ORIGINAL ARTICLE

Dynamic Distribution Adaptation Based Transfer Network for Cross Domain Bearing Fault Diagnosis

Yixiao Liao, Ruyi Huang, Jipu Li, Zhuyun Chen and Weihua Li[*]

Abstract

In machinery fault diagnosis, labeled data are always difficult or even impossible to obtain. Transfer learning can leverage related fault diagnosis knowledge from fully labeled source domain to enhance the fault diagnosis performance in sparsely labeled or unlabeled target domain, which has been widely used for cross domain fault diagnosis. However, existing methods focus on either marginal distribution adaptation (MDA) or conditional distribution adaptation (CDA). In practice, marginal and conditional distributions discrepancies both have significant but different influences on the domain divergence. In this paper, a dynamic distribution adaptation based transfer network (DDATN) is proposed for cross domain bearing fault diagnosis. DDATN utilizes the proposed instance-weighted dynamic maximum mean discrepancy (IDMMD) for dynamic distribution adaptation (DDA), which can dynamically estimate the influences of marginal and conditional distribution and adapt target domain with source domain. The experimental evaluation on cross domain bearing fault diagnosis demonstrates that DDATN can outperformance the state-of-the-art cross domain fault diagnosis methods.

Keywords: Cross domain fault diagnosis, Dynamic distribution adaptation, Instance-weighted dynamic MMD, Transfer learning

1 Introduction

As a critical part of modern equipment, bearing always works under harsh conditions and suffers from time-varying load, which results in a significant risk of failure [1]. Bearing failure is the main cause of the machinery breakdowns, and sometimes can lead to huge economic loss and severe casualties [2, 3]. To ensure the operational reliability of equipment, researchers have conducted related studies for bearing fault diagnosis and proposed many effective methods [4–8].

Among these methods, deep learning based methods have shown its excellent performance in recent years, which can learn diagnosis knowledge from large amount of labeled data and reduce the dependence on expertise [9–12]. Although deep learning based methods are not expertise-dependent, they are heavily data-dependent. Unfortunately, collecting labeled machinery failure data are expensive or even impossible [13]. Under such circumstance, transfer learning [14] begins to attract researchers' attention, which could transfer the related knowledge of fully labeled source domain to enhance the fault diagnosis performance in sparsely labeled or unlabeled target domain [15, 16].

Han et al. [17] introduced adversarial learning for feature distribution adaptation and transferred the source domain fault diagnosis model to the target domain. Guo et al. [18] utilized both Maximum mean discrepancy (MMD) and adversarial learning to adapt the feature distribution, which can transfer the fault diagnosis model to other machineries. Yang et al. [19] transferred fault diagnosis model from laboratory bearings to locomotive

*Correspondence: whlee@scut.edu.cn
School of Mechanical & Automotive Engineering, South China University of Technology, Guangzhou 510641, China
The original version of this article was revised: the affiliation has been updated.

Liao *et al. Chin. J. Mech. Eng.* (2021) 34:52

Page 2 of 10

bearings with MMD-based multi-layer feature alignment and pseudo label learning. Wang et al. [20] utilized conditional maximum mean discrepancy (CMMD) to align feature distribution for cross-domain bearing fault diagnosis. Li et al. [21] proposed a novel cross domain fault diagnosis method, which take full advantage of the availability of target domain health labels.

These transfer learning methods have achieved great success in unsupervised domain adaptation, which can build fault diagnosis models for unlabeled target domain. However, these works only tend to adapt either marginal or conditional distributions (MDA or CDA) between source and target domain. In practice, both marginal and conditional distribution discrepancies have significant but different influences on domain divergence [22]. Recently, researchers have carried out some work in joint distribution adaptation (JDA) [23, 24], which simultaneously adapt marginal and conditional distribution. Although these works have achieved better performance, they allocate equal weights to marginal and conditional distributions discrepancies, which cannot quantify the different contributions of these distributions discrepancies.

In this paper, a dynamic distribution adaptation based transfer network (DDATN) is proposed for cross domain bearing fault diagnosis, which utilizes the proposed instance-weighted dynamic maximum mean discrepancy (IDMMD) for dynamic distribution adaptation (DDA). The main contributions of the paper are as follows.

(1) Introduce DDA framework for cross domain bearing fault diagnosis, which can dynamically adjust the weights of marginal and conditional distributions discrepancies in domain adaptation.
(2) Propose a novel dynamic distribution discrepancy metric (IDMMD) for unsupervised DDA. IDMMD uses a novel dynamic factor estimation method to dynamically estimate the contributions of MDA and CDA, which further considers the contribution of CDA of each class. In addition, it takes the confidence of target domain pseudo labels into account when calculates the conditional distribution discrepancy.

The remainder of the paper are organized as follows. The theoretical and technical bases are introduced in Section 2. Section 3 describes the detail of the proposed DDATN, which has been experimental evaluated in Section 4. Finally, the conclusion is drawn in Section 5.

2 Preliminaries

2.1 Dynamic Distribution Adaptation

Marginal and conditional distributions have different contributions on domain divergence and their contributions dynamically change during the transfer learning procedures. To improve transfer learning performance, DDA [22] is proposed as a general transfer learning framework, which considers the different and ever-changing contributions of marginal and conditional distributions on domain divergence. In DDA, the dynamic distribution discrepancy has the general form as

$$
\begin{aligned}
D_d(\Omega_s, \Omega_t) = (1 - \mu)D(P_s, P_t) + \\
\mu \sum_{c=1}^{C} D^{(c)}(Q_s, Q_t),
\end{aligned}
\tag{1}
$$

where P_s and Q_s are marginal and conditional distributions of source domain Ω_s, respectively; P_t and Q_t are marginal and conditional distributions of source domain Ω_t, respectively; $D(P_s, P_t)$ is marginal distribution discrepancy, $D^{(c)}(Q_s, Q_t)$ is conditional distribution discrepancy for class c, C is the number of classes; μ is the dynamic weight which changes when the training goes on.

From Eq. (1), DDA degenerates to MDA and CDA when $\mu=0$ and $\mu=1$, respectively. Therefore, DDA can be regarded as a more general distribution adaptation framework.

2.2 Maximum Mean Discrepancy

Maximum mean discrepancy (MMD) [25] which is an effective distribution discrepancy metric widely used in transfer learning. Given datasets X_s and X_t sampled from distributions $P(X_s)$ and $P(X_t)$, the MMD between $P(X_s)$ and $P(X_t)$ is can be calculated as

$$
MMD_{\mathcal{H}}(P(X_s), P(X_t)) = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(x_i^s) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi(x_j^t) \right\|_{\mathcal{H}}^2,
\tag{2}
$$

where $x_i^s \in X_s$, $x_j^t \in X_t$; n_s and n_t are the numbers of samples in X_s and X_t, respectively; ϕ is a nonlinear mapping function in reproducing kernel Hilbert space (RKHS) $\mathcal{H}$.

From Eq. (2), the MMD is expressed as the distance in $\mathcal{H}$ between mean embeddings of X_s and X_t.

Liao *et al. Chin. J. Mech. Eng.* (2021) 34:52

Page 3 of 10

Figure 1 Framework of DDATN
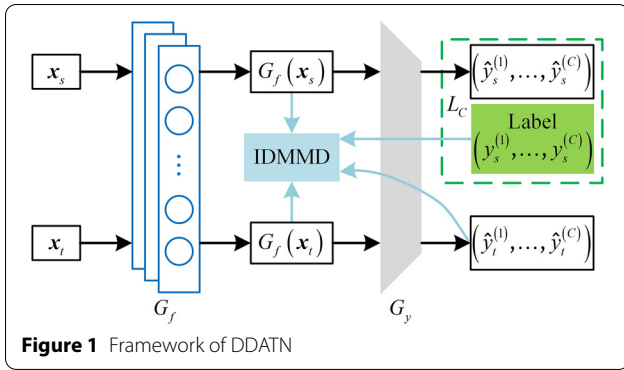
3 Dynamic Distribution Adaptation Based Transfer Network

In this paper, DDA is introduced for improving the cross domain bearing fault diagnosis performance. The framework of DDATN is shown in Figure 1. It consists of DDA and supervised learning. The DDA part aims to constrain the feature extractor G_f to extract domain-invariant features by minimizing the proposed dynamic distribution discrepancy IDMMD. The supervised learning part realized by minimizing the supervised loss L_C will guide the G_f to extract features which are discriminative for bearing health conditions, and train the effective classifier G_y which can accurately diagnosis bearing fault with these features.

3.1 Supervised Learning

The DDATN is proposed for unsupervised domain adaptation, which target domain data are totally unlabeled. The supervised learning is realized using labeled source domain data, whose loss can be defined as

$$L_C = \frac{1}{n_s} \sum_{i=1}^{n_s} J\big(G_y\big(G_f\big(\boldsymbol{x}_i^s\big)\big), \boldsymbol{y}_i^s\big), \tag{3}$$

where $J(\cdot, \cdot)$ is cross-entropy loss function, $\boldsymbol{y}_i^s$ is the labeled of source domain sample $\boldsymbol{x}_i^s$.

3.2 Instances-weighted Dynamic Maximum Mean Discrepancy (IDMMD)

In unsupervised domain adaptation, target domain cannot provide label information. The final fault diagnosis process can just be conducted by the shared classifier G_y which trained by labeled source domain data. To prevent the interference of target domain specific features and domain divergence, it is important to extract the domain-invariant features.

In DDATN, a novel dynamic distribution discrepancy metric IDMMD is proposed for constraining G_f to extract domain-invariant features. IDMMD based on the DDA framework, which considers the ever-changing contributions of marginal and conditional distribution discrepancy on domain divergence. In addition, IDMMD further considers the different contributions of conditional distributions discrepancies of different class, and the confidence of target domain samples' pseudo labels. The IDMMD between Ω_s and Ω_t is defined as:

$$IDMMD(\Omega_s,\ \Omega_t) = \left(1 - \sum_{c=1}^{C} \mu^{(c)}\right) IDMMD_M(\Omega_s, \Omega_t)$$
$$+ \sum_{c=1}^{C} \mu^{(c)} DMMD_C^{(c)}(\Omega_s, \Omega_t), \tag{4}$$

where $IDMMD_M$ is the marginal distributions discrepancy, $IDMMD_C^{(c)}$ is the conditional distributions discrepancy for class c, $\mu^{(c)}$ is the dynamic factor for $IDMMD_C^{(c)}$. They are defined as

$$IDMMD_M(\Omega_s, \Omega_t) = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi\big(\boldsymbol{x}_i^s\big) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi\big(\boldsymbol{x}_j^t\big) \right\|_{\mathcal{H}}^2, \tag{5}$$

$$IDMMD_C^{(c)}(\Omega_s, \Omega_t) = \left\| \frac{\sum_{i=1}^{n_s} y_i^{(c)} \phi\big(\boldsymbol{x}_i^s\big)}{\sum_{i=1}^{n_s} y_i^{(c)}} - \frac{\sum_{j=1}^{n_t} \hat{y}_j^{(c)} \phi\big(\boldsymbol{x}_j^t\big)}{\sum_{j=1}^{n_t} \hat{y}_j^{(c)}} \right\|_{\mathcal{H}}^2, \tag{6}$$

$$\mu^{(c)} = \frac{IDMMD_C^{(c)}(\Omega_s, \Omega_t)}{IDMMD_M(\Omega_s, \Omega_t) + \sum_{c=1}^{C} IDMMD_C^{(c)}(\Omega_s, \Omega_t)}, \tag{7}$$

where $y_i^{(c)}$ is the real one-hot label of $\boldsymbol{x}_i^s$ for class c, $\hat{y}_j^{(c)}$ is the prediction probabilities of $\boldsymbol{x}_j^t$ for class c.

The $IDMMD_M$ is the original form of MMD. For unsupervised domain adaptation, the labels of target domain samples which are necessary for calculating conditional distributions discrepancy is unavailable. Therefore, the predictions of Ω_t are regarded as its soft labels. Considering the confidence of the soft labels, different weights are allocated to different target domain samples while calculating $IDMMD_C^{(c)}$, which are their prediction probabilities for class c. Intuitively, MMD calculates the distance between the centers of two datasets in the embedded feature space. In target

domain, the center of each class will tend to be closer to the samples which have higher prediction probabilities with the proposed weight allocation. Therefore, the negative effects of misclassification will be diminished.

To quantify the ever-changing contributions of marginal and conditional distributions, the dynamic factors for each class are calculated as Eq. (7). The class has larger $IDMMD_C$ value will be allocated larger dynamic factor, and the dynamic factor of marginal distribution will be calculated as Eq. (4). The proposed dynamic factor allocation method aims at guiding the DDA to focus on the main cause of domain shift.

3.3 General Procedure of DDATN

As mentioned above, DDATN contain two parts: supervised learning and DDA. Therefore, the total loss function of DDATN can be defined as

$$L_{\text{total}} = L_C + \lambda \cdot IDMMD(\Omega_s, \Omega_t), \tag{8}$$

where λ is the trade-off factor.

The procedures of DDATN are presented in Figure 2 and summarized as follows.

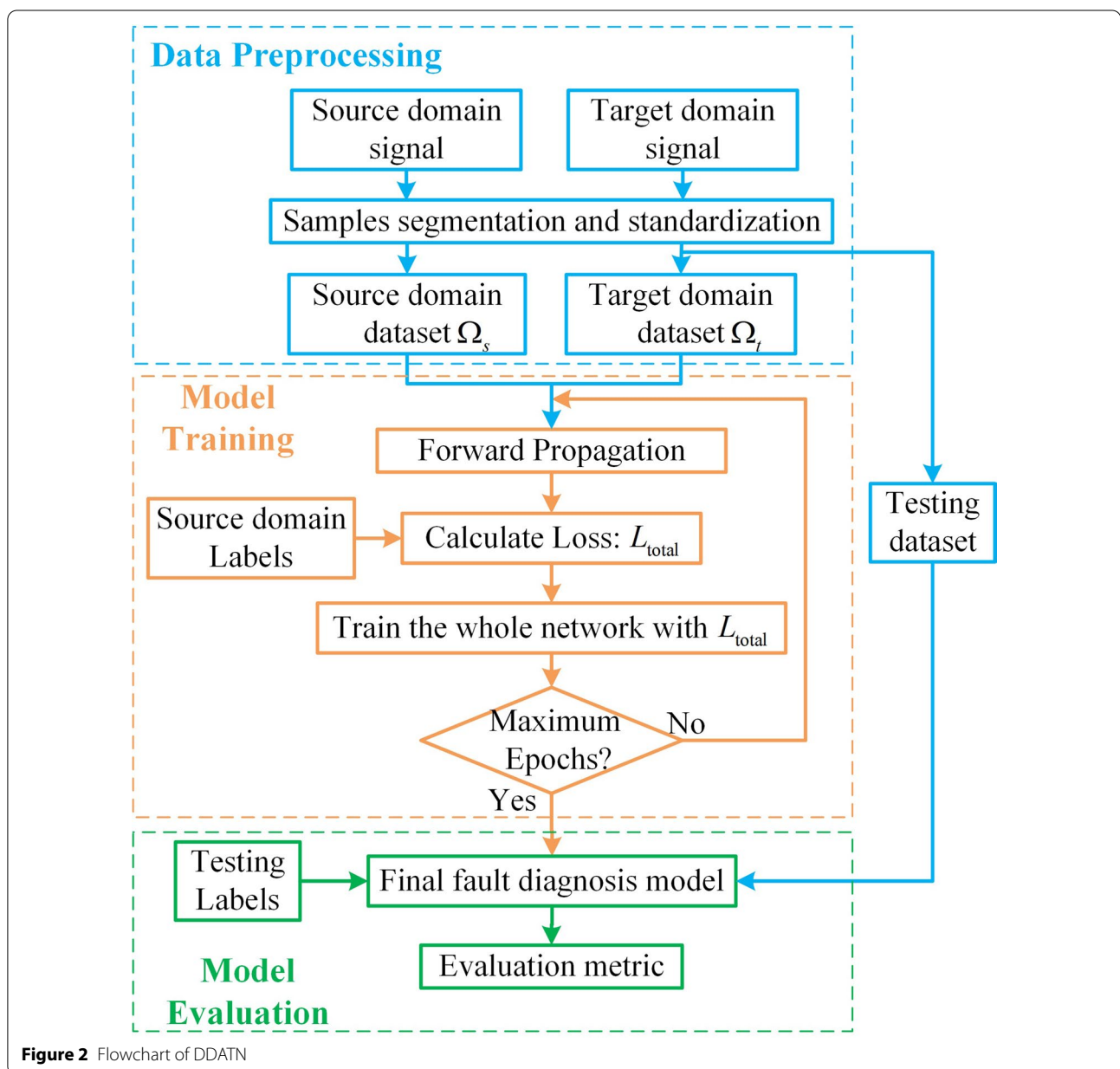(1) Datasets generation. The source and target domain signals are segmented and standardized to form
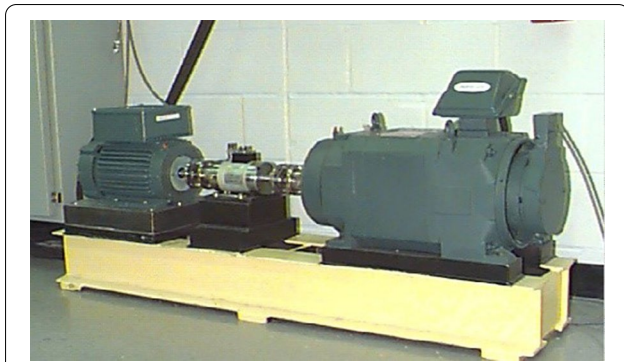

Figure 2 Flowchart of DDATN

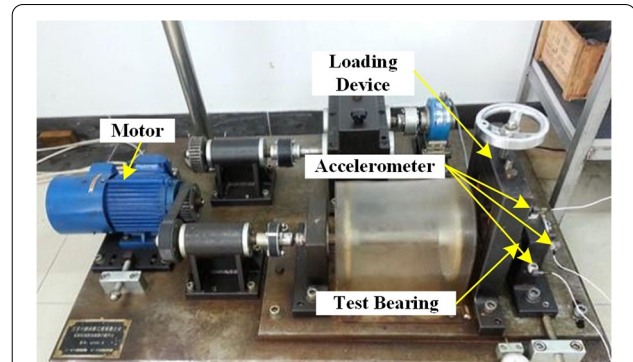Liao *et al. Chin. J. Mech. Eng.*   (2021) 34:52

Page 5 of 10

Figure 3 CWRU bearing test rig

Figure 4 Bearing test rig

Table 1 Details of the CWRU bearing dataset

Health conditions	Fault diameter /mm	Load /hp	Speed /(r/min)
Normal (N)	0.18/0.36/0.53	1	1772
Inner race fault (IR)		2	1750
Outer race fault (OR)		3	1730

Table 2 Details of the OL bearing dataset

Health conditions	Fault diameter /mm	Load /hp	Speed /(r/min)
Normal (N)	0.5	/	500/800/1100/1400
Inner race fault (IR)			
Outer race fault (OR)			

the source and target domain datasets (Ω_s and Ω_t), respectively.

(2) Batch optimization. Source and target domain samples batches are generated from Ω_s and Ω_t, respectively. These batches are forward propagated to calculate the loss L_{total}. The loss is back-propagated to update the whole network.

(3) Traverse datasets. Repeat step 2 to traverse the Ω_s and Ω_t.

(4) Iterative optimization. Repeat step 3 for n_{epochs} times.

(5) Model evaluation. Evaluate the final fault diagnosis model with testing dataset.

4 Experiment
4.1 Dataset Description
In this section, the CWRU [26] bearing dataset (CW) and the bearing dataset from our laboratory (OL) are utilized for verifying the effectiveness of the DDATN.

The test rig of CWRU bearing dataset is shown in Figure 3, which consists of a driven motor (left), a torque transducer and an encoder (middle), and a dynamometer (right). The test bearing is installed at the output side of the motor and support the motor shaft. This bearing test includes four health conditions with four fault diameters

(0.18, 0.36, 0.53, 0.071 mm), and it is conducted on four different loads (0, 1, 2, 3 hp) with sampling rates of 12 kHz and 48 kHz. The details of the used part of data whose sampling rate is 12 kHz are listed in Table 1.

The bearing test rig of our laboratory is shown in Figure 4. The test rig is driven by the motor, and the power transfer to the shaft which is supported by the test bearing with belt drive. The loading device exert radial force on the shaft to simulate the load of bearing. In this test, inner and outer race faults with size 0.5 mm are introduced to the test bearing by wire-electrode cutting. The acceleration signals are collected with sampling rate of 12 kHz. The details of this dataset are listed in Table 2.

For each health condition, 100 samples are segmented from the original vibration signals. Therefore, there are 300 and 1200 (300 for each speed) sampled from CW and OL bearing datasets, respectively. The length of the sample is set as 2048.

4.2 Comparison Setting
Thirty-six cross equipment tasks are conducted to verify the effectiveness of DDATN, which are listed in Table 3. For target domain dataset, half are used for training and the rest are served as testing dataset. In Table 3, S denotes source domain dataset, T denotes target domain training

Liao *et al. Chin. J. Mech. Eng.*   (2021) 34:52

Page 6 of 10

Table 3 Cross equipment tasks

Task	Training datasets		Testing dataset
	S	T	
1	CW0.18_1 (300)	OL500 (150)	OL500 (150)
2		OL800 (150)	OL800 (150)
3		OL1100 (150)	OL1100 (150)
4		OL1400 (150)	OL1400 (150)
5	CW0.18_2 (300)	OL500 (150)	OL500 (150)
6		OL800 (150)	OL800 (150)
7		OL1100 (150)	OL1100 (150)
8		OL1400 (150)	OL1400 (150)
9	CW0.18_3 (300)	OL500 (150)	OL500 (150)
10		OL800 (150)	OL800 (150)
11		OL1100 (150)	OL1100 (150)
12		OL1400 (150)	OL1400 (150)
13	CW0.36_1 (300)	OL500 (150)	OL500 (150)
14		OL800 (150)	OL800 (150)
15		OL1100 (150)	OL1100 (150)
16		OL1400 (150)	OL1400 (150)
17	CW0.36_2 (300)	OL500 (150)	OL500 (150)
18		OL800 (150)	OL800 (150)
19		OL1100 (150)	OL1100 (150)
20		OL1400 (150)	OL1400 (150)
21	CW0.36_3 (300)	OL500 (150)	OL500 (150)
22		OL800 (150)	OL800 (150)
23		OL1100 (150)	OL1100 (150)
24		OL1400 (150)	OL1400 (150)
25	CW0.53_1 (300)	OL500 (150)	OL500 (150)
26		OL800 (150)	OL800 (150)
27		OL1100 (150)	OL1100 (150)
28		OL1400 (150)	OL1400 (150)
29	CW0.53_2 (300)	OL500 (150)	OL500 (150)
30		OL800 (150)	OL800 (150)
31		OL1100 (150)	OL1100 (150)
32		OL1400 (150)	OL1400 (150)
33	CW0.53_3 (300)	OL500 (150)	OL500 (150)
34		OL800 (150)	OL800 (150)
35		OL1100 (150)	OL1100 (150)
36		OL1400 (150)	OL1400 (150)

dataset, OL 500 (150) denotes the OL bearing data with speed of 500 r/min and the number of samples are 150 (50 samples for each health condition). CW0.18_1(300) denotes the CWRU data with 0.18 mm fault diameter and 1 hp load, and the number of samples are 300 (100 samples for each health condition).

The structures of features extractor G_f and classifier G_y are presented in Table 4, where Conv1D denotes 1D convolutional layer, MP1D denotes 1D max pooling layer, FC denotes fully connected layer. The G_f and G_y are trained by Adam optimizer (learning rate$=0.001$, $\beta_1=0.9$, $\beta_2=0.999$). The tradeoff factor λ is set as 1.

The details of the comparison methods are as following. They use the same CNN structure as DDATN.

> Method 1 (DDC): Deep domain confusion (DDC) [27] is a deep transfer learning method proposed by Tzeng et al., which utilizes MMD for single-layer feature alignment.
> Method 2 (FTNN): Feature-based transfer neural network (FTNN) [19] is proposed by Yang et al, which applied MMD-based multi-layer feature alignment and pseudo label learning to transfer fault diagnosis knowledge from laboratory bearings to locomotive bearings.
> Method 3 (DTN) [23]: Deep transfer network (DTN) is a cross-domain fault diagnosis method proposed by Han et al. It utilizes MMD and CMMD to evaluate the marginal and conditional distributions discrepancies, respectively. They are given equal weights for single-layer feature joint distribution adaptation.
> Method 4 (IWC): IWC is derived from DDATN, which only takes the conditional part of IDMMD as the evaluation of the domain divergence.
> Method 5 (IWCM): IWCM is also derived from DDATN, which allocates equal weights to the marginal and conditional parts of the IDMMD.

4.3 Result and Discussion

The experiment is conducted on a computer with two E5-2630 v3 CPUs, a Nvidia GeForce RTX 2080 Ti GPU (11 GB memory), and 64 GB memory. To avoid the influence of randomness, each task is repeated 10 times. The mean accuracies, standard deviations, training and testing time are listed in Table 5. The overall accuracy curves of these methods are presented in Figure 5. In addition, for each method, the average accuracy and standard deviation among all tasks are also presented in Avg.

The comparison shows that IWC, IWCM and DDATN have better performances than other methods. DDC has the worst accuracy in in almost all tasks except task 14. FTNN and DTN are both derived from DDC, whereas they have different improving directions. FTNN extends the single-layer adaptation to multi-layer adaptation and

Liao *et al. Chin. J. Mech. Eng.*   (2021) 34:52

Page 7 of 10

Table 4 Structures of G_f and G_y

	Layer type	Activation function	Parameter name	Parameter size	Output size
G_f	Input	/	/	/	(2048,1)
	Conv1D_1	ReLU	Kernels	16×129	(1920,16)
	MP1D_1	/	Pool Size	8	(240,16)
	Conv1D_2	ReLU	Kernels	32×5	(240,32)
	MP1D_2	/	Pool Size	4	(60,32)
	Conv1D_3	ReLU	Kernels	32×5	(60,32)
	MP1D_3	/	Pool Size	4	(15,32)
	Conv1D_4	ReLU	Kernels	64×15	(1,64)
	Flatten	/	/	/	(64)
G_y	FC_1	Sigmoid	Nodes	32	(32)
	FC_2	Leaky_ReLU (α=0.2)	Nodes	16	(16)
	Output	SoftMax	Nodes	3	(3)

introduces pseudo label learning for further improvement, which has demonstrated to be effective in this case. DTN extends the MDA to JDA and achieves higher average accuracy than FTNN.

IWC achieves the average accuracy of 91.24% with standard deviation of 11.16%, which indicates that the proposed conditional distribution discrepancy metric is effective and robust. IWCM shows better performance than IWC in all tasks. The extension from CDA to JDA is proved valid while comparing IWCM with IWC. In addition, the IWCM can be regarded as the variation of DTN, which replaces the pseudo label strategy with instance-weighted strategy when calculating conditional distribution discrepancy. The comparison between IWCM and DTN demonstrates the effectiveness of the instance-weighted strategy.

Specifically, DDATN outperforms other methods in all tasks and achieves the highest average accuracy of 98.43% with lowest standard deviation, which indicates its superior effectiveness and robustness. In tasks 3, 7, 13, 21, 23, DDATN does not perform best, but it still gains a very close accuracy with the highest one. In tasks 4 (CW0.18_1 to OL1400) and 8 (CW0.18_2 to OL1400), the accuracies of DDATN are relatively low (67.07% for both tasks), which indicates that DDATN cannot gain very high accuracy in some transfer tasks. However, the accuracies of DDATN in these tasks still higher than other methods. In some difficult transfer tasks, DDATN may not be able to gain very high accuracy, but it can improve the performance to some degree.

In summary, the comparison indicates that the proposed conditional distribution discrepancy metric is effective and robust, whereas the extension from MDA, CDA and JDA to DDA can further improve the cross-domain fault diagnosis performance.

4.4 Feature Visualization

All the methods used in this experiment are feature-based transfer learning methods. To further evaluate the feature alignment performance of DDATN, t-distributed stochastic neighbor embedding (t-SNE) [28] is utilized for feature visualization. Tasks 33 and 36 are selected for visualization. For DDC, IWC, IWCM and DDATN, the feature visualizations are conducted on Flatten layer, whereas it is conducted on FC_2 layer for FTNN and DTN. In Figures 6 and 7, the legend consists of two parts: bearing health condition (outside the bracket) and the domain label (inside the bracket). For example, IR (T) denotes the inner race fault sample of the target domain. The marker of source and target domain samples are circle and triangle, respectively. The color represents the health condition, e.g., blue represent Normal (N), red represent Inner Race fault (IR), green represent Outer Race fault (OR).

In task 33, the features of IWC, IWCM and DDATN show good fusion of source and target domains, whereas great discriminability with respect to bearing health conditions is also observed. For other methods, their features still have related good interclass discriminability, but the aggregation of source and target domains is poor. Especially, the source and target domain samples can be linearly separated with the feature of DDC.
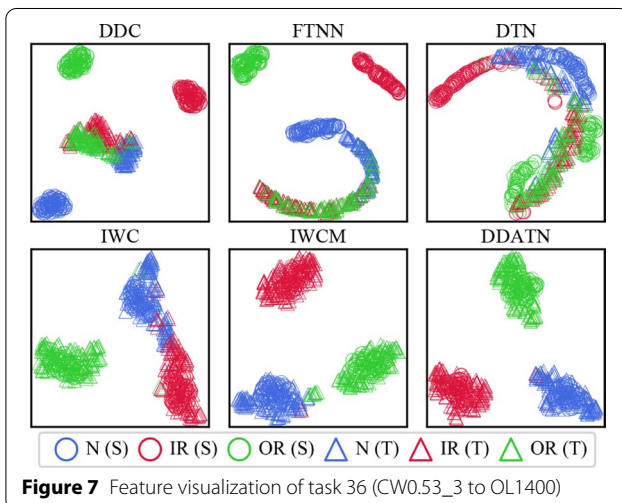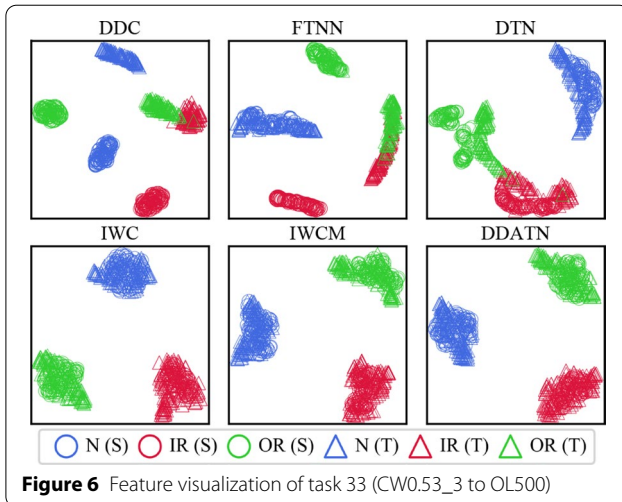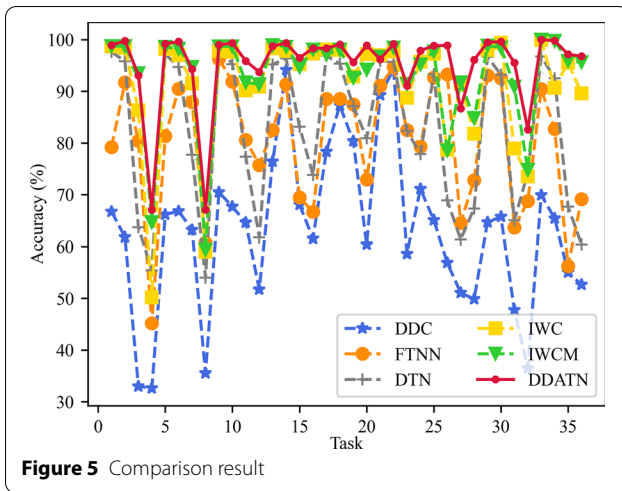
Liao *et al. Chin. J. Mech. Eng.* (2021) 34:52

Page 8 of 10

Table 5 Comparison results

Tasks	Methods					
	DDC	FTNN	DTN	IWC	IWCM	DDATN
Accuracy (%)						
1	66.73±3.49	79.20±9.82	97.60±3.80	98.67±1.83	98.73±1.55	**98.93±0.72**
2	61.80±2.41	91.73±3.22	95.80±4.96	98.40±2.25	98.73±2.21	**99.73±0.47**
3	32.97±2.70	80.40±8.60	63.70±15.97	86.27±7.32	**93.47±2.08**	93.07±1.64
4	32.60±1.84	45.20±18.08	55.40±9.61	50.20±7.58	64.67±5.38	**67.07±1.86**
5	66.13±2.72	81.33±7.77	97.20±2.66	98.22±3.83	98.60±2.36	**99.20±0.61**
6	66.80±5.11	90.47±5.43	94.70±5.57	97.01±3.10	98.20±2.25	**99.60±0.56**
7	63.20±3.78	87.93±9.22	77.73±9.90	91.53±3.21	**94.67±2.63**	94.27±2.83
8	35.53±2.61	60.73±17.34	54.00±7.48	59.00±10.77	59.40±9.26	**67.07±2.58**
9	70.47±5.36	96.27±2.04	98.47±2.39	97.77±3.79	98.67±2.24	**99.00±0.79**
10	67.67±5.65	91.87±3.12	95.27±3.89	97.60±2.25	98.67±2.20	**99.33±0.70**
11	64.60±2.21	80.60±6.43	77.33±8.57	90.20±5.37	91.73±3.16	**95.87±1.03**
12	51.73±2.67	75.73±8.53	61.73±6.49	90.87±5.91	91.33±5.75	**93.73±5.73**
13	76.40±9.78	82.47±4.35	95.20±9.56	98.30±1.95	**98.93±2.04**	98.73±1.39
14	94.13±2.26	91.33±3.72	96.37±5.69	97.80±1.69	98.60±1.52	**99.33±0.83**
15	68.07±2.54	69.47±5.94	83.13±6.46	95.27±2.25	94.67±2.90	**96.47±1.26**
16	61.53±3.08	66.67±4.92	73.80±7.17	97.27±3.23	98.07±2.19	**98.33±1.05**
17	78.33±3.97	88.53±4.99	96.87±4.59	98.07±4.55	97.13±3.08	**98.33±1.94**
18	87.27±5.63	88.53±2.06	95.33±3.36	98.01±2.37	98.07±1.97	**99.07±1.10**
19	80.33±6.43	87.47±3.74	87.13±2.83	92.87±2.35	92.60±2.28	**95.60±1.41**
20	60.40±3.15	72.93±5.61	80.93±5.35	97.20±3.12	94.20±4.79	**98.87±0.83**
21	89.27±7.28	91.07±3.93	95.87±3.51	**96.96±4.42**	96.67±5.69	96.20±3.34
22	94.20±2.42	94.67±1.54	95.67±2.42	97.73±2.11	98.47±1.63	**99.13±0.55**
23	58.60±3.94	82.47±6.56	82.33±4.17	88.73±6.09	**91.40±1.73**	91.07±3.36
24	71.13±3.21	79.27±5.59	77.87±7.32	95.73±3.81	95.20±1.96	**97.87±1.83**
25	65.07±1.97	92.67±2.58	93.30±5.32	97.26±3.31	97.73±2.20	**98.80±1.03**
26	56.87±2.69	93.27±2.56	68.90±2.06	78.67±6.93	78.53±7.63	**98.87±2.50**
27	51.07±1.84	64.53±6.86	61.33±2.79	91.47±4.48	**91.73±3.11**	86.67±4.96
28	49.87±2.43	72.80±3.44	67.33±4.42	81.80±10.48	84.80±10.88	**96.07±2.66**
29	64.67±2.06	93.07±4.03	96.67±3.46	98.00±4.53	98.80±2.43	**99.47±0.82**
30	65.73±2.04	92.53±2.61	93.20±9.05	99.40±0.66	98.47±4.16	**99.60±0.72**
31	47.73±4.36	63.67±9.19	65.03±2.67	78.93±12.77	90.87±4.38	**95.53±2.97**
32	36.47±2.29	68.73±9.74	74.20±12.79	73.53±12.84	74.80±10.66	**82.60±5.37**
33	69.93±6.72	90.40±6.61	96.73±1.59	99.93±0.22	**100.00±0.00**	100.00±0.00
34	65.40±1.87	82.73±7.53	92.53±6.15	90.67±11.88	99.73±0.34	**99.80±0.32**
35	55.00±3.10	56.23±3.69	67.67±9.22	95.53±1.81	95.87±1.33	**97.13±0.71**
36	52.67±3.05	69.13±3.22	60.40±3.00	89.60±7.52	95.60±2.23	**96.80±1.60**
Avg.	63.34±15.50	80.45±12.51	82.41±14.59	91.24±11.16	92.99±9.45	**95.20±7.85**
Training time (s)	28.9704	70.9475	80.3900	68.2972	78.5600	78.2453
Testing time (s)	0.0090	0.0043	0.0088	0.0081	0.0081	0.0084

The highest accuracies of each row are indicated in bold

In task 36, the features of IWC, IWCM and DDATN still show superior performance compared with the features of other methods, which still have good fusion of source and target domains. However, there has been a significant degeneration of the interclass separability of IWC and IWCM, whereas DDATN still hold excellent interclass separability. For DDC, FTNN and DTN, the interclass separability and intraclass aggregation are both

Liao *et al. Chin. J. Mech. Eng.*   (2021) 34:52

Page 9 of 10

Figure 5 Comparison result

Figure 6 Feature visualization of task 33 (CW0.53_3 to OL500)

Figure 7 Feature visualization of task 36 (CW0.53_3 to OL1400)

poor, whereas the fusion of source and target domains are hard to be observed.

The feature visualization demonstrates that DDATN can effectively adapt the target domain features distributions to that of source domain. The target samples can be accurately aggregated to the corresponding source cluster, and the extracted features shows good fusion of domains and excellent discriminability of bearing health conditions.

5 Conclusions

This article proposes a novel unsupervised domain adaptation method termed DDATN. It introduces DDA for cross domain bearing intelligent fault diagnosis. The DDA is realized by the proposed IDMMD, which combines novel dynamic factor estimation method and instance-weighted conditional distribution discrepancy metric. The cross domain bearing fault diagnosis experiment is conducted to verify the effectiveness of DDATN. DDATN achieved better performance than other state-of-the-art cross domain fault diagnosis methods. The results demonstrate that the proposed conditional distribution discrepancy metric and the dynamic factor calculation method are effective and robust for DDA. Therefore, DDATN can effectively adapt the target domain features distributions to that of source domain for better cross domain bearing fault diagnosis.

Authors' Contributions
WL was in charge of the whole trial; YL wrote the manuscript; RH, JL and ZC assisted with sampling and laboratory analyses. All authors read and approved the final manuscript.

Authors' Information
Yixiao Liao, born in 1992, is currently a PhD candidate at *School of Mechanical & Automotive Engineering, South China University of Technology, China*, where he received his bachelor's degree in 2015. His research interests include deep learning, transfer learning, and adversarial learning for fault diagnosis and prognostics.

Ruyi Huang, born in 1992, is currently a PhD candidate at *School of Mechanical & Automotive Engineering, South China University of Technology, China*. He received his B.S degree on mechanical engineering in *Qingdao University, China*, in 2014. His current research interests include deep learning and transfer learning methods for intelligent fault diagnostics and prognostics of rotating machinery.

Jipu Li, born in 1994, is currently a PhD candidate at *School of Mechanical & Automotive Engineering, South China University of Technology, China*. He received his B.S. and M.S. degree on mechanical engineering from *Lanzhou University of Technology, China*, in 2015 and 2018, respectively. His current research interests include deep learning and deep transfer learning methods for fault diagnosis and prognostics.

Zhuyun Chen, born in 1991, is currently a post doctor at *School of Mechanical & Automotive Engineering, South China University of Technology, China*. He received his B.S degree on mechanical design, manufacturing, and

Liao *et al. Chin. J. Mech. Eng.*   (2021) 34:52

Page 10 of 10

automation from *Nanjing Agricultural University, China*, in 2013. His current research interests include dynamic signal processing and deep learning methods for mechanical fault diagnosis and prognostics. His current research interests include deep learning and deep transfer learning methods for fault diagnosis and prognostics.

Weihua Li, born in 1973, is currently a professor at *School of Mechanical & Automotive Engineering, South China University of Technology, China*. His main research interests include nonlinear time series analysis, dynamic signal processing, autonomous driving, and machine learning methods for condition monitoring and health diagnosis of complex dynamical systems.

Competing Interests
The authors declare no competing financial interests.

References
[1] R N Liu, B Y Yang, E Zio, et al. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, 2018, 108: 33-47.
[2] Z Z Pan, Z Meng, Z J Chen, et al. A two-stage method based on extreme learning machine for predicting the remaining useful life of rolling-element bearings. *Mechanical Systems and Signal Processing*, 2020, 144: 106899.
[3] L L Cui, X Wang, H Q Wang, et al. Remaining useful life prediction of rolling element bearings based on simulated performance degradation dictionary. *Mechanism and Machine Theory*, 2020, 153: 103967.
[4] D Z Zhao, J Y Li, W D Cheng, et al. Generalized demodulation transform for bearing fault diagnosis under nonstationary conditions and gear noise interferences. *Chinese Journal of Mechanical Engineering*, 2019, 32: 7.
[5] Y T Hu, S Q Zhang, A Q Jiang, et al. A new method of wind turbine bearing fault diagnosis based on multi-masking empirical mode decomposition and fuzzy c-means clustering. *Chinese Journal of Mechanical Engineering*, 2019, 32: 46.
[6] G Q Jiang, P Xie, X Wang, et al. Intelligent fault diagnosis of rotary machinery based on unsupervised multiscale representation learning. *Chinese Journal of Mechanical Engineering*, 2017, 30(6): 1314-1324.
[7] F B Zhang, J F Huang, F L Chu, et al. Mechanism and method for the full-scale quantitative diagnosis of ball bearings with an inner race fault. *Journal of Sound and Vibration*, 2020, 488: 115641.
[8] J P Li, R Y Huang, G L He, et al. A deep adversarial transfer learning network for machinery emerging fault detection. *IEEE Sensors Journal*, 2020, 20(15): 8413-8422.
[9] S Y Shao, W J Sun , R Q Yan, et al. A deep learning approach for fault diagnosis of induction motors in manufacturing. *Chinese Journal of Mechanical Engineering*, 2017, 30(6): 1347-1356.
[10] H Q Wang, S Li, L Y Song, et al. An enhanced intelligent diagnosis method based on multi-sensor image fusion via improved deep learning network. *IEEE Transactions on Instrumentation and Measurement*, 2019, 69(6): 2648-2657.
[11] R Y Huang, J P Li, S H Wang, et al. A robust weight-shared capsule network for intelligent machinery fault diagnosis. *IEEE Transactions on Industrial Informatics*, 2020, 16(10): 6466-6475.
[12] Z Y Chen, K Gryllias, W H Li. Mechanical fault diagnosis using convolutional neural networks and extreme learning machine. *Mechanical Systems and Signal Processing*, 2019, 133: 106272.
[13] J P Li, R Y Huang, G L He, et al. A two-stage transfer adversarial network for intelligent fault diagnosis of rotating machinery with multiple new faults. *IEEE/ASME Transactions on Mechatronics*, 2020, doi: https://doi.org/10.1109/TMECH.2020.3025615.
[14] S J Pan, Q Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 2009, 22(10): 1345-1359.
[15] R Zhao, R Q Yan, Z H Chen, et al. Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, 2019, 115: 213-237.
[16] R Q Yan, F Shen, C Sun, et al. Knowledge transfer for rotary machine fault diagnosis. *IEEE Sensors Journal*, 2019, 20(15): 8374-8393.
[17] T Han, C Liu, W G Yang, et al. A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults. *Knowledge-Based Systems*, 2019, 165: 474-487.
[18] L Guo, Y G Lei, S B Xing, et al. Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE Transactions on Industrial Electronics*, 2018, 66(9): 7316-7325.
[19] B Yang, Y G Lei, F Jia, et al. An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. *Mechanical Systems and Signal Processing*, 2019, 122: 692-706.
[20] X Wang, C Q Shen, M Xia, et al. Multi-scale deep intra-class transfer learning for bearing fault diagnosis. *Reliability Engineering & System Safety*, 2020, 202: 107050.
[21] X Li, X D Jia, W Zhang, et al. Intelligent cross-machine fault diagnosis approach with deep auto-encoder and domain adaptation. *Neurocomputing*, 2020, 383: 235-247.
[22] J D Wang, Y Q Chen, W J Feng, et al. Transfer learning with dynamic distribution adaptation. *ACM Transactions on Intelligent Systems and Technology*, 2020, 11(1): 1-25.
[23] T Han, C Liu, W G Yang, et al. Deep transfer network with joint distribution adaptation: A new intelligent fault diagnosis framework for industry application. *ISA Transactions*, 2020, 97: 269-281.
[24] M Li, Z H Sun, W H He, et al. Rolling bearing fault diagnosis under variable working conditions based on joint distribution adaptation and SVM. *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, UK, Jul. 19-24, 2020: 1-8.
[25] A Gretton, K M Borgwardt, M J Rasch, et al. A kernel two-sample test. *The Journal of Machine Learning Research*, 2012, 13(1): 723-773.
[26] X S Lou, K A Loparo. Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mechanical Systems and Signal Processing*, 2004, 18(5): 1077-1095.
[27] E Tzeng, J Hoffman, N Zhang, et al. Deep domain confusion: Maximizing for domain invariance. arXiv preprint, http://arxiv.org/abs/1412.3474; 2014.
[28] L Maaten, G E Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 2008, 9: 2579-2605.