Article

Research on Rolling Bearing Fault Diagnosis Method Based on Generative Adversarial and Transfer Learning

Xin Pei [1], Shaohui Su [1,2], Linbei Jiang [1], Changyong Chu [1,2,3,*], Lei Gong [1] and Yiming Yuan [1,2]

[1] School of Mechanical Engineering, Hangzhou Dianzi University, Hangzhou 310018, China; 212010130@hdu.edu.cn (X.P.); sshhui@hdu.edu.cn (S.S.); 212010100@hdu.edu.cn (L.J.); alei@hdu.edu.cn (L.G.); gavie.12345@163.com (Y.Y.)
[2] Anji Intelligent Manufacturing Technology Research Institute, Hangzhou Dianzi University, Huzhou 313000, China
[3] State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430074, China
* Correspondence: kevin@hdu.edu.cn

Abstract: The diagnosis of rolling bearing faults has become an increasingly popular research topic in recent years. However, many studies have been conducted based on sufficient training data. In the real industrial scene, there are some problems in bearing fault diagnosis, including the imbalanced ratio of normal and failure data and the amount of unlabeled data being far more than the amount of marked data. This paper presents a rolling bearing fault diagnosis method suitable for different working conditions based on simulating the real industrial scene. Firstly, the dataset is divided into the source and target domains, and the signals are transformed into pictures by continuous wavelet transform. Secondly, Wasserstein Generative Adversarial Nets-Gradient Penalty (WGAN-GP) is used to generate false sample images; then, the source domain and target domain data are input into the migration learning network with Resnet50 as the backbone for processing to extract similar features. Multi-Kernel Maximum mean discrepancies (MK-MMD) are used to reduce the edge distribution difference between the data of the source domain and the target domain. Based on Case Western Reserve University's dataset, the feasibility of the proposed method is verified by experiments. The experimental results show that the average fault diagnosis accuracy can reach 96.58%.

1. Introduction

With the rapid development of industrial technology, the modernization level of industrial equipment is constantly improving, and the structure is becoming more complex. Casualties and heavy economic losses often accompany equipment failure. As the core component of the rotating mechanism, the rolling bearing supports the transmission elements through the rolling contact between the parts. Its working state plays a vital role in the reliability and safety of running equipment, so it is of great significance to study the fault diagnosis of rolling bearings.

In recent years, scholars at home and abroad have conducted ongoing fault detection research. The relevant research status is summarized as follows:

(1) Fault detection method based on traditional machine learning.

Those who use traditional machine learning for fault detection, such as Yang et al. proposed a SCADA raw signal reconstruction modeling technique using support vector regression, which showed higher performance in wind turbine fault detection, false alarm, and missed alarm control [1]. Hajji et al. developed an improved FDD technology for photovoltaic system failures, using the principal component analysis (PCA) technique for extracting and selecting the most relevant multivariate features and the supervised machine

learning (SML) classifiers are applied for fault diagnosis [2]. Zhang et al. established a data-driven wind turbine fault detection framework by combining random forest (R.F.) and extreme gradient lifting (XGBoost). Direct sensor signals or variables constructed by prior knowledge are classified. They conducted numerical simulations on three different types of wind turbines using the most advanced wind turbine simulator FAST [3]. SU et al. propose a distributed sensor-fault detection and diagnosis system based on machine learning algorithms, and its performance in fault detection is analyzed in terms of detection accuracy, the area under the ROC curve (AUC-ROC), false-positive rate, and F1 score [4]. Euldji et al. proposed a technique based on Vibration Analysis and the decision tree model. In the vibration signal analysis of the rolling bearing experiment, they used the classification algorithm to construct the decision tree and realize the development of the decision system [5]. Sharma et al. decomposed bearing fault signals under different working conditions by permutation entropy (P.E.) and flexible analytic wavelet transform (fawt) [6]. Xia et al. proposed an output hidden feedback Elman adaptive boosting-bootstrap aggregation algorithm under the comprehensive diagnosis framework, which demonstrates good diagnosis performance and high stability in different stages of rolling bearing fault [7]. Singh et al. used Stockwell transform (S.T.) to analyze stator current signals and diagnose various motor operating conditions. The extracted features were input into two different support vector machine (SVM) models, and good classification accuracy was achieved [8]. The traditional machine learning method is simple and fast but only suitable for small data occasions.

(2) Fault detection method based on deep learning and transfer learning.

The rapid development of deep learning and transfer learning can be significantly improved when applied to fault diagnosis research. Feng et al. proposed a local connection network (LCN) constructed by the normalized sparse automatic encoder (NSAE-LCN), which integrates feature extraction and fault recognition into a general learning program, and the algorithm is validated on two different vibration signal datasets [9]. Wang et al. proposed an extended deep belief network (EDBN), which combines valuable information and features in the original data to construct a dynamic fault classifier to extract the process data of dynamic characteristics consideration [10]. Guo et al. proposed a hierarchical learning rate-adaptive depth convolution neural network based on an improved algorithm and studied its application in bearing fault diagnosis and severity [11]. Andrew et al. studied the feasibility of deep learning technology in fault detection of industrial cold forgings. After collecting data from several common faults, the convolutional neural network classifier detects fault conditions with 99.02% accuracy and further classifies each fault with 92.66% accuracy [12]. Ewert et al. detected the possibility of mechanical damage to permanent magnet synchronous motor (PMSM) by analyzing the mechanical vibration supported by a shallow neural network (N.N.) [13]. Huang et al. developed a new dual network prognosis model of deep convolution neural network (DCNN) and multilayer perceptron (MLP). At the same time, one-dimensional time series data and two-dimensional image data are used as inputs to improve the performance of bearing life prediction [14]. Hadi et al. proposed a new method of analyzing speed by using discrete cosine transform (DCT) and used probabilistic neural network (PNN) to identify bearing failure, which achieved a better recognition effect than traditional SVM and ANN classifiers [15]. Khorram et al. proposed an end-to-end fault detection method. Using the equivalent time series as the input of the novel convolutional Long and Short memory recursive neural network (CRNN) to detect bearing faults with the highest accuracy in the shortest possible time [16]. Supervised learning is based on many labeled data, improving classification accuracy. However, labeled data will consume many materials and financial resources in real industrial scenarios. Therefore, semi-supervised, unsupervised, and migration learning represent the hot topics studied by more scholars [17–20].

(3) Application of Generative Adversarial Networks (GAN) in fault and image generation.

GAN was proposed by Goodfellow [21] in 2014. It has great potential in data generation. With its unsupervised characteristics, GAN and other variants have been widely used in image processing. Bowles generates synthetic samples with actual image appearance through GAN for medical segmentation tasks [22]. Loey et al. used the conditional generative adversarial networks (CGAN) based on the deep transfer learning model to generate more images for COVID-19 detection in chest CT scanning images [23]. Dixit et al. proposed a novel condition-assisted classifier GAN framework, which was combined with Model Agnostic Meta-Learning (MAML) to generate high-quality samples in the case of limited bearing fault data [24]. Li et al. proposed a semi-supervised FDD method to replace the GAN binary discriminator with a multi-class classifier, which realized efficient fault detection for building HVAC systems [25]. Zhou et al. redesigned the Generator and Discriminator of GAN, generated more fault samples using the global optimization scheme, and solved the high fault diagnosis error rate in imbalanced data due to uneven feature extraction [26]. Yan et al. adopted a custom generative adversarial network and proposed a GAN integration framework to rebalance the training data set of cooler AFDD [27].

(4)   Summary and analysis of research status.

The above work directly classifies and diagnoses the original one-dimensional data; or simply processes the data for diagnosis. It does not take into account that the proportion of standard data and fault data in the real industrial scenario is not neat, the sample size is small, and the experimental data are different from the actual industrial scenario data.

- Mechanical equipment is in regular operation, with few fault data and an imbalanced proportion of fault types, resulting in a poor fault diagnosis.
- The research object is the labeled fault data, while the data diagnosed in the industry are often the unlabeled data. In addition, the amount of unlabeled data is far greater than the amount of labeled data.
- There are differences in fault samples under different working conditions. It is necessary to establish different models to achieve accurate prediction, and a single model exhibits poor universality.

This paper proposes a composite semi-supervised fault diagnosis method combining continuous wavelet transform, Wasserstein Generative Adversarial Nets-Gradient Penalty (WGAN-GP), and transfer learning based on the analysis of the current research situation. Through the combination of continuous wavelet transform and WGAN-GP processing, and finally, through the fault diagnosis method of migration learning, the feasibility of the process is verified on the data set of Case Western Reserve University. The experimental results show that the method proposed in this paper can better solve the aforementioned problems.

2. WGAN-GP Based Transfer Learning Method

Figure 1 shows the network model. The steps of the diagnosis method are as follows:

Step 1: Since the relationship between frequency and time was not fully expressed in one-dimensional time domain signals, the original rolling bearing fault signals were processed. Continuous wavelet transformation transformed the rolling bearings faults from one-dimensional vibration data into two-dimensional time-frequency images. The labeled source domain images were divided into ten types of defects and named real images.

Step 2: The divided real images were taken as the input of the WGAN-GP network, and the internal network parameters of generators and discriminators were updated through the game between them in the network. There were many false sample images generated after reaching the Nash equilibrium [28]. These images that have reached the Nash equilibrium are called fake images.

Step 3: After the image generation in Step 2, a second selection of fake images is made. First, resnet50 is trained by using only the wavelet transformed images in the source domain as the training set and the verification set, and the training weight of the network is obtained. The training weight is added to the prediction script for Resnet50. The prediction script of resnet50 selects the false sample images with a probability of prediction accuracy

greater than 90%. After repeated selection, the reliability of the photos is relatively high. Finally, this part of the image is combined with the real image in Step 1 as a new source domain image.

Step 4: Migration learning adopts resnet50 as the backbone network, adds the Multi-Kernel Maximum mean discrepancies (MK-MMD) adaptation layer to the full connection layer of the resnet50 network, takes the fault pictures of the new source domain synthesized by Step 3 and tests the target domain as input. It then shares the parameters of the two parts, and realizes the detection of rolling bearing data in the target domain by migration learning.
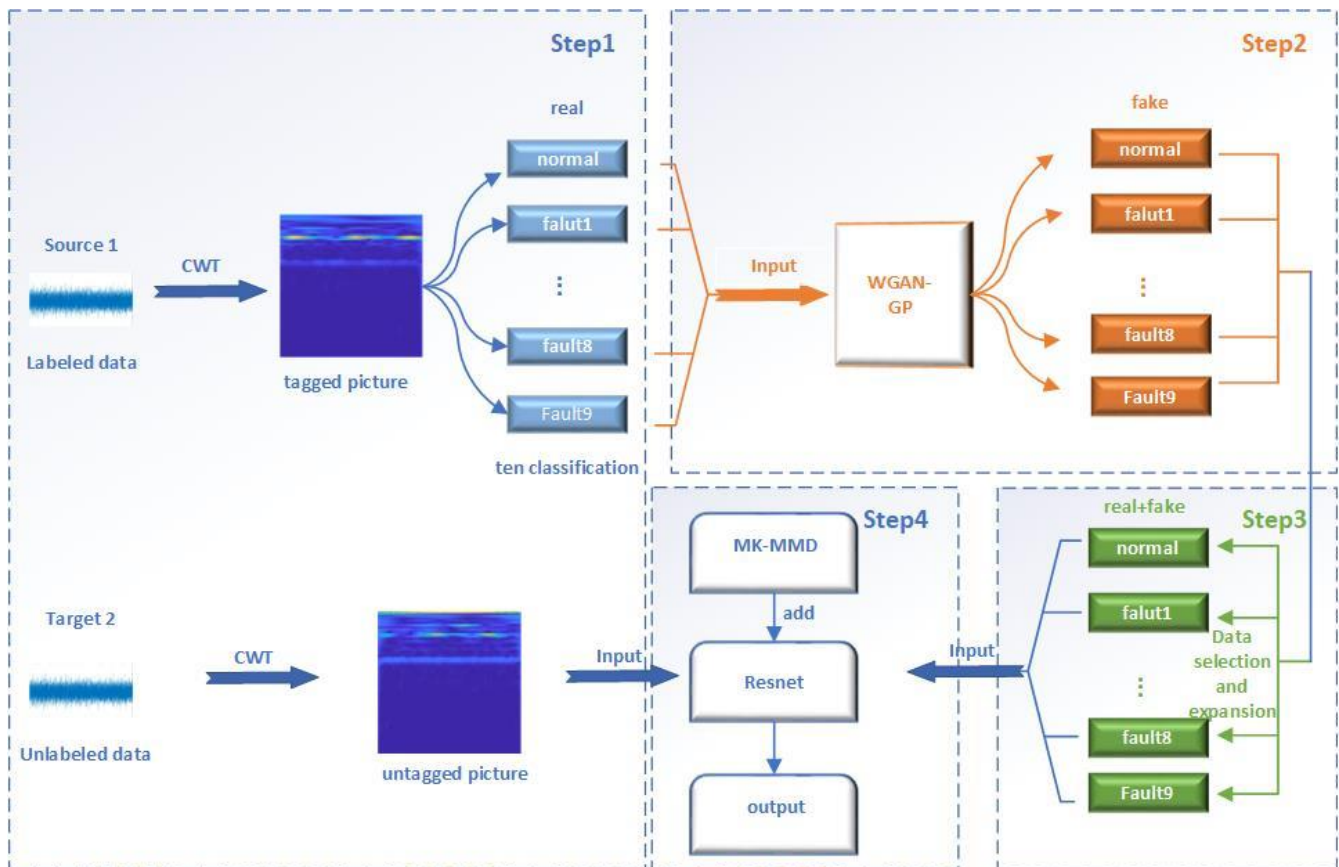

Figure 1. Fault diagnosis process diagram.

The idea of the Step 1–Step 4 scheme adopted in this paper is as follows:

1. In view of the problem that the amount of unlabeled data in real industrial scenes far exceeds the amount of labeled data, firstly, Step 1 applies the continuous wavelet transform to increase the extractable feature information. Secondly, the purpose of false generation through Step 2 is to expand the amount of labeled data. In the last Step 3, the second selection of the generated false sample data aims to improve the accuracy of the semi-supervised fault diagnosis designed in this paper.

2. Step 2 and Step 3 can also solve the problem of unbalanced proportion between normal and fault data in real industrial scenarios. After expanding unbalanced data, WGAN-GP is adopted to select samples with high reliability, to balance unbalanced group data.

3. In the real industrial situation, rolling bearings will work at different speeds, and the faults generated will be distributed in different working conditions, and the fault data in different working conditions must be different, so it is very inefficient to build different models to diagnose the faults in each working condition, Therefore, the Step 4 network model is proposed in this paper. Resnet50 is used as the backbone network, MK-MMD

is added to reduce the difference between different working conditions, and the target domain data of Step 1 and the source domain data of Step 3 are used as the input. Only the fault data in one working condition are used as the source domain to realize the fault diagnosis of the target domain in different working conditions.

3. Continuous Wavelet Transform of Data

Scholars have widely used Continuous wavelet transform to process mechanical fault signals [29,30]. Compared with the Fourier transform, the signal can be observed simultaneously in the time and frequency domains. Short-time Fourier transform is also a representative of the time-frequency method; compared with it, wavelet transform has the characteristics of window adaptive. It can better balance the frequency resolution and time resolution. The definition of continuous wavelet transform is that for any signal $f(t) \in L^2(R)$:

$$WT_\varphi f(a,b) = \int_{-\infty}^{+\infty} f(t)\overline{\varphi}_{a,b}(t)dt = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} f(t)\overline{\varphi}(\left(\frac{t-b}{a}\right))dt \tag{1}$$

In the formula, $a > 0$ is the scale parameter, whose function is to scale and transform basic wavelet $\varphi(t)$. The larger a is, the wider the waveform $\varphi(a/t)$ is. The constant factor $1/\sqrt{a}$ means that the $\varphi_{a,b}(t)$ energy under different a remains equal. b is the location parameter, and $\overline{\varphi}(t)$ is the complex conjugate of $\varphi(t)$. Continuous wavelet transform can also be rewritten as follows:

$$WT\_\varphi\, f(a,b) = \langle f(t), \varphi\_(a,b)\,(t)\,\rangle; \tag{2}$$

This inner product form gives the degree of similarity between the signal $f(t)$ and the template function $\varphi_{a,b}(t)$. The higher the degree of similarity, the greater the $WT_\varphi f(a,b)$. Combined with the actual needs, this paper needs to process the one-dimensional vibration fault data, so the wavelet amplitude must be smooth and continuous. Complex-valued wavelet with good phase expression ability should be selected to obtain the time and frequency domain characteristics. Combined with the above bearing fault diagnosis characteristics, this paper sets the Morlet wavelet, a complex exponential wavelet with non-orthogonality. There is a good balance between time and frequency localization, and a wide peak can contain positive and negative peaks, which have more fault vibration information. In this paper, continuous wavelet transform is applied to all fault data of rolling bearings. The sampling frequency is set to 12 KHZ and the sampling length to 1024. Due to the limited fault data, to obtain more training samples and avoid overfitting caused by a small amount of data, the fault data are enhanced by overlapping sampling. In addition, the scale of the overlapping piece is set to 0.5. After the wavelet transform, all the failure data are processed, four shown in Figure 2.
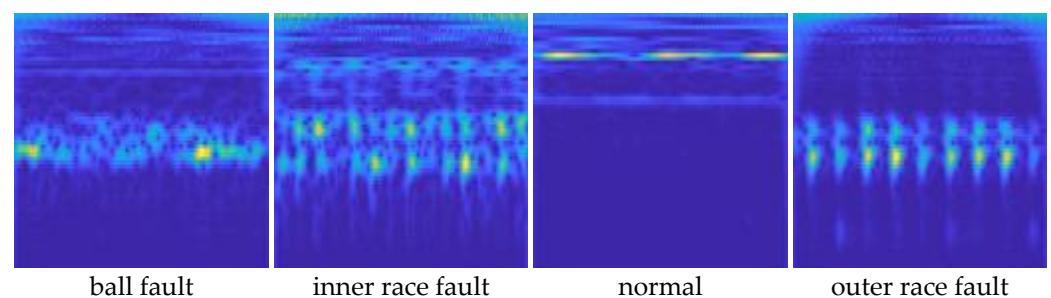

| ball fault | inner race fault | normal | outer race fault |

Figure 2. Four types of fault data.

4. Generation and Processing of Samples

4.1. Generative Adversarial Networks

Generative Adversarial Networks consist of Generator (G) and Discriminator (D) The generation model G achieves the purpose of deceiving the identification model D by generating data consistent with the samples' distribution as much as possible. The goal of identification model D is to distinguish between real data and false data. The two constantly update internal parameters to achieve Nash equilibrium through the mutual game, and finally, use the generic model to generate false data. GAN reduces the influence on the error by alternately training the two models, and its loss function is:

$$\min_G \max_D L(G, D) = E_x \sim p_{data}[\log D(x)] + E_z \sim p_z[\log(1 - D(G(z)))] \tag{3}$$

Most GANs have some defects, such as the collapse of generation mode, difficulty in model training, and instability of the training process. In this paper, WGAN-GP [31–33], a variant of GAN, is used to generate high-quality false samples, to solve the problems of imbalanced data types of fault samples and low sample numbers. WGAN replaces K.L. and J.S. divergence by introducing Wasserstein distance (Formula (4)).

$$W(P_1, P_2) = \frac{\ln f}{\gamma \sim \Pi(P_1, P_2)} \mathbb{E}_{(x,y)\sim\gamma}[\| x - y \|] \tag{4}$$

In the formula, the joint distribution of P_1 and P_2 is $\Pi(P_1, P_2)$. For γ joint distributions, real data x and generated false data y can be sampled from them and conform to $(x, y) \sim \gamma$, and the distance between samples is calculated $\| x - y \|$. Finally, the expected value of the samples $\mathbb{E}_{(x,y)\sim\gamma}[\| x - y \|]$ can be calculated. Based on WGAN, WGAN-GP adopts the gradient penalty strategy to replace the weight pruning strategy, which also realizes Lipschitz continuity and achieves more stable generated results and higher quality generated pictures. Its objective function is as follows:

$$L = \frac{E}{\tilde{x} \sim \mathbb{P}_g}[D(\tilde{x})] - \frac{E}{\tilde{x} \sim \mathbb{P}_r}[D(x)] + \lambda \frac{E}{\tilde{x} \sim \mathbb{P}_{\tilde{x}}}\left[\left(\| \nabla_{\tilde{x}} D(\tilde{x}) \|_2 - 1\right)^2\right] \tag{5}$$

where $\| \nabla_{\tilde{x}} D(\tilde{x}) \|_2$ is the weight of discrimination model D plus penalty, and λ is the penalty coefficient. WGAN-GP selects Adam optimizer to optimize the parameters in the training process.

Combined with the bearing fault diagnosis data, the WGAN-GP network structure is constructed, as shown in Figure 3. This paper uses WGAN-GP as the second step to process the vibration fault data after continuous wavelet transformation. The processed data are called real images. During the test, entering too much fault data at once would cause the generated model to crash. The information with unclear characteristics can only generate a few pictures, and there is even one fault type that does not create pictures. Therefore, fault samples are sent to WGAN-GP in batches for the same round of training. Figure 3 shows the whole training process. First, the Generator will process the randomly generated Gauss noise into the image, called fake images here, and then input the real and counterfeit images into the Discriminator to judge the real and unreal images. In the training process, the quality of the Generator and the accuracy of the Discriminator will increase continuously until the Discriminator fails to distinguish fake images, indicating that the training has achieved Nash equilibrium.

Figure 4 shows the loss function curve obtained after 2000 iterations of one sample data in this paper. It can be seen from the figure that the loss of the Discriminator reaches its peak briefly at the beginning of the iteration and then approaches zero after about 100 iterations, indicating that the authentication network has reached Nash equilibrium. After 1250 iterations, the generated network's fluctuation decreases, implying that the quality of the generated images tends to gradually stabilize.
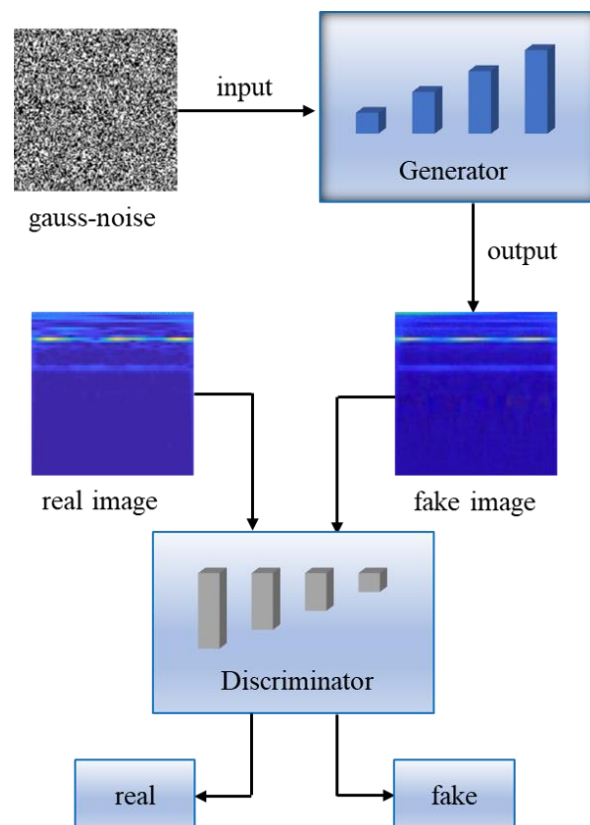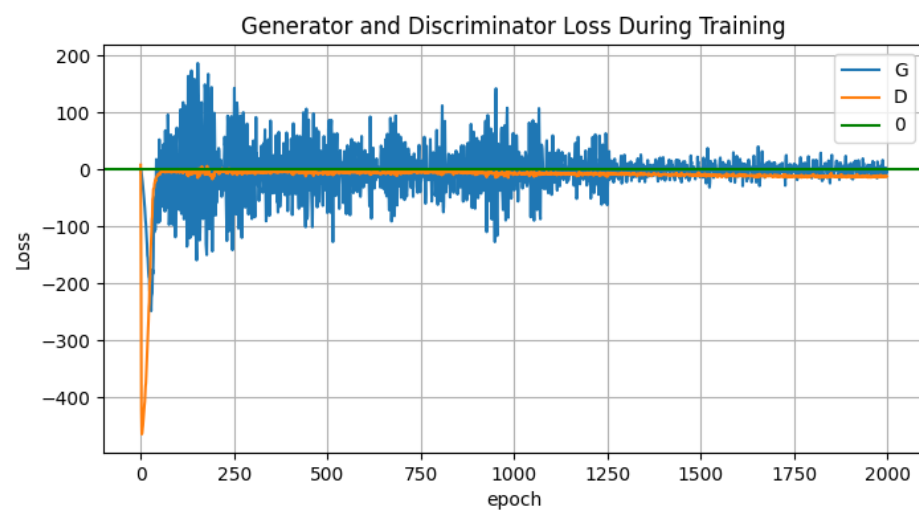
Figure 3. WGAN-GP structure diagram.

Figure 4. G/D loss function curve.

4.2. Convolutional Neural Network

Convolutional Neural Network (CNN) [34] is widely used in image processing, semantic image segmentation, and other fields since Lecun proposed it. It is an important part of deep learning. Its basic structure is generally composed of the input layer, convolution layer, pooling layer, full connection layer, and output layer. The convolution layer, as the core of the convolution neural network, is a special mathematical operation:

$$W_{output} = \frac{W_{iutput} - W_{filter} + 2P}{S} + 1 \quad H_{output} = \frac{H_{iutput} - H_{filter} + 2P}{S} + 1 \qquad (6)$$

where W and H represent the width and height of the image, P represents the number of pixel layers added to the image edge, and S denotes the step size of the convolution kernel.

With the development of convolutional neural networks, scholars have proposed more and more network models. Resnet [35] was proposed in 2015 and won the championship of ILSVRC2015. He et al. proposed a network structure called a residual block. The convolutional neural network solves the gradient vanishing problem by using residual mapping. The expression of a single residual block is as follows:

$$x_{l+1} = x_l + \mathcal{F}(x_1, W_l) \tag{7}$$

Resnet network contains two residual blocks, BasicBlock and Bottleneck, which correspond to two network structures, collectively referred to as BuildingBlock, as shown in Figure 5. The left network structure corresponds to a shallow network, such as Resnet34. For deeper networks, such as Resnet50, they use the Bottleneck structure to reduce the parameters of the overall network by stacking 1×1 convolution layers at the beginning and end, to reduce the training time.
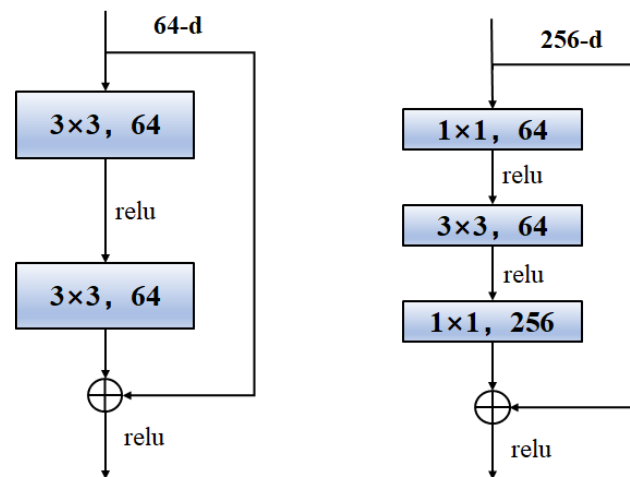

Figure 5. BasicBlock and Bottleneck.

This paper uses Resnet50 as the backbone network to train Step 3 and Step 4. Resnet50 adopts the Bottleneck structure in Figure 5. Compared with Resnet18 with lower network layers, the overall parameters are still within the acceptable range with better network performance. Using Resnet101 and other networks with a deeper network structure is unnecessary from the training results. When executing Step 3, the Resnet50 network is applied to the selected portion of the generated picture; that is, the composite standard is used to select the false samples of the fault. Firstly, the bearing fault data in the source domain are used as the training set and verification set of the Resnet50 network for training. Because the classification task of this paper is different from that of Resnet, a fully connected layer is added after the network structure of Resnet50 during training to match the bearing fault classification task. After training, a weight file is generated. Secondly, after WGAN-GP can generate stable fault pictures, it sends these pictures to the trained Resnet50 network for batch prediction. The fake images are taken as the input, the minimum prediction probability of the network is set to 90%, these pictures with high reliability are chosen as the final selected samples, and they are added to the source domain. The composite criteria adopted in this paper can ensure that the neural network can identify the generated false samples to improve fault diagnosis accuracy.

5. Deep Transfer Learning

Figure 6 shows the network structure of the migration phase proposed in this paper. The updated source domain data and the target domain data to be detected are simultane-

ously input into the deep migration learning network. Resnet50 is selected for the backbone network of deep transfer learning. Since the task structure of this paper is to deal with ten types of faults, the output node of the full connection layer is replaced by 10. Multi-kernel Maximum Mean Discrepancy (MK-MMD) is added to the full connection layer to measure and reduce the edge distribution difference between the source and target domains.
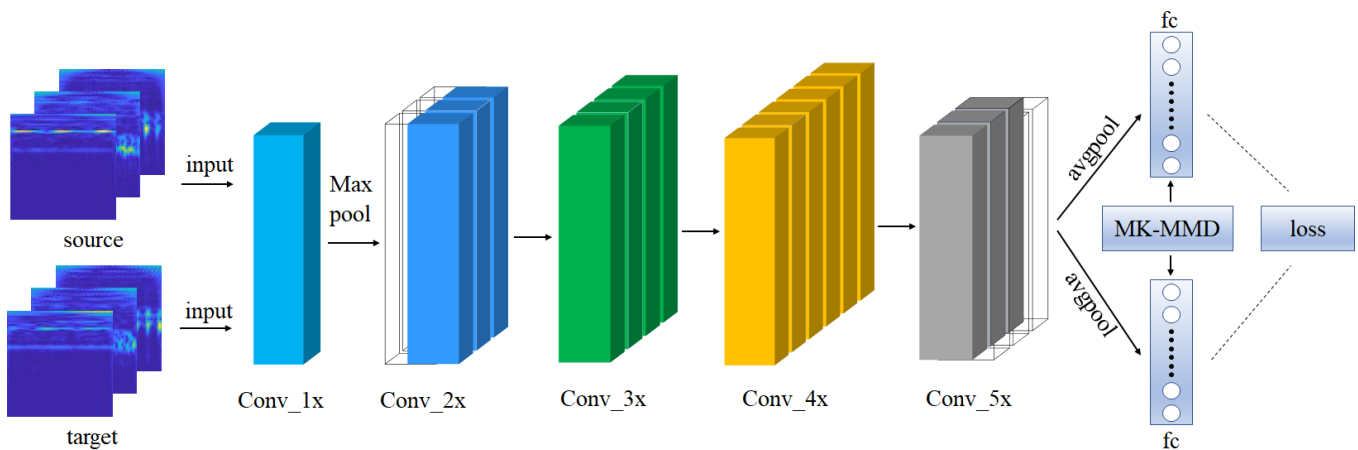

Figure 6. Transfer learning network structure.

Transfer learning is applying knowledge or model to another learning task. Improving the baseline performance, shortening the model development time, and improving performance helps the learning accomplish the target tasks. The success of migration learning is closely related to the correlation between jobs. Pan et al. defined transfer learning in 2009 [36]: The scholar explains the field D as a tuple of two elements, one of which is the marginal probability PX. In addition, X is represented as a sample data point, and the other is the feature space §. $X = \{\S_1, \S_2, \cdots, \S_n\}$, where $\S_i$ is a specific vector, and $X \in \S$, so there are:

$$D = \{\S, PX\} \tag{8}$$

Similarly, a task T can also be expressed as a tuple containing two elements. The formula is as follows:

$$T = \{\gamma, P\gamma|X\} \tag{9}$$

where γ represents the first element: feature space, and $P\gamma|X$ represents another element from a probabilistic point of view: the objective function f. According to the traditional machine algorithm, transfer learning can be divided into inductive, unsupervised, and direct push transfer.

As a typical representative of inductive transfer, deep transfer learning has two advantages over non-deep transfer learning. One is that it can meet the end-to-end needs in practical applications; the other is that it can automatically extract more expressive features. The deep transfer learning task consists of different learning types, such as domain adaptation, domain confusion, multi-task learning, one-time learning, zero sample learning, etc. Regarding domain adaptation, Tzeng et al. first proposed the DDC method [37]. Using AlexNet [38] as the backbone network, he added an adaptive layer between the feature classification layer and the extraction layer to align the data of the source domain and target domain as much as possible. The maximum mean deviation (MMD) he uses in his paper is widely used and is expressed as follows:

$$MMDX, Y = \|\frac{1}{\backslash}\sum_{\rangle=1}^{\backslash} \phi(\S_\phi) - \frac{1}{\Updownarrow}\sum_{|=1}^{\Updownarrow} \phi(\dagger_|)\|_H^2 \tag{10}$$

ϕ is a method to map the data of the source domain and the target domain to the reproducing kernel Hilbert space [39]. Through this formula, the distance between the

source domain and the target domain can be obtained and added to the loss function to improve the efficiency of migration learning.

MMD is correlated with the selected kernel function, and the performance of MMD will decrease when the kernel function performs poorly. The appearance of MK-MMD solves this problem well:

$$K \triangleq \left\{ k = \sum_{\sqcap=1}^{\Updownarrow} \beta_{\sqcap} k_{\sqcap} : \sum_{\sqcap=1}^{\Updownarrow} \beta_{\sqcap} = 1, \beta_{\sqcap} \geq 0, \forall \sqcap \right\} \tag{11}$$

MK-MMD assumes that the optimal kernel can be formed by the linear combination of multiple kernels, which avoids the difficulty of MMD in selecting the kernel function. Therefore, this paper chooses to add MK-MMD to the network so that the network can obtain better characterization ability. The loss function after adding MK-MMD is as follows:

$$L = \frac{min}{\theta} \frac{1}{N} \sum_{i=1}^{n} J(\theta(x_i), y_i) + \lambda L_{MK-MMD} \tag{12}$$

In the formula, the left side of the plus sign is the cross-entropy loss function of the source domain, and λ represents the weight of MK-MMD.

6. Experimental Verification

6.1. Data Set Introduction

This paper uses the bearing data set of Case Western Reserve University (CWRU) to verify the effectiveness and feasibility of the proposed method. As shown in Figure 7, the CWRU test-bed has a fan end bearing, a two-horsepower motor, drive end bearing, torque sensor, encoder, dynamometer, etc., from left to right. An acceleration sensor is placed above the bearing seat of the motor's fan end and drive end respectively to collect vibration acceleration signals of the faulty bearing. Vibration signals are collected by the 16-channel data recorder, and the power and speed are measured by the torque sensor and the encoder.
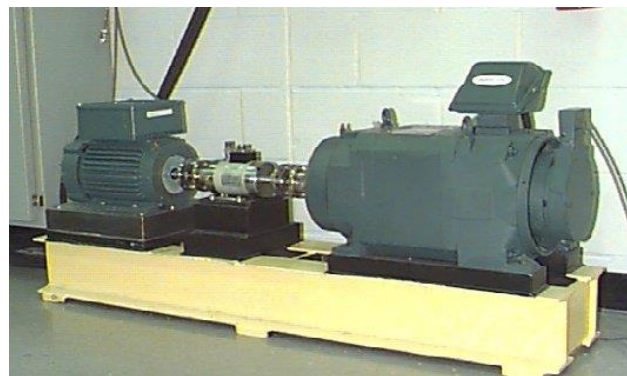

Figure 7. CWRU Bearing test bench.

In this experiment, electrical discharge machining (EDM) was used to cause single-point damage to bearings. The damage to the inner race, outer race, and rolling body of bearing under different loads is simulated. The bearing to be tested supports the motor's rotating shaft, the fan end bearing is SKF6203 with a sampling frequency of 12 KHZ, and the driving end bearing is SKF6205. The sampling frequency is 12 KHZ and 48 KHZ. The experiment has four rotating speeds: 1797 r/min, 1772 r/min, 1750 r/min, and 1730 r/min. The corresponding loads are 0 HP, 1 HP, 2 HP and 3 HP, respectively. The processed faulty bearings were re-loaded into the test motor, and the vibration acceleration signal data were recorded under the working conditions of 0, 1, 2, and 3 horsepower motor load, respectively. To meet the requirements of cross-working condition diagnosis in this paper, bearing fault signals at the driving end are selected at a sampling frequency of 12 KHZ.

This paper divides faults into ten categories according to the damaged parts and fault diameters of bearings, as shown in Table 1, which lists the fault data at each speed. The original data adopt the continuous wavelet transform method in Step 1 to obtain 10 types of fault pictures at each rate, 200 photos of each type, and a total of 8000 images. After the pictures in Step 2 and Step 3 are generated, each class selects 100 samples with high reliability to join as a new training set to train the network. Take 0 HP as an example; Table 1 compares the generated false sample image and the real sample image. The accuracy of fault identification is obtained by inputting the source and target domain into the migration network simultaneously through Step 4.

Table 1. Classification of fault types.

Fault Type	Ball	Ball	Ball	Inner Race	Inner Race
Fault diameter/″	0.007	0.014	0.028	0.007	0.014
Real sample images (200/type)					
Generate sample images (100/type)					
Fault type	inner race	normal	outer race	outer race	outer race
Fault diameter/″	0.028	0	0.007	0.014	0.028
Real sample images					
Generate sample images					

6.2. Experimental Verification under Cross-Working Conditions

Based on Experiment 1 above, to simulate the need for fault diagnosis in real industrial scenes, the experiment will verify the proposed method in two stages on the premise of cross-working conditions. In Experiment 1, four different rotational speeds were used as transfer learning conditions under cross-working conditions. For example, 0-1 indicates that 0 HP is the source domain and 1 HP is the target domain for training. A total of 12 migration schemes were set up in the experiment, and the control groups were DDC, Deepcoral [40], and DAN [41]. Under the same experimental conditions, Table 2 shows the experimental results.

To fully verify that this method still has a good effect in the case of imbalanced fault sample data, we add Experiment 2. First, based on Experiment 1, three groups with the highest accuracy were selected: 0-2 migration, 1-2 migration, and 3-2 migration. The average accuracy of these three transfer experiments under the four algorithms is 84.7%, 85.09%, and 84.76%, respectively. Secondly, three defect samples of ten types of defects are randomly reduced to 10% to simulate the state when the fault data are imbalanced. The algorithm proposed in this paper generates the remaining 90% false sample data through the WGAN-GP network of Step 2 and fills it into the source domain data set to keep the data set balanced. Finally, under the same conditions, compared with the other three algorithms, the accuracy results are shown in Table 3.

Table 2. The accuracy of fault diagnosis under different working conditions.

Transfer Patterns	DDC	Deep Coral	DAN	The Proposed Algorithm
0-1	55.75%	72.10%	97.45%	99.25%
0-2	72.45%	70.50%	96.65%	99.20%
0-3	71.50%	59.95%	89.95%	94.80%
1-0	56.20%	71.05%	97.15%	97.60%
1-2	74.20%	66.75%	99.50%	99.90%
1-3	54.15%	57.60%	97.00%	99.10%
2-0	69.00%	63.00%	89.20%	92.05%
2-1	68.70%	59.80%	93.75%	98.05%
2-3	60.95%	77.80%	98.70%	99.60%
3-0	61.30%	60.70%	83.80%	85.35%
3-1	70.60%	67.45%	89.95%	94.20%
3-2	69.50%	70.40%	99.30%	99.85%
average	65.36%	66.34%	94.37%	96.58%

Table 3. Experimental results with imbalanced data.

Transfer Patterns	DDC	Deep Coral	DAN	The Proposed Algorithm
0-2	51.05%	47.75%	87.80%	98.95%
1-2	53.25%	57.80%	91.30%	99.95%
3-2	66.95%	52.05%	91.45%	99.90%
average	57.08%	52.53%	90.18%	99.60%

In the real industrial scene, in addition to the problem of unbalanced data amount, the data collection process will also be interfered with by noise. Therefore, Experiment 3 was set up to verify the robustness of the proposed rolling bearing fault diagnosis method. On the premise of Experiment 2, an SNR of 5 dB and 10 dB was added to the target domain based on the unbalanced fault samples. The smaller the SNR, the greater the noise. With the same experimental conditions, the final algorithm pair is shown in Table 4.

Table 4. Experimental results in unbalanced data + noise environment.

Transfer Patterns/SNR = 5 dB	DDC	Deep Coral	DAN	The Proposed Algorithm
0-2	57.30%	65.65%	87.55%	98.65%
1-2	54.60%	50.65%	88.95%	99.75%
3-2	50.95%	69.10%	92.50%	99.95%
average	54.28%	61.80%	89.67%	99.45%
Transfer Patterns/SNR = 10 dB	DDC	Deep Coral	DAN	The Proposed Algorithm
0-2	59.55%	50.85%	89.55%	99.15%
1-2	50.20%	45.10%	93.15%	99.90%
3-2	47.35%	50.85%	95.00%	99.80%
average	52.37%	48.93%	92.57%	99.62%

6.3. Analysis of Experimental Results

As can be seen from the results in Table 2, the average accuracy of the proposed method under the 12 migration schemes is 96.58%, which is higher than 65.36% for DDC and 66.34% for Deepcoral, and 94.37% for DAN. It shows that after the source domain data samples are generated and expanded by this method, the new samples can be trained to achieve a better and more accurate network model. In Experiment 1, the average accuracy of the three migration schemes 0 HP–2 HP, 1 HP–2 HP, and 3 HP–2 HP when the target domain is 2 HP covers the top three, indicating that the fault feature extraction effect of 12 KHZ drive ends bearing under the load of 2 HP is the best.

Table 3 shows that when the fault sample data are imbalanced, the diagnostic accuracy of other methods decreases by more than 10% on average, indicating that the imbalanced data will lead to the imbalance of fault diagnosis of other models. The average accuracy of the proposed method is still 99.60%, which is only one sample data gap compared with 99.65% when the data are balanced. This method maintains good accuracy and reliability given the imbalanced data in the real industrial situation.

In Experiment 3, different SNR noises were added to the target domain based on Experiment 2 to simulate the possible data collection problems in real industrial conditions. Under the influence of different decibels, DDC, Deepcoral, and DAN showed great differences in accuracy. The average classification accuracy of DDC and Deepcoral at SNR = 5 is higher than that of SNR = 10, while DAN shows the opposite. In addition, the average classification accuracy of the proposed algorithm remains above 99% in the complex environment with unbalanced data and two different SNRs, which fully verifies the robustness of the proposed algorithm. It shows that the algorithm can effectively solve practical industrial problems.

7. Conclusions

The method proposed in this paper effectively solves three problems in the real industrial situation and is verified under the bearing data set of Case Western Reserve University. The following conclusions can be obtained:

(1) The data generated by WGAN-GP are selected by Resnet for the second time and then added to the source domain data to form the new source domain data, which can solve the problem of imbalanced fault data in the industry. In addition, in the case of data balance, it still can improve detection accuracy.

(2) The detected fault data should be unlabeled in real industrial situations. Resnet50 is used as the backbone network, and MK-MMD was added as the adaptation layer of the network so that the data of the source domain and the target domain can share parameters. It can effectively realize the detection of unmarked data.

(3) The fault data under different working conditions have other fault characteristics, so it is necessary to establish multiple network models to detect the data under different working conditions. In addition, if you use the transfer learning method, it can predict the fault data in any target domain with high accuracy under other source fault data.

Author Contributions: Conceptualization, X.P. and S.S.; methodology, X.P. and C.C.; validation, X.P. and S.S.; writing—original draft preparation, X.P. and L.G.; writing—review and editing, X.P. and S.S.; visualization, supervision, project administration, funding acquisition, S.S., Y.Y. and L.J. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. Yang, C.; Liu, J.; Zeng, Y.; Xie, G. Real-time condition monitoring and fault detection of components based on machine-learning reconstruction model. *Renew. Energy* **2019**, *133*, 433–441. [CrossRef]
2. Hajji, M.; Harkat, M.-F.; Kouadri, A.; Abodayeh, K.; Mansouri, M.; Nounou, H.; Nounou, M. Multivariate feature extraction based supervised machine learning for fault detection and diagnosis in photovoltaic systems. *Eur. J. Control* **2021**, *59*, 313–321. [CrossRef]
3. Zhang, D.; Qian, L.; Mao, B.; Huang, C.; Huang, B.; Si, Y. A data-driven design for fault detection of wind turbines using random forests and XGboost. *IEEE Access* **2018**, *6*, 21020–21031. [CrossRef]
4. Jan, S.U.; Lee, Y.D.; Koo, I.S. A distributed sensor-fault detection and diagnosis framework using machine learning. *Inf. Sci.* **2021**, *547*, 777–796. [CrossRef]
5. Euldji, R.; Boumahdi, M.; Bachene, M. Decision-making based on decision tree for ball bearing monitoring. In Proceedings of the 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-Being (IHSH), Boumerdes, Algeria, 9–10 February 2021; pp. 171–175.
6. Sharma, S.; Tiwari, S.; Singh, S. Integrated approach based on flexible analytical wavelet transform and permutation entropy for fault detection in rotary machines. *Measurement* **2021**, *169*, 108389. [CrossRef]
7. Xia, T.; Zhuo, P.; Xiao, L.; Du, S.; Wang, D.; Xi, L. Multi-stage fault diagnosis framework for rolling bearing based on OHF Elman AdaBoost-Bagging algorithm. *Neurocomputing* **2021**, *433*, 237–251. [CrossRef]
8. Singh, M.; Shaik, A.G. Incipient fault detection in stator windings of an induction motor using Stockwell transform and SVM. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 9496–9504. [CrossRef]
9. Jia, F.; Lei, Y.; Guo, L.; Lin, J.; Xing, S. A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines. *Neurocomputing* **2018**, *272*, 619–628. [CrossRef]
10. Wang, Y.; Pan, Z.; Yuan, X.; Yang, C.; Gui, W. A novel deep learning based fault diagnosis approach for chemical process with extended deep belief network. *ISA Trans.* **2020**, *96*, 457–467. [CrossRef]
11. Guo, X.; Chen, L.; Shen, C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement* **2016**, *93*, 490–502. [CrossRef]
12. Glaeser, A.; Selvaraj, V.; Lee, S.; Hwang, Y.; Lee, K.; Lee, N.; Lee, S.; Min, S. Applications of deep learning for fault detection in industrial cold forging. *Int. J. Prod. Res.* **2021**, *59*, 4826–4835. [CrossRef]
13. Ewert, P.; Orlowska-Kowalska, T.; Jankowska, K. Effectiveness analysis of PMSM motor rolling bearing fault detectors based on vibration analysis and shallow neural networks. *Energies* **2021**, *14*, 712. [CrossRef]
14. Huang, C.-G.; Huang, H.-Z.; Li, Y.-F.; Peng, W. A novel deep convolutional neural network-bootstrap integrated method for RUL prediction of rolling bearing. *J. Manuf. Syst.* **2021**, *61*, 757–772. [CrossRef]
15. Hadi Salih, I.; Babu Loganathan, G. Induction motor fault monitoring and fault classification using deep learning probablistic neural network. *Solid State Technol.* **2020**, *63*, 2196–2213.
16. Khorram, A.; Khalooei, M.; Rezghi, M. End-to-end CNN+ LSTM deep learning approach for bearing fault diagnosis. *Appl. Intell.* **2021**, *51*, 736–751. [CrossRef]
17. Xiao, D.; Huang, Y.; Qin, C.; Liu, Z.; Liu, C. Transfer learning with convolutional neural networks for small sample size problem in machinery fault diagnosis. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2019**, *233*, 5131–5143. [CrossRef]
18. Wen, L.; Gao, L.; Dong, Y.; Zhu, Z. A negative correlation ensemble transfer learning method for fault diagnosis based on convolutional neural network. *Math. Biosci. Eng.* **2019**, *16*, 3311–3330. [CrossRef]
19. Wang, G.; Zhang, M.; Xiang, L.; Hu, Z.; Li, W.; Cao, J. A multi-branch convolutional transfer learning diagnostic method for bearings under diverse working conditions and devices. *Measurement* **2021**, *182*, 109627. [CrossRef]
20. Li, X.; Zhang, W.; Ding, Q.; Li, X. Diagnosing rotating machines with weakly supervised data using deep transfer learning. *IEEE Trans. Ind. Inform.* **2019**, *16*, 1688–1697. [CrossRef]
21. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets—Advances in neural information processing systems 27. *arXiv* **2014**, arXiv:1406.2661v1. Available online: https://arxiv.org/pdf/1406.2661.pdf (accessed on 10 May 2022).
22. Bowles, C.; Chen, L.; Guerrero, R.; Bentley, P.; Gunn, R.; Hammers, A.; Dickie, D.A.; Hernández, M.V.; Wardlaw, J.; Rueckert, D. Gan augmentation: Augmenting training data using generative adversarial networks. *arXiv* **2018**, arXiv:1810.10863. Available online: https://arxiv.org/abs/1810.10863 (accessed on 5 April 2022).
23. Loey, M.; Manogaran, G.; Khalifa, N.E.M. A deep transfer learning model with classical data augmentation and CGAN to detect COVID-19 from chest CT radiography digital images. *Neural Comput. Appl.* **2020**, 1–13. [CrossRef] [PubMed]
24. Dixit, S.; Verma, N.K.; Ghosh, A. Intelligent fault diagnosis of rotary machines: Conditional auxiliary classifier GAN coupled with meta learning using limited data. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–11. [CrossRef]
25. Li, B.; Cheng, F.; Cai, H.; Zhang, X.; Cai, W. A semi-supervised approach to fault detection and diagnosis for building HVAC systems based on the modified generative adversarial network. *Energy Build.* **2021**, *246*, 111044. [CrossRef]
26. Zhou, F.; Yang, S.; Fujita, H.; Chen, D.; Wen, C. Deep learning fault diagnosis method based on global optimization GAN for unbalanced data. *Knowl. Based Syst.* **2020**, *187*, 104837. [CrossRef]
27. Yan, K.; Chong, A.; Mo, Y. Generative adversarial network for fault detection diagnosis of chillers. *Build. Environ.* **2020**, *172*, 106698. [CrossRef]

28. Alpcan, T.; Pavel, L. Nash equilibrium design and optimization. In Proceedings of the 2009 International Conference on Game Theory for Networks, Istanbul, Turkey, 13–15 May 2009; pp. 164–170.
29. Pucciarelli, G. Wavelet analysis in volcanology: The case of phlegrean fields. *J. Environ. Sci. Eng. A* **2017**, *6*, 300–307.
30. Torrence, C.; Compo, G.P. A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* **1998**, *79*, 61–78. [CrossRef]
31. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved training of wasserstein gans—Advances in neural information processing systems 30. *arXiv* **2017**, arXiv:1704.00028v3. Available online: https://arxiv.org/abs/1704.00028 (accessed on 5 January 2022).
32. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 214–223.
33. Arjovsky, M.; Bottou, L. Towards principled methods for training generative adversarial networks. *arXiv* **2017**, arXiv:1701.04862. Available online: https://arxiv.org/pdf/1701.04862.pdf (accessed on 10 February 2022).
34. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324.
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
36. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [CrossRef]
37. Tzeng, E.; Hoffman, J.; Zhang, N.; Saenko, K.; Darrell, T. Deep domain confusion: Maximizing for domain invariance. *arXiv* **2014**, arXiv:1412.3474. Available online: https://arxiv.org/pdf/1412.3474.pdf (accessed on 15 March 2022).
38. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*; Pereira, F., Burges, C.J., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates, Inc.: New York, NY, USA, 2012.
39. Borgwardt, K.M.; Gretton, A.; Rasch, M.J.; Kriegel, H.-P.; Schölkopf, B.; Smola, A.J. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics* **2006**, *22*, e49–e57. [CrossRef] [PubMed]
40. Sun, B.; Saenko, K. Deep coral: Correlation alignment for deep domain adaptation. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 443–450.
41. Long, M.; Cao, Y.; Wang, J.; Jordan, M. Learning transferable features with deep adaptation networks. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 97–105.