

Received June 6, 2020, accepted June 22, 2020, date of publication June 26, 2020, date of current version July 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3005243

Transfer Learning Method Based on Adversarial Domain Adaption for Bearing Fault Diagnosis

JIAJIE SHAO¹, ZHIWEN HUANG¹, AND JIANMIN ZHU¹

College of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

Corresponding author: Jianmin Zhu (jmzhu_usst@163.com)

This work was supported by the National Natural Science Foundation of China under Grant 51375289 and Grant 51775323.

ABSTRACT At present, most of the intelligent fault diagnosis methods of rolling element bearings require sufficient labeled data for training. However, collecting labeled data is usually expensive and time-consuming, and when the distribution of the test data is different from the distribution of the training data, the diagnostic performance will decrease. In order to solve the problem of unlabeled cross-domain diagnosis of bearings, this paper proposes an adversarial domain adaption method based on deep transfer learning. The short-time Fourier transform is used to transform the original data into a time-frequency image. The feature extractor is used to extract its deep features. The maximum mean discrepancy and domain confusion function are used for domain adaptation to extract domain-invariant features between two domains for cross-domain fault diagnosis. Experiments on two bearing datasets are carried out for validations. The results prove that the method in this paper is superior to other deep transfer learning methods. It shows the advantages of the improved method and can be used as an effective tool for cross-domain fault diagnosis.

INDEX TERMS Transfer learning, fault diagnosis, domain adaption, deep learning.

I. INTRODUCTION

Rolling bearings are widely used in machinery and equipment. With the development of modern industrial information and intelligence, emerging data-driven intelligent fault diagnosis methods have shown great advantages in processing complex mechanical signals [1]. A lot of research has been done on the technology of intelligent bearing fault diagnosis [2], [3]. In addition, the use of deep learning for bearing fault diagnosis is also emerging [4]–[6]. But these methods, like the traditional deep learning and machine learning, need to meet the following conditions: (1) the training and testing data need to follow the same distribution; (2) enough labeled data are needed to train the model and each task needs to be modeled separately. When the distribution is different or there is not enough data labels, the performance of these methods may drop sharply. With the increasing amount of data and data types, labeled data is difficult to obtain from some machines, but the device can accumulate a large amount of unlabeled data during long-term operation. Manually tagging the data or building each model from scratch takes time and effort.

The associate editor coordinating the review of this manuscript and approving it for publication was Tyson Brooks¹.

To solve this problem, transfer learning is a very effective method. Transfer learning is the application of the knowledge or patterns learned from a certain field or task to different but related domains or problems. It is used to solve the problems of scarce labeled data and different domain distribution of the target task, which is very prominent in the fields of natural language processing [7], computer vision [8], medical health and bioinformatics [9]. At present, applying transfer learning methods into fault diagnosis has become a research hotspot [10], [11]. Fine-tune is the most common method. Shao *et al.* [12] and Cao *et al.* [13] use the pre-trained deep convolution network to solve the problem that only a small amount of labeled data in training. They frozen the weights of some layers and fine-tuning the weights of the remaining layers, and a more accurate model is obtained. But fine-tune still can't solve the problem of unlabeled data.

In order to solve the problem of cross-domain fault diagnosis, several transfer learning based intelligent fault diagnosis methods have been proposed. Zhang *et al.* [14] transfer the learned knowledge from shallow models to pseudo labels and train the deep neural model for better generalization. Wen *et al.* [15] proposed a deep transfer learning method based on auto-encoder for fault diagnosis and tested on bearing datasets under different loading conditions. Xie *et al.* [16]

proposed a transfer component analysis based cross-domain feature fusion method for gearbox fault diagnosis under various operation conditions.

Specifically, domain adaption methods have received much attention in cross-domain fault diagnostic tasks. Most deep domain adaption methods use deep network to extract fault signal features and reduce the distance of domain distribution through adaptive layer. Lu *et al.* [17] minimized the maximum mean discrepancy between the two domains based on DNN network and added regularization to reduce the domain distribution and realized the health status recognition of rolling bearing under different working conditions. Li *et al.* used deep convolutional neural networks as the main architecture, minimized the multiple-kernel maximum mean discrepancy between two domains at the adaptive layer in [18], verified the effectiveness of the method under different load conditions. In [19], they proposed a deep generation neural network for fault diagnosis under different load conditions, which minimizes the maximum mean discrepancy between real source data and artificially generated false target data for cross-domain diagnosis. Zhang *et al.* [20] proposed a DACNN network that used source domain data to pre-train the source domain feature extractor, and fine-tune the parameters of untied adaptive layers of the target feature extractor during the back-propagation process. The accuracy of the method was verified on the bearings and gearboxes with different loads, but the accuracy and training speed of these methods are not high enough.

Nowadays, adversarial learning is promisingly used in domain adaptation, and there are few related researches. Guo *et al.* [21] and Li *et al.* [22] both optimized the feature extractor through the gradient reversal layer, directly maximizing the discriminator loss so that the domain discriminator could not distinguish the data of the two domains. This optimization corresponds to the true objective for generative adversarial networks [23], but early on during training the discriminator converges quickly, causing the gradient to vanish [24].

In order to further improve the diagnostic accuracy and training speed, and currently there is less attention on the cross-domain diagnosis of different position sensors, this paper proposes a deep adversarial network for cross-domain fault diagnosis. The proposed method starts from the idea of adversarial learning, uses the time-frequency image as the input, and uses the 50-layer deep residual network pre-trained on ImageNet as a feature extractor. Using domain confusion and multiple-kernel maximum mean discrepancy loss to minimize the domain disparity of the source and target data. We use two bearing datasets to verify the proposed method, cross-domain diagnosis of fault signals under different loads and different position sensors is used as the task. The experiments prove that the method proposed in this paper has faster training speed, higher accuracy and more reliable diagnostic results.

The main insights and contributions of this paper are summarized as follows.

1) A novel end-to-end cross-domain bearing fault diagnosis method is developed based on deep adversarial domain adaption framework by using raw signals of sensors. Significantly, unlike traditional methods, the proposed method combines adversarial learning and Multiple-kernel Maximum Mean Discrepancy, which can effectively extract domain-invariant features between two domains for cross-domain fault diagnosis and improve the ability to reduce domain discrepancy with unlabeled target data more effectively.

2) The 50-layer deep residual network pre-trained on ImageNet can accelerate the convergence on the target task, improve the accuracy and reduce overfitting. In this paper, the proposed method manages to provide reliable cross-domain diagnosis results and this exploration would promote the practical application of intelligent fault diagnosis.

The rest of the paper is organized as follows. Section II introduces the problems and processes to be solved. Section III gives the cross domain fault diagnosis method, and experimentally validated and investigated in Section IV. Finally, conclusions for this literature is made in Section V.

II. PRELIMINARIES

A. PROBLEM FORMULATION

In order to explain the problem to be solved, first we describe the problem with the terminology of transfer learning. There are two basic concepts of transfer learning: Domain and Task. Domain Δ consists of d -dimensional feature space \mathcal{X} and a marginal probability distribution $P(x)$, where $\Delta = \{\mathcal{X}, P(x)\}$, $x \in \mathcal{X}$. Given labeled data under one working condition as the source domain $\mathcal{D}_s = \{\mathcal{X}^s, P(x^s)\}$ and unlabeled data under other working condition as the target domain $\mathcal{D}_t = \{\mathcal{X}^t, P(x^t)\}$. Since the working conditions of the source and target domain are different, the data distribution in the two domains is also different, that is $P(x^s) \neq P(x^t)$.

The task \mathcal{T} is the goal of fault diagnosis learning. It consists of labels space Ψ and a prediction function $f(x)$, where $\mathcal{T} = \{\mathcal{Y}, f(x)\}$, $f(x) = P(y|x)$ is the conditional probability distribution, and $y \in \mathcal{Y}$. Since categories of different domain data are the same, it is assumed that their label spaces are the same, that is $\mathcal{Y}^s = \mathcal{Y}^t$, and their conditional probability distributions are also the same, that is $P(y^s|x^s) = P(y^t|x^t)$.

Through the training of source data samples, a non-linear mapping relationship from the feature space of source data \mathcal{X}^s to label space \mathcal{Y}^s is established: $C_s : \mathcal{X}^s \rightarrow \mathcal{Y}^s$, which is the acquired fault diagnosis knowledge. As shown in Fig. 1, due to the large distribution discrepancy between the data in the source domain and the target domain, the fault diagnosis knowledge learned from the source domain cannot accurately identify the category of the target domain. Aiming at this problem, this paper aims to construct a deep adversarial domain adaption network, adapting the distribution of data in the source and target domains, so that the fault diagnosis knowledge in the source domain can identify the status of the data from the target domain.

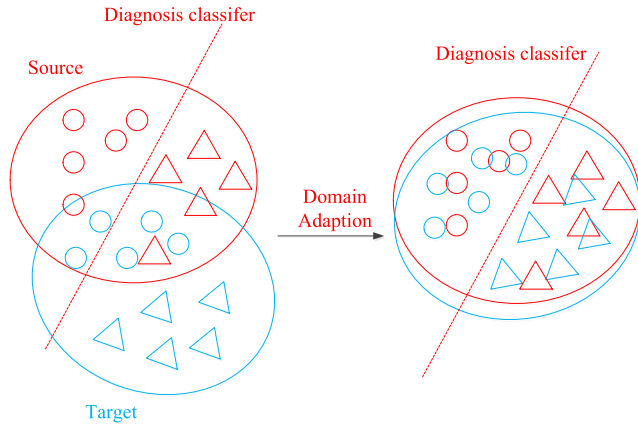


FIGURE 1. Fault diagnosis domain adaption diagram. The triangles and circles denote two different classes.

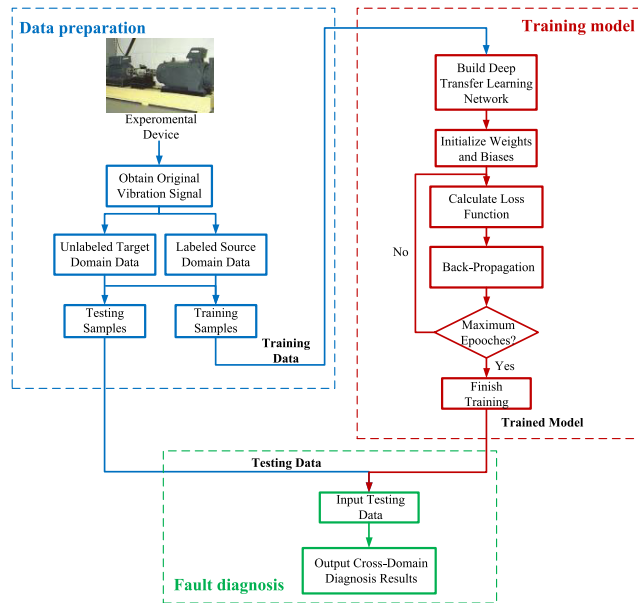


FIGURE 2. Cross-domain fault diagnosis flowchart.

B. DIAGNOSTIC PROCESS

This paper presents a flow chart of cross-domain diagnosis of rolling bearing, as shown in Fig. 2. First, the sensor is used to collect the original mechanical vibration signals, and the labeled source domain data and unlabeled target domain data are used as training samples. Then establish a suitable deep transfer learning network, and use the training samples as input to train the model. Finally, the unlabeled target domain data is used as testing samples to input the trained model, and cross-domain diagnosis results are obtained.

The performance of the model affects the results of cross-domain fault diagnosis. How to design a suitable deep transfer network is an important issue. The architecture of the network, initialization weights, and the domain adaptive loss function are very important. In this paper, we propose a deep adversarial domain adaption framework, the specific architecture is shown in Section III.

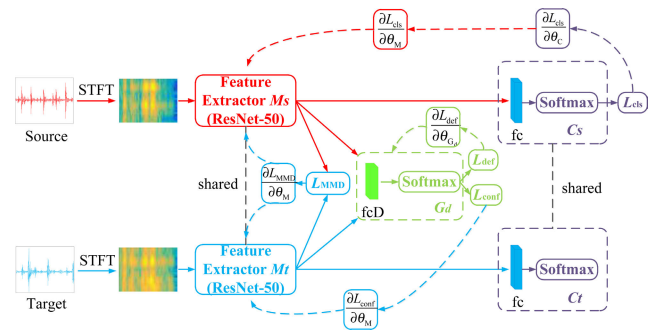


FIGURE 3. Proposed network architecture.

III. PROPOSED METHOD

A. NETWORK ARCHITECTURE

Aiming at the problem of cross-domain diagnosis of rolling bearing, this paper proposes a deep adversarial domain adaption network architecture, which consists of a feature extractor, a domain discriminator, and a health condition classifier, as shown in Fig. 3. The vibration signal is a non-stationary time-varying signal as a whole. Instead of focusing on time or frequency domain features alone, time-frequency analysis techniques are effective to investigate non-stationary signals [25], and they are promising on improving prognostic results [26]. As one of the most commonly used time-frequency analysis techniques, short-time Fourier transform (STFT) is widely used in vibration signal processing. Therefore, this paper uses STFT to visualize the signal collected by the sensor, and these images are input into the network to train the model.

The feature extractor M is composed of a 50-layer deep residual network (ResNet-50) [27], and pre-trains on ImageNet. The advantage of ImageNet dataset is high quality data annotation and it enable the model to learn the features that can be extended to other tasks in the problem domain. The deep residual network introduces the concept of residual unit, which is outstanding in image classification. The feature extractor M is used to extract deep features of the source and target domain, and the weights of the feature extractors of the source and target domain are shared.

The health condition classifier C is composed of one fully-connected layer and a softmax function. The features extracted by the feature extractor are used as input, and the softmax function is used to diagnose machine health conditions. The source domain and target domain classifiers also share weights.

Furthermore, there is a domain discriminator G_d for adversarial training, which is used to determine whether the input deep feature comes from source domain or target domain, and also consists of one fully-connected layer and a softmax function.

B. OPTIMIZATION OBJECTIVE

In order to complete the cross-domain diagnosis, the deep adversarial domain adaption network proposed in this paper has the following optimization objectives:

1) Generally, it is necessary to be able to identify the health conditions of the labeled source domain bearings. Therefore, the first optimization objective is to minimize the supervised classification loss of the source domain data and the cross-entropy loss is used in this study. Assume that there are K types of fault categories in the source domain samples. The source domain images and labels ($x_s, y_s \in \{1, \dots, K\}$) are used to train the source domain feature extractor M_s and health condition classifier C_s . The loss function is defined as follows,

$$\min_{M_s, C_s} \mathcal{L}_{cls}(\mathcal{D}_s, \mathcal{Y}_s) = -\mathbb{E}_{(x_s, y_s) \sim (\mathcal{D}_s, \mathcal{Y}_s)} \sum_{k=1}^K 1_{[k=y_s]} \log C_s(M_s(x_s)) \quad (1)$$

where $\mathbb{E}_{(x_s, y_s) \sim (\mathcal{D}_s, \mathcal{Y}_s)}$ means extracting an instance from the source domain distribution, and $1_{[k=y_s]}$ is an indicator function, if $k = y_s$, its value is 1, otherwise it is 0.

2) The domain discriminator G_d is used to classify whether the features come from the source domain or target domain. The second optimization objective is to minimize the domain discriminator loss of the two domains. Therefore, it is also optimized according to a standard supervised loss, where the labels indicate the origin domain, and also uses the cross-entropy loss function, defined below:

$$\min_{G_d} \mathcal{L}_{def}(\mathcal{D}_s, \mathcal{D}_t, M_s, M_t) = -\mathbb{E}_{x_s \sim \mathcal{D}_s} [\log G_d(M_s(x_s))] - \mathbb{E}_{x_t \sim \mathcal{D}_t} [\log(1 - G_d(M_t(x_t)))] \quad (2)$$

3) The domain discriminator is connected to the feature extractor to determine whether the extracted features come from source domain or target domain. If the domain discriminator cannot distinguish the features, these features are considered to be domain invariant. The usual method is to introduce the gradient reversal layer with reference to the generative adversarial network [23] to maximize the discriminator loss. In this paper, domain confusion loss [28] is introduced as a means to learn a representation that is domain invariant. By computing the cross entropy between the output predicted domain labels and a uniform distribution over domain labels, two domains are maximally confused to reduce the disparity of marginal probability distribution between them. Therefore, the third optimization objective is to minimize the domain confusion loss of the two domains. The used loss function is as follows:

$$\min_{M_s, M_t} \mathcal{L}_{conf}(\mathcal{D}_s, \mathcal{D}_t, G_d) = - \sum_{j \in \{s, t\}} \mathbb{E}_{x_j \sim \mathcal{D}_j} [\frac{1}{2} \log G_d(M_j(x_j)) + \frac{1}{2} \log(1 - G_d(M_j(x_j)))] \quad (3)$$

Ideally, we want to minimize (2) and (3) at the same time. However, the two losses are opposite: learning a good feature extractor means that the domain discriminator must do

poorly, and learning an effective domain discriminator means that the representation is not domain invariant. Rather than optimizing the parameters globally, we instead perform iterative updates for the two objectives given the fixed parameters from the previous iteration.

4) The extracted deep features affect the validity of the fault diagnosis. In order to reduce the distribution discrepancy between the two domains, the discrepancy is measured after the feature extractor. Many criteria can be used to estimate the discrepancy between distributions. Maximum Mean Discrepancy (MMD) [29] is the most frequently used non-parametric method in transfer learning to measure the distribution discrepancy between two domains. It measures the distance between two distributions in the reproducible kernel Hilbert space \mathcal{H}_k (RKHS) with a characteristic kernel k . It is a kernel learning method that maps the original variables into the RKHS space. The squared formulation of MMD is defined as:

$$d_k^2(\mathcal{D}_s, \mathcal{D}_t) = \left\| \frac{1}{m} \sum_{i=1}^m \phi(x_i^s) - \frac{1}{n} \sum_{j=1}^n \phi(x_j^t) \right\|_{\mathcal{H}_k}^2 \quad (4)$$

where $\phi(\cdot)$ is the nonlinear mapping from the original feature space to RKHS. The inner product in the RKHS space can be converted into a kernel function, so MMD can be calculated directly from the kernel function:

$$d_k^2(\mathcal{D}_s, \mathcal{D}_t) = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m k(x_i^s, x_j^s) - \frac{2}{mn} \sum_{i=1}^m \sum_{j=1}^n k(x_i^s, x_j^t) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n k(x_i^t, x_j^t) \quad (5)$$

where $k(\cdot, \cdot)$ is the characteristic kernel. In this paper, the characteristic kernel uses a Gaussian kernel function $k(u, v) = e^{-\frac{\|u-v\|^2}{2\sigma^2}}$. Multiple-kernel MMD (MK-MMD) assumes that the optimal kernel can be obtained linearly from multiple kernels, and the characteristic kernels associated

with the feature map ϕ , $k(x^s, x^t) = \langle \phi(x^s), \phi(x^t) \rangle$ is defined as the convex combinations of N_k kernels $\{k_u\}$ [30]:

$$\mathcal{K}\{k = \sum_{i=1}^{N_k} \beta_u k_u : \sum_{i=1}^{N_k} \beta_u = 1, \beta_u \geq 0, \forall u\} \quad (6)$$

where the constraints on the coefficients $\{\beta_u\}$ are imposed to guarantee that the derived multi-kernel k is characteristic. The multi-kernel k can leverage different kernels to enhance MK-MMD test, so as to a principled method for optimal kernel selection.

Therefore, the fourth optimization objective is to minimize the MK-MMD loss of the two domains. The features are mapped from the feature space to the RKHS space to minimize the distance between the two types of data. Define the loss function as:

$$\min_{M_s, M_t} \max_{k \in \mathcal{K}} \mathcal{L}_{MMD}(\mathcal{D}_s, \mathcal{D}_t) = d_k^2(\mathcal{D}_s, \mathcal{D}_t) \quad (7)$$

Since the parameters of the feature extractor and health condition classifier of the source and target domains are shared, let θ_M , θ_C and θ_G be the parameters of the feature extractor, health condition classifier and domain discriminator. Combining the above optimization objectives, the final optimization objective can be written as:

$$\mathcal{L}_{fea} = \min_{\theta_M, \theta_C} \mathcal{L}_{cls} + \lambda \mathcal{L}_{conf} + \gamma \mathcal{L}_{MMD} \quad (8)$$

$$\mathcal{L}_{critic} = \min_{\theta_G} \mathcal{L}_{def} \quad (9)$$

where λ and γ are the penalty coefficients for \mathcal{L}_{conf} and \mathcal{L}_{MMD} .

Based on the above loss function, training is performed using stochastic gradient descent (SGD) algorithm. In each training epoch, the parameters are updated as follows:

$$\theta_M \leftarrow \theta_M - \eta \left(\frac{\partial \mathcal{L}_{cls}}{\partial \theta_M} + \lambda \frac{\partial \mathcal{L}_{conf}}{\partial \theta_M} + \gamma \frac{\partial \mathcal{L}_{MMD}}{\partial \theta_M} \right) \quad (10)$$

$$\theta_C \leftarrow \theta_C - \eta \frac{\partial \mathcal{L}_{cls}}{\partial \theta_C} \quad (11)$$

$$\theta_G \leftarrow \theta_G - \eta \frac{\partial \mathcal{L}_{def}}{\partial \theta_G} \quad (12)$$

where η is the learning rate.

In particular, training the entire network from scratch requires a large amount of data and a long number of epochs, which is not possible for many practical scenarios. Therefore, we use the pre-trained model on ImageNet and fine-tune its parameters during the training process. Since the classifier is trained from scratch, we set its learning rate to 10 times that of the other layers. We use mini-batch stochastic gradient descent (SGD) with momentum of 0.9 and an learning rate annealing strategy in [31]. When training is complete, the trained model can be used to diagnose unlabeled target domains.

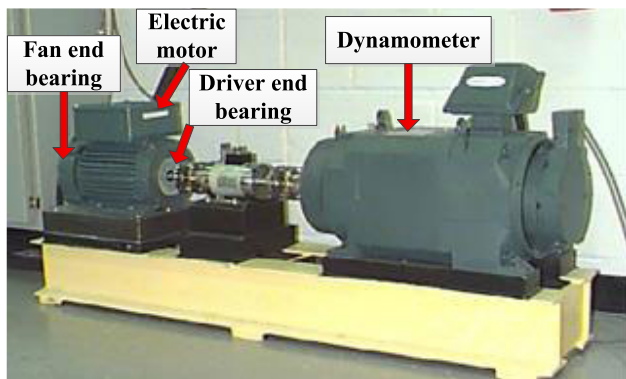


FIGURE 4. Experiment equipment in CWRU dataset.

IV. EXPERIMENTAL STUDY

A. DATASET

1) CWRU DATASET

The dataset used in this study is provided by the Bearing Data Center of the Case Western Reserve University [32]. Experiment equipment is shown in Fig. 4. Using the data

TABLE 1. Information of CWRU dataset.

Dataset	Label	0	1	2	3	4	5	6	7	8	9
CWRU	Type	Nor	OF	OF	OF	IF	IF	IF	BF	BF	BF
	Size(mil)	0	7	14	21	7	14	21	7	14	21

collected by the acceleration sensor on the driver end and the fan end, the health status is divided into 4 types: (1) healthy (Nor); (2) outer race fault (OF); (3) inner race fault (IF); (4) ball fault (BF). The three types of faults are also divided into different fault diameters of 7 mils, 14 mils, and 21 mils, for a total of 10 bearing conditions, as shown in Table 1. In addition, the data was collected under four load scenarios (0, 1, 2, 3hp), and the sampling frequency was 12kHz.

a: TRANSFER UNDER DIFFERENT LOADS

Working under different loads will result in inconsistent distribution of vibration data. In order to verify the accuracy of the improved method, the improved method is used to transfer under different load scenarios. Here we use the data from the driver end. There are 12 transfer tasks, i.e., $T_{01}, T_{02}, T_{03}, T_{10}, T_{12}, T_{13}, T_{20}, T_{21}, T_{23}, T_{30}, T_{31}$ and T_{32} , where T_{uv} denotes the scenario that the operating condition with u hp load is considered the source domain, and that with v hp load is the target domain. The source domain data is labeled and the target domain data is unlabeled.

b: TRANSFER UNDER DIFFERENT SENSOR POSITIONS

The data collected by sensors at different positions usually also have different features. We use the data collected by the sensors at the driver end and fan end under the same load to realize cross-domain diagnosis. There are 8 transfer tasks, i.e., $T_{0DF}, T_{0FD}, T_{1DF}, T_{1FD}, T_{2DF}, T_{2FD}, T_{3DF}$ and T_{3FD} , where T_{uDF} denotes the scenario that the data collected by the driver end sensor under the condition of u hp as the source domain, and the data collected by the fan end as the target domain for cross-domain diagnosis.

Short-time Fourier transform is performed on the vibration signal to obtain Time-frequency imaging. The number of signal sampling points is 1024, and the number of sampling points per frame is 256. Source and target domains take 200 time-frequency images for each type, for a total of 2000 images.

2) DDS BEARING DATASET

The DDS dataset was collected from the drivetrain dynamic simulator (DDS) [12]. This dataset contains 2 subdatasets, including bearing data and gear data, and we use the bearing data here. The different types of faults for bearings are shown in Table 2. The faults of bearing are investigated under two different operating conditions where rotating speed and load configuration are set as 20HZ-0V and 30HZ-2V. The improved method is used to transfer under these two working conditions, so there are two transfer tasks. Source and target domains take 200 images for each type, for a total of 1000 images.

TABLE 2. Information of DDS bearing dataset.

Dataset	Type	Description
DDS bearing	Normal	Health working state
	Ball	Crack occurs in the ball
	Inner	Crack occurs in the inner ring
	Outer	Crack occurs in the outer ring
	Combination	Crack occurs in the both inner and outer ring

B. COMPARED APPROACHES AND IMPLEMENTATION DETAILS

We compare the proposed improved method with the following methods, including Deep Domain Confusion (DDC) [33], Deep Adaptation Network (DAN) [30], Deep CORAL (D-CORAL) [34] and Reverse Gradient (RevGrad) [35].

We use labeled source examples as the source domain and unlabeled target examples as the target domain. Following standard evaluation protocols for unsupervised domain adaptation, the average classification accuracy and standard error over three random trials are reported for comparison. For MMD-based methods (DDC, DAN), we adopt Gaussian kernel with bandwidth set to median pairwise squared distances on the training data. All methods use pre-trained ResNet-50 as a feature extractor to extract deep features, based on the pytorch framework, and fine-tune from pytorch-provided models of ResNet.

In order to improve the optimization of SGD during training, the learning rate is using the following formula: $\eta = \eta_0 / (1 + \alpha p)^\beta$, where p is the training progress linearly changing from 0 to 1, $\eta_0 = 0.005$, $\alpha = 10$, $\beta = 0.75$. The weight decay in SGD is 5×10^{-4} , and momentum is 0.9. We fine-tune all convolutional and pooling layers and train the classifier layer via back propagation. Since the classifier and domain discriminator are trained from scratch, we set their learning rate to be 10 times that of the other layers. In [35], in order to suppress noisy activations early in training at the early stages of training, the penalty coefficients are gradually changed from 0 to 1 using the formula $2/e^{-10p} - 1$ instead of fixing it. Taking T_{30} as an example, the experiments are shown in Fig. 5, we set the penalty coefficients $\lambda = 0.5$ and $\gamma = 2/e^{-10p} - 1$. This progressive strategy significantly improves the performance of the model and simplify the choice of parameters.

C. RESULTS AND ANALYSIS

All experiments were performed on the SothisAI cloud platform using 32GB RAM and NVIDIA Tesla V100 GPU.

1) CWRU DATASET

(1) Transfer under different loads

The results of cross-domain diagnosis under different loads are shown in Table 3. All the 5 deep transfer learning methods we use can achieve an average accuracy of more than 90%. The proposed method in this paper outperforms other methods in the 12 transfer tasks, all the cross-domain fault diagnosis accuracies are over 98% and the final average

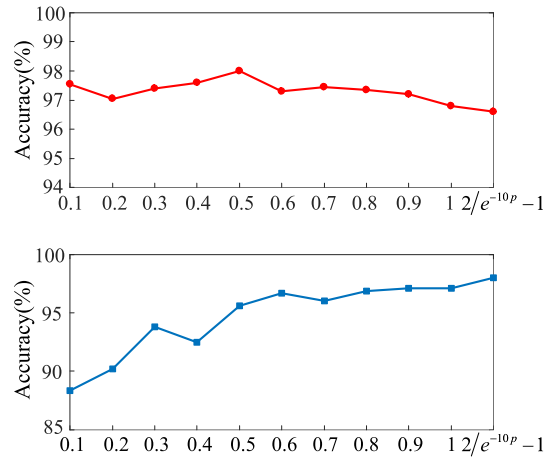


FIGURE 5. The performance comparison under different penalty coefficients.

accuracy rate is 99.2%, which proves the effectiveness of the proposed method. The same transfer task is used in [19]. Its average accuracy is 95.49%. The study in [17] is similar with the T_{03} transfer task. Four health conditions are considered, and the accuracy is 94.73%. The proposed method in this paper are better than them.

DDC and DAN minimize MMD loss and MK-MMD loss of the adaptive layer between two domains. These two methods can achieve good results when the distribution discrepancy is not large, but the performance goes down sharply when the distribution discrepancy is large. RevGrad is to maximize domain discriminator loss to reduce the distribution discrepancy. Deep Coral minimizes CORAL loss between two domains. The accuracy rate is slightly higher than the previous two methods, but it is still insufficient.

The proposed method can achieve good results when the distribution discrepancy is small, and the accuracy rate is much higher than other methods when the distribution difference becomes larger.

At the same time, we take the transfer task from 3hp to 0hp as an example and use the proposed method to compare the results of using a pre-trained feature extractor with using an unpre-trained feature extractor, as shown in Fig. 6 (a). The pre-trained network converges faster, has smaller oscillations, and improves accuracy significantly. Then we compared the results of minimizing domain confusion loss with maximizing domain discriminator loss, as shown in Fig. 6 (b). The results show that the domain confusion loss function used in this paper has higher accuracy.

In order to intuitively validate the effectiveness of the proposed method, we use the t-distributed stochastic neighbor embedding (t-SNE) [36] technique to map the high-dimensional features into a two-dimensional space. Using the task T_{30} as an example, and the results are shown in Fig. 7. It can be seen from the figure that the DDC, DAN, and D-Coral methods have obvious distribution discrepancy between the source and target domains. The classification

TABLE 3. Classification results (%) for CWRU dataset under different loads.

Method	T_{01}	T_{02}	T_{03}	T_{10}	T_{12}	T_{13}	T_{20}	T_{21}	T_{23}	T_{30}	T_{31}	T_{32}	Avg
DDC	98.3	96.8	88.9	95.9	99.8	98.2	91.9	97.8	99.6	77.9	93.4	99.3	94.8
DAN	99.1	98.4	89.5	96.2	98.6	96.5	91.1	97.0	98.9	83.4	93.1	98.1	95.0
RevGrad	98.8	96.3	86.6	96.9	99.6	98.0	92.5	97.1	99.8	88.1	93.5	99.7	95.6
D-Coral	97.9	97.9	94.4	97.2	99.6	98.6	95.7	97.6	99.4	88.8	96.5	97.2	96.7
proposed	99.4	99.7	98.4	99.1	99.9	99.4	99.3	98.5	99.8	98.0	98.7	99.9	99.2

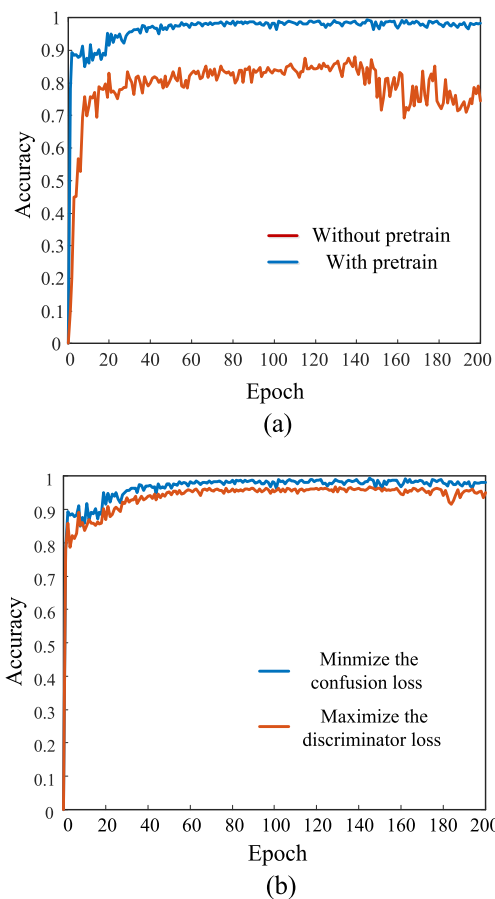


FIGURE 6. Comparison of results of different methods. (a) Comparison of whether to pre-train. (b) Comparison of minimizing confusion loss and maximizing discriminator loss.

model learned in the source domain cannot achieve good results in the target domain. The proposed method in this paper has a significantly closer distribution of the source and target domains, which can achieve better clustering results and increase the distance between classes. It is shown that the proposed method in this paper can shorten the distribution discrepancy between the source and target domain, and have significant effects on fault diagnosis under different loads. It can intuitively explain the effectiveness of the improved method.

a: TRANSFER UNDER DIFFERENT SENSOR POSITIONS

The results of cross-domain diagnosis under different position sensors are shown in Table 4. It can be seen that the

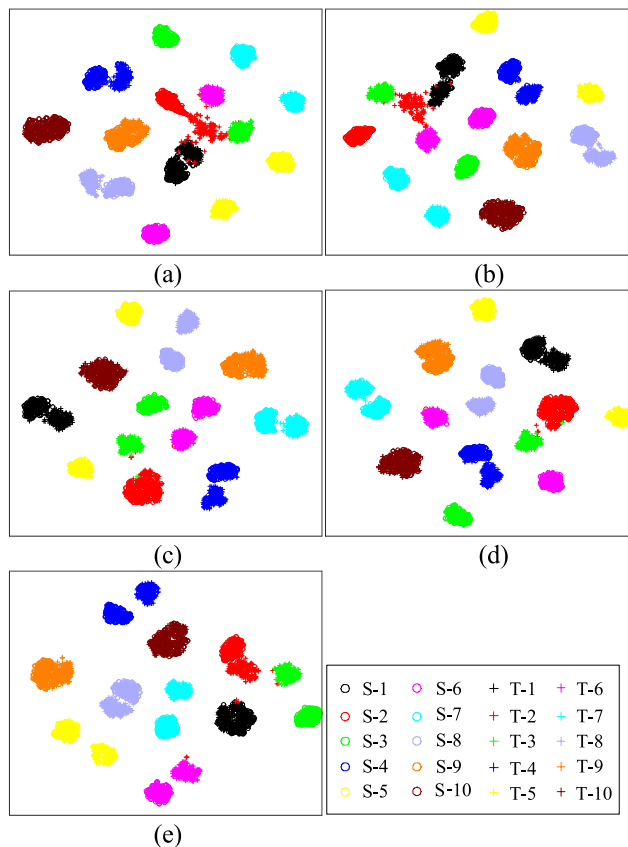


FIGURE 7. Using t-SNE to visualize features under different loads. (a) DDC, (b) DAN, (c) ReGrad, (d) D-Coral, (e) proposed.

TABLE 4. Classification results (%) for CWRU dataset under different sensor positions.

Method	T_{0DF}	T_{0FD}	T_{1DF}	T_{1FD}	T_{2DF}	T_{2FD}	T_{3DF}	T_{3FD}	Avg
DDC	97.4	79.8	92.7	86.0	98.5	89.7	99.3	86.7	91.3
DAN	93.9	81.4	93.5	89.0	97.8	91.7	97.6	86.3	91.4
RevGrad	99.7	81.3	95.1	86.4	99.9	93.0	99.9	87.0	92.8
D-Coral	98.6	77.2	94.5	86.8	99.8	92.0	94.6	81.0	90.6
proposed	99.9	97.6	98.9	97.0	99.9	98.3	99.9	97.3	98.6

accuracy of all transfer tasks of the proposed method is higher than 97%, and some tasks even reach 99.9%. The final average accuracy is 98.6%, which is better than other methods. Especially for the transfer task from the fan end to the drive end, the accuracy of other methods is significantly lower than that of the task from the drive end to the fan end under the same load. It shows that the proposed method has

better cross-domain diagnosis ability, and the cross-domain diagnosis task under different position sensors still has very good performance, further verifying the effectiveness and superiority of the method.

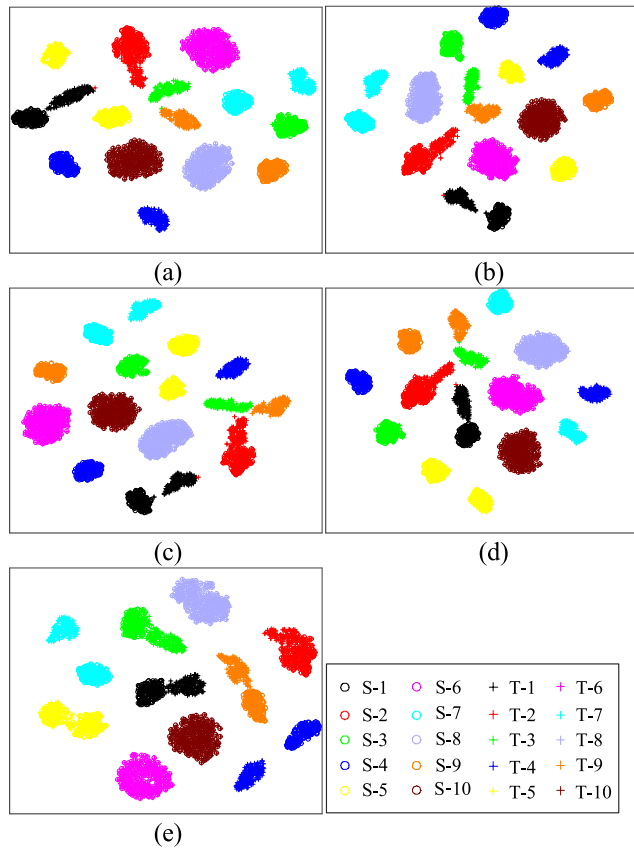


FIGURE 8. Using t-SNE to visualize features under different sensor positions. (a) DDC, (b) DAN, (c) RevGrad, (d) D-Coral, (e) proposed.

Taking the T_{0FD} task as an example, the visualization using the t-SNE technique is shown in Fig. 8. It can be clearly seen that the proposed method increases the distance between classes and reduces the distribution discrepancy. In this way, the trained cross-domain diagnosis model can improve the recognition of data samples, and other methods have obvious distribution discrepancy between different domains. It is intuitive to see the effectiveness of the proposed method for fault diagnosis under different sensor positions.

TABLE 5. Classification results (%) for DDS bearing dataset.

Transfer Task	DDC	DAN	RevGrad	D-Coral	proposed
20-0 → 30-2	72.6	76.4	75.7	79.7	87.8
30-2 → 20-0	72.8	74.1	74.8	76.0	84.5
Avg	72.7	75.3	75.3	77.9	86.2

2) DDS BEARING DATASET

The experimental results are shown in Table 5. We can observe that the proposed method is still able to achieve the best performance compared with other transfer learning methods in all transfer tasks. It can further validates the

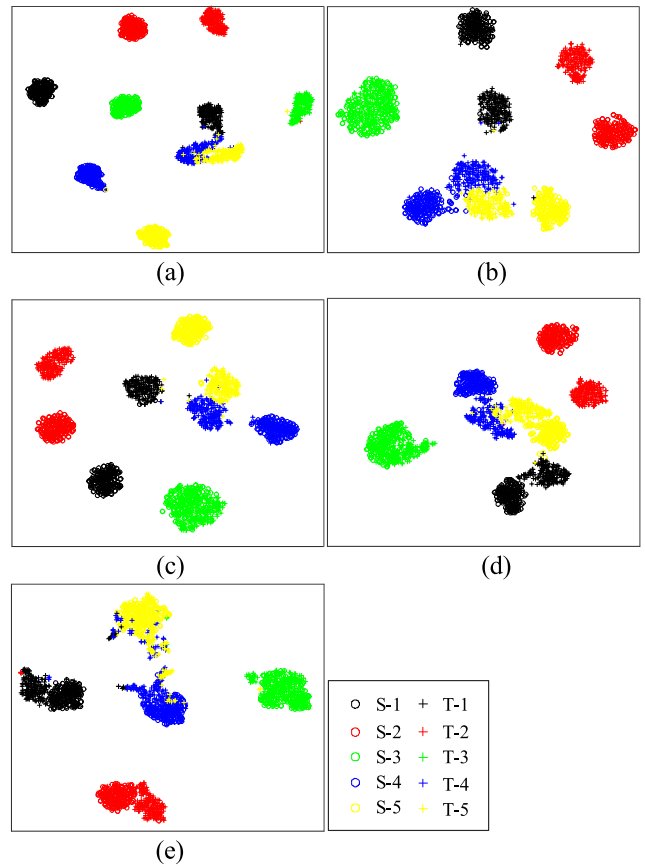


FIGURE 9. Using t-SNE to visualize features for DDS bearing dataset. (a) DDC, (b) DAN, (c) RevGrad, (d) D-Coral, (e) proposed.

effectiveness and superiority of the proposed method. Also, the task 20HZ-0V → 30HZ-2V is taken as an example to visualize feature distributions, as shown in Fig. 9. We can observe that for other transfer learning methods, the two domains of the same class are not projected into the same region, and the target domain samples merge into the regions of other classes. The proposed method makes distributions with the same category across different domains much closer and the features cluster the best where all the data samples of different conditions are separated well. And it proves that this method can reduce the domain distribution discrepancy. In this way, the necessity of the proposed method is demonstrated.

D. DISCUSSION

We proposed a new unsupervised domain adaption method for cross-domain fault diagnosis. Researching cross-domain bearing fault diagnosis under complex working conditions and multi-task transferring constitutes future work. Also, the model can minimize the global feature distribution, we can focus on the marginal feature distribution of each category like CAG [37] to improve the adaptation performance. Moreover, the model may need to make instant or deliberate decisions upon short or long input signals, a hybrid long and short fusion strategy [38] may also be promising for online decision in future work.

V. CONCLUSION

In this paper, we propose a transfer learning method based on the deep adversarial domain adaption framework to solve the problem of fault diagnosis with unlabeled bearing data. We use MK-MMD and domain confusion function as the loss of the network. After obtaining the labeled fault data of a machine, the bearing fault diagnosis model can be established based on the unlabeled data with different loads or different sensor positions, which saves the labor cost of the labeled data. Compared with other methods in CWRU dataset, the proposed method is better than other deep transfer learning methods, which proves the effectiveness and reliability of the proposed method. This method can be extended to other mechanical equipment fault diagnosis, and can promote the development of unlabeled fault diagnosis.

REFERENCES

- [1] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," *Mech. Syst. Signal Process.*, vol. 107, pp. 241–265, Jul. 2018.
- [2] V. T. Tran, B.-S. Yang, F. Gu, and A. Ball, "Thermal image enhancement using bi-dimensional empirical mode decomposition in combination with relevance vector machine for rotating machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 38, no. 2, pp. 601–614, Jul. 2013.
- [3] X. Zhang, B. Wang, and X. Chen, "Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine," *Knowl.-Based Syst.*, vol. 89, pp. 56–85, Nov. 2015.
- [4] X. Wang, Y. Qin, Y. Wang, S. Xiang, and H. Chen, "ReLU-Tanh: An activation function with vanishing gradient resistance for SAE-based DNNs and its application to rotating machinery fault diagnosis," *Neurocomputing*, vol. 363, pp. 88–98, Oct. 2019.
- [5] M. He and D. He, "Deep learning based approach for bearing fault diagnosis," *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 3057–3065, May 2017.
- [6] 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.
- [7] J. O' Neill and D. Bollegala, "Dropping networks for transfer learning," 2018, *arXiv:1804.08501*. [Online]. Available: <http://arxiv.org/abs/1804.08501>
- [8] V. M. Patel, R. Gopalan, R. Li, and R. Chellappa, "Visual domain adaptation: A survey of recent advances," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 53–69, May 2015.
- [9] D. S. Kermany, M. Goldbaum, W. Cai, C. C. S. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, and J. Dong, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122.e9–1131.e9, Feb. 2018.
- [10] R. Yan, F. Shen, C. Sun, and X. Chen, "Knowledge transfer for rotary machine fault diagnosis," *IEEE Sensors J.*, early access, Oct. 23, 2019, doi: [10.1109/JSEN.2019.2949057](https://doi.org/10.1109/JSEN.2019.2949057).
- [11] R. Zhang, H. Tao, L. Wu, and Y. Guan, "Transfer learning with neural networks for bearing fault diagnosis in changing working conditions," *IEEE Access*, vol. 5, pp. 14347–14357, 2017, doi: [10.1109/ACCESS.2017.2720965](https://doi.org/10.1109/ACCESS.2017.2720965).
- [12] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446–2455, Apr. 2019.
- [13] P. Cao, S. Zhang, and J. Tang, "Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning," *IEEE Access*, vol. 6, pp. 26241–26253, 2018.
- [14] J. Zhang, J. Tian, T. Wen, X. Yang, Y. Rao, and X. Xu, "Deep fault diagnosis for rotating machinery with scarce labeled samples," 2019, *arXiv:1907.09411*. [Online]. Available: <http://arxiv.org/abs/1907.09411>
- [15] 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.
- [16] J. Xie, L. Zhang, L. Duan, and J. Wang, "On cross-domain feature fusion in gearbox fault diagnosis under various operating conditions based on transfer component analysis," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Jun. 2016, pp. 1–6.
- [17] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, "Deep model based domain adaptation for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2296–2305, Mar. 2017.
- [18] X. Li, W. Zhang, Q. Ding, and J.-Q. Sun, "Multi-layer domain adaptation method for rolling bearing fault diagnosis," *Signal Process.*, vol. 157, pp. 180–197, Apr. 2019.
- [19] X. Li, W. Zhang, and Q. Ding, "Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks," *IEEE Trans. Ind. Electron.*, vol. 66, no. 7, pp. 5525–5534, Jul. 2019.
- [20] B. Zhang, W. Li, X.-L. Li, and S.-K. Ng, "Intelligent fault diagnosis under varying working conditions based on domain adaptive convolutional neural networks," *IEEE Access*, vol. 6, pp. 66367–66384, 2018.
- [21] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7316–7325, Sep. 2019.
- [22] X. Li, W. Zhang, N.-X. Xu, and Q. Ding, "Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places," *IEEE Trans. Ind. Electron.*, vol. 67, no. 8, pp. 6785–6794, Aug. 2020, doi: [10.1109/TIE.2019.2935987](https://doi.org/10.1109/TIE.2019.2935987).
- [23] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. NIPS*, 2015, pp. 2672–2680.
- [24] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2962–2971, doi: [10.1109/CVPR.2017.316](https://doi.org/10.1109/CVPR.2017.316).
- [25] X. Li, W. Zhang, and Q. Ding, "Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction," *Rel. Eng. Syst. Saf.*, vol. 182, pp. 208–218, Feb. 2019.
- [26] Y. Zhang, K. Xing, R. Bai, D. Sun, and Z. Meng, "An enhanced convolutional neural network for bearing fault diagnosis based on time–frequency image," *Measurement*, vol. 157, Jun. 2020, Art. no. 107667.
- [27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [28] E. Tzeng, J. Hoffman, T. Darrell, and K. Saenko, "Simultaneous deep transfer across domains and tasks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 4068–4076, doi: [65210.1109/ICCV.2015.463](https://doi.org/10.1109/ICCV.2015.463).
- [29] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Scholkopf, and A. J. Smola, "Integrating structured biological data by kernel maximum mean discrepancy," *Bioinformatics*, vol. 22, no. 14, pp. e49–e57, Jul. 2006.
- [30] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. 32nd Int. Conf. Mach. Learn.*, vol. 37, 2015, pp. 97–105.
- [31] Y. Zhu, F. Zhuang, J. Wang, J. Chen, Z. Shi, W. Wu, and Q. He, "Multi-representation adaptation network for cross-domain image classification," *Neural Netw.*, vol. 119, pp. 214–221, Nov. 2019.
- [32] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mech. Syst. Signal Process.*, vols. 64–65, pp. 100–131, Dec. 2015.
- [33] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," 2014, *arXiv:1412.3474*. [Online]. Available: <http://arxiv.org/abs/1412.3474>
- [34] B. Sun and K. Saenko, "Deep CORAL: Correlation Alignment for Deep Domain Adaptation," in *Computer Vision—ECCV 2016 Workshops (Lecture Notes in Computer Science)*, vol. 9915, G. Hua, H. Jégou, Eds. Cham, Switzerland: Springer, 2016.
- [35] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," 2014, *arXiv:1409.7495*. [Online]. Available: <http://arxiv.org/abs/1409.7495>
- [36] L. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2625, 2008.
- [37] J. Zhang, J. Tian, Y. Cao, Y. Yang, and X. Xu, "Deep time–frequency representation and progressive decision fusion for ECG classification," *Knowl.-Based Syst.*, vol. 190, Feb. 2020, Art. no. 105402.
- [38] Q. Zhang, J. Zhang, W. Liu, and D. Tao, "Category anchor-guided unsupervised domain adaptation for semantic segmentation," 2019, *arXiv:1910.13049*. [Online]. Available: <http://arxiv.org/abs/1910.13049>



JIAJIE SHAO received the B.S. degree in mechanical design, manufacture and automation from Jiangsu Normal University, Xuzhou, Jiangsu, China. He is currently pursuing the M.S. degree in mechanical manufacture and automation with the University of Shanghai for Science and Technology, Shanghai, China. His current research interests include fault diagnosis and transfer learning.



ZHIWEN HUANG received the B.S. degree in mechanical design, manufacture and automation from Anhui Polytechnic University, Wuhu, Anhui, China, in 2010, and the M.S. degree in mechatronic engineering from the University of Shanghai for Science and Technology, Shanghai, China, in 2013, where he is currently pursuing the Ph.D. degree in mechanical engineering. His research interests include fault diagnosis and prognosis, process monitoring, and intelligent control.



JIANMIN ZHU received the B.S. and M.S. degrees in fluid transmission and control from the Luoyang Institute of Technology, Luoyang, Henan, China, in 1991 and 1994, respectively, and the Ph.D. degree in mechanical manufacture and automation from the Huazhong University of Science and Technology, Wuhan, Hubei, China, in 2000.

From 2000 to 2008, he was a Professor with the Mechanical and Electrical Engineering College, Henan University of Science and Technology, Luoyang, Henan. Since 2008, he has been a Professor and a Ph.D. Supervisor with the School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai, China. He is the author of five books, more than 100 articles, and more than 20 inventions. His research interests include process monitoring and intelligent control of electromechanical systems, precision measurement technique, and complex system modeling and control.

Dr. Zhu received awards and honors, include Second Prize from the National Science and Technology Progress Award, the Ministry of Education Natural Science Award, and the Henan Province Science and Technology Progress Award.

...