

Research Article

A Deep Transfer Learning Method for Bearing Fault Diagnosis Based on Domain Separation and Adversarial Learning

Shoubing Xiang ^{1,2}, Jiangquan Zhang,¹ Hongli Gao ¹, Dalei Shi,¹ and Liang Chen ^{1,2}

¹School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China

²Robotics Engineering Laboratory for Sichuan Equipment Manufacturing Industry, Sichuan Engineering Technical College, Deyang 618000, China

Correspondence should be addressed to Hongli Gao; hongli_gao@swjtu.edu.cn

Received 9 March 2021; Revised 17 April 2021; Accepted 1 June 2021; Published 18 June 2021

Academic Editor: Chengwei Fei

Copyright © 2021 Shoubing Xiang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Current studies on intelligent bearing fault diagnosis based on transfer learning have been fruitful. However, these methods mainly focus on transfer fault diagnosis of bearings under different working conditions. In engineering practice, it is often difficult or even impossible to obtain a large amount of labeled data from some machines, and an intelligent diagnostic method trained by labeled data from one machine may not be able to classify unlabeled data from other machines, strongly hindering the application of these intelligent diagnostic methods in certain industries. In this study, a deep transfer learning method for bearing fault diagnosis, domain separation reconstruction adversarial networks (DSRAN), was proposed for the transfer fault diagnosis between machines. In DSRAN, domain-difference and domain-invariant feature extractors are used to extract and separate domain-difference and domain-invariant features, respectively. Moreover, the idea of generative adversarial networks (GAN) was used to improve the network in learning domain-invariant features. By using domain-invariant features, DSRAN can adopt the distribution of the data in the source and target domains. Six transfer fault diagnosis experiments were performed to verify the effectiveness of the proposed method, and the average accuracy reached 89.68%. The results showed that the DSRAN method trained by labeled data obtained from one machine can be used to identify the health state of the unlabeled data obtained from other machines.

1. Introduction

Intelligent fault diagnosis can be successful only when two conditions are met [1]. First, the model should be trained by a large amount of labeled fault data. Second, the training data and test data should have the same probability distribution. However, in practical applications, it is often difficult to meet both conditions. First, it is difficult to obtain labeled fault data from some machines [2]. These machines can avoid faults while in operation because unexpected failures may lead to catastrophic accidents and cause heavy losses. Moreover, it may take a long time for a machine to degrade from a healthy state to failure, making obtaining the fault data of the machine very time-consuming. Second, the data are currently labeled manually in most cases, so that the labeling of a large amount of data

is expensive and time-consuming. Third, the probability distributions of data obtained from different machines are different, and the classification performance of the intelligent fault diagnosis methods can be significantly weakened when the training and test data sets are obtained from different machines.

As a new machine learning method, transfer learning can make full use of the knowledge learned in the auxiliary domain to solve new but related tasks in the target domain [3], thereby solving problems without enough labeled data to train the model. Moreover, transfer learning brings different domains close to each other by learning domain-invariant features, thereby effectively reducing the differences in the data distribution between the source and target domains.

To date, transfer learning has been widely used in a variety of applications, and many intelligent fault diagnosis

methods based on deep transfer learning have been proposed. Zhang et al. [4] trained the domain adaptive ability of a bearing fault diagnosis model using high-order Kullback–Leibler (KL) divergence, and the effectiveness of the model was verified under different working conditions. Wang et al. [5, 6] proposed an adaptive spectrum mode extraction-based fault diagnosis method. Che et al. [7] proposed a deep transfer learning method for rolling bearing fault diagnosis under variable operating conditions. By combining model-based transfer learning with feature-based transfer learning, the versatility of the convolutional neural network (CNN) under variable operating conditions was improved. Wen et al. [8] also developed a deep transfer learning method for fault diagnosis that was tested on bearing data sets collected under different loading conditions. In addition, to solve the problem of the significantly degraded performance of the traditional intelligent fault diagnosis algorithm degrades for changes in the workload, Guo et al. [9] proposed a transfer learning method and verified it by experimenting on the fault diagnosis of the wind turbine gearbox.

Based on the abovementioned studies, currently available intelligent fault diagnosis methods based on transfer learning mainly focus on the conversion between different working and loading conditions. However, in engineering practice, it is difficult to obtain a large amount of labeled data from a machine for model training. Hence, it is of great scientific and practical engineering significance to study transfer fault diagnosis between machines, so that a model trained by labeled data obtained from one machine can be extended to unlabeled data obtained from other machines.

In this study, a deep transfer learning method for bearing fault diagnosis based on domain separation and adversarial learning, domain separation reconstruction adversarial networks (DSRAN), was proposed. Specifically, domain-difference and domain-invariant feature extractors were used to extract domain-difference and domain-invariant features, respectively, of the source and target domains. To ensure the integrity of the features, the two types of features were integrated and reconstructed, and then the training was based on the idea of the generative adversarial network (GAN). When the classifier can correctly classify data in the source domain, the model can be used in cross-domain applications using invariant features.

The major contributions of this study were summarized as follows:

- (1) A new deep transfer learning method for bearing fault diagnosis was proposed. By directly learning the nonlinear mapping relationship between the original vibration signals and the health state of the bearing, this method automatically extracted the fault features and identified bearing faults from end to end. The domain-difference and domain-invariant feature extractors are used to extract and separate, respectively, domain-difference and domain-invariant features effectively. Moreover, the idea of GAN was used to improve the network in learning domain-invariant features. By using domain-invariant features, the negative transfer can be avoided effectively.

- (2) Transfer fault diagnosis of bearings between machines was carried out. Traditional intelligent fault diagnosis methods collect the training and test data sets from the same machine. In this study, the training and test data sets were collected from different yet related machines, with the data samples in the test set being unlabeled. The exploration in the intermachine transfer can promote the application of intelligent fault diagnosis in engineering practice.

The structure of this paper is as follows: In Section 1, the research background of transfer fault diagnosis is introduced. In Section 2, the deep transfer diagnosis method is presented. Section 3 describes the experiments and results of transfer fault diagnosis on three bearing data sets. The conclusions are given in Section 4.

2. Transfer Diagnosis

To describe the transfer diagnosis problem in this study, certain definitions in transfer learning [10] need to be introduced. $\{x_i, y_i\}_{i=1}^n$ is a data set with n samples, where sample x_i is labeled as y_i , and $x_i \in \chi$, where χ denotes the sample space. $y_i \in Y$, $Y = \{1, 2, \dots, k\}$ refers to the labeled space. k is the number of health states. In addition, since the sample data follow the marginal probability distribution $P(\chi)$, a specific domain in transfer learning is defined as $D = \{\chi, P(\chi)\}$. Traditional intelligent fault diagnosis methods obtain the training set and test set from the same domain; therefore, the two data sets have the same feature space and probability distribution. However, transfer learning obtains the training set and the test set from the source and the target domains, respectively. The feature spaces of the two data sets can be the same or different, but the probability distributions are different.

Based on the above definitions of transfer learning, the transfer learning problem for bearing fault diagnosis between machines is described as follows:

- (1) The source domain D_s consists of the sample space χ^s of labeled data obtained from a machine and its marginal probability distribution $P(\chi^s)$, $D_s = \{\chi^s, P(\chi^s)\}$. n_s samples with labeled information are collected from the source domain, $\{x_i^s, y_i^s\}_{i=1}^{n_s}$.
- (2) The target domain D_t is composed of the sample space χ^t of unlabeled data obtained from other machines and its marginal probability distribution $P(\chi^t)$, $D_t = \{\chi^t, P(\chi^t)\}$. n_t samples with health states to be identified are extracted from the target domain D_t , that is, $\{x_i^t\}_{i=1}^{n_t}$.
- (3) There should be correlated fault information between the source and the target domains [10]. Moreover, data in the source and the target domains shared the same labeled space, $Y^s, Y^t \in Y$, and the data have different probability distributions, $P(\chi^s) \neq P(\chi^t)$.

The data from the source domain are used for training, and a nonlinear mapping relationship between the sample

space χ^s and the labeled space Y^s can be established, $f: \chi^s \rightarrow Y^s$. Since the data in the target and source domains have different distributions, the identification accuracy may be low if the fault diagnosis knowledge obtained from the source domain, f , is directly used to identify the health state of the unlabeled samples in the target domain. Therefore, in this study, a deep transfer diagnosis model was constructed to reduce the data distribution differences between the source and the target domains by learning domain-invariant features, so that the fault diagnosis knowledge f obtained from one machine can be used to identify the health state of the unlabeled data obtained from other machines.

3. Deep Transfer Diagnosis of Bearing Faults

The bearing fault data obtained from different machines share the same feature space. It is assumed that all of the domains consist of two types of features, namely, domain-invariant and domain-difference features. Domain-invariant features have the same or similar classification capabilities in different domains, while domain-difference features often have strong classification capabilities in one domain, but poor classification performance in another domain. In transfer fault diagnosis between domains, if the domain-difference features are transferred, a negative transfer will be accomplished [11]. Thus, the DSRAN method [12] proposed in this study has two main functions: (1) extracting domain-invariant features in different domains and (2) conducting transfer fault diagnosis based on the extracted invariant features.

3.1. DSRAN. To represent different domain features, the DSRAN method adopts the domain-difference and the domain-invariant modules to extract the domain-difference and domain-invariant features in a domain. The goal of network training is to apply the domain-invariant features obtained by training to different domains and prevent the model from being affected by the distribution differences between the domains. As shown in Figure 1, the DSRAN method is composed of the following six parts:

3.1.1. Target-Difference Feature Extractor. The target-difference feature extractor that is used to extract the domain-difference features of the target domain adopts the structure of a deep CNN. The parameters of the extractor are shown in Table 1.

3.1.2. Source-Difference Feature Extractor. The source-difference feature extractor has the same network structure as that of the target-difference feature extractor.

3.1.3. Domain-Invariant Feature Extractor. The domain-invariant feature extractor that is used to extract the domain-invariant features of the source and target domains has a network structure similar to that of the domain-difference feature extractors, as shown in Table 2.

3.1.4. Reconstructor. The reconstructor combines the domain-difference and domain-invariant features of the source or the target domain and then sends them into the convolutional self-encoder for decoding in order to reconstruct the original signals. Table 3 shows the network parameters of the reconstructor.

3.1.5. Discriminator. The discriminator determines whether the input sample is the original or reconstructed input. Table 4 shows the network parameters of the discriminator.

3.1.6. Classifier. The domain-invariant feature extractor extracts the shared invariant features between the source and target domains. Some of these domain-invariant features possess a strong ability of classifying the fault data in the target domain, while some have poor performance in such classification. The classifier can help the network extract domain-invariant features with strong classification ability. During the training stage, the classifier is used to classify the samples in the source domain to ensure that the training is performed in the expected direction. Once the training is completed, the classifier is used to classify the unlabeled data in the target domain.

3.2. Loss Functions. To train each module of the network effectively, five loss functions are established to constrain the training process.

3.2.1. Difference Loss. During training, sample x^s from the source domain is placed into the source-difference and the domain-invariant feature extractors, from which the source-difference feature f_u^s and source-invariant feature f_v^s are obtained. Similarly, after being processed by the two target feature extractors, the target-difference feature f_u^t and target-invariant feature f_v^t are obtained. To ensure good results, the DSRAN method must effectively avoid negative transfer by completely separating the domain-difference and the domain-invariant features. Therefore, the difference loss function is proposed to constrain the two features as follows:

$$L_d = \|f_u^{sT} f_v^s\|_F^2 + \|f_u^{tT} f_v^t\|_F^2, \quad (1)$$

where L_d calculates the similarities between f_u^s and f_v^s as well as between f_u^t and f_v^t . When $f_u^s = f_v^s$ and $f_u^t = f_v^t$, L_d reaches the maximum; when f_u^s is orthogonal to f_v^s , and f_u^t is orthogonal to f_v^t , L_d reaches the minimum. Therefore, by minimizing L_d , f_u^s and f_v^s as well as f_u^t and f_v^t can be completely separated.

3.2.2. Similarity Loss. Even when f_u^s and f_v^s as well as f_u^t and f_v^t are completely separated, f_u^s and f_v^t may not necessarily be transferred. Hence, the similarity loss function is proposed to improve the similarity between the two. The loss function in discriminant-adaptive neural network (DANN) [13] is applied, as shown in the following equation:

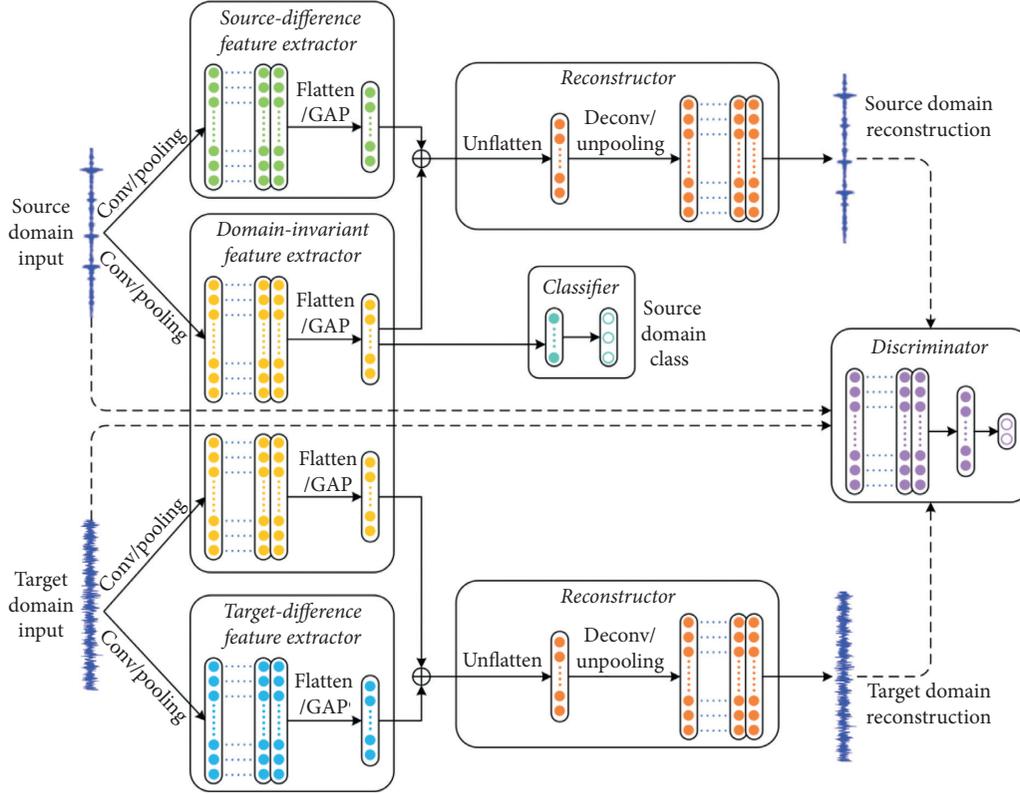


FIGURE 1: Overall structure of DSRAN.

TABLE 1: Network parameters of the domain-difference feature extractors.

Layer name	Kernel size/stride	Channel	Output size	Structure
Input	—	1	$1 \times 1,297$	—
Conv 1	$7/3$	32	1×431	—
Max pool	$3/2$	32	1×215	—
Conv 2	$5/3$	64	1×71	—
Block 3	3	64	1×35	$\begin{bmatrix} 1 \times 3, \text{stride}2 \\ 1 \times 3, \text{stride}1 \end{bmatrix} \times 1 \begin{bmatrix} 1 \times 3, \text{stride}1 \\ 1 \times 3, \text{stride}1 \end{bmatrix} \times 1$
Block 4	3	96	1×17	$\begin{bmatrix} 1 \times 3, \text{stride}2 \\ 1 \times 3, \text{stride}1 \end{bmatrix} \times 1 \begin{bmatrix} 1 \times 3, \text{stride}1 \\ 1 \times 3, \text{stride}1 \end{bmatrix} \times 1$
Block 5	3	128	1×8	$\begin{bmatrix} 1 \times 3, \text{stride}2 \\ 1 \times 3, \text{stride}1 \end{bmatrix} \times 1 \begin{bmatrix} 1 \times 3, \text{stride}1 \\ 1 \times 3, \text{stride}1 \end{bmatrix} \times 1$

TABLE 2: Network parameters of the domain-invariant feature extractor.

Layer name	Kernel size/stride	Channel	Output size	Structure
Input	—	1	$1 \times 1,297$	—
Conv 1	$7/3$	48	1×431	—
Max pool	$3/2$	48	1×215	—
Conv 2	$5/3$	72	1×71	—
Block 3	3	72	1×35	$\begin{bmatrix} 1 \times 3, \text{stride}2 \\ 1 \times 3, \text{stride}1 \end{bmatrix} \times 1 \begin{bmatrix} 1 \times 3, \text{stride}1 \\ 1 \times 3, \text{stride}1 \end{bmatrix} \times 1$
Block 4	3	96	1×17	$\begin{bmatrix} 1 \times 3, \text{stride}2 \\ 1 \times 3, \text{stride}1 \end{bmatrix} \times 1 \begin{bmatrix} 1 \times 3, \text{stride}1 \\ 1 \times 3, \text{stride}1 \end{bmatrix} \times 1$
Block 5	3	152	1×8	$\begin{bmatrix} 1 \times 3, \text{stride}2 \\ 1 \times 3, \text{stride}1 \end{bmatrix} \times 1 \begin{bmatrix} 1 \times 3, \text{stride}1 \\ 1 \times 3, \text{stride}1 \end{bmatrix} \times 1$

TABLE 3: Network parameters of the reconstructor.

Layer name	Kernel size	Stride	Channel	Output size
Input	—	—	1	128 + 152
Linear	—	—	1	8 × 256
Reshape	—	—	256	1 × 8
Deconv 1	3	2	128	1 × 17
Deconv 2	3	2	96	1 × 35
Deconv 3	3	2	64	1 × 71
Deconv 4	3	2	32	1 × 143
Deconv 5	5	3	16	1 × 431
Deconv6	7	3	1	1 × 1,297

TABLE 4: Network parameters of the discriminator.

Layer name	Kernel size	Stride	Channel	Output size
Input	—	—	1	1 × 1,297
Conv 1	7	3	8	1 × 431
Conv 2	5	3	16	1 × 143
Conv 3	3	2	32	1 × 71
Conv 4	3	2	65	1 × 35
Conv 5	3	2	96	1 × 17
Conv 6	3	2	128	1 × 8
Linear	—	—	1	1 × 1

$$L_s = \sum_{i=0}^{N_s+N_t} \{d_i \log \hat{d}_i + (1 - d_i) \log(1 - \hat{d}_i)\}, \quad (2)$$

where d_i denotes the real domain of the i -th input sample x_i and determines whether x_i belongs to the source domain ($d_i = 0$) or the target domain ($d_i = 1$); \hat{d}_i represents the predicted labeled value of the i -th input sample x_i , $\hat{d}_i \in \{0, 1\}$.

Similarity loss L_s can measure the difference between f_v^s and f_v^t . When L_s reaches the minimum, f_v^s and f_v^t are so similar to each other that they almost follow the same distribution pattern, and when the distributions of f_v^s and f_v^t are approximately the same, the features extracted by the domain-invariant feature extractor are domain-invariant. In this way, the classifier that is effective in the source domain can be transferred to the target domain.

3.2.3. Reconstruction Loss. Although the difference and the similarity loss functions can completely separate f_u^s from f_u^t and f_v^s from f_v^t and ensure that f_v^s and f_v^t follow the same or similar distribution patterns, the integrity of data in the source and target domains cannot be guaranteed. However, the reconstructor can ensure the integrity and validity of the features. To form an adversarial relationship between the reconstructor and the discriminator, the reconstructor is updated, and the reconstruction loss is defined. The binary cross-entropy loss serves as the reconstruction loss, as shown in the following equation:

$$L_r = -\frac{1}{N} \sum_{j=0}^N [z_j \log \hat{z}_j - (1 - z_j) \log(1 - \hat{z}_j)], \quad (3)$$

where N denotes the total number of reconstructed samples; z_j refers to the labels corresponding to the reconstructed samples, all of which are set to be 1; and \hat{z}_j represents the probability value of the j -th sample to label 1.

3.2.4. Discrimination Loss. Discriminator D is included to classify the original and reconstructed samples accurately, with the former labeled 1 and the latter 0. The discrimination loss is calculated by the binary cross-entropy loss function as shown in the following equation:

$$L_p = -\frac{1}{2M} \sum_{i=0}^{2M} [y_i \log \hat{y}_i - (1 - y_i) \log(1 - \hat{y}_i)], \quad (4)$$

where M denotes the total number of input samples; y_i stands for the real label of the i -th sample, which is 1 in the case of real samples and 0 in the case of reconstructed samples; and \hat{y}_i represents the probability value of the i -th sample to label $\{0, 1\}$.

The reconstructor and the discriminator are optimized in two steps. The discriminator is optimized in the first step. For the original sample x and the reconstructed sample \hat{x} , the discriminator optimization is as follows:

$$\begin{aligned} \max_D V(D, R) = & E_{x \sim p_x(x)} [\log(D(x))] \\ & + E_{\hat{x} \sim p_{\hat{x}}(\hat{x})} [\log(1 - D(G(\hat{x})))] \end{aligned} \quad (5)$$

For the original sample x , the predicted result is supposed to be as close to 1 as possible, that is, the greater the value of $D(x)$, the better the result. For the reconstructed sample \hat{x} , the predicted result should be as close to 0 as possible, that is, the smaller the value of $D(G(\hat{x}))$, the better the result.

Then, the reconstructor is optimized in the second step as follows:

$$\min_R V(D, R) = E_{\hat{x} \sim p_{\hat{x}}(\hat{x})} [\log(1 - D(G(\hat{x})))] \quad (6)$$

To reduce the difference between the reconstructed and original samples, the label of the reconstructed sample should be 1. At this point, $D(G(\hat{x}))$ should be as large as possible. To unify the form with Equation (5), $1 - D(G(\hat{x}))$ is minimized.

The following equation shows the combined optimization of the discriminator and reconstructor:

$$\begin{aligned} \min_R \max_D V(D, R) = & E_{x \sim p_x(x)} [\log(D(x))] \\ & + E_{\hat{x} \sim p_{\hat{x}}(\hat{x})} [\log(1 - D(G(\hat{x})))] \end{aligned} \quad (7)$$

3.2.5. Classification Loss. By classifying the data samples in the source domain, the training is monitored to ensure that the extracted domain-invariant features can accurately classify the data in both the source and the target domains. The following equation shows the loss function of the classifier:

$$L_c = - \sum_{i=0}^{N_s} y_i^s \cdot \log \hat{y}_i^s, \quad (8)$$

where N_s denotes the total number of samples in the source domain; y_i^s refers to the one-hot code of the label corresponding to the i -th sample in the source domain; and \hat{y}_i^s is the softmax output of the i -th sample in the source domain.

During the training process, the decreasing rate of L_c , L_d , L_s , L_p , and L_r may be inconsistent, causing the model to be dominated by a certain module [14]. Hence, need to use weight coefficients, for the entire network, the final loss function is given by

$$L = L_c + \alpha L_d + \beta L_s + \lambda L_p + \varphi L_r. \quad (9)$$

The ultimate optimization objective of the network is to minimize the abovementioned loss, where α , β , λ , and φ denote the weight coefficients of different loss functions; their function is to balance the decreasing rate of various losses. The loss with a faster decreasing rate corresponds to a smaller coefficient, while the loss with a slower decreasing rate corresponds to a larger coefficient. The specific value is generally determined by experiments. After many tests, the final value of α is 0.3; β is 1; λ is 1; and φ is 1.

4. Results and Analysis

4.1. Transfer Diagnosis Data Sets. Three data sets collected from three different rolling bearing experimental platforms were used as the bearing data in this study.

4.1.1. Case Western Reserve University (CWRU) Bearing Data Set. Collected by the Electrical Engineering Laboratory [15] of CWRU in the United States, the CWRU bearing data set is a universally accepted standard data set for fault diagnosis of rolling bearing. Figure 2 shows the CWRU rolling bearing test stand. The bearings used in the experiment included a normal bearing, bearing with fault at the fan end, and bearing with fault at the drive end. The faults were found in the inner ring, outer ring, or rolling element. The electrical discharge machining (EDM) technology was used to introduce single-point faults in the bearing. Faults with diameters ranging from 0.007 inches to 0.040 inches were introduced. In the experiment, the vibration signals of the bearings were collected by an accelerometer at a sampling frequency of 12 kHz, and the bearing fault data at the drive end were also collected by an accelerometer at a sampling frequency of 48 kHz.

Only the data sets of normal bearings and bearings at the drive end were used in this study. The bearings at the drive end were sampled at a frequency of 48 kHz, and the fault diameter was 0.014 inches. The motor load is 1 hp and has a speed of 1,772 r/min. The constructed data set included three types of faults: normal state (NR), inner ring fault (IR), and outer ring fault (OR).

4.1.2. Paderborn University Bearing Data Set. The Paderborn University bearing data set [16] was collected on the modularized experimental platform. The platform consisted



FIGURE 2: CWRU rolling bearing test stand.

of five modules: a motor, the torque measuring shaft, the rolling bearing test module, the flywheel, and a load motor. The flywheel and the loading device were used to simulate the inertia and load of the drive device. The rated torque of the motor is 6 Nm (power: 1.7 kW).

The data set contained two types of bearing damage data, namely, artificial and real damages. In this study, only the artificial damage data were used, including three types of faults, that is, NR, IR, and OR.

4.1.3. Xi'an Jiaotong University (XJTU-SY) Bearing Data Set. XJTU-SY rolling bearing accelerated life test data Set [17] contains the life cycle data of fifteen rolling bearings under three working conditions that was collected by Sumyoung Technology Co., Ltd. (Changxing City, Zhejiang Province, China) and a research team led by Professor Lei Yaguo from Xi'an Jiaotong University in a two-year rolling bearing accelerated life test. The experimental platform was mainly composed of the alternating-current (AC) motor, the hydraulic loading system, the digital force display, the revolving speed controller, the accelerometer, and the bearings. In the experiment, the vibration signals were sampled every 1 min at a frequency of 25.6 kHz, and the duration of each sampling was 1.28 s. The faults are worn inner ring, fractured outer ring, rolling elements, and cracked retainer of the tested bearings.

The data used in this study were collected at a speed of 2250 r/min and a radial force of 11 kN. The data set included three different bearing states, that is, NR, IR, and OR. For each state, there were 30 sample files.

4.2. Transfer Diagnosis Experiment. Currently, the most challenging part of fault diagnosis is intelligent fault diagnosis of machines with unlabeled data. To overcome this challenge, in this study, the classifier was first trained with labeled data obtained from one machine and then was used to classify the unlabeled data obtained from other machines.

In this section, six experiments were carried out on three rolling bearing data sets obtained from three different but related machines to study the transfer learning method for bearing fault diagnosis between machines. The data were vibration signals collected from different machines under different operating conditions. The types of bearing faults included NR, IR, and OR. Table 5 shows the distribution of the data sets in each transfer diagnosis experiment.

TABLE 5: Distribution of data sets in the transfer fault diagnosis experiments.

Transfer fault diagnosis experiments	Training data set	Test data set (30%)
CWRU→Paderborn	CWRU (labeled 100%) Paderborn (unlabeled 70%)	Paderborn
CWRU→XJTU-SY	CWRU (labeled 100%) XJTU-SY (unlabeled 70%)	XJTU-SY
Paderborn→CWRU	Paderborn (labeled 100%) CWRU (unlabeled 70%)	CWRU
Paderborn→XJTU-SY	Paderborn (labeled 100%) XJTU-SY (unlabeled 70%)	XJTU-SY
XJTU-SY→CWRU	XJTU-SY (labeled 100%) CWRU (unlabeled 70%)	CWRU
XJTU-SY→Paderborn	XJTU-SY (labeled 100%), Paderborn (unlabeled 70%)	Paderborn

The performance of the DSRAN method was evaluated based on the six transfer fault diagnosis experiments (Table 5). In each experiment, the data sets on the left and right sides of the arrow represent the data from the source and target domains, respectively. The training data set covered all labeled samples from the source domain and 70% unlabeled samples from the target domain, and the test data set contained the remaining 30% unlabeled samples from the target domain.

Figure 3 shows the results on the accuracy of DSRAN in the six transfer diagnosis experiments. The lowest accuracy of fault diagnosis was 85.06% occurred during the transfer from the XJTU-SY bearing data set to the CWRU bearing data set, and the highest accuracy of fault diagnosis was 93.15%, which was found during the transfer from the Paderborn bearing data set to the XJTU-SY bearing data set. The average accuracy of the six experiments was 89.68%, indicating that the DSRAN method proposed in this study can effectively train the classifier with the labeled data collected from one machine and then classify the unlabeled data collected from other machines.

4.3. Comparison of Different Methods. To further verify the effectiveness of the proposed DSRAN method, the same transfer diagnosis experiments were performed using the other five methods, that is, support vector machine (SVM), deep convolutional neural networks with wide first-layer kernels (WDCNN) [18], deep domain confusion (DDC) [19], deep reconstruction-classification networks (DRCN) [20], and DANN [13]. For comparison purposes, these methods were divided into three types (Table 6).

- (1) Traditional machine learning methods: The comparison aimed to demonstrate the differences between traditional machine learning methods that use manually extracted features and the deep learning methods that automatically extracted the features. With the original vibration signal as the input, the deep learning method performs an end-to-end fault diagnosis without the manual extraction of features.

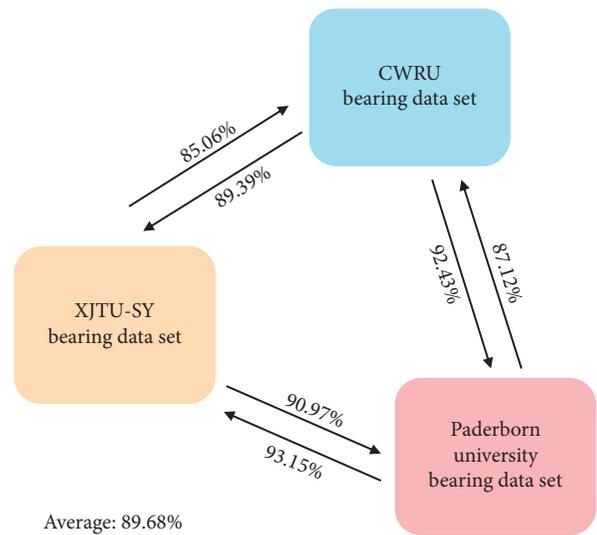


FIGURE 3: Accuracy of DSRAN in the six transfer diagnosis experiments.

- (2) Regular CNN methods: Regular CNN methods only use labeled samples from the source domain for training. However, the transfer learning method can train the model with both labeled samples from the source domain and some of the unlabeled samples from the target domain.
- (3) Classical transfer learning methods: Currently, available transfer learning methods have achieved good results in minimizing the distribution differences between the source and the target domains by maximum mean difference and domain-adversarial training, and networks can learn more and better domain-invariant features. With the integration of domain separation and adversarial learning, the DSRAN method can ensure that the learned features are domain-invariant.

Figure 4 shows the classification accuracy of the five methods in the six experiments. The proposed DSRAN method outperformed these methods in the six experiments.

TABLE 6: Comparison of different methods.

Method	Type	Feature extraction	Transfer method
SVM	Traditional machine learning	Manual extraction	None
WDCNN		Automatic extraction	None
DRCN	Regular CNN	Automatic extraction	Reconstruction
DDC		Automatic extraction	Maximum mean difference
DANN	Transfer learning	Automatic extraction	Domain-adversarial
DSRAN		Automatic extraction	Domain separation and adversarial learning

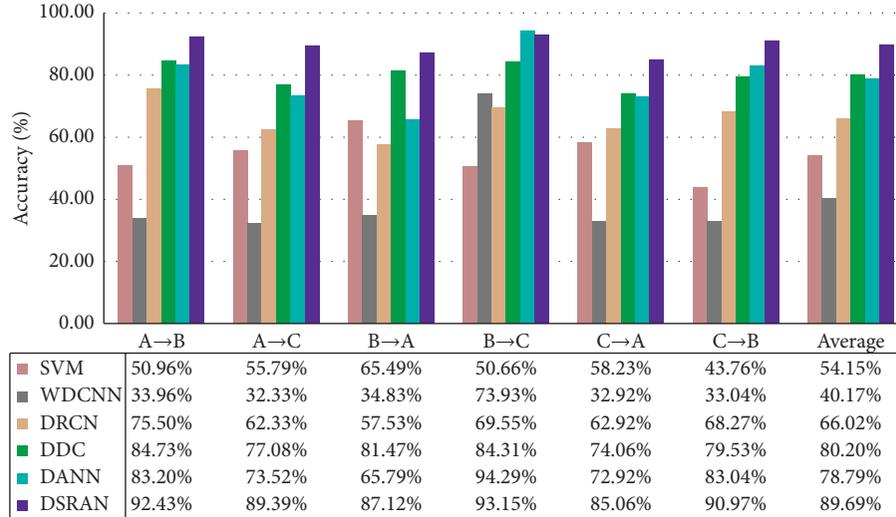


FIGURE 4: Transfer diagnosis results of the five control methods.

Specifically, the classification comparison showed the following results:

- (1) Compared with traditional machine learning methods, the deep learning algorithms showed higher accuracy, indicating that automatic feature extraction in deep learning was superior to manual feature extraction in machine learning. In particular, the classification accuracy of WDCNN was lower than that of SVM because WDCNN, which is designed to contain only a few simple convolutional layers, was unable to learn enough deep features. Moreover, for the deep learning methods that directly used the original signal for training, the distribution differences in the learned features were reduced.
- (2) Transfer learning methods generally had a higher classification accuracy than the regular CNN methods. This is largely because transfer learning methods can reduce the distribution differences between the data from the source domain and the data from the target domain, while regular CNN methods only use data from the source domain for training, without effectively utilizing the information collected from the target domain.
- (3) Compared with the other three widely used transfer learning methods, namely, DRCN, DDC, and DANN, the proposed DSRAN method realized higher classification accuracy in the six transfer

diagnosis experiments, indicating that DSRAN narrowed the distribution differences between data from the source and the target domains more effectively than the three extensively applied transfer learning methods. Moreover, the relatively high accuracy verified the practicability of the proposed DSRAN method.

5. Conclusions

In this study, transfer learning was used in intelligent bearing fault diagnosis between machines so that the intelligent fault diagnosis model trained by labeled data obtained from one machine can be used to identify the health state of the unlabeled data obtained from the other machines. The method solved the problems of the difficulty to obtain a large amount of labeled data from some machines for model training and of the fault diagnosis method trained with labeled data from one machine being not able to classify unlabeled data from other machines. The deep transfer method for bearing fault diagnosis proposed in this study was verified in six transfer fault diagnosis experiments, and the following conclusions were drawn:

- (1) By extracting the domain-invariant transfer fault features from data in the source and the target domains, DSRAN can reduce the differences in data distribution between different domains, so that the fault diagnosis knowledge obtained from one

machine can be used to identify the health state of the unlabeled data from other machines.

- (2) Compared with other methods, the proposed DSRAN method classified bearing faults more accurately, suggesting that the DSRAN method trained by labeled data obtained from one machine can effectively classify unlabeled data from other machines. Hence, DSRAN can be applied to fault diagnosis of machines with unlabeled data.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant nos. 51905452 and 51775452), the Fundamental Research Funds for the Central Universities (Grant nos. 2682019CX35 and 2682017ZDPY09), the China Postdoctoral Science Foundation (Grant no. 2019M663549), the Local Development Foundation guided by the Central Government (Grant no. 2020ZYD012), and the planning project of the Science & Technology Department of Sichuan Province (Grant no. 2019YFG0353).

References

- [1] 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 Transactions on Industrial Electronics*, vol. 66, no. 9, pp. 7316–7325, 2018.
- [2] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Burlington, MA, USA, 3rd edition, 2005.
- [3] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, pp. 1–40, 2016.
- [4] G. Zhang, H. Li, Y. Ran, and Q. Li, "A transfer learning model for bearing fault diagnosis," *Journal of Jilin University (Engineering and Technology Edition)*, vol. 50, no. 5, pp. 1617–1626, 2020.
- [5] Z. Wang, W. Zhao, W. Du, N. Li, and J. Wang, "Data-driven fault diagnosis method based on the conversion of erosion operation signals into images and convolutional neural network," *Process Safety and Environmental Protection*, vol. 149, pp. 591–601, 2021.
- [6] Z. Wang, N. Yang, N. Li, W. Du, and J. Wang, "A new fault diagnosis method based on adaptive spectrum mode extraction," *Structural Health Monitoring*, 2021, In press.
- [7] C. Che, H. Wang, Q. Fu, and X. Ni, "Deep transfer learning for rolling bearing fault diagnosis under variable operating conditions," *Advances in Mechanical Engineering*, vol. 11, no. 12, pp. 1–11, 2019.
- [8] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 1, pp. 136–144, 2017.
- [9] J. Guo, J. Wu, S. Zhang et al., "Generative transfer learning for intelligent fault diagnosis of the wind turbine gearbox," *Sensors*, vol. 20, no. 5, p. 1361, 2020.
- [10] Y. Lei, B. Yang, Z. Du, and N. Lu, "Deep transfer diagnosis method for machinery in big data era," *Journal of Mechanical Engineering*, vol. 55, no. 7, pp. 1–8, 2019.
- [11] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [12] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, and D. Erhan, "Domain separation networks," 2016, <https://arxiv.org/abs/1608.06019>.
- [13] Y. Ganin, E. Ustinova, H. Ajakan et al., "Domain-adversarial training of neural networks," *The Journal of Machine Learning Research*, vol. 17, no. 59, 2016.
- [14] S. Vandenhende, S. Georgoulis, W. Van Gansbeke, M. Proesmans, D. Dai, and L. Van Gool, "Multi-task learning for dense prediction tasks: a Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, In press.
- [15] CWRU, *The Case Western Reserve university Bearing Data Center Website*, CWRU, Cleveland, OH, USA, 2019, <https://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website>.
- [16] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, "Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification," in *Proceedings of the European Conference of the Prognostics and Health Management Society*, Bilbao, Spain, July 2016.
- [17] L. Yaguo, H. Tianyu, and W. Wang, L. Naipeng, Y. Tao, and Y. Jun, XJTU-SY rolling element bearing accelerated life test datasets: a tutorial," *Journal of Mechanical Engineering*, vol. 55, no. 16, pp. 1–6, 2019.
- [18] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, pp. 425–445, 2017.
- [19] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: maximizing for domain invariance," 2014, <https://arxiv.org/abs/1412.3474>.
- [20] M. Ghifary, W. B. Kleijn, M. Zhang, D. Balduzzi, and W. Li, "Deep reconstruction-classification networks for unsupervised domain adaptation," in *Proceedings of the Computer Vision-ECCV 2016*, pp. 597–613, Amsterdam, The Netherlands, October 2016.