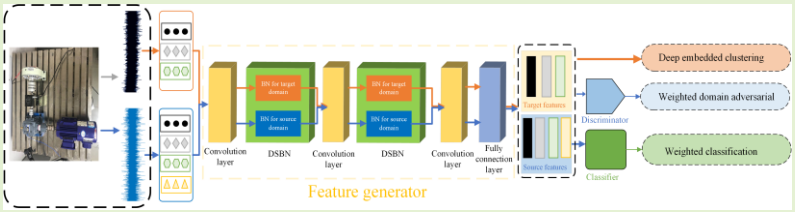
IEEE Sensors Council IEEE SENSORS JOURNAL, VOL. XX, NO. XX, MONTH X, XXXX

Clustering-guided novel unsupervised domain adversarial network for partial transfer fault diagnosis of rotating machinery

Hongru Cao, Haidong Shao, Bin Liu, Baoping Cai, and Junsheng Cheng

***Abstract*—Unsupervised partial transfer fault diagnosis studies of rotating machinery have practical significance, which still exists some challenges, for example, the learned domain-specific statistics and parameters usually influence the learning effect of target-domain features to some degree, and the relatively scattered target-domain features will lead to negative transfer. To overcome those limitations and further improve partial transfer fault diagnosis performance, a clustering-guided novel unsupervised domain adversarial network is proposed in this paper. Firstly, a novel unsupervised domain adversarial network is constructed using domain-specific batch normalization to remove domain-specific information to enhance alignment between source and target domains. Secondly, embedded clustering strategy is designed to learn tightly clustered target-domain features to suppress negative transfer in partial domain adaptation process. Finally, a joint optimization objective function is defined to balance different losses to improve the training and diagnosis performance. Two experimental cases of bevel gearbox and bearing are used to validate the effectiveness and superiority of the proposed method in solving unsupervised partial transfer fault diagnosis problems.**



***Index Terms*—embedded clustering strategy; joint optimization objective function; novel unsupervised domain adversarial network; partial transfer fault diagnosis; rotating machinery**

I. Introduction

Rotating machines applied in modern industries are becoming more safety-critical, and it is significant to prevent accidents by automatically and accurately identifying the health states of the components [1], [2]. In recent years, due to the excellent and automatic feature mining abilities, deep-learning-based techniques have attracted an increasing attention in intelligent fault diagnosis of rotating machines [3]-[6]. In most cases, the premise of existing deep-learning-based fault diagnosis methods is that the labeled training samples and testing samples should be sufficient and subject to the same distribution [7]. Nevertheless, in industrial situations, it is extremely difficult to acquire large amounts of labeled fault samples [8]. Thus, it remains a challenge on how to relax the premise to obtain a practical unsupervised fault diagnosis model.

This research is supported by the National Natural Science Foundation of China under Grant 51905160, and the Natural Science Fund for Excellent Young Scholars of Hunan Province under Grant 2021JJ20017. *(Corresponding author: Haidong Shao.)*

Hongru Cao, Haidong Shao, and Junsheng Cheng are with College of Mechanical and Vehicle Engineering, Hunan University, Changsha, China (e-mail: hongru@hnu.edu.cn; hdshao@hnu.edu.cn; chengjunsheng@hnu.edu.cn).

Bin Liu is with Department of Management Science, University of Strathclyde, Glasgow, UK (email: b.liu@strath.ac.uk).

Baoping Cai is with College of Mechanical and Electronic Engineering, China University of Petroleum, Qingdao, China (e-mail: caibaoping@upc.edu.cn).

As a commonly used feature transfer learning strategy, unsupervised domain adaptation methods can align the distribution discrepancies of source and target domains to extract the domain-invariant features, enabling to provide a new insight for the abovementioned problems [9]-[11]. In 2017, Wen *et al.* [12] applied maximum mean discrepancy (MMD) as a loss function in the sparse auto-encoder (SAE) fault diagnosis model to reduce the distribution discrepancy of the extracted features. In 2016, Lu *et al.* [13] adopted convolutional neural network (CNN) using MMD to strengthen the representative features. The effectiveness of the approach was validated in rolling bearing datasets. In 2019, Yang *et al.* [14] used the multi-Gaussian kernel-induced MMD in the diagnosis model, which further improved the robustness of MMD. In 2020, Jiao *et al.* [15] constructed a novel residual adversarial network for unsupervised cross-domain fault diagnosis, where the joint maximum mean discrepancy is used to simultaneously reduce the joint distribution shifts in the unsupervised domain adaptation process. In 2021, Xu *et al.* [16] proposed an intelligent fault diagnosis system, which constructs a domain discriminator through adversarial training to develop an unsupervised multisource transfer diagnosis approach. However, all of the methods are under the presumption that the label spaces of two domains are identical during the whole model training process.

However, for more general engineering scenarios, it is almost impossible to collect all the fault states in advance [17],

i.e., the label space of the target domain is a subset of label space of source domain, which is denoted as the partial transfer learning problem [18]. Partial transfer fault diagnosis of rotating machinery has attracted increased attentions since 2020 [19]-[22]. In 2020, Li *et al*. [23] used a class-weighted adversarial network-based method for partial transfer diagnosis, which added class-level weights on the corresponding source categories to measure the importance of each source category. In 2021, Li *et al*. [24] constructed a multiple classification module to extract fault diagnosis knowledge and two rolling bearing datasets are carried out for validation. In 2021, Liu *et al*. [25] proposed a SAE based weighted partial adversarial domain adaptation model to classify fault states. In 2021, Deng *et al*. [26] presented a novel adversarial network using two attention matrices to tackle the partial transfer fault diagnosis issue.

Nevertheless, the existing studies on partial transfer fault diagnosis of rotating machinery, to the best of our knowledge, focus on extracting the domain-invariant features for achieving satisfactory diagnosis performance in presence of the source labels, which may cause the following problems: (1) The extracted target-domain features are relatively scattered, and some of them may be wrongly transferred to the nearby source-domain features whose labels are different in the domain adaptation process, leading to negative transfer. (2) The learned shared statistics and parameters usually exhibit specific information of the source domain, which will influence the learning effect of target-domain features to some degree and result in domain misalignment [27].

In this article, to overcome the influence of negative transfer and improve partial transfer fault diagnosis performance, a clustering-guided unsupervised domain adversarial network is proposed. This article facilitates three significant contributions as follow.

(1) A novel unsupervised domain adversarial network is constructed using domain-specific batch normalization (DSBN) to remove domain-specific information to enhance alignment between source and target domain.

(2) Embedded clustering strategy is designed to learn tightly clustered target-domain features to suppress negative transfer in partial domain adaptation process.

(3) A joint optimization objective function is defined to balance different losses to improve the training and diagnosis performance.

In section 2, the article presents a brief overview of partial transfer fault diagnosis of rotating machinery. Section 3 presents the proposed method, and result analyses of two cases are introduced in section 4. Section 5 presents conclusions and the future work.

II. Introduction to Partial Transfer Fault Diagnosis of Rotating Machinery

Generally, a standard transfer learning task contains a source domain and a target domain, expressed as D_S and D_T, respectively, where $D_S = \{X^S, P_S(X)\}$, $D_T = \{X^T, P_T(X)\}$. The source domain $X^S = \{x_i^S, y_i^S\}_{i=1}^{N_S}$ contains N_S sufficient labeled samples, where x_i^S denotes the ith sample of source domain, y_i^S is the corresponding label of x_i^S; the label space of source domain Y_S contains different k types of health states, i.e., $Y_S = \{1, 2, 3, ..., k\}$; $P_S(X)$ is the marginal distribution of X^S. Similarly, the target domain $X^T = \{x_i^T, y_i^T\}_{i=1}^{N_T}$ contains N_T unlabeled samples and its label space Y_T contains the same k types of health states. It is important to note that the marginal distribution of target domain $P_T(X)$ is different from $P_S(X)$, however, the knowledge of source domain could be transferred to the target domain to facilitate its training [28], [29].

In this article, the partial transfer fault diagnosis model of rotating machinery is developed, in which Y_T is contained in Y_S, i.e., $Y_T \subseteq Y_S$. Furthermore, considering that the normal state is easy to collect in real scenarios of industry, it is assumed that both Y_S and Y_T should contain the normal state.

III. The Proposed Method

The proposed clustering-guided novel unsupervised domain adversarial network mainly contains four parts: construction of novel unsupervised domain adversarial network, design of embedded clustering strategy, definition of joint optimization objective function, and the general framework for the proposed method.

A. Construction of Novel Unsupervised Domain Adversarial Network

In this paper, to remove domain-specific interference information within the network and further enhance alignment between two domains, DSBN [30] is applied to construct the novel unsupervised domain adversarial network shown in Fig. 1, consisting of a feature generator, a domain discriminator and a classifier. DSBN uses two independent branches of BN, and each sample chooses one of the branches according to its domain in the network training phase [31]. More specifically, let x_i^D denote the input sample of the domain adversarial network, where $D = \{S, T\}$, (γ_S , β_S) denotes the domain-specific parameters of source domain, (γ_T , β_T) denotes the domain-specific parameters of target domain, and then DSBN layer can be expressed as

$$\mathrm{DSBN}_D(x_i^D; \gamma_D, \beta_D) = \gamma_D \cdot \hat{x}_i^D + \beta_D \tag{1}$$

$$\hat{x}_i^D = \frac{x_i^D - E[x_i^D]}{\sqrt{Var[x_i^D]}} \tag{2}$$

where $E[x_i^D]$ and $Var[x_i^D]$ represent the mean value and standard deviation of x_i^D, receptively. Before training phase, source and target samples are firstly simply split into different mini-batches. During the training phase, DSBN separately calculates each domain's mean and variance. At the testing

process, each domain's mean and variance are used for the samples in the corresponding domain.

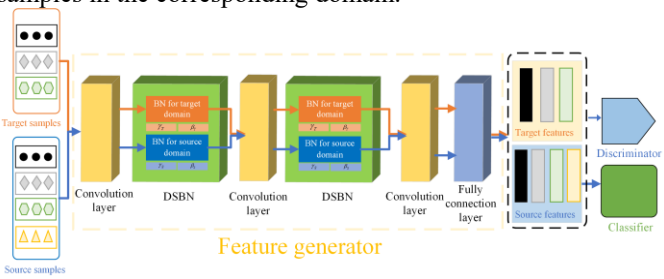


Fig. 1. Construction of novel unsupervised domain adversarial network using DSBN.

B. Design of Embedded Clustering Strategy

In this paper, in order to encourage feature generator to learn tightly clustered target-domain features and confirm that each cluster can be assigned to a unique class from the source domains in partial domain adaptation process, the following Kullback-Leibler divergence [32] is chosen as the embedded clustering loss to help soft assignment get close to the auxiliary target distribution, expressed as

$$\mathcal{L}_c = \mathrm{KL}(P\|Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{3}$$

where $\mathcal{L}_c$ is the embedded clustering loss, Q and P are the soft assignment and auxiliary distributions, respectively, $q_{ij}(q_{ij} \in Q)$ denotes the probability of target sample i being assigned to cluster j, $p_{ij}(p_{ij} \in P)$ is a predefined high-confident-assignment target distribution.

Assuming that the extracted feature of target domain $f(x_i^T)$ contains k clusters with each centroid of $\{\mu_j\}_{j=1}^k$ obtained by standard k-means technique, since we focus on the unsupervised partial transfer learning task, the labels of target samples are unavailable and a sub-set space of source domain labels are used instead. Therefore, it is hard to determine the value of k and calculate the initial cluster centroids of target domain. Here, we design a simple strategy to augment the target domain with the help of a portion of source-domain samples. For example, for each batch, we randomly take 60% as target samples while 40% as source samples (without using labels) to guarantee that the label space of augmented target domain is the same as the source domain. In this way, the partial domain adaptation problem is shift to the pseudo domain adaptation problem, naturally, and k is equal to the number of source classes. Besides, the embedded clustering strategy could make it difficult to distinguish the source and target samples, which can form a softer boundary for target set trained by the discriminator to reduce the effects of negative transfer.

When calculating q_{ij}, the student's $t-distribution$ [33] is used as an embedded kernel to measure the similarity between $f(x_i^T)$ and μ_j. Clearly, the closer to the cluster centroid μ_j, the higher the probability q_{ij}:

$$q_{ij} = \frac{(1+Sim(f(x_i^T),\mu_j)/\alpha)^{-\frac{\alpha+1}{2}}}{\sum_j (1+Sim(f(x_i^T),\mu_j)/\alpha)^{-\frac{\alpha+1}{2}}} \tag{4}$$

$$Sim(f(x_i^T),\mu_j) = \left\| f(x_i^T) - \mu_j \right\|^2 \tag{5}$$

where α is the degree of freedom of student's $t-distribution$, $Sim(\bullet)$ is Euclidean distance measurement to calculate the similarity of each $f(x_i^T)$ and μ_j.

In order to normalize the loss contribution of each centroid and prevent the hidden feature space from being distorted by large clusters, during our experiments, we compute $p_{ij}(p_{ij} \in P)$ by first raising q_{ij} to the second power and then normalizing by frequency each cluster, which can be defined as follows,

$$p_{ij} = \frac{q_{ij}^2 / \sum_i q_{ij}}{\sum_j (q_{ij}^2 / \sum_i q_{ij})} \tag{6}$$

where $\sum_i q_{ij}$ means soft cluster frequencies.

C. Definition of Joint Optimization Objective Function

In this paper, a joint optimization objective function is defined to balance different losses to improve the training and diagnosis performance, including embedded clustering loss, weighted domain classification loss, and weighted domain adaptation adversarial loss, expressed as:

$$\mathcal{L} = \lambda_c \mathcal{L}_c + \mathcal{L}_{cls} + \lambda_{adv} \mathcal{L}_{adv} \tag{7}$$

$$\mathcal{L}_{cls} = -\mathbb{E}_{(x_i^s, y_i^s)\in(X^S, Y^S)} \gamma_{y_i^s} \sum_{k=0}^{k-1} \mathrm{I}_{[y_i^s=k]} \log[C(G(x_i^s))] \tag{8}$$

$$\begin{aligned} \mathcal{L}_{adv} &= \mathbb{E}_{x_i^s \sim P_S(X)} \gamma_{y_i^s} \log[D(G(x_i^s))] \\ &+ \mathbb{E}_{x_i^t \sim P_T(X)} \log[D(1-G(x_i^s))] \end{aligned} \tag{9}$$

$$\gamma = \frac{1}{n_t} \sum_{i=1}^{N_t} y_i^t \tag{10}$$

$$\lambda_{adv} = \frac{1-e^{-\beta\rho}}{1+e^{-\beta\rho}} \tag{11}$$

where $\mathcal{L}$ is the joint optimization objective function, $\mathcal{L}_c$ represents the embedded clustering loss, which is given in Eq. (1), $\mathcal{L}_{cls}$ represents the weighted classification loss, $\mathcal{L}_{adv}$ represents the weighted domain adaptation adversarial loss, λ_c and λ_{adv} are the specifying parameters in the joint optimization objective function to balance the different losses, k is the total number of possible classes, $\mathrm{I}_{[\cdot]}$ is the indicator function, $G(\cdot)$ and $C(\cdot)$ are assumed to the generator and classifier, $D(\cdot)$ means partial adversarial domain discriminator, $\hat{y}_i^t$ is the predicted label of target sample i through Softmax function, γ is a k-dimensional weight

vector, which indicates the contribution of each source class [1]. $\gamma_{y_i^s}$ is the class weight of source sample x_i^S, β denotes as a constant, as the epoch number increases, ρ changes from 0 to 1 and λ_{adv} is progressive increasing from 0 to 1 in the training phase [34]. Specially, it is noteworthy that the clustering loss λ_C is fixed as 0 at the initial training stage. When the number of training epochs reaches to the preset epoch value, λ_C is activated and selected as a nonnegative constant. The training procedure is presented in Algorithm 1.

Algorithm 1. Training procedure of the clustering-guided novel unsupervised domain adversarial network

Inputs: Labeled source-domain samples $X^S = \{x_i^S, y_i^S\}_{i=1}^{N_S}$, unlabeled target-domain samples Y_S, network architecture, balance factors λ_C and λ_{adv}, batch size m, maximum epoch, preset epoch.

Outputs: The trained unsupervised domain adversarial network.

Begin: Randomly initialize the training parameter, training an initial network, while λ_C was fixed as 0 (epoch < preset epoch).

While not converged **do**

For epoch = preset epoch to maximum epoch **do**

Draw random minibatch $\{x_i^S, P_S(x_i^S)\}_{i=1}^m$, $\{X_i^T\}_{i=1}^m$

Compute $\mathcal{L}_{cls}$ by **Eq. (8)**.

Compute $\mathcal{L}_c$ by **Eq. (1)**.

Compute $\mathcal{L}_{adv}$ by **Eq. (9)**.

Optimize $\mathcal{L}$ with respect to **Eq. (7)**.

End for

D. The Overall Framework of The Proposed Method

Fig. 2 illustrates the framework of the proposed method, and the sequential diagnosis steps are given below.

Step 1: Divide the collected signals into source and target domains. Specifically, source and target domains both contain the normal state, while the target domain only contains partial fault states.

Step 2: Propose the clustering-guided novel unsupervised domain adversarial network for partial transfer fault diagnosis.

Step 2.1: Construct the novel unsupervised domain adversarial network. Apply DSBN to remove domain-specific information within the network and enhance alignment between source and target domain.

Step 2.2: Design the embedded clustering strategy to help learn tightly clustered target-domain features to suppress negative transfer in partial domain adaptation process.

Step 2.3: Define the joint optimization objective function to balance different losses to improve the training and diagnosis performance.

Step 3: Use source-domain and target-domain training samples to train and optimize the proposed network using Adam algorithm.

Step 4: Feed the target-domain testing samples into the trained network for validation.

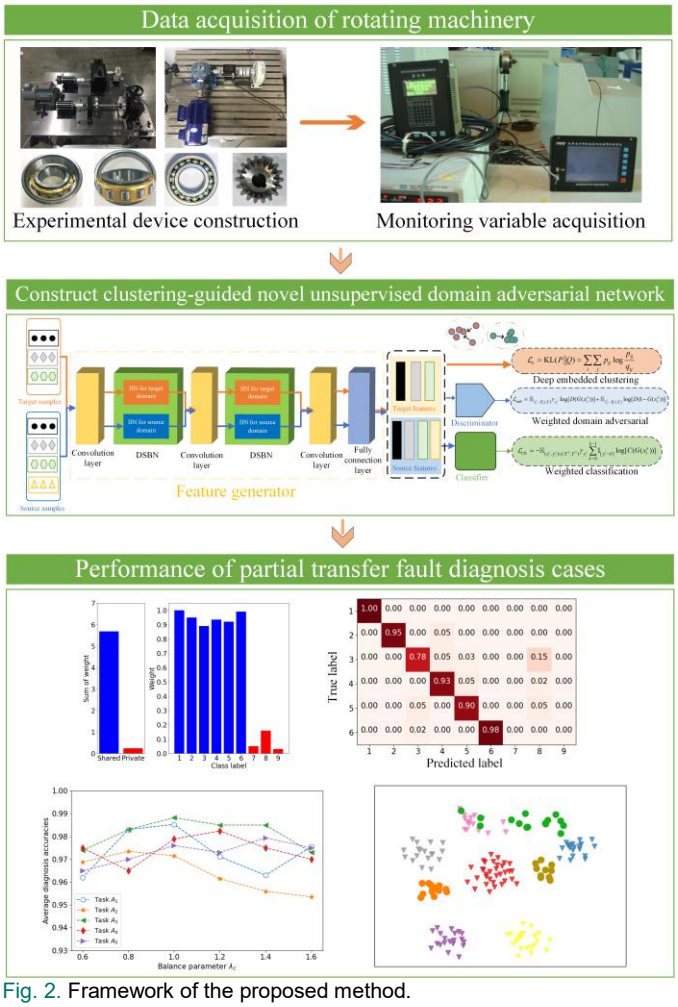


Fig. 2. Framework of the proposed method.

IV. Case Study

In this section, two cases of rotating machinery are implemented to validate the effectiveness and superiority of the proposed method.

1. Case I: Partial Transfer Fault Diagnosis Experiment on Bevel Gearbox

A. Data description of bevel gearbox

In Case I, the experimental acceleration signals are collected from a custom-built bevel gearbox fault simulation test rig shown in Fig. 3 [35]. Nine health states are simulated, including normal state, inner race fault states with different degrees, outer race fault states with different degrees and gear crack states with different degrees, as shown in Table I and Fig. 4. Two rotation speeds of the motor are 900 and 1200 rpm, respectively, and the sampling frequency is 10240 Hz. Each state from D_S and D_T has 210 samples and each sample consists of 1024 points. 1512 (168*9) samples are used as the total training samples while 378 (42*9) samples are used as the testing samples. The detailed settings of partial transfer learning tasks in Case I are presented in Table II.

TABLE I
DETAILS ABOUT THE NINE HEALTH STATES IN CASE I

Class label	Fault location	Fault size (mm)
1	Normal	\
2	Inner race fault	0.2
3	Inner race fault	0.4
4	Inner race fault	0.6
5	Inner race fault	0.8
6	Outer race fault	0.2
7	Outer race fault	0.4
8	Gear crack	1.6
9	Gear crack	0.6

TABLE II
THE SETTING OF THE NINE PARTIAL TRANSFER LEARNING TASKS IN **CASE I**

Task names	Transfer speeds (rpm)	Source classes Y_S	Target classes Y_T
$\mathbf{A_1}$	900 –> 1200	1,2,3,4,5,6,7,8,9	1,2,3,4,5,6,7,8,9
$\mathbf{A_2}$	900 –> 1200	1,2,3,4,5,6,7,8,9	1,2,3,4,5,6,7,8
$\mathbf{A_3}$	900 –> 1200	1,2,3,4,5,6,7,8,9	1,2,3,4,5,6,7
$\mathbf{A_4}$	900 –> 1200	1,2,3,4,5,6,7,8,9	1,2,3,4,5,6
$\mathbf{A_5}$	900 –> 1200	1,2,3,4,5,6,7,8,9	1,2,3,4
$\mathbf{A_6}$	1200 –> 900	1,2,3,4,5,6,7,8,9	1,2,3,4,5,6,7,8,9
$\mathbf{A_7}$	1200 –> 900	1,2,3,4,5,6,7,8,9	1,2,3,4,5,6,7,8
$\mathbf{A_8}$	1200 –> 900	1,2,3,4,5,6,7,8,9	1,2,3,4,5,6,7
$\mathbf{A_9}$	1200 –> 900	1,2,3,4,5,6,7,8,9	1,2,3,4

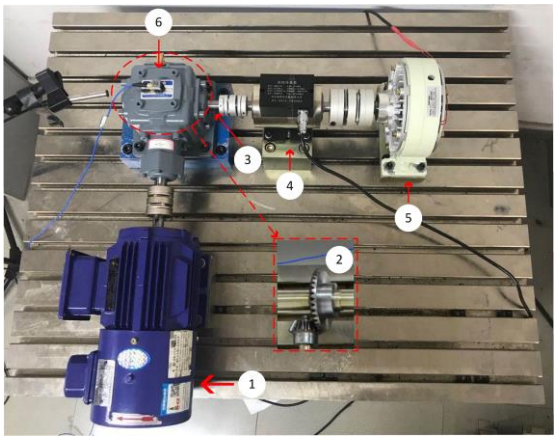


Fig. 3. Bevel gearbox fault simulation test rig in **Case I**. ① Motor; ② Bevel gears; ③ Testing rolling bear; ④ Torque sensor; ⑤ Loading; ⑥ Piezo-electric accelerometer.

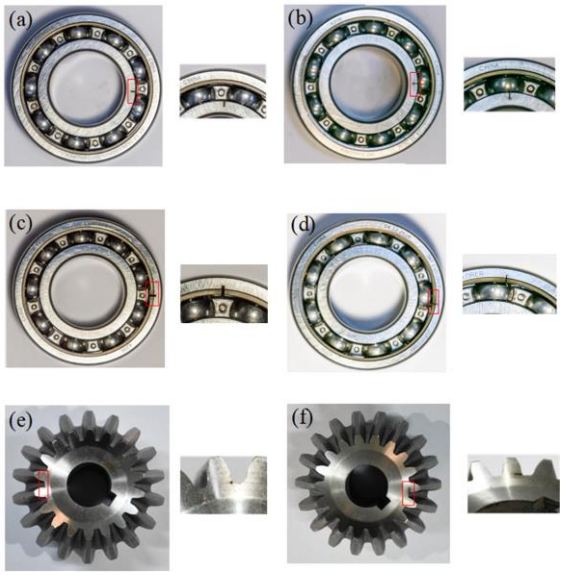


Fig. 4. Some pictures of the tested gears and bearings in **Case I**: (a) Label 2; (b) Label 3; (c) Label 6; (d) Label 7; (e) Label 8; (f) Label 9.

B. Comparisons with other partial transfer learning methods in Case I

To test the effectiveness of the method, some existing popular partial transfer learning methods are used for comparison, which include domain adaptation with MMD (DA-MMD) [1], partial adversarial domain adaptation (PADA) [36], importance weighted adversarial network (IWAN) [37], and CNN. Besides, Proposed method without DSBN (replace DSBN with BN in the proposed method) and Proposed method without clustering (remove the embedded clustering strategy in the proposed method) are also added in the experiments.

The partial transfer diagnosis performance of the proposed method for nine tasks is compared with the other six methods in Case I. To reduce the random error, each partial transfer diagnosis result is averaged over ten repeated validations. Specially, the average partial transfer diagnosis accuracies of each method reported after the last epoch (**denoted as L**) are shown in Table III. In addition, to quantitatively evaluate the negative transfer effect, maximum accuracies during the whole iterative process (**denoted as M**) are also listed in Table III.

In view of the comparison results, it can be found that (1) the **L** of the proposed method in the nine transfer tasks are 94.14%, 93.79%, 96.55%, 95.52%, 99.31%, 94.83%, 93.10%, 95.52%, 98.97%,98.97%, respectively. The proposed method achieves outstanding diagnosis performance compared with the other four transfer learning methods (CNN, DA-MMD, PADA, IWAN) on the classical transfer diagnosis tasks ($\mathbf{A_1}$, $\mathbf{A_6}$) and partial transfer diagnosis tasks. (2) Compared with Proposed the method without clustering, the total average of the proposed method has improved 3.96%, i.e., 95.75%, which shows that the embedded clustering strategy can suppress negative transfer in partial domain adaptation process and improve the diagnosis performance. Besides, it is noteworthy that the transferability of DA-MMD, PADA and IWAN have been degraded largely for Task A_8 because of negative transfer effect of private source samples (source samples whose label is not shared in the target). (3) Similarly, because DSBN can remove domain-specific information to enhance alignment between the two domains to some degree, the total average of the proposed method has improved 1.23% compared with Proposed method without DSBN.

TABLE III
TESTING DIAGNOSIS ACCURACIES IN DIFFERENT TRANSFER TASKS IN **CASE I**

Methods	Task names									Total average
	A_1	**A_2**	**A_3**	**A_4**	**A_5**	**A_6**	**A_7**	**A_8**	**A_9**	
CNN	L: 84.48 M: 86.55	L: 83.10 M: 84.83	L: 82.07 M: 83.37	L: 80.34 M: 82.07	L: 80.69 M: 83.10	L: 85.52 M: 86.55	L: 85.17 M: 88.28	L: 82.07 M: 89.31	L: 93.10 M: 93.45	84.06
DA-MMD	L:96.55 M: 96.55	L: 95.17 M: 96.21	L: 90.34 M: 94.48	L: 91.72 M: 95.52	L: 85.86 M: 90.34	L: 93.10 M: 94.83	L: 92.76 M: 94.14	L: 92.76 M: 95.52	L: 95.86 M: 97.93	92.68
PADA	L: 93.79 M: 94.48	L: 78.28 M: 83.10	L: 92.07 M: 93.45	L: 84.14 M: 91.72	L: 89.66 M: 95.86	L: 93.45 M: 95.86	L: 91.38 M: 94.83	L: 86.55 M: 94.48	L: 98.28 M: 98.97	89.73
IWAN	L: 95.17 M: 97.24	L: 90.34 M: 93.79	L: 93.10 M: 95.86	L: 92.41 M: 95.52	L: 94.83 M: 96.55	L: 92.76 M: 96.21	L: 91.03 M: 93.45	L: 89.66 M: 92.07	L: 96.21 M: 98.97	92.83
Proposed method without DSBN	L:93.79 M: 95.52	L: 90.69 M: 93.45	L: 95.17 M: 96.55	L: 94.14 M: 95.17	L: 97.59 M: 98.62	L: 94.83 M: 94.83	L: 92.07 M: 93.79	L: 94.48 M: 96.90	L: 97.93 M: 98.62	94.52
Proposed method without clustering	L:93.79 M: 95.29	L: 83.81 M: 87.62	L: 94.12 M: 96.76	L: 88.57 M: 93.57	L: 92.38 M: 96.19	L: 94.52 M: 96.43	L: 91.67 M: 94.52	L: 89.05 M: 95.24	L: 98.28 M: 98.62	91.79
Proposed method	**L: 94.14 M: 95.17**	**L: 93.79 M: 95.52**	**L: 96.55 M: 97.24**	**L: 95.52 M: 96.90**	**L: 99.31 M: 99.31**	**L: 94.83 M: 95.52**	**L: 93.10 M: 94.48**	**L: 95.52 M: 97.24**	**L: 98.97 M: 98.97**	**95.75**

Remarks: L represents the accuracies after the last epoch; M represents the maximum accuracies during the whole iterative process; Total average represents the average of "L" in the nine tasks.

In order to show the class weights assignment in shared and private classes on partial transfer diagnosis tasks, Fig. 5 shows the histograms of class weights and confusion matrixes learned by PADA, Proposed method without clustering, and the proposed method on Task A_4. For proposed method in Fig. 5 (c), shared classes can be given a much higher weight than private classes (the source classes not included in target domain label space) in the training process. Specifically, the class weights of the three private classes are only 0.05, 0.16 and 0.03, receptively. However, for Proposed method without clustering, the class weights of the private classes remain higher. The principal reason is that the embedded clustering strategy can tightly cluster the target-domain features in the same class while widen the relative distance between target-domain samples and private source-domain samples, thereby resulting in a great improvement of diagnosis accuracies. For PADA in Fig. 5 (a), the class weights of the three private classes are 0.26, 0.34 and 0.1, receptively. The private classes weights are still substantially larger than the proposed method. Furthermore, it can be found from the confusion matrix that PADA and Proposed method without clustering have categorized target-domain samples of Label-3, Label-7 and Lable-8 into one class, which greatly affects the transferability and even causes negative transfer because of the private source-domain samples. In Fig. 5 (c), the confusion matrix of the proposed method has best classification performance, which also demonstrates that the embedded clustering strategy can reduce the number of the target samples misclassified as private classes.

To visualize the effectiveness of embedded clustering strategy in the proposed method, t-SNE [38] is used in the second convolution layer to map the feature representations. For Task A_4, the visualizations of four methods are shown in Fig. 6 and Fig. 7. Some important points can be concluded: (1) in Fig. 6 (d), the proposed method obtains the best clustering performance for the transferable shared features. The same class of the two domains are completely merged into one cluster and each cluster has a clear boundary, which indicates that the embedded clustering strategy can effectively learn six tightly clustered target-domain features; In Fig. 7 (d), the two private classes, i.e., label-8 and label-9, are well separated from the shared classes, besides, only a few samples of label-7 are overlapped with the target cluster of label-3, which outperforms the other three methods. A possible reason could be that the DSBN can remove the domain-specific information, leading a better adapt to the target domain. (2) some target-domain samples in Proposed method without clustering and IWAN are too scattered to form a clear cluster boundary with other classes features, causing poor classification performance, such as label-2 and label-4; Besides, the same class of the source and target domains are not fully merged into one cluster, such as label-1 in Fig. 6 (c); (3) For DA-MMD in Fig. 7 (a), there are a significant overlapping between different classes in the two domains. The private source-domain features seriously obstruct the domain adaptation process, which leads to poor clustering performance.

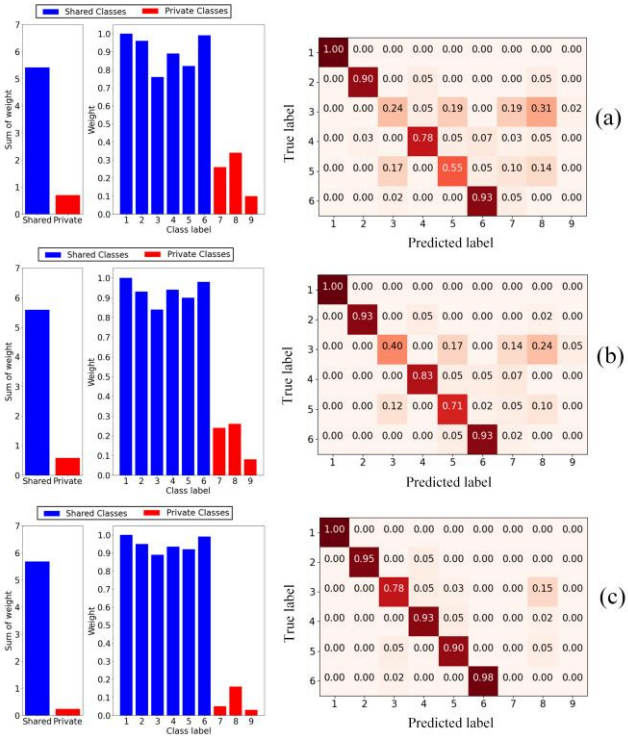


Fig. 5. Histograms of class weights and confusion matrixes on **Task A_4**. (a) PADA; (b) Proposed method without clustering; (c) The proposed method.

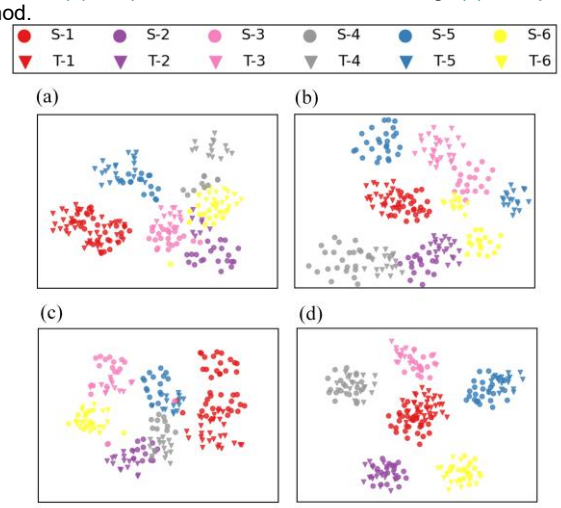


Fig. 6. Feature visualizations of the six transferable shared features in **Task A_4**: (a) DA-MMD; (b) Proposed method without clustering; (c) IWAN; (d) The proposed method. "T" is the target-domain features. "S" is the source-domain features.

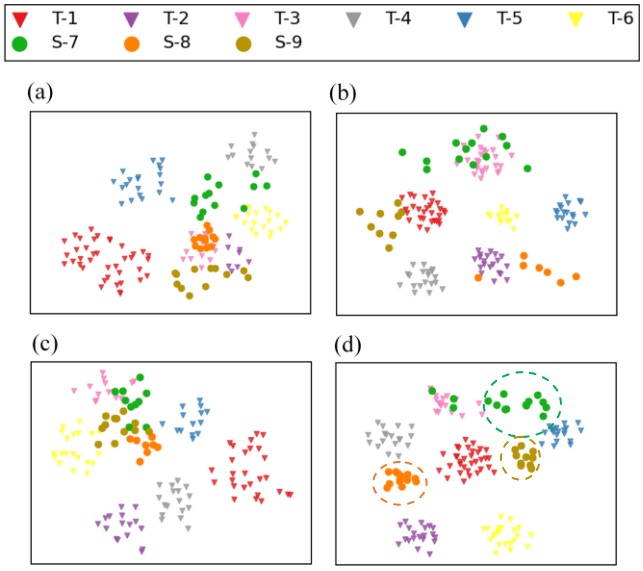


Fig. 7. Feature visualizations of the target-domain features and private source-domain features in **Task A_4**: (a) DA-MMD; (b) Proposed method without clustering; (c) IWAN; (d) The proposed method.

2. Case II: Partial Transfer Fault Diagnosis Experiment on Rolling Bearing Dataset

C. Data Description of Rolling Bearing

In this case, the acceleration signals of rolling bearing data are collected from QPZZ-II fault simulation test rig displayed in Fig. 8 [39]. The label space of source domain D_S contains four health states: normal (N), inner race fault (IF), rolling race fault (RF) and outer race fault (OF). The local defects of fault bearings are given in Fig. 9. Five partial transfer learning tasks are taken and the detailed setting is given in Table IV. Four rotating speeds are collected from the motor drive, i.e., 900, 1000, 1300 and 1500 rpm, respectively. In this case study, each state from D_S and D_T has 290 data samples and each sample contains 1024 sampling points. 928 (232*4) samples are used as the training samples while 232 (58*4) samples are used as the testing samples, respectively.

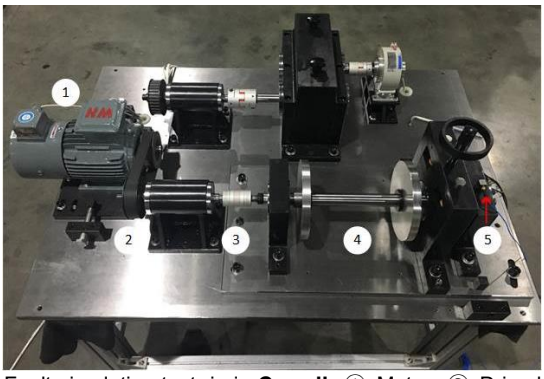


Fig. 8. Fault simulation test rig in **Case II**. ① Motor; ② Drive belt; ③ Coupling; ④ Revolving shaft; ⑤ Accelerometer.

TABLE IV
THE SETTING OF THE FIVE PARTIAL TRANSFER LEARNING TASKS IN **CASE II**

Task names	Transfer speeds (rpm)	Source classes Y_S	Target classes Y_T
B_1	900 –> 1500	N, IF, OF, RF	N, IF, OF, RF
B_2	900 –> 1500	N, IF, OF, RF	N, IF, OF
B_3	1300 –> 1000	N, IF, OF, RF	N, IF, OF, RF
B_4	1300 –> 1000	N, IF, OF, RF	N, IF, OF
B_5	1300 –> 1000	N, IF, OF, RF	N, IF

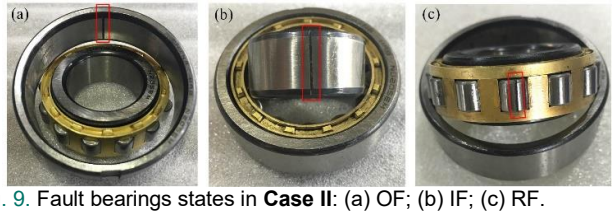


Fig. 9. Fault bearings states in **Case II**: (a) OF; (b) IF; (c) RF.

D. Result Analysis

In Table V, we compare the partial transfer diagnosis performance of seven different methods. Based on the results, the proposed method achieves the highest **L** among the compared methods, which are 98.53%, 97.35%, 98.24%, 99.41%, and 97.94% in the five transfer tasks, respectively. Moreover, these comparison results further prove the effectiveness and superiority of the embedded clustering strategy and DSBN in the proposed method. For example, the accuracies after the last epoch of the proposed method have improved 1.18%, 1.76%, 1.18% compared with Proposed method without DSBN in transfer Tasks B_1, B_2, B_5, respectively.

TABLE V
TESTING DIAGNOSIS ACCURACIES IN DIFFERENT TRANSFER TASKS IN **CASE II**

Methods	Task names					Total average
	B_1	**B_2**	**B_3**	**B_4**	**B_5**	
CNN	L: 70.58 M: 72.35	L: 57.65 M: 58.82	L: 74.11 M: 78.24	L: 84.41 M: 85.29	L: 73.24 M: 77.06	72.00
DA-MMD	L: 93.53 M: 94.12	L: 77.41 M: 87.94	L: 98.53 M: 98.53	L: 73.82 M: 84.71	L: 52.94 M: 94.71	79.25
PADA	L: 58.82 M: 70.29	L: 40.88 M: 76.47	L: 97.35 M: 98.24	L: 99.12 M: 99.12	L: 83.53 M: 94.71	75.94
IWAN	L: 84.71 M: 93.53	L: 69.12 M: 86.18	L: 94.12 M: 97.65	L: 98.24 M: 99.12	L: 93.24 M: 97.35	87.89
Proposed method without DSBN	L: 97.35 M: 98.24	L: 95.59 M: 96.76	L: 98.53 M: 98.82	L: 99.12 M: 99.71	L: 96.76 M: 98.53	97.47

Proposed method without clustering	L: 75.29 M: 78.24	L: 80.29 M: 85.00	L: 97.65 M: 98.53	L: 99.12 M: 99.41	L: 91.18 M: 95.88	88.71
Proposed method	**L: 98.53 M: 99.12**	**L: 97.35 M: 97.94**	**L: 98.24 M: 99.12**	**L: 99.41 M: 99.41**	**L: 97.94 M: 99.12**	**98.29**

Remarks: L represents the accuracies after the last epoch; M represents the maximum accuracies during the whole iterative process; Total average represents the average of "L" in the five tasks.

For more details, Fig. 10 shows the diagnosis accuracies of the compared methods by ten repeated validations in Task B_1. Each result is averaged by ten trials. The specific standard deviations of the six methods are 3.24%, 2.09%, 4.57%, 7.92%, 1.08%, and 0.84%, respectively. The proposed method obtains the minimum standard deviation, which demonstrates that the proposed method has good stability. Fig. 11 is the confusion matrices of the tenth validation of DA-MMD, PADA, IWAN and the proposed method. Moreover, Fig. 11 gives detailed information about misclassified states.

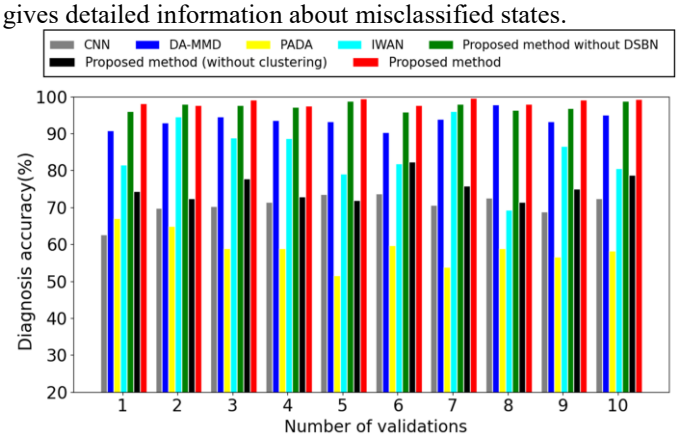


Fig. 10. Diagnosis accuracies of compared methods by ten repeated validations in **Task B_1**.

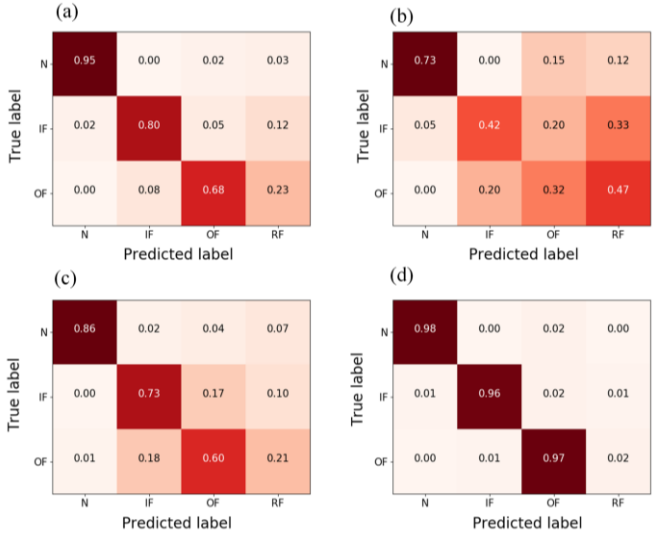


Fig. 11. Confusion matrices for the tenth repeated validation of **Task B_2**. (a) DA-MMD; (b) PADA; (c) IWAN; (d) The proposed method.

The above tasks are executed on a computer equipped with GeForce GTX 1650Ti, Intel Core i5-10400 and Pytorch 1.2 platform. The setting of the network structure is listed in Table VI, and the training parameters are given below.

The batch size is chosen from 64 to 256, initial learning rate is chosen from $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$, epoch number is 300, preset epoch is 20, the Adam optimization algorithm is chosen for training. Besides, we empirically find that setting the preset epoch to 20 leads to a good initialization point for the embedded clustering loss.

Furthermore, the influence of the balance parameter λ_c in the total optimization objective on the diagnosis performance of partial transfer learning tasks is studied. The balance parameter λ_c is searched from {0.6, 0.8, 1.0, 1.2, 1.4, 1.6}. Form the Fig. 12, as long as the balance parameter λ_c is within a reasonable range, the diagnosis performance of partial transfer learning tasks does not change significantly, which demonstrates that the proposed method has a robust convergence for the balance parameter λ_c.

TABLE VI
THE ARCHITECTURE OF THE PROPOSED NETWORK

Network modules	Layers	Parameters	Operations
Feature extractor	Input	1024 × 1	/
	Convolutional layer	Kernels: 15 × 1, Channel:16	/
	DSBN	/	/
	Max pooling layer	Kernels: 2, stride: 2	Rectified Linear Unit
	Convolutional layer	Kernels: 3 × 1, Channel:32	/
	DSBN	/	/
	Max pooling layer	Kernels: 2, stride: 2	/
	Fully connection layer	Output: 256 × 1	Rectified Linear Unit
Classifier	Fully connection layer	Output: k × 1	Softmax
Domain Discriminator	Fully connection layer	Output:128× 1	Rectified Linear Unit, Dropout 0.3
	Fully connection layer	Output:64× 1	Rectified Linear Unit, Dropout 0.3
	Fully connection layer	Output:2× 1	

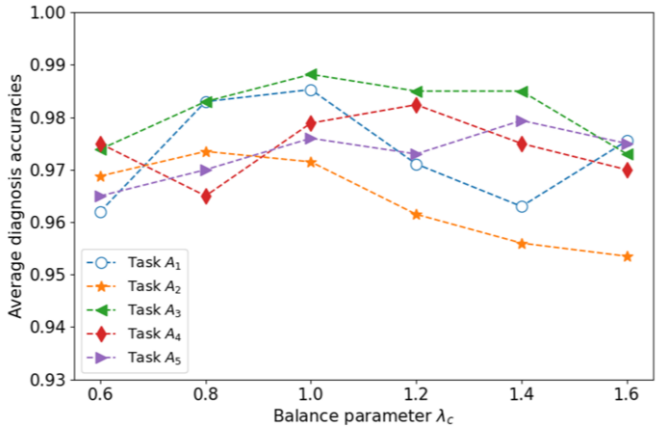


Fig. 12. Influence of λ_c on the five transfer tasks.

V. CONCLUSIONS

In order to alleviate the effects of negative transfer and improve the discriminative of extracted target-domain in partial transfer fault diagnosis of rotating machinery, we propose a clustering-guided novel unsupervised domain adversarial network. Novel unsupervised domain adversarial network is constructed and DSBN is applied to remove domain-specific information to enhance alignment between two domains. Embedded clustering strategy is designed to learn tightly clustered target-domain features to suppress negative transfer in partial domain adaptation process. Finally,

a joint optimization objective function is defined to balance different losses to improve the training and diagnosis performance.

A total of 14 partial transfer fault diagnosis tasks from two experimental datasets have verified that the proposed method can better reduce the influence of negative transfer in partial domain adaptation process, and achieve higher diagnosis accuracy compared to existing methods, which proves its potential applications in solving unsupervised partial transfer fault diagnosis problems. Future work will explore on how to accurately diagnosis fault modes under open-set domain adaptation and how to further design more robust embedded clustering strategies.

REFERENCES

[1] Z. Zhao et al., "Applications of Unsupervised Deep Transfer Learning to Intelligent Fault Diagnosis: A Survey and Comparative Study," *IEEE Transactions on Instrumentation and Measurement,* vol. 70, Art no. 3525828, Sep. 2021.

[2] W. Wang, Y. Lei, T. Yan, N. Li, and A. Nandi, "Residual Convolution Long Short-Term Memory Network for Machines Remaining Useful Life Prediction and Uncertainty Quantification," *Journal of Dynamics, Monitoring and Diagnostics,* vol. 1, no. 1, pp. 2-8, Nov. 2021.

[3] J. Sun, J. Wen, C. Yuan, Z. Liu, and Q. Xiao, "Bearing Fault Diagnosis Based on Multiple Transformation Domain Fusion and Improved Residual Dense Networks," *IEEE Sensors Journal,* vol. 22, no. 2, pp. 1541-1551, Jan. 2022.

[4] R. Wang, Z. Chen, S. Zhang, and W. Li, "Dual-Attention Generative Adversarial Networks for Fault Diagnosis Under the Class-Imbalanced Conditions," *IEEE Sensors Journal,* vol. 22, no. 2, pp. 1474-1485, Jan. 2022.

[5] H. Shao, J. Lin, L. Zhang, D. Galar, and U. Kumar, "A novel approach of multisensory fusion to collaborative fault diagnosis in maintenance," *Information Fusion*, vol. 74, pp. 65-76, Oct. 2021.

[6] C. Shen, X. Wang, D. Wang, Y. Li, J. Zhu and M. Gong, "Dynamic Joint Distribution Alignment Network for Bearing Fault Diagnosis Under Variable Working Conditions," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-13, Feb. 2021.

[7] R. Yan, F. Shen, C. Sun, and X. Chen, "Knowledge Transfer for Rotary Machine Fault Diagnosis," *IEEE Sensors Journal,* vol. 20, no. 15, pp. 8374-8393, Aug. 2020.

[8] X. Li, W. Zhang, H. Ma, Z. Luo, and X. Li, "Deep learning-based adversarial multi-classifier optimization for cross-domain machinery fault diagnostics," *Journal of Manufacturing Systems*, vol. 55, pp. 334-347, Apr. 2020.

[9] W. H. Li *et al.*, "A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges," *Mechanical Systems and Signal Processing,* vol. 167, Art no. 108487, Mar. 2022.

[10] Z. He, H. Shao, Z. Ding, H. Jiang, and J. Cheng, "Modified Deep Autoencoder Driven by Multisource Parameters for Fault Transfer Prognosis of Aeroengine," *IEEE Transactions on Industrial Electronics,* vol. 69, no. 1, pp. 845-855, Jan. 2022.

[11] Z. He, H. Shao, X. Zhong, and X. Zhao, "Ensemble transfer CNNs driven by multi-channel signals for fault diagnosis of rotating machinery cross working conditions," *Knowledge-Based Systems,* vol. 207, Art no. 106396, Nov. 2020.

[12] 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 Cybernetics-Systems*, vol. 49, no. 1, pp. 136-144, Jan. 2019.

[13] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, "Deep Model Based Domain Adaptation for Fault Diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 3, pp. 2296-2305, Mar. 2017.

[14] B. Yang, Y. G. Lei, F. Jia, and S. B. Xing, "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," *Mechanical Systems and Signal Processing,* vol. 122, pp. 692-706, May. 2019.

[15] J. Jiao, M. Zhao, J. Lin, and K. Liang, "Residual joint adaptation adversarial network for intelligent transfer fault diagnosis," *Mechanical Systems and Signal Processing,* vol. 145, Art no. 106962, Nov. 2020.

[16] D. Xu, Y. Li, Y. Song, L. Jia and Y. Liu, "IFDS: An Intelligent Fault Diagnosis System With Multisource Unsupervised Domain Adaptation for Different Working Conditions," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-10, Nov. 2021.

[17] Z. He, H. Shao, L. Jing, J. Cheng, and Y. Yang, "Transfer fault diagnosis of bearing installed in different machines using enhanced deep auto-encoder," *Measurement,* vol. 152, Art no. 107393, Feb. 2020.

[18] J. Zhang, Z. Ding, W. Li, and P. Ogunbona, "Importance Weighted Adversarial Nets for Partia Domain Adaptation," in *31st IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), 2018, pp. 8156-8164.

[19] S. Li et al., "Deep Residual Correction Network for Partial Domain Adaptation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 7, pp. 2329-2344, Jul. 2021.

[20] J. Hu, H. Tuo, C. Wang, H. Zhong, H. Pan, and Z. Jing, "Unsupervised satellite image classification based on partial transfer learning," *Aerospace Systems*, vol. 3, no. 1, pp. 21-28, Mar. 2020.

[21] P. Li, D. Z. Zhang, P. Chen, X. Liu, and A. Wulamu, "Multi-Adversarial Partial Transfer Learning With Object-Level Attention Mechanism for Unsupervised Remote Sensing Scene Classification," *IEEE Access*, vol. 8, pp. 56650-56665, Mar. 2020.

[22] Y. Tang, X. Wei, B. Zhao, and S. Huang, "QBox: Partial Transfer Learning With Active Querying for Object Detection," *IEEE Transactions on Neural Networks and Learning Systems*. Sep. 2021, doi: 10.1109/TNNLS.2021.3111621.

[23] X. Li, W. Zhang, H. Ma, Z. Luo, and X. Li, "Partial transfer learning in machinery cross-domain fault diagnostics using class-weighted adversarial networks," *Neural Networks,* vol. 129, pp. 313-322, Sep. 2020.

[24] X. Li and W. Zhang, "Deep Learning-Based Partial Domain Adaptation Method on Intelligent Machinery Fault Diagnostics," *IEEE Transactions on Industrial Electronics,* vol. 68, no. 5, pp. 4351-4361, May. 2021.

[25] Z. Liu, B. Lu, H. Wei, L. Chen, X. Li, and C. Wang, "A Stacked Auto-Encoder Based Partial Adversarial Domain Adaptation Model for Intelligent Fault Diagnosis of Rotating Machines," *IEEE Transactions on Industrial Informatics,* vol. 17, no. 10, pp. 6798-6809, Oct. 2021.

[26] Y. Deng, D. Huang, S. Du, G. Li, C. Zhao, and J. Lv, "A double-layer attention based adversarial network for partial transfer learning in machinery fault diagnosis," *Computers in Industry,* vol. 127, Art no. 103399, May. 2021.

[27] S. Udmale, S. Singh, R. Singh, and A. Sangaiah, "Multi-Fault Bearing Classification Using Sensors and ConvNet-Based Transfer Learning Approach," *IEEE Sensors Journal,* vol. 20, no. 3, pp. 1433-1444, Feb. 2020.

[28] J. Miao, J. Wang, D. Zhang and Q. Miao, "Improved Generative Adversarial Network for Rotating Component Fault Diagnosis in Scenarios With Extremely Limited Data," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-13, Art no. 3500213, Nov. 2022.

[29] H. Cao, H. Shao, X. Zhong, Q. Deng, X. Yang, and J. Xuan, "Unsupervised domain-share CNN for machine fault transfer diagnosis from steady speeds to time-varying speeds," *Journal of Manufacturing Systems,* vol. 62, pp. 186-198, Jan. 2022.

[30] K. Saito, D. Kim, S. Sclaroff, T. Darrell, and K. Saenko, "Semi-supervised Domain Adaptation via Minimax Entropy," in *IEEE/CVF International Conference on Computer Vision* (ICCV), 2019, pp. 8049-8057.

[31] W. Chang, T. You, S. Seo, S. Kwak, B. Han, and I. Soc, "Domain-Specific Batch Normalization for Unsupervised Domain Adaptation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), 2019, pp. 7346-7354.

[32] T. Lin, J. Lee, and W. Hsieh, "Robust mixture modeling using the skew t distribution," *Statistics and Computing,* vol. 17, no. 2, pp. 81-92, Jun. 2007.

[33] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised Deep Embedding for Clustering Analysis," in *33rd International Conference on Machine Learning*, 2016, pp. 478-487.

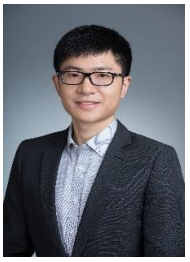
[34] Z. Zhao et al., "Unsupervised deep transfer learning for intelligent fault diagnosis: An open source and comparative study, " 2019. [Online]. Available: http://arxiv.org/pdf/1912.12528

[35] J. Wang et al., "Construction of a batch-normalized autoencoder network and its application in mechanical intelligent fault diagnosis,"

Measurement Science and Technology, vol. 30, no. 1, Art no. 015106, Jan. 2019.
[36] Z. Cao, L. Ma, M. Long, and J. Wang, "Partial Adversarial Domain Adaptation," in *15th European Conference on Computer Vision (ECCV)*, 2018, pp. 139-155.
[37] J. Zhang, Z. W. Ding, W. Q. Li, P. Ogunbona, "Importance Weighted Adversarial Nets for Partia Domain Adaptation," in *31st IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), 2018, pp. 8156-8164.
[38] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579-2605, Nov. 2008.
[39] Z. Di, H. Shao, and J. Xiang, "Ensemble deep transfer learning driven by multisensor signals for the fault diagnosis of bevel-gear cross-operation conditions," *Science China-Technological Sciences,* vol. 64, no. 3, pp. 481-492, Mar. 2021.

Hongru Cao received the B.S. degree in Mechanical and Electronic Engineering from Hebei Agricultural University, Baoding, China, in 2020. He is currently studying for the M.S. degree in Mechanical Engineering from Hunan University, Changsha, China. His research interests include unsupervised fault diagnosis, health management, and transfer learning.

Haidong Shao received the B.S. degree in Electrical Engineering and Automation and the Ph.D. degree in Vehicle Operation Engineering from Northwestern Polytechnical University, Xi'an, China, in 2013 and 2018, respectively. He is currently an Associate Professor in the College of Mechanical and Vehicle Engineering at Hunan University, Changsha, China. From 2019 to 2021, he was a Postdoctoral Fellow with the Division of Operation and Maintenance Engineering, Luleå University of Technology, Luleå, Sweden. From 2021 to 2022, he was a Visiting Scholar with the State Key Laboratory of Mechanical Transmissions, Chongqing University, Chongqing, China. His current research interests include fault diagnosis, and prognostics and health management.



Bin Liu received the B.S. degree in Automation from Zhejiang University, Hangzhou, China, in 2013, and the Ph.D. degree in Industrial Engineering from City University of Hong Kong, Hong Kong, China, in 2017. He is currently working as a lecturer in the Department of Management Science at University of Strathclyde, Glasgow, UK. Before that he was working as a Postdoctoral Fellow in University of Waterloo, Canada. His research interests include reliability and risk analysis, maintenance modeling, prognostic and health management, and decision making under uncertainty.

Baoping Cai (M'15) received the B.S. degree in Mechanical Design, Manufacturing, and Automation and the Ph.D. degree in Mechanical and Electronic Engineering from the China University of Petroleum, Qingdao, China, in 2006 and 2012, respectively. From 2013 to 2014, he was a Lecturer with the China University of Petroleum, where he is currently a Full Professor of Mechanical Engineering. His research interests include reliability engineering, fault diagnosis, risk analysis, and Bayesian networks methodology and application.

Junsheng Cheng received the B.S. degree in Mechanical and Electronic Engineering from Jilin University of Technology, Changchun, China, in 1991, the M.S. degree and the Ph.D. degree in Mechanical Engineering from Hunan University, Changsha, China, in 2000 and 2005, respectively. He is currently a Full Professor in College of Mechanical and Vehicle Engineering at Hunan University Changsha, China. His main research interests include mechanical fault diagnosis, signal processing, and vibration control.