**98.61%**
3-> 1	88.89%	88.61%	92.22%	85.83%	**97.22%**
3-> 2	96.94%	97.78%	96.39%	90.28%	**100.00%**
Average	93.61%	92.48%	95.23%	89.58%	**98.59%**

of different resolutions in WP-TFF(4), WP-TFF(5) and WP-TFF(6), a parallel network with three independent parameters is more effective than using a single network to extract depth features.

(3) Compared with WRNMM, the WRN18MM and WRN34MM models increase the network depth, but the classification accuracy of these two models does not improve significantly, but decreases in many transfer tasks, which is related to the smaller size of the WPT-TFFM and the dispersion of the state features mentioned in Section 3.2, and proves that our design of RNB-7 can more effectively extract the depth state features in the vibration signal WPT-TFFM.

(4) The classification accuracy of the WMGRNMM model outperforms other models in all types of transfer tasks, which proves the validity and reliability of the proposed WMGRNMM model.

Fig. 8 shows the variation of the state classification accuracy of various models with the number of epochs under different transfer tasks. All models converge after 4200 epochs, and the recognition accuracy is basically stable after 2000 epochs, reaching the optimal level. In general, the proposed WMGRNMM model outperforms the other models under various transfer tasks, followed by the WRNMM, and the results show that the WMGRNMM model is stable. The WRN18MM and WRN34MM models will outperform the WRNMM model under certain tasks, but due to the deeper network depth, it can easily lead to the overfitting problem. The overall performance is not as good as the WRNMM model.

To compare the adaptive ability of the depth features extracted by different models, we use the t-SNE method to analyze the depth features extracted by each model. Fig. 9(a)–(d) show the t-SNE feature analysis for each diagnostic model under the transfer task 3->1. The t-SNE plots

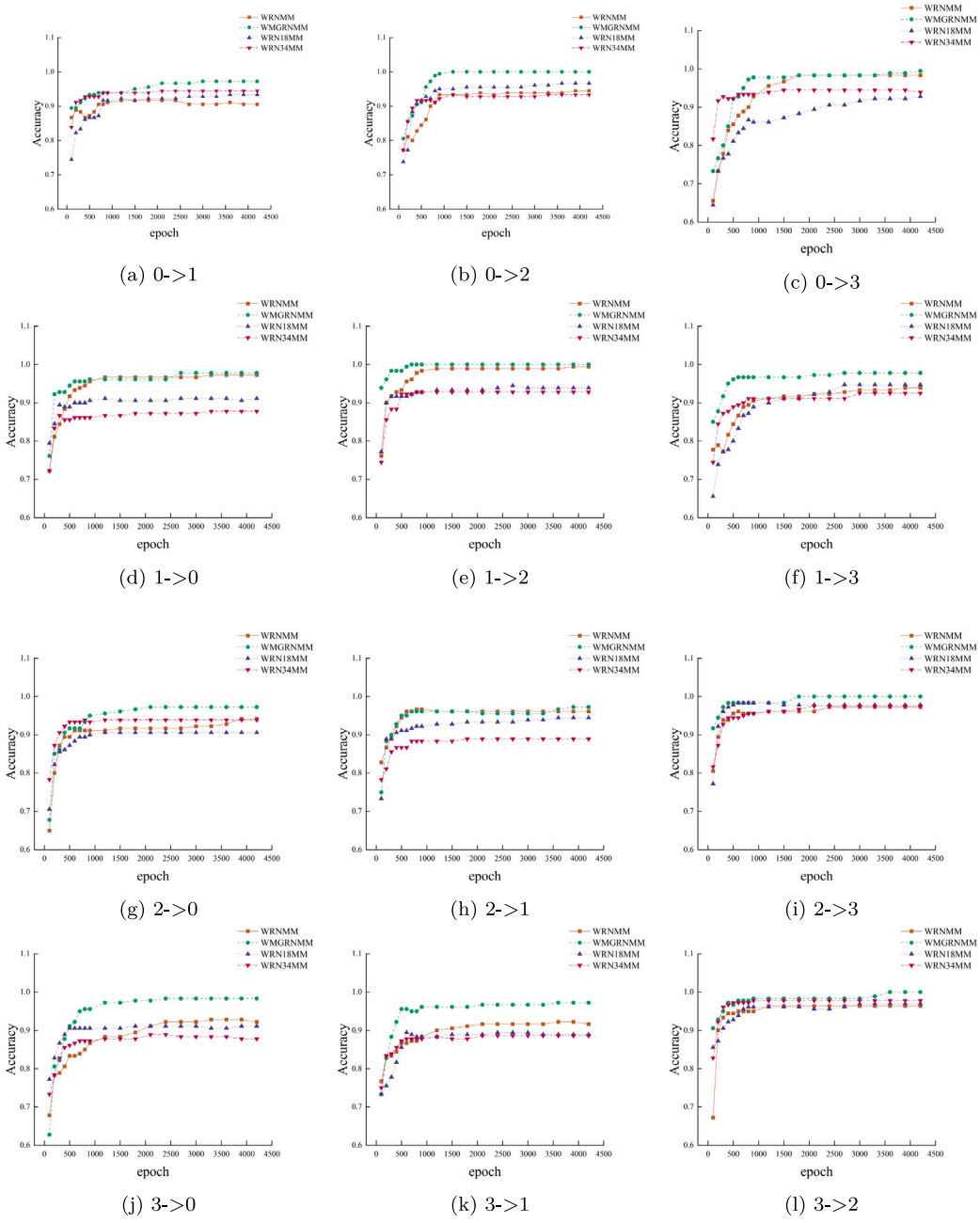


Fig. 8. Classification accuracy of each model with epochs in various transfer tasks.

show that, compared with WMGRN, the depth features extracted by the WMGRNMM model have better class separability after adding the MK-MMD loss, which improves the distance between different classes of signal samples. Compared to the WRNMM and WRN18MM, the depth features of the WMGRNMM model have better category separability.

To verify the effect of WPT in the diagnostic model, we remove the WPT step, directly convert the original bearing signal into a matrix and send it to MGRN for deep feature extraction, which is recorded as the diagnostic model MGRNMM. In contrast to WMGRNMM, MGRNMM does not have the WPT step. Similarly, several diagnostic models, such as RNMM, RN18MM, RN34MM, and MGRN, are constructed without WPT. The diagnostic results of each model are shown in Table 3, from which it can be seen that the classification accuracies of various models decrease by an average of 3–5 percentage points. This is consistent with our theoretical analysis in Section 3.1, further demonstrating that WPT-TFFM can effectively characterize the time–frequency information of vibrational signals and improve the performance of fault feature extraction of 2D convolutional kernels in MGRN.

4.1.3. Experimental results of noise immunity

In engineering practice, the bearing vibration signals obtained are often mixed with noise. As a result, the fault diagnosis model requires a certain anti-noise capability to verify the anti-noise capability of WMGRNMM model. Four signal-to-noise ratios of 3 dB, 5 dB, 8 dB and 10 dB are added to the original vibration signal to test the influence of the noise signal on the vibration model, and the signal-to-noise ratio *SNR* is calculated as follows.

$$SNR = 10\lg\left(\frac{P_{signal}}{P_{noise}}\right) \tag{8}$$

Where Psignal is the effective power of the signal and Pnoise is the effective power of the noise. The WMGRNMM model is used to diagnose and analyze the data with the addition of various signal to noise ratios.

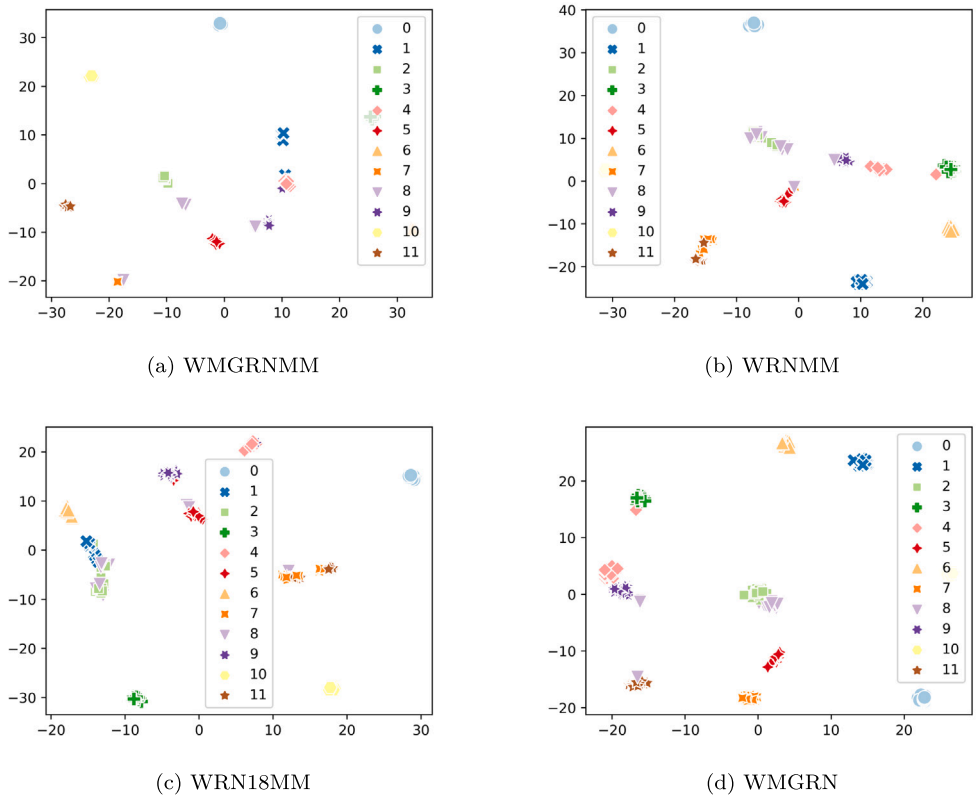


Fig. 9. t-SNE plots and confusion matrices in transfer task 3->1.

Table 3
Classification accuracies of diagnostic models without the WPT step on the CWRU experimental platform.

Transfer task	RN18MM	RN34MM	RNMM	MGRN	MGRNMM
0–1	90.56%	91.39%	88.89%	85.33%	**94.44%**
0–2	92.78%	89.72%	90.83%	82.78%	**97.22%**
0–3	91.11%	90.56%	94.72%	79.72%	**96.39%**
1–0	87.78%	84.72%	93.06%	91.39%	**93.33%**
1–2	89.72%	90.28%	96.94%	92.78%	**98.06%**
1–3	91.94%	86.94%	90.83%	80.28%	**94.44%**
2–0	88.06%	91.39%	90.56%	87.78%	**92.78%**
2–1	88.33%	86.39%	92.22%	88.33%	**93.61%**
2–3	93.06%	94.17%	94.17%	91.67%	**97.22%**
3–0	87.22%	83.89%	87.78%	81.94%	**95.56%**
3–1	86.11%	84.72%	88.61%	82.22%	**93.33%**
3–2	92.22%	93.89%	92.22%	86.94%	**96.67%**
Average	89.91%	89.01%	91.74%	85.93%	**95.25%**

Table 4
Classification accuracy of WMGRNMM model in various transfer tasks under different SNR conditions.

Transfer task	3 dB	5 dB	8 dB	10 dB	No noise added
0-> 1	94.72%	96.39%	**98.33%**	95.83%	97.78%
0-> 2	96.94%	**100.00%**	**100.00%**	99.72%	**100.00%**
0-> 3	98.61%	99.44%	**100.00%**	99.44%	99.72%
1-> 0	95.83%	96.39%	**97.50%**	97.22%	**97.50%**
1-> 2	99.17%	99.44%	99.72%	**100%**	**100.00%**
1-> 3	95.56%	96.39%	96.94%	96.67%	**97.50%**
2-> 0	95.28%	95.56%	95.28%	96.67%	**97.22%**
2-> 1	94.17%	96.39%	97.22%	**97.78%**	97.50%
2-> 3	98.33%	99.17%	98.89%	99.44%	**100.00%**
3-> 0	94.17%	96.11%	97.50%	98.06%	**98.61%**
3-> 1	93.89%	95.83%	96.39%	96.94%	**97.22%**
3-> 2	98.89%	99.17%	99.72%	99.44%	**100.00%**
Average	96.30%	97.52%	97.96%	98.12%	**98.59%**

Table 4 shows the comparison results without the addition of noise. It can be seen that there is no significant decrease in the accuracy of the WMGRNMM model for adding 8 dB and 10 dB of noise data. For 5 dB of added noise data, there is a slight decrease in the classification accuracy of the model, but it is still able to maintain an average accuracy of greater than 97%. For 3 dB of noise-added data, the WMGRNMM model accuracy was affected to some extent, with an average decrease of 2%. The results of the noise-added comparison experiments show that the WMGRNMM model accuracy is less affected by noise when the signal-to-noise ratio is greater than 5 dB, and the WMGRNMM model has good noise immunity.

4.1.4. Comparison with other diagnostic methods

We compare the WMGRNMM model proposed in this paper in the variable working condition environment with the TCA, JDA, and CNN models in various variable working condition modes. Table 5 shows the comparison results.

(1) That the average classification accuracy of the WMGRNMM model is much higher than that of the TCA, JDA, and CNN models, with the WMGRNMM model improving the average accuracy by 8% points over the TCA model, by more than 6% over the JDA models and by more than 11% over the CNN models.

(2) The average classification accuracy of the WMGRNMM model is higher than that of DANN, SF+SOF+HKL [42] and CNN-CORAL [43] models. The generalization ability of the depth features can be improved by introducing time–frequency feature processing of vibration signal into the deep transfer networks.

(3) In the variable condition mode, combined with WPT and MK-MMD, WMGRNMM can better extract the depth features reflecting the time–frequency characteristics of vibration signals and reduce the differences in the distribution of the depth features of the source and target domain data to improve the classification accuracy in the variable condition transfer task.

Table 5
Classification accuracy of different models in various transfer tasks.

Transfer task	TCA	JDA	CNN	DANN	SF+SOF+HKL [42]	CNN-CORAL [43]	WMGR-NMM
0-> 1	88.61%	91.67%	86.39%	97.22%	99.8%	98.5%	**97.78%**
0-> 2	93.33%	95.28%	85.83%	98.33%	99.86%	98.33%	**100.00%**
0-> 3	91.39%	93.06%	81.11%	91.94%	87.54%	99.07%	**99.72%**
1-> 0	92.50%	92.78%	92.50%	96.11%	88.5%	**98.40%**	97.50%
1-> 2	93.06%	95.28%	93.06%	**100.00%**	99.7%	99.00%	**100.00%**
1-> 3	87.50%	88.61%	82.22%	97.22%	**100.00%**	93.87%	97.50%
2-> 0	87.78%	89.17%	85.83%	89.72%	**99.59%**	99.53%	97.22%
2-> 1	88.89%	90.56%	86.39%	95.83%	**99.23%**	97.67%	97.50%
2-> 3	92.78%	94.17%	91.11%	84.72%	95.5%	98.40%	**100.00%**
3-> 0	91.39%	93.06%	83.33%	88.33%	95.17%	95.20%	**98.61%**
3-> 1	87.78%	88.33%	83.61%	89.44%	98.16%	**98.93%**	97.22%
3-> 2	95.28%	96.39%	89.72%	93.89%	97.81%	97.20%	**100.00%**
Average	90.86%	92.36%	86.76%	93.56%	96.74%	97.85%	**98.59%**

Table 6
Experimental datasets of MFS-RDS experimental platform.

	Rotational speed (r/min)	Normal	IRF			ORF			BF		
Damage(mm)	–	–	0.05	0.1	0.2	0.05	0.1	0.2	0.05	0.1	0.2
label	–	0	1	2	3	4	5	6	7	8	9
Dataset A	900	60	60	60	60	60	60	60	60	60	60
Dataset B	1200	60	60	60	60	60	60	60	60	60	60
Dataset C	1800	60	60	60	60	60	60	60	60	60	60

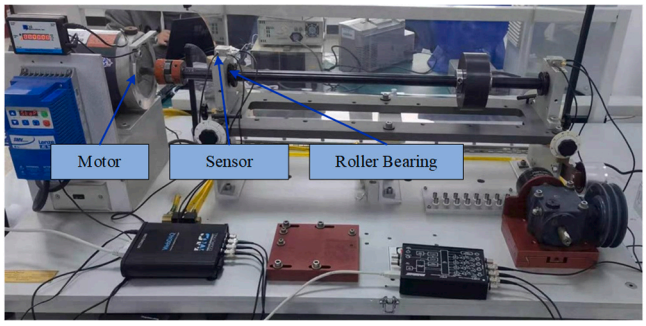


Fig. 10. Machinery Fault Simulator and Rotor Dynamics Simulator experimental platform.

4.2. Results of the MFS-RDS experimental platform

4.2.1. Platform description

We further analyze the validity of the model using the MFS-RDS experimental platform to test the transfer effect of the model at different speeds. As shown in Fig. 10, the bearing vibration signal at the motor drive end is acquired using a WebDAQ acquisition card with a signal sampling frequency of 8 kHz. The experimental bearing model is NSK's SER205-16, and three different degrees of damage of 0.05 mm, 0.1 mm, and 0.2 mm are engraved on the inner ring, outer ring, and rolling body of the bearing using a laser. As shown in Table 6, the experimental bench is set up with three different acquisition speed, 900 r/min, 1200 r/min, 1800 r/min, respectively. According to the different speed, from low to high set three working condition data set, they are recorded as A, B, C. Each working condition data set has normal bearing data and three different damage degree of inner ring, outer ring, rolling body data, a total of 10 different states of bearing vibration data.

Similarly, we use the vibration signal sequence of 4096 sampling points as a sample, and 60 samples of each bearing vibration signal are obtained for each of the single working condition states. For the three working condition data sets A, B and C, there are six working condition transfer task settings (A-> B, A-> C, B-> A, B-> C, C-> A, C-> B), where A-> B means that the training set (source domain) data set A is transferred to the test set (target domain) data set B.

Table 7
Classification accuracy of various network models of MFS-RDS experimental platform.

Transfer task	WRN18MM	WRN34MM	WRNMM	WMGRN	WMGRNMM
A->B	92.50%	90.17%	94.33%	83.83%	**97.83%**
A->C	91.17%	91.50%	93.67%	79.50%	**97.17%**
B->A	89.83%	87.67%	93.17%	81.83%	**96.33%**
B->C	90.67%	89.33%	92.83%	78.67%	**97.50%**
C->A	88.67%	86.67%	91.33%	82.17%	**95.83%**
C->D	94.33%	93.00%	95.17%	85.67%	**98.17%**
Average	91.20%	89.72%	93.42%	81.95%	**97.14%**

4.2.2. Results and analysis

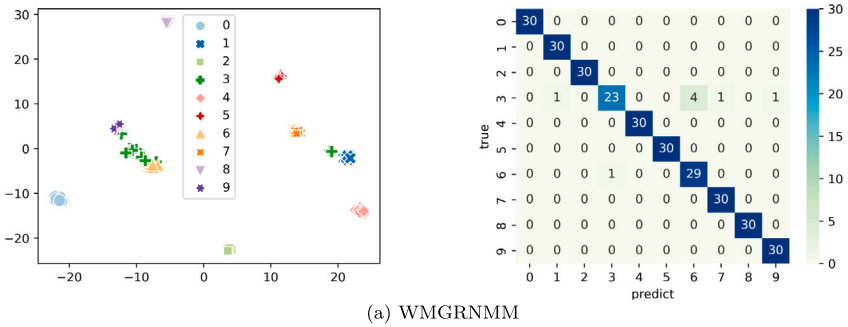
The fault transfer diagnosis results of the various fault diagnosis models described in Section 4.1.2 on the MFS-RDS experimental bench are shown in Table 7. From the table we can draw the following conclusions:

(1) In the MFS-RDS bearing data, the WMGRNMM model still maintains the highest accuracy across the work transfer tasks with an average accuracy of 97.14%, which is nearly 15% higher than the lowest WMGRN model. The results illustrate that using MK-MMD to optimize the depth features extraction network can effectively reduce the distribution differences between the depth features of the source and target domain data.

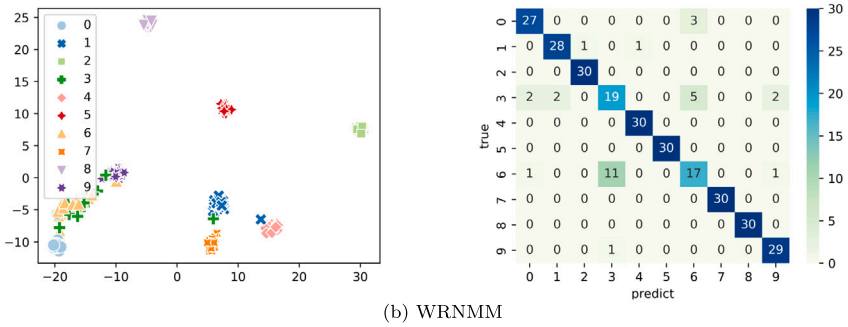
(2) The accuracy of the WMGRNMM model is higher than that of the WRNMM model in all work transfer tasks, further validating the advantage of MGRN in extracting the WPT-TFFM depth features of wavelet packets with different layer numbers.

(3) The transfer classification accuracy of the WMGRNMM model achieved the expected results on the MFS-RDS experimental bench, demonstrating the good adaptability and generalization ability of the WMGRNMM model on different experimental devices.

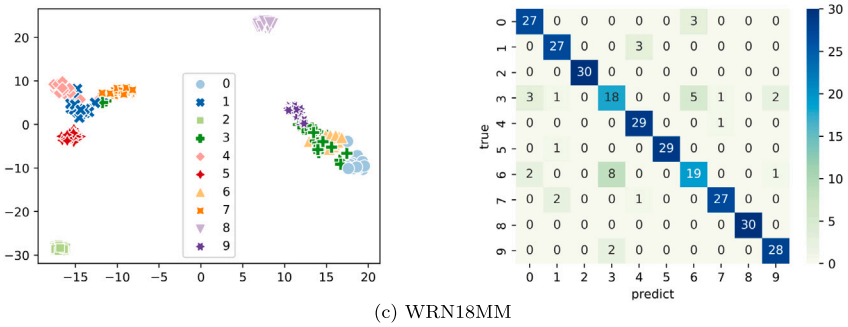
Fig. 11. shows the t-SNE depth features analysis plots and the confusion matrix of these models for the transfer task B->C. From the t-SNE results, we conclude that WMGRNMM can obtain the maximum inter-class distance and the minimum intra-class distance, resulting in better state differentiability of the depth features in the target domain. And the confusion matrix shows that the types and numbers of misclassifications are significantly reduced in WMGRNMM compared with other models. The experimental comparison results further show that WMGRNMM can better reduce the difference in kernel space distribution between the source and target domain depth features and



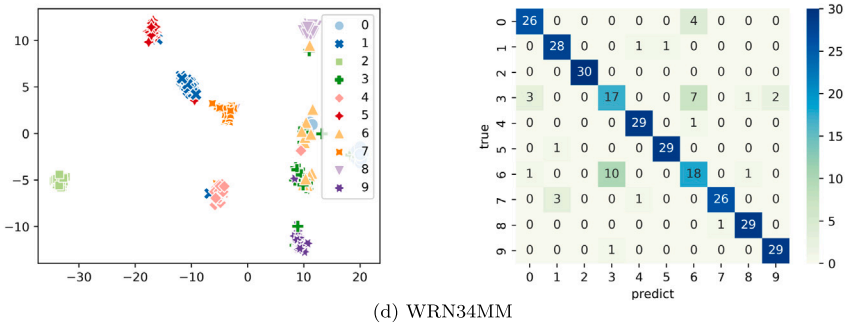
(a) WMGRNMM



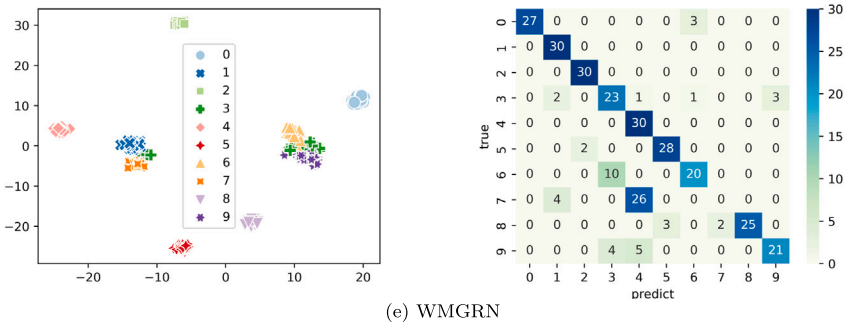
(b) WRNMM



(c) WRN18MM



(d) WRN34MM



(e) WMGRN

Fig. 11. Depth features t-SNE plots for the four model in work transfer task B->C.

Table 8
Classification accuracy of the model with different optimizer and learning rates.

Learning rate	Ada Delta	SGD	RMS Prop	Adam
0.0001	96.56%	96.67%	96.89%	**97.56%**
0.001	96.16%	96.38%	96.72%	**97.14%**
0.01	95.64%	96.19%	96.37%	**97.02%**
0.05	93.28%	93.53%	94.31%	**94.64%**
0.1	87.87%	89.72%	91.06%	**91.72%.**

Table 9
Classification accuracy of the model with different RNB network layers.

Transfer task	RNB-3s	RNB-5s	RNB-7s	RNB-9s	RNB-11s	RNB-13s
A->B	93.83%	96.33%	97.83%	**98.00%**	95.17%	93.50%
A->C	93.17%	93.83%	**97.17%**	95.67%	94.67%	92.83%
B->A	92.33%	**97.33%**	96.33%	95.17%	97.17%	96.67%
B->C	93.33%	**97.67%**	97.50%	96.33%	94.67%	92.67%
C->A	91.67%	94.33%	**95.83%**	**95.83%**	93.33%	91.17%
C->D	93.33%	96.17%	**98.17%**	95.83%	94.50%	92.67%
Average	92.94%	95.94%	**97.14%**	96.14%	94.92%	93.25%

has stronger fault diagnosis transfer capability compared with other models.

The effects of optimizer settings, the number of layers of the RBN network, and the number of MK-MMD Gaussian kernels on the classification accuracy of the WMGRNMM model are further analyzed.

(1) Analysis of the network optimizer and its learning rate settings

Adam, SGD, Ada Delta and RMS Pro are set as the optimizers of the WMGRNMM model respectively, and the classification accuracies at different learning rates are presented in Table 8. The results show that the highest accuracy can be obtained when the Adam is set as the optimizer of the WMGRNMM model. When the learning rate is smaller than or equal to 0.01, the model can maintain a high classification accuracy, and when the learning rate is greater than 0.01, the classification accuracy of the model starts to decrease. Therefore, in model applications, we should set a learning rate smaller than 0.01.

(2) Analysis of RNB network layer settings

Next, we experimentally analyze the effect of the number of RNB network layers on the classification accuracy of the model. The RNB networks with 3, 5, 7, 9, 11, and 13 layers are set for the WMGRNMM model, respectively, and are denoted as RNB-3s, RNB-5s, RNB-7s, RNB-9s, RNB-11s, and RNB-13s. As show in Table 9, when the number of RNB network layers is less than 9, the classification accuracy could be increased by increasing the number of network layers. On the other hand, when the number of layers is greater than 9, more layers would reduce the classification accuracy. This suggests that too many network layers can lead to a regression in the generalization ability of the model, which is analyzed in Section 3.2.

(3) Analysis of the number of MK-MMD kernels

To analyze the effect of the number of Gaussian kernels in MK-MMD on the classification accuracy of the model, 1, 3, 5, 7, and 9 Gaussian kernels are set for the WMGRNMM model, and the experimental results are shown in Table 10. The classification accuracy of the model is the highest when the number of Gaussian kernels is set to 5. This indicates that although the multi-kernel design of MK-MMD can improve the recognition accuracy of the model compared with MMD, too many Gaussian kernels can also lead to a decrease in the generalization ability of the model.

The depth feature plots are shown in Fig. 12. Fig. 12(a) shows the WPT-TFFMs for layers 4, 5, and 6 of a source domain sample and a target domain sample with IRF in transfer task B->C, which are set as the input of the deep network MGRN. Fig. 12(b) and (c) show the outputs of the last convolutional layer of the MGRN with Fig. 12(a) as inputs in the model of WMGRNMM and WMGRN respectively. It can be found that the depth features plots distribution of source and target domain are more similar in WMGRNMM model, which is beneficial to the domain adaptation ability of the MK-MMD.

Table 10
Classification accuracy of the model with different number of MK-MMD Gaussian kernels.

Transfer task	Kernel-1	Kernel-3	Kernel-5	Kernel-7	Kernel-9
A->B	94.17%	96.83%	**97.83%**	97.17%	95.67%
A->C	93.50%	96.33%	**97.17%**	95.67%	93.83%
B->A	**97.33%**	97.00%	96.33%	95.67%	95.33%
B->C	93.33%	96.83%	**97.50%**	96.83%	96.33%
C->A	91.17%	96.00%	95.83%	94.83%	**96.67%**
C->D	94.33%	97.17%	98.17%	**98.33%**	96.83%
Average	93.97%	96.69%	**97.14%**	96.42%	95.78%

Table 11
Classification accuracies of diagnostic models without WPT step on the MFS-RDS experimental platform.

Transfer task	RN18MM	RN34MM	RNMM	MGRN	MGRNMM
A->B	87.83%	85.83%	89.67%	77.83%	**93.83%**
A->C	86.67%	87.17%	90.17%	74.17%	**94.17%**
B->A	83.83%	82.17%	89.17%	76.83%	**92.33%**
B->C	85.83%	84.83%	88.33%	73.17%	**93.50%**
C->A	84.67%	81.83%	86.67%	76.83%	**91.83%**
C->D	90.17%	89.17%	90.67%	80.83%	**94.17%**
Average	86.50%	85.17%	89.11%	76.61%	**93.31%**

Table 12
Classification accuracy of WMGRNMM model in various transfer tasks under different SNR conditions.

Transfer task	3 dB	5 dB	8 dB	10 dB	No noise added
A->B	95.83%	96.83%	97.17%	97.33%	**97.83%**
A->C	95.50%	96.67%	**97.50%**	96.83%	97.17%
B->A	94.17%	95.83%	96.17%	96.00%	**96.33%**
B->C	95.87%	96.17%	97.33%	**97.83%**	97.50%
C->A	93.50%	95.33%	95.67%	94.83%	**95.83%**
C->B	96.33%	97.33%	98.00%	**98.17%**	**98.17%**
Average	95.20%	96.36%	96.97%	96.83%	**97.14%**

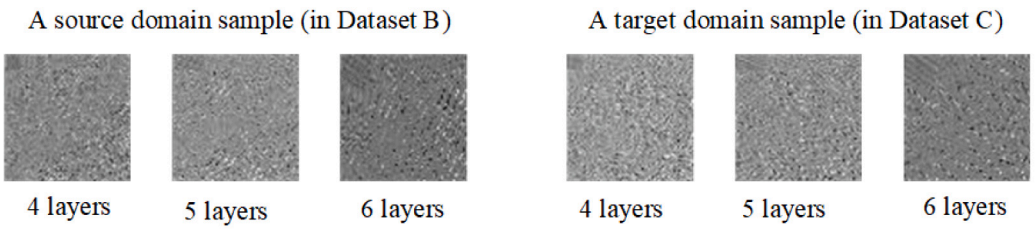
To analyze the effect of WPT step in the diagnostic model, Table 11 presents the classification accuracies of the RNMM, RN18MM, RN34MM, MGRN, and MGRNMM models, which are 5 percentage points lower than the results of models using WPT. The experimental results show that the proposed WPT-TFFM construction method can achieve effective extraction of the fault depth features of the vibration signal and improve the generalization ability of the diagnostic model.

To further verify the anti-noise ability of the WMGRNMM model, the anti-noise experiment is repeated under the MFS-RDS bearing experimental platform, and the two groups of anti-noise experiments are controlled under the same conditions. The results are shown in Table 12. It can be seen that for the 3 dB noise-added data, the diagnostic accuracy of the model decreases slightly, but the decrease is less than 2%, and for the 5 dB, 8 dB, 10 dB noise-added data, the accuracy of the WMGRNMM model does not decrease significantly. The experimental results show that the WMGRNMM model has good and stable anti-noise ability.

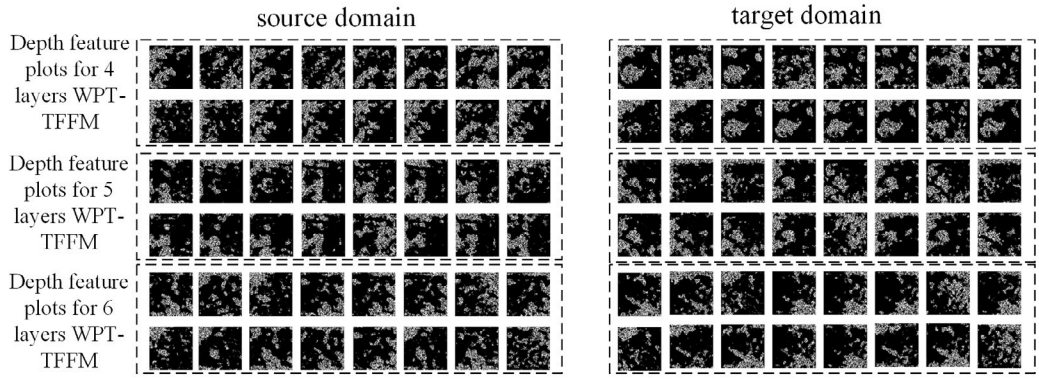
Table 13 shows the results of the effectiveness comparison between the WMGRNMM model and the TCA, JDA, CNN and DANN models in various transfer tasks under the MFS-RDS experimental platform. The comparison results show that the WMGRNMM model has significantly higher diagnostic accuracy than that of TCA, JDA, CNN and DANN models.

5. Conclusions

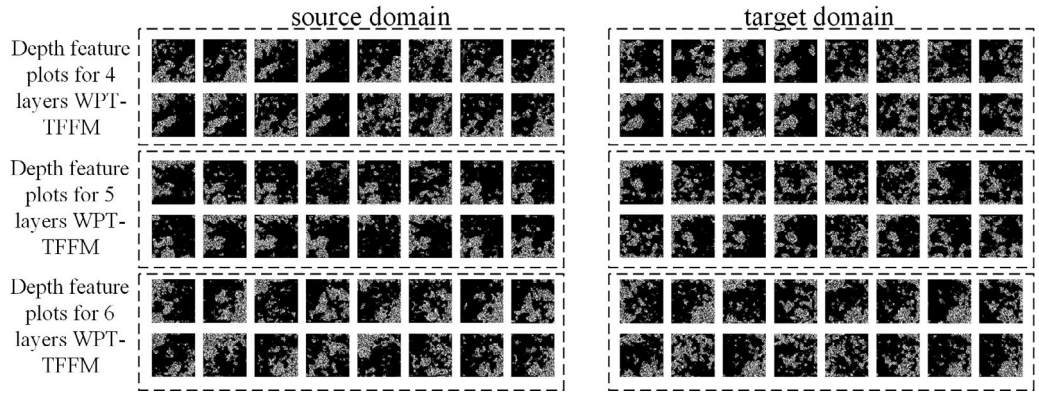
In this paper, we propose a depth transfer fault diagnosis model WMGRNMM. We first extract the time–frequency information of the vibration signal that contains the fault mechanism by WPT to construct the time–frequency feature map WPT-TFFM. Then a multi-group parallel Resnet network structure MGRN is designed to extract the depth features of WPT-TFFM. Finally, MK-MDD is introduced to evaluate



(a) The 3 channels inputs for MGRN (WPT-TFFMs of layers 4, 5 and 6)



(b) The 16*3 channels outputted depth features of the last convolution layer of MGRN in WMGRNMM



(c) The 16*3 channels outputted depth features of the last convolution layer of MGRN in WMGRN

Fig. 12. The depth feature plots of a source domain sample and a target domain sample with IRF in work transfer task B->C.

Table 13
Classification accuracy of different models in various transfer tasks.

Transfer task	TCA	JDA	CNN	DANN	WMGRNMM
A->B	88.83%	90.17%	81.33%	96.33%	**97.83%**
A->C	89.17%	89.50%	78.67%	91.94%	**97.17%**
B->A	86.50%	87.67%	79.17%	92.78%	**96.33%**
B->C	87.67%	88.17%	80.83%	94.17%	**97.50%**
C->A	85.33%	86.67%	78.33%	90.83%	**95.83%**
C->B	91.17%	92.83%	82.17%	95.28%	**98.17%**
Average	88.11%	89.17%	80.08%	93.56%	**97.14%**

the distribution differences of the depth features of the source and target domain data, and the optimization of the MGRN is realized by combining the classification loss of the source domain. After theoretical analysis and experimental results verification, the following conclusions can be obtained.

(1) The proposed vibration signal construction method, WPT-TFFM, can effectively extract the time–frequency features reflecting the fault state from the nonlinear and non-smooth vibration signal with high noise immunity, which is conducive to the extraction of subsequent depth features.

(2) The proposed ResNet-based depth features extraction network is able to solve the problems of small WPT-TFFM size and feature dispersion, and the MGRN network structure can more effectively extract depth features from multi-scale multi-resolution WPT-TFFM with different layers.

(3) By introducing MK-MMD to optimize the depth features extraction network, the classification accuracy of the diagnostic model under variable working condition scenarios can be significantly improved, and the experimental results of two bearing failure testbeds show that the WMGRNMM addresses the working condition transfer diagnosis problem with strong noise immunity under limited amount of training data.

CRediT authorship contribution statement

Xiao Yu: Conceptualization, Methodology, Software. **Zhongting Liang:** Data curation, Writing – original draft. **Youjie Wang:** Visualization, Investigation. **Hongshen Yin:** Programming, Software development. **Xiaowen Liu:** Oversight and leadership responsibility for the research, Development of methodology, Creation of models. **Wanli Yu:** Management and coordination responsibility for the research activity planning and execution. **Yanqiu Huang:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The present work is supported by the National Natural Science Foundation of China (No 52074273), Youth Science and Technology Fund of China University of Mining and Technology (No 2021QN1093), the "Smart Mine" Key Technology R&D Open Fund of China University of Mining and Technology and Zibo Mining Group Co., Ltd (No 2019LH08), National Key R&D Program of China (No 2017YFC0804401).

References

[1] W. Li, Z. Chen, G. He, A novel weighted adversarial transfer network for partial domain fault diagnosis of machinery, IEEE Trans. Ind. Inf. 17 (3) (2020) 1753–1762.
[2] J. Zhu, N. Chen, C. Shen, A new multiple source domain adaptation fault diagnosis method between different rotating machines, IEEE Trans. Ind. Inf. 17 (7) (2020) 4788–4797.
[3] B. Yang, Y. Lei, F. Jia, et al., An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings, Mech. Syst. Signal Process. 122 (2019) 692–706.
[4] W. Li, S. Zhang, S. Rakheja, Feature denoising and nearest-farthest distance preserving projection for machine fault diagnosis, IEEE Trans. Ind. Inf. 12 (1) (2015) 393–404.
[5] L. Cui, J. Huang, F. Zhang, Quantitative and localization diagnosis of a defective ball bearing based on vertical-horizontal synchronization signal analysis, IEEE Trans. Ind. Electron. 64 (11) (2017) 8695–8706.
[6] J. Dong, H. Li, Z. Fan, et al., Time-frequency sparse reconstruction of non-uniform sampling for non-stationary signal, IEEE Trans. Veh. Technol. 70 (11) (2021) 11145–11153.
[7] J. Guo, Z. Shi, D. Zhen, et al., Modulation signal bispectrum with optimized wavelet packet denoising for rolling bearing fault diagnosis, Struct. Health Monit. (2021) 984–1011.
[8] E.Q. Wu, D. Hu, P.Y. Deng, et al., Nonparametric bayesian prior inducing deep network for automatic detection of cognitive status, IEEE Trans. Cybern. 51 (11) (2020) 5483–5496.
[9] G. Yu, A concentrated time-frequency analysis tool for bearing fault diagnosis, IEEE Trans. Instrum. Meas. 69 (2) (2019) 371–381.
[10] B. Cai, Y. Liu, M. Xie, A dynamic-Bayesian-network-based fault diagnosis methodology considering transient and intermittent faults, IEEE Trans. Autom. Sci. Eng. 14 (1) (2016) 276–285.
[11] R. Liu, B. Yang, X. Zhang, et al., Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis, Mech. Syst. Signal Process. 75 (2016) 345–370.
[12] S. Bansal, S. Sahoo, R. Tiwari, et al., Multiclass fault diagnosis in gears using support vector machine algorithms based on frequency domain data, Measurement 46 (9) (2013) 3469–3481.
[13] S. Yin, J. Zhang, Fault diagnosis of the continuous stirred tank heater using fuzzy-possibilistic c-means algorithm, in: 2014 IEEE 23rd International Symposium on Industrial Electronics, ISIE, IEEE, 2014, pp. 2445–2450.
[14] K. Zhang, B. Tang, L. Deng, et al., A hybrid attention improved ResNet based fault diagnosis method of wind turbines gearbox, Measurement 179 (2021) 109491.
[15] Q. Qian, Y. Qin, Y. Wang, et al., A new deep transfer learning network based on convolutional auto-encoder for mechanical fault diagnosis, Measurement 178 (2021) 109352.
[16] K. Zhang, J. Wang, H. Shi, et al., A fault diagnosis method based on improved convolutional neural network for bearings under variable working conditions, Measurement 182 (2021) 109749.
[17] R. Zhao, R. Yan, Z. Chen, et al., Deep learning and its applications to machine health monitoring, Mech. Syst. Signal Process. 115 (2019) 213–237.
[18] Y. Zhang, X. Li, L. Gao, et al., A new subset based deep feature learning method for intelligent fault diagnosis of bearing, Expert Syst. Appl. 110 (2018) 125–142.
[19] J. Sun, C. Yan, J. Wen, Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning, IEEE Trans. Instrum. Meas. 67 (1) (2017) 185–195.
[20] X. Li, W. Zhang, Q. Ding, A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning, Neurocomputing 310 (2018) 77–95.
[21] W. Zhang, C. Li, G. Peng, et al., A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load, Mech. Syst. Signal Process. 100 (2018) 439–453.
[22] R. Liu, G. Meng, B. Yang, et al., Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine, IEEE Trans. Ind. Inf. 13 (3) (2016) 1310–1320.
[23] J. Pan, Y. Zi, J. Chen, et al., LiftingNet: A novel deep learning network with layerwise feature learning from noisy mechanical data for fault classification, IEEE Trans. Ind. Electron. 65 (6) (2017) 4973–4982.
[24] L. Jing, M. Zhao, P. Li, et al., A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox, Measurement 111 (2017) 1–10.
[25] J.H. Lee, J.H. Pack, I.S. Lee, Fault diagnosis of induction motor using convolutional neural network, Appl. Sci. 9 (15) (2019) 2950.
[26] D.T. Hoang, H.J. Kang, Convolutional neural network based bearing fault diagnosis, in: International Conference on Intelligent Computing, Springer, Cham, 2017, pp. 105–111.
[27] K. He, X. Zhang, S. Ren, et al., Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[28] T. Zhang, S. Liu, Y. Wei, et al., A novel feature adaptive extraction method based on deep learning for bearing fault diagnosis, Measurement 185 (2021) 110030.
[29] Y. Li, Y. Ren, H. Zheng, et al., A novel cross-domain intelligent fault diagnosis method based on entropy features and transfer learning, IEEE Trans. Instrum. Meas. 70 (2021) 1–14.
[30] H. Zheng, R. Wang, Y. Yang, et al., Intelligent fault identification based on multisource domain generalization towards actual diagnosis scenario, IEEE Trans. Ind. Electron. 67 (2) (2019) 1293–1304.
[31] F. Shen, Y. Hui, R. Yan, et al., A new penalty domain selection machine enabled transfer learning for gearbox fault recognition, IEEE Trans. Ind. Electron. 67 (10) (2020) 8743–8754.
[32] Z. Tong, W. Li, B. Zhang, et al., Online bearing fault diagnosis based on a novel multiple data streams transmission scheme, IEEE Access 7 (2019) 66644–66654.
[33] H. Zheng, Y. Yang, J. Yin, et al., Deep domain generalization combining a priori diagnosis knowledge toward cross-domain fault diagnosis of rolling bearing, IEEE Trans. Instrum. Meas. 70 (2020) 3501311.
[34] M. Ragab, Z. Chen, M. Wu, et al., Adversarial multiple-target domain adaptation for fault classification, IEEE Trans. Instrum. Meas. 70 (2020) 3500211.
[35] X. Li, W. Zhang, N.X. Xu, et al., Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places, IEEE Trans. Ind. Electron. 67 (8) (2019) 6785–6794.
[36] L. Guo, Y. Lei, S. Xing, et al., Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data, IEEE Trans. Ind. Electron. 66 (9) (2018) 7316–7325.
[37] V. Pandhare, X. Li, M. Miller, et al., Intelligent diagnostics for ball screw fault through indirect sensing using deep domain adaptation, IEEE Trans. Instrum. Meas. 70 (2020) 1–11.
[38] C. Chen, F. Shen, J. Xu, et al., Domain adaptation-based transfer learning for gear fault diagnosis under varying working conditions, IEEE Trans. Instrum. Meas. 70 (2020) 1–10.
[39] Y. Pan, R. Hong, J. Chen, et al., Performance degradation assessment of wind turbine gearbox based on maximum mean discrepancy and multi-sensor transfer learning, Struct. Health Monit. 20 (1) (2021) 118–138.
[40] C. Che, H. Wang, Q. Fu, et al., Deep transfer learning for rolling bearing fault diagnosis under variable operating conditions, Adv. Mech. Eng. 11 (12) (2019) 1–11.
[41] Bearings data center seeded fault test data, 2021, URL https://engineering.case.edu/bearingdatacenter/apparatus-and-procedures.
[42] W. Qian, S. Li, J. Wang, A new transfer learning method and its application on rotating machine fault diagnosis under variant working conditions, IEEE Access 6 (2018) 69907–69917.
[43] J. He, X. Li, Y. Chen, et al., Deep transfer learning method based on 1D-CNN for bearing fault diagnosis, Shock Vib. (2021).