

RESEARCH ARTICLE

A New Weight-Based Dual Domain Adaptation Transfer Model for Bearing Fault Diagnosis Under Noisy and Cross-Domain Conditions

MINGYUAN WANG¹, JISHUN LI^{1,2,3}, AND YUJUN XUE^{1,2,3}¹School of Mechatronics Engineering, Henan University of Science and Technology, Luoyang 471003, China²Henan Key Laboratory for Machinery Design and Transmission System, Henan University of Science and Technology, Luoyang 471003, China³Henan Provincial Collaborative Innovation Center for High End Bearings, Henan University of Science and Technology, Luoyang 471003, China

Corresponding author: Jishun Li (li-jishun@haust.edu.cn)

This work was supported in part by the National Key Research and Development Project under Grant 2018YFB2000501, and in part by the Special Project of Industrial Cluster in National Innovation Demonstration Zone under Grant 201200210400.

ABSTRACT The traditional fault diagnosis methods generally present poor diagnosis accuracy and robustness when faced with complex conditions that involve noisy interference and domain shift. Therefore, a new weight-based dual domain adaptation transfer model for bearing fault diagnosis is proposed. First, based on continuous wavelet transform, the temporal signals are transformed into time-frequency images (TFIs) for enhancing feature representation. Second, the TFIs are used as the input of the improved network which is based on dense and residual connections to complete feature extraction. Third, the proposed transfer model uses local maximum mean discrepancy (LMMD) to adjust data distribution between different working domains and batch nuclear-norm maximization (BNM) to improve the discriminability and diversity of the output matrix. Moreover, the weight controller is used to trade off the contributions of LMMD and BNM during training. Finally, the proposed frequency domain cut can be seen as a simple moving and cutting method to adjust each frequency spectrum of TFIs. In this process, the controlled weight factor is involved to further alleviate noise interference. Case studies show that the proposed model outperforms other methods and works well even in complex conditions mixed by noise and cross-domain.

INDEX TERMS Fault diagnosis, dual domain adaptation, transfer learning, rotor bearing.

I. INTRODUCTION

Rotating machinery plays a crucial role in various industries. The bearings are widely served in almost all rotating machinery, such as motor vehicles, wind turbines, high-speed trains, and mine hoists. Once the bearings fail without fault alarm in advance, it may cause huge economic losses and even the casualties of operators. Therefore, the effectiveness and robustness of bearing fault diagnosis have attracted significant attention in recent years [1], [2], [3].

With the rapid progress in artificial intelligence, there are many breakthroughs in machine learning techniques, such as object detection [4], [5], image classification [6], [7],

natural language processing [8], and control engineering [9]. Consequently, a large number of researchers have applied deep learning techniques to mechanical fault diagnosis [10], [11]. However, the supervised learning methods need numerous labeled samples as the input to train a unique model and a better detection performance only can be obtained in the dataset which has the same data distribution with training samples. In practical applications, the collection of labeled fault signals is difficult, especially for those key components in heavy machinery, such as the mine hoist spindle bearings. In addition, the collected samples often exhibit varying distributions due to the changes in working loads and rotating speeds during operation. Therefore, the data distribution alignment is necessary and crucial for cross-domain conditions. On the other hand, noise interference

The associate editor coordinating the review of this manuscript and approving it for publication was Lei Shu¹.

around intelligent sensors also has an adverse impact on fault diagnosis performance.

To solve the above problems, a lot of researchers focus on improving bearing fault diagnosis accuracy and have proposed various methods based on deep learning methods. Qian et al. proposed a deep discriminative transfer learning network to implement bearing fault transfer diagnosis, and achieved impressive results using three laboratory datasets even in cross-machine conditions [12]. Su et al. proposed a novel evidential deep learning-based adversarial network to improve the cross-domain fault diagnosis performance of rotary machinery, in which evidence-based fault identifier is adopted for known fault identification and exponential evidence score-based unknown estimation is developed for underlying unknown fault recognition [13]. Aiming at solving the negative effects of imbalanced datasets, Gu et al. developed a more robust generative adversarial network based on the self-attention mechanism and gradient penalty. Two experiments were conducted and the encouraging results serve as evidence of the effectiveness of bearing fault diagnosis under imbalanced conditions [14]. Li et al. combined the random sampling, huge kernel, and adaptive multiscale convolution to present an end-to-end fully convolutional network to realize noise suppression and feature extraction, and the state-of-the-art (SOTA) performance demonstrated the better noise adaptability of the proposed network [15]. However, the complex working conditions are not considered in the above methods, which is not in line with the real-world. To solve the discrepancy of data distribution between the source and target domains, the strategies of domain adaptation can be used to improve the generalization ability of trained model [16]. Yang et al. proposed a transfer network to extract features from the raw vibration signals and presented higher diagnosis performance in cross-domain scenarios. The experimental results using laboratory and locomotive bearings datasets demonstrated the maximum mean discrepancy (MMD) regularization term is effective [17]. To address the domain shift problems, Li et al. proposed a novel domain adaptation method based on deep convolutional network and MMD minimization. Without considering the effects of noise interference, the effectiveness of feature transferability is demonstrated through validation experiments with other related works [18]. In 2023, Qian et al. proposed a novel relationship transfer diagnosis framework to improve the abilities of anti-noise and domain adaptation for rotation machinery fault diagnosis. The presented task-irrelevant domain adaptation and task-relevant domain generalization are responsible for enhancing domain confusion and improving the generalization ability of fault classifier, respectively [19]. Based on the combination of the marginal and conditional distribution, Qin et al. proposed a novel deep joint distribution alignment model to simultaneously reduce the discrepancy in marginal and conditional distributions between source and target domains [20]. To avoid the disadvantages of the single model, Li et al. proposed an optimal ensemble deep transfer network for

bearing fault diagnosis based on the combination of transfer learning, domain adaptation and ensemble learning [21]. Based on deep residual network, multi-kernel MMD, and fault diagnosis classifier, Wan et al. proposed a novel domain adaptation model for bearing cross-domain fault diagnosis and demonstrated the superior transfer capability compared with the existing troubleshooting methods [22].

According to the discussion above and the real-world complex working condition, there are key problems in the current typical bearing fault transfer diagnosis: 1) Most current domain adaptation methods focus on aligning the data distribution in two domains but disregard the discriminability and diversity of the output matrix. 2) The independent anti-noise module is neglected, which is necessary, especially for intense noise interference. Aiming to solve these problems, a new weight-based dual domain adaptation transfer model is proposed in this paper to focus on improving the bearing fault diagnosis performance when faced with complex conditions mixed by noise and cross-domain. The main contributions and novelties of this paper are as follows:

- 1) Time-frequency images (TFIs) establishment method using continuous wavelet transform (CWT) and bilinear interpolation algorithm are presented for the improvement of bearing fault representation.
- 2) A new weight-based dual domain adaptation transfer model is proposed to further improve the cross-domain adaptability. LMMD is used for data distribution alignment. The discriminability and diversity of prediction results are enhanced by BNM. Different from the simple combination of LMMD and BNM, the proposed weight controller can dynamically trade off the contributions of LMMD and BNM through a unique exponential function.
- 3) The proposed frequency domain cut method uses the mean value in each frequency spectrum of TFIs as a “scissor” to cut the noise components. The controlled weight factor is involved in this process to further improve the bearing fault diagnosis performance.
- 4) Two case studies are conducted using laboratory and simulated bearing datasets to prove the effectiveness and robustness of the proposed method. Comparison experiments with other SOTA models show that the proposed method can achieve excellent performance even in complex conditions mixed by noise and domain shift.

This paper is organized as follows: The related works are reviewed in Section II. Section III is the methodology, including the establishment of the TFIs, the details of transfer model architecture, DenseNet-53 network structure, dual domain adaptation transfer model, and frequency domain cut method for denoising. In Section IV, case studies are conducted through four validations and two comparison experiments to demonstrate the effectiveness of the proposed method. Conclusions are further drawn in Section V.

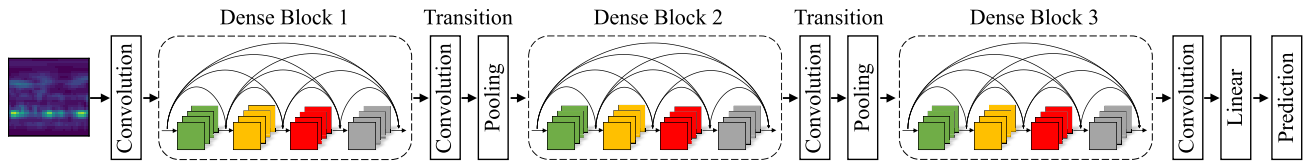


FIGURE 1. The sturcture of DenseNet network.

II. RELATED WORKS

A. DENSE CONVOLUTIONAL NETWORK

Dense convolutional network (DenseNet) [23] is proposed by Huang et al., aims to enhance feature fusion and improve the capability of feature extraction. The core idea of DenseNet is to connect every layer to each other layer in a feed forward fashion. Compared with the traditional neural network that has L connections, there are $L \times (L + 1)/2$ direct connections between each layer and its subsequent layer in DenseNet. The feature maps of all preceding layers are used as inputs for each layer, and its own feature maps are used as inputs for all following layers. DenseNet not only can avoid vanishing gradient, but also can reduce the scale of network, reuse and combine feature maps. As illustrated in Fig. 1, each Dense Block consists of numerous Dense units that are processed by the standard convolutions, batch normalization, and ReLU activation functions layers. Furthermore, the transition layers are designed between each Dense Block to change the size of feature maps via convolution and pooling so that the parameters of the established network can be reduced. To increase inference efficiency, a convolution layer is used as a bottleneck layer to reduce the number of input feature maps. The original DenseNet can be divided into different variations based on the number of layers, including DenseNet-121, DenseNet-169, DenseNet-201, and DenseNet-264.

B. LOCAL MAXIMUM MEAN DISCREPANCY

In recent years, transfer learning was introduced to fault diagnosis and provided a direction to solve the problem of lacking labeled samples, especially for the key component of large-scale mechanical equipment. In 2021, Zhu et al. proposed the local maximum mean discrepancy (LMMD) algorithm [24] and demonstrated it is one of the effective unsupervised domain adaptation methods. LMMD is an improved algorithm based on MMD [25], which can adjust the data distribution between source and target domains to achieve powerful cross-domain diagnosis capability. Similarly, the original feature maps of different domains are mapped to the reproducing kernel Hilbert space (RKHS) to complete the calculation of MMD. As for the advantage of LMMD, it not only takes the global distributions into account, but also aligns the data distributions within the same category of the different domains. In fact, the distance between different categories is enlarged and the distance between the same categories is closed. As shown in Fig. 2(a),

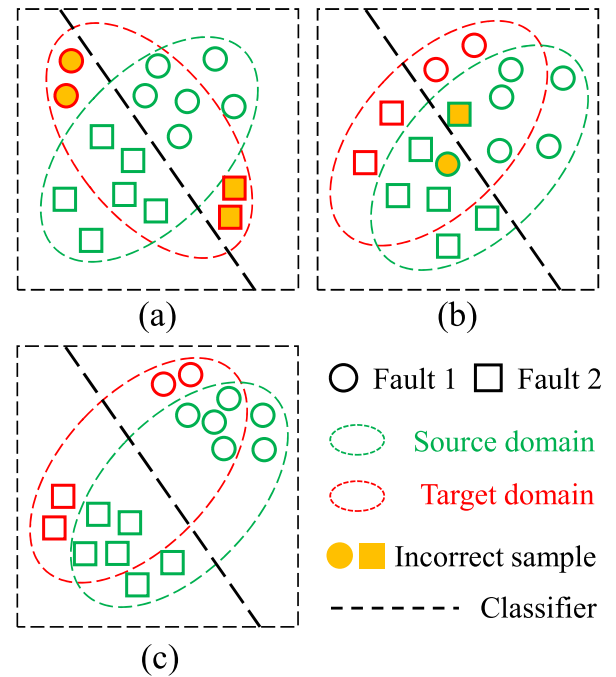


FIGURE 2. Intelligent fault diagnosis using different adaptation domain methods. (a) Without domain adaptation; (b) Domain adaptation based on MMD; (c) Domain adaptation based on LMMD.

the data distribution is different between source and target domains, it is difficult for a trained classifier to perform well in cross-domain tasks even if it can achieve an excellent performance in source domain. Following MMD method, the results depicted in Fig. 2(b), it can be seen that the number of incorrect samples is reduced because of the adjustment of global data distribution. However, a large number of samples are distributed around decision boundary and lead the distance between the same categories is small. It can be observed from Fig. 2(c) that all fault samples are separated successfully whatever source or target domains, LMMD method can decrease the sample density around the decision boundary by enlarging the distance between different categories. $\mathcal{L}_{lmmd}$ is the loss of LMMD and can be obtained through (1). The weight mechanism is used to distinguish different classes and complete the alignment of the same category. At the same time, the probability prediction (soft prediction) is used to measure the discrepancy between source domain D_s and target source domain D_t to alleviate the adverse impact due to the wrong

prediction.

$$\mathcal{L}_{lmm} = \frac{1}{K} \sum_{k=1}^K \left\| \sum_{x_i^s \in D_s} w_i^{sk} \Phi(x_i^s) - \sum_{x_j^t \in D_t} w_j^{tk} \Phi(x_j^t) \right\|_H^2 \quad (1)$$

where x_i^s and x_j^t denote the samples from D_s and D_t , respectively. K is the number of all categories. The weight of x_i^s and x_j^t belonging to class k is denoted by w_i^{sk} and w_j^{tk} . $\Phi(\cdot)$ is an kind of feature map function that can map samples from a low-dimensional space to a high-dimensional RKHS space H . During training, the cross-entropy loss function is used as classification loss, and $\mathcal{L}_{lmm}$ can be used as a regularization term to complete alignment tasks.

C. BATCH NUCLEAR-NORM MAXIMIZATION

Different with the LMMD method, Cui et al. focused on solving the problems of decision boundary with high data density and proposed batch nuclear-norm maximization (BNM) [26] method to reduce the ambiguous detection results under cross-domain scenarios. First, the prediction discriminability can be measured through the result of the Frobenius-norm (F-norm). For a binary classification task, there are four different output matrices O_1 , O_2 , O_3 , and O_4 , as shown in Fig. 3. True labels are presented on the left side and all samples of the first row are denoted by sample 1 which belongs to the normal class and the second is fault class. It can be observed that $\|O_1\|_F = 1.225 > \|O_2\|_F = 1.077$, that means the discriminability of O_1 is higher than O_2 , where $\|\cdot\|_F$ denotes the calculation result of F-norm. However, the output matrix is more prone to the case of O_3 in the early training stage. The F-norm maximization result of O_4 is shown in Fig. 3(d). Although $\|O_4\|_F = 1.223 > \|O_3\|_F = 1.010$ and the discriminability is improved, the prediction result of sample 2 is incorrect. Therefore, using BNM as the optimization objective directly at the early training stage is unreasonable.

Second, the diversity of prediction can be represented by the rank of output matrix. The reason is that the output vectors of different classes can be regarded as linearly independent with each other, whereas the output vectors of the same classes have the characteristic of approximately linear correlation. Consequently, the higher rank of output matrix, the more diversity of the prediction result. According to the literature [27], there is a positive correlation between Nuclear-norm and the rank of output matrix. Furthermore, literature [28] finds the F-norm result could be boosted through increase the Nuclear-norm of prediction matrix. Therefore, BNM uses the Nuclear-norm maximization method to both take the F-norm and matrix rank into account so that the ability of prediction discriminability and diversity can be improved.

$$\mathcal{L}_{bnm} = -\frac{1}{B^t} \|G(x_i^t)\|_* \quad (2)$$

Equation (2) is the optimization objective that only need target domain unlabeled samples X^t to maximize the output

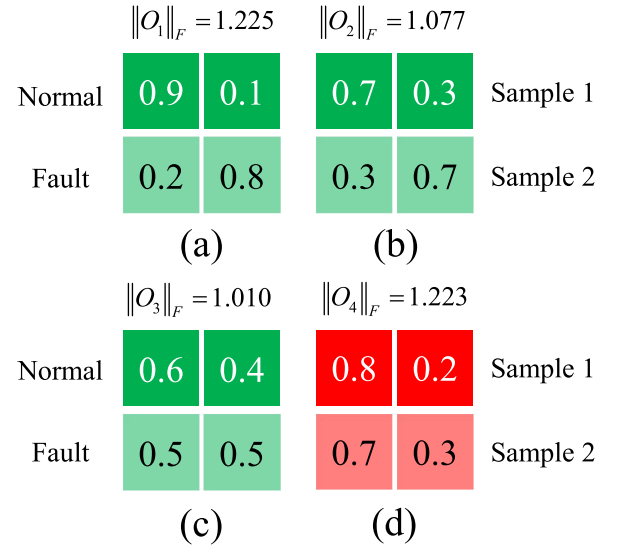


FIGURE 3. The output probability matrix and F-norm maximization results of different samples. (a), (b) and (c) are output matrices without F-norm maximization; (d) is O_4 matrix obtained following the O_3 F-norm maximization.

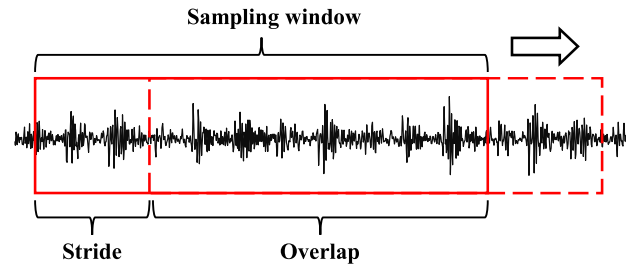


FIGURE 4. The overlap sampling method of sample augmentation. The red box represents the sampling window used to obtain samples that have the same points as the length of sampling window. Sample augmentation is realized by sliding the red box from left to right.

matrix which is the output of network G . B^t is the batch size of prediction results during training and $\|\cdot\|_*$ is the calculation result following Nuclear-norm. The negative sign is introduced to minimize the overall loss value. Like $\mathcal{L}_{lmm}$, the soft prediction is also used to complete the calculation of $\mathcal{L}_{bnm}$. Although the soft prediction has a negative impact during the early training stage, the proposed weight controller in this paper can be used as a solution to improve diagnosis performance and the details will be discussed in Section III-D.

III. METHODOLOGY

A. TIME-FREQUENCY FEATURE IMAGE

1) SAMPLE AUGMENTATION

The sample augmentation is a common and necessary strategy for model training and testing. In this paper, the overlap sampling method is conducted to achieve dataset establishment. Considering the bearing rotating speed and sampling frequency, 1024 signals are regarded as one sample

TABLE 1. Algorithm of time-frequency images establishment.

Step	Description
1	Original vibration signals preparation
2	Sample augmentation based on overlap sampling method
3	Transform 1D signals to 2D TFIs matrix based on CWT
4	Data transformation using max-min normalization
5	Images reconstruction based on bilinear interpolation algorithm
6	TFIs saving and end

to ensure sufficient features can be covered. Fig. 4 is a continuous vibration signals, and the length of sampling window is 1024, stride is set as 256. A large number of samples can be obtained by scanning the raw signals, and this is the foundation for the following TFIs establishment and validation.

2) IMAGE PROCESSING

Due to the advantage of image representation, the one-dimensional (1D) vibration signals need to be transformed into two-dimensional (2D) feature images. As shown in Table 1, the establishment of TFIs mainly consists of CWT [29], max-min normalization, and image reconstruction.

Following the signal preparation and sample augmentation, the next step is CWT using 1D temporal signals. During the TFIs establishment, CWT plays a vital role and can reflect the frequency change at different time points. Compared with original temporal signals and frequency spectrum, CWT has more features and the function is defined as follows.

$$\begin{aligned}
 CWT_x(s, \tau) &= \frac{1}{\sqrt{s}} \int x(t) \psi' \left(\frac{t - \tau}{s} \right) dt \\
 &= \int x(t) \psi'_{s, \tau}(t) dt \\
 &= \langle x(t), \psi_{s, \tau}(t) \rangle
 \end{aligned} \quad (3)$$

where $x(t)$ is the original temporal signal, s and τ are the stretching and translation parameters, respectively. $\psi_{s, \tau}(\cdot)$ is a mother wavelet of CWT and its function as follows.

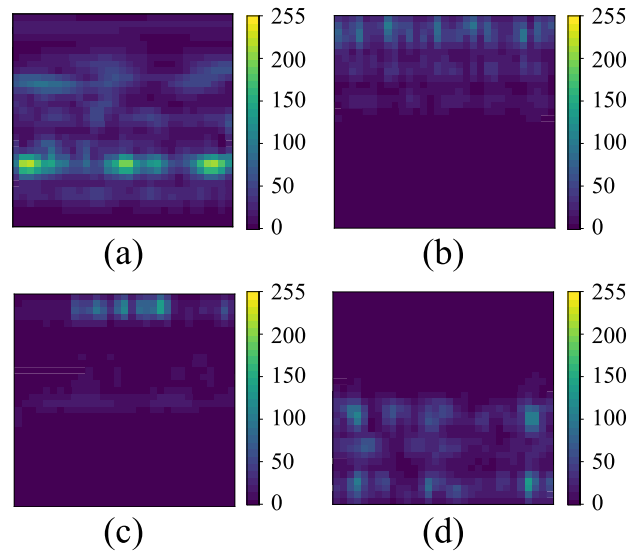
$$\psi_{s, \tau}(t) = \frac{1}{\sqrt{s}} \psi \left(\frac{t - \tau}{s} \right) \quad (4)$$

The choice of wavelet basis function is crucial for obtaining suitable results. In this paper, the Morlet is chosen as the wavelet basis function because it is similar to the fault impulse signal of rotating machinery. The result matrix of CWT can be obtained using a series of wavelets with different s and τ as filters to process temporal signals.

Data transformation is based on 2D CWT matrix $X(i)$. Due to the range of image pix is from 0 to 255. All values in $X(i)$ need to be conducted by max-min normalization which is defined as follows.

$$X(i)_{norm} = \frac{X(i)_{max} - X(i)}{X(i)_{max} - X(i)_{min}} \times 255 \quad (5)$$

where $X(i)$ is the i^{th} sample of the CWT matrix X , $X(i)_{max}$ and $X(i)_{min}$ are the maximum and minimum matrix of X , respectively.

**FIGURE 5.** Different time-frequency images following CWT, normalization, and matrix reconstruction. (a) Normal; (b) Inner race fault; (c) Ball fault; (d) Outer race fault.

The final step is the image reconstruction. In practice, different sampling frequencies and sampling window lengths lead to the shape of $X(i)_{norm}$ being different, which is not beneficial for a unified construction of a feature extraction network. Furthermore, the original size of matrix $X(i)_{norm}$ led to increasing the parameters of the diagnosis network and decreasing the prediction speed. Therefore, it is necessary to reshape $X(i)_{norm}$ to a smaller matrix $X(i)_{32 \times 32}$, which is a 32×32 square matrix. The reason of choosing 32 as matrix order is that the square of 32 is 1024, which is the foundation for the following feature images validation. Instead of using traditional methods, such as principal component analysis and nearest-neighbor interpolation, the bilinear interpolation algorithm [30] is used in this paper to complete the matrix reconstruction. Following CWT, normalization, and matrix reconstruction, four TFIs with different healthy states are shown in Fig. 5, including normal, inner race fault, ball fault, and outer race fault. The different frequency components of impulse excitation make it easy for healthy state recognition.

B. TRANSFER MODEL ARCHITECTURE

In this section, the transfer architecture for bearing fault diagnosis is presented. The training method and testing process is shown in Fig. 6, including dataset of source and target domains, DenseNet-53 for feature extraction, dual domain adaptation based on LMMD and BNM, frequency domain cut to reduce the interference of random noise.

Following the original signals collection, the sample augmentation was applied to realize the extension of samples. Furthermore, the 2D TFIs can be established through the CWT, data normalization, and matrix reconstruction based on 1D temporal signals.

The model training process is displayed on the left side with a green background. It can be observed that transfer

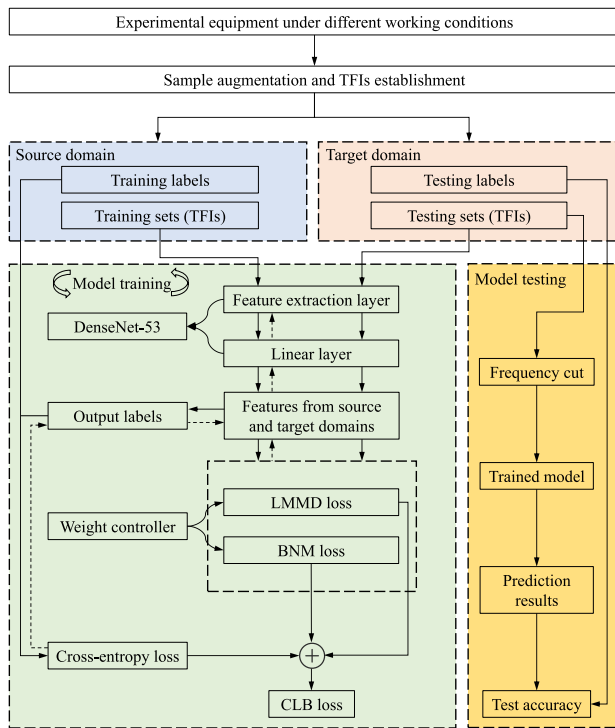


FIGURE 6. Flow chart of the proposed dual domain adaptation transfer model architecture.

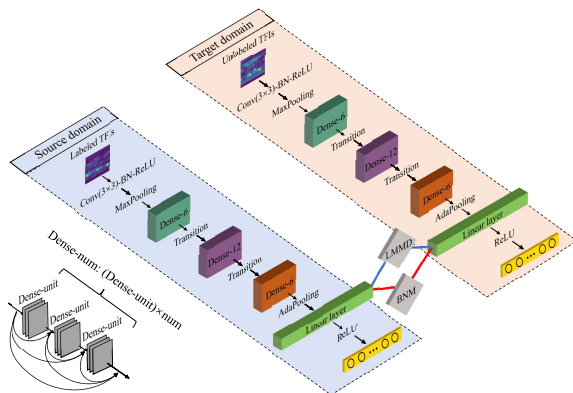


FIGURE 7. Detailed structure of proposed DenseNet-53.

model is an unsupervised method and involves labeled TFIs of source domain and unlabeled TFIs of target domain during training. First, the samples are fed into DenseNet-53 to complete feature extraction. The details of DenseNet-53 will be discussed in Section III-C. Second, to improve the ability of generalization and robustness, the dual domain adaptation method is applied to the linear layer for data distribution adjustment. After the linear layer, the extracted features have two flow trends, one is used to output labels for classification tasks, and another is for data distribution alignment between source and target domains. Section III-D will present the details of dual domain adaptation method which is based on

TABLE 2. The detailed parameters of DenseNet-53.

Layers	Configuration	Output size
Convolution	3×3 conv, stride 1	32×32
Pooling	3×3 max pooling, stride 2	16×16
Dense-6	$[1 \times 1 \text{ conv and } 3 \times 3 \text{ conv}] \times 6$	16×16
Transition	1×1 conv	16×16
Dense-12	$[1 \times 1 \text{ conv and } 3 \times 3 \text{ conv}] \times 12$	8×8
Transition	1×1 conv	8×8
Dense-6	$[1 \times 1 \text{ conv and } 3 \times 3 \text{ conv}] \times 6$	4×4
Linear Layer	Adaptive average pooling	1×1
Classification layer	256D Fully-connected	-
	Fully-connected, softmax	-

the LMMD and BNM, and weight controller. Finally, the losses of cross-entropy, LMMD, and BNM are combined as the optimization objective (CLB loss) to complete model training. Explanatorily, the cross-entropy loss is used for fault classification, LMMD is used for data distribution alignment, and BNM is used to improve the discriminability and diversity of prediction results.

The model testing process is displayed on the right side with an orange background. The labeled TFIs are used to verify the performance of trained model. First, the proposed frequency domain cut (FDC) method is conducted to alleviate the negative impacts of environmental noises. The details of FDC will be explained in Section III-E. Finally, the prediction results and fault diagnosis accuracy can be obtained.

C. NETWORK STRUCTURE ESTABLISHMENT

According to the number of layers in the entire network, DenseNet-53 is named in this paper like the typical DenseNet-121. DenseNet-53 is a simple deep convolutional network which designed to investigate the performance of the proposed dual domain adaptation method. The detailed structure of DenseNet-53 is shown in Fig 7. It can be observed that the shared network DenseNet-53 needs to process the labeled TFIs of source domain (blue background) and the unlabeled TFIs of target domain (yellow background) at the same time. First, the TFIs are processed by a 3×3 convolution, batch normalization and ReLU activation function one after another to extract the initial features. Consequently, the max pooling is used to reduce the size of feature maps. Second, three Dense blocks are used to reduce the computational complexity, including Dense-6, Dense-12 and Dense-6. For example, Dense-6 refers to a dense block that consists of 6 Dense-unit, and 1×1 and 3×3 convolution is the basis of Dense-unit. In the middle of each dense block, there are common transition layers which do convolution and pooling. Third, the adaptive pooling (AdaPooling) adjusts the shape of feature image to 1×1 and the total number of nodes is decided by the batch size during model training. Particularly, the following linear layer is used to complete the improvement of domain adaptation through the proposed dual domain adaptation method. The weight controller can be seen as a coordinator to ensure training stability and effectiveness.

Finally, the simple fully-connected layer has the same neural nodes as the classification tasks according to the practical application. The detailed parameters of DenseNet-53 are shown in Table 2.

D. WIGHT-BASED DUAL DOMAIN ADAPTATION TRANSFER MODEL

The concept of the proposed weight-based dual domain adaptation transfer model consists of two main parts, that is dual domain adaptation algorithm and weight controller. The function of the former is to realize the data distribution alignment between source and target domains using LMMD and BNM. The latter is used to avoid the negative impact of the two loss terms during training. It is well known that the loss function is most important to achieve a superior prediction model. As shown in (6), an improved loss function is proposed to complete the combination of LMMD and BNM. Despite the first standard cross-entropy term which is used for fault classification tasks, other terms are used for regularization. The second term is $\mathcal{L}_{lmmd}$ which focuses on aligning global and local data distribution between source and target domains. The third term is $\mathcal{L}_{bnm}$ which aims at boosting the discriminability and diversity of output matrix. The weight factors $\lambda_l \in (0, 1)$ and $\lambda_b \in (0, 1)$ are used to control the contributions of different regularization terms. The CLB loss function is defined as follows.

$$\begin{aligned} \mathcal{L}_{clb} = & -\frac{1}{n^s} \sum_{i=1}^{n^s} \sum_{k=1}^K l(y_i^s = k) \cdot \log C(M(x_i^s))_k \\ & + \frac{\lambda_l}{K} \sum_{k=1}^K \left\| \sum_{x_i^s \in D_s} w_i^{sk} \Phi(F(M(x_i^s))) \right. \\ & \left. - \sum_{x_j^t \in D_t} w_j^{tk} \Phi(F(M(x_j^t))) \right\|_H^2 \\ & - \frac{\lambda_b}{B'} \|F(M(x_i^t))\|_* \end{aligned} \quad (6)$$

where n^s is the sample number of source domain, K is the label or number of all bearing health states, $l(y_i^s = k)$ is used for class judgment. If the label of y_i^s belong to class k , the result is 1, otherwise, it is 0. $C(\cdot)$ represents the classifier at the end of network. $M(\cdot)$ refers to the feature extraction network before the linear layer involved. D_s and D_t denote source and target domains, respectively. w_i^{sk} and w_j^{tk} serve the same purpose as traditional LMMD method. $\Phi(\cdot)$ can map features to the RKHS to measure the discrepancy of data distribution. The output features from $F(\cdot)$ can be used to align data distribution using LMMD and enhance the diversity and discriminability of prediction using BNM. B' is the batch size of target domain samples. $\|\cdot\|_H^2$ denotes the square distance calculation between source and target domains in RKHS. $\|\cdot\|_*$ is Nuclear-norm.

The weight controller plays an important role in dual domain adaptation transfer model. Fig. 8 illustrates the

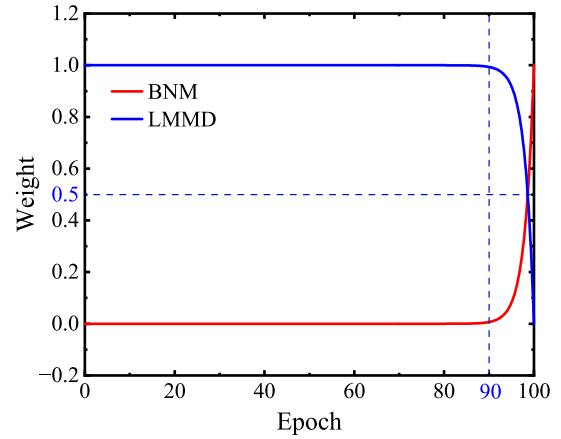


FIGURE 8. The strategy of weight controller during model training.

change strategy deployed by weight controller. The values of λ_l and λ_b are dynamically changing over 100 epochs during training process. It can be found that the change trend of blue and red curves exhibit an exponential change, as they are both controlled by (7) and (8).

$$\lambda_b = e^{\alpha(x-n_e)} \quad (7)$$

$$\lambda_l = -e^{\beta(x-n_e)} + 1 \quad (8)$$

where α and β are related to the change rate of weights and the default value is set as $\alpha = \beta = 0.5$. n_e is the number of training epochs and $\lambda_b + \lambda_l = 1$ no matter in which training stage. As shown in Fig. 8, LMMD takes on a prominent role in aligning the data distribution between source and target domains in the early 90% epochs. Consequently, BNM is rapidly involved in the remaining 10% epochs to improve the diversity and discriminability of the model predictions. The reason for this operation is that BNM with a low accuracy model has an adverse impact in the early training stage as discussed in Section II-C. In addition, as the training progresses, the accuracy of model tends to converge when trained using a single LMMD term. Further improvement can be achieved by substituting BNM with LMMD when the accuracy is converged.

E. FREQUENCY DOMAIN CUT

1) NOISE INTERFERENCE

The proposed frequency domain cut (FDC) method aims at alleviating the strong environmental noises that have serious negative effects on the bearing feature extraction and fault diagnosis. Therefore, the additive Gaussian white noise is added to the raw bearing vibration signals to simulate the real-working condition. Meanwhile, the intensity of noise is determined by signal-to-noise ratio (SNR). The value of SNR can be obtained as follows.

$$SNR = 10 \log_{10}(P_s/P_n) \quad (9)$$

where P_s and P_n denote the power of raw signal and added noise, respectively.

The effects of noise on the raw signals are visually presented in Fig. 9. The contaminated signals are generated by adding noise with SNR=0dB. It is apparent that temporal signals are highly sensitive to noise interference, as the presence of noise completely disrupts the underlying amplitude trends. Although the trend of the frequency spectrum is robust compared with temporal signals, there are a large number of low amplitude frequency components across the entire spectrum. The FDC method focuses on the adjustment of the frequency spectrum to minimize the influence of noise.

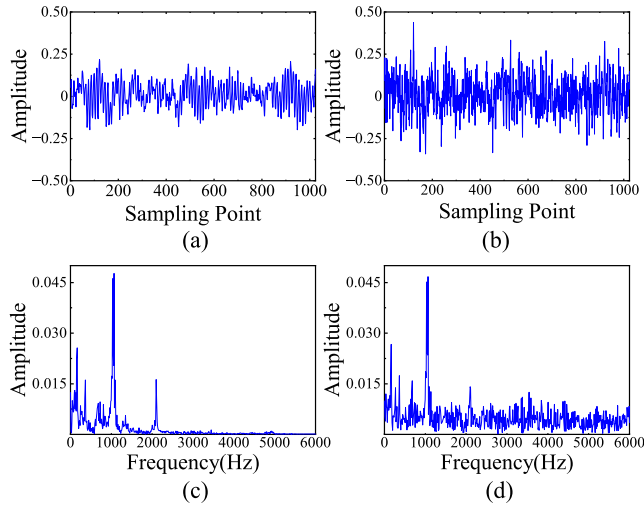


FIGURE 9. The comparison of raw and contaminated temporal signals and frequency spectrum. (a) raw temporal signals; (b) contaminated temporal signals; (c) frequency spectrum of raw signals; (d) frequency spectrum of contaminated signals.

2) NOISE SUPPRESSION

As the analysis above, data of frequency domain is chosen to minimize noise interference. Based on the theory of CWT, TFI is a discrete distribution of frequency components over the time scale. As shown in Fig. 10, the FDC is implemented in each frequency spectrum to minimize noise components. On the left side, a portion of TFI is visualized in a three-dimensional (3D) waterfall diagram. The curves with different colors represent frequency spectrum at different moments. The scissors denote FDC method, which removes the noise signals by cutting. On the right side, the details of cutting implementation detail is presented. It can be found that the lower portion of red frequency curve has been removed by cut line, which is determined by all amplitudes of this frequency spectrum. The mathematical calculation method is defined by (10).

$$X' = \max \left\{ 0, X - \frac{\lambda_c}{n} \text{sum}(X, \text{axis} = 0) \right\} \quad (10)$$

where X' is the output time-frequency matrix following FDC, X is the original time-frequency matrix added contaminated signals, and n is the number of amplitude point in one frequency spectrum. The function of $\text{sum}(X, \text{axis} = 0)/n$

is used to obtain the average amplitude value in x direction. The actual cut line can be obtained by multiplying the weight factor λ_c . $\max(\cdot)$ function is used to take itself if the results are greater than 0, If not, take 0. Although the pixel range of the TFI after FDC will no longer be 0-255, the diagnosis results are not affected. The reason is that the main fault features still remained in this process. Furthermore, the readjustment of pixel range using max-min normalization is not necessary in this step, because it is a linear transformation.

IV. CASE STUDIES

A. EVALUATION METRIC

To demonstrate the effectiveness of proposed methods, the common evaluation metric is used in the following experiments. The accuracy is used to access the diagnosis performance of dual domain adaptation models, and the calculation process is expressed as (11), and its input elements are described in Table 3. TP denotes the positive samples correctly predicted as positive. FP represents the negative samples incorrectly predicted as positive. Similarly, FN and TN can be obtained.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

TABLE 3. Confusion matrix for the accuracy calculation.

Predicted Class	True Class		
	Positive		Negative
	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

B. TRAINING SETUP

During the experiments, a NVIDIA A100-SXM4-40GB GPU was used to accelerate the processes of training. The system RAM size is 84.0 GB. For the programming language and deep learning platform, Python 3.10.12 and Pytorch 2.0.1 were used throughout all of the case studies. Furthermore, stochastic gradient descent (SGD) was used as the optimizer to guarantee the stability of gradient backpropagation and loss decrement results. The batch size, initial learning rate, momentum, and decay rate were set to 64, 0.01, 0.9, and 0.75 for all models, respectively. The training iteration is 100 epochs, and the weights of LMMD and BNM are controlled by the weight controller as discussed in Section III-D.

C. CASE STUDY I

1) DATASET DESCRIPTION

In this section, the public bearing experimental dataset of Case Western Reserve University (CWRU) [31] is used to verify the proposed dual domain adaptation method and the effectiveness of FDC. The superiority of every modules presented above will be discussed through following comparison experiments one after another. The bearing experimental platform is shown in Fig. 11, electric motor for

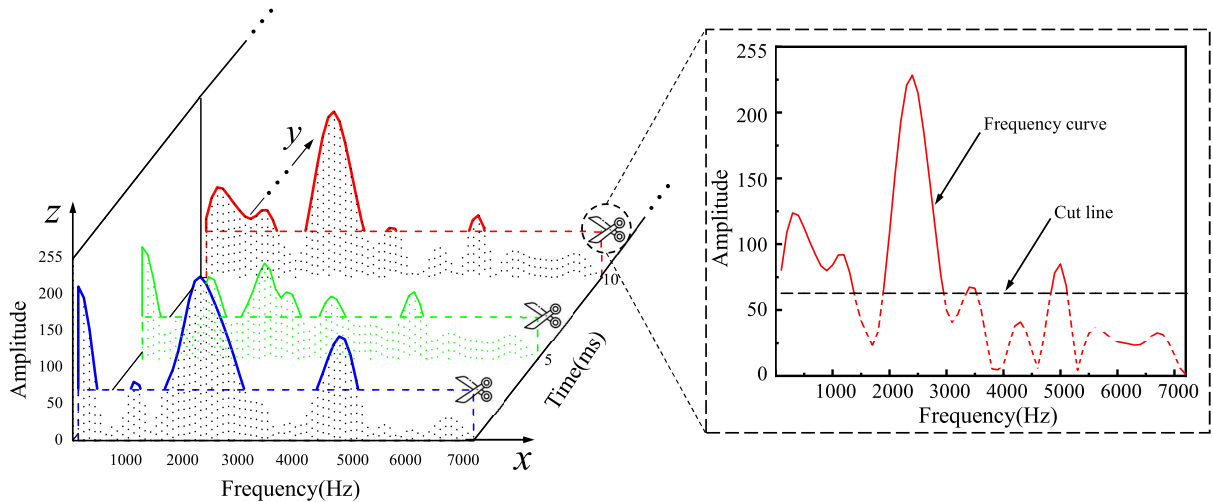


FIGURE 10. Implementation details of TFIs noise suppression based on FDC method.

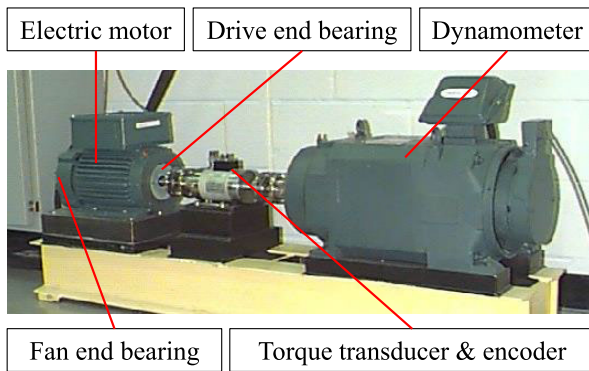


FIGURE 11. Bearing experimental platform of CWRU.

driving a shaft to rotate in a certain speed, drive end and fan end bearings are used to test, the dynamometer and torque transducer/encoder can adjust torque during test.

In this paper, the drive end bearings are selected to do comparison experiments and its type is SKF 6205-2RS JEM. As shown in Table 4, the range of driven motor load is from 0HP to 3HP. Four working conditions A, B, C, and D can be obtained according to the different rotational speeds and loads. All signals of drive bearing are collected at a sampling rate of 12000.

TABLE 4. The parameters of CWRU bearing experimental environment.

Condition	Load(HP)	Rotational speed(rpm)	Sampling rate(Hz)
A	0	1797	12000
B	1	1772	
C	2	1750	
D	3	1730	

In the following experiments, 13 different sets are established according to the type of bearing health state and the degree of damage. As described in Table 5, there are four

bearing health states: normal bearing, inner race faults, ball faults, and outer race faults that are also further divided into three categories according to the specific fault locations (3.00, 6.00, and 12.00 o'clock). Moreover, the degree of damage also is used to further more categories and the artificially generated bearing fault diameter range from 0.007" to 0.021". The sample augmentation and TFIs establishment are same as discussed in Section III-A. As shown in Table 5, each bearing set with fault has approximately 470 TFIs no matter which working condition it is. During experiments the labels are used to distinguish different bearing faults and complete classification tasks.

TABLE 5. Detailed description of CWRU dataset.

Labels	Health state	Fault size(inch)	Number of samples			
			A	B	C	D
C1	Normal	-	950	1887	1892	1894
C2	inner race fault	0.007"	471	474	474	477
C3		0.014"	473	473	473	472
C4		0.021"	474	472	473	474
C5		0.007"	476	471	472	472
C6	ball fault	0.014"	473	474	474	474
C7		0.021"	474	472	474	474
C8	outer race fault-3	0.007"	475	473	472	475
C9	outer race fault-6	0.007"	474	475	471	476
C10	outer race fault-9	0.007"	475	474	475	474
C11	outer race fault-3	0.021"	472	474	475	474
C12	outer race fault-6	0.021"	475	474	475	474
C13	outer race fault-9	0.021"	473	475	476	472

Based on the established CWRU dataset, four validations are implemented one after another in the following sections. As shown in Table 6, the comparison methods, objectives, and real-world applications are presented to give a comprehensive understanding of each validation. Furthermore, each validation has the actual value in real-world applications. For instance, the proposed DenseNet-53 also can be used for object detection, time-frequency images for other signal

analysis, CLB loss and weight controller for improving the performance in the working conditions with load and rotation speed changing, and frequency domain cut method for improving the ability of anti-noise. The details of different methods will be discussed in the following sections.

TABLE 6. Comprehensive description of different validation items and denotations of comparison methods.

Validation items	Denotations	Objectives	Applications
Network structure	ResNet-50	DenseNet-53 superiority	Object detection
	GhostNet		
	MobNet1		
	MobNet2		
	MobNet3		
	DenseNet-53		
Time-frequency images	TDIs FDIs TFIs	Find the best feature images	Signal analysis
Domain adaptation	Direct	Effectiveness of CLB loss and weight controller	Load change
	MMD		
	BNN		
	LMMD		
	Non-Controller		
	Proposed		
Frequency domain cut	Noise with different SNR	Anti-noise ability	Noisy interference

2) NETWORK SUPERIORITY VALIDATION

To demonstrate the superiority of proposed DenseNet-53 in cross-domain conditions, several SOTA classification networks were selected to conduct comparison experiments, namely, MobNet1 [32], MobNet2 [33], MobNet3 [34], GhostNet [35] and typical deep residual model ResNet-50 [36]. As described in transfer model architecture, the TFIs are used as the inputs of all models to achieve a trained model using source dataset and test the diagnosis performance in target dataset. And, in this process, there is no domain adaptation methods and noise interference are involved because the focus of this section is the network structure. As shown in Fig. 12 and Table. 7, the DenseNet-53 achieves the highest average accuracy of 96.27% and the lowest standard deviation of 2.91, which indicates that the dense connection structure is more suitable for the cross-domain scenario. Although the fault diagnosis accuracy of MobNet2 is 2.11% higher than DenseNet-53 in task C→A and 0.13% higher in task D→C, the result has a huge drop in task A→C that is 11.48% lower than the proposed network. It also can be observed that GhostNet and ResNet-50 perform undesirable diagnosis results in some tasks, especially in the difficult tasks between A and D. In summary, the proposed network is superior to other SOTA networks and can better handle domain shifts and realize accurate diagnosis.

3) TIME-FREQUENCY IMAGES VALIDATION

This section focuses on validating the diagnosis performance of TFIs compared with other feature images. As shown in Table. 8, the establishment of time domain images (TDIs) and frequency domain images (FDIs) have some differences with TFIs.

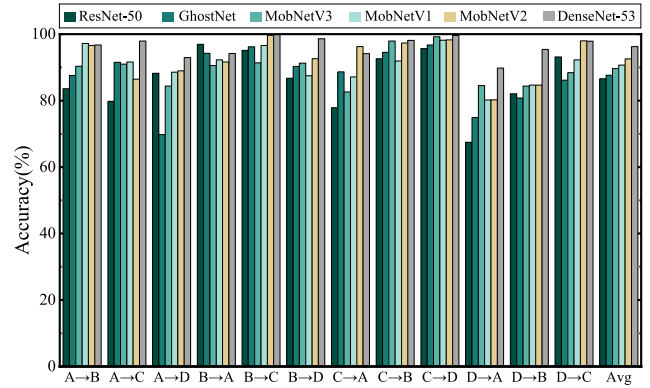


FIGURE 12. The performance of various transfer tasks with different diagnosis networks.

TABLE 7. The diagnosis accuracy(%) and statistical results of the comparison experiments with different diagnosis networks.

	ResNet-50	GhostNet	MobNet3	MobNet1	MobNet2	DenseNet-53
A→B	83.59	87.59	90.33	97.20	96.59	96.77
A→C	79.79	91.53	90.99	91.60	86.46	97.94
A→D	88.23	69.81	84.36	88.56	88.96	92.93
B→A	96.95	94.21	90.55	92.25	91.62	94.16
B→C	95.13	96.20	91.39	96.59	99.66	99.88
B→D	86.76	90.29	91.29	87.50	92.61	98.64
C→A	77.89	88.64	82.62	87.12	96.26	94.15
C→B	92.59	94.53	97.93	91.93	97.33	98.11
C→D	95.68	96.74	99.20	98.20	98.27	99.59
D→A	67.45	74.92	84.53	80.18	80.26	89.78
D→B	82.07	80.79	84.39	84.66	84.67	95.41
D→C	93.13	86.13	88.40	92.26	98.00	97.87
Avg	86.60	87.62	89.67	90.67	92.56	96.27
Std	8.48	8.16	5.01	5.12	5.99	2.91

TABLE 8. The establishment algorithms of the TDIs and FDIs.

Step	TDIs	FDIs
1		Original vibration signals preparation
2		Sample augmentation based on overlap sampling method
3	-	Frequency representation based on FFT
4		Transform 1D signals to 2D matrix
5		Data transformation using max-min normalization
6	-	Images reconstruction using bilinear interpolation

As for TDIs, the temporal signals are used directly to establish feature images. Different with TDIs, the process of FDIs has an additional step of fast fourier transform (FFT). The normalization and reconstruction are similar to TFIs, and the bilinear interpolation algorithm also is used to obtain a 32×32 feature image. Consequently, the obtained feature image is shown in Fig. 13, it can be seen that the TDI has an obvious characteristic of color alternation which is a representation of vibration signal fluctuation. Fig. 13(b) has a bright stripe that reflects the frequency component in the diagnosis signals. The TFI depicted in Fig. 13(c) can reflect the trend of specific frequency change over time.

According to Table 4, 12 domain adaptation tasks are designed to verify the effectiveness of TFIs, that is A→B, A→C, A→D, B→A, B→C, B→D, C→A, C→B, C→D, D→A, D→B and D→C. Take the A→B task as an example,

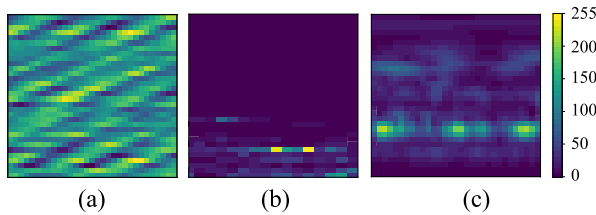


FIGURE 13. The comparison between with different feature images. (a) Feature image based on temporal signal; (b) Feature image based on frequency spectrum; (c) Feature image based on time-frequency representation.

all labeled samples of source domain A and unlabeled samples of target domain B are involved to train DenseNet-53. During testing process, labeled samples of target domain B will be used to verify the performance of trained model. It should be noted that the proposed dual domain adaptation method has not been implemented and target domain signals are not contaminated by noise, because the validation of TFIs is the focus in this section.

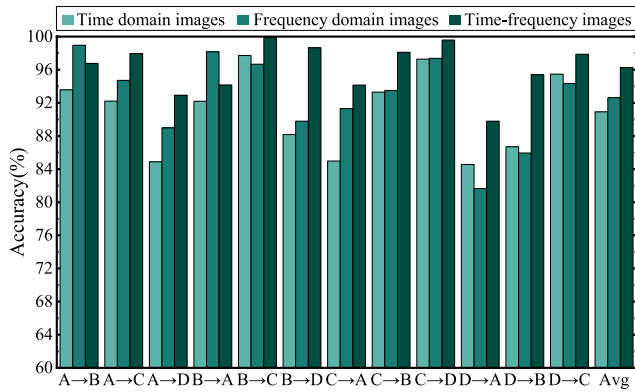


FIGURE 14. The performance of various transfer tasks using different feature images.

10-fold cross-validations were also conducted and the average accuracies were shown in Fig. 14 and Table 9. Although the accuracies of TFIs are not the best in certain tasks such as A→B and B→A, the whole performance outperforms than TDIs and FDIs. The average accuracy of TFIs method is 96.27%, 3.65% higher than FDIs and 5.35% higher than TDIs. The results indicate that TFIs can better represent the characteristics of original signals and are more suitable for domain adaptation. Therefore, the following experiments will be conducted using TFIs.

4) DOMAIN ADAPTATION METHOD VALIDATION

The main focus of this section is to verify the superiority of the proposed duan domain adaptation model and the effectiveness of weight controller. Similarly, the DenseNet-53 also is used to extract features from the input TFIs. Different domain adaptation comparison methods were implemented in the linear layer to realize the alignment of data distribution, including direct prediction without any domain adaptation techniques, transfer using single MMD [25], single BNM,

TABLE 9. The diagnosis accuracy(%) of the comparison experiments using different feature images.

	TDIs	FDIs	TFIs
A→B	93.57	98.96	96.77
A→C	92.21	94.72	97.94
A→D	84.87	88.98	92.93
B→A	92.20	98.16	94.16
B→C	97.72	96.66	99.88
B→D	88.18	89.79	98.64
C→A	84.98	91.31	94.15
C→B	93.29	93.49	98.11
C→D	97.28	97.37	99.59
D→A	84.57	81.64	89.78
D→B	86.69	85.95	95.41
D→C	95.47	94.35	97.87
Avg	90.92	92.62	96.27

single LMMD, and the proposed method. It is noted that the proposed method without weight controller (non-controller) is added as an ablation comparison group, in which the weights of BNM and LMMD are unchanged with training epoch and are set as constant 0.5. Furthermore, the noise is not added to testing samples of target domain since the capability of noise reduction will be discussed in following experiments.

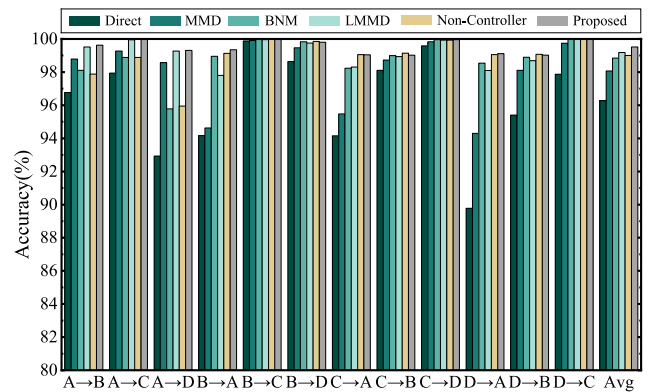


FIGURE 15. The performance of various transfer tasks using different domain adaptation methods.

TABLE 10. The diagnosis accuracy(%) of the comparison experiments using different domain adaptation methods.

	Direct	MMD	BNM	LMMD	Non-Controller	Proposed
A→B	96.77	98.79	98.11	99.52	97.88	99.63
A→C	97.94	99.27	98.89	99.95	98.89	99.97
A→D	92.93	98.57	95.78	99.27	95.95	99.31
B→A	94.16	94.63	98.96	97.80	99.13	99.35
B→C	99.88	99.91	100.00	99.96	100.00	100.00
B→D	98.64	99.47	99.83	99.75	99.85	99.80
C→A	94.15	95.48	98.24	98.30	99.05	99.04
C→B	98.11	98.73	98.99	98.94	99.14	99.02
C→D	99.59	99.83	99.96	99.93	99.93	99.96
D→A	89.78	94.30	98.54	99.10	99.05	99.11
D→B	95.41	98.11	98.90	98.69	99.07	99.02
D→C	97.87	99.74	100.00	100.00	100.00	100.00
Avg	96.27	98.07	98.85	99.18	99.00	99.52

The detailed results are shown in Fig. 15 and Table 10, compared with the direct prediction, the methods with domain adaptation perform an obvious improvement across all transfer tasks. Therefore, the implementation of domain

TABLE 11. The FDC diagnosis accuracies(%) of D→A* transfer task under different SNR and λ_c .

SNR	-2dB	0dB	2dB	4dB	6dB	8dB	λ_c
Without FDC	43.47	53.73	79.20	83.95	90.73	92.75	0
Following FDC	66.33	81.31	88.36	94.71	96.41	97.12	0.2
	75.57	81.29	83.32	85.91	95.96	96.43	0.4
	63.48	83.38	85.06	87.04	88.47	88.21	0.6
	61.46	65.83	84.19	84.34	86.68	83.41	0.8
	70.38	60.93	62.29	57.24	69.40	67.48	1
Best results	75.57	83.38	88.36	94.71	96.41	97.12	-

adaptation techniques is not only necessary but also crucial for real-world applications. The diagnosis average accuracy of LMMD is 99.18% and higher 1.11% than MMD, which demonstrates the distribution alignment of local subdomain and global domain are both important for improving diagnosis accuracies. Different with other methods, the BNM also achieves better results through enhancing the diversity and discriminability of the output matrix. Although the non-controller group can achieve better results in certain transfer tasks, the average accuracy is lower than LMMD. The reason is that BNM and LMMD affect with each other during training. Based on the combination of LMMD and BNM, the proposed method outperforms other models in all transfer tasks. It can be found in each task, the proposed method uses the weight controller to improve the diagnosis accuracies up to 99%. The average accuracy of the proposed model is 99.52%, which is 3.25% higher than the direct prediction model. On the other hand, the proposed dual domain adaptation method achieved a significant improvement of 9.33% compared to the direct prediction method in the challenging transfer task D→A. All of the results demonstrate the proposed method exhibit superior abilities of domain adaptation and feature recognition.

5) FREQUENCY DOMAIN CUT VALIDATION

The 3D visualization was conducted to find out the effectiveness of the proposed FDC method. One of the normal bearing samples with 1024 signals were used as the research instance. Fig. 16(a) is the 3D normalized TFI without noise contamination and the main frequency components are concentrated at around 1000 Hz. Fig. 16(b) presents 3D normalized TFI added Gaussian noise, where SNR=0 dB. It can be found that there are adverse effects on the frequency representation. It is also challenging for the trained model to complete useful feature extraction and fault diagnosis. Therefore, the preprocess method of noise reduction should be implemented during working in a noisy environment. Fig. 16(c) is the processed TFI using the proposed FDC method. It is worth noting that the effects due to noise interference are alleviated and the main frequency features are remained. Consequently, the FDC enables the proposed dual domain adaptation method to perform well even in complex working conditions, which will be demonstrated in the next section.

The impacts of different weight factors λ_c also were investigated on the performance improvement. For the

transfer task D→A, a complex transfer task is introduced and denoted D→A* which means testing dataset of A was added noise. As depicted in Table 11, there are 6 validation groups based on the different SNR values, ranging from -2 dB to 8 dB. Consequently, the FDC with different λ_c was implemented to find out the best diagnosis accuracy. The range of λ_c is set from 0 to 1 with an interval of 0.2. It should be noted that the FDC was not applied when $\lambda_c = 0$.

The results are shown in Fig. 17 and Table 11. Fig. 17(a) shows that the proposed FDC is an effective method to alleviate the effects of noise and improve the diagnosis performance. Furthermore, the greater the noise intensity, the better the effectiveness of noise reduction. For example, the performance of transfer task added -2dB noise has an remarkable improvement, the accuracy increased from 43.47% to 75.57%. The accuracy change trend can be found in Fig. 17(b), all curves are convex and have extreme points. Although the proposed FDC method can eliminate a part of noise when a suitable weight factor λ_c is used, it is important to note that the diagnosis accuracy will gradually decrease when λ_c is increased without limit. The reason is that too much cut will damage the main features of TFIs. Therefore, the actual working environment noise should be considered in order to set the optimal value of λ_c .

6) COMPARISON WITH OTHER STATE-OF-THE-ART MODELS UNDER COMPLEX CONDITIONS

Different with the previous studies, the comparison experiments conducted in this section will both take cross-domain and noise interference into account instead of analyzing the performance under a single condition. Based on the traditional transfer tasks, the samples of target domain are added noise to simulate the real working conditions. To simplify the experiment complexity, all of the mixed noise intensities in different transfer tasks are set to SNR=4dB. Similarly, 12 transfer tasks are conducted, including A→B*, A→C*, A→D*, B→A*, B→C*, B→D*, C→A*, C→B*, C→D*, D→A*, D→B*, and D→C*. To verify the effectiveness of the proposed method, the comparison experiments not only consist of the traditional algorithm but also the other state-of-the-art (SOTA) models. The details of them are as follows.

SVM [37] is a tradition algorithm and widely used in previous researchers. The radial basis function (RBF) is set as a kernel function in this experiment, and the cost parameter c , and kernel parameter γ are set as 1 and 0.1, respectively. CNN-Trans model is a combination of CNN and standard Transformer [38] based on Vision Transformer [39]. The CNN is used to extract features and reduce the image size. Besides, the self-attention mechanism of Transform is used to improve the dependencies between different tokens. The structure of network and training parameters are same as literature [40]. DeepCoral [41] is proposed by Sun et al., which can align correlations of layer activations in deep neural networks. After that the data distribution can be adjusted based on CORAL [42]. The SOTA performance can

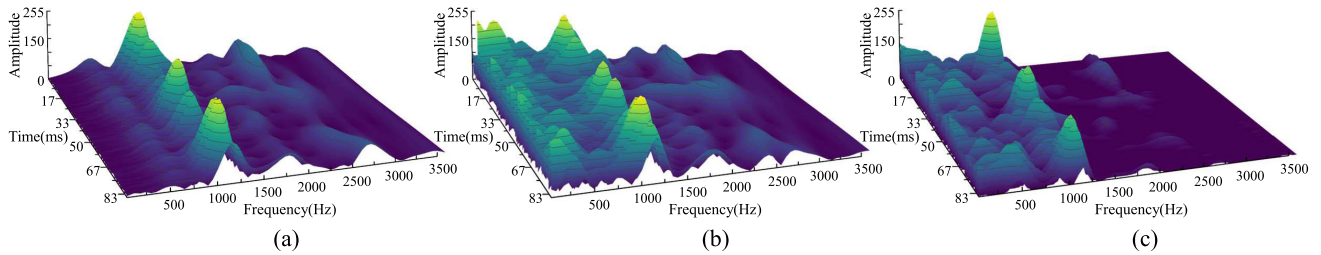


FIGURE 16. 3D visualization of FDC results. (a) Original 3D TFI; (b) 3D TFI with noise of SNR=0dB; (c) result following FDC method.

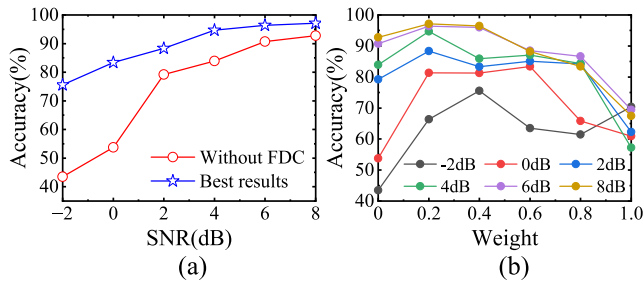


FIGURE 17. The performance curves of $D \rightarrow A^*$ transfer task under different SNR and λ_c . (a) best results following FDC and without FDC; (b) the trend of accuracies with the changes of λ_c .

be achieved on the standard benchmark datasets. DLTCNN [43] is another domain adaptation model based on MMD algorithm and temporal gray image features to improve classification accuracy. The trade-off parameters of λ and learning rate η are set as 0.5 and 0.01, respectively. TICNN [44] aims at processing temporal signals with training interference method. The unique neural network structure and batch setup achieve pretty high accuracy under cross-domain conditions. Dense-LBE-FDC denotes the proposed model in this paper, DenseNet-53 is used to extract features, LBE is the loss algorithm of proposed dual domain adaptation method, and FDC aims at reducing noise interference.

TABLE 12. The diagnosis accuracy(%) of complex transfer tasks using CWRU dataset.

	SVM	CNN-Trans	DeepCoral	DLTCNN	TICNN	Dense-LBE-FDC
A→B*	30.91	56.41	71.30	81.11	76.36	96.88
A→C*	23.05	47.47	76.91	76.76	84.12	95.99
A→D*	23.55	74.15	72.22	80.09	73.98	94.47
B→A*	43.08	51.96	69.96	76.66	80.85	96.01
B→C*	47.77	61.18	82.96	88.18	92.09	97.04
B→D*	46.50	54.52	78.71	85.61	88.94	96.20
C→A*	37.21	69.29	74.39	73.71	75.12	93.78
C→B*	46.12	63.63	82.66	82.57	88.19	95.71
C→D*	43.59	48.09	85.93	89.68	86.44	95.44
D→A*	40.94	47.05	73.84	66.44	75.39	94.71
D→B*	46.63	54.32	78.50	72.64	78.69	96.21
D→C*	48.09	81.98	88.66	84.58	77.47	97.64
Avg	39.79	59.17	78.00	79.84	82.05	95.84

10-fold cross-validations are also conducted in each model and the diagnosis results are presented in Fig. 19 and Table 12. Obviously, the Dense-LBE-FDC outperforms other SOTA algorithms with an average accuracy of 95.84%, which is

higher 13.79% than TICNN. It can be observed that the small fluctuations in results demonstrate that the Dense-LBE-FDC has more robustness against domain shift and noise interference. For the CWRU dataset, the TICNN trained by a small batch size achieves an average accuracy of 82.05% which is better than other SOTA models. By employing the single domain adaptation algorithm to align data distribution, the diagnosis accuracy of DLTCNN and DeepCoral is 79.84% and 78%, respectively. Therefore, the temporal features and single data distribution alignment are not enough for the complex transfer tasks. The CNN-Trans obtains a lower average accuracy of 59.17%. Although the self-attention mechanism is a better choice in the fields of natural language or sequence prediction, it is not suitable for scenarios involving noise interference and domain shift. Additionally, the traditional algorithm SVM is powerless to deal with these complex transfer tasks.

Generally, the inner mechanism of domain adaptation is hard to understand. To solve this problem, the performance visualization is conducted using the t-distributed stochastic neighbor embedding (t-SNE) [45] technique. The output high-dimensional features of last fully-connected layer are fed into the t-SNE algorithm to find out the representation discrepancy following the different models. The five clustering results of complex transfer task $D \rightarrow A^*$ are shown in Fig. 18. The t-SNE algorithm is configured with 1000 iterations and initializes the embedding using the principal component analysis (PCA). As shown in Fig. 18(e), although there are a few overlapped regions, the majority of the other features can be separated since the larger distance between different classes and the smaller distance between same classes. It can be observed from Fig. 18(d), it is impossible for a linear classifier to recognize C5 and C7. Besides, the clustering results of DLTCNN and DeepCoral show that feature points of each class have a large of confusion regions, which suggests that the single domain adaptation models will not perform well in discriminating between some classes. Fig. 18(a) demonstrates that the features following the CNN-Trans model are not divisible.

D. CASE STUDY II

1) BRIEF INTRODUCTION OF MINE HOIST EQUIPMENT

To further verify the effectiveness and robustness of the proposed algorithm, the second case was conducted using

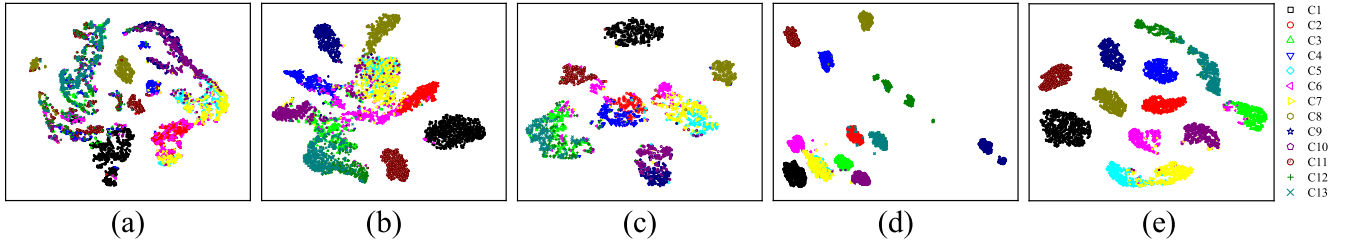


FIGURE 18. The clustering results of different SOTA models under $D \rightarrow A^*$ complex transfer task. (a) CNN-Trans; (b) DeepCoral; (c) DLTCNN; (d) TICNN; (e) proposed Dense-LBE-FDC.

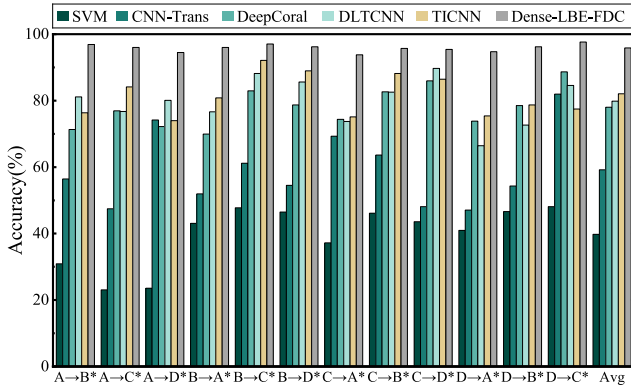


FIGURE 19. The diagnosis accuracy of complex transfer tasks using CWRU dataset.

the spindle bearing of mine hoist equipment. Unlike the laboratory dataset provided by CWRU, the real working conditions of the spindle bearing belongs to heavy radial loads, low rotation speeds, and intense noise interference. Considering the safety of mine hoist, the labeled samples in fault state are difficult to be obtained. Therefore, in this section, simulated fault signals of the spindle bearing will be used to conduct comparison experiments, and the simulation is based on the vibration model and geometry parameters of bearing.

Fig. 20 is an overview of the multi-rope friction mine hoist equipment. The mine hoist equipment is a typical rotor-bearing system, mainly consisting of inertia adjustment, gear reducer, asynchronous motor, friction pulley, spindle device, and supporting main bearing. As shown on the right local diagram of Fig. 20, the bearings are employed to provide support to the spindle unit when the friction pulley is rotating. Consequently, the load change of the friction pulley has a huge impact on the radial loads of bearing. Therefore, it is of great practical significance to carry out fault diagnosis research on hoist spindle bearings.

2) VIBRATION MODEL OF HOIST BEARING FAULT SIGNALS

The generation of bearing fault signals is based on the following vibration analysis. First, the impulse signals are generated by the interaction between rolling element and the damage in bearing inner or outer race. Second, the discrepancy of fault signals depends on the difference in

damage degree and the change of dynamic loads. In this paper, the simulated fault signals are based on the vibration model [46], and the fault signals calculation equation is defined as follows.

$$x(t) = \sum_i A_i s(t - iT_0) + n(t) \quad (12)$$

where $x(t)$ is the generated bearing fault signals, T_0 represents the repetitive cycle of the fault impulses, $n(t)$ is the environmental noise, such as the system vibration caused by brakes and hydraulic stations. A_i is an amplitude expression, $s(t)$ denotes an impulse exponential decay component. The details of A_i and $s(t)$ can be expressed as follows.

$$A_i = \begin{cases} A_m + A_0 \cos(2\pi f_r t) & \text{Inner} \\ A_0 + A_m \cos(2\pi f_r t) & \text{Outer} \end{cases} \quad (13)$$

$$s(t) = e^{-\beta t} \cos(2\pi f_n t) \quad (14)$$

where A_0 is the determinate load, which is determined by the lifting load, A_m represents the alternate load generated due to the rotation of the rotor, f_r denotes the rotational frequency, f_n is the resonance frequency of hoist rotor-bearing system, β is the decay coefficient. During working, the outer race is fixed and the inner race is rotating along with the spindle unit. The amplitude contributions of simulated fault vibration signals are opposite, which was explained in literature [46]. According to the relationship between damage point, rolling elements and races, different working conditions and fault types can be obtained by changing the parameters of vibration model. The detailed explanation is presented as follows.

- 1) Bearing fault type is determined by the impulse exponential decay component $s(t)$. The interval of impulse excitement and decay time varies depending on the damage degree and fault location of spindle bearing. Specifically, the changes of T_0 , β , and f_n can affect the fluctuation of fault signals.
- 2) Working conditions are determined by the amplitude term A_i . For mine hoist system, it is inevitable to work under different loads and rotating speeds. Furthermore, the signals distribution are different since the changes of determinate load A_0 and alternate load A_m . Therefore, the parameters of A_0 , A_m , and f_r can be used to simulate different working conditions.

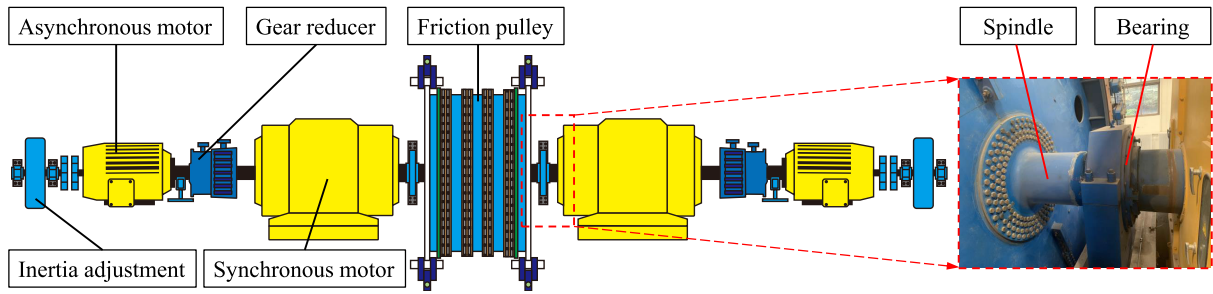


FIGURE 20. The overview of multi-rope friction mine hoist equipment and spindle main bearing.

TABLE 13. Specification parameters of hoist bearing SKF 240/750 ECA/W33.

Designation	d(mm)	D(mm)	f_r (Hz)	BPFI(Hz)	BPFO(Hz)
SKF 240/750 ECA/W33	750	109	0.8	13.042	10.958

TABLE 14. Description of simulation working conditions.

Conditions	A_0	A_m	$A_0 : A_m$	f_r (Hz)	Description
A	8	1	8	0.8	Heavy load
B	4	1	4	0.8	Medium load
C	2	1	2	0.8	Empty load

TABLE 15. Description of the simulation mine hoist spindle bearing dataset.

Labels	f_n	β	T_0	Fault description	Damage degree
C1	30	9			light
C2	60	6	1/BPFO	Outer race fault	medium
C3	90	3			severe
C4	30	9			light
C5	60	6	1/BPFI	Inner race fault	medium
C6	90	3			severe

TABLE 16. The diagnosis accuracy(%) of complex transfer tasks using simulation hoist spindle bearing dataset.

	A→B*	A→C*	B→A*	B→C*	C→A*	C→B*	Avg
SVM	32.04	32.33	32.47	32.51	32.37	32.29	32.34
CNN-Trans	62.19	41.79	69.57	41.54	49.10	64.74	54.82
DeepCoral	60.26	72.10	61.52	62.45	64.77	64.49	64.27
TICNN	66.56	79.96	66.19	83.12	66.99	78.92	73.62
DLTCNN	83.08	84.19	84.05	84.59	84.16	83.12	83.87
Dense-LBE-FDC	99.82	100.00	99.76	100.00	99.79	99.61	99.83

3) DESCRIPTION OF HOIST BEARING

Based on the above analysis, the SKF 240/750 ECA/W33 hoist spindle bearing is used to simulate different fault signals under three working conditions. According to the normal rotor speed of the spindle is range from 40-60r/min, the inner ring rotational frequency is set to 0.8 Hz. Using the outer fault frequency (15) and the inner fault frequency (16), the bearing parameters and fault frequency are shown in Table 13.

$$f_i = \frac{Z}{2} f_r \left(1 + \frac{d}{D} \cos \theta \right) \quad (15)$$

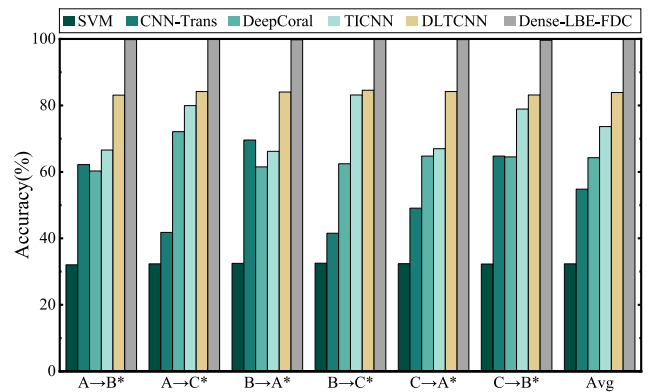


FIGURE 21. The diagnosis accuracy of complex transfer tasks using simulation hoist spindle bearing dataset.

$$f_o = \frac{Z}{2} f_r \left(1 - \frac{d}{D} \cos \theta \right) \quad (16)$$

where f_i is the ball pass frequency inner race (BPFI), f_o denotes ball pass frequency outer race (BPFO). Z is the number of rolling elements, d and D are the inner and outer diameter, respectively. θ denotes the contact angle. f_r denotes the rotating frequency of inner race.

4) ESTABLISHMENT OF TRANSFER TASKS AND BEARING DATASET

As shown in Table 14, three working conditions can be obtained through different combination of A_0 and A_m . As discussed above, A_0 is related to the weight of lifting objects and A_m is determined by multi factors, such as rotation speed, spindle stiffness, and installation error. Therefore, the ratio of $A_0 : A_m$ is used to simulate different working conditions instead of the exact value which is not easy to be obtained in practical engineering applications. In the mine hoist rotor-bearing system, the determinate load A_0 is much higher than the alternate load A_m . Therefore, three ratios of 8, 4, and 2 are used to simulate heavy, medium, and empty loads. Consequently, six complex transfer tasks can be established as A→B*, A→C*, B→A*, B→C*, C→A*, and C→B*.

Similarly, the hoist system resonance frequency f_n is not only related to the structure of support bearing, but is also

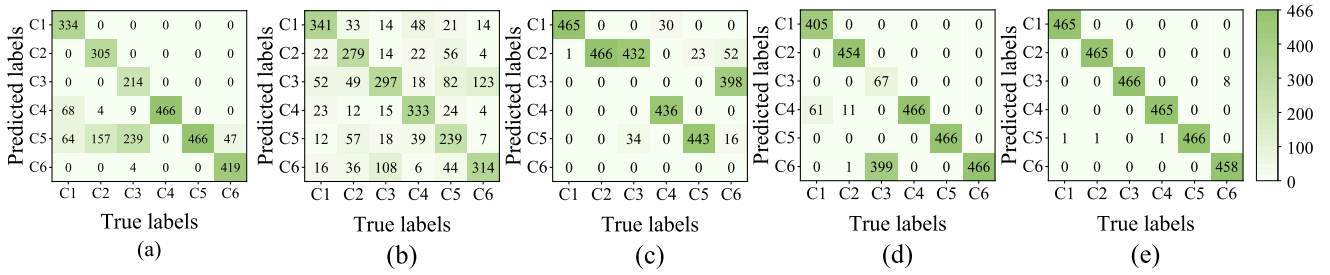


FIGURE 22. The confusion matrices of different SOTA models under the C→B* complex transfer task. (a) CNN-Trans; (b) DeepCoral; (c) TICNN; (d) DLTCNN; (e) proposed Dense-LBE-FDC.

influenced by the characteristics of the entire mine hoist equipment system. As shown in Table 15, different f_n are used to simulate three levels of bearing fault with different damage degrees: light, medium, and severe. Besides, the degree of fault damage can affect the duration of amplitude oscillation. A large value of β will cause a rapid oscillation, and vice versa. Therefore, the higher the damage level, the smaller the value of β is set. It can be seen that the simulation spindle bearing dataset is established by two fault types and three damage degrees. The sampling frequency is set to 256 Hz to ensure a appropriate number of signals to capture the rotating characteristics of the spindle bearing. The establishment of TFIs is same as the discussion in section III-A, the sampling length is 1024 and the stride is 256. Lastly, each simulation vibration signal consists of 120000 data points, and there are 466 TFIs in each category.

5) COMPARISON EXPERIMENTS AND ANALYSIS

Considering the complex working conditions of mine hoist, the more intense noise signals (SNR=-2 dB) are added to the samples of target domain. Moreover, the number of categories is different with CWRU dataset so that the classification sensitivity of the proposed model can be observed. As shown in Fig. 21 and Table 16, Dense-LBE-FDC also enjoys the best diagnosis performance than other models, and the average accuracy is 99.83% which is a huge improvement compared with the traditional algorithm SVM. Remarkably, the proposed method consistently achieves exceptional accuracy higher than 99% across all transfer tasks. The result of TICNN in B→C* has a large fluctuation than other scenarios indicates that the temporal signals are not stable for these complex conditions. Compared with the results in Case I, the discrepancy between different transfer tasks is smaller since all the environmental factors taken into account are impossible during bearing dataset simulation, such as the effects of surrounding temperature and humidity. Thus, the superiority of the proposed algorithm can be demonstrated through comparison analysis.

Similarly, the result visualization is conducted in this experiment using the complex transfer task of C→B*. Unlike the clustering analysis in CWRU dataset, the result

confusion matrices are shown in Fig. 22. The true and predicted labels are indicated on the horizontal and vertical axis, respectively. The diagnosis results are presented in the intersection. A notable observation is that the samples of C3 is difficult to diagnosis in a high accuracy. However, this problem can be solved by proposed Dense-LBE-FDC. Consequently, the bearing fault diagnosis has an obvious improvement to 99.61%.

V. CONCLUSION

This paper proposed a new weight-based dual domain adaptation transfer model for bearing fault diagnosis based on CWT, DenseNet, LMMD, and BNM. Furthermore, the FDC method is presented to reduce the noise interference. First, image processing based on CWT is conducted to transfer temporal signals to TFIs. Second, the DenseNet-53 is not only used to extract features but also to realize domain adaptation. Third, the proposed dual domain adaptation based LMMD and BNM is implemented to complete data alignment during training. At the same time, an exponential weight controller is presented and can balance the advantages of LMMD and BNM. Finally, the FDC is a cheap algorithm that can improve diagnosis accuracy through controlling the weight of the average cut line in frequency domain. The validations of feature images, domain adaptation, and noise reduction effectively demonstrate the superiority of TFIs, dual domain adaptation, and FDC method, respectively. The bearing fault diagnosis performance of the proposed method is demonstrated via the complex transfer tasks in two case studies. Using laboratory and simulation hoist spindle bearing dataset, the proposed model both outperforms than other SOTA algorithms. Through clustering analysis and confusion matrix visualization, it can be found that the results of proposed model are more divisible and robustness.

There are some limitations in this study. Due to the proposed method focus on solving the problem of domain shift and noise interference in the same machine, the cross-machine fault diagnosis condition is not suitable. As for the weight controller, the hyperparameters of α and β are used to control the changing speed of weights during training. The values of them need to be adjusted to achieve the best performance when the maximum number of iterations is

changed. Thus, the ability of cross-machine and the method of parameter self-adjusting are worth investigating in future studies.

DECLARATION OF CONFLICT OF INTEREST

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

REFERENCES

- [1] L. Wan, K. Gong, G. Zhang, X. Yuan, C. Li, and X. Deng, "An efficient rolling bearing fault diagnosis method based on spark and improved random forest algorithm," *IEEE Access*, vol. 9, pp. 37866–37882, 2021.
- [2] M. T. Pham, J.-M. Kim, and C. H. Kim, "Rolling bearing fault diagnosis based on improved GAN and 2-D representation of acoustic emission signals," *IEEE Access*, vol. 10, pp. 78056–78069, 2022.
- [3] P. Xu and L. Zhang, "A fault diagnosis method for rolling bearing based on 1D-ViT model," *IEEE Access*, vol. 11, pp. 39664–39674, 2023.
- [4] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.
- [5] L. Jiao, F. Zhang, F. Liu, S. Yang, L. Li, Z. Feng, and R. Qu, "A survey of modern deep learning-based object detection models," *IEEE Access*, vol. 126, pp. 128837–128868, 2019.
- [6] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10778–10787.
- [7] H. Sun, X. Zheng, and X. Lu, "A supervised segmentation network for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 30, pp. 2810–2825, 2021.
- [8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.
- [9] Z.-M. Zhai, M. Moradi, L.-W. Kong, B. Glaz, M. Haile, and Y.-C. Lai, "Model-free tracking control of complex dynamical trajectories with machine learning," *Nature Commun.*, vol. 14, no. 1, p. 5698, Sep. 2023.
- [10] Z. Zhang, F. Zhou, and S. Li, "A cross working condition multiscale recursive feature fusion method for fault diagnosis of rolling bearing in multiple working conditions," *IEEE Access*, vol. 10, pp. 78502–78518, 2022.
- [11] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.
- [12] Q. Qian, Y. Qin, J. Luo, Y. Wang, and F. Wu, "Deep discriminative transfer learning network for cross-machine fault diagnosis," *Mech. Syst. Signal Process.*, vol. 186, Mar. 2023, Art. no. 109884.
- [13] Z. Su, W. Jiang, B. Zhang, S. Feng, L. Cui, and Y. Qin, "Evidential deep learning-based adversarial network for universal cross-domain fault diagnosis of rotary machinery," *IEEE Sensors J.*, vol. 23, no. 19, pp. 22823–22831, Oct. 2023.
- [14] X. Gu, Y. Yu, L. Guo, H. Gao, and M. Luo, "CSWGAN-GP: A new method for bearing fault diagnosis under imbalanced condition," *Measurement*, vol. 217, May 2023, Art. no. 113014.
- [15] F. Li, L. Wang, D. Wang, J. Wu, and H. Zhao, "An adaptive multiscale fully convolutional network for bearing fault diagnosis under noisy environments," *Measurement*, vol. 216, Jul. 2023, Art. no. 112993.
- [16] S. Zhang, L. Su, J. Gu, K. Li, L. Zhou, and M. Pecht, "Rotating machinery fault detection and diagnosis based on deep domain adaptation: A survey," *Chin. J. Aeronaut.*, vol. 36, no. 1, pp. 45–74, Jan. 2023.
- [17] B. Yang, Y. Lei, F. Jia, and S. Xing, "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," *Mech. Syst. Signal Process.*, vol. 122, pp. 692–706, May 2019.
- [18] X. Li, W. Zhang, Q. Ding, and J.-Q. Sun, "Multi-layer domain adaptation method for rolling bearing fault diagnosis," *Signal Process.*, vol. 157, pp. 180–197, Apr. 2019.
- [19] Q. Qian, J. Zhou, and Y. Qin, "Relationship transfer domain generalization network for rotating machinery fault diagnosis under different working conditions," *IEEE Trans. Industr. Inform.*, vol. 19, no. 19, pp. 9898–9908, Sep. 2023.
- [20] Y. Qin, Q. Qian, J. Luo, and H. Pu, "Deep joint distribution alignment: A novel enhanced-domain adaptation mechanism for fault transfer diagnosis," *IEEE Trans. Cybern.*, vol. 53, no. 5, pp. 3128–3138, May 2023.
- [21] X. Li, H. Jiang, R. Wang, and M. Niu, "Rolling bearing fault diagnosis using optimal ensemble deep transfer network," *Knowl.-Based Syst.*, vol. 213, Feb. 2021, Art. no. 106695.
- [22] L. Wan, Y. Li, K. Chen, K. Gong, and C. Li, "A novel deep convolution multi-adversarial domain adaptation model for rolling bearing fault diagnosis," *Measurement*, vol. 191, Mar. 2022, Art. no. 110752.
- [23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
- [24] Y. Zhu, F. Zhuang, J. Wang, G. Ke, J. Chen, J. Bian, H. Xiong, and Q. He, "Deep subdomain adaptation network for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 4, pp. 1713–1722, Apr. 2021.
- [25] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. ICML*, Feb. 2015, pp. 97–105.
- [26] S. Cui, S. Wang, J. Zhuo, L. Li, Q. Huang, and Q. Tian, "Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3940–3949.
- [27] J. Jiao, K. Liang, C. Ding, and J. Lin, "Towards prediction constraints: A novel domain adaptation method for machine fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 18, no. 10, pp. 7198–7207, Oct. 2022.
- [28] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, Jan. 2010.
- [29] J. Gu, Y. Peng, H. Lu, X. Chang, and G. Chen, "A novel fault diagnosis method of rotating machinery via VMD, CWT and improved CNN," *Measurement*, vol. 200, Aug. 2022, Art. no. 111635.
- [30] H. Raveendran and D. Thomas, "Image fusion using LEP filtering and bilinear interpolation," 2014, *arXiv:1407.3986*.
- [31] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mech. Syst. Signal Process.*, vols. 64–65, pp. 100–131, Dec. 2015.
- [32] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
- [33] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
- [34] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1314–1324.
- [35] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1577–1586.
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [37] Z. Wang, Q. Liu, H. Chen, and X. Chu, "A deformable CNN-DLSTM based transfer learning method for fault diagnosis of rolling bearing under multiple working conditions," *Int. J. Prod. Res.*, vol. 59, no. 16, pp. 4811–4825, Aug. 2021.
- [38] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008.
- [39] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

- [40] M. Wang, J. Li, and Y. Xue, "A new defect diagnosis method for wire rope based on CNN-transformer and transfer learning," *Appl. Sci.*, vol. 13, no. 12, p. 7069, Jun. 2023.
- [41] B. Sun and K. Saenko, "Deep CORAL: Correlation alignment for deep domain adaptation," 2016, *arXiv:1607.01719*.
- [42] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *Proc. AAAI Conf. Artif. Intell.*, Mar. 2016, vol. 30, no. 1, pp. 1–8.
- [43] J. Zhu, N. Chen, and C. Shen, "A new deep transfer learning method for bearing fault diagnosis under different working conditions," *IEEE Sensors J.*, vol. 20, no. 15, pp. 8394–8402, Aug. 2020.
- [44] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–453, Feb. 2018.
- [45] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, Nov. 2008.
- [46] F. Cong, J. Chen, G. Dong, and M. Pecht, "Vibration model of rolling element bearings in a rotor-bearing system for fault diagnosis," *J. Sound Vibrat.*, vol. 332, no. 8, pp. 2081–2097, Apr. 2013.



MINGYUAN WANG received the B.S. degree in mechanical design, manufacturing and automation from the Henan University of Science and Technology, Luoyang, Henan, China, in 2013, where he is currently pursuing the M.S. degree with the Henan Key Laboratory for Machinery Design and Transmission System, School of Mechatronics Engineering.

His current research interests include transfer learning, and fault diagnosis and its applications to other science and engineering fields.



JISHUN LI received the Ph.D. degree in mechanical manufacturing and automation from Shanghai Jiao Tong University, Shanghai, China, in 1996.

He is currently a Professor and a Ph.D. Supervisor with the School of Mechatronics Engineering, Henan University of Science and Technology. He has presided over the National Basic Research Program of China and the National Key Research and Development Project. His research interests include fault diagnosis, signal processing, and

intelligent manufacturing.



YUJUN XUE received the Ph.D. degree in mechanical design and theory from Shanghai Jiao Tong University, Shanghai, China, in 2002.

He is currently a Professor and a Ph.D. Supervisor with the School of Mechatronics Engineering, Henan University of Science and Technology. He has published two books and more than 50 academic articles included in science citation index (SCI)/engineering index (EI). His research interests include measurement technique, mechanical

system design, and performance evaluation. He is also a member of the Chinese Mechanical Engineering Society.

...