

TRANSFER LEARNING ROLLING BEARING FAULT DIAGNOSIS METHOD BASED ON DEEP DOMAIN ADAPTIVE NETWORK

YU LIAO^{1,3}, JIAHAO GENG¹, LI GUO^{2,*}, BING GENG¹, KUN CUI¹ AND RUNZE LI⁴

¹School of Intelligent Science and Engineering
Hubei Minzu University
No. 39, Xueyuan Road, Enshi 445000, P. R. China
{yliao; 202130304; gengbing; 202330311}@hbmzu.edu.cn

²School of Electrical Engineering
Anhui Polytechnic University
No. 3, Beijing Middle Road, Wuhu 241000, P. R. China
*Corresponding author: lguo@ahpu.edu.cn

³College of Electronics and Information Engineering
Sichuan University
No. 29, Jiuyanqiao Wangjiang Road, Chengdu 610064, P. R. China

⁴College of Automation Engineering
Nanjing University of Aeronautics and Astronautics
No. 29, Jiangjun Road, Nanjing 211100, P. R. China
lirunze@nuaa.edu.cn

Received March 2024; revised July 2024

ABSTRACT. *The variation in motor operating conditions in nonsmooth dynamical systems leads to discrepancies in the distribution of condition monitoring data and label acquisition. To address this issue, we employ the transfer learning domain adaptation method and propose a deep domain adaptive network for assessing the state of motor rolling bearings in cross-working conditions. Initially, the vibration signals are preprocessed by FFT and SincNet to extract key features related to state assessment and perform feature fusion. Subsequently, a convolutional neural network is built to map the known label features of both the source and target domains into a shared space, enabling the extraction of common features. Employing the CBAM attention mechanism to enhance the network's feature expression capability aids in comprehending and utilizing input information more effectively. To enhance domain confusion, we combined the Maximum Mean Difference (MMD) and Correlation Alignment (CORAL) as novel measures for reducing discrepancies in the feature distribution of the same state under varying working conditions. Lastly, I-Softmax loss is utilized to improve network evaluation capability. Experimental results from multiple migration tasks conducted on two datasets demonstrate the excellent performance and adaptability of the proposed method in cross-condition fault diagnosis, surpassing the transfer learning method's focus on global domain adaptation.*

Keywords: Deep domain adaptation, Transfer learning, Fault diagnosis, Convolutional neural networks

1. **Introduction.** In recent years, the rapid development of power electronics technology and the promotion of new energy technologies have made the study of non-smooth power systems a pivotal area in the field of electricity. These advancements provide novel perspectives and solutions for constructing efficient, reliable, and sustainable power systems. Electric motors, critical in industrial manufacturing, transportation, medical equipment, energy, and other fields' applications, rely heavily on the integrity of components like

rolling bearings. These bearings, as crucial components of electric motors, play an important role in supporting and rotating the motor's rotor shaft. The complex operating environment of motors makes rolling bearings susceptible to factors such as long-term operation, load changes, and environmental conditions. This non-smoothness in motors can lead to various bearing failures, resulting in changes in operating conditions and severely impacting the normal operation and safety of equipment. According to relevant statistics, approximately 40% of motor failures are caused by bearing faults [1]. This underscores the importance of timely and accurate condition monitoring and fault diagnosis.

The rapid advancements in industrial big data and measurement technology have propelled deep learning-based state assessment methods to become highly anticipated research directions within the state assessment field [2,3]. These methods exploit deep learning technology to automatically extract state evaluation-related features, presenting significant advantages compared to conventional machine learning and signal analysis-based approaches. They can significantly reduce the reliance on human expertise and produce more accurate state evaluation results. Nonetheless, a crucial requirement for the successful application of deep learning methods is the availability of large-scale and high-quality annotated samples. In practical industrial settings, acquiring a sufficient quantity and quality of annotated samples is challenging, thereby constraining the enhancement of robustness and generalization ability of deep learning models. Moreover, state evaluation models based on deep learning typically necessitate training and test data to conform to a shared probability distribution. However, in practical applications, meeting this requirement is challenging. For instance, the data collected for motor monitoring inevitably exhibits considerable variations in data distribution due to factors such as workload, noise interference, and fault severity. Consequently, the lack of labeled samples imposes significant limitations, thereby intensifying the demand for advanced diagnostic techniques.

Recent strides in transfer learning and domain adaptation have been pivotal in addressing these challenges. Transfer learning, especially via domain adaptation methods, leverages unlabeled data from target domains to seamlessly connect diverse operational environments, significantly boosting the diagnostic models' ability to generalize across varied conditions. [4] developed a domain adaptation framework that utilizes adversarial training to align the feature distributions between different domains. This approach has been shown to significantly improve fault diagnosis accuracy by reducing the domain discrepancy through adversarial learning mechanisms. [5] introduced a multi-layer domain adaptation method that leverages deep learning architectures to extract transferable features across domains. This method enhances the model's robustness against varying operating conditions by ensuring that features relevant to fault diagnosis are consistently captured across different domains.

The concept of Generative Adversarial Networks (GANs) [6] has inspired methodologies like the one proposed by Ganin and Lempitsky [7], which utilizes a domain discriminator to blur the lines between source and target domains, enhancing domain adaptation through mechanisms such as gradient inversion layers. Furthermore, approaches like the Maximum Mean Difference (MMD) [8] have been utilized to quantify and minimize distribution differences. Concerning state evaluation, Yang et al. [9] employed a pre-trained ResNet18 (Residual Network) [10] as a feature extractor and calculated the MK-MMD (Multiple Kernel Maximum Mean Discrepancy) distance for each extracted feature from residual blocks. This approach effectively reduces the difference in condition distribution, thereby facilitating the evaluation of bearing operation status. Incorporating advanced discriminative feature learning with domain alignment, techniques developed by researchers such as An et al. [11] have shown promise. They integrate classification loss,

center-based discriminative loss, and correlation alignment loss during the training process across domains, which enhances the features' intra-class compactness and inter-domain separability, ultimately elevating the model's diagnostic accuracy.

The optimization and integration of these technologies have directly contributed to our research, and we have developed a deep adaptive network based on Convolutional Neural Networks (CNNs) [12]. This network utilizes Fast Fourier Transform (FFT), and Sinc-Net methods to extract vibration features signals from both time-domain and frequency-domain perspectives. The extracted features from the time-domain and frequency-domain are then fused. Additionally, convolutional neural networks are employed to extract time-frequency domain features that capture state evaluation-related features as well as common features between the source and target domains under varying operating conditions. We enhance the feature extraction capability of the convolutional neural networks by utilizing the attention mechanism of Convolutional Block Attention Module (CBAM) [13]. To achieve domain adaptation, we utilize the existing source domain labels to calculate pseudo labels for the target domain. The Maximum Mean Difference (MMD) and correlation comparison (CORAL) between the features of the two domains are calculated to match the conditional distribution distance. This approach effectively reduces the distribution differences of similar fault features under different working conditions and enhances the model's generalization ability for motor bearing state evaluation under varying operating conditions.

Despite these advancements, there remain challenges in achieving precise alignment of conditional distributions across different domains. We have innovatively developed a deep adaptive network that integrates both global and local alignment strategies to enhance the fault diagnosis capabilities in cross-condition scenarios. By addressing the issues of global and local feature alignment, our approach aims to significantly improve the generalization ability of fault diagnosis models across diverse operating conditions.

The remainder of this paper is organized as follows. Section 2 covers the preliminary knowledge on transfer learning. Section 3 presents the proposed model, including the network architecture and the loss functions used. Section 4 describes the datasets and experimental results, discussing the performance of the proposed method. Finally, Section 5 concludes the paper and outlines future research directions.

2. Preliminary Knowledge.

2.1. Problem description. Transfer Learning (TL) is an approach in the field of machine learning where the primary goal is to enhance performance on a different but related task by leveraging knowledge gained from one task. Traditional machine learning methods often assume that both the training data and the test data are drawn from the same distribution. However, in practice, this assumption does not always hold true.

In unsupervised domain adaptation, define a labeled source domain dataset $D_s = \{(x_{s_i}, y_{s_i})\}_{i=1}^{n_s}$, wherein, n_s represents the number of labeled samples in the source domain, x_{s_i} is the i th sample in the source domain, y_{s_i} represents the corresponding status evaluation type label, and $y_{s_i} \in R^C$, R^C represents label space, C indicates the number of fault categories; Define unlabeled target domain dataset $D_t = \{(x_{t_j})\}_{j=1}^{n_t}$, n_t indicates the number of unlabeled samples in the target domain, x_{t_j} represents the j th sample in the target domain. Assuming that the source domain data D_s and target domain data D_t have the same feature space and label space, but each follows different data probability distributions p , q , and $p \neq q$.

This article aims to design a deep domain adaptation model that integrates an improved measurement method combining MMD and CORAL to effectively reduce the distribution

difference between the source domain and target domain. The proposed method incorporates the distribution difference measurement between domains into the loss function of the neural network. By leveraging the superior feature extraction capability of deep models, it maps the sample features from the source domain and target domain to a high-dimensional feature space to align the data distributions of the two domains. Additionally, in the absence of labels in the target domain, the classifier trained on the source domain data is used to generate pseudo-labels for the target domain. These pseudo-labels are then combined with new common fault features to train the classifier in the target domain. This approach enhances the model's ability to recognize the operational status in the target domain.

2.2. Domain adaptation. In the evaluation of bearing states under variable conditions, domain adaptation is employed to map two datasets with different data distributions to a shared feature space. This process aligns the edge distribution of high-level features in this common feature space, effectively reducing the data distribution differences between the source and target domains. However, the full local adaptive method exclusively focuses on aligning the edge distribution of the two domains, overlooking the relationships between samples in these domains, which results in suboptimal migration performance.

In comparison with global domain adaptation, sample adaptation for the same category is more precise in aligning the distribution between identical categories. When the conditional distributions of the source and target domains are aligned, the edge distributions of the two domains also converge after adaptation. As depicted in Figure 1, after adaptation, identical samples in the two domains are aligned, while distinct samples are differentiated. At this stage, the overall data distribution of the source and target domains is harmonized. Consequently, mitigating the data distribution differences arising from changes in operating conditions can effectively enhance the accuracy of cross-operating-condition state evaluation.

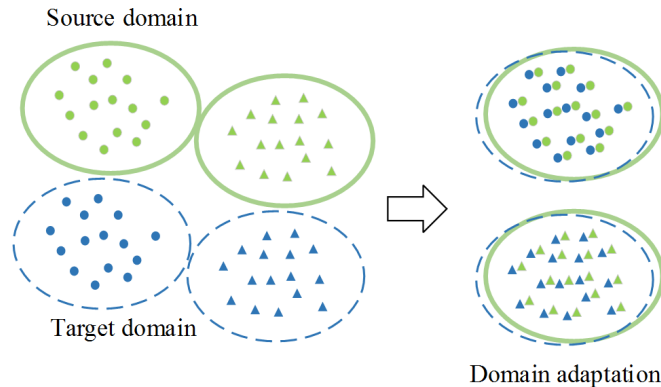


FIGURE 1. Domain adaptation process diagram

2.3. Maximum mean distance and correlation alignment. The Maximum Mean Discrepancy (MMD) is a statistic utilized to quantify the dissimilarity between two probability distributions. It involves mapping the samples from each distribution to a feature space and computing the mean of each distribution within that feature space. Subsequently, the MMD quantifies the disparity between the two distributions by evaluating the distance between these means. This distance is defined as follows:

$$\text{MMD}(X_s, X_t) = \left\| \frac{1}{n_s} \sum_{x_{s_i} \in D_s} \Phi(x_{s_i}) - \frac{1}{n_t} \sum_{x_{t_j} \in D_t} \Phi(x_{t_j}) \right\|_H^2 \quad (1)$$

In the formula, n_s and n_t indicate the number of samples from the source and target domains, H indicates Reproducing Kernel Hilbert Space (RKHS) [14]; $\Phi(*)$ represents the mapping function from the original feature space to RKHS. Just as MMD aligns the mean of the source and target domains in RKHS, CORAL aims to match second-order covariance statistics by calculating the covariance matrix of the source and target domain features. Then optimize the model by maximizing the correlation between matrices:

$$\text{CORAL}(X_s, X_t) = \frac{1}{4d^2} \|\text{Cov}_s - \text{Cov}_t\|_F^2 \quad (2)$$

In the formula, Cov_s and Cov_t represent the covariance matrices that characterize the features of the source domain sample and the target domain sample, respectively, and are defined as follows

$$\text{Cov}_s = \frac{1}{n_s - 1} \left(D_s^T D_s - \frac{1}{n_s} (\mathbf{1}^T D_s)^T (\mathbf{1}^T D_s) \right) \quad (3)$$

$$\text{Cov}_t = \frac{1}{n_t - 1} \left(D_t^T D_t - \frac{1}{n_t} (\mathbf{1}^T D_t)^T (\mathbf{1}^T D_t) \right) \quad (4)$$

In the formula, $\mathbf{1}$ represents a row vector with element 1.

Due to the lack of annotation in the target domain samples, target domain labels cannot be used. To reduce the impact of incorrect predictions, the network prediction model labels adopt a soft prediction method, and the classifier trained with source domain data is used to obtain target domain pseudo labels \hat{y}_{t_j} .

3. Deep Adaptive State Evaluation Model.

3.1. Deep adaptive network. The deep adaptive network primarily comprises two components: a feature extractor and a state evaluation classifier. The network framework is illustrated in Figure 2.

Firstly, for both source domain data $D_s = \{(x_{s_i}, y_{s_i})\}_{i=1}^{n_s}$ and target domain data $D_t = \{(x_{t_j}, \hat{y}_{t_j})\}_{j=1}^{n_t}$, we employed FFT [15] and SincNet [16] methods to process vibration data, transforming the original vibration signal into a frequency domain signal using FFT. This facilitates the capture of spectral information in the frequency domain of the vibration signal. The first layer of SincNet consists of a customized Sinc function [17]. This network can adaptively capture crucial time-domain information in the signal while avoiding disturbances from irrelevant frequency components, thereby preprocessing features in different states. By leveraging these two methods, we enhance the capability of vibration signal analysis, improve classification performance, and gain additional insights into signal characteristics.

Secondly, due to the advantages of CNN in feature learning, CNN was chosen as the basic structure for feature extraction in the source and target domains. The structural parameters of CNN are listed in Table 1. The feature extraction section includes five ‘‘Conv1D’’ blocks. Each ‘‘Conv1D’’ block consists of a convolutional layer, a Batch Normalization (BN) layer [18], and a max-pooling layer, except for the last ‘‘Conv1D-5’’ block, which includes an additional Global Average Pooling (GAP) layer [19]. The introduction of BN helps to accelerate the convergence of the network and alleviate overfitting problems. In addition, using weight sharing in the feature extraction network of the source and target domains to share partial weights of the model for better knowledge transfer can help adapt to the data from the feature extractor of the source domain to the target domain, thereby improving the generalization performance of the model.

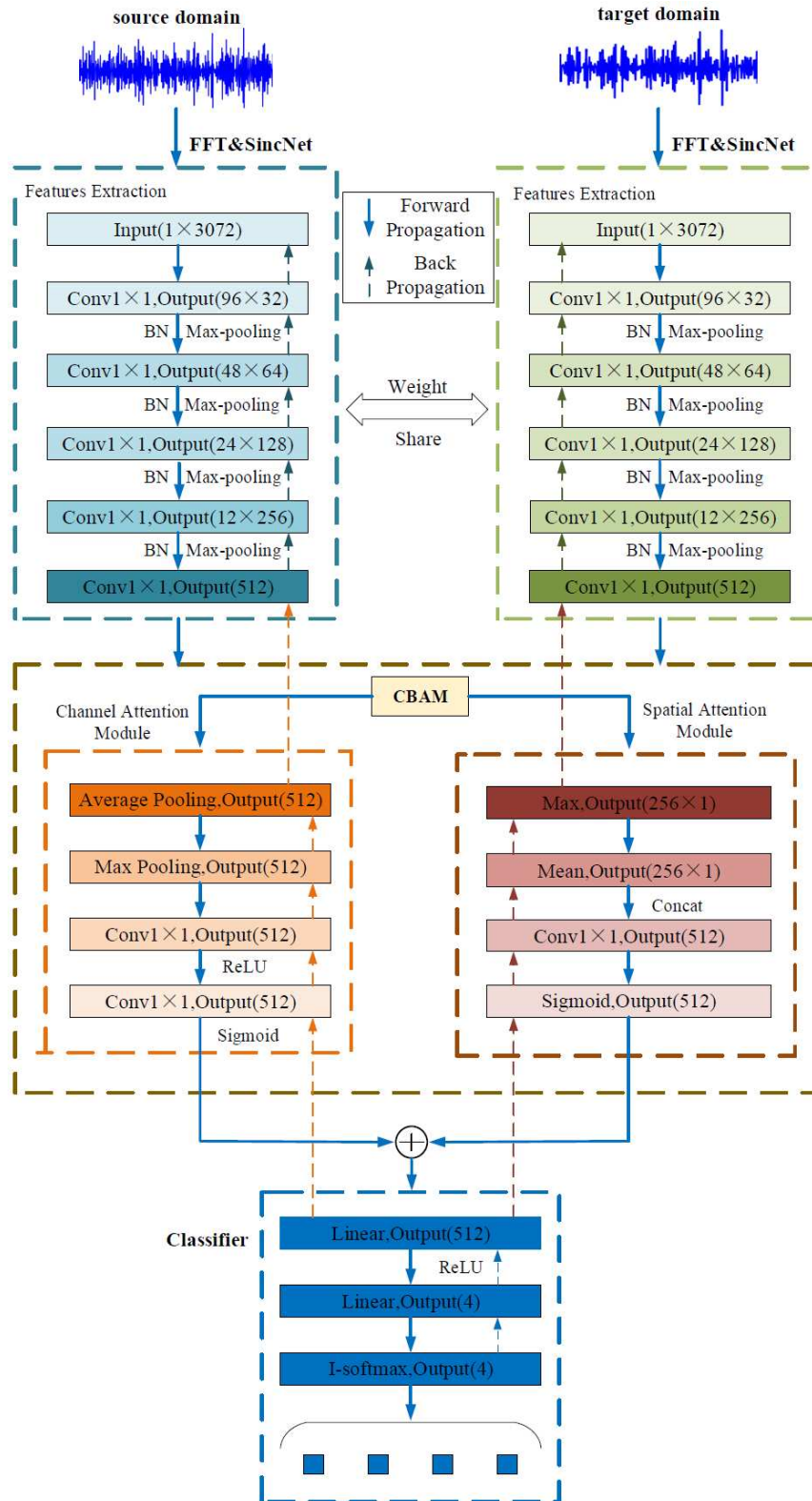


FIGURE 2. Deep adaptive state evaluation model framework

Considering the potential impact of noise in vibration signals on CNN feature extraction and enhancing CNN’s ability to extract features from data, we further adopt the CBAM attention mechanism as a part of the feature extractor component. CBAM considers both spatial attention and channel attention. Spatial attention helps the model focus

TABLE 1. CNN model parameters

Layer		Receptive field/ stride/channel	Input	Output
Conv1D-1	Convolution	64/16/32	(3072, 1)	(192, 32)
	BN	/	(192, 32)	(192, 32)
	Max-pooling	2/2/32	(192,32)	(96,32)
Conv1D-2	Convolution	3/1/64	(96, 32)	(96, 64)
	BN	/	(96, 64)	(96, 64)
	Max-pooling	2/2/64	(96, 64)	(48, 64)
Conv1D-3	Convolution	3/1/128	(48, 64)	(48, 128)
	BN	/	(48, 128)	(48, 128)
	Max-pooling	2/2/128	(48, 128)	(24, 128)
Conv1D-4	Convolution	3/1/256	(24, 128)	(24, 256)
	BN	/	(24, 256)	(24, 256)
	Max-pooling	2/2/256	(24, 256)	(12, 256)
Conv1D-5	Convolution	3/1/512	(12, 256)	(12, 512)
	BN	/	(12, 512)	(12, 512)
	Max-pooling	2/2/512	(12, 512)	(6, 512)
	GAP	/	(6, 512)	512

on the importance of different regions in the data, while channel attention helps the model focus on the importance of different feature channels. This comprehensive focus can better capture information in images. By dynamically adjusting the feature responses of different channels, it helps to extract more discriminative features related to faults, thereby improving the classification performance of the model.

For multi-classification tasks, the Softmax function is commonly used as the activation function in neural networks due to its probability interpretability and simplicity. However, the distribution of features in the source and target domains may differ, potentially leading to a decline in Softmax's performance in the target domain. [20] introduced the I-Softmax loss, which addresses these challenges by incorporating a regularization term that penalizes low-margin samples, thus encouraging the network to learn more discriminative and separable features. This loss function enhances the feature separability among different categories, particularly under domain adaptation scenarios. Hence, we have adopted the I-Softmax loss to facilitate the learning of more separable features and enhance accuracy in transfer tasks. This loss function proves effective in better distinguishing features among different categories, thereby augmenting the classification ability of the model.

3.2. Optimization objectives.

3.2.1. *Feature learning loss between samples.* To learn the state evaluation knowledge between deep features and fault types in the source and target domains, the cross-entropy loss of classification [21] is chosen as the loss of feature extraction in the CNN network to improve the learning of classification features. Assuming that the output from the data after passing through the classifier is $y_{s_i}^*$ and p_{s_i} , the evaluation classification loss sum of source domain and target domain data are respectively L_s and L_t :

$$L_s = -\frac{1}{n_s} \sum_{i=1}^{n_s} y_{s_i}^* \log(p_{s_i}) \quad (5)$$

where $y_{s_i}^*$ is the true label and p_{s_i} is the predicted probability for the source domain sample x_{s_i} , and n_s represents the total number of samples participating in loss calculation in the source domain.

$$L_t = -\frac{1}{n_t} \sum_{j=1}^{n_t} \dot{y}_{t_j} \log(p_{t_j}) \quad (6)$$

where \dot{y}_{t_j} is the pseudo label and p_{t_j} is the predicted probability for the target domain sample x_{t_j} , and n_t represents the total number of samples participating in loss calculation in the target domain.

By utilizing the superior feature extraction capability of deep neural networks, sample features from the source and target domains can be mapped to a common feature space, achieving alignment of data distribution between the two domains. In order to learn more discriminative state features, we introduce a new flexible loss function, namely the I-Softmax loss. This loss function can help us extract more state characteristic representations. I-Softmax loss is defined as:

$$L_{\text{I-Softmax}} = \begin{cases} -\frac{1}{n} \sum_{i=1}^n \log \left(\frac{e^{F^i(c)/h-g}}{e^{F^i(c)/h-g} + \sum_{j \neq c} e^{F^i(j)}} \right), & F^i(c) > 0 \\ -\frac{1}{n} \sum_{i=1}^n \log \left(\frac{e^{hF^i(c)-g}}{e^{hF^i(c)-g} + \sum_{j \neq c} e^{F^i(j)}} \right), & F^i(c) \leq 0 \end{cases} \quad (7)$$

F^i represents the feature vector output by the feature extractor, where $F^i(c)$ specifically refers to the element of the feature vector corresponding to the class label index c . Here, n represents the total number of feature vectors. The hyperparameters g and h are used to control the decision boundary, with g primarily adjusting the decision threshold and h modulating the discrimination margin between classes. According to relevant research, adjusting these parameters has been found to enhance the model's accuracy in state evaluation tasks [23].

3.2.2. Joint distribution adaptive loss. In order to generalize the model trained on the source domain well to the target domain, the joint distribution difference loss is applied to the network to reduce the distribution difference between the source domain and the target domain. The definition of joint distribution differential loss is

$$L_{\text{JDA}} = \left\| \mathbb{E}_{P(X_s)}[T(X_s)] - \mathbb{E}_{P(X_t)}[T(X_t)] \right\|^2 + \sum_{c=1}^C \left\| \mathbb{E}_{P(X_s|Y_s=c)}[T(X_s)]P(Y_s=c) - \mathbb{E}_{P(X_t|Y_t=c)}[T(X_t)]P(Y_t=c) \right\|^2 \quad (8)$$

In addition, due to the large amount of random noise in the data, the collected motor vibration signals approximately follow a Gaussian distribution. Therefore, in order to better implement the joint distribution adaptation mechanism and further improve the ability of domain confusion, the CORAL and MMD distribution difference measures are combined and named as a new measure of DDM(A, B):

$$\text{DDM}(A, B) = \text{MMD}(A, B) + \text{CORAL}(A, B) \quad (9)$$

Applying the DDM metric to the joint distribution adaptive loss, the final joint distribution adaptive loss is

$$L_{\text{JDA}} = \text{DDM}(T(X_s), T(X_t))$$

$$+ \sum_{c=1}^C \text{DDM}[(T(X_s)|Y_s = c)P(Y_s = c), (T(X_t)|Y_t = c)P(Y_t = c)] \quad (10)$$

Therefore, the final loss function of the network is

$$L = L_s + L_t + \gamma L_{\text{I-Softmax}} + \lambda L_{\text{JDA}} \quad (11)$$

where γ and λ , trade-off parameter of the loss function, are defined as

$$\gamma, \lambda = \frac{2}{1 + e^{-10t/T}} - 1 \quad (12)$$

In the formula, t represents the current number of iterations of the training, and T represents the total number of iterations.

Specifically, in Equation (11), $L_{\text{I-Softmax}}$ is the alignment loss between the source and target domains, used to reduce the difference in features between the two domains, weighted by γ ; L_{JDA} is the distribution loss between the source and target domains, used to maintain the performance in the source domain while better adapting to the target domain, weighted by λ .

4. Cross Domain State Evaluation Results and Analysis.

4.1. Introduction to experiment. To validate the proposed model's effectiveness, cross-condition state evaluation experiments were conducted on two datasets: the Case Western Reserve University bearing datasets (CWRU) [22] and the Jiangnan University bearing datasets (JNU) [23]. Both datasets were selected at different rotational speeds. The CWRU dataset includes three different rotational speeds with a sampling frequency of 12 kHz (1772, 1750, and 1730 r/min). The motor in this dataset exhibits three main state types: co-rolling body damage, outer ring damage, and inner ring damage. The JNU bearing dataset from Jiangnan University is a bearing state dataset collected by JNU, with a sampling frequency of 50 kHz. Vibration signals were collected using accelerometers at three different rotational speeds (600, 800, and 1000 r/min) during the experimental process, featuring a total of three state types.

Different rotational speeds are considered distinct domains for the state assessment task. Taking the JNU dataset as an example, data with rotational speeds of 600, 800, and 1000 r/min are designated as tasks A, B, and C, respectively. For instance, A \rightarrow B indicates that the source domain is vibration data with a rotational speed of 600 r/min, and the target domain is vibration data with a rotational speed of 800 r/min.

To assess the cross-domain evaluation effectiveness of the model, this paper selects widely used domain adaptive methods, including Deep Correlation Alignment (DCORAL) [24], Domain-Adversarial Neural Network (DANN) [25], Deep Domain Confusion (DDC) [8], Deep Discriminative Transfer Learning Network (DDTLN) [20], for comparative testing.

In the experiments, considering the differences between pseudo-labels and real labels on the one hand and reducing the impact of the I-Softmax loss function on the total loss function during training on the other hand, the parameters are set as follows: the model input data length is 3072, the training batch size is 256, and a total of 300 epochs are trained. The model parameters are optimized using the Adam optimization algorithm [26], and the initial value of the learning rate is set as 0.001.

The dataset utilized for the experiment comprises 1000 samples for each category in both the source and target domains, resulting in a total of 4000 samples for each domain. The training dataset includes samples from both the source and target domains, while the test dataset exclusively consists of target domain samples. Given the characteristic of having fewer actual fault state samples, a sliding sampling technique is employed to sample

the raw data and augment the fault state samples. Specifically, each sample comprises 3072 data points to ensure sufficient fault state information.

4.2. CWRU dataset experimental results and analysis.

4.2.1. *Description of the datasets.* The composition of the Case Western Reserve University bearing fault state simulation experiment platform is illustrated in Figure 3, which includes bearings, motors, torque sensors, encoders, power testers, and other electronic control equipment. The CWRU dataset encompasses four major categories: normal data, drive-side fault state data sampled at 12 kHz, fan-side fault state data, and combined fault state data. Acceleration sensors were employed to collect vibration signals from SKF6205-2RS deep groove ball bearings under diverse operating conditions, with the fault positions located in the inner ring, rolling element, and outer ring.

Single point failures were induced on the bearings using Electrical Discharge Machining (EDM), and the fault types included rolling element damage, outer ring damage, and inner ring damage. Throughout the bearing testing process, raw vibration signals were collected for various fault types, including Normal Condition (NC), Inner ring Fault (IF), Ball Fault (BF), and Outer ring Fault (OF).

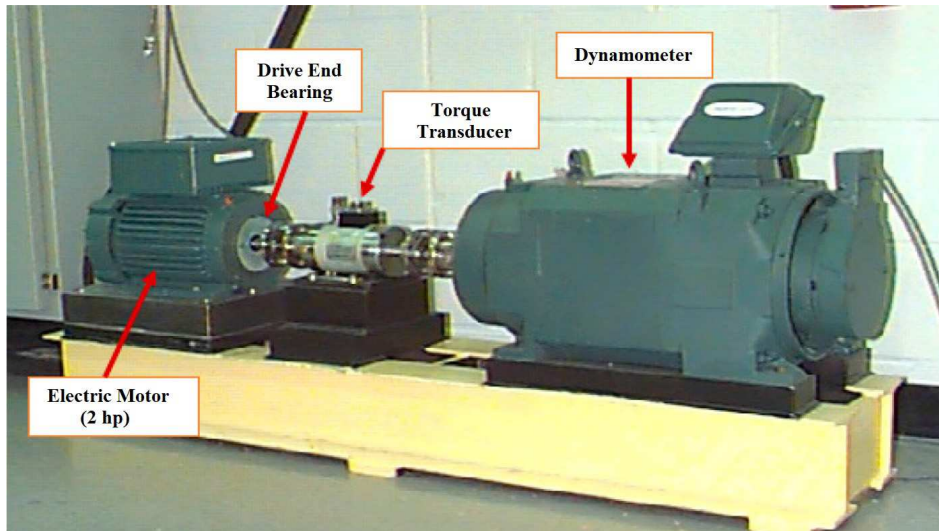


FIGURE 3. CWRU bearing failure testbed

TABLE 2. CWRU bearing data status classification

Dataset name	Rotation speed (r/min)	Status classification
A	1772	Normal Condition (NC), Inner ring Fault (IF), Ball Fault (BF), Outer ring Fault (OF)
B	1750	Normal Condition (NC), Inner ring Fault (IF), Ball Fault (BF), Outer ring Fault (OF)
C	1730	Normal Condition (NC), Inner ring Fault (IF), Ball Fault (BF), Outer ring Fault (OF)

4.2.2. *CWRU dataset cross-domain state assessment results and analysis.* Six cross-condition state assessment tasks were designed using the Case Western Reserve University dataset, and the correctness comparisons of state assessment for different methods are presented in Table 3.

TABLE 3. Experimental results (CWRU)

Method	Transfer of tasks accuracy (%)						Mean accuracy (%)
	A \rightarrow B	A \rightarrow C	B \rightarrow A	B \rightarrow C	C \rightarrow A	C \rightarrow B	
DCORAL	80.29	39.19	91.19	45.39	43.50	48.52	58.01
DANN	65.23	35.69	52.61	39.54	40.68	51.31	47.51
DDC	78.54	51.57	82.05	55.13	52.00	38.87	59.69
DDTLN	98.08	82.36	95.12	91.55	87.09	88.92	90.52
Proposed model	94.42	89.75	96.65	93.30	91.45	92.25	92.97

Based on the table, the evaluation results of the proposed method outperform other algorithms in five out of six transfer tasks. The proposed model demonstrates excellent performance across all types of task transfers, with an average accuracy of 92.97%, ranking first among the five methods. This indicates that the proposed model can maintain high accuracy in different task transfer scenarios, showcasing strong generalization ability.

From the experimental results, DDC assumes that the source and target domains have similar data distributions and attempts to make their feature representations as similar as possible. However, this approach results in degraded DDC performance due to variations in data distributions and feature representations under different working conditions. DANN, employing an adversarial approach, extracts domain-invariant features but solely focuses on global adaptation, neglecting cross-domain conditional distribution differences. This oversight leads to subdomain confusion and larger classification errors. The core concept of DCORAL is to align marginal distributions in domain adaptation by minimizing feature covariance differences between source and target domains. However, this concentration on marginal distribution alone proves insufficient for the domain adaptation task, resulting in the algorithm's performance degradation. While DDTLN achieves the highest accuracy (98.08%) in certain tasks, such as A \rightarrow B, it performs relatively poorly in other tasks. The superior performance of DDTLN in the A \rightarrow B task can be attributed to its robust feature extraction capabilities in scenarios with specific domain characteristics. However, its limitations in feature selection and generalization to varying domain conditions restrict its overall effectiveness across different tasks.

To further illustrate the superiority of the proposed method, t-distributed Stochastic Neighbor Embedding (t-SNE) [27] is employed to map the learned high-dimensional features to a two-dimensional space. In the case of task B \rightarrow A, the t-SNE maps obtained from the five models are presented in Figure 4. The proposed model demonstrates the minimum intra-class distance and the maximum inter-class distance. This is primarily attributed to its enhanced ability to select fault-related features and align the edge and conditional distributions of the target and source domains compared to existing Domain Adaptation (DA) models. These comparison results further affirm that the proposed model exhibits higher accuracy than typical DA methods.

4.3. JNU dataset experimental results and analysis.

4.3.1. *Description of the datasets.* The bearing fault state simulation platform of Jiangnan University is shown in Figure 5. The test rig is a fault diagnosis test for a centrifugal

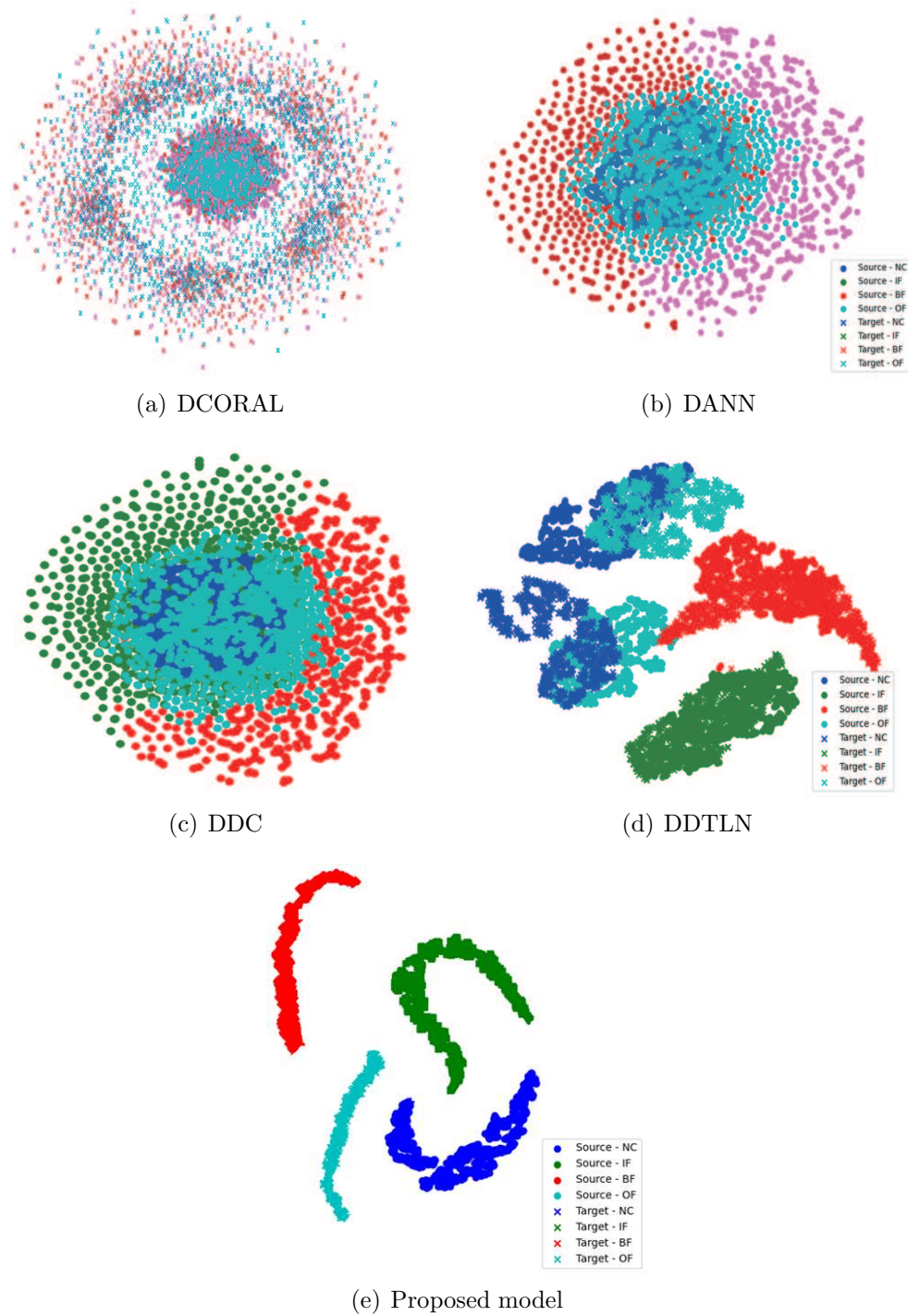


FIGURE 4. (color online) Five DA models on the CWRU dataset B \rightarrow A task t-SNE

fan system using a Mitsubishi SB-JR induction motor. The motor is a 3.7 KW three-phase induction motor with 220 V voltage and the number of poles of the motor is 4. The rated speed is 1800 rpm, and the sampling frequency is 50 kHz, the rotor is supported by two bearings, and there is a defective bearing in the whole experimental setup, which contains four states, including the normal state, inner-ring failure, outer-ring failure, and rolling-body failure. Accelerometers were used to collect vibration signals at three different speeds, 600, 800 and 1000 r/min.

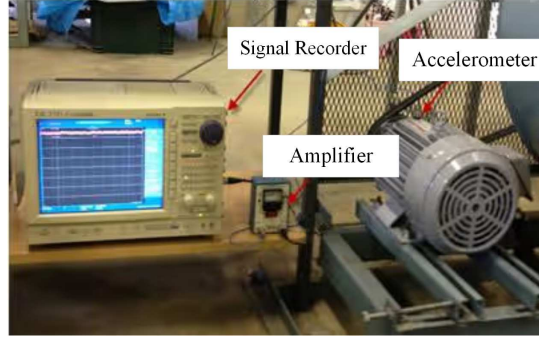


FIGURE 5. JNU bearing failure testbed

TABLE 4. JNU bearing data status classification

Dataset name	Rotation speed (r/min)	Status classification
A	600	NC, IF, BF, OF
B	800	NC, IF, BF, OF
C	1000	NC, IF, BF, OF

4.3.2. *JNU dataset cross-domain state assessment results and analysis.* Six cross-condition state assessment tasks were designed using the JNU dataset, and the comparison of state assessment correctness rates for different methods is presented in Table 5.

TABLE 5. Experimental results (JNU)

Method	Transfer of tasks accuracy (%)						Mean accuracy (%)
	A \rightarrow B	A \rightarrow C	B \rightarrow A	B \rightarrow C	C \rightarrow A	C \rightarrow B	
DCORAL	79.18	65.82	75.82	70.38	74.70	76.80	73.78
DANN	90.30	89.48	87.82	90.62	91.15	80.87	88.37
DDC	71.83	79.90	67.47	80.07	60.42	72.82	72.09
DDTLN	91.75	87.37	84.32	77.87	82.50	92.95	86.13
Proposed model	93.17	90.30	88.05	91.52	90.32	93.05	91.07

The t-SNE of the bearing experiment at JNU is shown in Figure 6.

From the above results, it can be seen that, except for the diagnostic results in the migration of work condition C \rightarrow A, which are slightly lower than DANN, the evaluation results of each migration task are better than those of other algorithms, and the average diagnostic correctness rate is 91.07%, which indicates that the method has a good generalization ability.

5. Conclusions. In this paper, we introduce a transfer learning network designed for cross-condition motor state assessment in nonsmooth dynamical systems. The network utilizes original vibration signals as inputs without any signal preprocessing or manual feature extraction. Furthermore, a CNN is constructed to capture more comprehensive state features, integrated with the CBAM to filter features and accelerate model training.

To enhance domain confusion, a distribution difference metric, comprising MMD and CORAL, is employed. This strategy improves the distribution matching between the source and target domains and aligns the conditional distributions of both domains. As a result, features under the same label from both domains are aggregated, and clear

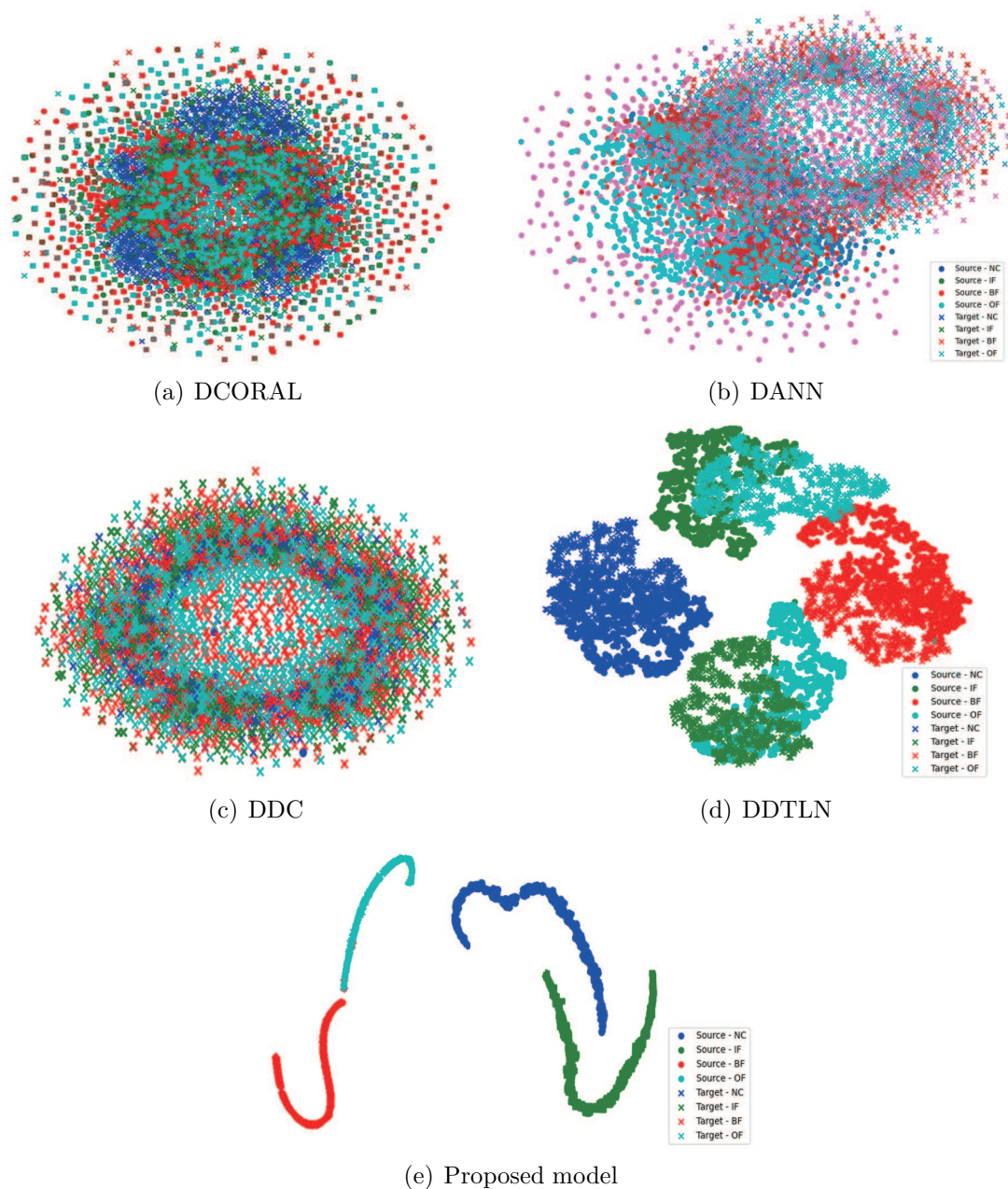


FIGURE 6. (color online) Five DA models on the JNU dataset $B \rightarrow A$ task t-SNE

distinctions emerge between features with different labels. Compared to algorithms that only align global distributions, our proposed model effectively adapts to the source and target domains of the migration task post-training, demonstrating favorable evaluation results and strong generalization capabilities, even in scenarios where training and test data distributions are skewed, and there is a lack of high-quality labeled data.

Future research could explore several directions to further enhance the model's effectiveness and applicability: investigating the use of more sophisticated or hybrid attention mechanisms to improve feature extraction and domain adaptation; extending the model to handle multiple source domains simultaneously, which could increase its robustness and adaptability under diverse operational conditions; and adapting and applying the proposed model to other types of machinery and equipment, validating its versatility and

effectiveness across different industrial settings. By pursuing these future research directions, the proposed model can be further refined and its applicability expanded, ultimately contributing to more robust and efficient state assessment systems in various industrial applications.

Acknowledgment. This work is partially supported by the National Natural Science Foundation of China under Grant 62020106003 and Grant 62263010, in part by the open fund of the National Key Laboratory of Helicopter Aeromechanics under Grant 2024-ZSJ-LB-02-04, the Natural Science Research Project of Anhui Province Universities under Grant 2023AH040121, and the High Level Talent Research Start-up Fund of Anhui Polytechnic University under Grant 2022YQQ052.

REFERENCES

- [1] H. Bian, X. Wang and Z. Deng, Brushless DC motor bearing fault diagnosis based on stator current, *Journal of Nanjing University of Aeronautics and Astronautics*, vol.52, no.2, pp.224-231, 2020.
- [2] M. R. Bhuiyan and J. Uddin, Deep transfer learning models for industrial fault diagnosis using vibration and acoustic sensors data: A review, *Vibration*, vol.6, no.1, pp.218-238, 2023.
- [3] Y. Yoo, H. Jo and S. W. Ban, Lite and efficient deep learning model for bearing fault diagnosis using the CWRU dataset, *Sensors*, vol.23, no.6, 3157, 2023.
- [4] J. Wang, R. Liu and J. Cao, Fault diagnosis method for unlabeled data based on multilayer domain adaptation, *Journal of Beijing University of Aeronautics and Astronautics*, vol.7, no.10, pp.1-13, 2023.
- [5] X. Song, W. Sun, G. Liu et al., Deep subdomain adaptive network electric machine rolling bearing fault diagnosis across operating conditions, *Journal of Electrotechnology*, pp.1-12, 2023.
- [6] J. Sun, Y. Wang, Y. Sun and F. Jin, Fault diagnosis method of migration learning based on antagonism generation network, *International Journal of Innovative Computing, Information and Control*, vol.19, no.6, pp.1953-1968, 2023.
- [7] Y. Ganin and V. Lempitsky, Unsupervised domain adaptation by backpropagation, *International Conference on Machine Learning*, pp.1180-1189, 2015.
- [8] E. Tzeng, J. Hoffman, N. Zhang et al., Deep domain confusion: Maximizing for domain invariance, *arXiv Preprint*, arXiv: 1412.3474, 2014.
- [9] B. Yang, Q. Li, L. Chen et al., Research on bearing fault diagnosis based on ResNet and domain adaptive, *Measurement and Control Technology*, vol.40, no.12, pp.31-39, 2021.
- [10] K. He, X. Zhang, S. Ren et al., Deep residual learning for image recognition, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.770-778, 2016.
- [11] J. An, P. Ai and D. Liu, Deep domain adaptation model for bearing fault diagnosis with domain alignment and discriminative feature learning, *Shock and Vibration*, vol.2020, 2020.
- [12] A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems*, vol.25, 2012.
- [13] S. Woo, J. Park, J. Y. Lee et al., CBAM: Convolutional block attention module, *Proc. of the European Conference on Computer Vision (ECCV)*, pp.3-19, 2018.
- [14] P. Ge, C. X. Ren, X. L. Xu et al., Unsupervised domain adaptation via deep conditional adaptation network, *Pattern Recognition*, vol.134, 109088, 2023.
- [15] B. Yang, Q. Li, L. Chen et al., Bearing fault diagnosis based on multilayer domain adaptation, *Shock and Vibration*, vol.2020, 2020.
- [16] M. Ravanelli and Y. Bengio, Speaker recognition from raw waveform with SincNet, *2018 IEEE Spoken Language Technology Workshop (SLT)*, pp.1021-1028, 2018.
- [17] D. Nguyen, *Adaptive Windowed Sinc Filter for Image Interpolation*, Master Thesis, San José State University, 2023.
- [18] X. Han, Z. Hu, M. Huang and L. Wang, Keras-based 3D-convolutional neural network bearing fault diagnosis, *ICIC Express Letters*, vol.16, no.4, pp.371-379, 2022.
- [19] F. Xie, L. Wang, H. Zhu et al., Research on a rolling bearing fault diagnosis method based on multi-source deep sub-domain adaptation, *Applied Sciences*, vol.13, no.11, 6800, 2023.
- [20] Q. Qian, Y. Qin, J. Luo et al., Deep discriminative transfer learning network for cross-machine fault diagnosis, *Mechanical Systems and Signal Processing*, vol.186, 109884, 2023.

- [21] J. Singh, M. Azamfar, A. Ainapure et al., Deep learning-based cross-domain adaptation for gearbox fault diagnosis under variable speed conditions, *Measurement Science and Technology*, vol.31, no.5, 055601, 2020.
- [22] W. A. Smith and R. B. Randall, Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study, *Mechanical Systems and Signal Processing*, vol.64, pp.100-131, 2015.
- [23] C. Li, L. Mo and R. Yan, Fault diagnosis of rolling bearing based on WHVG and GCN, *IEEE Transactions on Instrumentation and Measurement*, vol.70, pp.1-11, 2021.
- [24] B. Sun and K. Saenko, Deep CORAL: Correlation alignment for deep domain adaptation, in *Computer Vision – ECCV 2016 Workshops. ECCV 2016. Lecture Notes in Computer Science*, G. Hua and H. Jégou (eds.), Cham, Springer International Publishing, 2016.
- [25] Y. Ganin, E. Ustinova, H. Ajakan et al., Domain-adversarial training of neural networks, in *Domain Adaptation in Computer Vision Applications. Advances in Computer Vision and Pattern Recognition*, G. Csurka (ed.), Cham, Springer, 2017.
- [26] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *arXiv Preprint*, arXiv: 1412.6980, 2014.
- [27] S. Liu, H. Jiang, Z. Wu et al., Intelligent fault diagnosis of rotating machinery using a multi-source domain adaptation network with adversarial discrepancy matching, *Reliability Engineering & System Safety*, vol.231, 109036, 2023.

Author Biography



Yu Liao received his Bachelor's degree in Electrical Engineering and Automation from Hubei Minzu University in 2001, his Master's degree in Electrical Engineering from Huazhong University of Science and Technology in 2012, and his Ph.D. degree in Optical Engineering from Sichuan University in 2018. He is currently an associate professor at Hubei Minzu University. His current research interests include data-driven fault detection and diagnosis for complex mechatronic systems, multi-agent cooperative control, and intelligent system optimization.



Jiahao Geng obtained his Bachelor's degree in Electrical Engineering and Automation from Xingjian College of Science and Liberal Arts, Guangxi University, in 2019. He is currently a graduate student at the School of Intelligent Science and Engineering, Hubei Minzu University. His research interests mainly include fault diagnosis and deep learning.



Li Guo (Member, IEEE) received her Ph.D. degree from Sichuan University, Chengdu, China, in 2013. From 2016 to 2017, she worked as a visiting scholar in University of Groningen, the Netherlands. She is a professor with the School of Electrical Engineering at Anhui Polytechnic University. Her main research interest includes image processing and pattern recognition, data-driven fault detection and diagnosis.



Bing Geng obtained his Bachelor's degree in Electrical Engineering from Quancheng College of Jinan University in 2020. He is currently a graduate student at the School of Intelligent Science and Engineering, Hubei Minzu University. His research interests primarily include AI-based industrial fault detection.



Kun Cui obtained his Bachelor's degree in Measurement and Control Technology and Instrumentation from Qufu Normal University in 2021. He is currently a graduate student at the School of Intelligent Science and Engineering, Hubei Minzu University. His research interests mainly include the applications of artificial intelligence in fracture prediction.



Runze Li received his Bachelor's degree in Electrical Engineering and Automation from Renai College of Tianjin University in 2018 and his Master's degree in Electrical Engineering from Hubei Minzu University in 2021. He is currently pursuing a Ph.D. degree at Nanjing University of Aeronautics and Astronautics. His current research interests include data-driven fault detection and diagnosis, virtual reality simulation, and autonomous intelligent control.