


Article

Fusion Domain-Adaptation CNN Driven by Images and Vibration Signals for Fault Diagnosis of Gearbox Cross-Working Conditions

Gang Mao, Zhongzheng Zhang, Bin Qiao and Yongbo Li * 

MITT Key Laboratory of Dynamics and Control of Complex System, School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China; mg0207@yeah.net (G.M.); zhangzhongzheng@mail.nwpu.edu.cn (Z.Z.); qiaobin@mail.nwpu.edu.cn (B.Q.)

* Correspondence: yongbo@nwpu.edu.cn

Abstract: The vibration signal of gearboxes contains abundant fault information, which can be used for condition monitoring. However, vibration signal is ineffective for some non-structural failures. In order to resolve this dilemma, infrared thermal images are introduced to combine with vibration signals via fusion domain-adaptation convolutional neural network (FDACNN), which can diagnose both structural and non-structural failures under various working conditions. First, the measured raw signals are converted into frequency and squared envelope spectrum to characterize the health states of the gearbox. Second, the sequences of the frequency and squared envelope spectrum are arranged into two-dimensional format, which are combined with infrared thermal images to form fusion data. Finally, the adversarial network is introduced to realize the state recognition of structural and non-structural faults in the unlabeled target domain. An experiment of gearbox test rigs was used for effectiveness validation by measuring both vibration and infrared thermal images. The results suggest that the proposed FDACNN method performs best in cross-domain fault diagnosis of gearboxes via multi-source heterogeneous data compared with the other four methods.

Keywords: deep learning; fault diagnosis; multi-source heterogeneous fusion; gearbox; transfer learning



Citation: Mao, G.; Zhang, Z.; Qiao, B.; Li, Y. Fusion Domain-Adaptation CNN Driven by Images and Vibration Signals for Fault Diagnosis of Gearbox Cross-Working Conditions. *Entropy* **2022**, *24*, 119. <https://doi.org/10.3390/e24010119>

Academic Editor: Daniel Abasolo

Received: 15 November 2021

Accepted: 10 January 2022

Published: 13 January 2022

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



Copyright: © 2022 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

The gearboxes play an irreplaceable role in the mechanical power system, which usually works in harsh and complex environments [1,2]. The failure of gearboxes may cause unexpected accidents and economic losses. Therefore, accurately identifying and diagnosing faults is the key to ensuring that gearboxes normally operate [3].

The intelligent diagnosis method has received wide attention from researchers because of its ability to detect faults automatically and is not limited by manual experience. The methods based on deep learning are particularly prominent because they can adaptively learn the fault information hidden in the collected signals, such as long short-term memory network (LSTM) [4], recurrent neural network (RNN) [5] and convolutional neural network (CNN) [6]. In addition, some extended models based on standard deep learning models are proposed for rotating machinery fault diagnosis, such as deep convolutional auto-encoder (DCAE) [7], CNN with capsule network [8] and multiscale CNN [9], etc. Yao et al. [10] proposed a stacked inverted residual CNN (SIRCNN), which had stable and reliable fault diagnosis accuracy. Shao et al. [11] established an ensemble deep auto-encoder (EDAE), which consists of several DAEs with different activation functions. The results indicated that it has good accuracy in rolling bearing fault diagnosis. However, the methods mentioned above assume that adequate high-quality data collected from the concerned machine are available for estimating underlying data distributions. In addition, these methods need training and testing data drawn from the same probability distribution [12]. In actual applications, it is impractical to obtain a large amount of labeled data.

In addition, the performance of the aforementioned methods may decrease in recognizing unlabeled data collected from another machine or different working conditions due to data discrepancy [13].

Transfer learning can transfer learned knowledge to related machinery or fields, and it is widely applied to address the above-mentioned cross-domain fault diagnosis problem. Currently, parameter transfer and feature domain-adaptation are two popular transfer learning implementation methods. Parameter transfer is suitable for scenarios where only a few labeled data from the target domain are available, but it is not enough to train the model. Qian et al. [14] proposed a method for rolling bearing fault diagnosis under variable working conditions by transferring the parameters of the stack auto-encoder (SAE). Chen et al. [15] proposed the transfer neural network to diagnose the faults of the rotary machinery, which pre-trains a 1D-CNN with the source data and then uses the limited target data to fine-tune the model to obtain a transfer convolutional neural network. However, in most practical applications, there is no available labeled target data to participate in the model training process. Domain adaptation techniques based on feature transfer have been much preferred in this case. One implementation of domain adaptation is to add a domain adaptation term to the loss function, such as Maximum Mean Discrepancy (MMD) [16–18] and Wasserstein distance [19]. Another implementation of domain adaptation is through domain adversarial training, in which a feature extractor aims to extract common features from both source and target domain by adversarial training [20–22]. In addition to this, in order to further improve transfer and generalization capabilities of the models, multiple source domains of data are used to extract transferable features, which are used to diagnose the faults of rotating machinery [23–25].

However, most existing studies on transfer diagnosis mainly focus on single-channel signals with vibration signals as the mainstay. This is because the vibration signal can be collected by the acceleration sensor attached to the surface of the component, which is sensitive to the impact caused by structural damage, such as gear fracture and bearing outer race crack. For some non-structural faults, such as gear box oil shortages, vibration signals are not sensitive to them. These failures can also cause serious consequences and should not be ignored. Infrared thermal image can perfectly reflect non-structural fault information and is widely applied in fault diagnosis [26,27]. However, the single infrared thermal image is very sensitive and is easily affected by external factors such as oil temperature [28]. Therefore, the fault diagnosis method based on multi-source heterogeneous data fusion is an issue worthy of study. Bai et al. [29] proposed a method for coupling fault diagnosis of rotary machinery by using infrared images and vibration signals, in which the enhanced infrared thermal image and two-dimensional vibration signals are spliced and inputted into CNN to obtain final diagnosis result. Shao et al. [30] pre-trained multiple novel SAEs using multisensory signals from the source domain and finetuned each novel SAE using a target domain sample. The diagnosis result is obtained by a modified voting strategy. In the above research studies, multi-source heterogeneous signals are widely applied in fault diagnosis since they can supply abundant fault information. However, it is rare to use infrared thermal images and vibration signals to diagnose structured and unstructured failure states in unlabeled target domains.

This paper proposed a fusion domain-adaptation CNN (FDACNN) driven by images and vibration signals. An FDACNN consists of two main stages: data-level fusion and domain-adaptation network training. In the stages of data-level fusion, raw signals are transformed into frequency and squared envelope spectrum, and they are arranged into two-dimensional format. Two-dimensional format data are combined with the infrared thermal image to form fusion data samples for model training. In the stages of domain-adaptation network training, a features extractor, a domain discriminator and a state classifier are constructed. After a number of adversarial training, the domain invariant features can be extracted from fusion samples and used for the classification of health states. In actual industrial gearbox, both the accelerators and infrared camera can be installed to collect the infrared images and vibration signals. The vibration signal can be used to

effectively diagnose structural failures such as tooth breakage, tooth missing, and gear wear. Moreover, the infrared thermal image is sensitive to non-structural failures, such as oil shortage and oil temperature exorbitant. In this study, more comprehensive features can be extracted from infrared thermal image and vibration signals than a single sensor. Moreover, the proposed method has lower calculation complexity, which can rely in the online fault diagnosis of gearbox. The main contributions and insights of this study are listed below:

- (1) A data-level multi-source heterogeneous fusion scheme is proposed. The frequency and squared envelope spectrum can more clearly reflect fault information contained in the vibration signal. The fusion of the preprocessed vibration signal and the infrared thermal image makes the fault information in the training sample more abundant and obvious.
- (2) A fusion domain-adaptation CNN fault diagnosis method for gearboxes is explored. It can extract domain invariant features from the fusion information of vibration signals and infrared thermal images and implement gearbox fault diagnosis in an unlabeled target domain.

The rest of this article is arranged as follows: Section 2 presents preliminary and basic knowledge. The details of the proposed FDACNN are provided in Section 3. Section 4 validates the proposed method and analyzes the results. Finally, the conclusion in Section 5 brings the study to a close.

2. Preliminaries

2.1. Squared Envelope Spectrum

Demodulation analysis methods can extract and identify fault characteristic frequencies from the resonance band. Envelope analysis is widely applied to acquire the harmonics of characteristic frequencies, such as the squared envelope spectrum (SES). Since the faulty rotating machinery vibration signals usually contain second-order cyclostationary (CS2) components, they are often used to extract fault features. CS2 is commonly calculated using SES, and the formulas can be expressed as follows:

$$SES(\alpha) = \left| \frac{1}{L} \sum_{n=0}^{L-1} |\tilde{x}(n)|^2 e^{-j2\pi n\alpha/F_s} \right|^2 = \left| DFT\left(|\tilde{x}(n)|^2\right) \right|^2 \quad (1)$$

where α represents cyclic frequency, and F_s denotes the sample frequency. $\tilde{x}(n)$ is converted by the Hilbert transform from time-domain vibration signals. $DFT(\bullet)$ represents the discrete Fourier transform, and it is formulated as follows:

$$DFT[x(n)] = \sum_{n=0}^{L-1} x(n) e^{-j \cdot k \cdot n \frac{2\pi}{L}} \quad (2)$$

where $x(n)$ represents the signal sequence $[0 \ L-1]$. Thus, CS2 components is acquired with cyclic frequency α .

2.2. Convolutional Neural Network

Convolutional neural network is a typical deep feed-forward artificial neural network that can be used to process time sequences and images by convolution operation. This operation can reduce the number of weights and biases to decrease the complexity of the model. A standard CNN consists of convolution layer, pooling layer, fully connected layer and classification layer. In a convolutional layer, multiple convolutional kernels are used to convolution the input, and the weights and bias are shared between hidden neurons. The process in the convolutional layer can be expressed as follows:

$$\mathbf{z}_n^l = f^l\left(\sum_k \mathbf{x}_k^{l-1} * \mathbf{w}_n^l + \mathbf{b}_n^l\right) \quad (3)$$

where x_k^{l-1} is the k -th input sample in the l -1 layer. $*$ represents convolution operation. w_n^l and b_n^l denote the weight and corresponding bias, respectively. Additionally, $f^l(\cdot)$ represents the activation function.

A pooling layer usually follows the convolutional layer, and the subsampling operation is employed to reduce the spatial dimension for reducing overfitting risk. Mathematically, a maximum pooling operation is defined as follows:

$$po_j = \max_{i \in m_j} \{c_j(i)\} \tag{4}$$

where c_j represent the j -th location, and po_j is the output of the pooling. Moreover, average pooling and stochastic pooling are also usually used in pooling layer.

After several convolutional and pooling layers, the fully connected layer immediately converts the output matrix into a row or column. The last layer is usually served by a softmax output layer in which a softmax function is utilized to predict the probability of each target.

2.3. Deep Adversarial Convolution Neural Network

Generally, a deep adversarial convolution neural network usually consists of a feature extractor G_f , a domain discriminator G_d and a classifier G_c [13]. The feature extractor, which is a competitor in the DACNN, is typically served by several convolution blocks or fully connected layers. It can be expressed as $G_f = G_f(x, \theta_f) : x \rightarrow R^D$ with parameter θ_f , which indicates that the input sample x is transformed into D -dimensional features. In addition, the domain discriminator (binary classifier) is treated as the opponent, which is expressed as $G_d = G_d(G_f(x), \theta_d)$ with parameters θ_d . Inputting the source and target samples into the feature extractor and the output is further distinguished by the domain discriminator G_d . The binary cross entropy (BCE) loss is taken as objective function, which can be described as follows:

$$L(G_d(G_f(x_i)), d_i) = d_i \log \frac{1}{G_d(G_f(x_i))} + (1 - d_i) \times \log \frac{1}{1 - G_d(G_f(x_i))} \tag{5}$$

where d_i denotes the binary variable for x_i . By conducting adversarial training between the two parts, feature extractor G_f tends to extract common features from two types of data and makes the domain discriminator difficult to distinguish in terms of zero or one. Thus, the model can perform well on both the source and target datasets. Assuming n samples in the source domain dataset and $N-n$ samples in the target domain dataset, the objective function is expressed as follows:

$$E(\theta_f, \theta_d) = - \left(\frac{1}{n} \sum_{i=1}^n L_d^i(\theta_f, \theta_d) + \frac{1}{N-n} \sum_{i=n+1}^N L_d^i(\theta_f, \theta_d) \right) \tag{6}$$

where $L_d^i(\theta_f, \theta_d) = L_d(G_d(G_f(x_i, \theta_f), \theta_d), d_i)$, and this equation includes a maximization problem with respect to θ_d and a minimization problem with respect to θ_f .

Additionally, all the labeled samples should be supervised and trained to ensure the accuracy of the diagnosis in the adversarial procedure. Therefore, a classifier is established, and it is expressed as $G_y = G_y(G_f(x), \theta_y) : R^D \rightarrow R^L$ with parameters θ_y , in which L is the number of classes. Cross-entropy loss is applied in the Softmax function, and it can be described as follows.

$$L_y(G_y(G_f(x_i)), y_i) = \log \frac{1}{G_y(G_f(x_i))_{y_i}} \tag{7}$$

By adding the Equation (7) to objective function (6), the optimization objective can be expressed as follows:

$$E(\theta_f, \theta_y, \theta_d) = \frac{1}{n} \sum_{i=1}^n L_y^i(\theta_f, \theta_y) - \lambda \left(\frac{1}{n} \sum_{i=1}^n L_d^i(\theta_f, \theta_d) \right) + \frac{1}{N-n} \sum_{i=n+1}^N L_d^i(\theta_f, \theta_d) \quad (8)$$

where $L_y^i(\theta_f, \theta_y) = L_y(G_y(G_f(x_i, \theta_f), \theta_y), y_i)$. The entire training process of DANN is to optimize the parameters θ_f , θ_y and θ_d , and it is expressed as follows.

$$(\hat{\theta}_f, \hat{\theta}_y) = \underset{\theta_f, \theta_y}{\operatorname{argmax}} E(\theta_f, \theta_y, \hat{\theta}_d) \quad (9)$$

$$\hat{\theta}_d = \underset{\theta_d}{\operatorname{argmax}} E(\hat{\theta}_f, \hat{\theta}_y, \theta_d) \quad (10)$$

In the training stage, parameter updates are implemented in the opposite direction to the gradient in the adversarial process. The update in the domain discriminator G_d is to reduce the loss with the purpose of improving the discriminative ability. However, the update in the feature extractor G_f is to maximize the loss to fool the discriminator. In order to frame a flexible implement of the stochastic gradient descent (SGD) algorithms in the training of DACNN, we use a circuitous method by rewriting the loss as $L_d^i(\theta_f, \theta_d) = L_d(G_d(G_f(x_i, \theta_f), \theta_d), 1 - d_i)$ in updating the feature extractor parameters. Maximizing $L_d^i(\theta_f, \theta_d)$ can be accomplished by minimizing $L_d^i(\theta_f, \theta_d)$. During backpropagation, the features extractor takes the gradient of the recalculated $L_d^i(\theta_f, \theta_d)$ from the domain discriminator and updates its parameters with SGD. Overall, the update rates of parameters θ_f , θ_y and θ_d can be formulated as follows:

$$\theta_f \leftarrow \theta_f - \mu \left(\frac{\partial L_y^i}{\partial \theta_f} + \lambda \frac{\partial L_d^i}{\partial \theta_f} \right) \quad (11)$$

$$\theta_d \leftarrow \theta_d - \mu \frac{\partial L_d^i}{\partial \theta_d} \quad (12)$$

$$\theta_y \leftarrow \theta_y - \mu \frac{\partial L_y^i}{\partial \theta_y} \quad (13)$$

where μ represents the learning rate.

By the above optimization process, DACNN tends to train a feature extractor G_f that can extract suitable representations from input samples (either source domain datasets or target domain datasets), which can be classified accurately by the classifier G_y but weakens the ability of the domain discriminator G_d to differentiate a sample from the source or target domain datasets. In the phases of testing, domain insensitive features will be extracted by feature descriptor G_f and fed into the classifier G_y to identify the states immediately.

3. The Proposed Method

3.1. Data-Level Fusion

Data-level fusion is a relatively direct fusion method that can retain the effective fault information hidden in the measured signals and reduce the complexity of the model. Therefore, a data-level fusion strategy is designed in this study to fuse one-dimensional vibration signals and two-dimensional infrared thermal images.

In general, the measured time-domain vibration signals are feeble and inadequate, particularly in the early stages of a failure. Although many deep learning methods only use raw time-domain signals for fault diagnosis, they rely heavily on deep learning structure. It is considered that frequency features play a significant role in rotating machinery failure diagnostics. The frequency domain signal and the squared envelope spectrum, in particular,

are widely used in classical signal processing methods [31,32]. Therefore, measured raw signals are transformed to acquire frequency domain signals by fast Fourier transform (FFT) and CS2 by squared envelope spectrum in this study. Then, they are reshaped into two-dimensional matrix as part of the input of the convolution layer as shown in Figure 1.

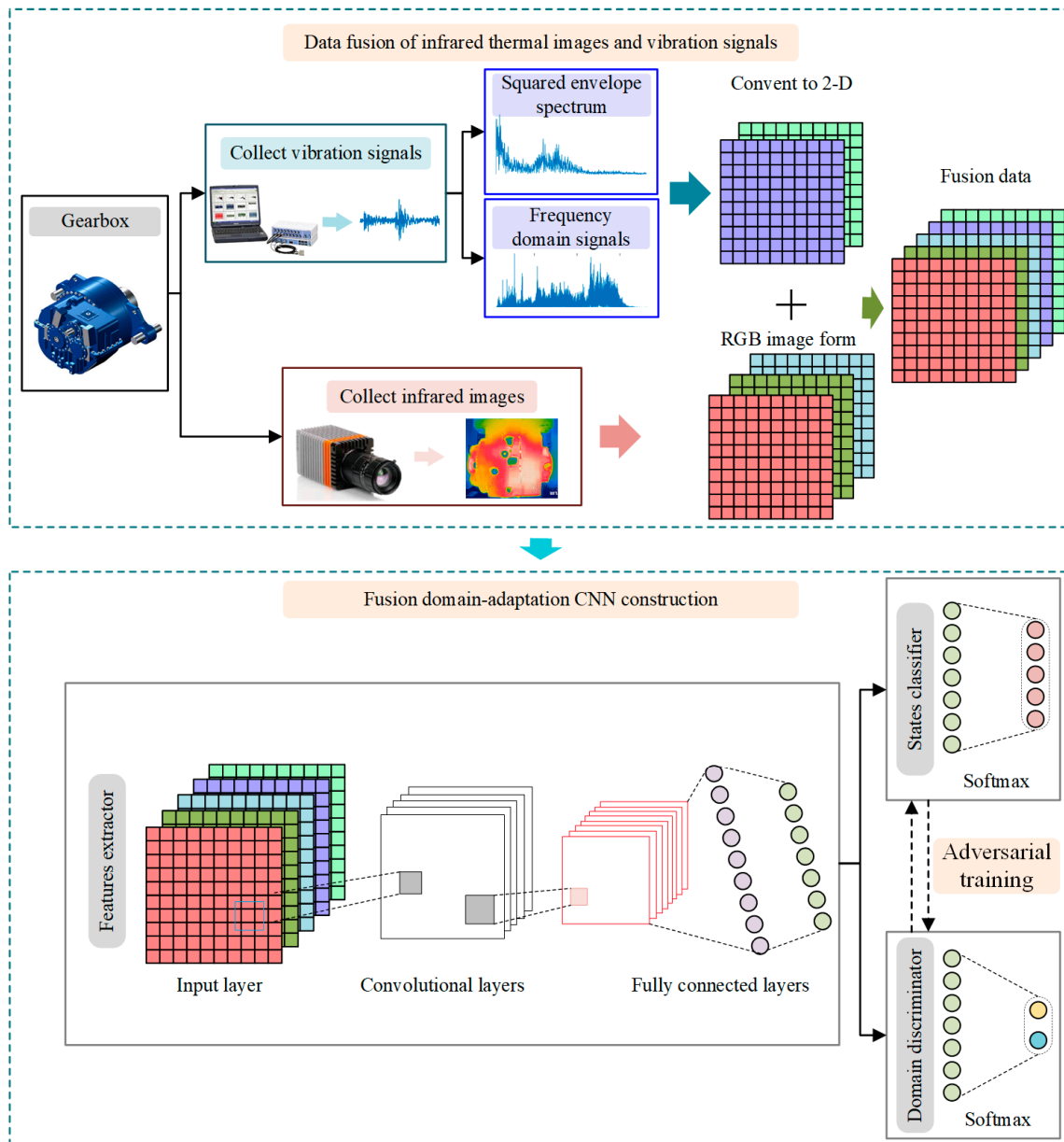


Figure 1. The procedures of the proposed method.

After that, the RGB 3-channels of each infrared thermal image are combined with the 2-dimensional frequency domain signal and the squared envelope spectrum to synthesize a 5-channel data. The 5-channel fusion data will be used for subsequent domain adaptation CNN training.

3.2. Fusion Domain-Adaptation CNN Construction

Domain adaptation based on adversarial networks is an effective approach for cross-domain fault diagnosis. In this section, fusion 5-channel data are used to train a domain-adaptation CNN for cross-domain fault diagnosis of gearboxes.

Firstly, a feature extractor is constructed, which contains multiple convolutional blocks and several fully connected layers. The feature extractor is used to extract domain-insensitive features from 5-channel fusion data. The extracted features are fed into the state classifier to recognize health states. Meanwhile, the extracted features are inputted into the domain discriminator to distinguish whether they are from the source or target domain.

In the training process, the feature extractor minimizes the state classification loss and maximizes the domain discrimination loss so that it can extract features that are not only not sensitive to the domain and make the state classifier easy to classify. As the opponent, the domain discriminator aims to minimize domain discrimination loss so that it can distinguish the feature from the source or target domain. Finally, through a large amount of adversarial training, the diagnosis model composed of feature extractor and state classifier can recognize fault states accurately with multi-source heterogeneous data under cross-working conditions.

3.3. Procedures of Proposed Fusion Domain-Adaptation CNN

This section presents the summaries of the proposed fusion domain-adaptation CNN as shown in Figure 1. The main procedures are described as follows.

Collect the infrared thermal images and raw vibration signals from the concerned gearboxes under different working conditions and divide them into labeled source domain samples and unlabeled target domain samples.

Convert the raw time-domain signals into frequency domain signals and the squared envelope spectrum and arrange them into matrixes.

Fuse the RGB 3-channels of infrared thermal image and two matrixes (frequency domain and squared envelope spectrum) to obtain 5-channel fusion samples.

Train the FDACNN model using 5-channel fusion samples by adversarial training.

Test the performance of the proposed FDACNN model by using the remaining samples from the target domain.

4. Experimental and Result Discussion

4.1. Dataset Descriptions

Test data are from a compound gear failure experiment performed on a helical gearbox called Spectra Quest Mechanical Failure Simulator (MFS) from Northwestern Polytechnical University lab [33,34]. The experiment system and the layout of the experiment rig are shown in Figure 2a,b, respectively. The experiment system mainly consists of an AC motor, two gearboxes and a generator. The infrared camera is fixed on the front of gearbox 1 to collect the infrared thermal image. The detailed parameters of the infrared camera are listed in Table 1. Vibration signals are collected by an acceleration sensor mounted on the surface of the gearbox 1. The sample frequency is 12.8 kHz, and the motor speed is 3000 rpm. In this experiment, there are two kinds of lubricating oil, i.e., EP 320 and EP 100. EP 320 lubricant viscosity is 320cSt @ 40 °C, EP 100 lubricant viscosity is 100cSt @ 40 °C. The EP 320 lubricant is applied in this article.

In this study, five different health states are introduced, including a normal state, two structural fault states (TB 50 and TB 100) and two non-structural fault states (OS 1500 and OS 2000). “TB 50” and “TB 100” refer to 50% and 100% tooth breakage in driving gear, respectively. Based on the baseline oil of 2600 mL, “OS 1500” and “OS 2000” refer to the reduction of 600 mL and 1100 mL of oil from GB1, respectively. Vibration signals and infrared images were collected under four different loads of 0%, 30%, 70% and 100% (L0, L30, L70 and L 100). For each load, vibration signals in each state were divided into 800 samples with 2048 data points. Four-hundred and eighty samples were randomly selected as tests, and the remaining 320 samples were used to train. Similarly, 480 infrared images were used to train, and the other 320 were used to test. The size of each infrared thermal image is 64×32 . The details of the dataset are listed in Table 2.

4.2. Implementation Details

At first, data-level fusion strategy is used to fuse infrared thermal images and vibration signals. The measured raw samples in five different health states are transformed to acquire frequency domain signals by FFT and CS2 by squared envelope spectrum in this study. Under L0 load, the three kinds of signal waveforms of different health states are shown in Figure 3. As shown in Figure 3, the left column is the measured raw signals, and the middle column and the right column are the corresponding spectral distribution and squared envelope spectrum, respectively. It can be observed that the time-domain characteristics and frequency domain characteristics of each health state are relatively similar and difficult to distinguish. Then, frequency domain signals and CS2 sequence are arranged into $2 \times 64 \times 32$ formats.

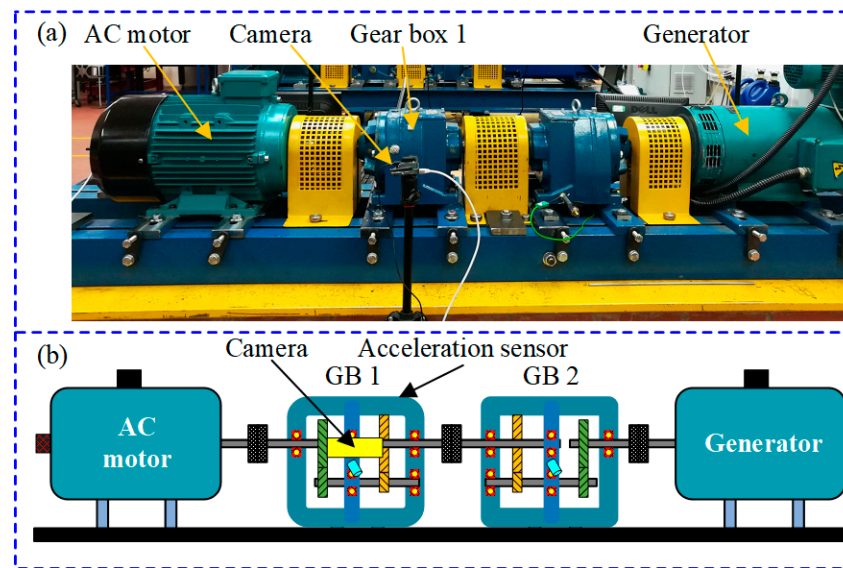


Figure 2. The gear box fault simulator system: (a) the experimental test rig; (b) the layout of the test rig.

Table 1. The detailed parameters of the infrared camera.

Parameters	Values
Alg type	PHE
Frame rate	25 fps
Temperate measurement range	$-25\text{ }^{\circ}\text{C}\sim 260\text{ }^{\circ}\text{C}$
Environment temperature	$18.9\text{ }^{\circ}\text{C}$
Thermal sensitivity	$0.050\text{ }^{\circ}\text{C}$
Image resolution	384×288
Contrast	50
Brightness	50
Gain	2
Palette	rainbow

Table 2. 5 health states of gearbox.

Label	Health States	The Number of Training/Testing Samples
1	Normal	480/320
2	TB 50	480/320
3	TB 100	480/320
4	OS 1500	480/320
5	OS 2000	480/320

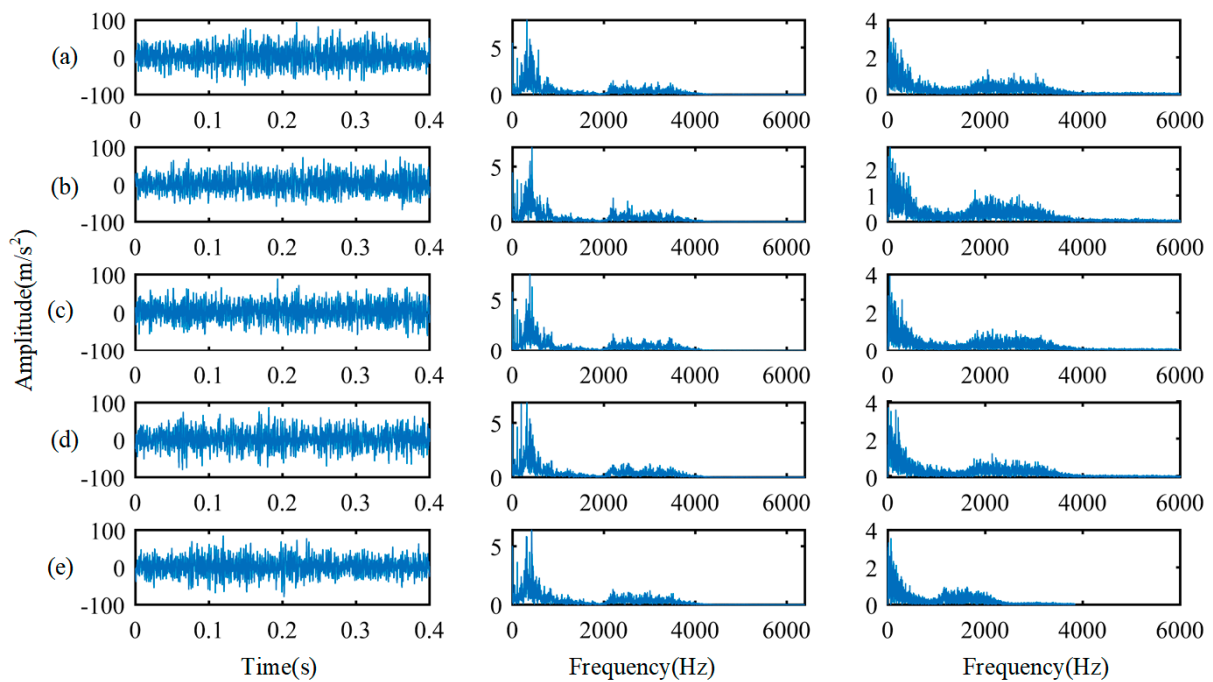


Figure 3. Raw signals, spectral distribution and squared envelope spectrum of different health states. (a) Normal; (b) TB 50; (c) TB 100; (d) OS 1500; (e) OS 2000.

Each infrared thermal image has RGB channels, i.e., $3 \times 64 \times 32$ formats. The collected infrared thermal images of each health state under L0 are shown in Figure 4. From Figure 4, we can observe that the images of normal, OS 1500 and OS 2000 are relatively similar, and the images of TB 50 and TB 100 are relatively similar. However, it is still very difficult to visually distinguish concrete health states. RGB channels of the infrared thermal image ($3 \times 64 \times 32$) will be combined with the frequency domain signals and CS2 ($2 \times 64 \times 32$), and the 5-channel fusion samples ($5 \times 64 \times 32$) are obtained to train DACNN.

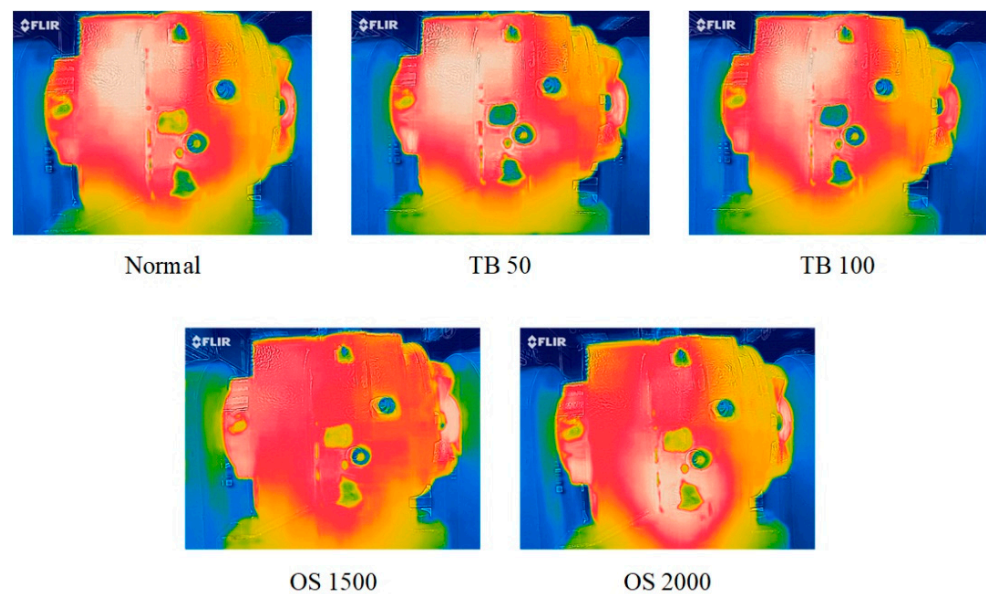


Figure 4. The infrared thermal image of different health states.

In DACNN, a feature extractor, a domain discriminator and a state classifier were constructed, and the structures of those are listed in Table 3. The DACNN is trained by adversarial training using fusion data. In order to illustrate the robustness of the proposed method, multiple test tasks are designed. The concrete setting of different tasks and the results are listed in Table 4. It can be observed that the proposed method has good performance among the five test tasks, especially the accuracy of reaching 100.00% in T1, T3 and T4. In task T5, the transfer span is larger from load L0 to L100. The accuracy of the proposed method can still reach 96.67%. This suggests that the fusion of infrared thermal images and vibration signals to implement cross-domain fault diagnosis has good performance, and the result is relatively robust.

Table 3. The structures of features extractor, domain discriminator and states classifier.

Model	Layer	Filter Number	Size of Kernel	Output Size	Stride	Padding	Active Function
Features extractor	Conv2d 1	8	3×3	$8 \times 62 \times 30$	[1,1]	0	ReLU
	BN 1	8	-	$8 \times 62 \times 30$	-	-	-
	MaxPool2d 1	8	2×2	$8 \times 31 \times 15$	[2,2]	-	-
	Conv2d 2	16	3×3	$16 \times 29 \times 13$	[1,1]	0	ReLU
	BN 2	16	-	$16 \times 29 \times 13$	-	-	-
	MaxPool2d 2	16	2×2	$16 \times 14 \times 6$	[2,2]	-	-
	FC 1	-	-	680	-	-	-
	FC 2	-	-	300	-	-	ReLU
	FC 3	-	-	56	-	-	ReLU
FC 4	-	-	28	-	-	ReLU	
States classifier	FC	-	-	5	-	-	Softmax
Domain discriminator	FC	-	-	1	-	-	Softmax

Table 4. Result of different test tasks.

Tasks	Source Domain	Target Domain	Accuracy (%)
T1	L0	L30	100.00%
T2	L0	L70	98.98%
T3	L30	L70	100.00%
T4	L70	L0	100.00%
T5	L0	L100	96.67%

4.3. Methods Comparison and Results Discussion

In this section, four methods are employed for comparison on the test tasks T1~T5 to illustrate the superiority of the proposed FDACNN. The details of compared methods are described as below.

The standard DANN method is utilized for comparisons [35]. Training data are sequences that convert fusion data into one dimension.

Another popular domain-adaptation method based on maximum mean discrepancy (DA-MMD) is applied for comparisons [36]. The train data and network structure used in DA-MMD are consistent with the proposed method.

The DACNN model was trained using single vibration signals (DACNN-SV). This model is trained using frequency domain signals and CS2 sequence ($2 \times 64 \times 32$ formats).

The DACNN model was trained using single infrared thermal image (DACNN-SI). This model is trained using RGB 3-channels of infrared thermal images ($3 \times 64 \times 32$ formats).

The results of different comparison methods are listed in Table 5. Meanwhile, in order to compare the results more intuitively, the bar diagram of results is shown in Figure 5. It can be observed that DACNN_SV has the lowest accuracy in five tasks: 47.38%, 40.88%, 67.44%, 35.56% and 35.44%, respectively. This is due to the fact that the vibration signal is not sensitive to two non-structural faults, which makes it impossible to classify, and the

resulting accuracy is low. DACNN_SI has a good performance on T1 and T3, but it is not satisfactory in other test tasks; in particular, the accuracy of T5 is only 83%. It illustrates that infrared thermal images can be used to identify structural and non-structural failure states effectively, but they are susceptible to environmental interference. From the results of DANN and DA-MMD, it can be observed that the accuracy in all test tasks is lower than the proposed FDACNN. This is because two-dimensional data fusion can effectively maintain the fault information contained in the infrared image and vibration signal. Meanwhile, the adversarial domain adaptation network can enable the extracted extractor to extract target domain features that are easy to distinguish.

Table 5. The results of different comparisons methods.

Tasks	DANN	DA-MMD	DACNN_SV	DACNN_SI	Proposed FDACNN
T1	98.98%	100.00%	47.38%	100%	100.00%
T2	92.96%	97.12%	40.88%	95.69%	98.98%
T3	97.38%	98.97%	67.44%	100%	100.00%
T4	80.00%	92.69%	35.56%	97.94%	100.00%
T5	60.00%	85%	35.44%	83%	96.67%

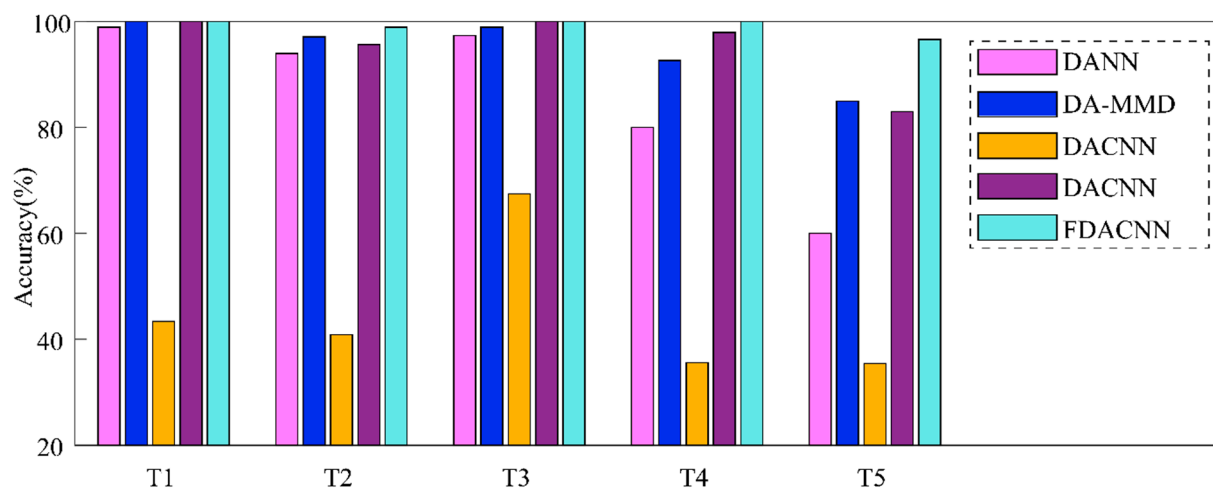


Figure 5. Bar diagram of results in different test tasks.

Additionally, in order to demonstrate the ability of the feature extractor to extract domain invariant features, principal analysis (PCA) is used to map extracted features into 2-dimensional space. Figure 6 shows 2-dimensional visualizations in different test tasks, in which PCA 1 and PCA 2 denote first and second principal components, respectively. It can be observed that the points with the same color are clustered in T1 and T3, and the point clusters of different colors are obviously isolated. In T2, T4 and T5, only a few points with the same color are confused, and most of them with the same color are relatively concentrated. Therefore, extracted features are relatively separable in all test tasks. It suggests that the trained feature extractor has the ability to extract distinguishable features from the unlabeled target domain fusion samples.

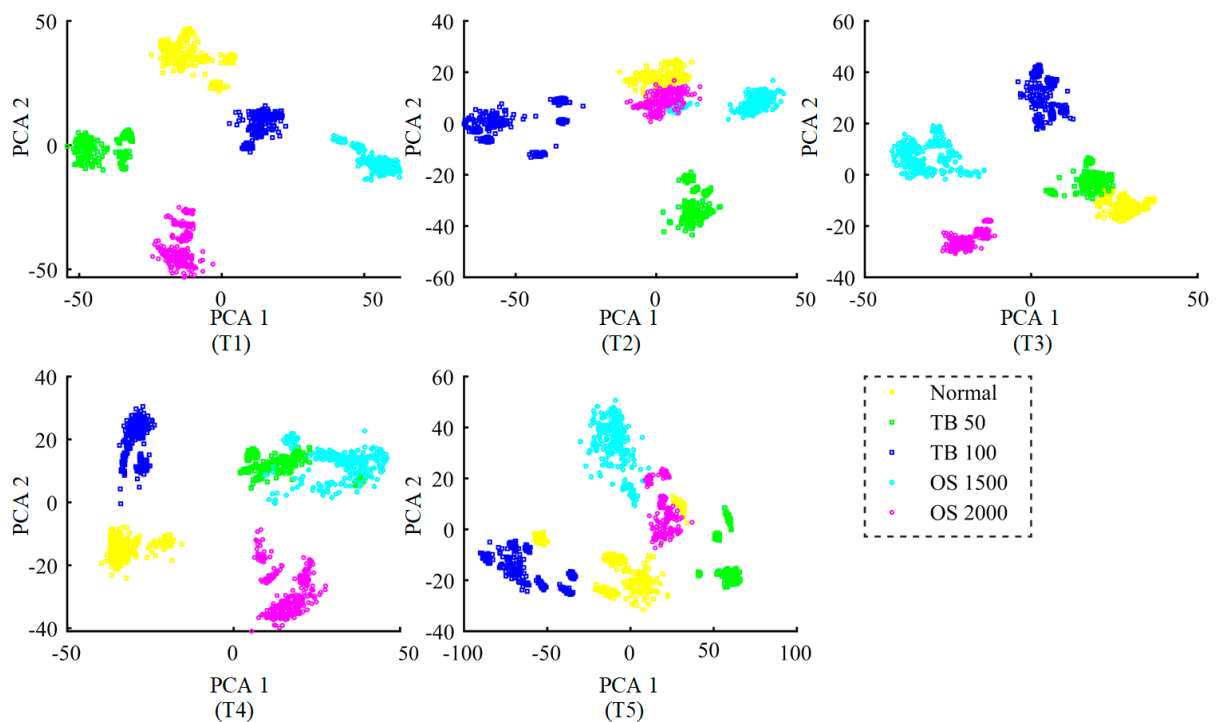


Figure 6. Feature visualization of different test tasks.

5. Conclusions

This study focuses on cross-domain fault diagnosis of gearboxes via multi-source heterogeneous data. Infrared thermal images and vibration signals are fused to characterize the health states of the gearbox, which can effectively recognize structural and non-structural faults. Moreover, the domain-adaptation neural network is trained via adversarial training using fusion data samples to extract the common transfer knowledge, called the FDACNN method. By performing this, the proposed FDACNN method can be used to recognize the unlabeled target domain samples of gearbox. For validation, the proposed FDACNN method is used to analyze gearbox multi-source heterogeneous data measured under various operating conditions. Moreover, we compare the FDACNN method with four other relevant methods to confirm its superiority in cross-domain fault diagnosis of gearboxes under various operating conditions. The results demonstrate that the proposed method obtains highest classification accuracy among four methods.

Author Contributions: Data curation, Z.Z.; Methodology, Y.L.; Writing—original draft, G.M.; Writing—review & editing, B.Q. All authors have read and agreed to the published version of the manuscript.

Funding: The research is supported by National Natural Science Foundation of China under Grant 51805434 and 12172290 and Key Laboratory of Equipment Research Foundation under Grant 6142003190208.

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 that they have no conflict of interest.

References

1. Kong, X.; Mao, G.; Wang, Q.; Ma, H.; Yang, W. A multi-ensemble method based on deep auto-encoders for fault diagnosis of rolling bearings. *Measurement* **2020**, *151*, 107132. [[CrossRef](#)]
2. Karabacak, Y.E.; Özmen, N.G.; Gümüsel, L. Intelligent worm gearbox fault diagnosis under various working conditions using vibration, sound and thermal features. *Appl. Acoust.* **2022**, *186*, 108463. [[CrossRef](#)]
3. Miao, Y.; Zhang, B.; Lin, J.; Zhao, M.; Liu, H.; Liu, Z.; Li, H. A review on the application of blind deconvolution in machinery fault diagnosis. *Mech. Syst. Signal Process.* **2022**, *163*, 108202. [[CrossRef](#)]
4. Shi, J.; Peng, D.; Peng, Z.; Zhang, Z.; Goebel, K.; Wu, D. Planetary gearbox fault diagnosis using bidirectional-convolutional LSTM networks. *Mech. Syst. Signal Process.* **2022**, *162*, 107996. [[CrossRef](#)]
5. Liu, H.; Zhou, J.; Zheng, Y.; Jiang, W.; Zhang, Y. Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Trans.* **2018**, *77*, 167–178. [[CrossRef](#)] [[PubMed](#)]
6. 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](#)]
7. Yu, J.; Liu, X. One-dimensional residual convolutional auto-encoder for fault detection in complex industrial processes. *Int. J. Prod. Res.* **2021**, *196*, 1–20. [[CrossRef](#)]
8. Li, D.; Zhang, M.; Kang, T.; Li, B.; Xiang, H.; Wang, K.; Pei, Z.; Tang, X.; Wang, P. Fault diagnosis of rotating machinery based on dual convolutional-capsule network (DC-CN). *Measurement* **2022**, *187*, 110258. [[CrossRef](#)]
9. Zhu, J.; Chen, N.; Peng, W. Estimation of bearing remaining useful life based on multiscale convolutional neural network. *IEEE Trans. Ind. Electron.* **2019**, *66*, 3208–3216. [[CrossRef](#)]
10. Yao, D.; Liu, H.; Yang, J.; Li, X. A lightweight neural network with strong robustness for bearing fault diagnosis. *Measurement* **2020**, *159*, 107756. [[CrossRef](#)]
11. Shao, H.; Jiang, H.; Lin, Y.; Li, X. A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders. *Mech. Syst. Signal Process.* **2018**, *102*, 278–297. [[CrossRef](#)]
12. 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](#)]
13. Han, T.; Liu, C.; Yang, W.; Jiang, D. A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults. *Knowl. Based Syst.* **2019**, *165*, 474–487. [[CrossRef](#)]
14. Qian, W.; Li, S.; Wang, J.; Xin, Y.; Ma, H. A New Deep Transfer Learning Network for Fault Diagnosis of Rotating Machine Under Variable Working Conditions. In Proceedings of the 2018 Prognostics and System Health Management Conference (PHM-Chongqing), Chongqing, China, 26–28 October 2018; pp. 1010–1016.
15. Chen, Z.; Gryllias, K.; Li, W. Intelligent Fault Diagnosis for Rotary Machinery Using Transferable Convolutional Neural Network. *IEEE Trans. Ind. Inform.* **2019**, *16*, 339–349. [[CrossRef](#)]
16. Yang, B.; Lei, Y.; Jia, F.; Xing, S. An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. *Mech. Syst. Signal Process.* **2019**, *122*, 692–706. [[CrossRef](#)]
17. Xu, K.; Li, S.; Wang, J.; An, Z.; Qian, W.; Ma, H. A novel convolutional transfer feature discrimination network for imbalanced fault diagnosis under variable rotational speed. *Meas. Sci. Technol.* **2019**, *30*, 105107. [[CrossRef](#)]
18. 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.* **2017**, *49*, 136–144. [[CrossRef](#)]
19. Zhang, M.; Wang, D.; Lu, W.; Yang, J.; Li, Z.; Liang, B. A Deep Transfer Model With Wasserstein Distance Guided Multi-Adversarial Networks for Bearing Fault Diagnosis under Different Working Conditions. *IEEE Access* **2019**, *7*, 65303–65318. [[CrossRef](#)]
20. Zhang, Z.; Li, X.; Wen, L.; Gao, L.; Gao, Y. Fault Diagnosis Using Unsupervised Transfer Learning Based on Adversarial Network. In Proceedings of the 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE), Vancouver, BC, Canada, 22–26 August 2019; pp. 305–310.
21. Zhang, B.; Li, W.; Hao, J.; Li, X.-L.; Zhang, M. Adversarial adaptive 1-D convolutional neural networks for bearing fault diagnosis under varying working condition. *arXiv* **2018**, arXiv:1805.00778.
22. Wang, B.; Shen, C.; Yu, C.; Yang, Y. Data Fused Motor Fault Identification Based on Adversarial Auto-Encoder. In Proceedings of the 2019 IEEE 10th International Symposium on Power Electronics for Distributed Generation Systems (PEDG), Xi'an, China, 3–6 June 2019; pp. 299–305.
23. Yang, S.; Kong, X.; Wang, Q.; Li, Z.; Cheng, H.; Yu, L. A multi-source ensemble domain adaptation method for rotary machine fault diagnosis. *Measurement* **2021**, *186*, 110213. [[CrossRef](#)]
24. He, Z.; Shao, H.; Zhong, X.; Zhao, X. Ensemble transfer CNNs driven by multi-channel signals for fault diagnosis of rotating machinery cross working conditions. *Knowl. Based Syst.* **2020**, *207*, 106396. [[CrossRef](#)]
25. Shi, Y.; Deng, A.; Ding, X.; Zhang, S.; Xu, S.; Li, J. Multisource domain factorization network for cross-domain fault diagnosis of rotating machinery: An unsupervised multisource domain adaptation method. *Mech. Syst. Signal Process.* **2022**, *164*, 108219. [[CrossRef](#)]
26. Xin, L.; Haidong, S.; Hongkai, J.; Jiawei, X. Modified Gaussian convolutional deep belief network and infrared thermal imaging for intelligent fault diagnosis of rotor-bearing system under time-varying speeds. *Struct. Health Monit.* **2021**, *99*, 8957. [[CrossRef](#)]

27. Choudhary, A.; Mian, T.; Fatima, S. Convolutional neural network based bearing fault diagnosis of rotating machine using thermal images. *Measurement* **2021**, *176*, 109196. [[CrossRef](#)]
28. Jia, Z.; Liu, Z.; Vong, C.-M.; Pecht, M. A rotating machinery fault diagnosis method based on feature learning of thermal images. *IEEE Access* **2019**, *7*, 12348–12359. [[CrossRef](#)]
29. Bai, T.; Yang, J.; Yao, D.; Wang, Y. Information Fusion of Infrared Images and Vibration Signals for Coupling Fault Diagnosis of Rotating Machinery. *Shock. Vib.* **2021**, *2021*, 6622041. [[CrossRef](#)]
30. Di, Z.; Shao, H.; Xiang, J. Ensemble deep transfer learning driven by multisensor signals for the fault diagnosis of bevel-gear cross-operation conditions. *Sci. China Technol. Sci.* **2021**, *64*, 481–492. [[CrossRef](#)]
31. Xu, L.; Chatterton, S.; Pennacchi, P. Rolling element bearing diagnosis based on singular value decomposition and composite squared envelope spectrum. *Mech. Syst. Signal Process.* **2021**, *148*, 107174. [[CrossRef](#)]
32. Kia, S.H.; Henao, H.; Capolino, G.-A. Efficient digital signal processing techniques for induction machines fault diagnosis. In Proceedings of the 2013 IEEE Workshop on Electrical Machines Design, Control and Diagnosis (WEMDCD), Paris, France, 11–12 March 2013; pp. 232–246.
33. Li, Y.; Gu, J.X.; Zhen, D.; Xu, M.; Ball, A. An evaluation of gearbox condition monitoring using infrared thermal images applied with convolutional neural networks. *Sensors* **2019**, *19*, 2205. [[CrossRef](#)]
34. Yongbo, L.; Xiaoqiang, D.; Fangyi, W.; Xianzhi, W.; Huangchao, Y. Rotating machinery fault diagnosis based on convolutional neural network and infrared thermal imaging. *Chin. J. Aeronaut.* **2020**, *33*, 427–438.
35. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* **2016**, *17*, 2030–2096.
36. Lu, W.; Liang, B.; Cheng, Y.; Meng, D.; Yang, J.; Zhang, T. Deep model based domain adaptation for fault diagnosis. *IEEE Trans. Ind. Electron.* **2016**, *64*, 2296–2305. [[CrossRef](#)]