A Coarse-to-Fine Bi-Level Adversarial Domain Adaptation Method for Fault Diagnosis of Rolling Bearings

Zhao-Hua Liu, *Senior Member, IEEE*, Liang Chen, Hua-Liang Wei, Ying Zhang, Lei Chen, and Ming-Yang Lv

***Abstract*—Recently, domain adaptation has been used to solve the fault diagnosis problem in rolling bearings. However, most of the existing methods only align the distribution of domains, and ignore the fine-grained information of the same fault categories in different domains, which leads to the degradation of diagnostic performance. To address such domain difference issues, this paper proposes a novel coarse-to-fine bi-level adversarial domain adaptation approach (C2FADA) for bearings fault diagnosis. Firstly, a sparse auto-encoder (SAE) is used to extract features from raw data (containing both the source and target domains), and a Kullback-Leibler (KL) divergence term is then introduced to measure the discrepancy between the features from the source domain and the target domain. Secondly, a bi-level adversarial module is established to gradually align different domains at the domain level (with a coarse-grained model) and the class level (with a fine-grained approach) to tackle the domain shift issue, and enable the classifier to learn the domain invariant representation features. Thirdly, a spectral norm regularization constraint term is introduced to improve the stability of adversarial training process by mitigating the effect of adversarial perturbations. results show that the classification performance of the proposed C2FADA method is better than the compared existing peer methods.**

***Index Terms*—Fault diagnosis, rolling bearings, domain adaptation , bi-level adversarial learning, sparse representation, machine learning, deep learning,transfer learning.**

I. Introduction[1]

In recent years, the rapid development of the internet of things, big data, sensor technology and industrial wireless networks have driven the progress of intelligent manufacturing and Industry 4.0 [1]. In order to meet the needs of modern industrial production, machinery equipment has been developed rapidly towards the direction of large-scale, complex, intelligent and efficient. Rolling bearings, as an important support component for rotating machinery, are one of the key parts that are vulnerable to wear after long-term work under heavy loads. Therefore, timely and effective detection of rolling bearing failures is extremely important [2].

With the development of artificial intelligence (AI) techniques especially machine learning, more and more advanced smart diagnosis approach has been applied to the domain of rotating machinery fault diagnosis in current years [3]. The vibration signals of rolling bearings have the characteristic of cyclostationarity due to their periodic operation mode [4], [5]. In addition, when the rolling bearings break down, the vibration signals may have obvious impulsive characteristic [6]. Mining the information of these vibration signals can improve the fault diagnosis level of rolling bearings. Intelligent fault diagnosis methods for rolling bearings, especially data-driven methods, have been developed more widely and rapidly than ever before in industry [7]. The commonly used methods include support vector machines (SVMs) [8], deep belief networks (DBNs) [9], back propagation (BP) neural networks [10], sparse autoencoder (SAE) [11] and so on. Such approaches build discriminative classifiers through learning from historical data. These methods usually work well for some specific classification and diagnosis tasks. However, the majority of the current popular data-driven fault diagnosis approach has a common hypothesis, which is the labeled training data and the unlabeled test data are have the same distribution [12]. In fact, in real industry, the working conditions of rotating machinery are usually variable with the production demand, and thus the operation of rolling bearings is carried out under different conditions due to the changes in speed and load, which directly influences the vibration characteristics of bearings. In the operating environment with multiple alternative working conditions, the vibration data of bearings cover a large amount of health and fault information collected in different working conditions and complex environmental factors. The fault category mapped by the vibration data is unknown and changeable. In addition, complex working conditions often mean that the test data are collected in a working state different from the one of training data. Therefore, the training and test data no longer completely obey the same distribution, making it difficult for the model trained by train data to perform well on the test data [13].

Domain adaptation (DA), as a new effective machine

[1]Manuscript received April 22, 2022; revised June 28, 2022, September 11, 2022, and accepted October 02, 2022. This work was supported in part by the National Key Research and Development Project of China under Grant 2019YFE0105300, in part by the National Natural Science Foundation of China under Grant 61972443 and Grant 62103143, in part by the Hunan Provincial Key Research and Development Project of China under Grant 2022WK2006,and in part by the Hunan Provincial Hu-Xiang Young Talents Project of China under Grant 2018RS3095.

Z.-H Liu, L. Chen, L. Chen and M.-Y Lv are with the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, China (e-mail:zhaohualiu2009@hotmail.com; 13069302167@163.com;chenlei@hnust.edu.cn; 1040133@hnust.edu.cn).

H. -L. Wei is with the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S1 3JD, U.K. (e-mail: w.hualiang@sheffield.ac.uk).

Y. Zhang is with the School of Computer Science, Northwestern Polytechnical University,Xi'an,710129, China.(email:ying_zhang@nwpu.edu.cn).

learning strategy closely pertaining to transfer learning, can be used to obtain the most useful information from the training data (source domain) and transfer it to other data of interest (target domain). Zhao *et al*. [14] proposed a joint distribution adaptation network for fault diagnosis of rolling bearings. This method precisely matched the distribution matching, and extracted the domain-invariant features of the source and target domains by adversarial learning. In [15], a new semi-supervised fault diagnosis method was presented by mixing convolutional neural network (CNN), correlation alignment loss, and transfer component analysis to classify the fault characteristics. Xu *et al*. [16] designed a metric transfer learning framework (MTLF), which can utilize the internal information between instances in different domains more effectively by learning instance weights and distance metrics at the same time.

Although existing fault diagnosis methods such as DA have achieved fine test effectiveness, the primary disadvantage consists in the assumption that the test data corresponding to all different equipment health conditions can be obtained and used for training. When such a requirement is not satisfied, the existing methods may produce invalid diagnosis. Recently, more and more adversarial adaptation methods have been introduced to minimize the difference distance between domains through an adversarial objective of the domain discriminator [17], [18]. These methods are highly related to the well-known generative adversarial networks (GANs) [19], in which two models are concurrently trained: a generative model and a discriminative model; the models are used to capture the data distribution and calculate the probability that the sample came from the different data domain, respectively. The generator is trained to produce data in a way that confuses the discriminator, which in turn tries to distinguish them from real data. Chen *et al.* [20] proposed a domain adversarial transfer network (DATN) using deep CNNs with an asymmetric encoder model, together with domain adversarial training technology, to successfully solve the problem of fault diagnosis in the presence of a large domain shift in data distribution. Li *et al.* [21] proposed an improved domain adaptation intelligent fault diagnosis method, where the maximum mean difference and domain adversarial training are used to train the two feature extractors of feature space distance and domain adaptation, respectively. In this way, the feature representation capability is enhanced. She *et al.* [22] proposed a deep conditional adversarial diagnosis method based on weighted entropy minimization. This method applies the transferability weight of the sample to the entropy minimization loss, and solves the problem of model collapse and difficulty of sample transfer in the adversarial domain adaptation training. In [23], Fu *et al.* designed a feature enhanced GAN to extract the refined features by embedding the self-attention mechanism in the residual network, and constructed an auxiliary classifier to classified generated and unlabeled samples. In [24], a deep transfer learning model was proposed to detect the rolling bearing faults. In this model, the Wasserstein GANs were introduced to calculate the discrepancy between the domains, and the minimum singular was applied to capture the effective fault information. However, the training process of adversarial adaptation methods is often unstable and easy to converge in advance, which may affect the fault diagnosis effect of the models.

Due to its ingenious structure and excellent performance, the adversarial domain adaptation method is gradually becoming one of the mainstream approaches. Undoubtedly, the rapid development of generative adversarial models provides a powerful tool for addressing the fault diagnosis problem with imbalance data, but the combination of the traditional fault diagnosis methods and adversarial domain adaptation only considers the distribution difference between the source and target domains, the resulting performance may not be good enough. This is because fault diagnosis tasks are far more than just aligning the distribution of source and target domains.

Recently, a domain adaptation network for cross-domain fine-grained recognition has been widely used in pattern recognition and object detection [25]-[27]. The main idea is to explore the commonalities between the fine-grained information of existing image datasets and a large amount of unlabeled data. On this basis, the concept of fine-grained information has been introduced into the field of fault diagnosis. It is known that coarse-grained information has the following disadvantage: the coarse-grained process of some nonlinear measurement methods (e.g. multi-scale permutation entropy) only consider low-frequency information. To overcome such a drawback, in [28] a fine-to-coarse multi-scale permutation entropy measurement method is proposed, which can provide low-frequency and high-frequency information to improve bearing fault diagnosis performance. In [29], a coarse-to-fine weak fault detection method for rotating machinery was proposed, and a variational modal decomposition-based coarse-to-fine decomposition strategy was designed to obtain the optimal mode of rotating machinery and extract its weak repetitive transients.

In many real fault diagnosis tasks, it not only needs to align the data distribution of the source and target domains to lessen the difference between domains, but also consider the specific category information of the target sample to provide more accurate diagnosis results. It is essential to establish an effective method to achieve more comprehensive DA.

In this article, a framework of bi-level adversarial domain adaptation based on coarse-to-fine features (called C2FADA) is designed for bearings fault diagnosis under variable operating conditions. In this framework, the relation information between the source and target domains is used to ensure the success of domain adaptation and fault classification. It uses coarse-grained domain discriminator and fine-grained domain discriminator to perform domain-oriented and class-oriented alignment, respectively. Specifically, the C2FADA model structure consists of four modules: a coarse-grained domain discriminator, a fine-grained domain discriminator, a feature generator, and a standard machine learning classifier. The feature generator can generate domain-invariant feature representations that retain the distinguishing structure. Each of the two discriminators obtains features from the feature generator and performs mini-max games with the generator

respectively. The coarse-grained domain discriminator learns the features of source and target domains and maps these features into the domain label space (called coarse-grained feature representation), and the fine-grained domain discriminator approaches the predicted labels of the classifier to learn the domain invariant features of each fault category, so as to transform the features of the source domain into the fault label space (called fine-grained feature representation). Afterwards, the domain-invariant features learned by the model can effectively transfer the knowledge to the target domain from the source domain, so as to obtain more accurate classification performance. The proposed C2FADA model considers the problem of domain alignment and class alignment in practical industry, and can well solve the fault diagnosis of rolling bearings. the main contributions are as follows:

1) To effectively solve the domain shift problem in fault diagnosis under variable operating conditions, a new C2FADA framework is proposed. In this framework, a bi-level adversarial module is established to gradually align different domains at domain level and class level, enabling the classifier to capture the domain invariant representation features.
2) The proposed coarse-to-fine bi-level adversarial network can better distinguish the fault structure, maximize the intra-class distance, and minimize the inter-class distance.
3) To improve the stability of the GAN training process, a spectral norm regularization (SNR) scheme is introduced. Regular constraints are introduced based on the spectral norm of the neural network parameter matrix to make the training process more stable and quick to converge.

The remainder of this paper is organized as follows. Section II provides the preliminaries. The proposed C2FADA method is presented in Section III. In Section IV, plenty of comparative experiments are validated. Finally, conclusion is in Section V.

II. Preliminaries

A. Problem Formulation

The vibration data of rolling bearings under variable working conditions generally have large differences, which make bearing fault diagnosis an extremely challenging problem. In order to overcome this challenge, a novel C2FADA method is proposed in this paper. The method is based on feature transfer, and the feature mapping process of the DA method is presented in Fig. 1. According to the principle of feature transfer, it is assumed that the following assumptions are satisfied:

1) The fault diagnosis tasks in different domains are the same; all faults of rolling bearings can be categorized into a limited number of classes.
2)Owing to variable working conditions, the data distribution of the source and target domains may be different, but there are similarities between data in the two domains.
3)There are lots of labeled samples that can be used to structure the fault types in the source domain, while only unlabeled data are available in the target domain.

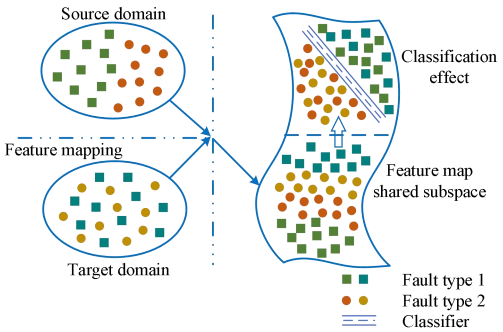


Fig. 1. An illustration of feature-based DA method.

A domain, denoted by $\mathcal{D}=\{\mathcal{X}, P(X)\}$, consists of two main components: a feature space of inputs $\mathcal{X}$, and a marginal probability distribution of inputs $P(X)$, where the sample set $X=\{x_1,x_2,\ldots x_n\}\in\mathcal{X}$. If the source domain $\mathcal{D}_s$ and target domain $\mathcal{D}_T$ are different, they have different data spaces and marginal distributions.

A task $\mathcal{T}$, denoted by $\mathcal{T}=\{\mathcal{Y}, f(x)\}$, contains two parts: a label space $\mathcal{Y}$ and a target prediction function $f(x)$. The function is unknown but can be learned from data in the source domain. From a probabilistic viewpoint, $f(x)$ can be considered as a conditional probability distribution $P(y|x)$, where $y\in\mathcal{Y}$. In the fault diagnosis classification of this article, $\mathcal{Y}$ is the set of all labels. For example, the values of y are either true or false for a binary classification task.

In this paper, the situation where there is one $\mathcal{D}_s$ and one $\mathcal{D}_T$ is considered. More specifically, let the labeled $\mathcal{D}_s$ data be $\mathcal{D}_s=\{(x_i,y_i)\}_{i=1}^{n_s}$, where $x_i\in\mathcal{X}_s$ is the input and $y_i\in\mathcal{Y}_s$ is the corresponding output. Analogously, let the unlabeled $\mathcal{D}_T$ data be $\mathcal{D}_T=\{x_j\}_{j=1}^{n_t}$, where the input $x_j\in\mathcal{X}_T$. $\mathcal{D}_s$ and $\mathcal{D}_T$ are sampled from joint distributions $P(X,Y)$ and $Q(X,Y)$, respectively, and due to the domain shift, $P\neq Q$. The task is to learn a function $f(\bullet)$ from $\mathcal{D}_s$. Afterwards, the forecast function built on $\mathcal{D}_s$ will be applied to classify the unlabeled samples of $\mathcal{D}_T$.

B. Spectral Norm Regularization

In recent years, GANs have been drawing growing attention in the field of machine learning and successfully applied to various types of tasks [30]. However, GANs are still very difficult to train. In the training process, the discriminator may enter the ideal state early and can always distinguish between true and false samples. Therefore, for such a premature convergence case, gradient information would become less useful for proceeding and improving the model performance. To overcome the premature convergence issue and stabilize the

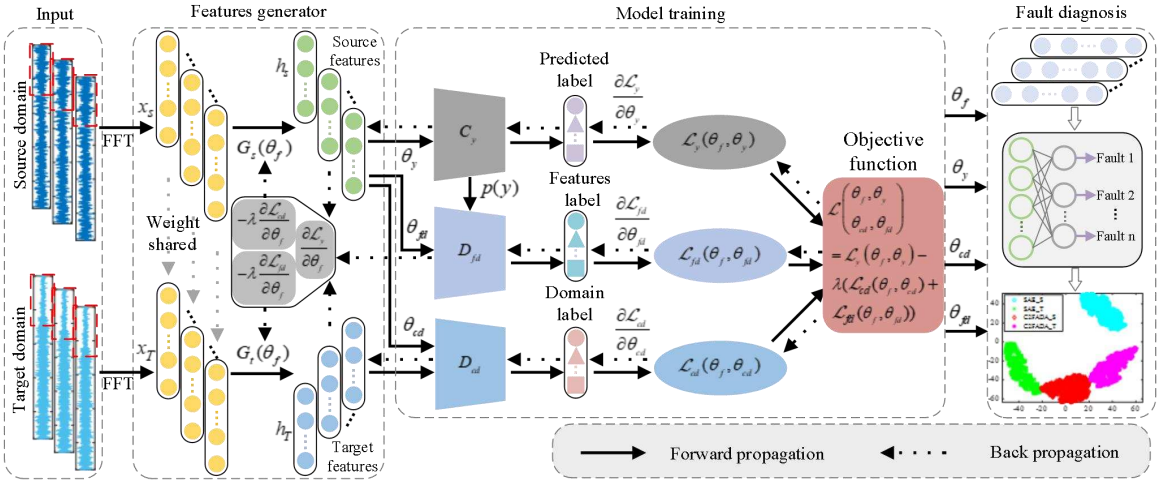


Fig. 2. The framework of the proposed C2FADA method.

training of discriminator networks, a simple and effective spectral norm regularization (SNR) method is introduced.

For the discriminator network $D(x)$, its input-output relationship can be expressed as

$$x_l = f_l\left(W_l x_{l-1} + b_l\right) \qquad l = 1,2,\ldots,L \tag{1}$$

where x_{l-1} is the input of the l-th layer, $f_l(\bullet)$ is the activation function (ReLU), W_l and b_l denote the layer weight matrix and bias vector, respectively, and the set of parameters can also be written as $\Theta = \{W_l, b_l\}_{l=1}^{L}$. The spectral norm of the discriminator network parameter matrix $W_{\Theta,x}$ can be represented as $\sigma(W_{\Theta,x})$. For a given vector x, $f_l(\bullet)$ is equivalent to a diagonal matrix $H_{\Theta,x}$. If the corresponding element in x_{l-1} is positive, the element on the diagonal is equal to one; otherwise, it is equal to zero [31]. Therefore, the discriminant network $D(x)$ can be expressed as the multiplication of multiple matrices, and its input-output relationship can be expressed as

$$D(x) = W_{\Theta,x}x = H_{\Theta,x}^{L} W_L H_{\Theta,x}^{L-1} W_{L-1} \cdots H_{\Theta,x}^{1} W_1. \tag{2}$$

Note that the ReLU function satisfies the 1-Lipschitz constraint, so $\sigma(H_{\Theta,x}^{l}) \le 1$ for every $l \in \{1,2,\ldots,L\}$. Therefore, we have

$$\begin{aligned}\sigma\left(W_{\Theta,x}\right) &\le \sigma\left(H_{\Theta,x}^{L}\right)\sigma\left(W_L\right)\sigma\left(H_{\Theta,x}^{L-1}\right)\sigma\left(W_{L-1}\right) \\ &\cdots\sigma\left(H_{\Theta,x}^{1}\right)\sigma\left(W_1\right) \le \prod_{l=1}^{L}\sigma\left(W_l\right).\end{aligned} \tag{3}$$

From this, it is concluded that to bound the spectral norm of the parameter matrix $W_{\Theta,x}$, is sufficient to bound the spectral norm of W_l for every $l \in \{1,2,\ldots,L\}$. To this end, the spectral norm is added as a regular term to the loss function of multilayer neural networks.

III. Proposed Method

This section presents the C2FADA method in detail. The section includes four parts, namely, the framework overview, the method of feature extraction, the domain alignment module, and the optimization algorithm. DA is a core of the proposed C2FADA method, therefore considerable attention is paid to it.

A. Bi-Level Adversarial Domain Adaptation Model

Considering the problem that rolling bearings are prone to failure in complex work conditions, to find an effective method to diagnose the faults of rolling bearings timely, a C2FADA approach is proposed. As shown in Fig. 2, the framework of this approach includes four modules: 1) the feature generator G using SAE; 2) the coarse-grained domain discriminator D_{cd}; 3) the fine-grained domain discriminator D_{fd}; 4) the fault classifier C_y with *softmax* loss.

As mentioned in section II, S and T are used to respectively denote the source and target domains. Meanwhile, the number of classes is denoted by N_{class}. The input data (features) and the associated class labels are presented by $X = \{x_1, x_2, \ldots x_n\}$ and a vector $Y = [y_1, y_2, \ldots, y_{N_{class}}]^T$, respectively.

In contrast to the existing methods of aligning different domains, we argue that the category information should be considered and used to facilitate the fault diagnosis task, thereby highlighting the distribution of each category to further refine feature alignment. As shown in Fig. 2, two modules are developed to conquer the challenges in this task. Firstly, in order to avert the performance degradation resulted from the domain shift, a domain-level alignment module, that is, the coarse-grained alignment module D_{cd}, is proposed. This module is used to align the distribution between two domains. Secondly, it is known that the objects in fault type recognition are usually different in the vibration data, a fine-grained alignment module D_{fd} is proposed accordingly. By aligning the two domains at a fine-grained level, and learning the domain invariant features of each fault class, this can better fulfill a coarse-to-fine transfer of feature knowledge.

B. Sparse Auto-encoder for Feature Extraction

A large amount of raw vibration data of rolling bearings can be obtained, but the collected data can become noisy after the

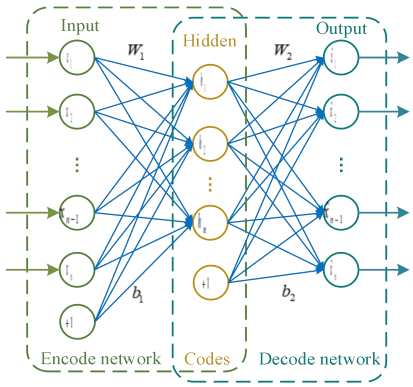


Fig. 3. The structure of the proposed SAE.

bearings work for a long time [32]. Therefore, it is required to pre-process the data and extract useful features. SAE provides a good approach to automatically process information and extract efficient features from unlabeled data. Thus, this paper uses SAE as the feature generator, which, working together with a sparse constraint, can mine more essential and discriminative features as well as avoid feature redundancy.

As shown in Fig. 3, the sparse auto-encoder is an unsupervised feature learning neural network with three layers, where the input layer represents the data inputs, the hidden layer represents the learned features, and the same dimension output layer represents the reconstructed inputs. The encoder network transforms the original data from a high dimensional space into hidden representation codes with lower dimension, while the decoder network can reconstruct the original inputs from the learned hidden codes without the need of label information.

Given a total of n samples, denoted by $X=\{x_1,x_2,\ldots x_i,\ldots x_n\}$, where each sample is defined in N dimensional space, that is, $x_i \in \mathbb{R}^{N\times 1}$. The encoding mapping function is

$$h_\theta = f_\theta(W_1 X + b_1) \tag{4}$$

where h_θ denotes the hidden layer feature, $f_\theta(\bullet)$ is the activation function (sigmoid function), W_1 and b_1 are the weight matrix and corresponding bias vector of the encode network. Similarly, the role of the decoder is to reconstruct X from h_θ, and the decoding mapping function is

$$\hat{X} = f_\theta\left(W_2 h_\theta + b_2\right) \tag{5}$$

where W_2 and b_2 are the weight matrix and corresponding bias vector in the decoding network. The learning process of SAE is to minimize the loss function [11]

$$J_{Spare}\left(\theta_f\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2}\left\|x_i-\hat{x}_i\right\|^2\right) + \beta\sum_{j=1}^{s} KL\left(\rho\left\|\hat{\rho}_j\right.\right) \tag{6}$$

$$KL\left(\rho\left\|\hat{\rho}\right.\right) = \rho\log\frac{\rho}{\hat{\rho}} + (1-\rho)\frac{1-\rho}{1-\hat{\rho}}. \tag{7}$$

where $\theta_f = \{W_1, b_1, W_2, b_2\}$ is the parameter set, s denotes the neurons number, β controls the weight of the sparsity penalty term, ρ represents the sparsity parameter, $\hat{\rho}$ denotes the average activation value of the j-th hidden unit, and the standard KL divergence function is used to realize the sparse representation of SAE hidden layer features.

SAE improves the performance of the traditional auto-encoder. It can learn the representative features of original data in a complex data environment while reducing the dimensionality efficaciously. In consequence, it is especially applicable for fault diagnosis under variable conditions.

C. Domain Alignment Module

Different domains are continuously aligned at the domain level and the class level. Specifically, we first reduce the differences between different domains on the whole, and then further align the same categories under different domains. This is different from most existing research, which usually only focuses on the domain-level alignment but ignores the class-level alignment. This is the merit of the proposed method.

1) Domain-level alignment

The features extracted generally have domain shifts between the source and target domains (features are denoted by h_S and h_T, respectively, in Fig. 2). Therefore, an adversarial learning method is adopted to achieve alignment at the domain-level. The process of adversarial domain adaptation is a zero-sum game process between the feature extractor and the domain discriminator. Specifically, the domain discriminator distinguishes the source and target domains features through learning, and maps the feature representations to coarse-grained label space [0, 1] to represent the source and target domains samples. The output of the domain discriminator is:

$$D_{cd}(h_\theta, \theta_{cd}) = sigmoid(W_3 h_\theta + b_3) \tag{8}$$

where W_3 and b_3 represent the weight matrix and corresponding bias vector.

Owing to the instability of GAN training process, spectral norm regularization is introduced. In this method, regular constraints are introduced from the perspective of the spectral norm of the neural network parameter matrix to make the training process more stable and easier to be converged. The logistical loss of the domain-level alignment is defined as:

$$\begin{aligned}\mathcal{L}_{D_{cd}}\left(\theta_f,\theta_{cd}\right) &= \frac{1}{n_S+n_T}\sum_{i=1}^{n_S+n_T}\mathcal{L}_{D_{cd}}\left(D_{cd}\left(G_f\left(x_i,\theta_f\right)\right), cd_i\right) \\ &+ \frac{\mu}{2}\sum\sigma\left(W_{D_{cd}}\right)^2 \\ &= -\frac{1}{n_s+n_T}\sum_{i=1}^{n_s+n_T}\Big[\, cd_i \log D_{cd}\left(G_S\left(x_i,\theta_f\right)\right) \\ &+ \left(1-cd_i\right)\times\log\left(1-D_{cd}\left(G_T\left(x_i,\theta_f\right)\right)\right)\Big] \\ &+ \frac{\mu}{2}\sum\sigma\left(W_{D_{cd}}\right)^2\end{aligned} \tag{9}$$

where G_s and G_T represent the feature extractors of the source and target domains respectively (a shared weight strategy is adopted for the two extractors here), θ_f is the parameter of the feature extractor (SAE), D_{cd} is a binary domain classifier (similar to the discriminator in GAN), and cd_i denotes the

coarse-grained label of x_i, whose value is 1 if x_i belongs to the source domain and 0 if target domain, μ is a regularization factor, $W_{D_{cd}}$ is the weight matrix of the discriminator D_{cd}.

2) Class-level alignment

In addition to domain-level alignment, class-level alignment should also be built to ensure that the same category features from different domains are close, so as to achieve more comprehensive DA. In order to achieve the alignment in class-level and learn the domain invariant features of each fault type, a fine-grained domain discriminator is built. Since the label knowledge of the source domain reflects the type of failure, the classifier G_y can provide predicted fault labels for each sample. Features extracted by the feature generator G_f can be divided into N_{class} types of fine-grained feature representations. The output of the class discriminator is represented as:

$$D_{fd}(h_s,\theta_{fd})=\frac{1}{2n_s}\sum_{i=1}^{n_s}\left(\log\frac{\exp\left(\theta_i^T h_i\right)}{\sum_{j=1}^{N_{class}}\exp(\theta_j^T h_i)}-p(y)\right)^2 \quad (10)$$

where θ_{fd} is the parameter of the discriminator, and $p(y)$ denotes the class probability output by classifier C.

Similar to the coarse-grained alignment module, class-level domain alignment is also realized by the adversarial learning method, and a regularization based on the spectral norm is imposed. According to the predicted labels, the class-level loss is defined as follows:

$$\begin{aligned}\mathcal{L}_{D_{fd}}\left(\theta_f,\theta_{fd}\right)&=\frac{1}{N_{class}}\frac{1}{n_S+n_T}\sum_{c=1}^{N_{class}}\sum_{i=1}^{n_S+n_T}\mathcal{L}_{D_{fd}}\left(D_{fd}\left(G_f\left(x_i,\theta_f\right)\right),fd_i\right)\\&\quad+\frac{\mu}{2}\sum\sigma\left(W_{D_{cd}}\right)^2\\&=-\frac{1}{N_{class}}\frac{1}{n_S+n_T}\sum_{c=1}^{N_{class}}\sum_{i=1}^{n_S+n_T}\Big[fd_i\log D_{fd}\left(G_S\left(x_i,\theta_f\right)\right)\\&\quad+\left(1-fd_i\right)\times\log\left(1-D_{fd}\left(G_S\left(x_i,\theta_f\right)\right)\right)\Big]\\&\quad+\frac{\mu}{2}\sum\sigma\left(W_{D_{cd}}\right)^2\end{aligned} \quad (11)$$

where D_{fd} represents the fine-grained domain discriminator, fd_i is the fine-grained domain labels of x_i, and $W_{D_{fd}}$ denotes the weight matrix of D_{fd}.

D. Softmax Classifier for Fault Diagnosis

The softmax classifier has been widely used for multi-class classification tasks [33]. In the source domain, given $D_s=\{(x_i,y_i)\}_{i=1}^{n_s}$, where $x_i\in\mathbb{R}^{N\times1}$ denotes the labeled training sample, and $y_i\in\{1,2,\ldots,N_{class}\}$ represents the corresponding label. For an input sample x_i, the softmax function is used to calculate, by regression analysis, the probability $p\left(y_i=\hat{y}_i\mid x_i\right)$ for each label $\hat{y}_i\left(\hat{y}_i=1,2,\ldots,N_{class}\right)$. The hypothesis function is given by:

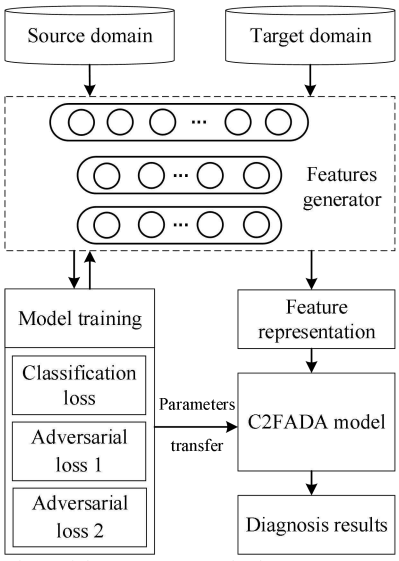


Fig. 4. The flowchart of the C2FADA method.

$$h_\theta\left(x_i\right)=\begin{bmatrix}p\left(y_i=1\right)\\p\left(y_i=2\right)\\\vdots\\p\left(y_i=N_{class}\right)\end{bmatrix}=\frac{1}{\sum_{k=1}^{N_{class}}\exp\left(\theta_k^T x_i\right)}\begin{bmatrix}\exp\left(\theta_1^T x_i\right)\\\exp\left(\theta_2^T x_i\right)\\\vdots\\\exp\left(\theta_{N_{class}}^T x_i\right)\end{bmatrix} \quad (12)$$

where $\theta=\left[\theta_1,\theta_2,\ldots,\theta_{N_{class}}\right]^T$ represents the softmax parameters.

The cross-entropy function $\mathcal{L}_y$ is adopted as the classification loss function [34], and $\mathcal{L}_y$ is defined as:

$$\begin{aligned}\mathcal{L}_y\left(\theta_f,\theta_y\right)&=\frac{1}{n_s}\sum_{i=1}^{n_s}\mathcal{L}_y\left(G_y\left(G_f\left(x_i,\theta_f\right)\right),y_i\right)\\&=-\frac{1}{n_s}\left[\sum_{i=1}^{n_s}\sum_{k=1}^{N_{class}}1\{y_i=k\}\log\frac{\exp\left(\theta_k^T G_f\left(x_i\right)\right)}{\sum_{j=1}^{N_{class}}\exp(\theta_j^T G_f\left(x_i\right))}\right]\end{aligned} \quad (13)$$

where $1\{y_i=k\}$ is the indicator function, whose equals to 1 (if the condition is true) or 0 (otherwise).

E. Diagnosis Procedure

Fig. 4 presents a simple flowchart of the C2FADA-based fault diagnosis process.

In the training phase, the proposed model is trained using the labeled source domain data and unlabeled target domain data. The parameters of each module are learned by optimizing the objective function. Then, the fault diagnosis model based on C2FADA is obtained.

In the testing phase, the target sample is firstly input into the parameter-sharing feature generator to obtain the feature representation. Then, the obtained features are input into the fault classifier to obtain the fault category of the target domain sample. Thereby, the diagnostic result of the target domain sample is determined.

F. Optimization Algorithm

By integrating equations (9), (11), and (13) together, the overall loss function is as follows:

$$\begin{aligned}&\mathcal{L}\left(\theta_f,\theta_y,\theta_{cd},\theta_{fd}\right)\\&\quad=\mathcal{L}_y\left(\theta_f,\theta_y\right)-\lambda\left(\mathcal{L}_{cd}\left(\theta_f,\theta_{cd}\right)+\mathcal{L}_{fd}\left(\theta_f,\theta_{fd}\right)\right)\end{aligned}\tag{14}$$

where λ is the tradeoff parameter that controls the level of adversarial domain adaptation.

The goal of optimization is to find a set of optimal parameters $(\theta_f,\theta_y,\theta_{cd},\theta_{fd})$ so that G_y minimizes the loss of the classification of all labels, and at the same time, D_{cd} and D_{fd} maximize the loss of the domain labels. Hence, the optimization objective can be written as:

$$\left(\hat{\theta}_f,\hat{\theta}_y\right)=\arg\min_{\theta_f,\theta_y}\mathcal{L}\left(\theta_f,\theta_y,\theta_{cd},\theta_{fd}\right)\tag{15}$$

$$\left(\hat{\theta}_{cd},\hat{\theta}_{fd}\right)=\arg\max_{\theta_{cd},\theta_{fd}}\mathcal{L}\left(\theta_f,\theta_y,\theta_{cd},\theta_{fd}\right)\tag{16}$$

In the training stage, the adversarial between minimizing the loss of the label predictor and maximizing the loss of the domain discriminator is the transfer process of the model. Among this process, the model automatically extracts the features of transfer between different domains. Actually, *G* and *D* are connected through a gradient reversal layer (GRL) to learn the invariant features of the domain [35]. GRL only affects the gradient calculation in the backward pass because it is placed between *G* and *D*.

$$\begin{aligned}\theta_f&\leftarrow\theta_f-\alpha\left(\frac{\partial\mathcal{L}_y}{\partial\theta_f}-\lambda\left(\frac{\partial\mathcal{L}_{cd}}{\partial\theta_{cd}}+\frac{\partial\mathcal{L}_{fd}}{\partial\theta_{fd}}\right)\right)\\\theta_y&\leftarrow\theta_y-\alpha\frac{\partial\mathcal{L}_y}{\partial\theta_y}\\\theta_{cd}&\leftarrow\theta_{cd}-\alpha\frac{\partial\mathcal{L}_{cd}}{\partial\theta_{cd}}\\\theta_{gd}&\leftarrow\theta_{gd}-\alpha\frac{\partial\mathcal{L}_{gd}}{\partial\theta_{gd}}\end{aligned}\tag{17}$$

Stochastic gradient descent (SGD) method can be applied to address the optimization problems (15) and (16). In the training process, the parameters are updated according equation (17), where α represents the learning rate, $-\lambda$ denotes the gradient reversal.

IV. Experiments

Two real datasets are considered to test the performance of the proposed method. The modeling setups and results for the two datasets are detailed in Sections A&B and Section C, respectively.

A. Set-ups of CWRU Dataset

1) Datasets

The experiment builds the multi-condition rolling bearing dataset based on the bearing data center of Case Western Reserve University [36]. The bearing vibration data of 0~3hp load were recorded during the experiment, and the vibration data of different working conditions includes four bearing states, namely normal (NR), fault of inner race (FIR), fault of outer race (FOR), and fault of rolling ball (FRB). Each fault state corresponds to four fault diameters of 0.007 inches, 0.014 inches, 0.021 inches, and 0.028 inches, respectively.

In this article, two domains are generated according to the load of the motor. Specifically, the source domain consists of data obtained from four bearing states under the same motor load, $\mathcal{D}_S=\left\{NR^0,FIR_1^0,FOR_1^0,FRB_1^0,FIR_2^0,FOR_2^0,FRB_2^0\right\}$, with fault diameters of 0.007 and 0.014 inches. The target domain consists of data obtained from four bearing states under another motor load, $\mathcal{D}_T=\left\{NR^1,FIR_1^1,FOR_1^1,FRB_1^1,FIR_2^1,FOR_2^1,FRB_2^1\right\}$, with fault of the two types of (0.007 and 0.014 diameters). The task is to get the status label of the unlabeled data in the target domain, $Y=[y_1,y_2,\ldots,y_{N_{class}}]^T$. Therefore, according to the two fault diameters and four different loads (0~3hp), the bearing conditions are split into several cases, resulting in six types of domain shift tasks, details of which are shown in Table I. The fast Fourier transform (FFT) is performed on the signal samples, and the first half of the frequency coefficients are retained as the input of the model.

TABLE I
Tasks and Their Associated Fault Labels

Task	Domain shift	NR label	0.007 inches fault labels	0.014 inches fault labels
H0→H1	0hp→1hp	1	2, 3, 4	5, 6, 7
H0→H2	0hp→2hp	1	2, 3, 4	5, 6, 7
H0→H3	0hp→3hp	1	2, 3, 4	5, 6, 7
H1→H2	1hp→2hp	1	2, 3, 4	5, 6, 7
H1→H3	1hp→3hp	1	2, 3, 4	5, 6, 7
H2→H3	2hp→3hp	1	2, 3, 4	5, 6, 7

2) Baselines and Settings

The proposed C2FADA approach was compared with three traditional methods without using DA and other six methods using DA. These compared methods are: BP method [10], Softmax method, SAE method [11], domain adaptation in fault diagnosis (DAFD) method [37], MTLF method [16], double-level adversarial domain adaptation network (DL-ADAN) method [38], deep adversarial domain adaptation (DADA) method [32], deep convolution domain-adversarial transfer learning (DCDATL) method [39], and CNN-based C2FAFA (C2FADA-CNN) method.

The three traditional supervised methods, BP, Softmax, and SAE, have been widely used for fault diagnosis applications. Among the DA methods, the DAFD method reduces the differences between different domains by learning the transferable features between domains, while strengthening the identifiable features in the original data; the DL-ADAN method combines domain discriminator and classifier to bridge differences between domains through adversarial training; the MTLF method learns the instance weights and Mahalanobis distance to minimize the inter-class distance and maximize the intra-class distance for the target domain; the DADA method uses the SAE as the feature extractor and the adversarial training is performed using the GRL; the DCDATL method designs a deep residual network to extract the features from two domains and the joint distribution of the samples from two domains is utilized for domain-adversarial training; the C2FADA-CNN method uses the CNN as the feature extractor,

TABLE II
THE TEST RESULTS (%) OF DIFFERENT METHODS ON THE CWRU DATASET

Methods	H0→H1		H0→H2		H0→H3		H1→H2		H1→H3		H2→H3		Average (%)
	Average (%)	STDEV	Average (%)	STDEV	Average (%)	STDEV	Average (%)	STDEV	Average (%)	STDEV	Average (%)	STDEV	
BP	81.375	0.008	82.698	0.043	71.012	0.030	85.188	0.032	84.246	0.046	73.681	0.041	79.700
Softmax	76.811	0.063	71.320	0.057	79.545	0.051	69.771	0.057	71.564	0.067	66.571	0.047	72.597
SAE	84.372	0.045	80.018	0.035	84.814	0.047	83.261	0.041	84.099	0.046	82.512	0.035	83.179
MTLF	82.665	0.040	68.544	0.037	77.862	0.042	83.671	0.047	85.542	0.039	84.824	0.042	80.518
DAFD	96.024	0.054	90.131	0.043	95.503	0.036	94.274	0.034	90.290	0.053	97.321	0.030	93.924
DL-ADAN	95.446	0.027	94.304	0.032	96.339	0.027	95.232	0.031	90.036	0.053	93.232	0.038	94.098
DADA	95.393	0.009	95.500	0.031	94.696	0.056	93.661	0.045	90.643	0.032	93.375	0.036	93.878
DCDATL	91.786	0.051	94.250	0.050	95.107	0.049	95.607	0.053	94.250	0.059	97.429	0.053	94.738
C2FADA-CNN	96.607	0.031	96.357	0.033	95.071	0.027	97.393	0.024	94.375	0.028	97.839	0.030	96.274
Proposed method	**98.050**	**0.022**	**98.498**	**0.024**	**97.452**	**0.020**	**96.339**	**0.026**	**97.007**	**0.021**	**98.500**	**0.023**	**97.641**

which is different with the proposed C2FADA method using the SAE as the feature extractor.

This work uses the classification accuracy of the target task as the performance evaluation index, which is popular in fault diagnosis [11], [37]. The calculation formula is defined as

$$Acc = \frac{\left|x : x \in \mathcal{D}_T \cap prediction(y) = actual(y)\right|}{\left|x : x \in \mathcal{D}_T\right|} \tag{18}$$

where $\mathcal{D}_T$ denotes the dataset of the target domain, $prediction(y)$ denotes the labels predicted by the classifier, $actual(y)$ are the actual labels of x.

In addition, the diagnostic performance of the seven DA methods is also evaluated by Macro_F1 score index, which can be calculated as follows:

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \tag{19}$$

$$Macro\_F1 = \frac{\sum_{i=1}^{N_{class}} F1_i}{N_{class}} \tag{20}$$

where *precision* represents the ratio of the number of positive samples correctly classified to the number of positive samples determined by the classifier, and *recall* denotes the proportion of correctly classified samples to the total number of samples.

B. Results and Analysis

1) Experimental Results

In the work, each experiment was conducted ten times, and the average accuracy and the standard deviation of the classification results of all tasks was recorded. The comparison of the results from the nine contrasted methods and the proposed C2FADA method, for the six diagnosis tasks, are shown in Table II, where the "Average" and "STDEV" represent the average accuracy and the standard deviation of the diagnosis results, respectively. It can be easily discovered that the test and diagnosis performance of the proposed method under different tasks outperforms other listed methods, and the test accuracies are basically above 95%, which clearly verifies the availability of the C2FADA method. Specifically, the accuracy of the baseline Softmax method is only 72.597%, which is 25.044% lower than that of the proposed method. BP acquires the best accuracy of 85.188% amidst the three methods without DA, which is 12.453% lower than the proposed approach. The accuracy of SAE diagnosis results is above 80%, even so, there is still a certain distance from the required diagnosis. In these methods with DA, the DAFD method shows the higher accuracy of 97.321%, but is still lower than the proposed method.

The DL-ADAN, DADA, and DCDATL methods adopt the adversarial training to learn the domain information and obtain better performance than simple domain adaption methods, and the accuracy of these methods is over 90%. However, these methods only align the distribution of domains, and ignore the fine-grained information of the same fault categories from different domains, and thus obtain lower accuracy. Different with the proposed method using SAE to learn the features of the source and target domains, the C2FADA-CNN method uses the CNN as the feature extractor. From Table II, it shows that the C2FADA-CNN method obtains similar performance with the proposed method, and even achieves higher accuracy (97.393%) than the proposed method on the H1→H2 task. However, on the most tasks, the accuracy of the proposed method is slightly higher than that of the C2FADA-CNN method, which means that the SAE can more efficiently extract the domain features in the adversarial training. Additionally, the proposed method achieves better stability results compared with other comparison approaches for multiple-class classification tasks. This can be interpreted that the adoption of DA improves the capability of the model classifier, and the adoption of spectral norm regularization imposed on the adversarial training makes it more effective in reducing the domain shift. In addition, from Table II, it can be also seen that the standard deviation of the diagnosis results of the proposed method is smaller than that of other methods on most tasks, especially for other DA methods. That means that the proposed method has more stable performance in fault diagnosis of rolling bearing.

In order to further verify the superiority of the proposed method, the comparison results for four DA methods based on

TABLE III
THE MACRO_F1 SCORE RESULTS (%) OF FOUR DA METHODS ON THE CWRU DATASET

Methods	H0→H1 (%)	H0→H2 (%)	H0→H3 (%)	H1→H2 (%)	H1→H3 (%)	H2→H3 (%)	Average (%)
MTLF	80.228	70.032	71.174	72.883	83.414	82.580	76.719
DAFD	93.401	89.021	92.953	92.224	91.336	95.924	92.477
DL-ADAN	94.006	94.104	95.319	93.886	90.547	91.511	93.229
DADA	94.281	95.003	94.621	92.908	90.152	92.968	93.322
DCDATL	91.219	93.886	93.739	94.597	93.195	96.286	93.820
C2FADA-CNN	96.532	95.996	94.720	96.537	93.837	97.371	95.832
Proposed method	**97.255**	**97.329**	**97.264**	**94.826**	**95.371**	**98.902**	**96.825**

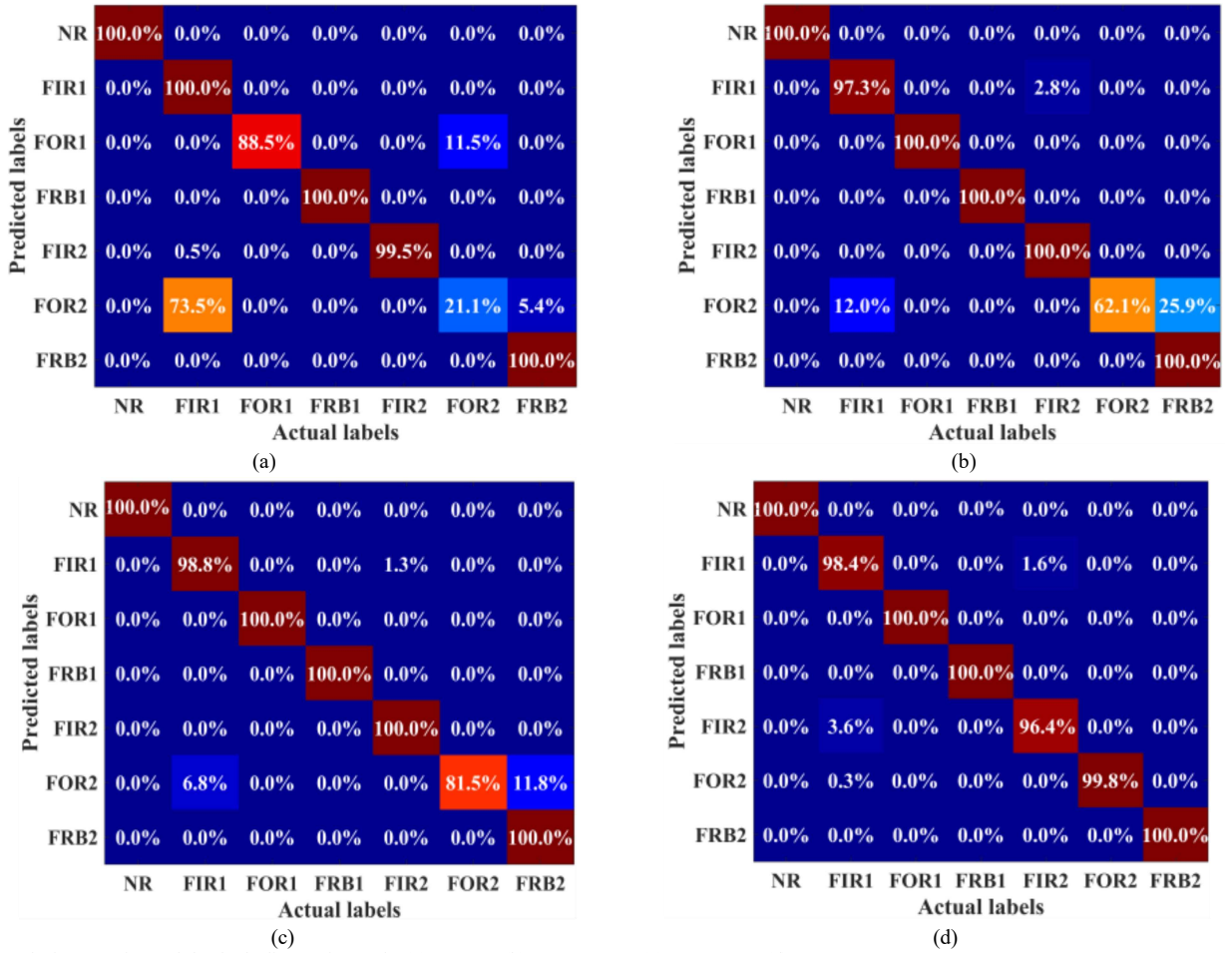


Fig. 5. Confusion matrices of the fault diagnosis result. (a) DAFD; (b) DL-ADAN; (c) DCDATL; (d) C2FADA.

Macro_F1 score following (20) are shown in Table III. It can be clearly observed from the table that the performance of the proposed method is better than the other six DA methods.

In order to understand for which cases the performance is (and is not) significantly improved by introducing DA, the confusion matrices obtained by four DA methods (namely, DAFD, DL-ADAN, DCDATL, and C2FADA) on the H0→H1 task are shown in Fig. 5. For DAFD, it can be readily observed that there are the classification errors in category FOR2, and the accuracy of FOR2 is only 21.1%. The main reason for such situation is that their fault categories are similar, but only the fault severity is different; this makes the data samples located in the decision boundary prone to being misclassified. The DL-ADAN and DCDATL methods have also the different classification errors in category FOR2. For the proposed method, although there are a few misclassifications in the diagnosis of FIR1 and FIR2, the overall accuracy rate remains above 96%. The results show that the overall performance of the proposed C2FADA method is superior to other compared methods, and in the task of diagnosis, this method has a significant superiority for obtaining more accurate recognition performance.

2) Feature Visualization

In order to definitely present the matching degree of the data distribution before and after the DA, an effective technology t-SNE [40] is adopted to map the high-dimensional samples in

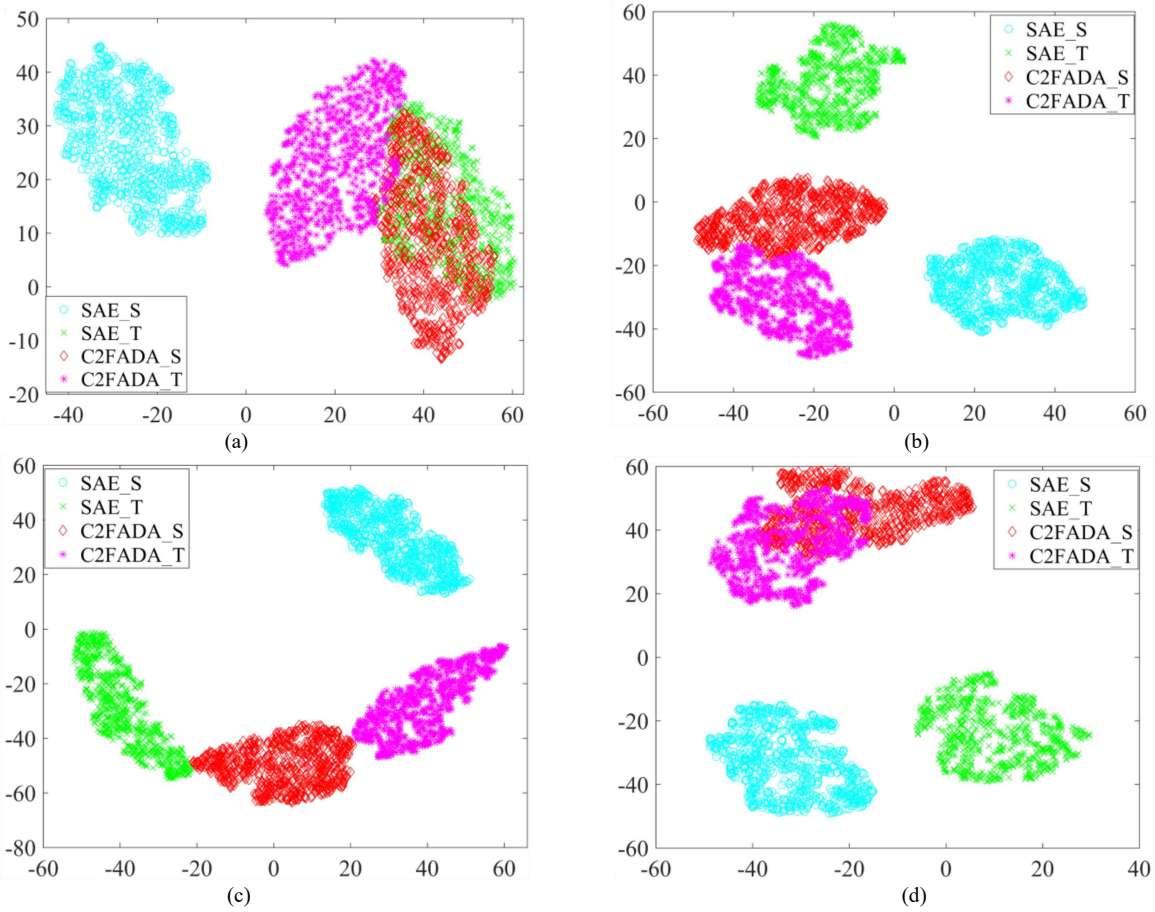


Fig. 6. Feature visualization via t-SNE. (a) NR; (b) FIR; (c) FOR; (d) FRB.

the learned feature space to the two-dimensional space, enabling the visualization of the high-dimensional data. The 0.007 inches fault diameter in the H0→H1 diagnosis task is taken as an example, and the visualization effects of the four bearing states features are shown in Fig. 6. It is obvious from Fig. 6(a) that the discrepancy between the features extracted by the proposed model (features from $\mathcal{D}_s$ are marked by the magenta asterisks and features from $\mathcal{D}_T$ are marked by the red diamonds) is smaller than the features of SAE (points marked by the cyan circle are from $\mathcal{D}_s$ and the green cross are from $\mathcal{D}_T$). Similar observations can also be made in the other three figures. The observed discrepancy between the results produced by the proposed method and SAE can be explained that the method of this article proposed clusters the data of the same category and separates different categories, thus makes the learned feature cluster better and more efficient for reducing the domain shift. At the same time, the alignment of the source and the target domain is well realized. Therefore, the model trained on the source task can be commendably applied to the fault diagnosis of the target task.

3) Effect of Feature Dimension

The model proposed in this article mainly relies on SAE as the generator to acquire the hidden features of the original data. Therefore, the choice of the neurons number in the hidden layer (the input feature dimension) will greatly influence the performance of machine learning. In this subsection, the influence of feature dimension on fault diagnosis of variable

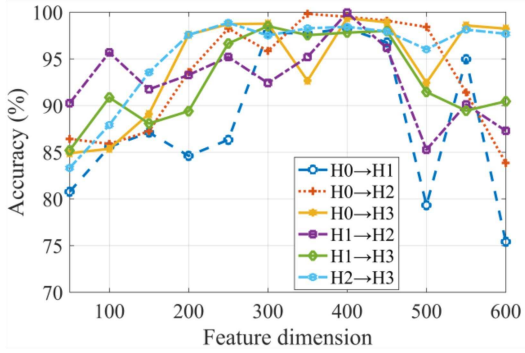


Fig. 7. Effects of the features dimension in the feature extractor.

conditions is studied, and the results are shown in Fig. 7. The original sample dimension is 600, so the feature space is [1, 600]. The model diagnostic performance is sensitive to feature dimensions. Specifically, as the feature dimension increases, the diagnosis accuracy initially shows a rising trend, but then the accuracy drops somewhere, and the optimal dimension was identified to be between 350 and 450. This can be explained as low-dimensional representations usually cannot capture enough representative information for data generation and classification, while high-dimensional representations are prone to overfitting. According to this result, the feature dimension was chosen to be 400. The experimental results show that this choice is appropriate and can guarantee a good performance.

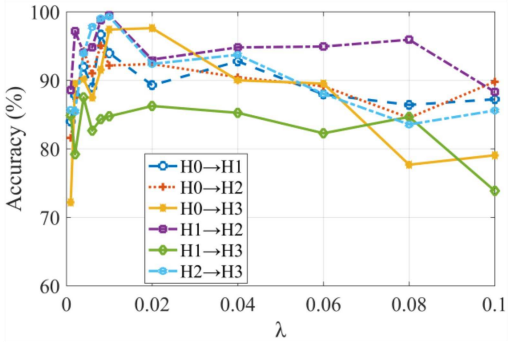


Fig. 8. Effect of the trade-off parameter λ on the overall performance.

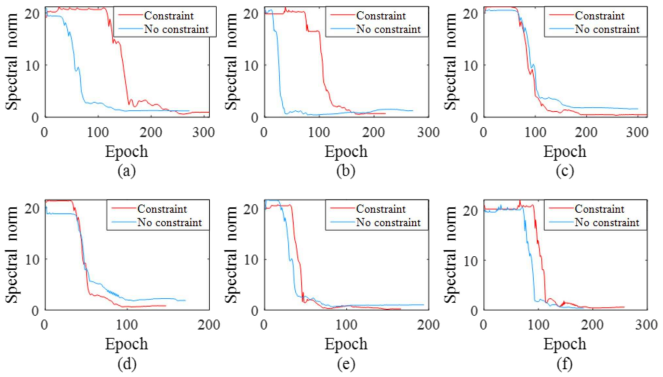


Fig. 9. Spectral norm of discriminator weight matrix. (a) H0→H1; (b) H0→H2; (c) H0→H3; (d) H1→H2; (e) H1→H3; (f) H2→H3.

4) Effect of Trade-off Parameter

The trade-off parameter λ in the overall loss function (14) controls the adversarial domain adaptation level. Here, the sensitivity of the overall objective function to the change in λ is investigated. Experiments are carried out on six tasks of the CWRU dataset, and the experimental results are shown in Fig. 8. It can be observed that when λ changes in the range [0.004, 0.014], the overall performance of the model is relatively good. It is worth noting that when λ=0.01, the model shows the best performance. Therefore, this value is selected as the value of the trade-off parameter in this paper.

5) Spectral Norm Analysis of Discriminator Weights

In order to illustrate the influence of the spectral norm regularization on the GAN training process, this subsection provides a visualization of the training process of the spectral norm of the discriminator model, as shown in Fig. 9. It can be observed that the spectral norm of the discriminator gradually decreases with the training, and finally stabilizes after a number of iterations. This means that the effect of constraints is gradually weakened as the algorithm converges, which is consistent with what is expected from the design of the discriminator. It should be pointed out that the results in the figure are based on the C2FADA method, where the regularization constraints are removed on the premise of ensuring higher accuracy. Take Fig. 9(a) as an example, where the red curve and blue curve represent constrained and unconstrained conditions, respectively. It can observe that when there are no constraints, the model stabilizes after 70 iterations. However, after constraints are introduced to the model, it stabilizes after 140 iterations. This can be explained that due to the introduction of the spectral norm regularization method, it imposes gradient constraints on the parameter matrix in the model discriminant network, which limits the convergence speed of the discriminator, thereby improves the training stability of the entire network. The direct result of these effects is to improve the data generation quality of the generator during the game between the generator network and the discriminator network.

In addition, the influence of spectral norm regularization on fault diagnosis results of the six tasks is shown in Table IV, where the column "Without SNR (%)" represents the fault diagnosis accuracy obtained by the proposed method without using the spectral norm regularization, and correspondingly, the column "With SNR (%)" means the fault diagnosis accuracy obtained by the proposed method using the spectral norm regularization. From this table, it can been seen that the proposed method using the spectral norm regularization can obtain higher fault diagnosis accuracy. Take the H0→H1 task as an example. The accuracy of the proposed method without using the spectral norm regularization is 93.286%, which is 4.081% lower than that of the proposed method using the spectral norm regularization. For the six tasks, the average accuracy is 94.070%, which is 3.66% lower than that of the proposed method using the spectral norm regularization. Therefore, adding the spectral norm regularization to the proposed method can better train the model and realize better fault diagnosis performance.

TABLE IV
INFLUENCE OF SPECTRAL NORM REGULARIZATION ON FAULT DIAGNOSIS RESULTS OF THE SIX TASKS

Task	Without SNR (%)	With SNR (%)
H0→H1	93.286	98.050
H0→H2	96.482	98.498
H0→H3	94.035	97.452
H1→H2	93.161	96.339
H1→H3	92.005	97.007
H2→H3	95.450	98.500

6) Convergence Performance

The convergence performance of the model was evaluated by calculating the loss function of different methods. The comparison results of the loss functions of different methods under variable working conditions are shown in Fig. 10. It can be seen from the figure that the proposed method can achieve fast and stable convergence in the task of rolling bearing diagnosis, and obtain the minimum loss value. Although the results given by other methods also show good convergence

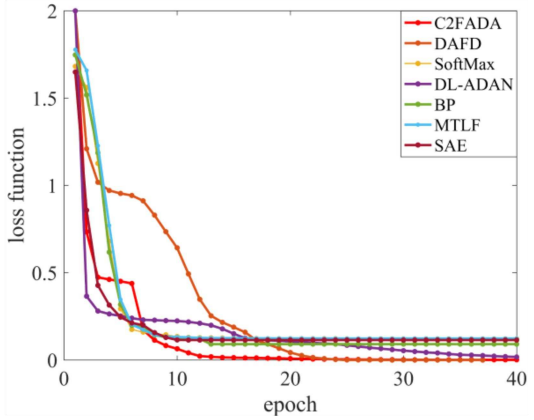


Fig. 10. The loss function of the training process.

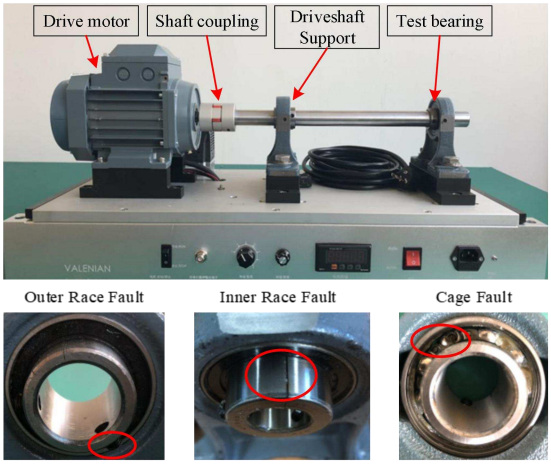


Fig. 11. Bearings test platform, and three types of faults in the dataset.

performances, the classification errors are larger. These demonstrate that the proposed method is more effective than the compared methods for cross-domain fault diagnosis.

C. Validation on Rolling Bearing Test Platform Dataset

In order to further verify the performance of the proposed method, a rolling bearing fault test platform is established and the bearing vibration data required for the experiment is collected. In the experiment, the vibration data of five bearing states were tested, including normal (NR), fault of inner race (FIR), fault of outer race (FOR), fault of rolling ball (FRB), fault of bearing cage (FBC), and fault of composite bearing (FCB). For each state, the bearing operated under three disparate speeds (1800RPM, 2100RPM, and 2400RPM), and the sampling frequency is 48000Hz. The collected data includes the shaft coupling end and the non-driving end (that is, the test bearing end), and the test platform is shown in Fig. 11.

Corresponding to the CWRU dataset, the fault and normal conditions are considered according to different fault locations and operating speeds. In addition, new types of bearing cage fault and composite fault have been added. Six transfer tasks are tested, and the details of the task sample are described in Table V.

TABLE V
DETAILS OF THE ROLLING BEARINGS TEST PLATFORM DATASET

Transfer task	Source domain (RPM)	Target domain (RPM)	Number of samples
P1→P2	1800	2100	6492
P2→P1	2100	1800	6492
P1→P3	1800	2400	6492
P3→P1	2400	1800	6492
P2→P3	2100	2400	6492
P3→P2	2400	2100	6492

Compared with the CWRU dataset, the bearing platform dataset is more in line with industrial scenarios because it contains bearing cage fault and composite bearing fault. Therefore, the difficulty of fault diagnosis has also increased. The experimental results are shown in Fig. 12, and compared with six different methods. It can be observed that the fault diagnosis accuracy of the platform dataset is generally lower than that of the CWRU dataset. However, compared with other methods, this method can still achieve the best performance under different operating conditions, which further verifies the availability and meliority of this method.

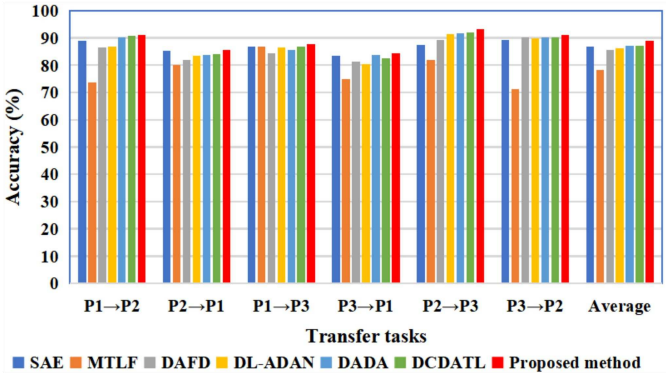


Fig. 12. Classification results of the six tasks on the bearing test platform.

V. CONCLUSIONS

This paper proposed a novel C2FADA method to address the problem of domain difference in rolling bearings under variable conditions. This method adopts a novel framework by making good use of inter-domain fault feature information and under-domain fault category information, and combines source domain label information to optimize the feature representation and parameters of the learning model. Compared with traditional methods and several existing adaptation methods, the combination of domain-level alignment and class-level alignment can better identify the failure types of rolling bearings and enhance the accuracy of the diagnosis results. Meanwhile, the introduction of the spectral norm regularization makes the training process of the model more stable and shows better convergence property. Through feature visualization, it illustrates the availability of the C2FADA method in reducing the distribution difference between domains. Apart from this, the proposed method can achieve more comprehensive domain adaptation without need of manual feature extraction, and this makes it more suitable for generalization to practical applications.

The main limitation of the proposed method is that it needs enough source domain data and corresponding labels to be constructed to train the model. So, the next step of the research is to develop methods and algorithms that can be used to effectively train good models when the data and labels are not balanced.

REFERENCES

[1] J. Wan, S. Tang, D. Li, S. Wang, C. Liu, H. Abbas, and A. V. Vasilakos, "A manufacturing big data solution for active preventive maintenance," *IEEE Trans. Ind. Inform.*, vol. 13, no. 4, pp. 2039-2047, Aug. 2017.

[2] Z.He , H.Shao ,J. Lin , J.Cheng , Y.Yang , "Transfer fault diagnosis of b earing installed in different machines using enhanced deep auto-encoder, " . Measurement, vol.152, Art no.107393,2020.

[3] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Proc.*, vol. 108, pp. 33-47, 2018.

[4] R. -B. Sun, F. -P. Du, Z. -B. Yang, X. -F. Chen, and K. Gryllias, "Cyclostationary analysis of irregular statistical cyclicity and extraction of rotating speed for bearing diagnostics with speed fluctuations," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1-11, Art no. 3514011, 2021.

[5] B. Hou, D. Wang, Y. Wang, T. Yan, Z. Peng, and K. -L. Tsui, "Adaptive weighted signal preprocessing technique for machine health monitoring," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1-11, Art no. 3504411, 2021.

[6] Y. Sun and J. Yu, "Adaptive sparse representation-based minimum entropy deconvolution for bearing fault detection," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1-10, Art no. 3513010, 2022.

[7] J. Li, R. Huang, G. He, Y. Liao, Z. Wang, and W. Li, "A two-stage transfer adversarial network for intelligent fault diagnosis of rotating machinery with multiple new faults," *IEEE/ASME Trans. Mechatron.*, vol. 26, no. 3, pp. 1591-1601, Jun. 2021.

[8] F. Deng, S. Guo, R. Zhou, and J. Chen, "Sensor multifault diagnosis with improved support vector machines," *IEEE Trans. Autom. Sci. Eng.,* vol. 14, no. 2, pp. 1053-1063, Apr. 2017.

[9] H. Shao, H. Jiang, F. Wang, and Y. Wang, "Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet," *ISA Trans.,* vol. 69, pp. 187-201, 2017.

[10] J. Li, X. Yao, X. Wang, Q. Yu, and Y. Zhang, "Multiscale local features learning based on BP neural network for rolling bearing intelligent fault diagnosis," *Meas.*, vol. 153, Art no. 107419, 2020.

[11] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst. Man Cybern. Syst.,* vol. 49, no. 1, pp. 136-144, Jan. 2019.

[12] Z. -H. Liu, L. -B. Jiang, H. -L. Wei, L. Chen and X. -H. Li, "Optimal Transport-Based Deep Domain Adaptation Approach for Fault Diagnosis of Rotating Machine," *IEEE Trans. Instrum. Meas*, vol. 70, pp. 1-12, 2021, Art no. 3508912, doi: 10.1109/TIM.2021.3050173.

[13] Y. Zhou, Y. Dong, H. Zhou, and G. Tang, "Deep dynamic adaptive transfer network for rolling bearing fault diagnosis with considering cross-machine instance," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1-11, Art no. 3525211, 2021.

[14] K. Zhao, H. Jiang, K. Wang, and Z. Pei, "Joint distribution adaptation network with adversarial learning for rolling bearing fault diagnosis," *Knowl.-Based Syst.*, vol. 222, Art no. 106974, 2021.

[15] X. Li, Z. Zhang, L. Gao, and L. Wen, "A new semi-supervised fault diagnosis method via deep coral and transfer component analysis," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 6, no. 3, pp. 690-699, Jun. 2022.

[16] Y. Xu, S. J. Pan, H. Xiong, Q. Wu, R. Luo, H. Min, and H. Song, "A unified framework for metric transfer learning," *IEEE Trans. Knowl. Data Eng.,* vol. 29, no. 6, pp. 1158-1171, Jun. 2017.

[17] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*., 2017, pp. 2962-2971.

[18] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F.-Y. Wang, "Generative adversarial networks: Introduction and outlook," *IEEE/CAA J. Autom. Sinica.,* vol. 4, no. 4, pp. 588-598, 2017.

[19] W. Wan, S. He, J. Chen, A. Li, and Y. Feng, "QSCGAN: An un-supervised quick self-attention convolutional GAN for LRE bearing fault diagnosis under limited label-lacked data," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1-16, Art no. 3527816, 2021.

[20] Z. Chen, G. He, J. Li, Y. Liao, K. Gryllias, and W. Li, "Domain adversarial transfer network for cross-domain fault diagnosis of rotary machinery," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 8702-8712, Nov. 2020.

[21] Y. Li, Y. Song, L. Jia, S. Gao, Q. Li, and M. Qiu, "Intelligent fault diagnosis by fusing domain adversarial training and maximum mean discrepancy via ensemble learning," *IEEE Trans. Ind. Inform.,* vol. 17, no. 4, pp. 2833-2841, Apr. 2021.

[22] D. She, M. Jia, and M. Pecht, "Weighted entropy minimization based deep conditional adversarial diagnosis approach under variable working conditions," *IEEE/ASME Trans. Mechatron.*, vol. 26, no. 5, pp. 2440-2450, Oct. 2021.

[23] W. Fu, X. Jiang, C. Tan, B. Li, and B. Chen, "Rolling bearing fault diagnosis in limited data scenarios using feature enhanced generative adversarial networks," *IEEE Sensors J.*, vol. 22, no. 9, pp. 8749-8759, May 2022.

[24] J. He, M. Ouyang, Z. Chen, D. Chen, and S. Liu, "A deep transfer learning fault diagnosis method based on WGAN and minimum singular value for non-homologous bearing," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1-9, Art no. 3509109, 2022.

[25] Y. Zheng, D. Huang, S. Liu, and Y. Wang, "Cross-domain object detection through coarse-to-fine feature adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit*., 2020, pp. 13763-13772.

[26] Z. Chai and C. Zhao, "A fine-grained adversarial network method for cross-domain industrial fault diagnosis," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 3, pp. 1432-1442, Jul. 2020.

[27] J. Miao, B. Zhang, and B. Wang, "Coarse-to-fine joint distribution alignment for cross-domain hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 12415-12428, 2021.

[28] Z. Huo, Y. Zhang, L. Shu, and M. Gallimore, "A new bearing fault diagnosis method based on fine-to-coarse multiscale permutation entropy, laplacian score and SVM," *IEEE Access*, vol. 7, pp. 17050-17066, 2019.

[29] X. Jiang, J. Wang, J. Shi, C. Shen, W. Huang, and Z. Zhu, "A coarse-to-fine decomposing strategy of VMD for extraction of weak repetitive transients in fault diagnosis of rotating machines," *Mech. Syst. Signal Process.*, vol. 116, pp. 668-692, 2019.

[30] C. Wang, C. Xu, X. Yao, and D. Tao, "Evolutionary generative adversarial networks," *IEEE Trans. Evolut. Comput.*, vol. 23, no. 6, pp. 921-934, Dec. 2019.

[31] Y. Yoshida and T. Miyato, "Spectral norm regularization for improving the generalizability of deep learning," 2017, arXiv:1705.10941.

[32] Z.-H. Liu, B.-L. Lu, H.-L. Wei, L. Chen, X.-H. Li, and M. Rätsch, "Deep adversarial domain adaptation model for bearing fault diagnosis," *IEEE Trans. Syst. Man Cybern. Syst.,* vol. 51, no. 7, pp. 4217-4226, Jul. 2021.

[33] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.,* vol. 63, no. 5, pp. 3137-3147, May 2016.

[34] Q. Jiang, X. Yan, and B. Huang, "Deep discriminative representation learning for nonlinear process fault detection," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 3, pp. 1410-1419, Jul. 2020.

[35] R. Huang, J. Li, Y. Liao, J. Chen, Z. Wang, and W. Li, "Deep adversarial capsule network for compound fault diagnosis of machinery toward multidomain generalization task," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1-11, Art no. 3506311, 2021.

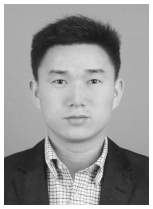
[36] K. Loparo. *Case western reserve university bearing data center,* (2013), [online]. Available: http://csegroups.case.edu/bearingdatacenter/pages/12k-drive-end-bearing-fault-data.

[37] W. Lu, B. Liang, Y. Cheng, D. Meng, and T. Zhang, "Deep model based domain adaptation for fault diagnosis," *IEEE Trans. Ind. Electron.,* vol. 64, no. 3, pp. 2296-2305, Mar. 2017.

[38] J. Jiao, J. Lin, M. Zhao, and K. Liang, "Double-level adversarial domain adaptation network for intelligent fault diagnosis," *Knowl.-Based Syst.*, vol. 205, Art no. 106236, 2020.

[39] F. Li, T. Tang, B. Tang, and Q. He, "Deep convolution domain-adversarial transfer learning for fault diagnosis of rolling bearings," *Measurement*, vol. 169, Art. no. 108339, 2021.

[40] V. D. M. Laurens and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.,* vol. 9, no. 2605, pp. 2579-2605, 2008.



Zhao-Hua Liu (*M'16, SM'2022*) received M.Sc. degree in computer science and engineering, and the PhD. degree in automatic control and electrical engineering from the Hunan University, China, in 2010 and 2012, respectively. He worked as a visiting researcher in the Department of Automatic Control and Systems Engineering at the University of Sheffield, United Kingdom, from 2015 to 2016.

He is currently a Professor of Automatic Control and Systems with the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, China. His current research interests include intelligent information processing and control of wind turbines, computational intelligence and learning algorithms design, machine learning aided fault diagnosis and prognosis, parameter estimation and intelligent control of permanent-magnet synchronous machine drives, and fault diagnosis based intelligent operations and maintenance for electric power equipment.
Dr. Liu has published more than 50 research papers in refereed journals and conferences, including IEEE TRANSACTIONS / JOURNAL / MAGAZINE. He is a regular reviewer for several international journals and conferences.

Liang Chen received the B.Eng. degree in automation from the Henan University, Kaifeng, China, in 2018. He received M.Sc. degree in automatic control and electrical engineering from Hunan University of Science and Technology, Xiangtan, China ,in 2022.

His current research interests include transfer learning, deep learning algorithm design and fault diagnosis of wind turbine.

Hua-Liang Wei received the Ph.D. degree in automatic control from the University of Sheffield, Sheffield, U.K., in 2004.

He is currently a senior lecturer with the Department of Automatic Control and Systems Engineering, the University of Sheffield, Sheffield, UK. His research focuses on identification and modelling for complex nonlinear systems, model and parameter estimation, signal processing, feature engineering, pattern recognition and classification, and interpretable machine learning, among others.

Ying Zhang (S'16-M'20) received the Ph.D. degree in computer science and technology from College of Computer Science and Electronic Engineering, Hunan University (HNU), Changsha, China, in 2020.

He is currently an Associate Professor with the School of Computer Science at Northwestern Polytechnical University, Xi'an, China. He was a Visiting Scholar with the Department of Electrical and Computer Engineering, Technische Universität Dresden (TUD), Dresden, Germany, from February 2018 to March 2018, as well as with the University of Michigan, Dearborn, MI, USA, from October 2018 to October 2019. His research interests are in the areas of parameter/state estimation, information fusion, situation awareness, modeling, decision-making, energy-aware optimization and intelligent control with applications to electric, connected and autonomous vehicles (e-CAVs), intelligent transportation system (ITS), unmanned aerial vehicles (UAVs) and other autonomous systems. Dr. Zhang has authored or coauthored more than 15 papers in IEEE journals and conferences.

Lei Chen received M.S. degree in computer science and engineering, and the PhD. degree in automatic control and electrical engineering from the Hunan University, China, in 2012 and 2017, respectively.

He is currently a Lecturer with the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, China. His current research interests include deep learning, network representation learning, information security of industrial control system and industrial big data analysis.

Ming-Yang Lv received the Ph.D. degree in control theory and engineering with Hunan University, Changsha, China, in 2020.

He is currently a Lecturer with the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, China. His research interests include chaos theory and application, and modeling for complex process industries based on machine learning and deep learning.