Knowledge-Based Systems 279 (2023) 110960

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Duplex adversarial domain discriminative network for cross-domain partial transfer fault diagnosis

Fuqiang Liu [a,b,*], Wenlong Deng [b], Chaoqun Duan [a,c], Yi Qin [a,b], Jun Luo [a,b], Huayan Pu [a,b]

[a] State Key Laboratory of Mechanical Transmissions, Chongqing University, Chongqing, 400044, China
[b] College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China
[c] School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China

ARTICLE INFO

Article history:
Received 28 July 2023
Received in revised form 23 August 2023
Accepted 30 August 2023
Available online 9 September 2023

Keywords:
Joint adaptive distribution
Duplex adversarial network
Partial domain adaptation
Rotating machinery
Fault diagnosis

ABSTRACT

Domain-adaptation technologies have been widely developed for mechanical fault diagnosis. Most related methods assume the same label space between the source and target domains, whereas the partial domain adaptation problem is more widespread and challenging. Moreover, most previous studies focused on data distribution alignment but failed to address fine-grained distributional differences of subdomains. Hence, this study proposes a novel duplex adversarial deep discriminative network (DADDN) for fault diagnosis in cross-domain partial transfer cases. First, a dual-domain adversarial attention mechanism is adopted in the DADDN to discriminate the label space and evaluate the transferability of the data samples. Second, a new metric function of the distribution difference is employed to learn both the fine-grained and discriminative features of the data samples and a weight-joint adaptive distribution mechanism is proposed to improve the domain confusion capability. Third, a center-of-balance weighting strategy is utilized to expand the categories of the target domain and reduce label space asymmetry. In addition, to avoid negative transfer, a weighting mechanism is designed to filter outlier categories, and a dynamic adaptive factor is introduced to weigh the importance of the marginal and conditional distributions. Finally, the effectiveness, robustness, and progressiveness of the proposed method are validated using several partial-transfer diagnostic tasks based on three rotating machinery datasets.

1. Introduction

Rotating machinery, which is the most typical mechanical component, is widely utilized in industries and mechanical products such as aerospace, intelligent manufacturing, power transmission, and underwater vehicles. However, the long-term operation of rotating machinery in harsh operating environments results in performance deterioration and increases the risk of system failure; therefore, the intelligent diagnosis of rotating machinery must be performed to prevent accidents [1]. Owing to the rapid development of sensing technologies and the Internet of Things, researchers have extensively investigated data-driven intelligent diagnostic algorithms [2]. In particular, the method represented by deep learning has received significant attention owing to its high diagnosis rate, fast response, and reduced dependence on prior knowledge. In this regard, remarkable results have been achieved in the field of intelligent diagnosis [3], e.g., the establishment of convolutional neural networks [4,5], recurrent neural networks [6], long short-term memory [7], and deep belief networks [8]. To ensure accurate fault diagnosis, the devised methods must satisfy two conditions. First, a significant amount of labeled data is required for model training and testing. Second, the training and test data must have the same distribution [9]. However, in the case of data sampling from rotating machinery, the equipment typically operates smoothly for long periods, thus rendering it difficult to obtain a sufficient amount of sample data with fault labels. Furthermore, the machine operating conditions, signal sampling frequency, and workload can cause significant differences in the distributions of the acquired training and test data [10].

Transfer learning (TL) techniques have been shown effective in the field of intelligent diagnosis to address the issues of sparse labeling and significant distributional differences in sample data [11]. The TL technique extracts feature information from target samples by learning feature knowledge from a source domain with sufficient labeled data and then transferring this knowledge to a target domain with limited labeled data [12]. Among the developed methods, domain adaptation (DA) is a subset of TL and has yielded significant results in cross-domain transfer tasks [13]. The typically used DA methods can be classified into those based on statistical metric distance [14–18] and

* Corresponding author at: State Key Laboratory of Mechanical Transmissions, Chongqing University, Chongqing, 400044, China.
E-mail address: liufq@cqu.edu.cn (F. Liu).

https://doi.org/10.1016/j.knosys.2023.110960
0950-7051/

those based on domain confrontation mechanisms [19–24]. The first type achieves domain-invariant feature extraction using a metric distance function to reduce the distribution differences between two domains, whereas the second type achieves shared feature extraction using an adversarial game strategy between the feature extractor and domain discriminator. For instance, the local maximum mean difference (LMMD) [20] and multiple kernel maximum mean discrepancy (MK-MMD) [21] are integrated into the domain adversarial network to improve the domain confusion ability and realize the distribution alignment of sub-domain categories. The conditional weighted transmission Wasserstein distance method [15] has been proven to yield better results than other typical covariance methods. Qian et al. [16] proposed a deep discriminative transmission network that combined the maximum mean difference (MMD) and correlation aligned metric distance methods to compose joint adaptive distributions to improve domain mixing. Shi et al. [19] proposed a multi-scale feature fusion network that integrates a transferable duplex attention mechanism into a domain adversarial network to improve the transferability of fused shared features.

DA-based TL techniques have yielded remarkable results. However, the target domain space is typically much smaller than the source domain space in actual situations. This is particularly evident in a specific DA scenario, i.e., partial domain adaptation (PDA), which poses a significant problem in industrial applications [25,26]. Research pertaining to PDA has been conducted. For example, Qian et al. [27] constructed a multiscale domain confrontation network to achieve local domain adaptation (DA) based on a domain confrontation network, instance weights, and category weights. Zhang et al. [28] proposed a novel digital twin-driven approach to reduce the feature distribution discrepancy between simulated and measured data. Yang et al. [29] proposed a contrastive learning-assisted conditional alignment strategy to reduce the difference between the conditional and marginal distributions of domains. The critical classes and samples identifying the network [30] and maximum cosine similarity [31] were used to identify the most relevant source classes and critical target samples to quantify the transferability of the samples. Although PDA-based methods have been proven effective in partial transfer tasks, they typically rely on instance and category weights to extract domain-invariant features, whereas the finer details of the sample features are disregarded. In addition, the asymmetry of the domain label space typically blurs the feature decision boundaries and reduces transferability, thus resulting in a negative transfer.

To solve the aforementioned issues via PDA [19,32], this study proposes a duplex adversarial deep discriminative network (DADDN) to manage fine-grained information in PDA problems. Sample transferability is quantified by distinguishing outliers in the source domain through a dual-domain adversarial mechanism. A center-of-balance weighting strategy is used to mitigate the asymmetry in the domain label space. A joint adaptive weight distribution is employed to reduce the distance between the two domains and achieve domain alignment. The effectiveness of the proposed method is verified by transferring fault diagnosis tasks across domains on three rotating machinery datasets. The main contributions of this study are as follows:

1. To account for the nuanced information in the subdomain categories, a duplex domain adversarial attention mechanism is developed to enhance the generalization and domain confusion capabilities of the model.
2. Considering the distribution differences of the different sample categories, a measurement function for the weighted conditional MMD is constructed, and a weight-joint distribution adaptation is constructed to align the probability distributions in different domains.
3. A dynamic adaptation factor is designed to dynamically measure the importance and correlation of marginal and conditional probability distributions to achieve distributional alignment between domains.

The remainder of this paper is organized as follows: Section 2 presents the preliminaries. In Section 3, a novel methodology is proposed. In Section 4, transfer experiment verification and analysis are presented. Finally, Section 5 concludes the paper.

2. Preliminaries

2.1. Problem formulation

Several domain concepts are introduced in this section to provide a clear explanation of the PDA problem. The domain $\mathcal{D} = \{(X, Y), \mathrm{P}(X, Y)\}$ is first defined, where $(X, Y) \in R^n$ represents n-dimensional feature spaces with $(\mathbf{x}, \mathbf{y}) = \{(x, y)_1, (x, y)_2, \ldots, (x, y)_n\} \in (X, Y)$ and $\mathrm{P}(X, Y)$ represents the joint probability distribution of $\mathcal{D}$. The labeled source domain is defined as $\mathcal{D}_s = \left\{(x_i^s, y_i^s)\right\}_{i=1}^{n_s}$, where (x_i^s, y_i^s) represents the data samples. The label space category of the source domain is defined as $|C_S|$. The unlabeled target domain data set $\mathcal{D}_t = \left\{(x_i^t)\right\}_{i=1}^{n_t}$ is sampled from the target domain distribution $\mathrm{P}^t(X, Y)$ with category $|C_T|$.

As shown in Fig. 1, different transfer fault diagnosis scenarios are illustrated. In the unsupervised domain adaptation (UDA) setting, the source and target domains have a common label space $C_S = C_T$, different marginal distributions $P(X_s) \neq P(X_t)$, and different conditional distributions $P(Y_s|X_s) \neq P(Y_t|X_t)$. In the PDA setting, the target label space is a subset of the source label space $C_T \subseteq C_S$. Hence, the source domain can be classified into two subdomains, i.e., the shared label space $C_S \cap C_T$ and the outlier label space $C_S \setminus C_T$, where the set of classes situated in the shared label space is defined as the shared classes $\mathfrak{R}_{C_S \cap C_T}$, whereas outlier classes are defined $\mathfrak{R}_{C_S \setminus C_T}$. For the standard PDA, the source and target domains satisfy the following constraints: $C_T \subseteq C_S$ and $P^s\{(X_s, Y_s)\} \neq P^t\{(X_t, Y_t)\}$. The label space of the target domain is a subset of the source domain, which renders it difficult to transfer and share domain knowledge. Therefore, this study is performed to identify a novel metric function approach and network framework to achieve domain-shared feature extraction in PDA scenarios.

2.2. Domain adversarial neural network (DANN)

Inspired by generative adversarial networks, domain adversarial neural networks (DANN) [33] and their variants have been effectively applied to various TL problems. Specifically, a DANN facilitates the acquisition of transferable knowledge by extracting domain-invariant features to minimize distributional differences between the source and target domains. The DANN comprises three key components: a feature extractor $\mathrm{G_f}$, a domain discriminator $\mathrm{G_d}$, and a classifier $\mathrm{G_c}$, whose parameters are expressed as ϑ_f, ϑ_d, and ϑ_c, respectively. The feature extractor $\mathrm{G_f}$ and domain discriminator $\mathrm{G_d}$ adopt a game process to reduce the domain distribution difference; the classifier $\mathrm{G_c}$ achieves an accurate prediction of samples by minimizing the classification loss. Mathematically, the optimization objective of the DANN is defined as [27]

$$\begin{aligned}\mathscr{L}_{adv}\left(\vartheta_f, \vartheta_d, \vartheta_c\right) = &\frac{1}{n_s}\sum_{\mathbf{x}_i \in \boldsymbol{D}_S} \mathcal{L}_c\left(\mathrm{G_c}\left(\mathrm{G_f}\left(\mathbf{x}_i\right)\right), \mathbf{y}_i\right) \\ &- \frac{\lambda}{n_s + n_t}\sum_{\mathbf{x}_i \in \boldsymbol{D}_S \cup \boldsymbol{D}_T} \mathcal{L}_d\left(\mathrm{G_d}\left(\mathrm{G_f}\left(\mathbf{x}_i\right)\right), \mathbf{d}_i\right),\end{aligned} \tag{1}$$

where $\mathcal{L}_c$ and $\mathcal{L}_d$ represent the cross-entropy losses for $\mathrm{G_f}$ and $\mathrm{G_d}$, respectively; n_s and n_t are the minimum batch sizes in the

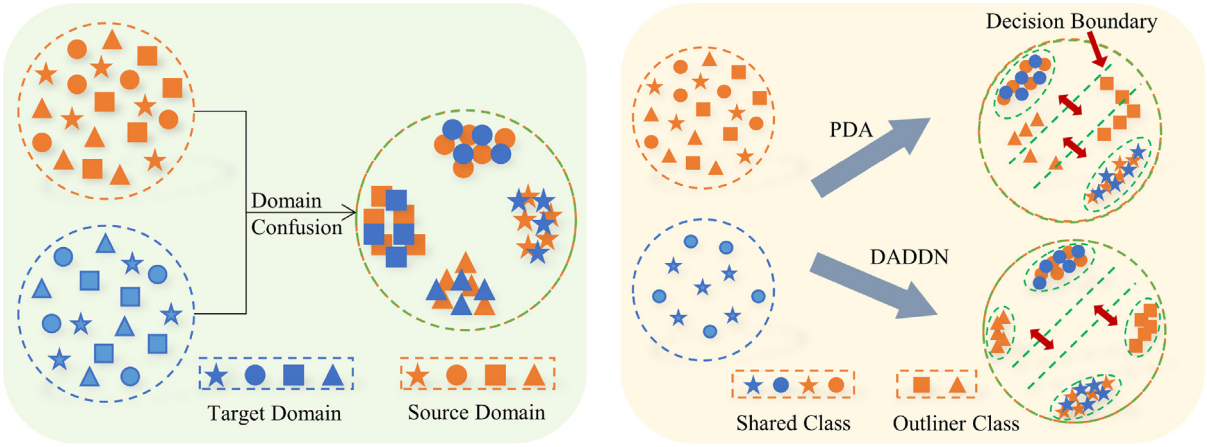

Fig. 1. Feature alignment process of different methods on unsupervised domain adaptation (UDA) and PDA tasks; different shapes represent different categories, and orange and blue represent source and target domains, respectively. (a) Feature distribution alignment effect of DA method in UDA scenario. (b) Feature distribution alignment effect of PDA-based and DADDN method proposed herein in partial transfer scenario.

source and target domains, respectively; and λ is a trade-off coefficient. After minimax game optimization is performed, the domain-shared features are extracted. In Eq. (1), the parameters ϑ_f, ϑ_d, and ϑ_c are defined as follows:

$$\begin{aligned}(\hat{\vartheta}_f, \hat{\vartheta}_c) &= \arg\min_{\vartheta_f, \vartheta_c} E\left(\vartheta_f, \vartheta_c, \hat{\vartheta}_d\right), \\ (\hat{\vartheta}_d) &= \arg\max_{\vartheta_d} E\left(\hat{\vartheta}_f, \hat{\vartheta}_c, \vartheta_d\right).\end{aligned} \tag{2}$$

By introducing a gradient inversion layer to update the model parameters, the DANN becomes widely applicable to UDA scenarios, where the source and target domains share the same label space.

2.3. Local maximum mean discrepancy

The metric function is a key factor in achieving DA. MMD is a simple and efficient metric that is typically used to quantify the distribution difference between the source and target domains. It evaluates the mean of the kernel function metric projected into the regenerated kernel hilbert space (RKHS) and minimizes the mean to reduce the distribution difference between these domains. Mathematically, the MMD can be formulated as [21]

$$D_{\mathscr{H}}(P_s, P_t) \triangleq \left\| \mathbf{E}_{x^s \sim P_s}[\phi(X^s)] - \mathbf{E}_{x^t \sim P_t}[\phi(X^t)] \right\|^2_{\mathscr{H}}, \tag{3}$$

where X^s and X^t are the instances in the source and target domains, respectively; P_s and P_t are the distribution spaces of the source and target domains, respectively; $\phi(\cdot)$ maps primary features into RKHS; $\mathscr{H}$ is the RKHS associated with kernel $k\left(x^s, x^t\right) = \left\langle \phi\left(x^s\right), \phi\left(x^t\right) \right\rangle$. Moreover, $D_{\mathscr{H}}(P_s, P_t) = 0$ only if $P_s = P_t$.

Currently, most MMD-based domain-adaptation techniques use a global distribution alignment method to estimate the differences in the distributions of different domains. However, these approaches overlook the granular information of closely related subdomains within the same category, thus resulting in confusion between related categories in both the source and target domains. Hence, LMMD is used to align the related subdomains by matching both global and local distributions within the same category and assigning weight values to the samples. It is mathematically defined as [20]

$$D_{\mathscr{H}}(P_s, P_t) = \frac{1}{C} \sum_{c=1}^{C} \left\| \sum_{x_i^s \in \mathcal{D}^s} w_i^{sc} \phi(x_i^s) - \sum_{x_j^t \in \mathcal{D}^t} w_j^{tc} \phi(x_j^t) \right\|^2_{\mathscr{H}}, \tag{4}$$

where x_i^s and x_i^t are the ith samples in the source and target domains, respectively; and w_i^{sc} and w_j^{tc} represent the weights of x_i^s and x_i^t belonging to category c, respectively. In addition, $\sum_{i=1}^{N_s} \omega_i^s = \sum_{j=1}^{N_t} \omega_j^t = 1$ and $\sum_{x_i \in \mathcal{D}} \omega_i^c \phi(x_i)$ is the weighted sum of category c. The weight w_i^c for a sample x_i can be computed as follows:

$$w_i^c = \frac{y_{ic}}{\sum_{(x_i, y_i) \in D} y_{ic}}, \tag{5}$$

where y_{ic} represents the probability of a sample belonging to category c. The LMMD is embedded in the fully connected layer of the label classifier and the activation function of this layer includes distributions $\left\{z_i^{sl}\right\}_{i=1}^{N_s}$ and $\left\{z_j^{tl}\right\}_{j=1}^{N_t}$. $\hat{D}_{\mathscr{H}}(P_s, P_t)$ is defined as an unbiased estimator of $D_{\mathscr{H}}(P_s, P_t)$ and can be written as

$$\begin{aligned}\mathcal{L}_{\mathrm{LMMD}} = \hat{D}_{\mathscr{H}}(P_s, P_t) = \frac{1}{C} \sum_{c=1}^{C} \Bigg\| &\sum_{i=1}^{N_s} \sum_{j=1}^{N_s} w_i^{sc} w_j^{sc} k\left(z_i^{sl}, z_j^{sl}\right) \\ &+ \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} w_i^{tc} w_j^{tc} k\left(z_i^{tl}, z_j^{tl}\right) \\ &- 2 \sum_{i=1}^{N_s} \sum_{j=1}^{N_t} w_i^{sc} w_j^{tc} k\left(z_i^{sl}, z_j^{tl}\right) \Bigg\|^2_{\mathscr{H}}.\end{aligned} \tag{6}$$

3. Proposed method

3.1. Network architecture

To address the diagnostic task in the PDA scenario, a novel DADDN model is proposed based on the network structure of the DANN. The detailed architecture of the model is illustrated in Fig. 2. In addition to the basic feature extractor G_f, domain discriminator G_d, and label classifier G_c, a domain classifier fusion module, weight-joint adaptive mechanism, and class balanced

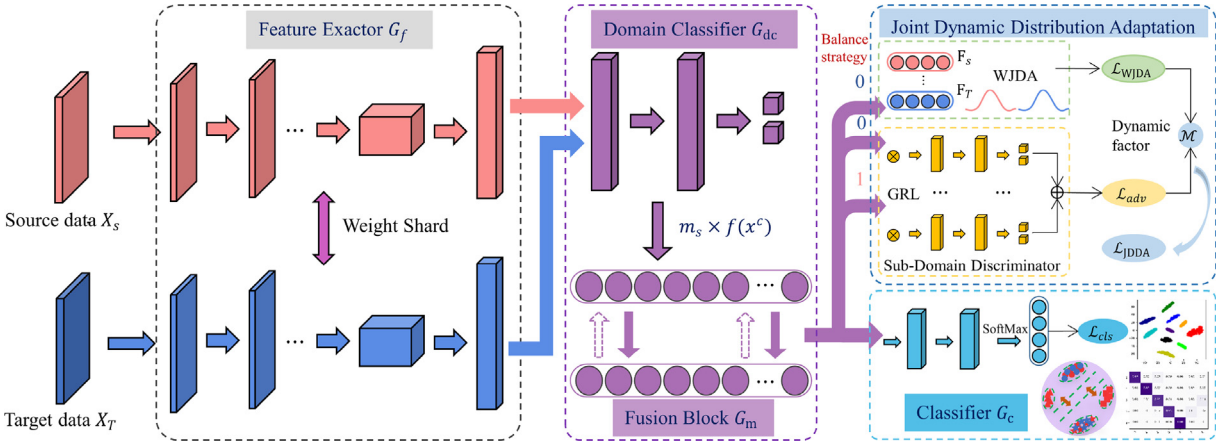

Fig. 2. In the proposed DADDN network structure, the joint dynamic distribution adaptive loss $\mathcal{L}_{\text{JDDA}}$ is computed using the output of the weight-joint domain adaptive loss and the subdomain discriminator and adjusted by a dynamic factor $\mathcal{M}$; the weight parameter m_s is computed by the domain classifier.

Table 1
Structural parameter details of DADDN.

Module	Layers	Operators	Stride/Padding	Output
Feature extractor	Conv1D-1	Convolution	64/16	(512,32)
		BN	–	(512,32)
		LeakyReLu	–	(512,32)
		Max-pooling	2/2	(256,32)
	Conv1D-2	Convolution	5/1	(256,64)
		BN	–	(256,64)
		LeakyReLu	–	(256,64)
		Max-pooling	2/2	(128,64)
	Conv1D-3	Convolution	3/1	(128,128)
		BN	–	(128,128)
		LeakyReLu	–	(128,128)
		Max-pooling	2/2	(64,128)
	Residual block1	Convolution BN LeakyReLu	(3/1) × 2	(32,256)
	Residual block2	Convolution BN LeakyReLu	(3/1) × 2	(16,512)
	GAP	–	16/1	512
	Flatten	Linear	–	512
Fusion block	Fc1	Linear	–	512
Lable classifier	Fc2	Linear	–	512
		Dropout (0.5)	–	256
	Fc3	Linear	–	128
		Softmax	–	Category
Domain discriminator (Domain Classifier)	Fc4	Linear	–	512
		Dropout (0.5)	–	256
		LeakyReLu	–	256
	Fc5	Linear	–	128
		Sigmoid	–	128
		Linear	–	1

weight center mechanism are introduced to balance the distribution between different domains and improve the robustness and generalization of the model. The specific network parameters are listed in Table 1. This network contains three Convolutional 1D (Conv1D) blocks, two residual blocks, a global average pooling layer, and five fully connected layers. Each Conv1D block comprises a convolution layer, batch normalization (BN) layer, LeakyReLU activation function, and maximum pooling layer. The dropout and BN layers can accelerate network convergence and reduce overfitting.

3.2. Proposed DADDN mechanism

In this study, a dual-domain adversarial attention mechanism is first employed to effectively discriminate outlier samples in the source domain and quantify the transferability of the samples through the fusion module. Subsequently, the fused feature information is used to complement the target samples through the balanced center weighting mechanism to alleviate the asymmetry of the label space. Finally, the obtained feature information is input into a weighted joint distribution adaptation and subdomain discriminator to facilitate domain distribution alignment, thus resulting in well-defined feature decision boundaries.

3.2.1. Dual-domain adversarial attention mechanism

In conventional domain-adaptive mechanisms, assigning the same weight to samples from different domain categories is unreasonable. During domain alignment, the over-weighting of samples in uncertain classification boundaries typically hinders classification alignment between domains. Therefore, to improve the transferability of samples during the diagnostic process, different weight values are assigned to different types of feature information. This study adopts the information entropy criterion to analyze and measure the feature information extracted by the feature extractor G_f. Specifically, the set of weight parameters m_s is obtained based on the dynamic attention mechanism, and its mathematical formula is as follows:

$$m_s(E_s) = 1 + \exp(-E_s \times \tau),$$
$$E_s = \sum_{i=1}^{n} \left(H\left(G_{dc}\left(G_f(x_i^s)\right)\right) + H\left(G_{dc}\left(G_f(x_i^t)\right)\right) \right), \tag{7}$$

where G_{dc} refers to the domain classifier; n represents the batch size; and x_i^s and x_i^t denote examples in the source and target domains, respectively. The information entropy criterion $H(g) = -\sum_{c=1}^{C} g_c \log(g_c)$ is used to measure the amount of information possessed by a sample. The higher the value, the easier it is to confuse the corresponding feature information. Therefore, larger weight values should be assigned when extracting domain-invariant features. Using fully connected layers and softmax functions, different classes of dynamic parameter weight sets m_s are provided to quantify sample transferability. The coefficient τ, which serves as a regulator, is used to increase the stability of

the network during training. This prevents the weight parameter $m_s\{f(x^s)\}$ from becoming extremely large to achieve a local optimum during training.

During the forward propagation of the DADDN, the dynamic weight parameters are assigned to the feature vector G_f for fusion, and the resulting parameters are fused by the fully connected feature layer G_m. During the backward propagation of the DADDN, a domain discriminator loss $\mathcal{L}_{dc}$ with dynamic weight parameters is generated for classification training. The loss function used to describe this process is explained in the optimization objective subsection.

3.2.2. Weight-joint distribution adaptation

Currently, most domain-adaptive distributions extract domain-invariant features by minimizing the domain distribution without considering the fine-grained information in the subdomains of the source and target domains, the label space asymmetry of the source and target domains in the PDA scenario. Hence, this study adopts the LMMD to evaluate the global marginal distribution difference of domains and proposes a novel distribution metric function (conditional weighted maximum mean difference, CWMMD) to evaluate the conditional distribution difference between subdomains. To further improve domain confusion, LMMD and CWMMD are combined as a new metric, i.e., weight-joint distribution adaptation (WJDA).

The CWMMD uses a non-linear kernel function and subdomain category weights to compute the distribution differences between domains [25]. Specifically, CWMMD projects feature information into the RKHS and measures the projected distance between $P^s(X|Y)$ and $P^t(X|Y)$. Let ϕ and ψ represent the nonlinear mappings for X and Y, respectively. The conditional distribution $P(X|Y)$ is a vector space in which the conditional kernel mean is embedded in the RKHS. For each sample x_i, its conditional kernel mean $\boldsymbol{\mu}_{X|\mathbf{y}}$ can be defined as

$$\boldsymbol{\mu}_{X|\mathbf{y}} = \mathbb{E}_{\mathbf{X}|\mathbf{y}}[\psi(\mathbf{X})] = \int_{\Omega} \psi(\mathbf{x}) \mathrm{d}P(\mathbf{X} \mid \mathbf{y}) = \mathbf{C}_{X|Y}\phi(\mathbf{y}), \tag{8}$$

where $C_{X|Y}$ denotes the conditional kernel mean of $P(X|Y)$, which can be generated as follows:

$$\begin{aligned} C_{X|Y} &= \mathcal{C}_{XY}\mathcal{C}_{YY}^{-1}, \\ C_{XY} &= \mathbb{E}_{XY}\left[\phi\left(X\right) \otimes \psi\left(Y\right)\right] - \mu_X \otimes \mu_Y. \end{aligned} \tag{9}$$

In Eq. (9), C_{XY} is a cross-covariance operator: $\mathcal{F} \rightarrow \mathcal{G}$ and $\otimes$ is the tensor product operator. Assuming a dataset $\mathcal{D}_{XY}$ with independent and identically distributed samples from $P(X, Y)$, the unbiased estimate of $C_{X|Y}$ can be obtained as follows:

$$\widehat{C}_{X|Y} = \Psi\left(\Phi^{\top}\Phi + \lambda\mathbf{I}\right)^{-1}\Phi^{\top} = \Psi\left(K + \lambda\mathbf{I}\right)^{-1}\Phi^{\top}, \tag{10}$$

where $\Psi = [\psi(x_1), \psi(x_2), \ldots, \psi(x_n)]$, $\Phi = [\phi(y_1), \phi(y_2), \ldots, \phi(y_n)]$, $K = \Phi^{\top}\Phi$ is the kernel function, and λ denotes the regularization parameter.

The projection distance between $C^s_{X|Y}$ and $C^t_{X|Y}$ in the RKHS can be obtained using the CWMMD. The weight w_i^c of the subdomain category is added to compute the distance between the domains. Similar to the MMD, the CWMMD loss is defined as

$$\mathcal{L}_{\mathrm{CWMMD}} = \left\| w_i^s C^s_{X|Y} - w_i^t C^t_{X|Y} \right\|^2_{\mathcal{F}\otimes\mathcal{G}}, \tag{11}$$

where w_i^s and w_j^t belonging to class c denote the weights of x_i^s and x_j^t, respectively.

Owing to the unlabeled features of the target domain, the pseudo-label vector, denoted as $\hat{y}_i^t$, can be obtained from the fully connected layer of the label classifier G_c. Using this approach, an unbiased estimate of the CWMMD between the source and target domains can be defined as

$$\begin{aligned} \widehat{\mathcal{L}}_{\mathrm{CWMMD}} &= \left\| w_i^s \widehat{C}^s_{X|\mathbf{Y}} - w_i^t \widehat{C}^t_{\widehat{X}|\mathbf{Y}} \right\|^2_{\mathcal{F}\otimes\mathcal{G}} \\ &= \left\| w^{sc}\Psi_s\left(K_s + \lambda\mathbf{I}\right)^{-1}\Phi_s^{\top} - w^{tc}\Psi_t\left(K_t + \lambda\mathbf{I}\right)^{-1}\Phi_t^{\top} \right\|^2_{\mathcal{F}\otimes\mathcal{G}} \\ &= \frac{1}{C}\sum_{c=1}^{C}\Bigg\| \sum_{i=1}^{n_s}\sum_{j=1}^{n_s} w_i^{sc} w_j^{sc} K_s \widetilde{K}_s^{-1}\mathcal{L}_s\widetilde{K}_s^{-1} \\ &\quad + \sum_{i=1}^{n_t}\sum_{j=1}^{n_t} w_i^{tc} w_j^{tc} K_t \widetilde{K}_t^{-1}\mathcal{L}_t\widetilde{K}_t^{-1} \\ &\quad - 2\sum_{i=1}^{n_s}\sum_{j=1}^{n_t} w_i^{sc} w_j^{tc} K_{st} \widetilde{K}_s^{-1}\mathcal{L}_{st}\widetilde{K}_t^{-1} \Bigg\|^2_{\mathcal{F}\otimes\mathcal{G}}, \end{aligned} \tag{12}$$

where $K_s = \Phi_s^{\top}\Phi_s$, $\widetilde{K}_s = K_s + \lambda I$, and $\mathcal{L}_s = \Psi_s^{\top}\Psi_s$ are defined on the dataset $\mathcal{D}_s = \left\{(x_i^s, y_i^s)\right\}_{i=1}^{n_s}$, with K_t, $\mathcal{L}_t$, and $\widetilde{K}_t$ defined on the dataset $\mathcal{D}_t = \left\{(x_i^t)\right\}_{i=1}^{n_t}$.

Once the CWMMD mechanism is combined with the LMMD mechanism, the joint adaptive distribution with weights can be expressed as

$$\begin{aligned} \mathcal{L}_{\mathrm{WJDA}} &= \lambda_1 \left\| \mathbb{E}_{\mathbf{x}\sim P_s} T(x^s) - \mathbb{E}_{\mathbf{x}\sim P_t} T(x^t) \right\|^2_{\mathscr{H}} \\ &\quad + \lambda_2 \left\| \mathbb{E}_{\mathbf{x}|\mathbf{y}^s}\phi(x^s, y^s) - \mathbb{E}_{\mathbf{x}|\mathbf{y}^t}\phi(x^t, \hat{y}^t) \right\|^2_{\mathcal{F}\otimes\mathcal{G}} \\ &= \lambda_1\mathcal{L}_{\mathrm{LMMD}} + \lambda_2\mathcal{L}_{\mathrm{CWMMD}}. \end{aligned} \tag{13}$$

Here, the tradeoff parameters λ_1 and λ_2 are utilized to balance the importance of the marginal and conditional probability distributions, respectively.

3.2.3. Balanced center weighting mechanism

In PDA scenarios, the sample label space of the target domain is a subset of the source domain, and the samples can be easily misplaced and matched during the domain alignment process, thus resulting in a negative transfer. A balanced center weighting strategy was constructed from the perspectives of class-level weights, balanced adversarial alignment, and uncertain target complementary functions to alleviate the asymmetry of the sample label space.

(1) Class-level weights The DA model is typically trained with labels from the source domain, thus resulting in a pretrained feature extractor G_f and label classifier G_c. Subsequently, the target sample is fed into the classifier G_c to predict the label $\hat{\mathbf{y}}_i^t$. In the PDA scenarios, the label space of the target domain belongs to the common space $\mathfrak{R}_{C_S\cap C_T}$ in the source domain label space, which is significantly different from the outlier sample $\mathfrak{R}_{C_S\setminus C_T}$. Therefore, the probability of the target sample being predicted as an outlier by the source classifier is sufficiently low. In this study, the probability distribution obtained by inputting the target sample into the source classifier G_c is regarded as the class-level weight of the target sample to measure the contribution of different classes of samples in the transfer process. This is mathematically expressed as

$$w_c = \frac{1}{n_t}\sum_{i=1}^{n_t}\hat{y}_i^t. \tag{14}$$

The weight w_c of $|\mathcal{C}_S|$ is subject to the condition $\sum_{c=1}^{|\mathcal{C}_s|} w_c = 1$. Since the weight elements of w_c from some categories are relatively small, the weight vector w_c is normalized as $w_c \leftarrow w_c/\max(w_c)$.

(2) Balanced adversarial alignment During PDA, DA-based models typically exhibit inconsistency between the target samples and source domain outliers, thus resulting in a negative transfer. Although the class-level weighting process can increase the weight of common samples and decrease the weight of outliers, it essentially relies on the probability distribution of the output of the source classifier, whereas the distribution shift between domains is disregarded. To achieve balance between the label spaces of the source and target samples, an extension strategy for balanced adversarial alignment is devised in this study. Specifically, the source samples are randomly selected as target expansion samples, which complemented the target samples in the domain confrontation. Based on the balanced weighted adaptive loss function, the domain discriminator is designed as follows:

$$\mathcal{L}_{adv}^{h} = -\frac{1}{n_s+n_t}\sum_{h=1}^{H}\sum_{k=1}^{n_s+n_t}\mathcal{L}_D^h\left(G_d^h\left(Z_k^h \mathrm{G_m}\left(\mathrm{G_f}(x_i)\right)\right), d_i\right), \tag{15}$$

where

$$\begin{aligned}\mathcal{L}_D^h = &-\frac{1}{n_t}\sum_{j=1}^{n_t} w\left(x_i^s\right) m\left(y_i^s\right)\log\left(\mathrm{G_d}\left(\mathrm{G_m}\left(\mathrm{G_f}(x_i^s)\right)\right)\right)\\ &-\frac{1}{n_t}\sum_{j=1}^{n_t} w(x_j^t)\log\left(1-\mathrm{G_d}\left(\mathrm{G_m}\left(\mathrm{G_f}(x_i^t)\right)\right)\right)\\ &-\frac{\rho}{n_s}\sum_{i=1}^{n_s} w(x_i^s)m(y_i^s)\log\left(1-\mathrm{G_d}\left(\mathrm{G_m}\left(\mathrm{G_f}(x_i^s)\right)\right)\right);\end{aligned} \tag{16}$$

H is the total number of label categories of the input samples; the corresponding subdomain discriminator of category h is G_d^h, with its cross-entropy loss function being $\mathcal{L}_D^h$; Z_k^h represents the expected classifier score for a specified category; and d_i is the domain category label of the input sample x_i. To solve the asymmetry problem of category labels, an augmentation rate parameter ρ is introduced, which gradually decreases to 0 during the training iterations. In the early stages of training, numerous relevant input samples are required to augment the target samples and balance the class labels, whereas the augmentation requirements gradually decrease over time, thus resulting in a lower value for ρ.

(3) Supplementary function of uncertain objective When domain alignment strategies are used for domain-shared feature extraction, certain target samples must match the outlier source samples on the classification decision boundary. This results in the misclassification of such source samples because the assigned larger weights result in increased uncertainty. Hence, this paper introduces the supplementary function of uncertain objectives to suppress such uncertainty by denoting the wrong category of the source sample on the decision boundary and reducing the prediction score of the wrong category. The confidence-weighted supplementary function of the uncertain objective sample is expressed as follows:

$$\mathcal{L}_{\mathrm{cot}} = \frac{1}{n_s(K-1)}\sum_{i=1}^{n_s} w\left(y_i^s\right)\mathcal{L}_{\mathrm{wce}}\left(\mathrm{G_c}\left(\mathrm{G_m}\left(\mathrm{G_f}(x_i^s)\right)\right), y_i\right), \tag{17}$$

where

$$\mathcal{L}_{\mathrm{wce}}\left(\hat{y}, y\right) = (1-\hat{y}_a)^{\eta}\sum_{j=1, j\neq g}^{K}\frac{\hat{y}_j}{1-\hat{y}_a}\log\left(\frac{\hat{y}_j}{1-\hat{y}_a}\right); \tag{18}$$

K represents the number of categories in the source domain, η the hyperparameter, y_i^s the truth label, and a the index of the truth category in $\hat{y}_a$.

3.2.4. Joint dynamic distribution adaptation

The marginal and conditional distributions between the samples of different classes contribute differently in the DA process. To accurately estimate their relative importance, this study introduces a dynamic and adjustable factor that ensures that the entire model is more specific during transition training, thereby preventing negative transfers.

The dynamic adaptive factor is determined by the marginal and conditional distributions. The $\mathcal{A}$-distance is employed as an indirect measure of the influence of these two components and is defined as

$$\begin{aligned}\mathcal{A}_g\left(\mathcal{D}_s, \mathcal{D}_t\right) &= 2\left(1-2\left(\mathcal{L}_{\mathrm{WJDA}}\right)\right),\\ \mathcal{A}_l\left(\mathcal{D}_s^h, \mathcal{D}_t^h\right) &= 2\left(1-2\left(\mathcal{L}_{adv}^h\right)\right),\end{aligned} \tag{19}$$

where $\mathcal{D}_s^h$ and $\mathcal{D}_t^h$ represent the category c in the source and target domains, respectively. Thus, the dynamic adaptive factor can be defined as

$$\mathcal{M} = \frac{\mathcal{A}_g\left(\mathcal{D}_s, \mathcal{D}_t\right)}{\mathcal{A}_g\left(\mathcal{D}_s, \mathcal{D}_t\right) + \frac{1}{K}\sum_{k=1}^{K}\mathcal{A}_l\left(\mathcal{D}_s^h, \mathcal{D}_t^h\right)}. \tag{20}$$

During the training process, extreme situations where $\mathcal{A}_g(\mathcal{D}_s, \mathcal{D}_t) \to 0$ or $\frac{1}{K}\sum_{k=1}^{K}\mathcal{A}_l\left(\mathcal{D}_s^h, \mathcal{D}_t^h\right) \to 0$ can occur and will adversely affect the overall calculation. Hence, Eq. (20) can be optimized exponentially as follows:

$$\mathcal{M} = \frac{e^{\mathcal{A}_g(\mathcal{D}_s, \mathcal{D}_t)}}{e^{\mathcal{A}_g(\mathcal{D}_s, \mathcal{D}_t)} + e^{\frac{1}{K}\sum_{k=1}^{K}\mathcal{A}_l(\mathcal{D}_s^h, \mathcal{D}_t^h)}}. \tag{21}$$

Subsequently, the loss function of the joint dynamic distribution adaptation (JDDA) can be defined as

$$\mathcal{L}_{\mathrm{JDDA}} = (1-\mathcal{M})\,\mathcal{L}_{\mathrm{WJDA}} + \mathcal{M}\mathcal{L}_{\mathrm{adv}}. \tag{22}$$

The following subsection discusses the effect of the loss function.

3.3. Optimization and training objective

The goal of the diagnostic task is to accurately classify the defect information in the target samples. For this purpose, the input target samples are classified using the trained source classifier $\mathrm{G_c}$ to obtain the expected accurate label information. By combining it with the category weight w_c, the cross-entropy loss function for optimizing the classifier can be expressed as

$$\mathcal{L}_{\mathrm{cls}} = \frac{1}{n_s}\sum_{i=1}^{n_s} w\left(y_i^s\right) L_y\left(\mathrm{G_c}\left(\mathrm{G_m}\left(\mathrm{G_f}(x_i^s)\right)\right), y_i^s\right). \tag{23}$$

To ensure that the transition task proceeds as intended and that the model possesses reliable discrimination capability, entropy minimization is incorporated into the model to mitigate false matches caused by uncertain factors in the source samples. The objective function for entropy minimization can be expressed as

$$\mathcal{L}_{\mathrm{ent}} = -\frac{1}{n_t}\sum_{i=1}^{n_t}\sum_{c=1}^{|\mathcal{C}_s|}\hat{y}_{i,c}^t \log\hat{y}_{i,c}^t, \tag{24}$$

where $\hat{y}_{i,c}^t$ represents the predicted label of target sample x_i^t.

For DA, a dual-adversarial strategy is used in this study. The information entropy weight m_s and domain classifier $\mathrm{G_{dc}}$ are used to globally align the feature information (the first alignment). After the feature information is fused and weighted, the discriminator is used for local alignment (the second alignment). This dual adversarial strategy further extracts domain-invariant information from the samples, thus allowing the discrimination loss function of $\mathrm{G_{dc}}$ to be expressed as

$$\mathcal{L}_{\mathrm{dc}} = -\frac{1}{n_s+n_t}\sum_{k=1}^{n_s+n_t} m_s L_D\left(\mathrm{G_{dc}}\left(\mathrm{G_f}(x_i)\right), d_i\right). \tag{25}$$

The objective function of the proposed DADDN comprises five components: label classifier loss $\mathcal{L}_{\text{cls}}$, joint adaptive loss $\mathcal{L}_{\text{JDDA}}$, entropy minimization loss $\mathcal{L}_{\text{ent}}$, domain classifier loss $\mathcal{L}_{\text{dc}}$, and objective complimentary loss $\mathcal{L}_{\text{cot}}$. The overall loss function of the DADDN is expressed as

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + \alpha\mathcal{L}_{\text{dc}} + \mathcal{L}_{\text{JDDA}} + \beta_1\mathcal{L}_{\text{ent}} + \beta_2\mathcal{L}_{\text{cot}}, \tag{26}$$

where α, β_1, and β_2 are hyperparameters used to balance the learning process. During the training process, based on Eq. (26), the optimization of the confrontation domain adaptive parameters ϑ_f, ϑ_c, and ϑ_d can be performed as follows:

$$\begin{aligned}
\vartheta_f &\leftarrow \vartheta_f - \xi\left(\frac{\partial\mathcal{L}_{\text{cls}}}{\partial\vartheta_f} - \lambda_1\frac{\partial\mathcal{L}_{\text{JDDA}}}{\partial\vartheta_f} - \lambda_2\frac{\partial\mathcal{L}_{\text{dc}}}{\partial\vartheta_f}\right),\\
\vartheta_c &\leftarrow \vartheta_c - \xi\frac{\partial\mathcal{L}_{\text{cls}}}{\partial\vartheta_c},\\
\vartheta_d &\leftarrow \vartheta_d - \xi\left(\frac{\partial\mathcal{L}_{\text{JDDA}}}{\partial\vartheta_d} + \frac{\partial\mathcal{L}_{\text{dc}}}{\partial\vartheta_d}\right),
\end{aligned} \tag{27}$$

where ξ denotes the learning rate and λ_1 and λ_2 are the balancing parameters. The detailed training process of the DADDN algorithm is presented in Algorithm 1.

Algorithm 1 Training process of DADDN algorithm

Input: Labeled source domain data $\mathcal{D}_s = \left\{(x_i^s, y_i^s)\right\}_{i=1}^{n_s}$, unlabeled target domain data $\mathcal{D}_t = \left\{(x_i^t)\right\}_{i=1}^{n_t}$, minimum batch size n_c, model training epoch N, learning rate ξ, and weight balance parameter λ.

for epoch in range(0,N) **do**
  for batch-size in range(0,n_c) **do**
    The minimum batch sample data x_i^s and x_i^t are from $\mathcal{D}_s$ and $\mathcal{D}_t$, respectively.
    Place x_i^c into G_f to obtain $f(x^c)$ and $m_s\{f(x^c)\}$; subsequently, place them into G_{dc} to obtain $\mathcal{L}_{dc}$, and implement weighted mixing in G_m.
    The fused information $f(m_s x^c)$ is inserted into G_c to obtain $\mathcal{L}_{\text{clc}}$ and $\mathcal{L}_{\text{WJDA}}$.
    $f(m_s x^c)$ is used in G_d^h with a balanced weighting strategy to generate $\mathcal{L}_{\text{adv}}$.
    Calculate the dynamic adaptive factor $\mathcal{M}$ from $\mathcal{L}_{adv}$ and $\mathcal{L}_{\text{WJDA}}$, and derive $\mathcal{L}_{\text{JDDA}}$.
    Calculate $\mathcal{L}_{\text{cot}}$ and $\mathcal{L}_{\text{ent}}$.
    Obtain the DADDN optimization objective function as follows: $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + \alpha\mathcal{L}_{\text{dc}} + \mathcal{L}_{\text{JDDA}} + \beta_1\mathcal{L}_{\text{ent}} + \beta_2\mathcal{L}_{\text{cot}}$.
    Update the gradient and parameters ϑ_f, ϑ_c, and ϑ_d.
  end for
  Update the dynamic weight parameter $\mathcal{M}$.
end for
return Diagnostic results for target samples.

4. Experimental study

To confirm the diagnostic capability of the DADDN in cross-domain PDA scenarios, we performed an experimental verification using three typically used rotating machinery datasets.

4.1. Dataset descriptions

(1) Case 1 The first bearing data were obtained from the bearing rig of the Case Western Reserve University (CWRU) [34], as shown in Fig. 3. The bearing failures were classified into three categories: i.e. ball (B), inner race (IR), and outer race (OR), each with three different sizes of bearing diameter damage (7, 14, and 21 mils). The data used were obtained from the motor-driven side at a sampling frequency of 12 K. The dataset comprised four rotational speeds, i.e., 1797, 1772, 1750, and 1730 rpm. Consequently, four different domains, denoted as A_0, A_1, A_2, and A_3 were constructed based on the speed variation. To evaluate the proposed method, 12 cross-domain tasks were formulated; the task specifications are listed in Table 2.

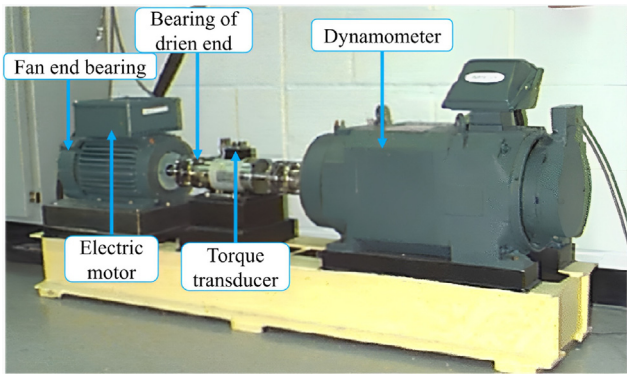

Fig. 3. Acquisition platform for CWRU dataset.

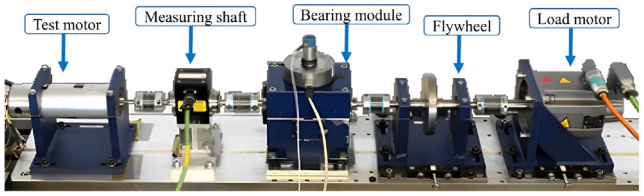

Fig. 4. Acquisition platform for PU dataset.

(2) Case 2 The second bearing dataset was obtained from the bearing test rig at Paderborn University (PU) [35], as shown in Fig. 4. This dataset contained two types of bearing data: artificially damaged and accelerated-life damaged. In this study, only the accelerated-life damage dataset was considered. The sampling frequency used to obtain the data was 64 K. The data types included health (H), outer race (OR), inner race (IR), and a combination of OR and IR, where each comprised two damage levels based on the load size and motor rotation speed. The dataset was classified into four operating conditions, namely B_0, B_1, B_2, and B_3, which were determined by the disparity between the moment and radial force. Therefore, 12 cross-domain transfer tasks were established based on different operating environments. The details are presented in Table 2.

(3) Case 3 The dataset was obtained from a classical mechanical system simulator at Southeast University (SEU) [36] in China and comprised two subdatasets focusing on bearings and gears, as shown in Fig. 5. In this study, only the gear dataset was used, which included the health, chipped, miss, surface, and root categories. The speed load included two different operating conditions, i.e., 20 Hz–0 V and 30 Hz–2 V, resulting in two different domains. By creating different missing scenarios, seven cross-domain transfer tasks were constructed to evaluate the proposed method. Detailed information regarding these tasks is presented in Table 3.

4.2. Comparative methods and experimental setup

To illustrate the advantages of the proposed DADDN model, we compared it with the following DA methods:

Table 2
Description of fault diagnosis tasks on CWRU and PU bearing databases.

Name	Tasks	Work conditions	Source domain	Target domain
CWRU	A_0 A_1 A_2 A_3	0 hp/1797 rpm 1 hp/1772 rpm 2 hp/1750 rpm 3 hp/1730 rpm	H, IR07,IR14,IR21, B07,B14,B21, OR07,OR14,OR21	H, IR07,IR14,IR21, B07,B14,B21
PU	B_0 B_1 B_2 B_3	1500 rpm/0.7 N m/1000 N 900 rpm/0.7 N m/1000 N 1500 rpm/0.1 N m/1000 N 1500 rpm/0.7 N m/400 N	H(K004), OR(KA04,KA30), IR(KI16,KI21), IR+OR(KB23,KB27)	H(K004), OR(KA04,KA30), IR(KI16,KI21)

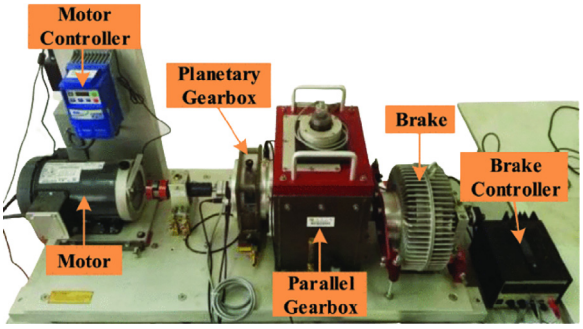

Fig. 5. Acquisition platform for SEU dataset.

Table 3
Fault diagnosis task description on the SEU gearbox dataset.

Tasks	Source domain	Target domain	Labels
Work conditions	20 Hz-0 V	30 Hz-2 V	
C_1	H,C,M,S,R	H,C,M,S,R	
C_2	H,C,M,S,R	H,C,M,S	Health(H):0,
C_3	H,C,M,S,R	H,C,M,R	Chipped(C):1,
C_4	H,C,M,S,R	H,C,S,R	Miss(M):2,
C_5	H,C,M,S,R	H,C,M	Surface(S):3,
C_6	H,C,M,S,R	H,C,S	Root(R):4
C_7	H,C,M,S,R	H,C,R	

(i) Baseline: The baseline contains only the feature extractor and label classifier, i.e., no DA module, and the source and target domain data are used for training and testing, respectively.
(ii) Deep adaptive network (DAN) [17]: The MK-MMD metric function is built into the baseline network to learn the domain-invariant features and facilitate DA.
(iii) DANN [33]: Compared with the baseline, the DANN learns domain-invariant features from the sample data through an adversarial mechanism.
(iv) Conditional domain adversarial network [18]: The auxiliary measurement function of the predictive label is added to the label classifier in the DANN.
(v) Partial adversarial domain adaptation (PADA) [23]: Based on DANN, partial transfer problems are solved by applying low-category weights to outlier samples.
(vi) Importance weighted adversarial network (IWAN) [24]: An auxiliary domain discriminator is designed based on the DANN to reduce the weight of outlier samples and achieve a partial transition.
(vii) BA3US [32]: Based on the DANN, it adds balanced adversarial alignment and adaptive uncertainty suppression to mitigate the class balance in PDA scenarios.

To ensure an unbiased cross-domain diagnostic evaluation, the same network architecture and optimization parameters as the DADDN model were used in these comparative methods. The classification accuracy of the target samples served as an evaluation metric for model performance. To accelerate the network training and convergence processes, the Adam optimization algorithm with a learning rate of $\xi = 0.001$ was employed. Using the StepLR learning rate decay method, the decay rate was $\gamma \times \xi$ and the number of iterations performed in the training was 5 to 10 times. The batch size for the training samples was set to 128 and the maximum epoch was set to 100. The tradeoff parameter for conflict resolution was calculated as follows: $\lambda = 2/\left(1 + e^{-10\times \mathrm{epoch}/\max}\right) - 1$ [16].

To prevent "test leakage" and ensure sufficient sample data, the same preprocessing strategy was used for the three rotating machinery datasets. A sliding window segmentation approach (window size of 2048 and repetition rate of 60%) was used to generate samples for each domain. The source and target domains comprised 2400 and 1800 samples, respectively, and each contained 1024 data points, to obtain sufficient fault information. In addition, each domain dataset was segmented into training and test sets at a ratio of 4:1. Considering the time–frequency domain information of the vibration signal, this study uses preprocessed samples of fast Fourier transform data as the input of the diagnostic model. To ensure the reliability of the diagnostic task and eliminate the randomness of the transfer tasks, each transfer task was repeated 10 times. All methods were implemented using the PyTorch framework in Python 3.9 with an Nvidia 2080 GPU.

4.3. Results and analysis

4.3.1. Case1: PDA-tasks at different speeds

Table 4 and Fig. 6(a) present the diagnostic results of the proposed DADDN method and other comparative methods on the CWRU-bearing dataset for cross-domain transfer tasks. The source and target domains comprised 10 and 7 label categories, respectively. Four different speeds were implemented, which resulted in 12 transfer tasks. An analysis of the results revealed the following:
(i) The diagnostic accuracy of the DA method surpassed that of the baseline approach, although only source domain samples were used. This indicates the effectiveness of the DA module for various cross-domain transfer tasks.
(ii) IWAN and BA3US achieved suboptimal results using an auxiliary discriminator and balanced weighting mechanism to eliminate irrelevant outlier categories in the source domains. Nevertheless, their average diagnostic accuracies were 12.26% and 14.73% lower than that of the proposed DADDN method, respectively.
(iii) DADDN consistently exhibited the highest diagnostic accuracy in all PDA tasks, thus demonstrating its exceptional generalization ability and stability in various transfer scenarios.

4.3.2. Case 2: Different operating environments

The goal of the cross-domain partial transfer tasks under different operating environments is to emulate real-world machine-operating scenarios. The proposed DADDN and the other methods were experimentally verified on the PU-bearing dataset, and the diagnostic results are presented in Table 5 and Fig. 6(b).
(i) DADDN demonstrated better diagnostic performance than the other compared methods in each transfer task, with an average

Table 4
Test accuracy (%) of failure diagnosis on CWRU bearing dataset.

Tasks	A_{0-1}	A_{0-2}	A_{0-3}	A_{1-0}	A_{1-2}	A_{1-3}	A_{2-0}	A_{2-1}	A_{2-3}	A_{3-0}	A_{3-1}	A_{3-2}	Average
Baseline	80.121	70.399	55.121	88.367	89.149	80.555	78.125	74.392	80.642	60.825	64.93	72.482	74.5923
DAN	82.79	76.909	78.414	91.35	97.251	88.932	85.463	89.184	85.208	72.544	73.229	82.864	83.6781
CDAN	84.722	89.459	80.425	89.481	92.847	90.833	75.892	83.615	90.399	67.001	76.367	80.868	83.4924
DANN	87.152	77.864	63.454	78.683	86.805	88.425	75.558	82.118	78.993	70.424	65.104	75.781	77.5300
PADA	82.142	75.998	72.873	78.608	90.682	89.670	75.669	81.488	84.548	55.357	62.720	81.539	77.6078
BA3US	90.833	85.091	79.926	94.642	97.604	87.326	82.388	92.013	90.017	66.824	74.110	82.147	85.2434
IWAN	89.409	84.427	80.468	97.366	99.891	90.342	94.447	87.986	97.899	72.879	77.582	79.756	87.7043
DADDN	**100**	**100**	**100**	**100**	**100**	**100**	**99.776**	**100**	**100**	**99.776**	**100**	**100**	**99.9626**

Table 5
Fault diagnosis test accuracy on PU bearing dataset (%).

Tasks	B_{0-1}	B_{0-2}	B_{0-3}	B_{1-0}	B_{1-2}	B_{1-3}	B_{2-0}	B_{2-1}	B_{2-3}	B_{3-0}	B_{3-1}	B_{3-2}	Average
Baseline	53.124	90.82	71.484	68.281	66.015	51.953	94.726	50.585	84.687	78.671	51.875	80.208	70.2024
DAN	57.304	91.943	79.562	74.414	69.335	62.5	97.695	62.76	85.625	88.91	51.43	90.478	75.9963
CDAN	57.91	93.115	77.343	70.273	76.757	64.296	92.334	70.312	84.104	77.081	57.421	80.625	75.1309
DANN	58.78	84.248	74.374	70.41	68.359	60.812	91.601	62.5	78.476	79.375	56.054	81.171	72.18
PADA	65.82	83.203	76	73.554	76.513	51.875	88.867	68.554	86.523	84.293	59.765	80.429	74.6163
BA3US	63.037	96.972	78.468	84.13	82.177	79.57	98.906	60.585	87.062	93.437	60.449	93.828	81.5517
IWAN	68.31	86.523	83.687	85.117	82.666	66.343	96.406	73.593	90.656	91.679	57.929	94.257	81.4305
DADDN	**81.64**	**97.265**	**88.281**	**91.992**	**88.281**	**83.593**	**99.414**	**83.398**	**95.468**	**97.656**	**81.64**	**95.117**	**90.3124**

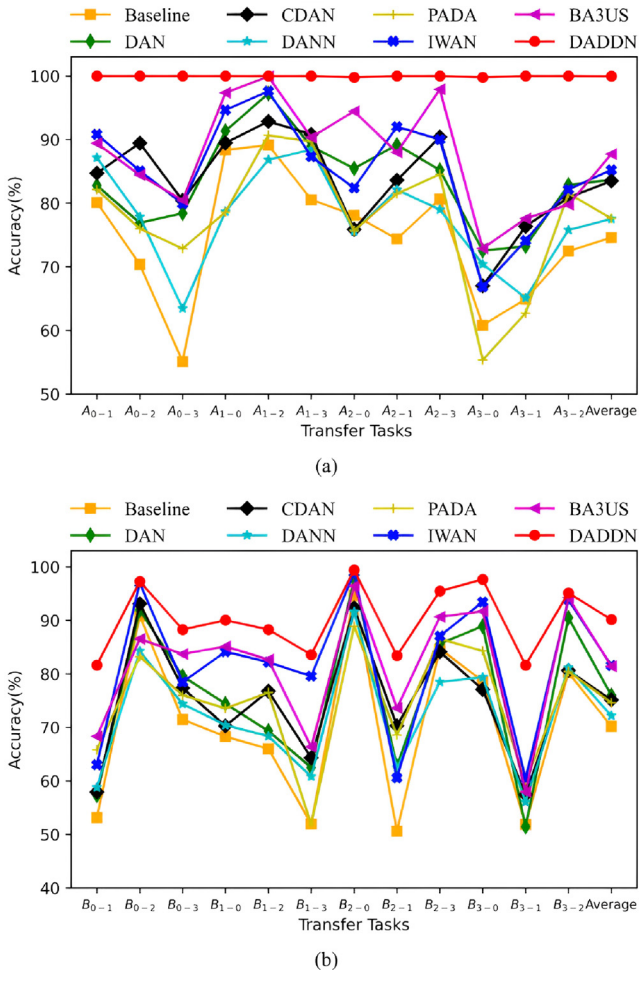

Fig. 6. Diagnostic accuracy across 12 cross-domain partitions: (a) CWRU dataset; (b) PU dataset.

accuracy of 90.312% and the smallest standard deviation, thus reflecting its stability. The diagnostic performance was not affected by the operating environment.

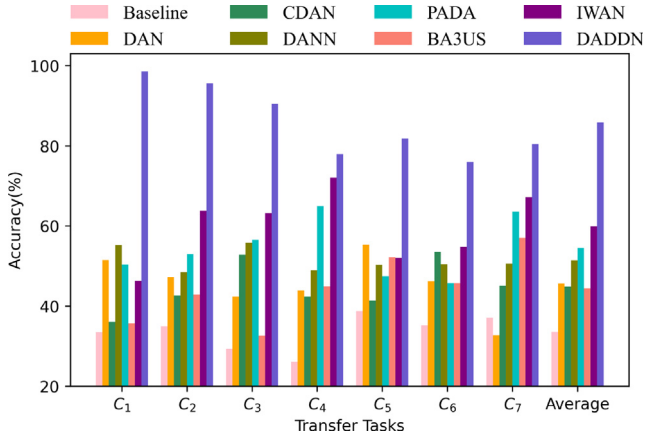

Fig. 7. Accuracy of diagnostic testing across seven cross-domain transfer tasks using SEU Gearbox dataset.

(ii) In some migration tasks, the diagnostic accuracy of the baseline was higher than that of the other DA-based methods, thus indicating that during the DA process, the outlier samples in the source domain were mismatched, which resulted in a negative transfer during model training.
(iii) Compared with the other three well-established partial transfer models (PADA, BA3US, and IWAN), DADDN achieved the highest diagnostic accuracy in each task. This indicates that in the PDA scenario, the DADDN can extract subtle features in the sample and obtain a clear feature decision boundary.

4.3.3. Case 3: Different missing scenarios

To comprehensively assess the diagnostic efficacy of different methods for cross-domain partial transfer tasks, we conducted tests on a planetary gear SEU dataset under different missing scenarios. Specifically, seven PDA tasks were created by manipulating the number of target sample categories. The diagnostic results are presented in Table 6 and Fig. 7. The diagnostic results indicate that the comparison methods achieved unsatisfactory diagnostic results in each transfer task. This can be attributed to the complex nature of the defect data in the planetary gear dataset, which resulted in highly similar feature information across different defect categories. Conventional domain-adaptive methods

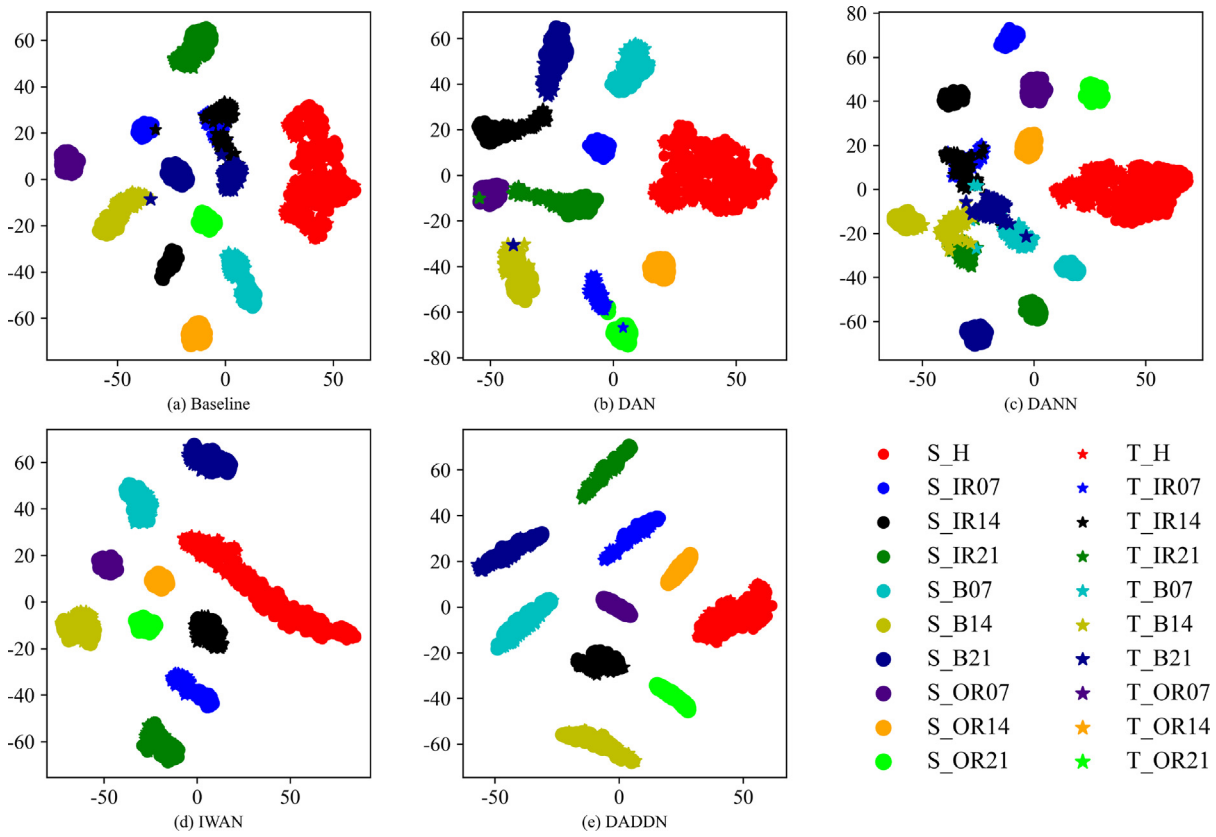

Fig. 8. Under the 2 hp → 3 hp (A_{2-3}) cross-domain transfer task, t-SNE visualizes five models as follows: (a) Baseline, (b) DAN (c) DANN, (d) IWAN, and (e) DADDN.

Table 6
Fault diagnosis accuracy on SEU gearbox dataset (%).

Tasks	C_1	C_2	C_3	C_4	C_5	C_6	C_7	Average
Baseline	33.463	35	29.375	26.093	38.769	35.152	37.109	33.5658
DAN	51.432	47.239	42.343	43.854	55.273	46.191	32.776	45.5868
CDAN	36.067	42.656	52.812	42.343	41.406	53.515	45.117	44.8451
DANN	55.208	48.437	55.781	48.906	50.284	50.488	50.585	51.3845
PADA	50.39	52.968	56.562	64.921	47.406	45.703	63.574	54.5034
BA3US	35.667	42.838	32.625	44.895	52.201	45.703	57.031	44.4228
IWAN	46.293	63.75	63.203	72.031	51.953	54.758	67.187	59.8821
DADDN	**98.567**	**95.625**	**90.468**	**77.968**	**81.835**	**75.976**	**80.468**	**85.8445**

cannot differentiate effective decision boundaries within an extracted feature space, thus ultimately resulting in unsatisfactory diagnostic results. By contrast, the proposed DADDN achieved an impressive average diagnostic rate of 85. 844%, which was more than 30% higher than those of other DA methods, thus demonstrating that the DADDN can effectively discriminate between the outlier and shared classes in the source and target domains through a dual-domain adversarial attention mechanism. In addition, the DADDN uses joint adaptive distributions to align subdomain-related categories, thus allowing for a more nuanced capture of feature information.

In summary, the results from the three different cross-domain partial transfer tasks, which were characterized by different speeds, operating conditions, and missing scenarios, showed that the proposed DADNN approach outperformed other methods based on DA in terms of diagnostic accuracy and model generalizability owing to the following three factors: (i) the balanced weighting strategy of the target domain addresses any adverse effect resulting from label space asymmetry; (ii) the category weighting mechanism of the source classifier effectively filters out outlier categories, thereby avoiding misalignment and negative transfers; and (iii) the joint adaptive distribution captures fine-grained information and uses it to achieve a well-defined decision boundary of the feature space.

4.3.4. Feature visualization analysis

To further validate the usefulness and advantages of the proposed method, four representative methods (baseline, DANN, DAN, and IWAN) were selected; additionally, T-distribution stochastic neighbor embedding (t-SNE) visualization and a confusion matrix were adopted to visually evaluate the diagnostic results of the transfer tasks. Fig. 8 shows the embedded features obtained by both the representative methods and the proposed DADNN for diagnostic task A_{2-3}. Fig. 9 shows the feature visualization results for transfer task B_{1-0}. Fig. 10 shows the feature visualization results for transfer task C_2.

Based on the results above, the following conclusions can be inferred. The DADDN exhibits exceptional performance in all three transfer visualization tasks, thus further validating the effectiveness of the dual-domain confrontation mechanism and fine-grained information matching method in cross-domain transfer tasks. In particular, the conventional DA-based approach is susceptible to outliers, thus resulting in false matches and a shift in the decision boundary of the feature space, which undermines its ability to distinguish between common and outlier classes in these two domains. Regarding the gearbox dataset, the representative methods showed insufficient domain adaptability, i.e., they failed to align related categories in different domains, thus highlighting the sensitivity of the conventional DA methods to complex information in cross-domain transfer tasks. Conversely, the proposed DADDN exhibited superior robustness to different missing scenarios, thus verifying its stability and generalization.

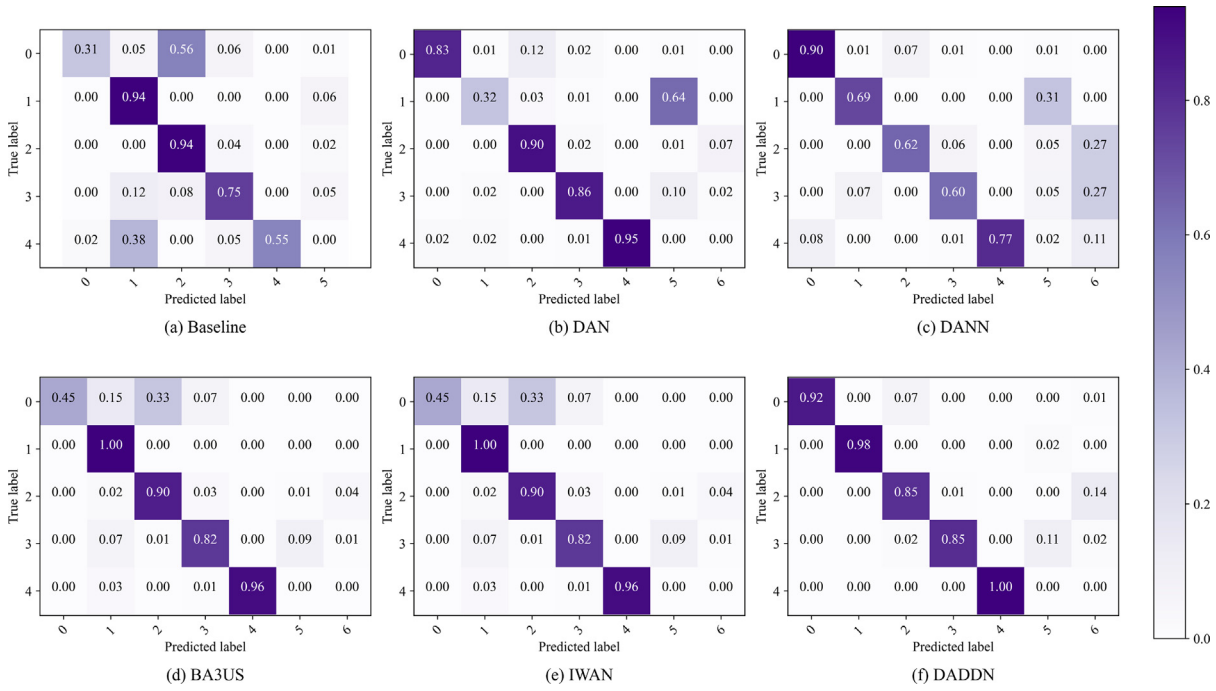

Fig. 9. Under the B_{1-0} cross-domain transfer task, the confusion matrix visualizes six models as follows: (a) Baseline, (b) DAN, (c) DANN, (d) BA3US, (e) IWAN, and (f) DADDN.

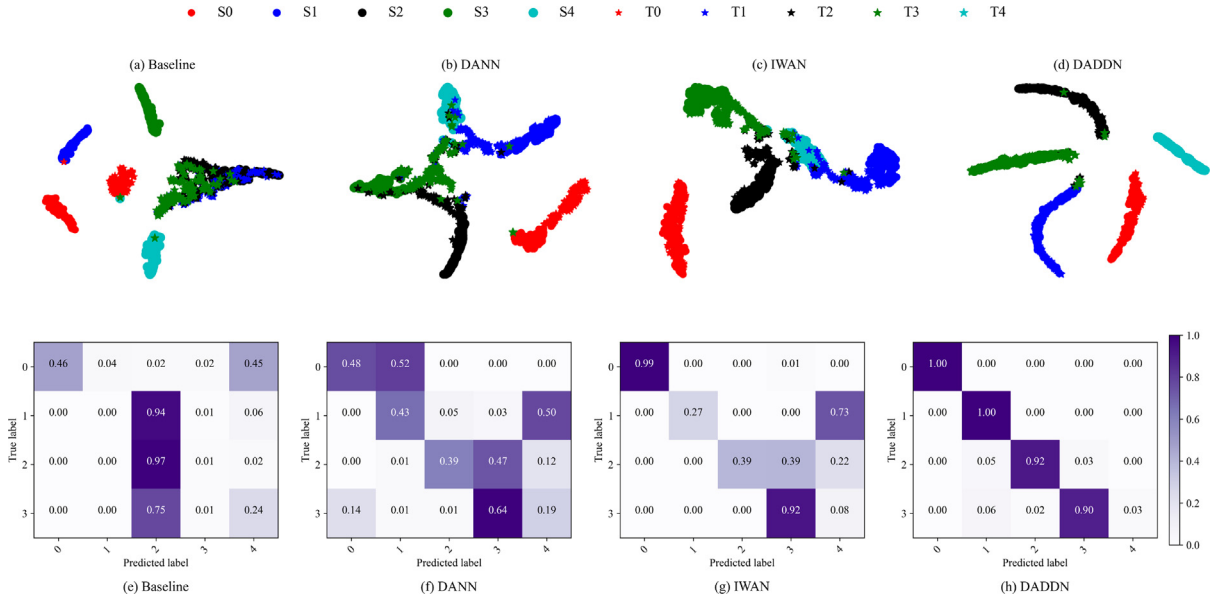

Fig. 10. Under the C_1 cross-domain transfer task, t-SNE and the confusion matrix visualize four model features as follows: Baseline for (a) and (e), DANN for (b) and (f), IWAN for (c) and (g), and DADDN for (d) and (h).

4.4. Ablation experiments

Next, we verified the effectiveness of different components of the proposed DADDN in the transfer task through ablation experiments. Specifically, three transfer tasks, namely A_{0-1}, B_{0-1}, and C_1, were selected as representatives of transfer cases in the different datasets. The CWMMD, JDDA, and DC components were removed from the DADDN, and their diagnostic accuracies were compared based on different diagnostic tasks to determine the effectiveness of each component. The details are provided in Table 7 and Fig. 11, from which the following information can be obtained:

(i) The absence of the JDDA component resulted in reductions of 10.94%, 14.65%, and 44.79% in the A_{0-1}, B_{0-1}, and C_1 tasks, respectively. This demonstrates the effectiveness of the JDDA in improving DA.

(ii) In the C_1 task, the average diagnostic accuracy in the absence of these components decreased by 30.99%, thus indicating that the proposed approach can effectively extract domain-invariant features and facilitate successful model transfer when managing large and complex data.

(iii) The absence of these three components resulted in varying degrees of reduced diagnostic performance in all three tasks, thereby highlighting the importance of effectively integrating

Table 7
Diagnostic accuracy of diagnostic ablation experiments for various components (%).

Tasks	Results	DADDN	Remove CWMMD	Remove JDDA	Remove $\mathcal{L}_{dc}$	Average
A_{0-1}	Accuracy	100%	95.54%	89.06%	92.63%	92.41%
	Variety		**−4.46%**	**−10.94%**	**−7.37%**	**−7.59%**
B_{0-1}	Accuracy	81.64%	69.92%	66.99%	74.61%	70.50%
	Variety		**−11.72%**	**−14.65%**	**−7.03%**	**−11.14%**
C_1	Accuracy	98.57%	60.55%	53.78%	88.41%	67.58%
	Variety		**−38.02%**	**−44.79%**	**−10.16%**	**−30.99%**

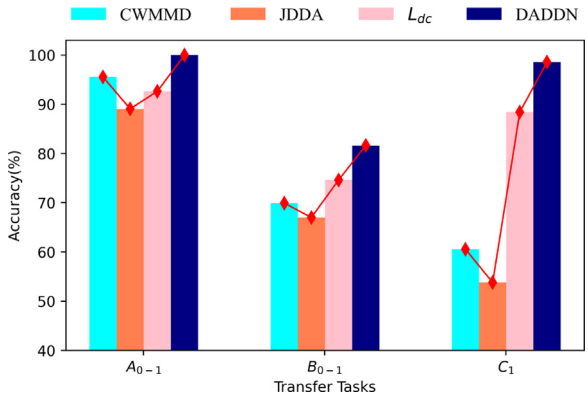

Fig. 11. Diagnostic accuracies of different ablation experiments for cross-domain transition scenarios A_{0-1}, B_{0-1}, and C_1.

these components to improve the feature learning capabilities of the proposed methodology.

5. Conclusion

Herein, we presented a novel transfer learning framework, i.e., the DADDN, which addressed the challenge of cross-domain partial transfer in mechanical fault diagnosis. The DADDN incorporates a dual-domain adversarial attention mechanism and a joint adaptive distribution with a weighting strategy. The former effectively separates the label space and evaluates the transferability of samples, whereas the latter improves domain confusion by detecting highly specific and discriminative subdomain features. In addition, we employed balanced center weighting mechanisms, category weighting mechanisms, and dynamic adaptive factors to remove outlier samples from the source domain, which resulted in clear feature decision boundaries. The DADDN delivered remarkable performance and robustness on the three datasets of rotating machines for PDA mapping.

This study was performed based on a known label space with certain limitations. In actual industrial applications, the label space is typically partially or completely unknown, i.e., open-set DA. In the future, we will integrate DA and multisource data fusion algorithms into the DADDN to investigate and improve the diagnostic application of the DADDN for open-set DA. In addition, the interpretability of the model established will be further investigated.

CRediT authorship contribution statement

Fuqiang Liu: Writing – review & editing, Visualization, Supervision, Funding acquisition. **Wenlong Deng:** Writing – original draft, Software, Methodology, Conceptualization. **Chaoqun Duan:** Validation, Software, Investigation. **Yi Qin:** Writing – review & editing, Funding acquisition. **Jun Luo:** Writing – review & editing, Funding acquisition. **Huayan Pu:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

The authors would like to thank the State Key Laboratory of Mechanical Transmissions of Chongqing University, China (Grant Nos. SKLMT-MSKFKT-202120 and SKLMT-ZZKT-2022M01) and the National Natural Science Foundation of China (Grant No. 62033001).

References

[1] X. Li, W. Zhang, N.-X. Xu, Q. Ding, Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places, IEEE Trans. Ind. Electron. 67 (8) (2019) 6785–6794.
[2] Q. Chen, J. Cao, S. Zhu, Data-driven monitoring and predictive maintenance for engineering structures: Technologies, implementation challenges, and future directions, IEEE Internet Things J. (2023).
[3] S. Sengupta, S. Basak, P. Saikia, S. Paul, V. Tsalavoutis, F. Atiah, V. Ravi, A. Peters, A review of deep learning with special emphasis on architectures, applications and recent trends, Knowl.-Based Syst. 194 (2020) 105596.
[4] A. Choudhary, R.K. Mishra, S. Fatima, B. Panigrahi, Multi-input CNN based vibro-acoustic fusion for accurate fault diagnosis of induction motor, Eng. Appl. Artif. Intell. 120 (2023) 105872.
[5] D. He, C. Liu, Z. Jin, R. Ma, Y. Chen, S. Shan, Fault diagnosis of flywheel bearing based on parameter optimization variational mode decomposition energy entropy and deep learning, Energy 239 (2022) 122108.
[6] M.-y. Zhong, Q.-y. Yang, Y. Liu, B.-y. Zhen, B.-b. Xie, et al., EEG emotion recognition based on TQWT-features and hybrid convolutional recurrent neural network, Biomed. Signal Process. Control 79 (2023) 104211.
[7] N.S. Ranawat, J. Prakash, A. Miglani, P.K. Kankar, Performance evaluation of LSTM and bi-LSTM using non-convolutional features for blockage detection in centrifugal pump, Eng. Appl. Artif. Intell. 122 (2023) 106092.
[8] Z. Jin, D. He, Z. Wei, Intelligent fault diagnosis of train axle box bearing based on parameter optimization VMD and improved DBN, Eng. Appl. Artif. Intell. 110 (2022) 104713.
[9] T. Han, C. Liu, W. Yang, D. Jiang, Deep transfer network with joint distribution adaptation: A new intelligent fault diagnosis framework for industry application, ISA Trans. 97 (2020) 269–281.
[10] Z. Zhao, Q. Zhang, X. Yu, C. Sun, S. Wang, R. Yan, X. Chen, Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study, IEEE Trans. Instrum. Meas. 70 (2021) 1–28.
[11] G. Li, L. Chen, J. Liu, X. Fang, Comparative study on deep transfer learning strategies for cross-system and cross-operation-condition building energy systems fault diagnosis, Energy 263 (2023) 125943.
[12] F. Liu, Y. Chen, W. Deng, M. Zhou, Entropy-optimized fault diagnosis based on unsupervised domain adaptation, Mathematics 11 (9) (2023) 2110.
[13] Y. Zhang, Z. Ren, K. Feng, K. Yu, M. Beer, Z. Liu, Universal source-free domain adaptation method for cross-domain fault diagnosis of machines, Mech. Syst. Signal Process. 191 (2023) 110159.
[14] Z. Su, J. Zhang, J. Tang, Y. Wang, H. Xu, J. Zou, S. Fan, A novel deep transfer learning method with inter-domain decision discrepancy minimization for intelligent fault diagnosis, Knowl.-Based Syst. 259 (2023) 110065.

[15] K. Zhao, F. Jia, H. Shao, A novel conditional weighting transfer wasserstein auto-encoder for rolling bearing fault diagnosis with multi-source domains, Knowl.-Based Syst. 262 (2023) 110203.
[16] Q. Qian, Y. Qin, J. Luo, Y. Wang, F. Wu, Deep discriminative transfer learning network for cross-machine fault diagnosis, Mech. Syst. Signal Process. 186 (2023) 109884.
[17] M. Long, Y. Cao, J. Wang, M. Jordan, Learning transferable features with deep adaptation networks, in: International Conference on Machine Learning, PMLR, 2015, pp. 97–105.
[18] M. Long, Z. Cao, J. Wang, M.I. Jordan, Conditional adversarial domain adaptation, Adv. Neural Inf. Process. Syst. 31 (2018).
[19] Y. Shi, A. Deng, M. Deng, M. Xu, Y. Liu, X. Ding, A novel multiscale feature adversarial fusion network for unsupervised cross-domain fault diagnosis, Measurement 200 (2022) 111616.
[20] Y. Liu, Y. Wang, T.W. Chow, B. Li, Deep adversarial subdomain adaptation network for intelligent fault diagnosis, IEEE Trans. Ind. Inform. 18 (9) (2022) 6038–6046.
[21] L. Wan, Y. Li, K. Chen, K. Gong, C. Li, A novel deep convolution multi-adversarial domain adaptation model for rolling bearing fault diagnosis, Measurement 191 (2022) 110752.
[22] C. Shen, J. Tian, J. Zhu, J. Shi, Z. Zhu, D. Wang, A new multisource domain bearing fault diagnosis method with adaptive dual-domain obfuscation weighting strategy, IEEE Trans. Instrum. Meas. 72 (2023) 1–11.
[23] Z. Cao, L. Ma, M. Long, J. Wang, Partial adversarial domain adaptation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 135–150.
[24] J. Zhang, Z. Ding, W. Li, P. Ogunbona, Importance weighted adversarial nets for partial domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8156–8164.
[25] P. Ge, C.-X. Ren, X.-L. Xu, H. Yan, Unsupervised domain adaptation via deep conditional adaptation network, Pattern Recognit. 134 (2023) 109088.
[26] A. Sahoo, R. Panda, R. Feris, K. Saenko, A. Das, Select, label, and mix: Learning discriminative invariant feature representations for partial domain adaptation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 4210–4219.
[27] Q. Qian, Y. Qin, J. Luo, S. Wang, Partial transfer fault diagnosis by multiscale weight-selection adversarial network, IEEE/ASME Trans. Mechatronics 27 (6) (2022) 4798–4806.
[28] Y. Zhang, J. Ji, Z. Ren, Q. Ni, F. Gu, K. Feng, K. Yu, J. Ge, Z. Lei, Z. Liu, Digital twin-driven partial domain adaptation network for intelligent fault diagnosis of rolling bearing, Reliab. Eng. Syst. Saf. 234 (2023) 109186.
[29] C. Yang, Y.-M. Cheung, J. Ding, K.C. Tan, B. Xue, M. Zhang, Contrastive learning assisted-alignment for partial domain adaptation, IEEE Trans. Neural Netw. Learn. Syst. (2022).
[30] S. Li, K. Gong, B. Xie, C.H. Liu, W. Cao, S. Tian, Critical classes and samples discovering for partial domain adaptation, IEEE Trans. Cybern. (2022).
[31] P. Guo, J. Zhu, Y. Zhang, Selective partial domain adaptation, in: 33rd British Machine Vision Conference, BMVC, 2022, pp. 1–13.
[32] J. Liang, Y. Wang, D. Hu, R. He, J. Feng, A balanced and uncertainty-aware approach for partial domain adaptation, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI, Springer, 2020, pp. 123–140.
[33] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, Domain-adversarial training of neural networks, J. Mach. Learn. Res. 17 (1) (2016) 1–35.
[34] W.A. Smith, R.B. Randall, Rolling element bearing diagnostics using the case western reserve university data: A benchmark study, Mech. Syst. Signal Process. 64 (2015) 100–131.
[35] C. Lessmeier, J. Kimotho, D. Zimmer, W. Sextro, KAt-data center, chair of design and drive technology, in: Chair of Design and Drive Technology, Paderborn Univ, 2019.
[36] S. Shao, S. McAleer, R. Yan, P. Baldi, Highly accurate machine fault diagnosis using deep transfer learning, IEEE Trans. Ind. Inform. 15 (4) (2018) 2446–2455.