Rotating machinery fault diagnosis by deep adversarial transfer learning based on subdomain adaptation

Jiajie Shao¹*, Zhiwen Huang¹*, Yidan Zhu², Jianmin Zhu¹* and Dianjun Fang³

Abstract

Rotating machinery fault diagnosis is very important for industrial production. Many intelligent fault diagnosis technologies are successfully applied and achieved good results. Due to the fact that machine damages usually happen under different working conditions, and manual scale labeled data are too expensive, domain adaptation has been developed for fault diagnosis. However, the current methods mostly focus on global domain adaptation, the application of subdomain adaptation for fault diagnosis is still limited. A deep transfer learning method is proposed for rotating machinery fault diagnosis in this study, where subdomain adaptation and adversarial learning are introduced to align local feature distribution and global feature distribution separately. Experiments are performed on two rotating machinery datasets to verify the effectiveness of this method. The results reveal that this method has outstanding mutual migration ability and can improve the diagnostic performance.

Keywords

Subdomain adaptation, adversarial training, deep learning, transfer learning, fault diagnosis

Introduction

Rotating machinery is currently used in various industries, and timely and effective diagnosis of rotating machinery faults is also very important.¹ With the maturity of the concept and technology of big data and intelligent manufacturing, intelligent fault diagnosis methods have been applied.²,³ Deep learning is also widely used to solve fault diagnosis due to its powerful feature learning ability.⁴–⁶ The success of these methods depends on the large amount of labeled data available for supervised learning, and they require training and testing data come from the same distribution. However, manual large scale labeled data are too expensive and sometimes cannot be collected in practice. Also, due to the fact that the working conditions of the machine usually change in different tasks, there is a general distribution difference between training and testing data in the real industry. The predictive models trained using these deep representations on one dataset cannot be well generalized to novel tasks.

Transfer learning can mine domain invariant essential features and structures between two different but
interrelated domains, which enables supervised information such as labeled data to be migrated and reused between domains. Recently, transfer learning methods have been increasingly used in rotating machinery fault diagnosis. Shao et al. used a pre-trained network and fine-tune strategy to achieve fast and accurate fault state classification. He et al. proposed a deep transfer auto-encoder for fault diagnosis of small samples under different working conditions. However, these methods still require the labels of the target domain.

Domain adaptation is a representative method in transfer learning that learns domain-invariant features without using target labels to bridge source domain and target domain. Because deep networks could learn more transferable features, embedding domain adaptation method into deep learning can better match the distributions across domains. Minimizing domain distribution discrepancy is the most popular approach in cross-domain fault diagnostics and the maximum mean discrepancy (MMD) is commonly used as a metric. Li et al. proposed a domain adaptation architecture and minimized multi-kernel MMD between two domains to realize cross-domain fault diagnosis. Wen et al. also used MMD with a three-layer sparse auto-encoder for bearing fault diagnosis. Zhang et al. used a domain adaptive convolutional neural networks, which combined MMD and fine-tune strategy for fault diagnosis under different working conditions.

Recently, adversarial domain adaptation integrates adversarial learning and domain adaptation, has been successfully embedded in deep networks to minimize the discrepancy distance of the domain discriminator by an adversarial objective. Guo et al. used adversarial adaptation and minimization of MMD to design a deep convolutional domain adaptation network for bearing fault diagnosis. Besides, Li et al. utilized adversarial learning on rotating machinery fault diagnosis. They also combined adversarial training and parallel data strategy to distinguish machine health conditions.

However, these domain adaptation methods only align global distributions between source and target domains, regardless of crucial information for each category. As a result, we cannot guarantee that samples from different domains but from the same category will be mapped near the feature space, since there is no labeled information for target domain. Subdomain adaptation can accurately align the class conditional distributions. As shown in Figure 1, when the categories of mechanical faults are close, such as different fault diameters of the same fault type, after global domain adaptation the overall distributions of source domain and target domain are close, but the features of different fault categories are too close to be accurately classified. After subdomain adaptation, different mechanical fault categories are divided into subdomains and the same mechanical fault can be aligned accurately. Subdomain adaptation can explore the dependency between the features and categories to capture the underlying multi-mode structures of data distributions. Xu et al. proposed a conditional domain adaptation method based on domain adversarial neural network for cross-domain fault diagnosis. Yu et al. proposed a simulation data-driven domain adaptation approach which align the marginal distribution and conditional distribution between simulation data and realistic monitoring data.

In this study, a deep transfer learning method based on subdomain adaptation and adversarial learning is proposed for rotating machinery fault diagnosis. We use domain confusion to align global feature distribution and use local maximum mean discrepancy to align local feature distribution. Also, we pre-train the feature
extractor on ImageNet dataset and utilize fine-tune strategy to speed up training process and improve accuracy. The experimental results reveal that this method can greatly improve the ability to extract the transferable features between the two domains, and can improve the diagnostic performance.

In the remainder, Section 2 introduces the problem formulation and domain adaptation. Section 3 presents the proposed method in detail, and the verification and research of the proposed algorithm are carried out in Section 4. In the end, Section 5 provides a conclusion of this paper.

Preliminary work

Problem formulation

In this study, the relationship between relevant fault subcategories is considered in the rotating machinery fault transfer diagnostic problem. In order to explain the concerned problem, some symbols, and definitions are first given in this section.

Let $D_s = \{(x_s^i, y_s^i)\}_{i=1}^{n_s}$ and $D_t = \{(x_t^i, y_t^i)\}_{i=1}^{n_t}$ denote the source and target domain samples from data distribution $P(x^s)$ and $P(x^t)$ respectively, where $x^s_i$ and $x^t_i$ are collected signals from different operation conditions or monitoring location, $y^s_i \in \mathbb{R}^c$ and $y^t_i \in \mathbb{R}^c$ are corresponding machine health condition labels, $n_s$ and $n_t$ are the number of the source and target domains fault samples, and $n_c$ is the number of rotating machinery fault categories. Particularly, the machine health condition label spaces of the source and target domains are identical in this study. However, due to the differences in operation conditions or sensor installation locations, there are great differences in feature distribution between rotating machinery fault monitoring data in the two domains, that is, $P(x^s) \neq P(x^t)$.

Domain adaptation

As previously mentioned, recent studies reveal that deep transfer learning can reduce the shifts in the distributions of monitoring signals from two different domains and learn transferable representations of machinery fault feature simultaneously. But these methods mainly focus on aligning the overall distributions of two domains, which ignore the relationship between subdomains in different domains of the same class. Actually, when machinery fault monitoring signals are collected from different operation conditions or sensor installation location, not only overall distributions of two domains is different, but also the relationship between subdomains in different domains of the same class is also variant. For this purpose, this study proposes a novel deep adversarial transfer learning method. This method uses adversarial learning and subdomain adaptation to realize rotating machinery transfer fault diagnostic, which aligning both global and local feature distribution of source and target domain simultaneously. This study considers the relationships between subdomains in different domains of the same class in machinery transfer fault diagnostic field.

Proposed method

Network architecture

In rotating machinery fault diagnosis field, the purpose of this study is to design a deep transfer learning network to not only reduce the shifts in overall distributions of source and target domains, but also reduce the shifts in local distributions within the same category of two domains, and learn transferable feature representations between different domains simultaneously. The architecture of the proposed deep adversarial transfer learning method based on subdomain adaptation is shown as Figure 2. The method consists of two feature extractors $F_s$ and $F_t$, a domain discriminator $D$, and a classifier $C$, which involves source domain and target domain. The feature extractor is used for learning high-level representations. The high-level features extracted by the feature extractor can use local maximum mean discrepancy to align local feature distribution. The classifier is designed for health state classification and the domain discriminator is designed for domain confusion loss to align global feature distribution.

Feature extractor. In this study, machinery fault monitoring signals are collected by accelerometer from
different operation conditions or installation locations. Considering vibration signal is a non-stationary time-varying signal in the machine running process, this study firstly applies short-time Fourier transform (STFT) to obtain time-frequency spectrum of raw signals, which is then fed into the feature extractors as inputs.27,28

Due to the different input spaces in fault diagnosis, two feature extractors $F_i$ and $F_t$ are employed to the source domain sample $x'$ and target domain sample $x'$, which have identical network structures. Moreover, considering the brilliant achievements of residual networks, which have identical network structures. Moreover, considering the brilliant achievements of residual networks, the 50-layers deep residual network is used as feature extractors.30 However, this will make the classifier identify the health status of the source domain correctly, supervised learning with the source labeled data is critical which is objective 1. To align the global feature distribution, we use adversarial learning, that is objective 2 and objective 3.

In this study, feature extractors $F_i$ and $F_t$ are pretrained on the ImageNet dataset, and the weights of two feature extractors are shared. Also, using ImageNet pre-training can accelerate the convergence on the target task and reduce over-fitting. Finally, the high-level feature representations of the input raw vibration signals from different operation conditions or installation locations can be obtained as $x_j^t = F_t(x')$ and $x_j^s = F_i(x')$ respectively.

Classifier. The classifier $C$ takes the obtained high-level feature representations $x_j^t$ and $x_j^s$ as input to diagnosis rotating machinery health conditions, which consists of a fully-connected layer and an output layer. Concretely, one fully-connected layer adopts ReLU as activation unit, while the output layer utilizes Softmax function as activation unit. In this way, the final classification of machinery health condition can be carried out.

Domain discriminator. The domain discriminator $D$ is a two-class classifier with adversarial learning, consisting of one fully-connected layer and an output layer. Similarly, one fully-connected layer utilizes ReLU as activation unit, while the output layer takes Softmax function as activation unit to identify whether high-level representation features come from source or target domain. Furthermore, in order to mine domain-invariant features, the adversarial learning is employed to train the domain discriminator.

Optimization objective

In this study, the proposed deep adversarial transfer learning method comprises four optimization objects. In order to make the classifier identify the health status of the source domain correctly, supervised learning is applied to maximize the domain recognition loss to reduce these four optimization objectives in detail.

Objective 1. In order to recognize rotating machinery health conditions, the proposed network should have the ability to learn discriminative features from the source domain supervised samples. Therefore, the source supervision is utilized to minimize the classification error. Concretely, the cross-entropy loss is regarded as the source supervision loss, which is defined as follow,

$$L_{ss} = -\frac{1}{n_s} \sum_{i=1}^{n_s} \sum_{j=1}^{n_c} 1[y_i^s = j] \log \frac{e^{yo_{C_i,j}}}{\sum_{j=1}^{n_c} e^{yo_{C_i,j}}}$$

where $x_{C_i,j}$ means the $j$-th output element of the $i$-th source domain sample in the classifier module, $y_i^s$ represents the corresponding label of rotating machinery health condition, $n_c$ is the number of health condition categories, and $1[\cdot]$ is an indicator function.

Objective 2. The domain discriminator is firstly applied to recognize whether the high-level features are from source domain or target domain. Therefore, domain recognition is introduced to minimize the domain recognition error between the two domains. Similarly, the cross-entropy loss is regarded as the domain recognition loss, which is defined as follow,

$$L_{dr} = -\frac{1}{n_t} \sum_{i=1}^{n_t} d_i \log \frac{e^{yo_{D_i,j}}}{\sum_{j=1}^{n_c} e^{yo_{D_i,j}}} - \frac{1}{n_t} \sum_{i=1}^{n_t} (1-d_i) \log \frac{1 - e^{yo_{D_i,j}}}{\sum_{j=1}^{n_c} e^{yo_{D_i,j}}}$$

where $x_{D_i,j}$ and $x_{D_i,j}$ means the $j$-th output element of the $i$-th source domain sample and target domain sample respectively in the domain discriminator module, and $d_i$ represents the corresponding ground truth domain label.

Objective 3. For learning domain-invariant features, the adversarial learning is employed to train domain discriminator. In general, the gradient reversal layer is applied to maximize the domain recognition loss to extract domain-invariant features like generative adversarial network.30 However, this will make the
discriminator converge too fast, and make the gradient disappear. In this study, we employ domain confusion loss as the adversarial loss to learn actual mapping. The cross-entropy loss function is used to train the map with uniform distribution and it predicts the input binary domain label to encourage them to predict as close as possible to a uniform distribution on the binary labels. The loss seeks to learn domain invariance to confuse the two domains. Hence, the third optimization objective is calculated as follow,

\[ L_{dc} = -\frac{1}{n_s} \sum_{i=1}^{n_s} \frac{1}{2} \log \left( \sum_{j=1}^{2} e^{h,ij} \right) - \frac{1}{n_t} \sum_{i=1}^{n_t} \frac{1}{2} \log \left( 1 - \sum_{j=1}^{2} e^{h,ij} \right) \]  

(3)

Generally, we need to simultaneously minimize equations (2) and (3) for the representation and the domain classifier parameters. Nevertheless, these two losses are directly opposed. Learning fully domain invariant feature extractors represent the domain discriminator does a poor job. While learning a high-performance domain discriminator represents the features learned by feature extractors are not domain invariant. We are not globally optimizing the parameters, but given the fixed parameters of the previous iteration, iterative update of the two objectives is performed. In this way, the loss can ensure that the adversarial discriminator views the two domains equally.

**Objective 4.** Maximum mean discrepancy (MMD)\(^{15}\) is a non-parametric distance estimation, and it is widely used in the field of rotating machinery fault diagnosis to measure the difference between target and source distribution. However, the previous MMD-based deep fault diagnosis transfer learning algorithms only focused on the global distributions, neglecting the relationship between subdomains in different domains of the same class in machinery transfer fault diagnostic field. Considering the relationship between related subdomains, it is of great significance to align the distributions of the related subdomains in the two domains of the same class. Therefore, the local maximum mean discrepancy (LMMD) is introduced to align distributions of the related subdomains, it assumes each sample belongs to weight \(w^c\), which is defined as follow,\(^{23}\)

\[ L_{sa} = \frac{1}{n_c} \sum_{c=1}^{n_c} \left[ \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} w^c_i w^c_j k(z^c_i, z^c_j) - 2 \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} w^c_i w^c_j k(z^c_i, z^c_j) \right] \]

(4)

where \(w^c_i\) and \(w^c_j\) mean the weight of \(x_i^c\) and \(x_j^c\) belonging to category \(c\), and \(z^c\) is the \(l\)-th \((l \in L = \{1, 2, ..., |L|\})\) layer activation. For sample \(x_i\), \(w^c_i\) is computed as follow,

\[ w^c_i = \frac{y_{ic}}{\sum_{j=1}^{n_t} y_{jc}} \]

(5)

where \(y_{ic}\) is the \(c\)th entry of vector \(y_i\). In source domain, the true label \(y_i^c\) is a one-hot vector to calculate \(w^c\). In target domain, there is no labeled data, but the output of the network is a probability distribution which can describes the probability of assigning \(x_i\) to each category. Therefore, it uses \(y_i^c\) as the probability to compute \(w^c\) for target domain.

When these optimization objects are built, the network is trained by stochastic gradient descent (SGD) algorithm. Due to the fact that the parameters of the feature extractors and classifiers in the two domains are shared, three modules are used: the feature extractor, domain discriminator, and classifier, whose parameters are denoted as \(\theta_F\), \(\theta_D\), and \(\theta_C\), respectively. Instead of global optimization parameters, we choose to optimize this objective in stages and use the following unconstrained optimization:

\[ L_1 = \min_{\theta_F, \theta_C} L_{ss} + \lambda L_{dr} + \gamma L_{sa} \]

(6)

\[ L_2 = \min_{\theta_D} L_{dc} \]

(7)

Equation (8) only updates \(\theta_F\) and \(\theta_C\) and equation (9) only updates \(\theta_D\). Also \(\lambda\) and \(\gamma\) are the penalty coefficients for \(L_{dr}\) and \(L_{sa}\). These updates ensure that we can learn domain-invariant representation.

Based on the equations (8) and (9), the parameters can be updated in each training epoch as follows:

\[ \theta_F \leftarrow \theta_F - \eta \left( \frac{\partial L_{ss}}{\partial \theta_F} + \lambda \frac{\partial L_{dr}}{\partial \theta_F} + \gamma \frac{\partial L_{sa}}{\partial \theta_F} \right) \]

(8)

\[ \theta_C \leftarrow \theta_C - \eta \frac{\partial L_{sa}}{\partial \theta_C} \]

(9)

\[ \theta_D \leftarrow \theta_D - \eta \frac{\partial L_{dc}}{\partial \theta_D} \]

(10)

where \(\eta\) is the learning rate.

In this way, the model can learn the domain invariant features from the two domains in the training process. Then the trained model can predict the unlabeled target samples according to these features.

**Experimental study**

**Dataset descriptions**

**CWRU dataset.** The CWRU dataset is came from the Case Western Reserve University Bearing Data
Center. It uses acceleration sensor to monitor the bearing with single-point faults damaged by electrical discharge machines (EDMs) and is widely used in fault diagnosis research. The experimental device is shown in Figure 3. The dataset includes four kinds of health conditions: Normal operation (Nor), inner race fault (IF), outer race fault (OF), and ball fault (BF). Different kinds of faults are also manually created with different fault diameters of 7, 14, and 21 mils. Hence, ten health states are diagnosed. The dataset includes four motor load conditions (0 HP, 1 HP, 2 HP, and 3 HP), and are acquired from the drive and fan end, respectively. To verify the advancement of our proposed method, the migration was carried out under both different load scenarios and different sensor positions. STFT is used for time-frequency imaging representation of vibration signal. About 200 images are taken for each category of health statuses in the two domains, totaling 2000 images, respectively.

**PHM 2009 challenge dataset.** The PHM2009 dataset is published by the PHM Society. It uses acceleration sensor to monitor the mixed faults of gear, bearing, and rotating shaft. At the same time, there are many kinds of fault states in each component, which makes the fault diagnosis of equipment a certain challenge. The dataset contains two sets of gears: spur and helical gears. The experiment device and schematic of the gearbox are shown in Figure 4.

In this study, the spur gearbox dataset is taken to test the accuracy of our proposed method. It has eight different health conditions under low and high load and 30, 35, 40 Hz speed. The Signal under each health condition is collected at a sampling frequency of 66.67 kHz and the acquisition time is 4 s. The label information of the dataset is shown in Table 1. To verify the advancement of our method, the migration was carried out under different load scenarios and different speeds. STFT is used for time-frequency imaging representation of vibration signal. About 200 images are taken for each category of health statuses in the two domains, totaling 2000 images, respectively.

| Label | Gear | 32T | 96T | 48T | 80T | Bearing | IS:IS | ID:IS | OS:IS | IS:OS | ID:OS | OS:OS | Shaft | Input | Output |
|-------|------|-----|-----|-----|-----|---------|------|------|------|------|------|------|-------|-------|-------|
| 1     | Nor  | Nor | Nor | Nor | Nor | Nor     | Nor  | Nor  | Nor  | Nor  | Nor  | Nor  | Nor   | Nor   | Nor   |
| 2     | Chipped | Nor | Eccentric | Nor | Nor | Nor     | Nor  | Nor  | Nor  | Nor  | Nor  | Nor  | Nor   | Nor   | Nor   |
| 3     | Nor  | Nor | Eccentric | Nor | Nor | Nor     | Nor  | Nor  | Nor  | Nor  | Nor  | Nor  | Nor   | Nor   | Nor   |
| 4     | Nor  | Nor | Eccentric | Broken | Ball | Nor     | Nor  | Nor  | Nor  | Nor  | Nor  | Nor  | Nor   | Nor   | Nor   |
| 5     | Chipped | Nor | Eccentric | Broken | Inner Ball | Outer | Nor  | Nor  | Nor  | Nor  | Nor  | Nor  | Nor   | Nor   | Nor   |
| 6     | Nor  | Nor | Nor | Nor | Broken | Inner Ball | Outer | Nor  | Nor  | Nor  | Nor  | Nor  | Balance | Nor   | Nor   |
| 7     | Nor  | Nor | Nor | Nor | Nor | Inner | Nor   | Nor  | Nor  | Nor  | Nor  | Nor  | Keyway | Nor   | Nor   |
| 8     | Nor  | Nor | Nor | Nor | Nor | Nor | Ball | Outer | Nor  | Nor  | Nor  | Nor  | Balance | Nor   | Nor   |

IS: Input Shaft; IS: Input Side; ID: Idler Shaft; OS: Output Side; OS: Output Shaft; Nor: Normal.
each category of health statuses in the two domains, totaling 1600 images, respectively.

**Compared methods and training details**

The proposed method is compared with other deep transfer learning methods: Deep Domain Confusion (DDC), Deep Adaptation Network (DAN), Domain-Adversarial Neural Networks (DANN), Deep CORAL (D-CORAL), and our previous work DADA-TL. DDC embeds MMD into an adaptation layer to learn domain invariant features. DAN uses multi-kernel MMD to align different distributions optimally to learn transferable features. DANN makes the domain discriminator unable to distinguish the source and target domain through adversarial training, thereby improving domain adaptability. Deep Coral uses CORAL loss to match source and target domains. We use ResNet-50 as the feature extractor for the above methods. We follow standard evaluation protocols for unsupervised domain adaptation, comparing the average accuracy of each method in three random experiments. For all MMD-based methods and our proposed method, we adopt Gaussian kernel with bandwidth set to median pairwise squared distances on the training data.

We use PyTorch framework to implement all transfer learning methods, and fine-tune ResNet models provided by PyTorch. The models have been pre-trained on the ImageNet 2012 dataset. The layers of feature extractor are fine-tuned and the layers of domain discriminator and classifier are trained from scratch via back propagation. Therefore, the learning rate of domain discriminator and classifier are set ten times of feature extractor. We use the learning rate annealing strategy in DANN, it is adjusted during SGD with 0.9 momentum and describing by the following expression: \( \eta = \eta_0 / (1 + ap)^b \), where \( p \) is the linearly changing from 0 to 1 in the training progress, \( \eta_0 = 0.005, \alpha = 10, \beta = 0.75 \). For suppressing noisy activations at the early stages of training, instead of fixing the penalty coefficients, we change them from 0 to 1 by a progressive schedule: \( 2/e^{-10p} - 1 \). The gradual strategy importantly stabilizes parameter sensitivity to a large extent and simplifies model selection.

**Results and analysis**

**CWRU dataset.** There are a total of twelve transfer tasks under different load scenarios, where \( T_{uv} \) is to take the data of \( u \) HP as the source domain, and the data of \( v \) HP as the target domain. Figure 5 shows the classification accuracy of this dataset under different loads. The proposed method is superior to other comparison methods in all transfer tasks. The accuracy of some tasks is even improved to 100%, and the average accuracy is 99.7%. It can be noticed that the accuracy of other deep transfer learning methods will decrease a lot when the difference between the two domains increase, such as \( T_{03} \) and \( T_{30} \). However, the proposed method greatly improves the accuracy of transfer tasks with large difference between the two domains. These results illustrate the key significance of subdomain adaptation in cross-domain fault diagnosis, and show that the proposed method can learn more transferable fault features to effectively realize domain adaptation.

![Figure 5. The classification accuracy of CWRU dataset under different loads.](image-url)
Furthermore, the confusion matrix of the proposed method for task $T_{30}$ is shown in Figure 6. It is noted that the misclassification only occurs in ball fault, and most of them are misclassified only in fault diameters. The other seven health conditions can be classified exactly. This verifies the effectiveness and superiority of this method in cross-domain fault diagnosis tasks. The proposed method can improve the accuracy in the case of the domain differences.

There are eight transfer tasks under different sensor locations, where $T_{uDF}$ represents the data of $u$ HP collected in the driver end as source domain, and the fan end as target domain. The classification accuracy under different sensor positions is shown in Figure 7. Although the CWRU dataset is usually simple for fault diagnosis, the problem of domain adaptation between different sensor positions can be more challenging. It can be seen that the accuracy of other methods under different sensor positions is lower than that under different loads, but the proposed method obtains good performance in all tasks. The accuracy of all transfer tasks is above 99.6%, and the average accuracy is 99.9%. It shows that the effectiveness of the method is obvious. Particularly, in the migration task from fan to drive end under the same load, the performance of other transfer learning methods is obviously lower than that from drive to fan end. After adding subdomain adaptation, the effect is obviously improved. The results of this method have broad application prospects in fault diagnosis.

**PHM 2009 challenge dataset** To further verify the advancement of our method, PHM2009 dataset is used. Its multi-class mixed faults make the fault diagnosis transfer task more challenging. Table 2 shows the accuracy of this dataset under different speeds, and Table 3 shows the classification accuracy under different loads. About 30–35 means that the data collected at 30 Hz is used as source domain and 35 Hz as target domain. The 30L-H represents the data with the low load of 30 Hz speed is used as source domain and the data with the high load of 30 Hz speed is used as target domain. It can be seen from the results that in the hybrid fault diagnosis transfer tasks, the performance of the global

![Figure 6. The confusion matrix of task $T_{30}$.](image)

![Figure 7. The classification accuracy of CWRU dataset under different sensor locations.](image)
domain adaptation is not satisfactory, and the performance is significantly reduced in different transfer tasks. Especially for migration under different loads, the best average accuracy of the global domain adaptation method is lower than 70%. The proposed method using subdomain adaptation can greatly increase the accuracy of classification. The average accuracy is 85.9% under different loads and 97.5% under different speeds, which greatly improves the performance. In general, the above results all illustrate that our method can effectively realize fault diagnosis. In addition, we take 40L-H task as an example to compare the performance of domain confusion loss and the gradient reversal layer as shown in Figure 8. The result shows that using domain confusion loss has higher accuracy than gradient reversal layer and is more suitable for our network architecture.

To intuitively analyze the performance of our method, t-SNE\textsuperscript{38} technique is applied to visualize the features extracted by the feature extractor from the two domains into a two-dimensional map. We visualized the 40L-H task, and the results are displayed in Figure 9. For global domain adaptation methods, they focus on marginal domain adaptation, when the categories of mechanical faults are close, the fault features in source and target domains are not aligned very well and some features are hard to classify. Although these methods can improve the distribution difference, it is still not satisfactory. For our proposed method, we can find that the fault features of the same category in two domains are aligned very well. Fault features in these two domains with the same category are very close, and fault features with different categories are also scattered. Subdomain adaptation can obtain more fault category information, which can effectively improve the performance of cross-domain fault diagnosis. The results suggest that the proposed method is more effective to reduce the distribution discrepancy of the two domain, and intuitively illustrate the high-performance of our method.

**Conclusions and future work**

This paper presents a deep adversarial transfer learning method on rotating machinery fault diagnosis. Unlike the previous method, the paper uses domain confusion

| Table 2. Classification accuracy (%) on the PHM 2009 dataset under different speeds. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Method          | 30–35 | 30–40 | 35–30 | 35–40 | 40–30 | 40–35 | Avg  |
|-----------------|--------|--------|--------|--------|--------|--------|------|
| DDC             | 91.8   | 80.4   | 72.2   | 96.5   | 49.6   | 83.4   | 79.0 |
| DAN             | 99.6   | 86.1   | 83.8   | 98.7   | 65.7   | 94.8   | 88.1 |
| DANN            | 96.2   | 64.6   | 73.6   | 95.6   | 62.3   | 86.8   | 79.9 |
| D-Coral         | 96.3   | 70.0   | 79.8   | 86.0   | 49.9   | 72.3   | 75.7 |
| DADA-TL         | 99.2   | 84.2   | 89.6   | 96.8   | 66.2   | 98.6   | 89.1 |
| Proposed        | 100    | 98.7   | 99.5   | 100    | 86.8   | 99.9   | 97.5 |

| Table 3. Classification accuracy (%) on the PHM 2009 dataset under different loads. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Method          | 30L-H | 30H-L | 35L-H | 35H-L | 40L-H | 40H-L | Avg  |
|-----------------|--------|--------|--------|--------|--------|--------|------|
| DDC             | 65.3   | 49.3   | 43.9   | 52.3   | 48.1   | 35.4   | 49.1 |
| DAN             | 66.5   | 63.2   | 61.2   | 62.1   | 64.4   | 61.4   | 63.1 |
| DANN            | 53.5   | 62.1   | 54.0   | 58.6   | 48.4   | 58.2   | 55.8 |
| D-Coral         | 59.0   | 55.6   | 57.8   | 60.9   | 49.6   | 61.9   | 57.5 |
| DADA-TL         | 69.8   | 69.1   | 69.4   | 62.7   | 73.1   | 69.2   | 68.9 |
| Proposed        | 87.4   | 87.3   | 86.8   | 75.9   | 96.5   | 81.2   | 85.9 |

![Figure 8. Comparison of using domain confusion loss and gradient reversal layer.](image-url)
and local maximum mean discrepancy to align global distributions and subdomain distributions simultaneously. By comparing with other transfer learning methods on two rotating machinery datasets, our proposed method improves the ability of extracting domain-invariant and transferable features and greatly improve the rate of accuracy. Furthermore, in the transfer tasks under various complex working conditions, the proposed method achieves the best results. It proves this method is an effective way to address the problem of the unlabeled data in practical industrial application. Furthermore, this method can also extend to the fault detection of other mechanical systems too. The future work can be focused on the imbalanced data problem and more transfer scenarios. Besides, different time-frequency analysis methods will affect the accuracy of the model, we will try to research it in the future.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is supported by National Natural Science Foundation of China (Grant No. 51775323).

ORCID iD
Jianmin Zhu https://orcid.org/0000-0002-7544-9721

Figure 9. The t-SNE visualization of features: (a) DAN, (b) DADA-TL, and (c) the proposed method.

References
1. Lei Y, Lin J, He Z, et al. A review on empirical mode decomposition in fault diagnosis of rotating machinery. Mech Syst Signal Process 2013; 35: 108–126.
2. Gao X, Wei H, Li T, et al. A rolling bearing fault diagnosis method based on LSSVM. Adv Mech Eng. Epub ahead of Print 1 January 2020. DOI: 10.1177/1687814019899561.
3. Wang Y, Wei Z and Yang J. Feature trend extraction and adaptive density peaks search for intelligent fault diagnosis of machines. IEEE Trans Ind Inform 2019; 15: 105–115.
4. Jian Y, Qing X, He L, et al. Fault diagnosis of motor bearing based on deep learning. Adv Mech Eng. Epub ahead of Print 13 September 2019. DOI: 10.1177/1687814019875620.
5. Che C, Wang H, Fu Q, et al. Deep transfer learning for rolling bearing fault diagnosis under variable operating conditions. Adv Mech Eng. Epub ahead of Print 30 December 2019. DOI:10.1177/1687814019897212.
6. Jing L, Zhao M, Li P, et al. A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. Measurement 2017; 111: 1–10.
7. Pan SJ and Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2010; 22: 1345–1359.
8. Yan R, Shen F, Sun C, et al. Knowledge transfer for rotary machine fault diagnosis. IEEE Sens J 2020; 20: 8374–8393.
9. Shao S, McAleer S, Yan R, et al. Highly accurate machine fault diagnosis using deep transfer learning. IEEE Trans Ind Inform 2019; 15: 2446–2455.
10. He Z, Shao H, Zhang X, et al. Improved deep transfer auto-encoder for fault diagnosis of gearbox under variable working conditions with small training samples. IEEE Access 2019; 7: 115368–115377.
11. Long M, Wang J and Jordan M. Deep transfer learning with joint adaptation networks. In: Proceedings of the 34th international conference on machine learning, Sydney, Australia, PMLR 2017, vol. 70, pp.2208–2217.
12. Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks. In: Proceedings of the 27th international conference on neural information processing systems, Montreal, QC, Canada, 08–13 December 2014, pp. 3320–328. California: NIPS.
13. Xu G, Liu M, Jiang Z, et al. Online fault diagnosis method based on transfer convolutional neural networks. IEEE Trans Instrum Meas 2020; 69: 509–520.
14. Lu W, Liang B, Cheng Y, et al. Deep model based domain adaptation for fault diagnosis. IEEE Trans Ind Electron 2017; 64: 2296–2305.
15. Borgwardt KM, Gretton A, Rasch MJ, et al. Integrating structured biological data by kernel maximum mean discrepancy. Bioinformatics 2006; 22: e49–e57.
16. Li X, Zhang W, Ding Q, et al. Multi-layer domain adaptation method for rolling bearing fault diagnosis. Signal Process 2019; 157: 180–197.
17. Wen L, Gao L and Li X. A new deep transfer learning based on sparse auto-encoder for fault diagnosis. IEEE Trans Syst Man Cybern Syst 2019; 49: 136–144.
18. Zhang B, Li W, Li XL, et al. Intelligent fault diagnosis under varying working conditions based on domain adaptive convolutional neural networks. *IEEE Access* 2018; 6: 66367–66384.

19. Tzeng E, Hoffman J, Saenko K, et al. Adversarial discriminative domain adaptation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, Honolulu, HI, 21–26 July 2017, pp. 2962–2971, New York: IEEE.

20. Guo L, Lei Y, Xing S, et al. Deep convolutional transfer learning network: a new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE Trans Ind Electron* 2019; 66: 7316–7325.

21. Li X, Zhang W, Ding Q, et al. Diagnosing rotating machines with weakly supervised data using deep transfer learning. *IEEE Trans Ind Inform* 2020; 16: 1688–1697.

22. Li X, Zhang W, Xu NX, et al. Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places. *IEEE Trans Ind Electron* 2020; 67: 6785–6794.

23. Zhu Y, Zhuang F, Wang J, et al. Deep subdomain adaptation network for image classification. *IEEE Trans Neural Netw Learn Syst* 2021; 32: 1713–1722.

24. Pei Z, Cao Z, Long M, et al. (2018). Multi-adversarial domain adaptation. In: *32nd AAAI Conference on Artificial Intelligence*, New Orleans, LA, 02–07 February 2018, pp. 3934–3941, Palo Alto: Assoc.

25. Xu NX and Li X. (2019). Cross-domain machinery fault diagnosis using adversarial network with conditional alignments. In: *10th IEEE Prognostics and System Health Management Conference (PHM-Qingdao)*, Qingdao, China, 25–27 October 2019, pp. 1–5, New York: IEEE.

26. Yu K, Fu Q, Mu H, et al. Simulation data driven weakly supervised adversarial domain adaptation approach for intelligent cross-machine fault diagnosis. *Struct Health Monit* 2021; 20: 2182–2198.

27. Li X, Zhang W and Ding Q. Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. *Reliab Eng Syst Saf* 2019; 182: 208–218.

28. Zhang Y, Xing K, Bai R, et al. An enhanced convolutional neural network for bearing fault diagnosis based on time–frequency image. *Measurement* 2020; 157: 107667.

29. He K, Zhang X and Ren S. Deep residual learning for image recognition. In: *Proceedings of the 2016 IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, 2016, pp.770–778, New York: IEEE.

30. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: *NeurIPS Proceedings*, 2015, pp. 2672–2680.

31. Smith WA and Randall RB. Rolling element bearing diagnostics using the case western reserve university data: a benchmark study. *Mech Syst Signal Process* 2015; 64–65: 100–131.

32. Phm data challenge 2009, https://www.phmsociety.org/competition/PHM/09.

33. Tzeng E, Hoffman J, Zhang N, et al. Deep domain confusion: maximizing for domain invariance. *arXiv*, 2014, pp. 1–9. http://arxiv.org/abs/1412.3474

34. Long M, Cao Y, Wang J, et al. (2015). Learning transferable features with deep adaptation networks. In *Proceedings of the 32nd international conference on machine learning, PMLR*, 2015, vol.37, pp. 97–105.

35. Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks. In: Csurka G (ed.) Domain adaptation in computer vision applications. Advances in computer vision and pattern recognition. Cham, Switzerland: Springer, 2017, pp.189–209.

36. Sun B and Saenko K. Deep CORAL: correlation alignment for deep domain adaptation. In: Hua G and Je´gou H (eds) *Computer vision—ECCV 2016 workshops (lecture notes in computer science)*. Amsterdam, Netherlands, 08–16 October 2016, pp.443–450, Berlin: Springer-Verlag.

37. Shao J, Huang Z and Zhu J. Transfer learning method based on adversarial domain adaption for bearing fault diagnosis. *IEEE Access* 2020; 8: 119421–119430.

38. Maaten L and Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008; 9: 2579–2625.

39. Yu K, Ma H, Han H, et al. Second order multisynchrosqueezing transform for rub-impact detection of rotor systems. *Mech Mach Theory* 2019; 140: 321–349.

40. Yu K, Lin TR, Ma H, et al. A combined Polynomial chirplet transform and synchroextracting technique for analyzing nonstationary signals of rotating machinery. *IEEE Trans Instrum Meas* 2020; 69: 1505–1518.