

## Article

# A Weighted Subdomain Adaptation Network for Partial Transfer Fault Diagnosis of Rotating Machinery

Sixiang Jia, Jinrui Wang, Xiao Zhang and Baokun Han \*

College of Mechanical and Electronic Engineering, Shandong University of Science and Technology, Qingdao 266590, China; sixiang\_j@163.com (S.J.); wangjr33@163.com (J.W.); zhangxiao9789@163.com (X.Z.)

\* Correspondence: bk\_han@163.com; Tel.: +86-139-5325-2858

**Abstract:** Domain adaptation-based models for fault classification under variable working conditions have become a research focus in recent years. Previous domain adaptation approaches generally assume identical label spaces in the source and target domains, however, such an assumption may be no longer legitimate in a more realistic situation that requires adaptation from a larger and more diverse source domain to a smaller target domain with less number of fault classes. To address the above deficiencies, we propose a partial transfer fault diagnosis model based on a weighted subdomain adaptation network (WSAN) in this paper. Our method pays more attention to the local data distribution while aligning the global distribution. An auxiliary classifier is introduced to obtain the class-level weights of the source samples, so the network can avoid negative transfer caused by unique fault classes in the source domain. Furthermore, a weighted local maximum mean discrepancy (WLMMD) is proposed to capture the fine-grained transferable information and obtain sample-level weights. Finally, relevant distributions of domain-specific layer activations across different domains are aligned. Experimental results show that our method could assign appropriate weights to each source sample and realize efficient partial transfer fault diagnosis.



**Citation:** Jia, S.; Wang, J.; Zhang, X.; Han, B. A Weighted Subdomain Adaptation Network for Partial Transfer Fault Diagnosis of Rotating Machinery. *Entropy* **2021**, *23*, 424. <https://doi.org/10.3390/e23040424>

Academic Editor: Yongbo Li

Received: 12 March 2021

Accepted: 31 March 2021

Published: 1 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** domain adaptation; partial transfer; fault diagnosis; subdomain; rotating machinery

## 1. Introduction

As indispensable parts of rotating machinery, the fault identification and diagnosis of bearings and gears are crucial for the normal operation of the machinery. Since traditional fault diagnosis methods rely on manual processing of vibration signals, it is difficult to explore the depth of fault diagnosis knowledge. With the widely application in industry and academia of deep learning technology, it is possible to mine effective diagnosis knowledge from massive amounts of fault data [1–3]. Therefore, such methods have been extensively applied in fault diagnosis of rotating machinery [4–7].

Li et al. [8] proposed a fault diagnosis framework based on multi-scale permutation entropy (MPE) and multi-channel fusion convolutional neural networks (MCFCNN). Since it considers the structure and spatial information between different sensor measurement points, the fault diagnosis with high accuracy and speed is realized. Valtierra-Rodriguez et al. [9] proposed a methodology based on convolutional neural networks for automatic detection of broken rotor bars by considering different severity levels. This method applies a notch filter to remove the fundamental frequency component of the current signal, and the short-time Fourier transform (STFT) is used to obtain time-frequency plane. Experimental results show that the methods is capable of identifying the healthy condition of the induction motor. However, the distributions of the collected datasets may different due to the change of the operating environments. The diagnostic knowledge in the original training data will no longer be fully applicable to the new testing data when the working condition changes [10–14]. In this case, the fault diagnosis methods under variable working conditions based on transfer learning come into being. Recently, some transfer learning-based methods have been developed to solve cross domain fault diagnosis problem. Mao et al. [15] proposed a deep dual temporal

domain adaptation (DTDA) model which could recognize whether an early fault occurs and achieve an earlier detection location and lower false alarm rate. An et al. [16] proposed to apply the maximum mean discrepancy (MMD) based on multiple kernels to intelligent fault diagnosis, and the features of different layers were involved in the domain adaptation process. Wang et al. [17] presented a deep adaptive adversarial network (DAAN) which could narrow the discrepancy to learn domain-invariant features. Chen et al. [18] proposed an unsupervised domain adaptation method which could maximize the mutual information between the target feature space and the entire feature space and minimize the feature-level discrepancy between the two domains. Hasan et al. [19] proposed a multitask-aided transfer learning-based diagnostic framework. This method applies multitask learning-based convolutional network to identify working conditions, and then identifies health status of the rolling element bearings based on transfer learning. In a word, transfer learning techniques provide an efficient solution to cross domain fault diagnosis problems.

Although transfer learning-based methods have made great progress, partial transfer fault diagnosis problem has not been well solved. The partial transfer diagnosis means that the number of fault types in the test data is less than that in the training data. Since the machine is in a healthy working state most of the time, the test data may contain only a few types of fault data. That is, the distribution of two domains is different and the label space of target domain is a subset of that of the source domain [20–22]. As many different health types as possible can be involved by training data through a long period of data accumulation, while it is difficult to guarantee the symmetry of health types in testing data and training data. Therefore, this setting is closer to engineering practice compared with the scenario for which the standard domain adaptation is targeted. Since most of the transfer fault diagnosis methods use all source samples for domain adaptation, the unique types of source samples can enable the network to learn false classification knowledge during domain adaptation, which is the major challenge in partial transfer fault diagnosis. Actually, partial transfer problem has been studied in the field of target detection and computer vision. Cao et al. [23] proposed a selective adversarial network (SAN) to facilitate positive transfer by selecting the source samples highly correlated with the target samples. Chen et al. [24] proposed reinforced transfer network (RTNet) which could apply both high-level and pixel-level information to solve partial transfer problem. In addition, importance weighted adversarial nets [25] and example transfer network (ETN) [26] also obtained excellent performance in the image classification task. These works have laid a solid foundation for solving the problem of partial transfer in mechanical fault diagnosis.

Recently, the partial transfer problem has made initial progress in fault diagnosis. Jiao et al. [27] applied weighted cross entropy loss to give smaller weight to the unique source samples, and such weight is determined by the predicted outputs of two classifiers [28]. Li et al. [29] presented a weighted adversarial transfer network (WATN) which used adversarial training to reweight the source domain samples. Yang et al. [30] proposed a deep partial transfer learning network (DPTL-Net) which could learn domain-asymmetry factor to weight the source samples and finally block unnecessary knowledge. The previous partial domain adaptation methods mainly tried to get the weight of the source samples from a global perspective without considering the relationships between two subdomains [31] in source and target domains, which is not conducive to obtaining the fine-grained transferable information in each type of data. To solve the above problem, this paper proposed a weighted subdomain adaptation network (WSAN) to improve the efficiency of partial transfer diagnosis of machinery. All the samples are divided into class-level subdomains, and the subdomain distributions of deep features in multiple layers are aligned. In order to block the samples of outlier source types, an auxiliary classifier is introduced to conduct adversarial training with the feature generator to obtain the class-level weights. To achieve weighted subdomain adaptation, we propose a weighted local maximum mean discrepancy (WLMMD) to measure the Hilbert-Schmidt norm between kernel mean embedding of empirical distributions between relevant subdomains. The main innovations of this work are summarized as follows:

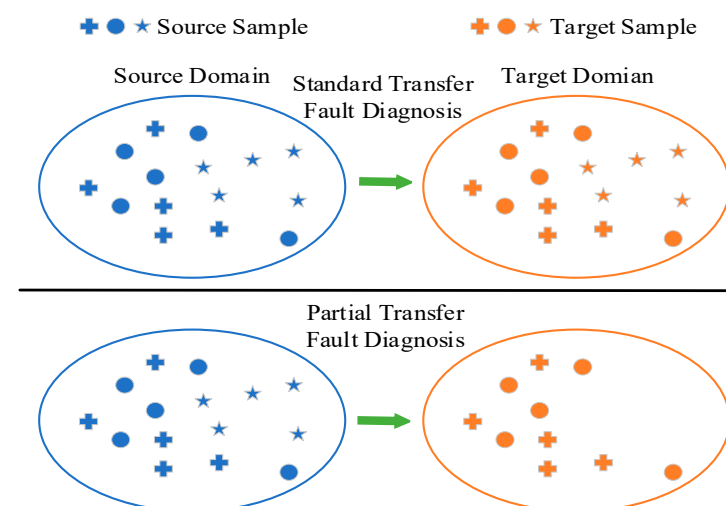
- (1) A WSAN framework is presented to solve the partial transfer fault diagnosis problem. Relevant subdomains are built to capture fine-grained transferable information and avoid negative transfer caused by redundant source samples.
- (2) The class-level weights are obtained through the adversarial training between the auxiliary classifier and the feature generator. WLMMD is designed to measure the distribution discrepancy between relevant subdomains and obtain fine-grained transferable information. As a result, proper alignment of relevant subdomains in specific activation layers is realized.

The remainder of this work begins with the background of theory in Section 2. In addition, Section 3 provides an introduction to the methodology presented, and Section 4 applies the proposed model to partial transfer fault diagnosis and verifies the advantages of the model by comparing other methods. Finally, some conclusions are drawn in Section 5.

## 2. Theoretical Background

### 2.1. Partial Transfer Fault Diagnosis

For standard domain adaptation-based frameworks, target domain  $D_t$  and source domain  $D_s$  are collected under different but related working conditions [26]. As shown in the upper part of Figure 1, the job of standard transfer fault diagnosis is to facilitate a knowledge transfer from the labeled source data  $\{X_s, C_s\}$  to the unlabeled target dataset  $X_t$ . However, different from the closed transfer fault diagnosis, the source label space  $C_s$  and target label space  $C_t$  are different in partial transfer diagnosis problem. In the bottom part of Figure 1, there are more source classes than target classes, i.e.,  $C_t \subseteq C_s$ . In addition, it should be noted that the sample types in the target domain do not deviate from the scope of the source domain, which ensures the authority of the diagnostic knowledge in source domain. The purpose of partial transfer fault diagnosis is to find the categories associated with the source domain and classify them accurately.

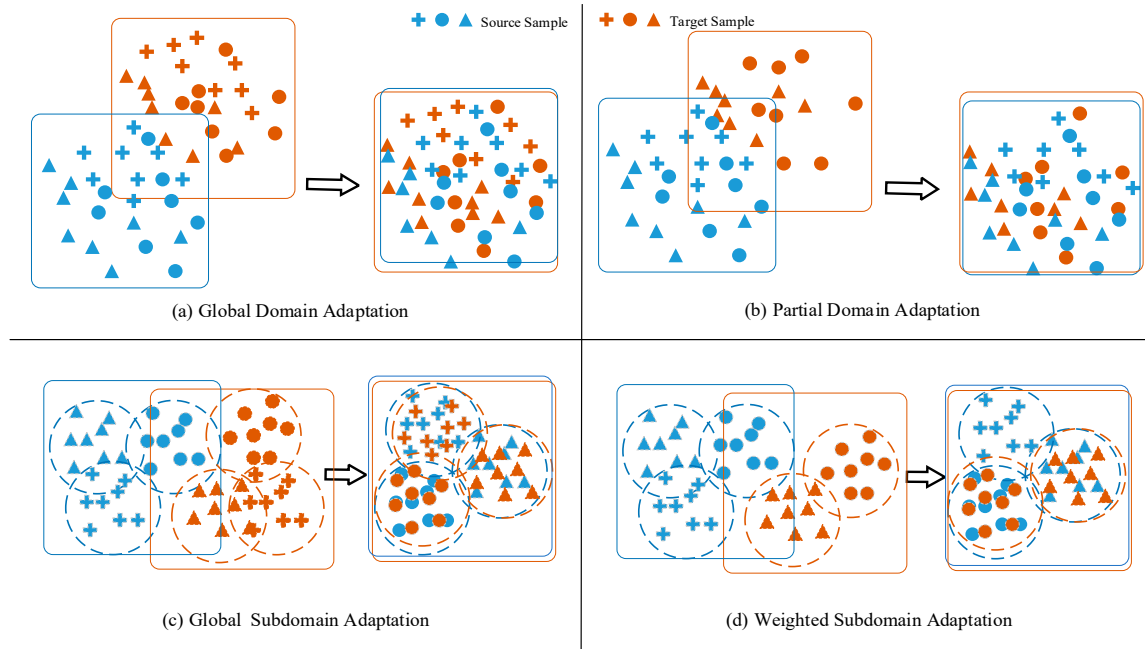


**Figure 1.** Comparison of standard transfer fault diagnosis and partial transfer fault diagnosis.

### 2.2. Subdomain Adaptation

The source and target domains may consist of some subdomains that can be defined according to different criteria, such as class or category. For partial transfer fault diagnosis, the number of sample types in the source domain must be no less than that in the target domain, so is practicable to delimit the subdomains based on the number of types in the source domain, although this may not be appropriate for the target domain, but it ensures alignment of local data distribution discrepancy. As can be seen from Figure 2a,b, it is difficult to match two data distributions directly in the process of global or partial domain adaptation. In Figure 2c,d, subdomain adaptation is of superior feature representation ability because the fine-grained transferable information within the subdomains is uti-

lized [31]. However, the problem with this is that the data in the target domain is unlabeled, which prevents target domain from being partitioned. Fortunately, we take the prediction probability output of the model for the target samples as pseudo-labels to divide them into some subdomains. In this way, subdomain adaptation enables the model to focus more on local data distribution differences.



**Figure 2.** Comparison of standard domain adaptation (a,b) and subdomain adaptation (c,d). The box represents the data distribution range and the arrows represent the domain or subdomain adaptation process. The dotted circles represent the divided subdomains.

### 2.3. Weighted Local Maximum Mean Discrepancy

In the field of transfer learning, MMD [32] is a common nonparametric metric that measures the discrepancy between two distributions. It takes the mean embeddings of two distributions in a Reproducing Kernel Hilbert Space (RKHS) as a distance calculation to avoid the density estimation. MMD can be defined as:

$$d_{\mathcal{H}}(\mathcal{D}_s, \mathcal{D}_t) \triangleq \|\mathbf{E}_p[\phi(\mathbf{x}^s)] - \mathbf{E}_q[\phi(\mathbf{x}^t)]\|_{\mathcal{H}}^2, \quad (1)$$

where  $\phi(\cdot)$  is the feature mapping function that maps the original data to RKHS  $\mathcal{H}$ . Therefore, an estimate of the MMD compares the square distance between the empirical kernel mean embeddings as:

$$\begin{aligned} \hat{d}_{\mathcal{H}}(\mathcal{D}_s, \mathcal{D}_t) &= \left\| \frac{1}{n_s} \sum_{\mathbf{x}_i \in \mathcal{D}_s} \phi(\mathbf{x}_i) - \frac{1}{n_t} \sum_{\mathbf{x}_j \in \mathcal{D}_t} \phi(\mathbf{x}_j) \right\|_{\mathcal{H}}^2 \\ &= \frac{1}{n_s^2} \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} k(\mathbf{x}_i^s, \mathbf{x}_j^s) + \frac{1}{n_t^2} \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} k(\mathbf{x}_i^t, \mathbf{x}_j^t) \\ &\quad - \frac{2}{n_s n_t} \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} k(\mathbf{x}_i^s, \mathbf{x}_j^t) \end{aligned} \quad (2)$$

where  $\hat{d}_H(p, q)$  is an unbiased estimator of  $d_H(p, q)$ .  $n_s$  and  $n_t$  are the number of source samples and target samples, respectively.

Most previous domain adaptation methods apply MMD to narrow the distribution discrepancy without considering the internal distribution of the data. However, such methods may result in poor alignment because the relationship between related subdomains is ignored. Furthermore, these methods also fail to selectively involve source samples in the adaptation process due to the asymmetry of data types across the two domains. Considering the above problems, we propose the WLMMD to achieve weighted subdomain adaptation:

$$d_{\mathcal{H}}(\mathcal{D}_s, \mathcal{D}_t) \triangleq \mathbf{E}_c \left\| \mathbf{E}_{p^{(c)}}[\phi(\mathbf{x}^s)] - \mathbf{E}_{q^{(c)}}[\phi(\mathbf{x}^t)] \right\|_{\mathcal{H}}^2, \quad (3)$$

where  $\mathbf{x}^s$  and  $\mathbf{x}^t$  are the instances in  $D_s$  and  $D_t$ , and  $p^{(c)}$ , and  $q^{(c)}$  are the distributions of  $D_s^{(c)}$  and  $D_t^{(c)}$ , respectively. So we can calculate an unbiased estimator of WLMMD as:

$$\hat{d}_{\mathcal{H}}(\mathcal{D}_s, \mathcal{D}_t) = \frac{1}{C} \sum_{k=1}^C \left\| \sum_{\mathbf{x}_i^s \in \mathcal{D}_s} w_i^{sk} \phi(\mathbf{x}_i^s) - \sum_{\mathbf{x}_j^t \in \mathcal{D}_t} w_j^{tk} \phi(\mathbf{x}_j^t) \right\|_{\mathcal{H}}^2, \quad (4)$$

where  $w_i^{sk}$  and  $w_j^{tk}$  denote the weights of  $\mathbf{x}_i^s$  and  $\mathbf{x}_j^t$  belonging to class  $k$ , respectively. Obviously,  $\sum_{i=1}^{n_s} w_i^{sk} = \sum_{i=1}^{n_t} w_i^{tk} = 1$ , and  $w_i^k$  for the sample  $x_i$  can be computed as:

$$w_i^k = \frac{y_{ik}}{\sum_{(x_j, y_j) \in \mathcal{D}} y_{jk}}, \quad (5)$$

where  $y_{ic}$  is the  $k$ -th entry of vector  $\mathbf{y}_i$ . Since the source samples are labeled with a one-hot vector, we can directly calculate the weight  $w_i^{sk}$  by the labels. Although the samples of the target domain are unlabeled, it is feasible to use pseudo labels to partition related subdomains. Note that the predicted output  $\hat{y}_i^t$  given by the classifier can be used as pseudo target labels which measures the probability that the target sample belongs to the corresponding category.  $\hat{y}_i^t$  can be regarded as the probability of assigning  $\mathbf{x}_i^t$  to each of the  $C$  classes, and the weight  $w_i^{tk}$  of target samples could be acquired. Thus, we can approximate Equation (5) as:

$$\begin{aligned} \hat{d}_l(\mathcal{D}_s, \mathcal{D}_t) = \frac{1}{C} \sum_{k=1}^C & \left[ \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} w_i^{sk} w_j^{sk} k(\mathbf{z}_i^{sl}, \mathbf{z}_j^{sl}) \right. \\ & + \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} w_i^{tk} w_j^{tk} k(\mathbf{z}_i^{tl}, \mathbf{z}_j^{tl}) \\ & \left. - 2 \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} w_i^{sk} w_j^{tk} k(\mathbf{z}_i^{sl}, \mathbf{z}_j^{tl}) \right] \end{aligned} \quad (6)$$

where  $\mathbf{z}^l$  is the  $l$ th layer activation of  $L$  layers. By using Equation (6), the distribution discrepancy between the two subdomains at a particular activation layer can be calculated.

### 3. Proposed Method

#### 3.1. Weighted Subdomain Adaptation Network

In order to achieve efficient partial transfer fault diagnosis, we design a novel weighted subdomain adaptation network (WSAN). The details of the proposed model are clearly presented in Figure 3. The feature generator  $G$  is a deep structure based on one dimensional convolutional neural network (1D-CNN) that is expected to extract domain invariant deep features. The auxiliary classifier  $C_A$  is set to obtain the class-level weights of the source samples, which is achieved by adversarial training. After acquiring class-level weights,

weighted subdomain adaptation can be carried out in activation layers of the classifier C based on WLMMD. The objective function can be written as:

$$\begin{aligned}
 F_0(\theta_G, \theta_C) &= \frac{1}{n_s} \sum_i^{n_s} L_c(G(x_{si}), y_{si}) \\
 &\quad - \frac{\lambda_0}{n_s + n_t} \sum_{x_i \in (\mathcal{D}_s \cup \mathcal{D}_t)} L_d(D(G(x_i)), d_i) \\
 &\quad + \gamma_0 \hat{d}_l(\mathcal{D}_s, \mathcal{D}_t)
 \end{aligned} \tag{7}$$

where  $\lambda_0$  and  $\gamma_0$  are the penalty coefficients,  $y_{si}$  and  $d_i$  are the source sample label and domain label, and  $L$  denotes the condition prediction loss.

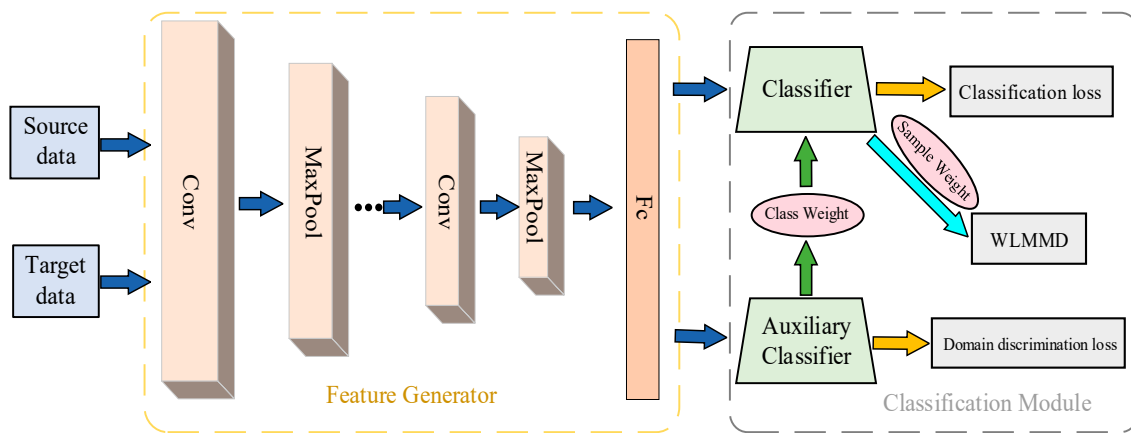


Figure 3. The structural composition of the proposed model.

### 3.2. Adversarial Training-Based Class-Level Weights Obtaining

Due to the asymmetry of the fault classes in the two domains, samples of redundant types in the source domain may cause a negative transfer. Therefore, these redundant subdomains must be selected to block the classification knowledge that is unfavorable to the recognition of target samples. Inspired by generative adversarial networks (GAN), we set up an auxiliary classifier  $C_A$  to play the mini-max game with the feature generator. Specifically, given input  $x_s$  or  $x_t$  with the label 1 or 0, after multiple layers of extraction, the feature generator  $G$  narrows the domain shift to make classifier cannot distinguish the true source of the input sample. The auxiliary classifier is trained to give the correct label. The objective of the adversarial training can be defined as:

$$\begin{aligned}
 \minmax_{G \ C_A} \mathcal{L}(C_A, G) &= \frac{1}{n_s} \sum_{i=1}^{n_s} \log(C_A(G(x_{si}))) \\
 &\quad + \frac{1}{n_t} \sum_{j=1}^{n_t} \log(1 - C_A(G(x_{tj})))
 \end{aligned} \tag{8}$$

The distribution differences of the deep features of shared fault types will be narrowed in the training process, so the auxiliary classifier will be unable to distinguish samples of these types and give an output close to 0, while the output of the unique source samples will be close to 1. The aim of adversarial training is to learn the relative importance of source samples, suggesting that the outlier samples should be assigned a relatively small weight. Therefore, the weight function is inversely related to  $C_A(G(x))$  and the importance weights function can be defined as:

$$w_c(x) = 1 - C_A(G(x)), \tag{9}$$



After obtaining the class weights of the source samples, the overall objective can be rewritten as:

$$\begin{aligned}
 F(\theta_G, \theta_C) &= \frac{1}{n_s} \sum_i^{n_s} \sum_{\mathbf{x}_{si} \in \mathcal{D}_s^j} w_{cj} L_c(G(\mathbf{x}_{si}), \mathbf{y}_{si}) \\
 &+ \frac{\lambda_1}{n_s} \sum_{i=1}^{n_s} \sum_{\mathbf{x}_{si} \in \mathcal{D}_s^j} w_{cj} \log(C_A(G(\mathbf{x}_{si}))) \\
 &+ \frac{\lambda_2}{n_t} \sum_{j=1}^{n_t} \log(1 - C_A(G(\mathbf{x}_{tj}))) \\
 &+ \gamma \hat{d}_l(\mathcal{D}_s, \mathcal{D}_t)
 \end{aligned} \tag{10}$$

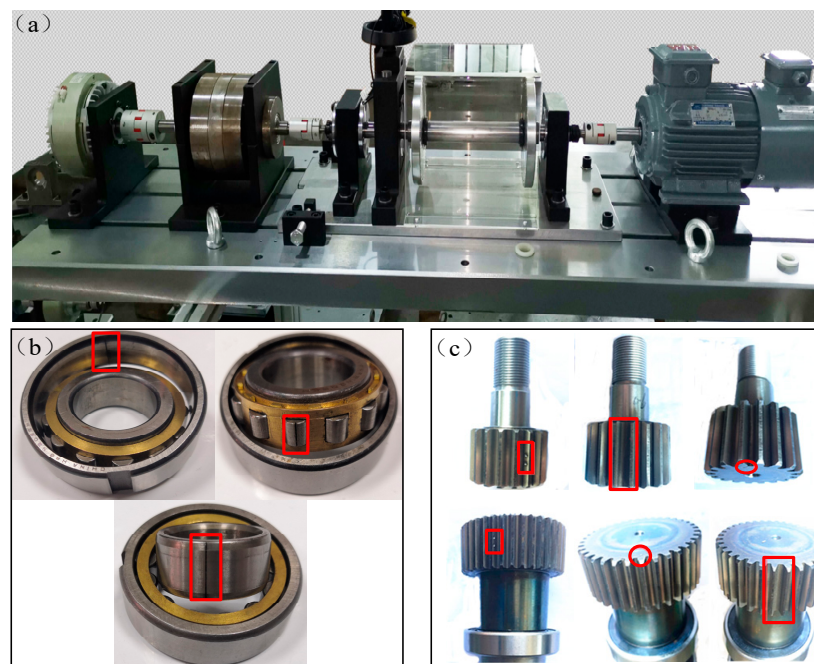
$$\begin{aligned}
 F(\theta_{C_A}) &= -\frac{1}{n_s} \sum_{i=1}^{n_s} \sum_{\mathbf{x}_{si} \in \mathcal{D}_s^j} w_{cj} \log(C_A(G(\mathbf{x}_{si}))) \\
 &- \frac{\lambda}{n_t} \sum_{j=1}^{n_t} \log(1 - C_A(G(\mathbf{x}_{tj})))
 \end{aligned} \tag{11}$$

where  $w_{cj}$  and  $\mathcal{D}_s^j$  denote the weights and samples for the  $j$ -th source class,  $\mathbf{y}_s$  is the source sample labels and  $\gamma$  is a penalty coefficient.

## 4. Experiments

### 4.1. Dataset Introduction

The proposed framework is verified with the datasets collected in our laboratory to validate the performance in partial transfer fault diagnosis. Figure 4a indicates the experimental equipment used in our laboratory. The platform consists of a motor, two balancing rotors, two bearing seats, a planetary gearbox, and a magnetic brake for controlling load. Vibration sensors are installed on fixed holders at both ends of the gearbox, and the sampling frequency is 25.6 kHz.



**Figure 4.** Representation for rotating machinery fault diagnosis test bed in our laboratory (a) and different types of bearing fault (b) and gear fault (c). Red box indicates the location of damage.

#### (1) Bearing fault dataset

Five health conditions are involved in the bearing fault dataset, namely, normal, inner ring fault, outer ring fault, rolling element fault, and combined fault of the rolling element and outer ring. The fault parts are shown in Figure 4b. There are two damage sizes for each type of fault, specifically, 0.2 and 0.4 mm. Thus, the bearing fault dataset contains samples of nine health types, namely, NC, IF1, IF2, OF1, OF2, RF1, RF2, RO1, and RO2. Four datasets are obtained under different loads, specifically, L1 (80 N), L2 (60 N), L3 (40 N), and L4 (20 N). And the engine speed is 2000 r/min.

#### (2) Gear fault dataset

As shown in Figure 4c, the gear fault dataset contains samples of seven health types, namely, normal condition (NC), sun gear fracture (SF), sun gear pitting (SP), sun gear wear (SW), planet gear fracture (PF), planet gear pitting (PP), and planet gear wear (PW). We collected three datasets at different rotational speeds (without load), specifically, S1 (2200 r/min), S2 (2000 r/min), and S3 (1800 r/min).

The number of each type of samples is 500. Thus, the number of samples in bearing and gear datasets are 4500 and 3500, respectively. In order to give full play to the feature extraction and weight learning ability of the proposed method, 40% of the samples were used for training and the remaining for testing.

### 4.2. Compared Methods

To show the superior performance of the proposed model, four comparative methods are adopted as follows:

- (1) Supervised training without classification knowledge transfer is adopted as a basic comparative method (Basic), and it obtained classification knowledge only from the source domain samples.
- (2) Domain adaptation framework based on multiple kernel variant of maximum mean discrepancy (MKMMD) [16]: Efficient kernel method is adopted in different layers of the network, and excellent performance was achieved on the global domain adaptation task.
- (3) Deep subdomain adaptation network (DSAN) [31]: As a typical global domain adaptation approach, it does not include class-level weight acquisition, that is, the auxiliary classifier is not adopted in the network. In this method, local maximum mean difference (LMMD) is used for effective subdomain adaptation.
- (4) Example transfer network (ETN) [26]: It is an adversarial discriminative domain adaptation method, and the adversarial training is adopted to obtain the weights of the source samples. Similar to the proposed method, an auxiliary domain discriminator and an auxiliary classifier are adopted to obtain the sample weights in the source domain.

### 4.3. Implementation Details

As detailed in Table 1, we randomly discarded a number of fault types to design different partial transfer diagnosis tasks on the basis of the two fault diagnosis datasets. For the dataset, each sample consists of 2400 data points, then fast Fourier transformation (FFT) is applied to transform the time-domain signal to frequency-domain signal that contains 1200 Fourier coefficients. The structure of the framework are illustrated in Table 2. The learning rate is set as 0.0001, and the maximum training epoch is 1000. In order to avoid the effects of random cause, we conducted 10 experiments on each task. The running steps of the proposed model are shown in Algorithm 1. In the test process, the spectral data of the target domain can be directly input into the model for classification. The code programming of the model is implemented on the Pytorch platform.



**Algorithm 1:** Weighted Subdomain Adaptation Network (WSAN)**Model:** Feature generator  $G$ ; Auxiliary classifier  $C_A$ ; Classifier  $C$ .**Input:** Labeled source data  $\{X_s, C_s\}$  and unlabeled target data  $X_t$ .For  $i$  in epochs:**Step 1:** The feature generator  $G$  outputs the high-dimensional features of the two domains and inputs them into the feature generator  $G$  and classifier  $C$ .**Step 2:** Auxiliary classifier  $C_A$  obtains the class-level weights. The classifier gives prediction probability output on the target samples and obtains sample-level weights to guide WLMMD to perform subdomain adaptation.**Step 3:** Train the feature generator  $G$  and classifier  $C$  to obtain the optimal parameters  $\hat{\theta}_G$  and  $\hat{\theta}_C$  by minimizing  $F(\theta_G, \theta_C)$ ;**Step 4:** Train the auxiliary classifier  $C_A$  to obtain the optimal parameters  $\hat{\theta}_{C_A}$  by minimizing  $F(\theta_{C_A})$ ;**Table 1.** Descriptions of the diagnosis tasks.

Dataset	Task	Transfer	Target	Source
Bearing	B1	L1→L2	All types	All types
	B2	L1→L3	NC, IF1, IF2, OF1, OF2, RF1, RF2, RO2	
	B3	L1→L4	NC, IF1, IF2, OF2, RF1, RO2	
	B4	L2→L3	NC, IF1, OF2, RF1, RO1	
	B5	L2→L1	NC, OF1, OF2, RO1	
	B6	L3→L4	IF1, OF1, RF1	
	B7	L3→L1	IF1, IF2	
	B8	L4→L3	RO2	
Gear	G1	S1→S2	NC, SF, SP, SW, PF, PP	All types
	G2	S1→S3	NC, SF, SP, SW, PF	
	G3	S2→S3	SF, SP, SW, PF	
	G4	S2→S1	SF, SP, SW	
	G5	S3→S1	NC, SP	
	G6	S3→S2	PW	

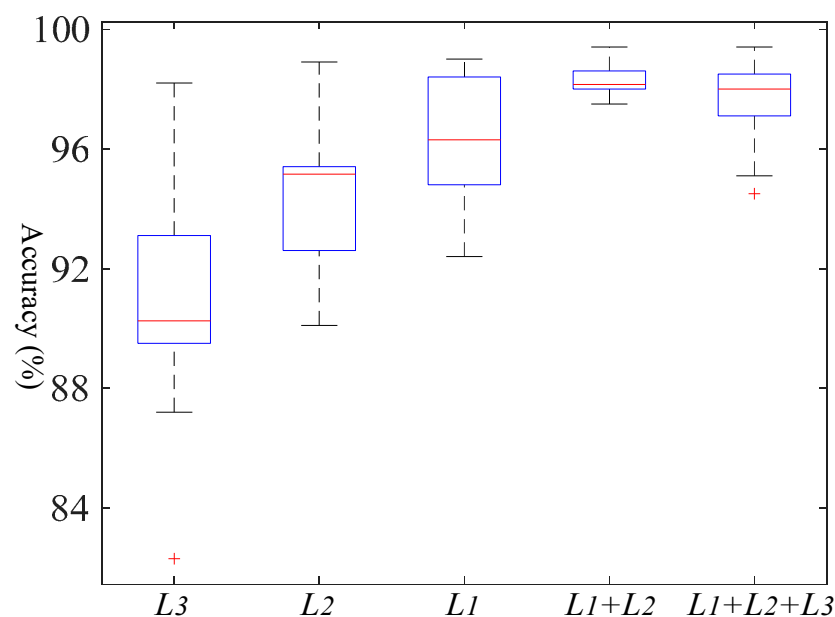
**Table 2.** Structural composition of the proposed model.

Module	Layer	Size	Channels × Kernel Size	Stride	Activation
Feature Generator	Input data	1200	/	/	/
	Conv	8 × 300	8 × 3	4	ReLU
	MaxPool	8 × 150	8 × 2	2	/
	Conv	16 × 150	16 × 3	1	ReLU
	MaxPool	16 × 75	16 × 2	2	/
	Conv	32 × 75	32 × 3	1	ReLU
	MaxPool	32 × 37	32 × 2	2	/
	Conv	32 × 37	32 × 2	1	ReLU
	MaxPool	32 × 17	32 × 2	2	/
	Fc	512	/	/	/
Classifier	Fc	256	/	/	/
	Fc	128	/	/	/
	Fc	C	/	/	Softmax
Auxiliary Classifier	Fc	256	/	/	/
	Fc	128	/	/	/
	Fc	1	/	/	Sigmoid

#### 4.4. Experimental Results

As mentioned in Section 2, the deep features in different activation layers of the model are involved in subdomain adaptation. In order to obtain the best performance for subdomain adaptation, the deep features with dimensions 128, 256 and 512 in fully

connected layers of the classifier and feature generator (named L1, L2 and L3, respectively) are extracted and combined for comparison. B7 task was selected to verify the combination of different layers, and the experiment was conducted for 15 times. As shown in Figure 5, it is clear that L1 achieves the best performance during the single layer, while L3 performs the worst. It means that the model needs to carry out deep operation to extract more separable domain invariant features. In the multi-layer combination, L1 + L2 performed better than single layer while L1 + L2 + L3 has a lower performance than L1 + L2. This indicates that some non-invariant features may exist in the shallow layers and using subdomain adaptation to align these features will degrade the performance of the model. Therefore, we apply the combination of L1 + L2 for the designed tasks.



**Figure 5.** Boxplot for the performance of different layer combinations under the same task.

The average accuracy of the proposed method and the comparison method in all tasks are detailly shown in Table 3. In general, our method obtains the highest average accuracy and the lowest standard deviation. This indicates that WSAAN has excellent and stable performance in both global and partial domain adaptation tasks. Since the basic approach does not include any domain adaptation operations, it obtains the worst performance on all tasks. MKMMD achieves the highest accuracy on the non-partial transfer fault diagnosis task B1, but performed poorly on the partial transfer tasks. This indicates that the domain adaptation methods based on MMD has superior performance in the fault diagnosis task under variable conditions, but it is not feasible to directly apply it to partial transfer scenarios. DSAN performed better than MKMMD in most tasks, and its average accuracy is 4.2% higher than that of MKMMD. But it still lags behind the other two partial transfer methods because it does not carry out any weight learning operation. WSAAN achieved an average accuracy of 97.7%, which was 4.7% higher than ETN, 13.2% higher than DSAN, and 17.4% higher than MKMMD.

It can be noted that ETN and WSAAN, as two domain adaptation methods with weighted learning, perform significantly better than other methods in partial transfer diagnosis tasks. In addition, it can be found that the proposed method gets more ahead of ETN with the increasing degree of domain class asymmetry. For task B2, the accuracy of WSAAN is 1.4% higher than that of ETN, while WSAAN is 5.6% higher than that of ETN on task B8. The same phenomenon can be observed for tasks G1 and G6.

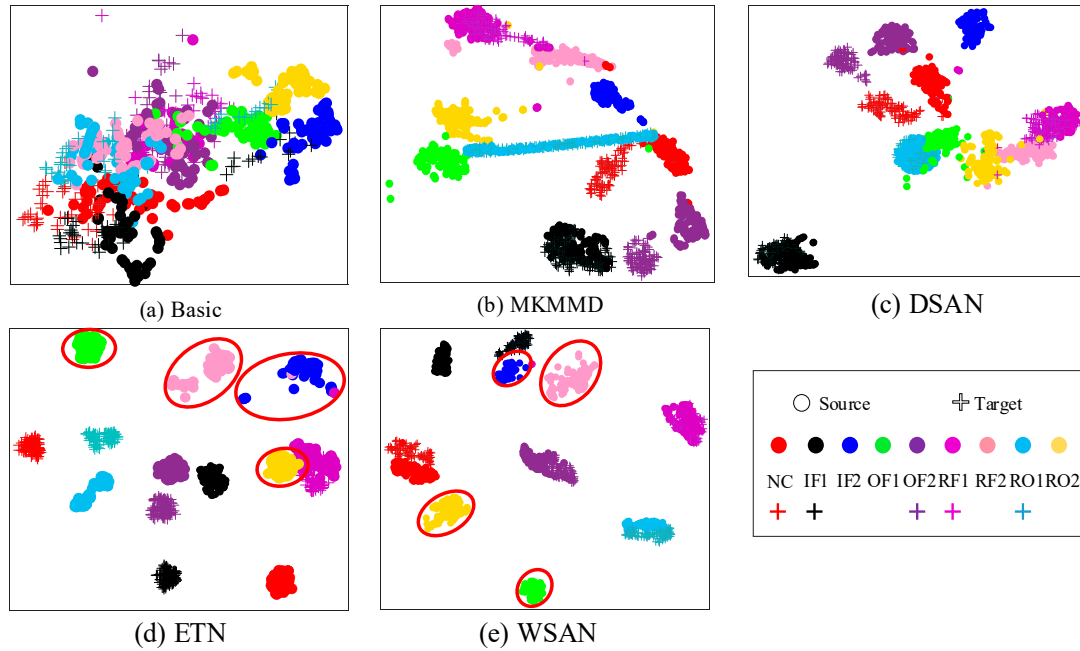
**Table 3.** Experimental results of the average testing accuracies in all tasks (%).

Task	Basic	MKMMD	DSAN	ETN	WSAN (Ours)
B1	73.1 ( $\pm 2.1$ )	99.5 ( $\pm 0.1$ )	98.3 ( $\pm 0.3$ )	98.8 ( $\pm 0.2$ )	<b>99.4 (<math>\pm 0.1</math>)</b>
B2	65.5 ( $\pm 4.5$ )	90.1 ( $\pm 0.3$ )	92.6 ( $\pm 0.8$ )	97.1 ( $\pm 0.3$ )	<b>98.5 (<math>\pm 0.1</math>)</b>
B3	63.7 ( $\pm 3.4$ )	87.2 ( $\pm 0.8$ )	90.3 ( $\pm 1.1$ )	95.5 ( $\pm 0.5$ )	<b>98.6 (<math>\pm 0.2</math>)</b>
B4	59.1 ( $\pm 5.8$ )	80.4 ( $\pm 2.1$ )	90.1 ( $\pm 1.5$ )	92.5 ( $\pm 0.6$ )	<b>98.5 (<math>\pm 0.2</math>)</b>
B5	59.7 ( $\pm 4.6$ )	82.3 ( $\pm 2.0$ )	85.4 ( $\pm 2.6$ )	93.0 ( $\pm 0.7$ )	<b>97.9 (<math>\pm 0.2</math>)</b>
B6	64.0 ( $\pm 5.8$ )	79.6 ( $\pm 3.9$ )	80.1 ( $\pm 3.9$ )	91.4 ( $\pm 1.0$ )	<b>98.0 (<math>\pm 0.1</math>)</b>
B7	45.2 ( $\pm 8.1$ )	68.2 ( $\pm 7.3$ )	75.2 ( $\pm 5.4$ )	88.5 ( $\pm 1.1$ )	<b>97.8 (<math>\pm 0.3</math>)</b>
B8	42.4 ( $\pm 6.5$ )	63.1 ( $\pm 8.0$ )	78.9 ( $\pm 7.2$ )	92.4 ( $\pm 0.6$ )	<b>98.0 (<math>\pm 0.2</math>)</b>
G1	60.2 ( $\pm 4.4$ )	88.1 ( $\pm 0.3$ )	90.6 ( $\pm 0.5$ )	98.4 ( $\pm 0.6$ )	<b>99.2 (<math>\pm 0.1</math>)</b>
G2	55.1 ( $\pm 5.0$ )	85.2 ( $\pm 0.8$ )	90.0 ( $\pm 1.1$ )	92.5 ( $\pm 0.9$ )	<b>97.1 (<math>\pm 0.1</math>)</b>
G3	50.7 ( $\pm 5.3$ )	80.4 ( $\pm 2.1$ )	84.1 ( $\pm 1.5$ )	94.3 ( $\pm 1.7$ )	<b>98.2 (<math>\pm 0.2</math>)</b>
G4	52.0 ( $\pm 3.8$ )	82.3 ( $\pm 2.0$ )	85.4 ( $\pm 2.6$ )	92.4 ( $\pm 2.0$ )	<b>95.6 (<math>\pm 0.3</math>)</b>
G5	41.2 ( $\pm 4.0$ )	69.6 ( $\pm 3.9$ )	71.2 ( $\pm 3.5$ )	88.5 ( $\pm 1.8$ )	<b>95.1 (<math>\pm 0.3</math>)</b>
G6	35.4 ( $\pm 5.6$ )	68.2 ( $\pm 5.3$ )	70.2 ( $\pm 4.0$ )	88.1 ( $\pm 0.9$ )	<b>95.5 (<math>\pm 0.6</math>)</b>
Average	55.2	80.3	84.5	93.0	<b>97.7</b>

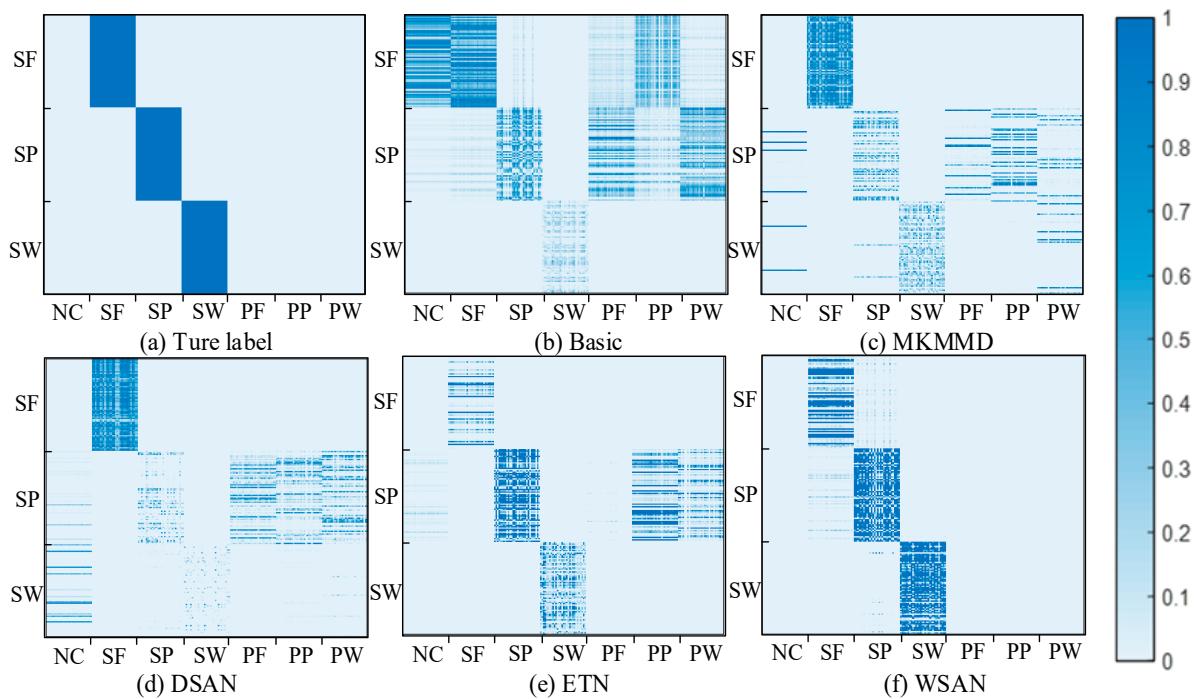
To demonstrate the feature classification effect of our method intuitively, the high-dimensional features extracted of the model are processed with the well-known  $t$ -SNE [33] technology for dimension reduction. The dimension reduction results of B3 are shown in Figure 6. In Figure 6a, we can see that the feature separability and clustering effect obtained by the basic method are inefficient. Features become separable but shared types and outliers are still cannot be distinguished in Figure 6b,c when domain adaptation is adopted. Although MKMMD and DSAN perform efficient global domain adaptation, the existence of outlier types would enable the model to extract classification knowledge that is not applicable in the target domain. This also indicates that the global adaptation methods only pays attention to the alignment of the two domains, but does not consider the relationship between the subdomains within the domain. In Figure 6d, it can be seen that ETN basically separates outlier samples but the alignment of shared type features is not accurate enough, which indicates that the classifier cannot carry out effective sample-level alignment after obtaining class-level weights and it may leads to inaccurate classification. There are some confusions between the source samples of RO2 and RF1 types. In this case, ETN may treat the RF1 samples as outliers and filter out some useful classification knowledge. For the proposed method, precise alignment of the related subdomains is performed while blocking the outlier types in Figure 6e. After obtaining accurate class-level weights, WSAN can use the proposed WLMMD to perform effective subdomain alignment which involves the sample-level weights learning.

In order to further explore how the weights learned affect the alignment of deep features, the similarity matrix of source and target features in deep layer is drawn on task G4. According to [30], the similarity matrix can be calculated by  $G(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 200)$  wherein  $x_i \in D^s$  and  $x_j \in D^t$ . Figure 7a shows the actual correspondence between the source and target labels. In Figure 7b, only the samples of SW type can be identified to a certain extent, while the features extracted from the other two target types of samples are highly similar to various source types, which is extremely unfavorable for classification. Obviously, the deep features extracted by the basic method are chaotic due to the lack of domain adaptation operation. In Figure 7c,d, the corresponding samples of SP and SW types have low similarity degree, and some of the samples have great similarity with other types. Consequently, global domain adaptation methods may extract fuzzy deep features when dealing with partial transfer problem. Figure 7e shows that ETN can assign large weight to shared types, but there are still some outlier samples with large weights, resulting in a higher similarity between target features of SP and source features of PP and PW. By comparison, Figure 7f indicates that WSAN obtains more accurate weights, which is reflected in the large similarity between the extracted features of the target domain and corresponding features of source domain, and only a few samples are weakly similar to other source types. In general, the proposed method

can make the shared samples fully participate in the subdomain adaptation and block outliers. Thus, the extracted domain invariant features own high similarity among the corresponding shared types.



**Figure 6.** *t*-SNE visualization results of (a) Basic, (b) MKMMD, (c) DSAN, (d) ETN, and (e) WSAN in task B3. The samples circled in red are outlier types.



**Figure 7.** Similarity matrix of learned features by (a) true label, (b–e) comparison methods, and (f) the proposed method. The abscissa and ordinate represent the source sample sequence and target sample sequence, respectively. The depth of the color indicates the similarity between the corresponding samples.

## 5. Conclusions

A weighted subdomain adaptation network (WSAN) is presented to solve partial transfer fault diagnosis problem of machinery. Different from the previous global domain adaptation approaches, we divide all samples into different subdomains according to sample types of the source domain, and design WLMMD to perform accurate subdomain alignment. In addition, in order to obtain class-level weights, an additional auxiliary classifier is set up to conduct adversarial training with the feature generator. Under the guidance of class-level weights, the prediction probability output of the target domain by the classifier is used as the sample-level weights, so that the model could capture fine-grained transferable information within the relevant subdomains. The optimal layer combination was found by exploring the performance of the deep features in different activation layers participating the subdomain adaptation. The best diagnostic performance can be obtained under the combination of fully connected layers (L1 + L2) with dimensions 128 and 256. Experimental results on the bearing and gear datasets collected in our laboratory indicates that the average accuracy of the proposed method on the designed fault diagnosis task is 97.7%, which is higher than that of several comparison methods. This means WSAN could solve the partial transfer fault diagnosis problem more efficiently compared several popular methods. *t*-SNE dimension reduction and correlation matrix show that WSAN can learn accurate weights and carry out accurate weighted subdomain adaptation.

Although the proposed weighted subdomain adaptation approach achieves superior performance on the partial transfer fault diagnosis tasks, the laboratory works on the premise that the target data is available during training. It is difficult to guarantee the performance of such a model under unknown working conditions. Such approaches may fail when we need real-time diagnosis. However, this problem may be solved with the help of domain generalization technology [34], and we will explore this issue in depth in our future work.

**Author Contributions:** Writing—review and editing, S.J.; visualization, J.W.; supervision, X.Z.; project administration, B.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research was supported by the National Natural Science Foundation of China (52005303), the Natural Science Foundation of Shandong Province (ZR201911100329), the Project of China Postdoctoral Science Foundation (2019M662399), and the Postdoctoral Innovation Project of Shandong Province (202003029).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Hoang, D.-T.; Kang, H.-J. A survey on deep learning based bearing fault diagnosis. *Neurocomputing* **2019**, *335*, 327–335. [[CrossRef](#)]
2. Li, Y.; Yang, Y.; Li, G.; Xu, M.; Huang, W. A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mRMR feature selection. *Mech. Syst. Signal Process.* **2017**, *91*, 295–312. [[CrossRef](#)]
3. Lei, Y.; Yang, B.; Jiang, X.; Jia, F.; Li, N.; Nandi, A.K. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mech. Syst. Signal Process.* **2020**, *138*, 106587. [[CrossRef](#)]
4. Shao, H.; Jiang, H.; Zhao, H.; Wang, F. A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. *Mech. Syst. Signal Process.* **2017**, *95*, 187–204. [[CrossRef](#)]
5. Wang, J.; Li, S.; An, Z.; Jiang, X.; Qian, W.; Ji, S. Batch-normalized deep neural networks for achieving fast intelligent fault diagnosis of machines. *Neurocomputing* **2019**, *329*, 53–65. [[CrossRef](#)]
6. Han, B.; Wang, X.; Ji, S.; Zhang, G.; Jia, S.; He, J. Data-enhanced stacked autoencoders for insufficient fault classification of machinery and its understanding via visualization. *IEEE Access* **2020**, *8*, 67790–67798. [[CrossRef](#)]
7. Li, Y.; Wang, X.; Liu, Z.; Liang, X.; Si, S. The entropy algorithm and its variants in the fault diagnosis of rotating machinery: A review. *IEEE Access* **2018**, *6*, 66723–66741. [[CrossRef](#)]



8. Li, H.; Huang, J.; Yang, X.; Luo, J.; Zhang, L.; Pang, Y. Fault diagnosis for rotating machinery using multiscale permutation entropy and convolutional neural networks. *Entropy* **2020**, *22*, 851. [[CrossRef](#)]
9. Valtierra-Rodriguez, M.; Rivera-Guillen, J.R.; Basurto-Hurtado, J.A.; De-Santiago-Perez, J.J.; Granados-Lieberman, D.; Amezcua-Sanchez, J.P. Convolutional neural network and motor current signature analysis during the transient state for detection of broken rotor bars in induction motors. *Sensors* **2020**, *20*, 3721. [[CrossRef](#)]
10. Wang, J.; Feng, W.; Chen, Y.; Yu, H.; Huang, M.; Yu, P.S. Visual domain adaptation with manifold embedded distribution alignment. In Proceedings of the 26th ACM international conference on Multimedia, Seoul, Korea, 22–26 October 2018; pp. 402–410.
11. Zhu, Y.; Zhuang, F.; Wang, J.; Chen, J.; Shi, Z.; Wu, W.; He, Q. Multi-representation adaptation network for cross-domain image classification. *Neural Netw.* **2019**, *119*, 214–221. [[CrossRef](#)]
12. Jia, S.; Wang, J.; Han, B.; Zhang, G.; Wang, X.; He, J. A Novel transfer learning method for fault diagnosis using maximum classifier discrepancy with marginal probability distribution adaptation. *IEEE Access* **2020**, *8*, 71475–71485. [[CrossRef](#)]
13. Ren, Y.; Li, Y.; Wang, X.; Wang, S.; Si, S. A Novel Bearing Fault Diagnosis Method Based on Multi-scale Transfer Sample Entropy. In Proceedings of the 11th International Conference on Prognostics and System Health Management (PHM-2020 Jinan), Jinan, China, 23–25 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 232–236.
14. Wu, Z.; Jiang, H.; Zhao, K.; Li, X. An adaptive deep transfer learning method for bearing fault diagnosis. *Measurement* **2020**, *151*, 107227. [[CrossRef](#)]
15. Mao, W.; Sun, B.; Wang, L. A new deep dual temporal domain adaptation method for online detection of bearings early fault. *Entropy* **2021**, *23*, 162. [[CrossRef](#)] [[PubMed](#)]
16. An, Z.; Li, S.; Wang, J.; Xin, Y.; Xu, K. Generalization of deep neural network for bearing fault diagnosis under different working conditions using multiple kernel method. *Neurocomputing* **2019**, *352*, 42–53. [[CrossRef](#)]
17. Wang, J.; Ji, S.; Han, B.; Bao, H.; Jiang, X. Deep adaptive adversarial network-based method for mechanical fault diagnosis under different working conditions. *Complexity* **2020**, *2020*, 6946702. [[CrossRef](#)]
18. Chen, J.; Wang, J.; Zhu, J.; Lee, T.H.; De Silva, C. Unsupervised cross-domain fault diagnosis using feature representation alignment networks for rotating machinery. *IEEE/ASME Trans. Mechatron.* **2020**. [[CrossRef](#)]
19. Hasan, M.J.; Sohaib, M.; Kim, J.-M. A multitask-aided transfer learning-based diagnostic framework for bearings under inconsistent working conditions. *Sensors* **2020**, *20*, 7205. [[CrossRef](#)]
20. Cao, Z.; Ma, L.; Long, M.; Wang, J. Partial adversarial domain adaptation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 135–150.
21. Li, S.; Liu, C.H.; Lin, Q.; Wen, Q.; Su, L.; Huang, G.; Ding, Z. Deep residual correction network for partial domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**. [[CrossRef](#)]
22. Chen, J.; Wu, X.; Duan, L.; Gao, S. Domain adversarial reinforcement learning for partial domain adaptation. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**. [[CrossRef](#)]
23. Cao, Z.; Long, M.; Wang, J.; Jordan, M.I. Partial transfer learning with selective adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2724–2732.
24. Chen, Z.; Chen, C.; Cheng, Z.; Jiang, B.; Fang, K.; Jin, X. Selective transfer with reinforced transfer network for partial domain adaptation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 12706–12714.
25. Zhang, J.; Ding, Z.; Li, W.; Ogunbona, P. Importance weighted adversarial nets for partial domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8156–8164.
26. Cao, Z.; You, K.; Long, M.; Wang, J.; Yang, Q. Learning to transfer examples for partial domain adaptation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2985–2994.
27. Jiao, J.; Zhao, M.; Lin, J.; Ding, C. Classifier inconsistency-based domain adaptation network for partial transfer intelligent diagnosis. *IEEE Trans. Ind. Inform.* **2019**, *16*, 5965–5974. [[CrossRef](#)]
28. Saito, K.; Watanabe, K.; Ushiku, Y.; Harada, T. Maximum classifier discrepancy for unsupervised domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3723–3732.
29. Li, W.; Chen, Z.; He, G. A novel weighted adversarial transfer network for partial domain fault diagnosis of machinery. *IEEE Trans. Ind. Inform.* **2020**, *17*, 1753–1762. [[CrossRef](#)]
30. Yang, B.; Lee, C.-G.; Lei, Y.; Li, N.; Lu, N. Deep partial transfer learning network: A method to selectively transfer diagnostic knowledge across related machines. *Mech. Syst. Signal Process.* **2021**, *156*, 107618. [[CrossRef](#)]
31. Zhu, Y.; Zhuang, F.; Wang, J.; Ke, G.; Chen, J.; Bian, J.; Xiong, H.; He, Q. Deep subdomain adaptation network for image classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**. [[CrossRef](#)]
32. Dziugaite, G.K.; Roy, D.M.; Ghahramani, Z. Training generative neural networks via maximum mean discrepancy optimization. *arXiv* **2015**, arXiv:1505.03906.
33. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
34. Matsuura, T.; Harada, T. Domain generalization using a mixture of multiple latent domains. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 11749–11756.