

Article

Improved Adversarial Transfer Network for Bearing Fault Diagnosis under Variable Working Conditions

Jun Wang^{1,2}, Hosameldin Ahmed², Xuefeng Chen¹, Ruqiang Yan¹ and Asoke K. Nandi^{1,2,*} 

¹ Department of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China; jun.wang@brunel.ac.uk (J.W.); chenxf@xjtu.edu.cn (X.C.); yanruqiang@xjtu.edu.cn (R.Y.)

² Department of Electronic and Electrical Engineering, Brunel University London, London UB8 3PH, UK; hosameldin.abdallahahmed@brunel.ac.uk

* Correspondence: asoke.nandi@brunel.ac.uk

Abstract: Bearings are one of the critical components of rotating machinery, and their failure can cause catastrophic consequences. In this regard, previous studies have proposed a variety of intelligent diagnosis methods. Most existing bearing fault diagnosis methods implicitly assume that the training and test sets are from the same distribution. However, in real scenarios, bearings have been working in complex and changeable working environments for a long time. The data during their working processes and the data used for model training cannot meet this condition. This paper proposes an improved adversarial transfer network for fault diagnosis under variable working conditions. Specifically, this paper combines an adversarial transfer network with a short-time Fourier transform to obtain satisfactory results with the lighter network. Then, this paper employs a channel attention module to enhance feature fusion. Moreover, this paper designs a novel domain discrepancy hybrid metric loss to improve model transfer learning performance. Finally, this paper verifies the method's effectiveness on three datasets, including dual-rotor, a Case Western Reserve University dataset and the Ottawa dataset. The proposed method achieves average accuracy, surpassing other methods, and shows better domain alignment capabilities.

Keywords: bearing intelligent diagnosis; adversarial learning; transfer learning; improved adversarial transfer network



Citation: Wang, J.; Ahmed, H.; Chen, X.; Yan, R.; Nandi, A.K. Improved Adversarial Transfer Network for Bearing Fault Diagnosis under Variable Working Conditions. *Appl. Sci.* **2024**, *14*, 2253. <https://doi.org/10.3390/app14062253>

Academic Editors: Haidong Shao, Janet Lin and Liangwei Zhang

Received: 25 January 2024

Revised: 29 February 2024

Accepted: 2 March 2024

Published: 7 March 2024



Copyright: © 2024 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/>).

1. Introduction

Bearings are critical rotating machinery components and one of these machines' weakest links [1,2]. Their performance directly affects the system's stable operation and production efficiency [3,4]. Taking an aerospace engine as an example, as shown in Figure 1, rolling bearings are easy to start at low temperatures and have small friction losses, wide operating ranges and strong resistance to oil cutoffs; engines use rolling bearings as main bearings [5,6]. The main bearings support the engine rotor in the engine and transfer rotor loads. The main bearing is divided into the thrust bearing and the traveling bearing. Thrust bearings carry axial loads and control the axial clearance of the rotor and stator. Travel bearings bear radial loads and control radial clearance. When any bearing fails, the aircraft may be forced to land; otherwise, the aircraft may be destroyed, and personnel may lose their lives [7,8]. Therefore, developing efficient and accurate bearing fault diagnosis methods has become necessary to ensure equipment reliability and reduce maintenance costs.

In the past decades, previous studies have proposed various methods for intelligent bearing fault diagnosis [9,10], which can be divided into two categories: model-based methods and data-driven methods [11]. The model-based methods extract features from raw data according to prior knowledge and construct mathematical models to evaluate the health state of bearings [12,13]. However, model-based methods rely too much on prior knowledge, so their robustness could be better. Data-driven methods refer to establishing mappings from fault space to feature space by using a large amount of training data

without any knowledge and experience [14]. With the recent development of sensing and intelligent computing technology, it has become more accessible and easier to collect massive amounts of data, and data-driven intelligent diagnosis methods have attracted increased attention [15].

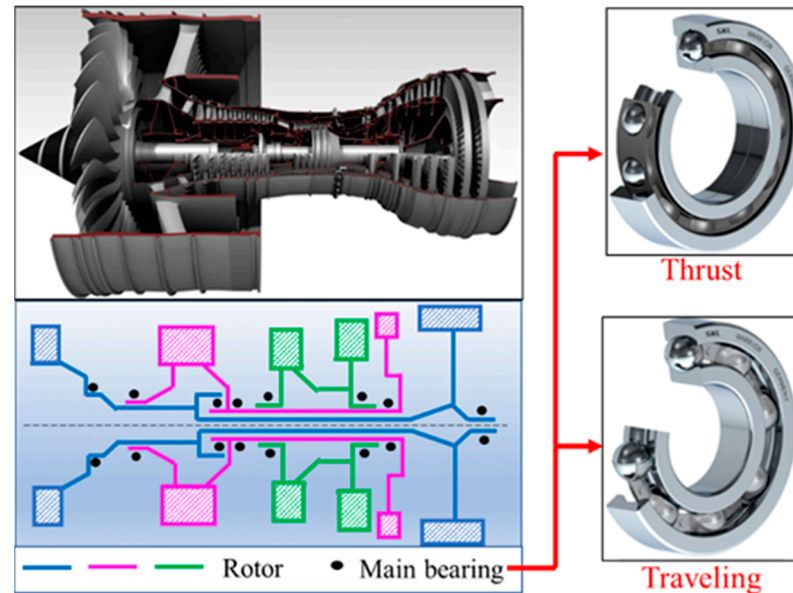


Figure 1. Two kinds of main bearings in aeroengine.

Data-driven methods have swept through the field of intelligent bearing fault diagnosis due to their end-to-end diagnosis mode [16]. For instance, Ahmed et al. [17] used a sparse autoencoder to learn over-complete sparse representations of datasets compressed by compressed sampling. Linshan et al. [18] proposed a novel convolution neural network (CNN) model named the Gramian Time-Frequency Enhancement Network (GTFE-Net) for bearing fault diagnosis. Kaicheng et al. [19] combined Improve Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and CNN to enhance recognition of the state of rolling bearings. Diwang et al. [20] improved the performance of CNN by guiding the CNN design with the physical characteristics of bearing acceleration signals. Wang et al. [21] presented an improved spiking neural network for inter-shaft bearing fault diagnosis. Sinitsin et al. [22] designed a novel hybrid CNN-MPL model-based bearing fault-diagnosis method. He et al. [23] proposed a new framework based on small labeled infrared thermal images and an enhanced convolutional neural network transferred from a convolutional autoencoder.

Although data-driven methods have been applied to bearing fault diagnosis, most assume that the source dataset, which indicates the dataset used for the training model, and the target dataset, which indicates the datasets used for the testing model, are from the same distribution [24,25]. However, the bearing usually works under complex and changing conditions in actual scenarios, so the assumption cannot be achieved [26,27]. To address this problem, transfer learning (TL) emerged [28,29]. Transfer learning improves performance in the target domain by transferring knowledge from different but related source domains [30,31]. In this way, the reliance on copious amounts of target domain data for building learners can be reduced [32,33]. A feasible solution is to find standard latent features through feature transformation and use them as a bridge for knowledge transfer, converting each original feature into a new feature representation for knowledge transfer [34,35]. Wen et al. [36] reduced the distribution discrepancy across the two domains by minimizing the maximum mean discrepancy (MMD). Sinno et al. [37] designed a novel metric to measure the discrepancy between source and target datasets. Chunran et al. [38] proposed a class-level matching TL network to match source and target domain data. Wang et al. [39] used domain-adversarial training based on the Wasserstein distance to

learn domain-invariant features from the raw signal. Cui et al. [40] combined distance-based domain adaptation and adversarial-based domain adaptation to propose a multi-adversarial joint distribution adaptation network.

This paper proposes an improved adversarial transfer network (IATN) for bearing fault diagnosis under variable working conditions, inspired by the effectiveness of distance-based and adversarial-based domain adaptation. Specifically, this paper combines an adversarial transfer network with a short-time Fourier transform (STFT) [41] to obtain satisfactory results with lighter networks. Then, this paper employs an attention module to enhance feature fusion. Moreover, this paper designs a novel domain discrepancy hybrid metric loss to improve model transfer learning performance. Finally, this paper verifies the method's effectiveness on a private dual-rotor dataset, a Case Western Reserve University (CWRU) dataset [42] and the Ottawa dataset [43]. The main contributions are as follows:

- (1) This paper combines adversarial TL with the traditional signal processing method, STFT, which allows us to obtain satisfactory results with a much lighter network.
- (2) This paper employs a channel attention module to enhance feature fusion.
- (3) This paper designs a novel domain discrepancy hybrid metric to measure the discrepancy between features of the source dataset and target dataset features.
- (4) This paper evaluates the proposed approach on three test benches' single-source tasks and multi-source tasks. The results demonstrate that the proposed method performs better than other methods.

The structure of this paper is as follows: In Section 2, related knowledge is introduced. Section 3 presents the proposed method in detail. Section 4 provides information about the experiment and analysis of the results. Section 5 provides further discussion, and Section 6 provides conclusions.

2. Preliminary Information

2.1. Related Works

In industrial applications, bearings are key rotating machinery components, and monitoring and fault diagnosis of their condition is crucial to ensuring the stable operation of equipment. Especially under changing working conditions, traditional fault diagnosis methods are often difficult to adapt to different operating environments, resulting in a decline in diagnostic performance. In recent years, deep learning, as a powerful end-to-end strategy, has been widely used to diagnose bearing faults intelligently. This section reviews significant recent research progress in the intelligent diagnosis of bearing faults under variable operating conditions.

In recent years, multi-scale feature learning has shown strong potential in bearing fault diagnosis. For example, Chen et al. [44] proposed a method based on multi-scale convolutional neural networks to effectively extract bearing fault features for fault diagnosis. Hu et al. [45] proposed a novel fault diagnosis method based on enhanced multi-scale sample entropies and balanced adaptation regularization for fault diagnosis needs under different working conditions. To solve the problem of inconsistent training and test data distribution, Guo et al. [46] explored the application of transfer learning technology in bearing fault diagnosis. They proved its effectiveness in cross-domain fault diagnosis. Saha et al. [47] employed transfer learning and random forest classification to exemplify the potential of combining deep learning techniques with traditional algorithms in bearing fault diagnosis. Moreover, researchers have also widely discussed data enhancement methods based on generative adversarial networks (GANs). Wang et al. [48] developed a data augmentation strategy based on GANs to improve the model's generalization ability under varying working conditions. Shi et al. [49] encapsulated a shared feature extractor, a label predictor and a series of domain discriminators to propose an adversarial multi-source data subdomain adaptation (AMDSA) model. Zhu et al. [50] present a deep subdomain adaptation network (DSAN) by aligning the relevant subdomain distributions of domain-specific layer activations across domains based on a local maximum mean. Meng et al. [51] explore the application of graph neural networks in bearing fault diagnosis.

In addition, ensemble learning methods improve diagnostic accuracy by combining the prediction results of multiple learners and have also received increasing attention. Tong et al. [52] proposed a multi-sensor information fusion method based on ensemble learning to automatically learn fault-related information from multi-sensor signals and provide accurate diagnosis results. Chen et al. [53] proposed an explainable learning framework to further improve the accuracy and robustness of bearing fault diagnosis under variable working conditions.

Through these studies, one can see the rapid development of intelligent diagnostic technology in bearing fault detection. These techniques improve the accuracy of fault diagnosis and enhance the model’s adaptability in the face of changing operating conditions.

2.2. Domain-Adversarial Adaptation Network

DAT maps two different domains into a common subspace to eliminate the differences between the domains [54]. As Figure 2 shows, a domain-adversarial adaptation network consists of three parts—feature extractor θ_f , classifier θ_c and domain discriminator θ_D [55].

Given a labeled source dataset D_s and an unlabeled target dataset D_T , both are input into θ_f , and then the features of D_s are input into θ_c and obtain the predicted class. According to the predicted label and the true label, class loss L_C is calculated. Moreover, the features of D_s and D_T are input into θ_D to obtain the predicted domain, followed by the domain loss L_D .

In the training process of domain-adversarial adaptation networks, all the parameters are updated by Equation (1):

$$\begin{aligned} \theta_f &\leftarrow \theta_f - \alpha \left(\frac{\partial L_C}{\partial \theta_f} - \beta \frac{\partial L_D}{\partial \theta_f} \right) \\ \theta_C &\leftarrow \theta_C - \alpha \frac{\partial L_C}{\partial \theta_C} \\ \theta_D &\leftarrow \theta_D - \alpha \frac{\partial L_D}{\partial \theta_D} \end{aligned} \tag{1}$$

where α and β are positive constants. When this paper updates the parameters, this paper multiplies the gradient of L_D to θ_f by a negative constant $-\beta$, which prevents θ_f from extracting features for domain discrimination.

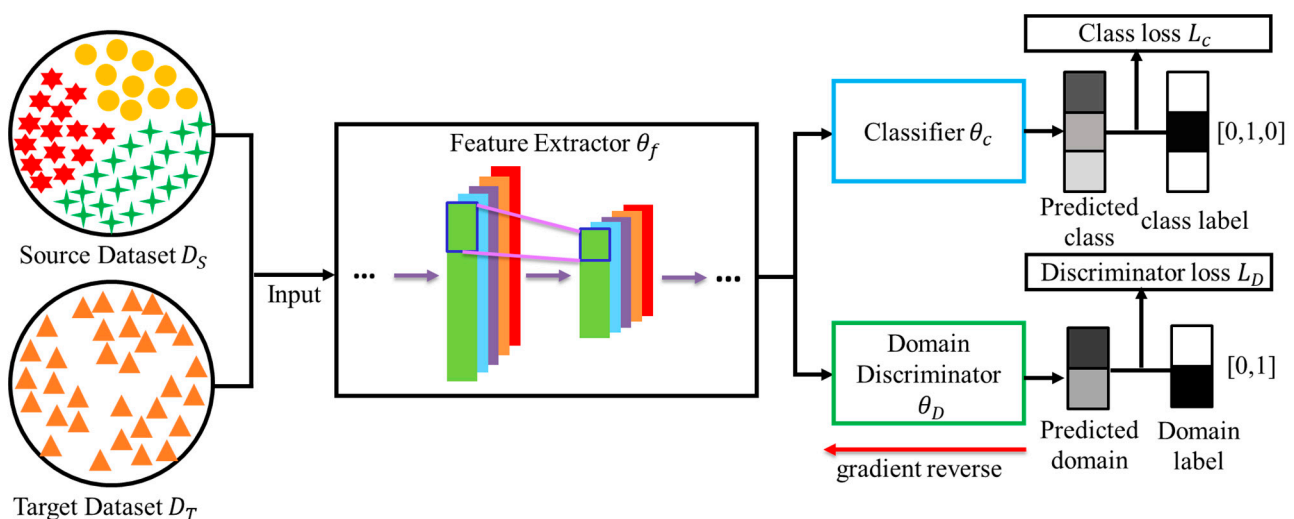


Figure 2. The structure of a domain-adversarial adaptation network. The lighter the output color, the closer the value is to 0; the darker the output color, the closer the value is to 1.

In general, domain-adversarial adaptation networks can be thought of as a two-player game. In this game, the players are the domain discriminator θ_D and the feature extractor θ_f [56]. The target of θ_D is to identify the domain of each sample, but θ_f strives to deceive θ_D . After long-term adversarial training, θ_f specifically extracts domain-invariant features.

2.3. Maximum Mean Difference (MMD)

The purpose of TL is to apply the knowledge learned in the source domain to a different but related target domain. Essentially, it is to find a loss function that minimizes the distance between domain features and target domain features. Therefore, MMD is used to measure the difference in data distribution between the two domains [57]. Given two distributions, X and Y , MMD is defined as follows:

$$\text{MMD}(X, Y) = \left\| \frac{1}{n} \sum_{i=1}^n \phi(x_i) - \frac{1}{m} \sum_{j=1}^m \phi(y_j) \right\|_H^2 \quad (2)$$

where H means that this distance is measured by mapping the data into the regenerative Hilbert space using the kernel function $\phi(x)$. m , n are the sample numbers of two distributions, X and Y .

In a word, the meaning of MMD is to map variables onto a high-dimensional space through a kernel function and then find the expected difference between two distributed random variables after mapping.

3. IATN for Bearing Fault Diagnosis under Variable Working Conditions

Like domain-adversarial adaptation networks, the proposed IATN also consists of three parts—feature extractor θ_f , classifier θ_c and domain discriminator θ_D . Meanwhile, there are some improvements in the proposed approach.

3.1. Preprocess

Compared with the raw time-domain signal, the time-frequency image has both time and frequency domain information and contains more fault information. In addition, the features of time-frequency images are more prominent and require lower feature extraction capabilities of the network. In fault diagnosis research, relevant studies have proved that time-frequency domain input is better than time-domain input [58]. By converting the raw 1D time-domain signal into a 2D time-frequency diagram, the features of the signal are enhanced, allowing us to use lighter downstream networks for classification.

Therefore, this paper performs STFT on the collected raw signals. STFT selects a time-frequency localized window function, assumes that the signal is stationary (pseudo-stationary) within a brief time interval and moves the window function to calculate the frequency spectrum at different moments [59]. STFT is defined as follows:

$$\text{STFT}_x(t, f) = \int_{-\infty}^{+\infty} x(\tau) g^*(\tau - t) e^{-j2\pi f \tau} d\tau \quad (3)$$

where $x(t)$ is the raw signal, and $g^*(\tau - t)$ is the window function.

Ideally, STFT can provide time and frequency information at the same time so that the frequency distribution of the signal over time can be observed, which is crucial for fault diagnosis. By converting a signal into its time-frequency representation, STFT can reveal structural features that may not be apparent in the original time-domain signal. This conversion can improve the separability between different categories, such as inner rings, outer rings, rolling elements and different degrees, such as standard, minor and major faults, so shallow neural networks can achieve high-accuracy classification.

However, in actual scenarios, various factors often interfere with bearing signals, including background noise, structural resonance, amplitude modulation, etc. At this time, the classification of time-frequency diagrams may involve highly nonlinear and complex data structures, especially when there are multiple fault types and changing work conditions. Shallow neural network performance cannot model such complexity and nonlinearity. In contrast, deep neural networks could be better at extracting high-level features from complex data. Deep neural networks extract abstract features that are difficult to observe directly through multi-layer nonlinear transformations to achieve high-precision classification of time-frequency images.

3.2. Feature Extractor

As shown in Figure 3, differently from previous studies, the input of IATN is a 2D time-frequency diagram rather than a 1D vibration signal. Therefore, the feature extractor consists of a 2D CNN. Since this paper performs STFT on the raw signal, the feature extractor can use a lighter network. The structure of the feature extractor of the model is shown in Table 1.

Table 1. The structure of feature extractor.

Layer	Kernel	Channel
Conv2D	3×3	16
BatchNorm	/	16
MaxPool	/	16
Relu	/	16
Conv2D	3×3	32
BatchNorm	/	32
MaxPool	/	32
Relu	/	32

Moreover, the hyperparameters of IATN for all tasks are shown in Table 2.

Table 2. Hyperparameters used in IATN.

Hyperparameter	First Appeared	Value
d	(4)	32
w	(4)	64
β	(14)	0.04
γ	(13)	0.1

In addition, since convolution only operates in a local space, obtaining enough information to extract the relationship between channels takes much work. Therefore, this paper introduces the channel attention mechanism to determine the importance of each feature channel, thereby enhancing the weight of essential features and reducing the weight of irrelevant information [60].

The channel attention mechanism includes three steps: squeezing, excitation and scaling. As shown in Figure 3, the number of channels of the feature X_2 of the input channel attention layer is C_2 , and the data are expressed as $H_2 \times W_2$. The squeeze operation applies global average pooling to convert the data into Z , with a shape of $1 \times 1 \times C_2$. This process expands the receptive field and encodes the entire spatial feature on a channel into a global feature, as shown in (4).

$$Z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (4)$$

where z_c denotes the c -th element of matrix Z , and u_c denotes the c -th channel of C_2 .

Subsequently, excitation obtains the feature weight of each channel through two fully connected layers, as shown in (5):

$$S = F_{ex}(Z, W) = \sigma(g(Z, W)) = \sigma(W_2 \delta(W_1 Z)) \quad (5)$$

where $\sigma(x)$ denotes sigmoid function, $\delta(x)$ denotes the Relu function.

Finally, scaling multiplies the obtained feature weights S with the input X_2 channel by channel, as shown in (6):

$$X_3^c = F_{scale}(u_c, s_c) = s_c u_c \quad (6)$$

where X_3^c denotes the c -th channel of X_3 .

After that, in the training, both the samples from source datasets and the target dataset are input into the feature extractor, and then the feature extractor outputs source features F_S and F_T .

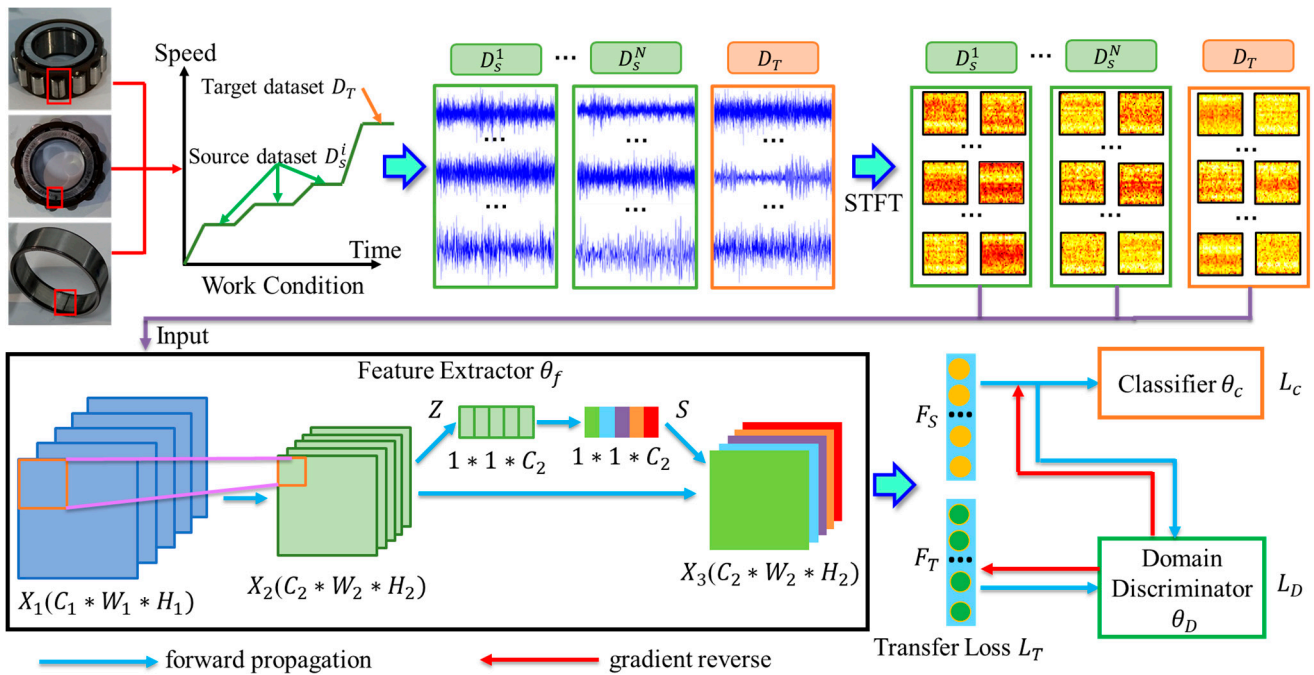


Figure 3. IATN for bearing fault diagnosis under variable working conditions.

3.3. Transfer Loss

As mentioned in Section 2, the most used transfer loss function is currently MMD. One of the most important concepts of MMD is the kernel function $\phi(x)$. In MMD, $\phi(x)$ is fixed. The Gaussian kernel function is usually chosen as the kernel function because the Gaussian kernel function can map data onto an infinite-dimensional space:

$$k(x, x') = \exp\left(-\|x - x'\|^2 / (2\sigma^2)\right) \quad (7)$$

Theoretically, any selection of a kernel function can determine the MMD of two different distributions. But in some extreme cases, this kernel function is determined to make the MMD distance of two different distributions exceedingly small. To avoid this situation, MK-MMD is proposed.

$$\mathcal{K} \triangleq \left\{ k = \sum_{u=1}^m \beta_u k_u : \sum_{u=1}^m \beta_u = 1, \beta_u \geq 0, \forall u \right\} \quad (8)$$

By using multi-kernel functions, this paper can better represent the differences in data distribution in high-dimensional space and improve the representation ability of model features.

$$\text{MK-MMD}(X, Y) = \left\| \frac{1}{n} \sum_{i=1}^n \mathcal{K}(x_i) - \frac{1}{m} \sum_{j=1}^m \mathcal{K}(y_j) \right\|_H^2 \quad (9)$$

Moreover, to further enhance the domain adaptation performance of the transfer loss, this paper designs a novel domain difference measure combining MK-MMD and feature center distance (FCD). FCD refers to calculating the distance between the target domain feature clustering center and the source domain feature clustering center. Assuming that there are m samples $X^k = [X_1, X_2 \dots X_m]$ of the k -th category in the source domain, and

there are n samples $Y^k = [Y_1, Y_2 \dots Y_m]$ in the target domain judged to be the k -th category by the model, the definition of FCM is shown in (10):

$$L_{FCM} = \frac{1}{k} \sum_{k=1}^K \dot{s}_k = \frac{1}{k} \sum_{k=1}^K \left| \overline{X^k} - \overline{Y^k} \right| \quad (10)$$

where K is the number of categories. By combining (9) and (10), the novel transfer loss is defined as (11):

$$L_T = L_{MK-MMD} + L_{FCM} \quad (11)$$

The combination of MMD and FCD enables the model to use different feature alignment strategies for feature transfer. FCD promotes local feature alignment, and MMD is used for global feature alignment. The integration of these two loss functions combines local and global feature alignment, thereby enhancing domain adaptation performance.

3.4. Training Process

As Figure 3 shows, the total loss function consists of three parts, transfer loss L_T , class loss L_C and domain loss L_D . They are defined as follows:

$$\begin{aligned} L_C &= \text{MSE}(Y_C, Y_C^{label}) \\ L_D &= \text{MSE}(Y_D, Y_D^{label}) \\ L_{total} &= \gamma L_T + L_C + L_D \end{aligned} \quad (12)$$

where MSE is the mean squared error, Y_C and Y_D are the output of the classifier and domain discriminators and Y_C^{label} and Y_D^{label} are the class label and domain label.

Since there is gradient reverse in the backpropagation update process of domain loss L_D , the model parameters are updated as shown in (13):

$$\begin{aligned} \theta_f &\leftarrow \theta_f - \alpha \left(\frac{\partial L_C}{\partial \theta_f} - \beta \frac{\partial L_D}{\partial \theta_f} + \gamma \frac{\partial L_T}{\partial \theta_f} \right) \\ \theta_c &\leftarrow \theta_c - \alpha \left(\frac{\partial L_C}{\partial \theta_c} \right) \\ \theta_D &\leftarrow \theta_D - \alpha \left(\frac{\partial L_D}{\partial \theta_D} \right) \end{aligned} \quad (13)$$

3.5. Dataset

To evaluate the performance of the proposed method more comprehensively, this paper verifies its effectiveness on three datasets.

The first dataset is private. As Figure 4 shows, it is a dual-rotor test bench to simulate dual-rotor engine operation. The low-pressure rotor (LR) and the high-pressure rotor (HR) are driven by different motors and rotate in the same direction at different speeds. The LR is supported by the No. 1 and No. 4 bearings, and the high-pressure rotor HR is supported by the No. 2 and 3 bearings. The inter-shaft bearing is No. 3, and the experimental bearing is between the two rotors. The outer ring of the inter-shaft bearing rotates with the HR at a higher speed, and the inner ring rotates with the LR at a lower speed.

This paper collects vibration data under four working conditions (WDs), as shown in Table 3. Each working condition includes bearing data in nine different health states, including normal state, three degrees of inner-ring failure, three degrees of rolling-element failure and two different degrees of outer-ring failure. The sampling frequency is 20.48 k Hz. In addition, this paper uses a sliding window of a length of 1024 to sample the raw signal without overlap and collect 100 samples for each health state under each operating condition.

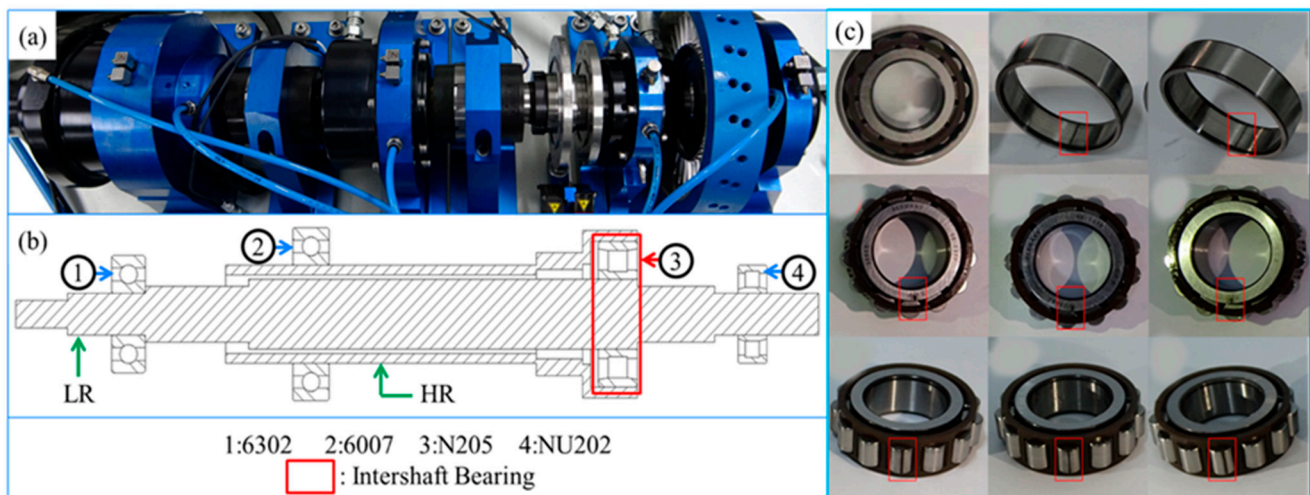


Figure 4. Dual-rotor test bench. (a) The test rig; (b) engineering sketch; (c) bearing failure mode.

Table 3. The working conditions of dual-rotor test bench.

Number	LR Speed (rpm)	HR Speed (rpm)
D0	2000	4000
D1	3000	6000
D2	4000	8000
D3	5500	10,000

The second dataset is the public Case Western Reserve University dataset (CWRU). The CWRU dataset has four working conditions: 1797 rpm (R0), 1772 rpm (R1), 1750 rpm (R2) and 1730 rpm (R3). Under all working conditions, this paper selects normal data and three different degrees of fault data of the inner ring, outer ring and rolling elements collected at the driving end as experimental data. The sampling frequency is 12,000 Hz.

The third dataset is the public Ottawa dataset. The Ottawa dataset has four working conditions: increasing speed (O0), decreasing speed (O1), increasing then decreasing speed (O2) and decreasing then increasing speed (O3). Under all working conditions, this paper selects normal data and fault data of the inner ring and outer ring. The sampling frequency is 200,000 Hz.

The training epoch is uniformly set to 300 using the Adams optimizer, and the learning rate is 0.001. The code is written through the Pytorch 2.2 framework and runs on the GPU of GPTX1650.

4. Results

Here this paper compares the performance of the IATN with CNN [61], DANN [62], AMDSA, DASN, ISAE-CSDF [63] and CDGATLN [64] models on the three datasets, and the results are shown in Tables 4–6. In addition, the experiment in this paper is repeated 20 times to eliminate the influence of randomness.

One can generally make three observations: (1) the IATN has the highest accuracy on most tasks. (2) the IATN has the most minor variances. (3) The more similar the working conditions of the source and target domain data are, the better the transfer learning effect will be.

In Table 4, the diagnosis accuracy of the IATN is significantly higher than the other six methods on most tasks. In the task of D0→D3, all the methods have the lowest accuracy. The accuracies of the CNN, DANN, AMDSA, DASN, ISAE-CSDF and SA-SN-DCGAN are 58.75%, 61.82%, 65.28%, 60.32%, 65.64% and 66.82%, respectively, but the accuracy of the IATN is 70.62%. In the task of D0, D1, D3→D2, the accuracies of the CNN, DANN, TL and DSAN are 95.12%, 96.16%, 95.44%, 93.66%, 96.08% and 96.06%, respectively, but the

accuracy of the IATN is 98.44%. In the last row, this paper calculates the average accuracy of the five methods on all tasks, and the average accuracy of the IATN is 91.89%, which is 6.82%, 4.77%, 1.06%, 2.52%, 2.58% and 1.23% higher than the CNN, DANN, ISAE-CSDF and SA-SN-DCGAN, respectively. As for variance, the average variance of the IATN is 1.10%, which is significantly lower than the averages of the other six methods, which are 2.29%, 1.72%, 1.52%, 1.91%, 1.35% and 1.49%, respectively.

Table 4. Performance comparison of dual-rotor test bench.

Task	Model Performance (%)						
	CNN	DANN	AMDSA	DSAN	ISAE-CSDF	CDGATLN	IATN
D1→D0	84.22 ± 1.21	85.22 ± 1.08	89.24 ± 0.98	89.20 ± 1.04	89.78 ± 1.22	89.92 ± 1.14	90.76 ± 1.06
D2→D0	68.80 ± 2.42	72.80 ± 3.42	79.26 ± 1.82	76.84 ± 2.42	78.24 ± 1.79	79.94 ± 1.62	82.82 ± 1.58
D3→D0	68.67 ± 1.46	72.76 ± 3.18	79.32 ± 1.74	78.75 ± 2.36	76.58 ± 1.58	78.75 ± 2.36	80.76 ± 1.38
D1, D2→D0	84.11 ± 1.82	87.14 ± 1.69	91.42 ± 1.58	91.26 ± 1.78	90.68 ± 1.58	91.36 ± 1.68	92.44 ± 1.28
D1, D3→D0	86.44 ± 2.05	89.84 ± 2.16	90.72 ± 1.46	91.13 ± 1.68	89.94 ± 1.02	91.24 ± 1.08	92.31 ± 0.76
D2, D3→D0	78.89 ± 2.28	81.86 ± 1.06	89.92 ± 1.88	88.98 ± 2.56	86.73 ± 1.42	90.28 ± 1.52	91.89 ± 1.26
D1, D2, D3→D0	87.33 ± 3.79	89.67 ± 0.96	92.56 ± 1.24	93.12 ± 1.36	90.54 ± 1.06	92.02 ± 1.12	93.42 ± 0.88
D0→D1	88.89 ± 1.67	89.16 ± 1.56	92.12 ± 1.72	92.02 ± 1.92	89.49 ± 1.82	92.12 ± 1.26	92.82 ± 1.08
D2→D1	88.26 ± 1.88	89.06 ± 1.68	95.86 ± 1.94	95.46 ± 2.12	94.62 ± 1.58	95.46 ± 2.12	97.24 ± 1.32
D3→D1	80.33 ± 1.24	81.33 ± 1.12	81.48 ± 1.68	81.86 ± 1.94	81.28 ± 1.26	81.86 ± 1.94	82.62 ± 1.12
D0, D2→D1	97.49 ± 1.40	97.89 ± 0.62	98.02 ± 1.72	98.08 ± 1.64	97.77 ± 0.98	97.08 ± 1.24	98.10 ± 0.86
D0, D3→D1	90.67 ± 1.38	92.04 ± 1.28	94.32 ± 1.72	96.28 ± 1.98	94.76 ± 1.44	96.58 ± 1.78	97.52 ± 1.28
D2, D3→D1	90.72 ± 2.38	92.06 ± 1.42	92.58 ± 1.20	96.38 ± 1.04	94.82 ± 0.94	96.28 ± 1.02	97.13 ± 0.80
D0, D2, D3→D1	97.67 ± 1.24	98.06 ± 0.82	98.00 ± 0.52	98.02 ± 0.66	97.78 ± 0.58	98.02 ± 0.76	98.12 ± 0.46
D0→D2	73.42 ± 3.42	75.72 ± 3.12	98.24 ± 2.08	79.36 ± 2.36	78.24 ± 1.96	79.06 ± 2.46	79.26 ± 1.82
D1→D2	91.11 ± 2.14	92.48 ± 2.08	95.62 ± 0.75	95.46 ± 0.92	95.26 ± 0.68	94.62 ± 0.90	96.78 ± 0.76
D3→D2	87.18 ± 1.04	87.78 ± 0.94	92.12 ± 1.52	90.16 ± 2.02	91.20 ± 1.16	92.76 ± 1.42	94.58 ± 1.06
D0, D1→D2	92.24 ± 1.58	93.92 ± 1.76	96.88 ± 1.06	96.84 ± 1.38	95.58 ± 0.92	96.82 ± 0.98	97.12 ± 0.76
D0, D3→D2	89.94 ± 1.88	91.72 ± 1.06	94.52 ± 1.74	93.24 ± 2.12	93.56 ± 1.42	94.28 ± 1.52	96.13 ± 1.08
D1, D3→D2	95.12 ± 0.88	96.16 ± 1.42	95.44 ± 0.98	93.66 ± 1.38	96.02 ± 0.88	97.68 ± 0.68	98.44 ± 0.44
D0, D1, D3→D2	96.76 ± 1.02	96.72 ± 1.42	96.06 ± 1.42	96.16 ± 1.98	96.08 ± 0.82	96.06 ± 1.58	97.36 ± 1.02
D0→D3	58.78 ± 5.68	61.82 ± 3.28	65.28 ± 2.92	60.32 ± 3.88	65.64 ± 2.96	66.82 ± 3.28	70.62 ± 2.18
D1→D3	75.12 ± 2.66	77.24 ± 3.08	77.42 ± 2.40	74.78 ± 3.18	79.54 ± 2.62	80.18 ± 2.78	81.68 ± 1.88
D2→D3	86.08 ± 2.14	90.06 ± 2.02	93.64 ± 1.48	93.82 ± 2.26	91.54 ± 1.28	94.42 ± 0.76	97.89 ± 0.48
D0, D1→D3	70.24 ± 5.98	75.24 ± 2.46	84.28 ± 1.62	76.17 ± 2.14	84.21 ± 1.56	88.12 ± 1.64	90.16 ± 1.18
D0, D2→D3	92.12 ± 3.56	94.66 ± 1.06	97.16 ± 1.04	97.26 ± 1.24	93.53 ± 0.96	95.16 ± 1.04	97.62 ± 0.64
D1, D2→D3	89.33 ± 2.16	91.53 ± 1.48	95.68 ± 0.88	95.92 ± 1.88	93.52 ± 0.92	94.02 ± 1.08	96.74 ± 0.74
D0, D1, D2→D3	92.24 ± 3.68	95.48 ± 1.09	96.12 ± 1.68	96.18 ± 2.26	93.88 ± 1.42	96.78 ± 1.16	97.33 ± 1.08
Average	85.07 ± 2.29	87.12 ± 1.72	90.83 ± 1.52	89.37 ± 1.91	89.31 ± 1.35	90.66 ± 1.49	91.89 ± 1.10

Table 5. Performance comparison of CWRU.

Task	Model Performance (%)						
	CNN	DANN	AMDSA	DSAN	ISAE-CSDF	SA-SN-DCGAN	IATN
R1→R0	93.10 ± 0.38	94.35 ± 0.98	96.02 ± 1.20	95.62 ± 1.32	94.68 ± 0.52	95.22 ± 0.72	97.76 ± 0.36
R2→R0	86.30 ± 0.46	88.82 ± 0.82	90.82 ± 1.52	90.42 ± 1.88	90.26 ± 0.89	91.22 ± 0.88	94.82 ± 0.58
R3→R0	83.76 ± 1.22	86.86 ± 1.68	88.72 ± 1.98	84.26 ± 2.14	90.13 ± 1.76	87.06 ± 1.54	93.76 ± 1.28
R0→R1	97.25 ± 0.36	97.66 ± 0.62	98.40 ± 0.92	97.88 ± 1.04	98.26 ± 0.72	97.28 ± 0.32	99.40 ± 0.26
R2→R1	97.22 ± 0.48	99.02 ± 0.46	99.30 ± 0.65	99.02 ± 0.78	98.52 ± 0.48	99.22 ± 0.78	99.60 ± 0.18
R3→R1	94.76 ± 0.82	94.96 ± 0.58	98.24 ± 0.95	97.76 ± 1.14	96.28 ± 0.78	97.06 ± 0.72	97.86 ± 0.62
R0→R2	91.22 ± 0.66	91.76 ± 1.66	96.28 ± 1.65	95.64 ± 1.68	93.58 ± 0.92	94.64 ± 1.14	97.64 ± 0.84
R1→R2	99.96 ± 0.04	99.98 ± 0.02	99.90 ± 0.04	99.96 ± 0.04	99.98 ± 0.02	99.96 ± 0.04	100.00 ± 0.00
R3→R2	99.94 ± 0.06	99.96 ± 0.04	99.95 ± 0.04	99.92 ± 0.06	99.98 ± 0.02	99.92 ± 0.08	100.00 ± 0.00
R0→R3	81.56 ± 1.24	86.26 ± 1.36	89.24 ± 1.53	90.26 ± 1.62	87.38 ± 0.92	90.06 ± 1.42	92.56 ± 1.06
R1→R3	95.32 ± 0.48	94.82 ± 0.68	98.38 ± 0.92	98.82 ± 1.08	97.66 ± 0.68	98.22 ± 0.88	99.52 ± 0.48
R2→R3	99.90 ± 0.08	99.90 ± 0.10	99.90 ± 0.10	99.92 ± 0.08	99.95 ± 0.05	99.92 ± 0.08	100.00 ± 0.00
Average	93.36 ± 0.52	94.52 ± 0.75	96.26 ± 0.96	95.79 ± 1.07	95.55 ± 0.64	95.81 ± 0.72	97.74 ± 0.47

In Tables 5 and 6, one could also find similar discoveries by comparing the performance of the IATN and the other methods. The average accuracy of the IATN on all CWRU tasks is 97.74%, which is 4.38%, 3.22%, 1.48%, 1.95%, 2.19% and 1.93% higher than the CNN, DANN, AMDSA, DASN, ISAE-CSDF and SA-SN-DCGAN, respectively. The average variance of the IATN on all CWRU tasks is 0.47%, which is lower than the other six methods, which

are 0.52%, 0.75%, 0.96%, 1.07%, 0.64% and 0.72%, respectively. The average accuracy of the IATN on all Ottawa dataset tasks is 81.11%, which is 6.77%, 4.23%, 1.42%, 3.78%, 3.28% and 0.84% higher than the CNN, DANN, AMDSA, DASN, ISAE-CSDF and SA-SN-DCGAN, respectively. The average variance of the IATN on all Ottawa dataset tasks is 1.24%, which is lower than those of the other six methods, which are 1.56%, 1.97%, 2.10%, 2.38%, 1.28% and 1.34%, respectively.

Table 6. Performance comparison of Ottawa dataset.

Task	Model Performance (%)						
	CNN	DANN	AMDSA	DSAN	ISAE-CSDF	SA-SN-DCGAN	IATN
O1 → O0	64.20 ± 1.82	65.38 ± 2.48	66.81 ± 2.57	65.58 ± 2.95	65.76 ± 1.86	66.64 ± 1.72	67.53 ± 1.42
O2 → O0	58.30 ± 1.86	62.82 ± 2.52	67.84 ± 2.62	72.26 ± 3.01	64.86 ± 1.42	68.52 ± 1.78	70.36 ± 1.46
O3 → O0	70.26 ± 1.52	71.86 ± 1.68	73.79 ± 1.86	63.82 ± 2.04	72.58 ± 1.26	73.36 ± 1.14	74.76 ± 1.08
O0 → O1	60.25 ± 1.76	63.76 ± 2.62	66.94 ± 2.68	64.04 ± 3.14	64.36 ± 1.62	65.78 ± 1.68	68.53 ± 1.56
O2 → O1	84.32 ± 1.28	85.08 ± 1.46	87.98 ± 1.56	85.82 ± 1.71	86.92 ± 0.78	87.12 ± 1.08	88.33 ± 0.76
O3 → O1	85.26 ± 1.22	86.42 ± 1.56	87.24 ± 1.72	86.72 ± 1.98	87.08 ± 1.18	97.76 ± 1.14	88.76 ± 1.28
O0 → O2	48.76 ± 2.56	54.36 ± 2.68	61.9 ± 3.04	55.75 ± 3.29	57.18 ± 1.62	62.54 ± 1.88	65.67 ± 1.84
O1 → O2	76.26 ± 1.24	80.88 ± 1.82	84.84 ± 1.84	81.33 ± 2.14	81.88 ± 1.02	83.76 ± 0.94	86.83 ± 0.98
O3 → O2	93.94 ± 1.28	94.98 ± 1.54	95.91 ± 1.64	95.25 ± 1.82	95.58 ± 0.92	95.92 ± 0.88	96.38 ± 0.86
O0 → O3	70.56 ± 1.84	73.56 ± 1.98	76.84 ± 2.32	73.92 ± 2.57	74.38 ± 1.82	76.26 ± 1.82	78.48 ± 1.78
O1 → O3	85.32 ± 1.28	87.72 ± 1.88	89.54 ± 1.94	87.64 ± 2.27	87.56 ± 1.29	88.88 ± 1.28	90.46 ± 1.18
O2 → O3	94.70 ± 1.12	95.76 ± 1.46	96.76 ± 1.52	95.87 ± 1.72	95.85 ± 0.68	96.42 ± 0.82	97.26 ± 0.78
Average	74.34 ± 1.56	76.88 ± 1.97	79.69 ± 2.10	77.33 ± 2.38	77.83 ± 1.28	80.27 ± 1.34	81.11 ± 1.24

The comparison results demonstrate that the IATN significantly improves bearing fault diagnosis under variable WDs. Moreover, when the WDs of the source and target domains are similar, the test accuracy obtained is also higher. For example, the accuracy of all methods on the D2 → D3 task is higher than that on the D0 → D3 task. Data similarity may be higher when operating conditions are similar, making it easier to transfer fault diagnosis knowledge. This suggests that data similarity may be higher when WDs are similar, making it easier to transfer fault diagnosis knowledge.

5. Discussion

5.1. Running Process Comparison

To evaluate the performance of the model more comprehensively, this paper compares the running process of 300 epochs of different methods on the D0 → D1 task, the D0, D1 → D2 task and the D0, D1, D2 → D3 task, the results are shown in the Table 7 and Figure 5.

Table 7. Comparison of running-times of dual-rotor dataset.

Task	Running Time (s)				
	CNN	DANN	ISAE-CSDF	CDGATLN	IATN
D0 → D1	276	280	508	526	410
D0, D1 → D2	284	291	564	602	417
D0, D1, D2 → D3	287	295	765	826	601

According to Table 7, it is obvious that the running time of the IATN is more than those of the CNN and DANN but less than those of the ISAE-CSDF and CDGATLN. By comparing the running time of the IATN with those of the CNN and DANN, it can be concluded that the transfer process is much more time-consuming than the adversarial process.

Moreover, the training curves of the IATN, CNN and DANN are presented in Figure 5. In Figure 5, there are three subfigures. Figure 5a presents the training curves of the D0 → D1 task; Figure 5b presents the training curves of the D0, D1 → D2 task; and Figure 5c presents the training curves of the D0, D1, D2 → D3 task.

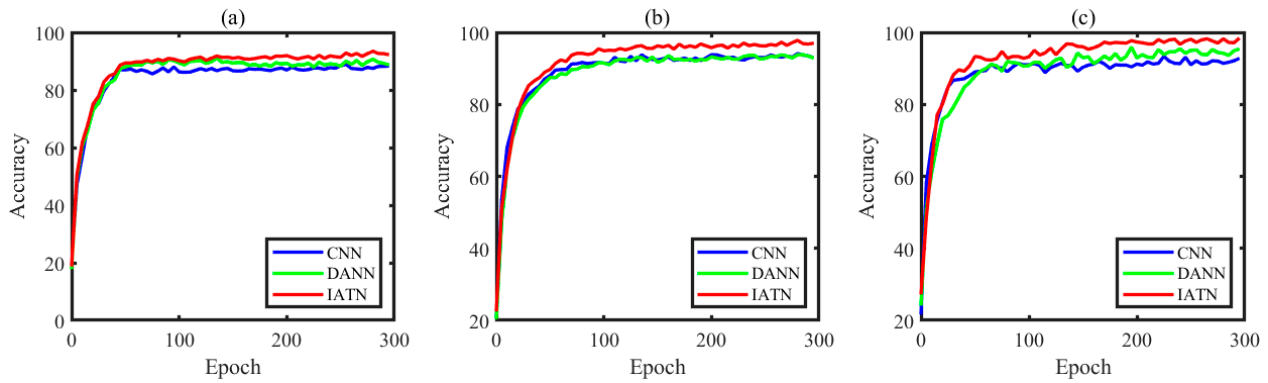


Figure 5. Training curves. (a) D0→D1, (b) D0, D1→D2, (c) D0, D1, D2→D3.

In Figure 5, the blue lines represent the accuracy of the CNN; the green lines represent the accuracy of the DANN, and the red lines represent the accuracy of the IATN. According to Figure 5, it is obvious that although the calculation process of the IATN is more complex, this does not affect its convergence. On the three tasks with different numbers of source domains, the convergence speed of the IATN is like that of the CNN and DANN.

5.2. Feature Visualization

According to Tables 4–6, it is obvious that the IATN performs better than the other methods in extracting domain-invariant features. To demonstrate the superiority of the IATN more intuitively in extracting domain-invariant features, this paper performs feature visualization of the D2→D3 task. Specifically, the source domain features and target domain features extracted by each model are reduced to two dimensions for visualization with t-distributed stochastic neighbor embedding (t-SNE) [65]. The results are shown in Figure 6.

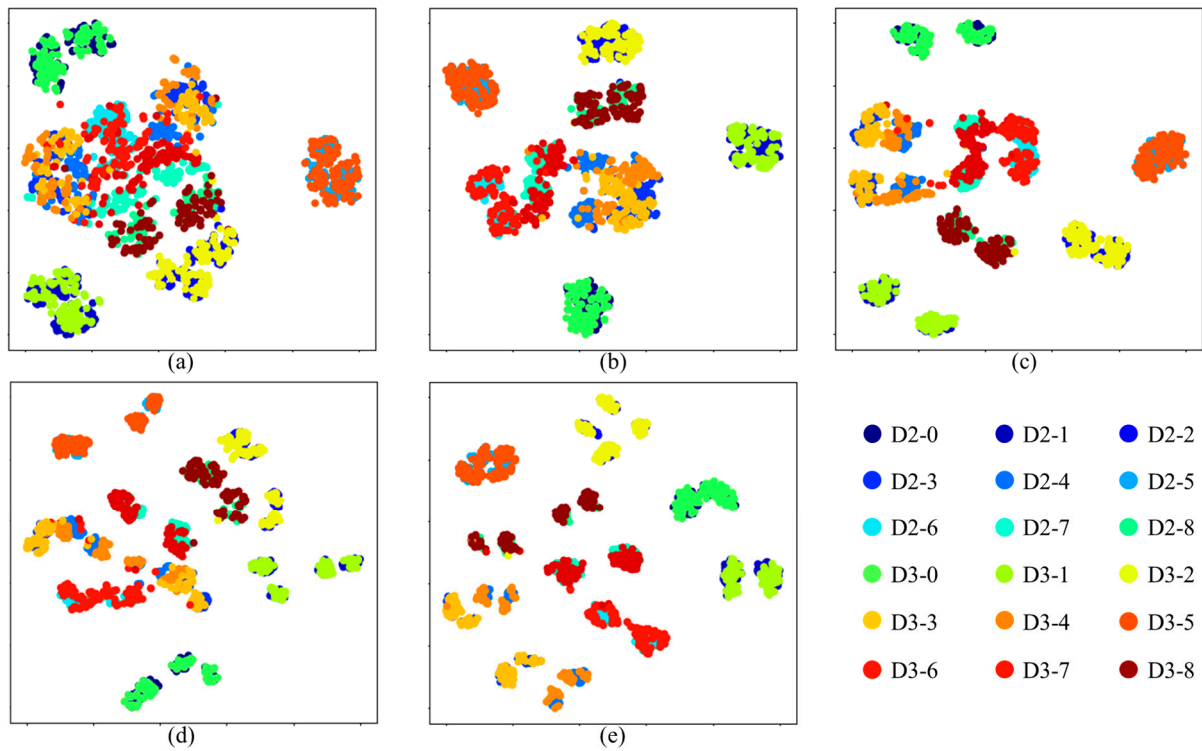


Figure 6. t-SNE feature visualization. (a) CNN. (b) DANN. (c) ISAE-CSDF. (d) CDGATLN. (e) IATN. D2-0 indicates data collected for category “0” fault bearings in D2 working conditions.

In Figure 6, there are five subfigures. Figure 6a is the feature visualization of the CNN, Figure 6b is the feature visualization of the DANN, Figure 6c is the feature visualization of the ISAE-CSDF, Figure 6d is the feature visualization of the CDGATLN and Figure 6e is the feature visualization of the IATN. In each subfigure, there are 18 clusters; 9 belong to D2 tasks and 9 belong to D3 tasks. There are two criteria for judging model performance. One is the degree of separation of the nine fault clusters in each WD, and the other is the degree of overlap of the same fault clusters in the two working conditions.

According to Figure 6, it is obvious that there is significant overlap between different fault clusters of source domain features and target domain features extracted by the CNN, DANN, ISAE-CSDF and CDGATLN, which initially have completely separate clusters, but there is still aliasing at the boundaries of some categories. In Figure 6e, all the clusters of each WD are separated entirely, and all the same fault clusters of D2 and D3 are basically overlapped, which indicates that the IATN has strong domain-invariant feature extraction capabilities.

5.3. Data Analysis

Based on the observations in Tables 4 and 5, the closer the working conditions are, the more similar the distribution of the collected data will be. This paper calculates the MK-MMD and MMD between the data collected under different working conditions to verify the hypothesis. The results are shown in Table 8.

Table 8. The distribution connections of different domains.

Task	MK-MMD	MMD	Mean Acc
D0↔D1	0.438	0.122	91.79
D0↔D2	1.085	0.312	90.03
D0↔D3	1.782	0.511	75.74
D1↔D2	0.316	0.087	97.01
D1↔D3	1.175	0.253	82.15
D2↔D3	0.321	0.084	96.23
R0↔R1	0.061	0.011	98.58
R0↔R2	0.072	0.010	96.23
R0↔R3	0.071	0.014	93.16
R1↔R2	0.052	0.008	99.80
R1↔R3	0.065	0.013	98.69
R2↔R3	0.054	0.012	100
O0↔O1	0.854	0.198	68.03
O0↔O2	0.927	0.209	68.01
O0↔O3	0.833	0.194	76.62
O1↔O2	0.207	0.045	87.58
O1↔O3	0.029	0.006	89.61
O2↔O3	0.021	0.004	96.82

According to Table 8, it is evident that the more similar the working conditions of the source domain data and the target domain data are, the lower the MK-MMD and MMD between them are. To express the connection between the distribution difference between source and target domain data and the model performance more intuitively, we perform correlation analysis on them according to (14):

$$\rho_{x,y} = \frac{\text{Cov}(X,Y)}{\sqrt{D(X)D(Y)}} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \quad (14)$$

The correlation coefficient between the MK-MMD of dual-rotor tasks and their transfer learning accuracy is defined as ρ_1 ; the correlation coefficient between the MMD of dual-rotor tasks and their transfer learning accuracy is defined as ρ_2 ; the correlation coefficient between the MK-MMD of CWRU tasks and their transfer learning accuracy is defined as ρ_3 ; the correlation coefficient between the MMD of CWRU tasks and their transfer learning

accuracy is defined as ρ_4 ; the correlation coefficient between the MK-MMD of Ottawa dataset tasks and their transfer learning accuracy is defined as ρ_5 ; the correlation coefficient between the MMD of Ottawa dataset tasks and their transfer learning accuracy is defined as ρ_6 . According to (11), it can be calculated that $\rho_1 = -0.9523$, $\rho_2 = -0.9044$, $\rho_3 = -0.8456$, $\rho_4 = -0.5045$, $\rho_5 = -0.9574$ and $\rho_6 = -0.9549$.

It is obvious that there exists a strong negative correlation. Therefore, it can be summarized that the more similar the working conditions of the source domain data and the target domain data are, the lower the MK-MMD and MMD between them are, and the better the transfer learning results will be.

5.4. Hyperparameter Optimization

The IATN has two main hyperparameters, the inverse gradient coefficient β and the transfer learning coefficient γ . To study the impact of their values on model performance, this paper adjusts k and θ to find the optimal values. Here, this paper sets β and γ to different values on D1→D2 and D2→D3 tasks and conducts 10 repeated experiments to obtain their average accuracy. The results are shown in Figure 7. In Figure 7, the abscissa of each graph represents the value of the hyperparameter, the ordinate represents the accuracy, the blue line represents the experimental results of D1→D2 and the red line represents the experimental results of D2→D3.

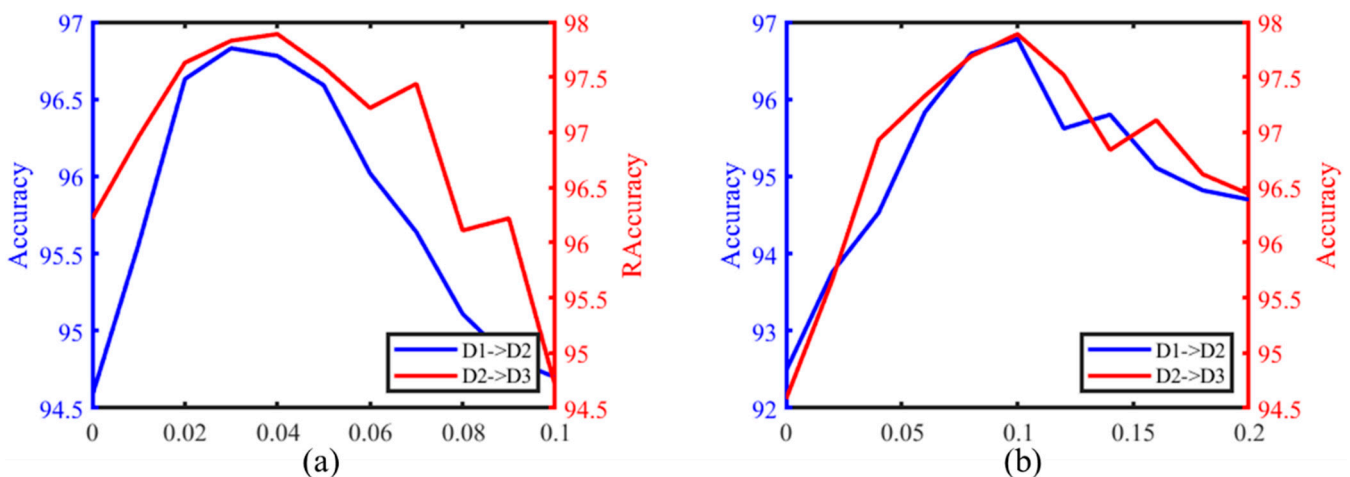


Figure 7. The testing accuracy of β and γ on D1→D2 and D2→D3 tasks. (a) The impact of different values of β on D1→D2 and D2→D3 tasks; (b) the impact of different values of γ on D1→D2 and D2→D3 tasks.

In Figure 6a, it is obvious that the optimal parameters are different for different tasks. For the D1→D2 task, the best β is 0.03; for the D2→D3 task, the best β is 0.04. For the D1→D2 task, the best β is 0.03; for the D2→D3 task, the best β is 0.04. In Figure 6b, the optimal γ for both tasks is 0.1.

In summary, this paper chooses $\beta = 0.04$ and $\gamma = 0.1$ as the experimental parameters.

6. Conclusions

This paper proposes an IATN for bearing fault diagnosis under variable WDs. Specifically, this paper combines adversarial transfer networks with an STFT to obtain satisfactory results with lighter networks. Then, this paper employs a channel attention module to enhance feature fusion, allowing the model to emphasize useful features and suppress less-useful features selectively. Moreover, this paper designs a novel domain discrepancy hybrid metric loss by combining MK-MMD and FCD. FCD promotes local feature alignment, and MMD is used for global feature alignment. The integration of these two loss functions combines local and global feature alignment. Finally, this paper verifies the method's effectiveness on three datasets, a private dual-rotor dataset, a CWRU dataset and

the Ottawa dataset. The IATN achieves an average accuracy that is significantly higher than those of the other six methods, CNN, DANN, AMDSA, DASN, ISAE-CSDF and SA-SN-DCGAN. The results indicate that the IATN has strong domain-invariant feature extraction capabilities.

In the future, we will pay more attention to the interpretability of the transfer process and study what knowledge is transferred during the transfer learning process.

Author Contributions: J.W. contributed to conceptualization, formal analysis, investigation, methodology and writing—original draft; H.A. contributed to methodology, validation and writing—review and editing; X.C. contributed to funding acquisition, project administration and writing—review and editing; R.Y. contributed to funding acquisition, project administration and data curation; and A.K.N. contributed to supervision, validation and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Natural Science Foundation of China (No. 52175116), Major Research Programs of the Natural Science Foundation of China (No. 92060302), the Research Foundation of the Higher Educational Key Laboratory for Flexible Manufacturing Equipment Integration of Fujian Province, the Xiamen Institute of Technology, the National Key Science and Technology Infrastructure Opening Project Fund for Research and Evaluation facilities for Service Safety of Major Engineering Materials and the Aeronautical Science Foundation (No. 2019ZB070001). Also, this work was supported in part by the Royal Society award (number IEC\NSFC\223294) to Asoke K. Nandi. Jun Wang acknowledges the financial support from the Innovative Leading Talents Scholarship and Brunel University London.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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

References

- Ahmed, H.; Nandi, A.K. *Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines*; John Wiley & Sons: Chichester, UK, 2020; ISBN 978-1-119-54462-3.
- Desnica, E.; Ašonja, A.; Radovanović, L.; Palinkaš, I.; Kiss, I. Selection, Dimensioning and Maintenance of Roller Bearings. In Proceedings of the 31st International Conference on Organization and Technology of Maintenance (OTO 2022), Osijek, Croatia, 12 December 2022; Blažević, D., Ademović, N., Barić, T., Cumin, J., Desnica, E., Eds.; Lecture Notes in Networks and Systems. Springer: Cham, Switzerland, 2022; Volume 592, pp. 210–227. [\[CrossRef\]](#)
- Liu, S.; Chen, S.; Chen, Z.; Gong, Y. Fault Diagnosis Strategy Based on BOA-ResNet18 Method for Motor Bearing Signals with Simulated Hydrogen Refueling Station Operating Noise. *Appl. Sci.* **2024**, *14*, 157. [\[CrossRef\]](#)
- Patil, A.A.; Desai, S.S.; Patil, L.N.; Patil, S.A. Adopting Artificial Neural Network for Wear Investigation of Ball Bearing Materials Under Pure Sliding Condition. *Appl. Eng. Lett.* **2022**, *7*, 81–88. [\[CrossRef\]](#)
- Xu, Z.; Ji, F.; Ding, S.; Zhao, Y.; Zhou, Y.; Zhang, Q.; Du, F. Digital Twin-Driven Optimization of Gas Exchange System of 2-Stroke Heavy Fuel Aircraft Engine. *J. Manuf. Syst.* **2021**, *58*, 132–145. [\[CrossRef\]](#)
- Mu, X.; Wang, Y.; Yuan, B.; Sun, W.; Liu, C.; Sun, Q. A New Assembly Precision Prediction Method of Aeroengine High-Pressure Rotor System Considering Manufacturing Error and Deformation of Parts. *J. Manuf. Syst.* **2021**, *61*, 112–124. [\[CrossRef\]](#)
- He, Y.; Hu, M.; Feng, K.; Jiang, Z. An Intelligent Fault Diagnosis Scheme Using Transferred Samples for Intershaft Bearings under Variable Working Conditions. *IEEE Access* **2020**, *8*, 203058–203069. [\[CrossRef\]](#)
- Berghout, T.; Benbouzid, M. Diagnosis and Prognosis of Faults in High-Speed Aeronautical Bearings with a Collaborative Selection Incremental Deep Transfer Learning Approach. *Appl. Sci.* **2023**, *13*, 10916. [\[CrossRef\]](#)
- Ahmed, H.O.A.; Nandi, A.K. Connected Components-Based Colour Image Representations of Vibrations for a Two-Stage Fault Diagnosis of Roller Bearings Using Convolutional Neural Networks. *Chin. J. Mech. Eng.* **2021**, *34*, 37. [\[CrossRef\]](#)
- Ahmed, H.; Nandi, A.K. Three-Stage Hybrid Fault Diagnosis for Rolling Bearings with Compressively Sampled Data and Subspace Learning Techniques. *IEEE Trans. Ind. Electron.* **2019**, *66*, 5516–5524. [\[CrossRef\]](#)
- Zuo, L.; Zhang, L.; Zhang, Z.-H.; Luo, X.-L.; Liu, Y. A Spiking Neural Network-Based Approach to Bearing Fault Diagnosis. *J. Manuf. Syst.* **2021**, *61*, 714–724. [\[CrossRef\]](#)
- Li, Y.; Wang, S.; Deng, Z. Intelligent Fault Identification of Rotary Machinery Using Refined Composite Multi-Scale Lempel-Ziv Complexity. *J. Manuf. Syst.* **2020**, *61*, 725–735. [\[CrossRef\]](#)
- Ahmed, H.; Nandi, A.K. Compressive Sampling and Feature Ranking Framework for Bearing Fault Classification with Vibration Signals. *IEEE Access* **2018**, *6*, 44731–44746. [\[CrossRef\]](#)

14. Wang, Z.; Liang, B.; Guo, J.; Wang, L.; Tan, Y.; Li, X. Fault Diagnosis Based on Residual–Knowledge–Data Jointly Driven Method for Chillers. *Eng. Appl. Artif. Intell.* **2023**, *125*, 106768. [[CrossRef](#)]
15. Pan, J.; Zi, Y.; Chen, J.; Zhou, Z.; Wang, B. LiftingNet: A Novel Deep Learning Network with Layerwise Feature Learning from Noisy Mechanical Data for Fault Classification. *IEEE Trans. Ind. Electron.* **2018**, *65*, 4973–4982. [[CrossRef](#)]
16. Wen, L.; Li, X.; Gao, L.; Zhang, Y. A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method. *IEEE Trans. Ind. Electron.* **2018**, *65*, 5990–5998. [[CrossRef](#)]
17. Ahmed, H.O.A.; Wong, M.L.D.; Nandi, A.K. Intelligent Condition Monitoring Method for Bearing Faults from Highly Compressed Measurements Using Sparse Over-Complete Features. *Mech. Syst. Signal Process.* **2018**, *99*, 459–477. [[CrossRef](#)]
18. Jia, L.; Chow, T.W.S.; Yuan, Y. GTFE-Net: A Gramian Time Frequency Enhancement CNN for Bearing Fault Diagnosis. *Eng. Appl. Artif. Intell.* **2023**, *119*, 105794. [[CrossRef](#)]
19. Zhao, K.; Xiao, J.; Li, C.; Xu, Z.; Yue, M. Fault Diagnosis of Rolling Bearing Using CNN and PCA Fractal-Based Feature Extraction. *Measurement* **2023**, *223*, 113754. [[CrossRef](#)]
20. Ruan, D.; Wang, J.; Yan, J.; Gühmann, C. CNN Parameter Design Based on Fault Signal Analysis and Its Application in Bearing Fault Diagnosis. *Adv. Eng. Inform.* **2023**, *55*, 101877. [[CrossRef](#)]
21. Wang, J.; Li, T.; Sun, C.; Yan, R.; Chen, X. Improved Spiking Neural Network for Intershaft Bearing Fault Diagnosis. *J. Manuf. Syst.* **2022**, *65*, 208–219. [[CrossRef](#)]
22. Sinitsin, V.; Ibryaeva, O.; Sakovskaya, V.; Ereemeeva, V. Intelligent Bearing Fault Diagnosis Method Combining Mixed Input and Hybrid CNN-MLP Model. *Mech. Syst. Signal Process.* **2022**, *180*, 109454. [[CrossRef](#)]
23. Zhiyi, H.; Haidong, S.; Xiang, Z.; Yu, Y.; Junsheng, C. An Intelligent Fault Diagnosis Method for Rotor-Bearing System Using Small Labeled Infrared Thermal Images and Enhanced CNN Transferred from CAE. *Adv. Eng. Inform.* **2020**, *46*, 101150. [[CrossRef](#)]
24. Huo, C.; Jiang, Q.; Shen, Y.; Qian, C.; Zhang, Q. New Transfer Learning Fault Diagnosis Method of Rolling Bearing Based on ADC-CNN and LATL under Variable Conditions. *Measurement* **2022**, *188*, 110587. [[CrossRef](#)]
25. Zheng, B.; Huang, J.; Ma, X.; Zhang, X.; Zhang, Q. An Unsupervised Transfer Learning Method Based on SOCNN and FBNN and Its Application on Bearing Fault Diagnosis. *Mech. Syst. Signal Process.* **2024**, *208*, 111047. [[CrossRef](#)]
26. Li, C.; Zhang, S.; Qin, Y.; Estupinan, E. A Systematic Review of Deep Transfer Learning for Machinery Fault Diagnosis. *Neurocomputing* **2020**, *407*, 121–135. [[CrossRef](#)]
27. Liu, Z.; Su, S.; Cao, J.; Yu, P. Transfer Learning for Bearing Fault Diagnosis Based on Improved Residual Network. In Proceedings of the 2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA), Changchun, China, 24–26 February 2023; pp. 1214–1216. [[CrossRef](#)]
28. Liu, Y.; Wang, Y.; Chow, T.W.S.; Li, B. Deep Adversarial Subdomain Adaptation Network for Intelligent Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2022**, *18*, 6038–6046. [[CrossRef](#)]
29. Öztürk, C.; Taşyürek, M.; Türkdamar, M.U. Transfer Learning and Fine-Tuned Transfer Learning Methods’ Effectiveness Analyse in the CNN-Based Deep Learning Models. *Concurr. Comput. Pract. Exp.* **2023**, *35*, e7542. [[CrossRef](#)]
30. Nguyen, C.T.; Van Huynh, N.; Chu, N.H.; Saputra, Y.M.; Hoang, D.T.; Nguyen, D.N.; Pham, Q.-V.; Niyato, D.; Dutkiewicz, E.; Hwang, W.-J. Transfer Learning for Wireless Networks: A Comprehensive Survey. *Proc. IEEE* **2022**, *110*, 1073–1115. [[CrossRef](#)]
31. Chen, K.-F.; Lee, M.-C.; Lin, C.-H.; Yeh, W.-C.; Lee, T.-S. Multi-Fault and Severity Diagnosis for Self-Organizing Networks Using Deep Supervised Learning and Unsupervised Transfer Learning. *IEEE Trans. Wirel. Commun.* **2024**, *23*, 141–157. [[CrossRef](#)]
32. Azari, M.S.; Flammini, F.; Santini, S.; Caporuscio, M. A Systematic Literature Review on Transfer Learning for Predictive Maintenance in Industry 4.0. *IEEE Access* **2023**, *11*, 12887–12910. [[CrossRef](#)]
33. Bu, K.; He, Y.; Jing, X.; Han, J. Adversarial Transfer Learning for Deep Learning Based Automatic Modulation Classification. *IEEE Signal Process. Lett.* **2020**, *27*, 880–884. [[CrossRef](#)]
34. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *Proc. IEEE* **2021**, *109*, 43–76. [[CrossRef](#)]
35. Liu, D.; Zhang, J.; Wu, H.; Liu, S.; Long, J. Multi-Source Transfer Learning for EEG Classification Based on Domain Adversarial Neural Network. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2023**, *31*, 218–228. [[CrossRef](#)]
36. Wen, L.; Gao, L.; Li, X. A New Deep Transfer Learning Based on Sparse Auto-Encoder for Fault Diagnosis. *IEEE Trans. Syst. Man Cybern. Syst.* **2019**, *49*, 136–144. [[CrossRef](#)]
37. Pan, S.J.; Kwok, J.T.; Yang, Q. Transfer Learning via Dimensionality Reduction. In Proceedings of the 23rd National Conference on Artificial Intelligence—Volume 2, Chicago, IL, USA, 13–17 July 2008; AAAI Press: Washington, DC, USA, 2008; pp. 677–682.
38. Huo, C.; Jiang, Q.; Shen, Y.; Lin, X.; Zhu, Q.; Zhang, Q. A Class-Level Matching Unsupervised Transfer Learning Network for Rolling Bearing Fault Diagnosis under Various Working Conditions. *Appl. Soft Comput.* **2023**, *146*, 110739. [[CrossRef](#)]
39. Wang, Y.; Sun, X.; Li, J.; Yang, Y. Intelligent Fault Diagnosis with Deep Adversarial Domain Adaptation. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–9. [[CrossRef](#)]
40. Cui, Z.; Cao, H.; Ai, Z.; Wang, J. A Multi-Adversarial Joint Distribution Adaptation Method for Bearing Fault Diagnosis under Variable Working Conditions. *Appl. Sci.* **2023**, *13*, 10606. [[CrossRef](#)]
41. Guo, Y.; Zhou, X.; Li, J.; Ba, R.; Xu, Z.; Tu, S.; Chai, L. A Novel and Optimized Sine–Cosine Transform Wavelet Threshold Denoising Method Based on the sym4 Basis Function and Adaptive Threshold Related to Noise Intensity. *Appl. Sci.* **2023**, *13*, 10789. [[CrossRef](#)]

42. Case Western Reserve University. Bearing Data Center. Available online: <https://engineering.case.edu/bearingdatacenter/apparatus-and-procedures> (accessed on 25 January 2024).
43. Huang, H.; Baddour, N. Bearing Vibration Data Collected under Time-Varying Rotational Speed Conditions. *Data Brief* **2018**, *21*, 1745–1749. [[CrossRef](#)] [[PubMed](#)]
44. Chen, J.; Huang, R.; Zhao, K.; Wang, W.; Liu, L.; Li, W. Multiscale Convolutional Neural Network With Feature Alignment for Bearing Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–10. [[CrossRef](#)]
45. Hu, Q.; Si, X.; Qin, A.; Lv, Y.; Liu, M. Balanced Adaptation Regularization Based Transfer Learning for Unsupervised Cross-Domain Fault Diagnosis. *IEEE Sens. J.* **2022**, *22*, 12139–12151. [[CrossRef](#)]
46. Guo, L.; Lei, Y.; Xing, S.; Yan, T.; Li, N. Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines with Unlabeled Data. *IEEE Trans. Ind. Electron.* **2019**, *66*, 7316–7325. [[CrossRef](#)]
47. Saha, D.; Hoque, M.E.; Chowdhury, M.E.H. Enhancing Bearing Fault Diagnosis Using Transfer Learning and Random Forest Classification: A Comparative Study on Variable Working Conditions. *IEEE Access* **2024**, *12*, 5986–6000. [[CrossRef](#)]
48. Wang, H.; Li, P.; Lang, X.; Tao, D.; Ma, J.; Li, X. FTGAN: A Novel GAN-Based Data Augmentation Method Coupled Time-Frequency Domain for Imbalanced Bearing Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–14. [[CrossRef](#)]
49. Shi, J.; Wang, X.; Lu, S.; Zheng, J.; Dong, H.; Zhang, J. An Adversarial Multisource Data Subdomain Adaptation Model: A Promising Tool for Fault Diagnosis of Induction Motor Under Cross-Operating Conditions. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–14. [[CrossRef](#)]
50. 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.* **2021**, *32*, 1713–1722. [[CrossRef](#)] [[PubMed](#)]
51. Meng, Z.; Zhu, J.; Cao, S.; Li, P.; Xu, C. Bearing Fault Diagnosis Under Multisensor Fusion Based on Modal Analysis and Graph Attention Network. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–10. [[CrossRef](#)]
52. Tong, J.; Liu, C.; Bao, J.; Pan, H.; Zheng, J. A Novel Ensemble Learning-Based Multisensor Information Fusion Method for Rolling Bearing Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–12. [[CrossRef](#)]
53. Chen, Z.; Qin, W.; He, G.; Li, J.; Huang, R.; Jin, G.; Li, W. Explainable Deep Ensemble Model for Bearing Fault Diagnosis under Variable Conditions. *IEEE Sens. J.* **2023**, *23*, 17737–17750. [[CrossRef](#)]
54. Sun, S.; Yeh, C.-F.; Hwang, M.-Y.; Ostendorf, M.; Xie, L. Domain Adversarial Training for Accented Speech Recognition. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 4854–4858. [[CrossRef](#)]
55. Li, Y.; Wu, H.; Kang, Y.; Guo, Y.; Cui, Z.; Xing, J.; Wang, Q.; Meng, J. Channel Discrepancies Adaptive Modulation Recognition Using Domain Adversarial Training. In Proceedings of the 2021 Asia-Pacific International Symposium on Electromagnetic Compatibility (APEMC), Nusa Dua, Bali, Indonesia, 27–30 September 2021; pp. 1–4. [[CrossRef](#)]
56. Ma, X.; Mou, X.; Wang, J.; Liu, X.; Geng, J.; Wang, H. Cross-Dataset Hyperspectral Image Classification Based on Adversarial Domain Adaptation. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 4179–4190. [[CrossRef](#)]
57. Qin, R.; Lu, C. Research on Measurement Methods of Transferability between Different Domains in Transfer Learning. In Proceedings of the 2019 CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS), Xiamen, China, 5–7 July 2019; pp. 926–931. [[CrossRef](#)]
58. Zhang, Q.; Deng, L. An Intelligent Fault Diagnosis Method of Rolling Bearings Based on Short-Time Fourier Transform and Convolutional Neural Network. *J. Fail. Anal. Prev.* **2023**, *23*, 795–811. [[CrossRef](#)]
59. Wang, X.; Ying, T.; Tian, W. Spectrum Representation Based on STFT. In Proceedings of the 2020 13th International Congress on Image and Signal Processing, BioMedical Engineering, and Informatics (CISP-BMEI), Chengdu, China, 17–19 October 2020; pp. 435–438. [[CrossRef](#)]
60. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [[CrossRef](#)] [[PubMed](#)]
61. Zhang, W.; Li, C.; Peng, G.; Chen, Y.; Zhang, Z. A Deep Convolutional Neural Network with New Training Methods for Bearing Fault Diagnosis under Noisy Environment and Different Working Load. *Mech. Syst. Signal Process.* **2018**, *100*, 439–453. [[CrossRef](#)]
62. Di, Y.; Yang, R.; Huang, M. Fault Diagnosis of Rotating Machinery Based on Domain Adversarial Training of Neural Networks. In Proceedings of the 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE), Kyoto, Japan, 20–23 June 2021; pp. 1–6. [[CrossRef](#)]
63. Luo, S.; Huang, X.; Wang, Y.; Luo, R.; Zhou, Q. Transfer Learning Based on Improved Stacked Autoencoder for Bearing Fault Diagnosis. *Knowl.-Based Syst.* **2022**, *256*, 109846. [[CrossRef](#)]
64. Wu, Z.; Jiang, H.; Liu, S.; Liu, Y.; Yang, W. Conditional Distribution-Guided Adversarial Transfer Learning Network with Multi-Source Domains for Rolling Bearing Fault Diagnosis. *Adv. Eng. Inform.* **2023**, *56*, 101993. [[CrossRef](#)]
65. Senanayake, D.A.; Wang, W.; Naik, S.H.; Halgamuge, S. Selforganizing Nebulous Growths for Robust and Incremental Data Visualization. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 4588–4602. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.