applied sciences

MDPI

Article

MCRCNet: A Bearing Fault Diagnosis Method for Unknown Faults Based on Transfer Learning

Guangyuan Xu [1,2,3], Ruifeng Guo [1,2,3], Zhenyu Yin [1,2,3,*] and Feiqing Zhang [1,3]

[1] Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China; xuguangyuan18@mails.ucas.ac.cn (G.X.); grf@sict.ac.cn (R.G.); qing34120212@163.com (F.Z.)
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] Liaoning Key Laboratory of Domestic Industrial Control Platform Technology on Basic Hardware & Software, Shenyang 110168, China
* Correspondence: congmy@163.com

Abstract: Bearing fault diagnosis in actual working conditions often faces the problem that unknown type faults cannot be identified, which seriously restricts the practical application of fault diagnosis technology. To solve this problem, this paper proposes a bearing fault diagnosis method based on transfer learning. Firstly, this paper designs a feature extraction network, the Multi-scale Convolution-Convolutional Reconstruction Network (MCRCNet), which incorporates a multi-scale feature extraction module to extract bearing fault features at multiple scales, thereby enhancing the extraction ability of key information. Secondly, this paper designs an improved convolutional reconstruction module AcConv (Adaptive Convolution reconstruction), which highlights key feature information and reduces redundant features by reconstructing the feature map. Furthermore, this paper also modifies the loss function to improve the performance in the case of data imbalance, and introduces the Wasserstein distance to optimize the adversarial training process. The proposed method is experimentally verified on Case Western Reserve University, Jiangnan University, and laboratory datasets. The experimental results show that the method has good performance in most tasks and has good generalization ability, which provides a feasible solution for the research of bearing fault diagnosis.

Keywords: fault diagnosis; transfer learning; convolutional reconstruction; loss function; Wasserstein distance

check for updates

Received: 27 December 2024
Revised: 12 January 2025
Accepted: 14 January 2025
Published: 18 January 2025

Citation: Xu, G.; Guo, R.; Yin, Z.; Zhang, F. MCRCNet: A Bearing Fault Diagnosis Method for Unknown Faults Based on Transfer Learning. *Appl. Sci.* **2025**, *15*, 921. https://doi.org/10.3390/app15020921

Copyright: © 2025 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

In recent years, research on mechanical fault diagnosis based on computer technology has been receiving increasing attention [1–3] as the key components of machinery and equipment, bearings, and their running state directly affect the performance and reliability of the entire system [4–6]. Bearing fault diagnosis is of great significance in industrial production as it can detect potential problems in time, prevent sudden equipment shutdown, reduce maintenance costs, and enhance production efficiency. Accurate fault diagnosis can guide maintenance personnel to take targeted measures to extend the service life of the equipment and ensure production safety [7–9].

The bearing fault diagnosis approach based on computer technology is capable of achieving precise fault recognition by rapidly processing vast amounts of data and applying sophisticated algorithms. Moreover, it can minimize the impact of human factors and enhance the consistency of diagnosis. These advantages significantly elevate the level of equipment management and maintenance decision making, thereby playing a crucial role in enhancing industrial production efficiency.

In recent years, deep learning, given its advantages in feature extraction, has been increasingly incorporated into the domain of bearing fault diagnosis. Relevant researchers have also put forward various enhancement approaches to enhance the feature extraction capacity by ameliorating the feature extraction network, thereby enhancing the accuracy rate of bearing fault diagnosis [10,11]. Zhang et al. [12] incorporated the extended convolutional network into the domain of bearing fault diagnosis. Through augmenting the perception field of the convolution kernel and appending the global residual block, the accuracy of bearing fault diagnosis in a noisy environment was enhanced. Deng et al. [13] introduced a network architecture that integrates information across multiple granular levels to effectively diagnose and locate faults in multi-bearing systems using supplementary bearing vibration data. The experiments they conducted revealed that the incorporation of multi-granularity fused information significantly boosts the system's ability to extract and identify features from intricate signals and improves its resilience to noisy environments.

In addition to the above methods, Wang et al. [14] developed a reinforcement learning-based methodology focused on enhancing deep features for diagnosing faults in rolling bearings. Their technique leverages the Elu activation function in conjunction with a convolutional attention mechanism, which facilitates the extraction of vital global information and strengthens the capability to enhance feature representation. Tang et al. [15] put forward a bearing fault diagnosis approach based on Fault Characteristic Region (FFR) detection. The crux lies in establishing an FFR proposal network for multiple fault characteristic regions from the training samples to achieve the diagnosis of composite fault types. Tang et al. [4] proposed a trusted multi-scale secondary neural network with an embedded attention mechanism for bearing fault diagnosis. By introducing secondary neurons featuring an embedded attention mechanism to optimize the convolution operation, the feature extraction capability of the standard CNN was enhanced. Peng et al. [16] proposed a diagnosis method based on Convolutional Neural Network Self-feature Fusion and Transformer Data Transformation, which converts heterogeneous data into a unified format through the global self-attention mechanism of the transformer and automatically focuses on important information in different input paths. Thus, fault features can be extracted more comprehensively and accurately. Although the above methods based on deep learning have achieved satisfactory results, there are notable flaws in their application. Firstly, the training data and test data are required to adhere to the same distribution, meaning that all types and all conditions of load data need to be encompassed in the training data to achieve accurate fault diagnosis. Nevertheless, this is challenging to accomplish under actual working conditions, so researchers have commenced attempts to achieve the diagnosis of unknown working conditions and unknown types of faults through other approaches.

Researchers have widely adopted transfer learning in the realm of bearing fault diagnosis due to its effective cross-domain adaptation capabilities. This approach involves segregating data into two distinct categories: the source domain and the target domain. In the context of our study on bearing fault diagnosis, the source domain comprises data that are accessible for training models, while the target domain includes data that must be analyzed and diagnosed under real-world conditions. Liu et al. [17] conducted supervised classification on source domain data and implemented batch nuclear norm maximization on target domain data to mitigate label offset effects. Tian et al. [18] proposed a new adversarial transmission network guided by class aggregation (ATN-CA) for the state diagnosis of bearings with small samples. The introduced local difference feature representation enables ATN-CA to concentrate on the distinctive characteristics of bearings, preventing a single neural network from overlooking crucial fault information. Additionally, the proposed class aggregation guidance strategy leverages the semantic information from the signals to dynamically adjust both the marginal and conditional distributions of the source and target

datasets. This adjustment effectively reduces the distribution gap for the same class across different domains, thereby facilitating accurate transfer diagnosis. Li et al. [19] developed a deep transfer learning model that integrates multi-domain information fusion with Multi-kernel Maximum Mean Difference (MK-MMD). Their approach includes constructing a multi-domain feature extractor using a multi-attention mechanism and performing feature integration through a Bidirectional Gated Recurrent Unit (BiGRU) network. The MK-MMD component is employed to minimize the distribution gap between the source and target domains, thereby achieving domain-invariant feature representations.

The aforementioned methods have achieved good results by incorporating transfer learning techniques into fault diagnosis. At the same time, some researchers have also proposed fault diagnosis methods based on transfer learning. Zhong et al. [20] incorporated the finite element simulation data of bearing faults into the domain adversarial network and used the fault evolution mechanism to guide the adversarial training. The proposed algorithm establishes both global and subdomain classifiers, substituting the label predictions for target domain data within each subdomain with simulated signals to minimize information transfer. By redesigning the loss function, it effectively extracts shared features of identical bearing types across various operating conditions, thereby enhancing robustness against domain adversarial challenges. Saha et al. [21] used the VGG16 model based on transfer learning as the feature extractor and the random forest as the classifier to realize the adaptation to different working environments and enhance the fault classification performance under different working conditions. Xing et al. [22] introduced the Dual-Domain Signal Cooperative Transmission Network (DSCTN) for diagnosing bearing faults across different domains under variable speed conditions. The network is built on the Swin Transformer architecture and utilizes dual-domain signals. They developed a multi-level fusion feature extraction module that effectively captures essential features from both time and frequency domains, thereby diminishing the effects of speed fluctuations. Liu et al. [23] introduced a fault diagnosis system tailored for cross-domain applications by integrating a deep transfer convolutional neural network with a supervised joint matching strategy. The process begins with applying continuous wavelet transform to the raw vibration signals from bearings, converting them into time-frequency representations. Subsequently, they fine-tuned a deep transfer convolutional neural network to adeptly extract complex features across different domains. To address domain adaptation challenges, the researchers developed a novel supervised joint matching method, which combines the alignment of feature distributions with instance reweighting based on the maximum margin criterion. Finally, they trained an intelligent diagnostic model capable of accurately predicting the labels of feature data in the target domain.

The above methods based on transfer learning are aimed at the situation where the fault types in the source domain and the target domain are the same. However, in the actual working conditions, there are often some unknown fault types, that is to say, the fault data cannot be obtained, which will cause the fault types in the target domain that do not exist in the source domain. At the same time, there will be some fault types in the source domain that do not occur in the actual working conditions. This problem can be abstracted as the Open Set Domain Adaptation problem in transfer learning, which was originally highlighted by Busto et al. [24]. This problem reflects the limitations of traditional closed set domain adaptation for image classification, and a method suitable for both closed set and open set scenarios was proposed. Saito et al. [25] proposed Open Set Domain Adaptation by Backpropagation (OSDABP). Fu et al. [26] improved the distance calculation method for the core problem and proposed Improved Open Set Doamin Adaptation With Backpropagation. These methods are aimed at the field of image recognition. Wang et al. [27] proposed a method for OSDA problems in the field of bearing fault diagnosis, which involved a new

feature separation method. The above methods only adjust the domain adaptation methods and do not propose suitable feature extraction methods for this problem.

Based on the previous research, this paper studies the bearing fault diagnosis method using transfer learning to solve the problem of low accuracy in recognizing unknown fault types. The main contributions of this paper are as follows:

1. Construction of a bearing fault diagnosis model. By constructing a fault diagnosis model of unknown type bearing based on transfer learning and defining the problem as an Open Set Domain Adaptation problem, this paper solves the problem that unknown type bearing faults cannot be classified, effectively improves the generalization ability of the model, and improves the accuracy rate of fault diagnosis.
2. Design of a network structure based on multi-scale feature extraction and an improved convolutional reconstruction module. Through the extraction of fault features of different scales and the reconstruction of bearing data feature maps, the extraction efficiency of key features is improved, and the extraction of redundant features is reduced, which ultimately improves the efficiency of transfer learning.
3. Design of an improved loss function. By improving the structure of the loss function, the performance of the proposed method is improved under the condition of data imbalance, and the Wasserstein distance is introduced to improve the efficiency of adversarial training.

2. Preliminaries

2.1. Reconstruction Convolution

The convolution operation of deep learning shows its advantages in feature extraction, but the traditional convolution method has large redundancy, which causes the model to extract too many redundant features, and these redundant features also affect the final accuracy of the model. Therefore, researchers have begun to try to improve the convolution operation in order to achieve the purpose of reducing redundant features in the spatial and channel dimensions. The convolutional reconstruction method based on space and channel reconstruction is a typical representative. This method has two parts, a spatial reconstruction unit and a channel reconstruction unit, and enhances the feature expression ability by operating in two aspects: space and channel, as shown in Figures 1 and 2.

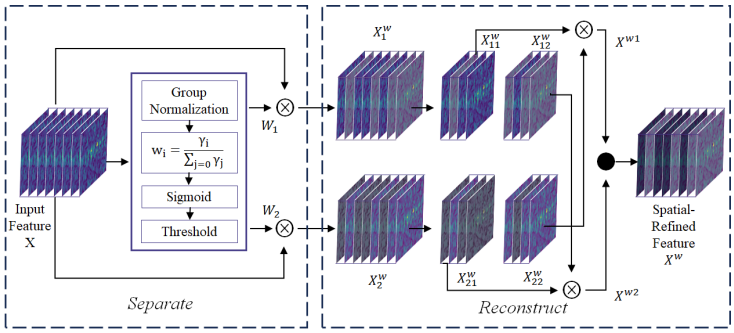


Figure 1. Spatial reconstruction unit.

The spatial reconstruction unit primarily addresses the issue of spatial feature redundancy. The method it employs involves using group normalization to evaluate the information content of the feature maps and separates the feature maps based on the amount of information. Specifically, given an intermediate feature map $X \in R$, the input feature X is first normalized by subtracting the mean μ and dividing by the standard deviation σ, as shown in the following equation:

$$X_{out} = \mathrm{GN}(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta. \tag{1}$$

In the equation, the mean and standard deviation of X are denoted by μ and σ. ε is a small positive constant. The purpose of adding it is to make the division more stable. For affine transformations, we use γ and β, which are trainable. For the measurement of pixel variance, we use the parameter γ from the GN layer. A larger value of this parameter indicates richer spatial information. Subsequently, through the sigmoid function, the re-weighted feature map is mapped within the range (0, 1). The gating method uses a threshold: values greater than the threshold are set to 1, corresponding to $W1$; otherwise, to $W2$. The specific process is described in the following equation:

$$W = \mathrm{Gate}(\mathrm{Sigmoid}(W_\gamma(\mathrm{GN}(X)))). \tag{2}$$

Finally, the input feature X is multiplied by $W1$ and $W2$, respectively, to obtain two weighted features: the feature with more information X_1^ω and the feature with less information X_2^ω. In this way, the input features are successfully divided into two parts: X_1^ω has informative and expressive spatial content, while X_2^ω has almost no information and is considered redundant.

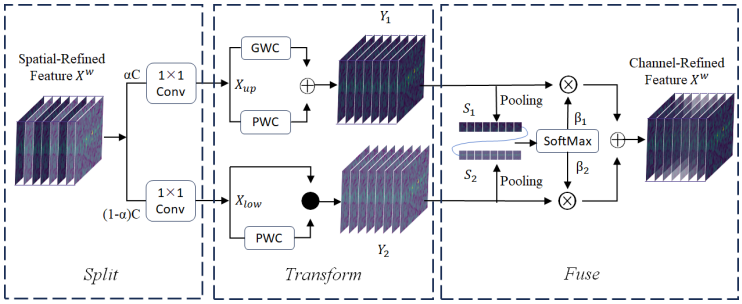


Figure 2. Channel reconstruction unit.

The reconstruction operation process adds the informative features to the less informative ones to generate more informative features, thus saving space. The cross-reconstruction operation is used to fully combine the two different weighted information features to enhance the information flow between them. Then the cross-reconstructed features $X^{\omega 1}$ and $X^{\omega 2}$ are concatenated to obtain the spatially fine feature map X^ω. The subsequent calculation process is shown in the equation below

$$\begin{cases} X_1^\omega = W_1 \otimes X \\ X_2^\omega = W_2 \otimes X \\ X_{11}^\omega \oplus X_{22}^\omega = X^{\omega 1} \\ X_{21}^\omega \oplus X_{12}^\omega = X^{\omega 2} \\ X^{\omega 1} \cup X^{\omega 2} = X^{\omega 1} \end{cases}, \tag{3}$$

where $\otimes$ is multiplication by element, $\oplus$ is summation by element, and $\cup$ is concatenation. After the spatial reconstruction unit is applied to the intermediate input feature X, it not only separates the information features from the features with less information, but also reconstructs them to enhance the representative features and suppress the redundant features in the spatial dimension. However, the spatial fine feature map X^ω is still redundant in the channel dimension.

After the spatial reconstruction unit, the channel reconstruction unit is added to realize the reconstruction of features by segmenting, transforming, and fusing features from the perspective of channels. The segmentation operation splits the input spatial refined feature

X^{ω} into two parts—one part is the channel number αC, the other part is the channel number $(1-\alpha)C$—and then the number of channels of the two groups of features is compressed by the 1×1 convolution kernel to obtain X_{up} and X_{low}, respectively. The transformation operation takes the input X_{up} as the input of "rich feature extraction", performs GWC (Gate-Way Convolution), and PWC (Point-Wise Convolution), respectively, and then adds to obtain the output $Y1$. The input X_{low} is used as the supplement of "rich feature extraction", performs PWC, and obtains $Y2$ by taking the union of the obtained result and the original input. The fusion operation uses a simplified SKNet approach from adaptation to merge $Y1$ and $Y2$. Specifically, the global average pooling is first used to combine the global spatial information and channel statistical information to obtain the pooled $S1$ and $S2$. Then, Softmax is performed on $S1$ and $S2$ to obtain the feature weight vectors β_1 and β_2. Finally, the feature weight vectors are used to obtain the output $Y = \beta_1 \times Y_1 + \beta_2 \times Y_2$, and Y is the extracted feature of the channel. A schematic diagram of the channel reconstruction unit is as follows.

After completing the design of the two units, the two units are fused and placed in a sequential manner. Specifically, for the intermediate input feature X in the bottleneck residual block, the spatial refinement feature X^{ω} is first obtained by the spatial reconstruction operation, and then the channel refinement feature Y is obtained by the channel reconstruction operation. This method makes full use of the spatial redundancy and channel redundancy between features, and finally, reduces the redundancy between the intermediate feature maps and enhances the feature representation of the convolution operation.

2.2. Open Set Domain Adaptation by Backpropagation

Traditional transfer learning is faced with the closed set recognition task, that is, the source domain and the target domain contain exactly the same data types. However, in practical problems, it is more often the case that the target domain contains the target types that do not exist in the source domain, and the data types in the source domain do not all appear in the target domain. Therefore, the open set domain adaptive method is proposed. The method mainly faces two problems: one is that the unknown class cannot be separated from the target domain due to lack of prior knowledge; the other is that the difference between the source domain and the target domain cannot be solved due to drift, and the domain feature alignment method cannot be used to solve the problem.

Among these solutions, Open Set Domain Adaptation by Backpropagation is very representative, and its core idea is to separate known classes and unknown classes in the target domain. The feature extractor wants the data distribution to be consistent, while the classifier wants to separate the target domain and source domain data, and classify the unknown category into the same class. Obviously, there is an adversarial relationship between the feature extractor and the classifier.

Therefore, the method uses an adversarial training approach to obtain a classifier and a generator, where the classifier is trained to form a boundary between the source and target samples, while the generator is trained to keep the target samples away from the boundary. And the classifier is set to an output $K+1$ dimensional vector, where K represents the classification probability of the known class, and the last dimension refers to the probability of the unknown class, that is, to unify all unknown types into one class.

In the field of bearing fault diagnosis, this problem can be described as follows: the source domain and the target domain contain different types of bearing faults, among which different types can include different working conditions, different fault types and different fault size. In actual working conditions, only partial bearing fault data can be obtained in most cases. Sometimes it is possible to obtain all fault class data, but it is

difficult to obtain all fault size and working conditions for each fault class, and sometimes it is impossible to obtain all fault class data.

We explain this problem more clearly in Figure 3, where different shapes represent different fault categories, different colors represent different fault sizes, and the color shades represent load conditions. Among them, the four figures a, b, c, and d show the four cases where the source domain and target domain contain the same fault type, the target domain contains unknown fault class, the target domain contains an unknown fault size, and the target domain contains both an unknown fault class and fault size. To solve this practical problem in the field of bearing fault diagnosis, we use the OSDABP method to solve and improve it.

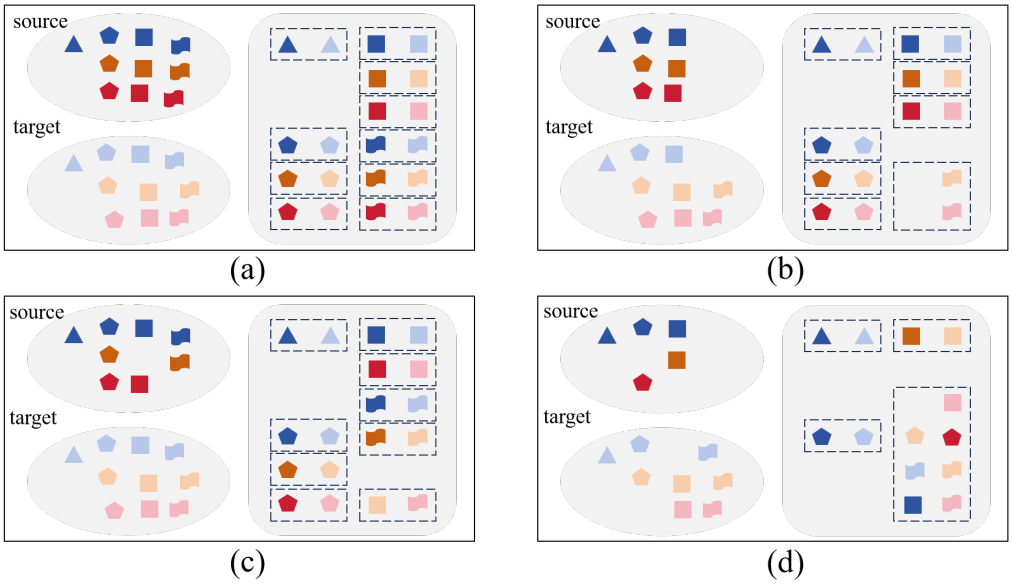


Figure 3. Open Set Domain Adaptation by Backpropagation. (**a**) source domain and target domain contain the same fault type; (**b**) target domain contains unknown fault class; (**c**) target domain contains an unknown fault size; (**d**) target domain contains both an unknown fault class and fault size.

3. Proposed Method

3.1. Overview of the Proposed Model's Architecture

Based on the above analysis of unknown bearing fault diagnosis problems and the research into the existing method, we propose a method of unknown type bearing fault diagnosis based on transfer learning: the Multi-scale Convolution-Convolution Reconstruction Network. The overall structure of the method is shown in Figure 4, including data acquisition and preprocessing, feature extraction, loss function, and diagnostic result display.

Specifically, firstly, the data collected by the sensor are preprocessed, and then the feature extraction network is used to extract the fault features and reconstruct the feature map, and the training process is optimized based on the improved loss function. This section is the core of the model, comprising a multi-scale feature extraction module and a convolutional reconstruction module. In the multi-scale feature extraction module, based on the characteristics of bearing vibration signals under different operating conditions, we selected four convolutional modules with scales ranging from 1 to 7 to extract features at various scales. This enables our method to accurately distinguish different fault locations and fault diameters, facilitating the diagnosis of unknown fault types.

In the convolutional reconstruction module, we reconstruct convolutions in both spatial and channel dimensions to reduce the extraction of redundant features and prevent features generated by changes in load and rotational speed from being mistakenly extracted.

This enhances the model's generalization capability. Additionally, we have improved the loss function to increase the training efficiency under conditions such as class imbalance and optimized the adversarial target generation process during backpropagation. Finally, the collected bearing vibration data are detected to realize the fault identification.

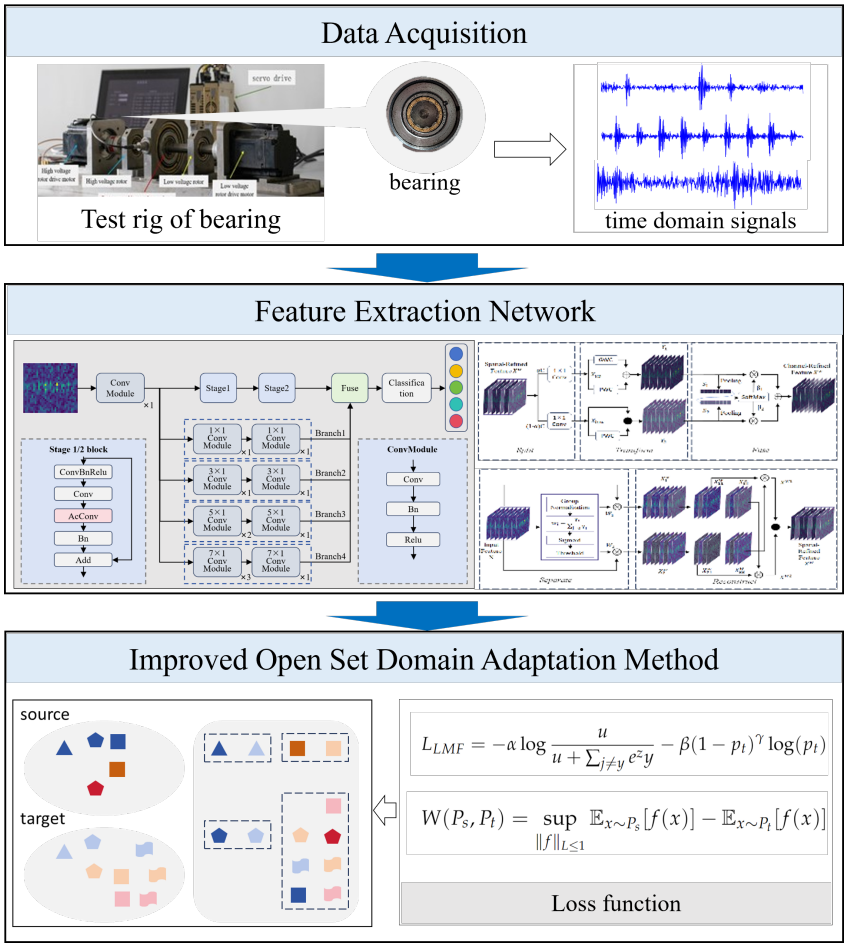


Figure 4. Overall framework structure of the proposed model.

3.2. Improved Feature Extraction Network

In transfer learning, higher requirements are put on the feature extraction ability of deep learning models. This is because the core of methods such as the domain adaptive method in transfer learning is to minimize the difference between the distribution of features in the source domain and the target domain through adversarial domain obfuscation branching, that is, to make it as difficult as possible to distinguish between the features of the two domains, and ultimately to achieve a feature extractor that can have a better representation of both domains. In order to achieve this goal, it is necessary to extract as many effective features as possible and reduce redundant features in the feature extraction stage.

The bearing vibration signal has its own characteristics. Its features include the amplitude intensity of the spectral signal, characteristic peak value, frequency band range, and other types; different layers may learn different features. At the same time, different bearings have different mechanical properties, and their feature sizes are also different. The impact of these characteristics on the feature acquisition process mainly exists in two aspects: one is that different features in the scale of the differences in the size of the convolution kernel also bring more complex requirements, and the second is that there are some differences in the importance of the signal features on different channels.

Based on the above analysis, we design a feature extraction method that is more suitable for transfer learning, which mainly consists of two core ideas: one is to introduce a parallel multi-scale feature extraction structure, which enhances the extraction of features at different scales through multi-scale convolution kernels to strengthen the adaptability of features at different scales; the other is to introduce an improved convolutional reconstruction method, which reconstructs the features from the spatial and channel dimensions, to highlight important features. At the same time, we streamline the number of network layers to avoid overfitting, that is, we try to extract the common features of the bearings. Our feature extraction method is described in detail below.

Our feature extraction network is based on the ResNet architecture, and its core mainly consists of a basic convolutional module and a parallel multi-branch structure including multi-scale convolutions. The basic convolutional module consists of two convolutional layers, batch normalization, ReLU activation, and AcConv modules, and includes a residual join structure. The convolution layer in the module realizes feature extraction, and the AcConv module reconstructs the extracted features to highlight the main features.

Our proposed AcConv module is improved on the basis of the original ScConv module. The main improvement is that in the channel reconstruction part, the group convolution (GWC) is improved, and the convolution operation with variable convolution kernel size is modified to better adapt to different feature scales. At the same time, for the space reconstruction part, it is more important to set the feature map discrimination threshold as a learnable parameter to distinguish those features better. The multi-branch part of multi-scale convolution mainly contains four independent branches, with the four branches existing in parallel with the basic convolution module. Branch 1 uses convolution with convolution size 1 to capture simple features, Branch 2 uses convolution size 3 to focus more on moderately complex features, and branches 1 and 2 contain only one convolutional layer. Branches 3 and 4 use convolution kernels of size 5 and 7, respectively, to extract more complex features. Meanwhile, to enhance the feature extraction capability, we use two convolution kernels in branches 3 and 4, respectively, and add skip joins.

The overall structure of the network is an efficient combination of two core contents. First, feature extraction is carried out through the initial convolutional layer, and the processed data are entered into two basic convolutional modules and parallel multi-branch structure as input features, respectively. Then, feature fusion is carried out for five output features, in order to better improve the overall feature expression ability and highlight important features. In the process of feature fusion, weighted fusion is adopted to set the weight parameters for the output of each branch and the output of the basic convolutional block and perform weighted summation, so that the fused features contain multi-scale and multi-level information, which effectively improves the network performance. The specific structure of our network is shown in Figure 5.

3.3. Improved Open Set Domain Adaptation Method

For the open set domain adaptation problem, one of our goals is to correctly classify the known target samples into the corresponding known class, and the second goal is to train the classifier to draw a boundary between the unknown and known target samples. For the first objective, we use LMFLoss as the loss function. In the actual working condition, there are differences in the frequency of each bearing fault, and the amount of different types of fault data that can be collected is not the same when the fault occurs, so there is bound to be data imbalance. In order to better realize the correct classification of imbalanced known samples, we improve the loss function. By setting separate weights for different classes, it strengthens the learning of the minority class samples. We adopt LMFLoss for implementation, which jointly uses two methods: Focal Loss and LDAM Loss. Focal Loss

solves the problem of unbalanced class distribution by dynamically adjusting the loss weight of training hard and easy samples. Its core idea is that in classification tasks, most samples can easily be correctly classified, and only a few hard samples need more attention from the model. Therefore, Focal Loss introduces a dynamic adjustment factor to reduce the loss contribution of easy to classify samples and increase the loss contribution of difficult to classify samples, so as to achieve the purpose of the focused training model. The formula is as follows:

$$Focal_{Loss} = -(1 - p_t)^{\gamma} \log(p_t), \tag{4}$$

where, p_t is the probability that the sample belongs to the true class, and γ is the adjustment factor that controls the loss weight of easy/hard samples

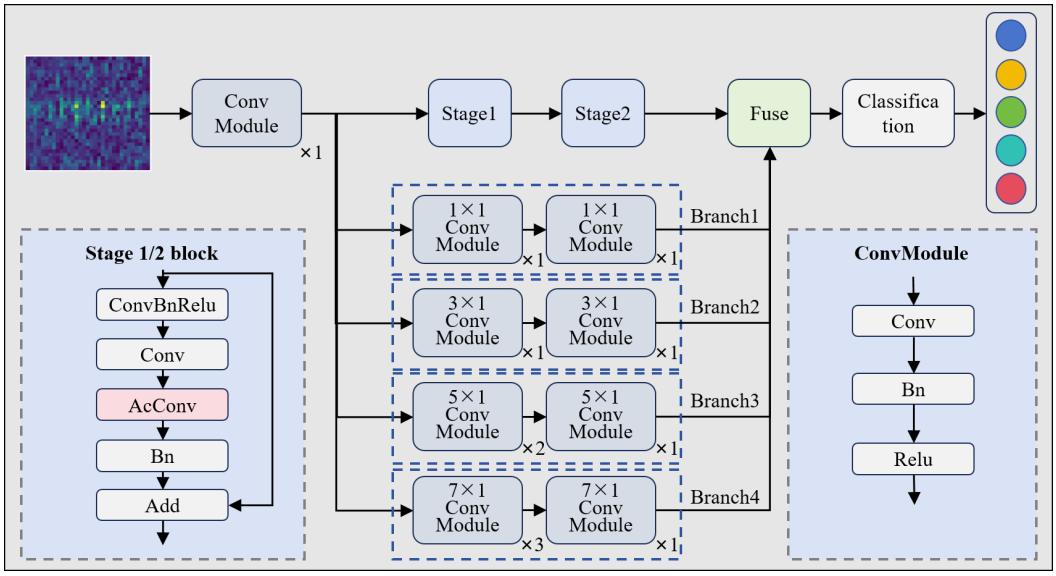


Figure 5. Improved feature extraction network.

LDAM Loss sets different decision boundaries for different classes from the perspective of label distribution adaptation. The specific method includes two parts. The first part is to set a label-aware boundary, which sets a different decision boundary margin for each class, using a smaller margin for the minority class and a larger margin for the majority class. In this way, the minority class samples are easier to classify correctly. The second part, adaptive loss scaling, also introduces an adaptive loss scaling factor that is inversely proportional to the frequency of the class to which the sample belongs. In this way, the loss contribution of the minority class samples can be increased, and the model can learn these categories better. The equation is as follows:

$$L_{LDAM}((x, y), f) = -\log \frac{u}{u + \sum_{j \neq y} e^{z} y}. \tag{5}$$

The LMF Loss we used combines the two loss functions and makes full use of the prediction probability information and label distribution information of samples, which enables the model to better mine the valuable information contained in the data, thus improving the overall classification performance. At the same time, these two strategies can also enhance each other and form a synergistic effect. The formula is as follows:

$$L_{LMF} = -\alpha \log \frac{u}{u + \sum_{j \neq y} e^{z} y} - \beta(1 - p_t)^{\gamma} \log(p_t). \tag{6}$$

For the second goal, we use the idea of generative adversarial training to achieve the classification of unknown classes. The purpose of this process is to train a classifier to draw a boundary between unknown and known target samples. To achieve this goal, we need to measure the difference between the source domain and the target domain. The objective function based on the Wasserstein distance can well measure the distance between the two distributions when there is no overlap or little overlap between the source domain and the target domain, and effectively prevent the phenomenon of gradient disappearance. Its formula is as follows, where P_s and P_t denote the data distribution in the source and target domains, respectively:

$$W(P_s, P_t) = \sup_{\|f\|_{L\leq 1}} \mathbb{E}_{x\sim P_s}[f(x)] - \mathbb{E}_{x\sim P_t}[f(x)]. \quad (7)$$

4. Experiments

To validate the effectiveness of our proposed method, we selected three different datasets, including the CWRU dataset, the JNU dataset, and a dataset collected in our laboratory, for verification. This was done to demonstrate that our proposed model exhibits excellent performance under various conditions, indicating good generalization capabilities. The mechanical structures of the equipment, data collection frequencies, speeds, and loads of these three datasets are all different. In the experimental design, we focused on different fault locations and fault diameters as the basis for classifying fault types. Factors such as the rotational speed and load were used as interference factors to verify the generalization ability of our model under different conditions. Below, we will describe our experimental process in detail, as well as our analysis of the experimental results.

4.1. Dataset Description

1. CWRU dataset
   We choose the CWRU [28] dataset, a widely used dataset in the field of bearing fault diagnosis, to verify our proposed method. The experimental data are collected from the experimental platform, which is driven by a motor and sampled by acceleration sensors mounted on the bearing at the driving end and the fan end. In our experiment, we use the data of the driving end bearing with a sampling frequency of 12 kHz, which includes three fault categories, namely, the normal condition (N) and outer race fault (OF), inner race fault (IF), and ball fault (BF), where each fault category contains three damage levels of 0.007, 0.014, and 0.021 inches, respectively, The experimental data are collected under four different load conditions of 0–4 hp. These experimental conditions can simulate well the actual working conditions of different loads, different fault types, and different fault damage. The specific data labels are shown in Table 1.

Table 1. CWRU dataset information.

Label	Bearing Status	Fault Diameter (inches)
0	N	-
1	BF	7
2	BF	14
3	BF	21
4	IF	7
5	IF	14
6	IF	21
7	OF	7
8	OF	14
9	OF	21

2. JNU dataset
The bearing dataset of Jiangnan University [29] is designed specifically for bearing fault diagnosis and predictive maintenance research. The test bed is mainly composed of an acceleration sensor, motor, bearing, and conveyor belt. The dataset is collected at a sampling rate of 50 kHz and contains three different speed conditions of 600 RPM, 800 RPM, and 1000 RPM. The data types are divided into inner ring fault (IF), outer ring fault (OF), rolling element fault (BF), and normal state (N) to support various fault diagnosis and machine learning research scenarios. In the experiment, all faults were also cut manually. In this paper, we use the data of 600 and 1000 revolutions to conduct experiments. The specific data labels are shown in Table 2.

Table 2. JNU and VMS dataset information.

Label	Bearing Status
0	N
1	IF
2	OF
3	BF

3. VMS dataset
To address the insufficiency in data types of public datasets and to better verify the methods we proposed, we designed the Variable Mechanical Structure Bearing Experimental Platform and carried out data collection on it. The experimental platform mainly encompasses a motor drive system, rotor system, and support system. Among them, the motor drive system employs two SGMJV-04ADD6S AC servo motors produced by Yaskawa of Japan, with a rated torque of 1.27 N·m, a power of 1.2 kW, a rotational speed of 3000 r/min, and a current of 2.7 A. The servo unit adopts SGDV-2R8A01 produced by Yaskawa of Japan, with a power supply of 3PH AC 200–230 V 50/60 Hz, 2 A. The rotor system consists of three parts, namely, the high-speed rotor, low-speed rotor, and differential rotor. In this experiment, we utilized the data of the differential rotor. This part selects the bearing model NU202EM, which is a type of cylindrical roller bearing, mainly constituted by the inner ring, outer ring, and cylindrical rolling elements between them. To simulate different fault types of the bearing, electrical discharge machining technology was used to cut the bearing, including inner ring faults, outer ring faults, and rolling element faults. When collecting data, we set the sampling frequency uniformly at 51.2 kHz and set the rotational speeds of the two groups of data to 677 RPM and 1797 RPM, respectively. Compared with the experimental platforms of the CWRU dataset and the JNU dataset, the position of the bearing to be tested on our experimental platform is farther from the drive end and passes through the support structure. Its mechanical structure can better simulate the characteristics of some more complex mechanical equipment, and the rotational speed differences of the two groups of collected data are greater, enabling a wider range of simulation of actual working conditions. This experimental platform is shown in Figure 6. The specific data labels are shown in Table 2.

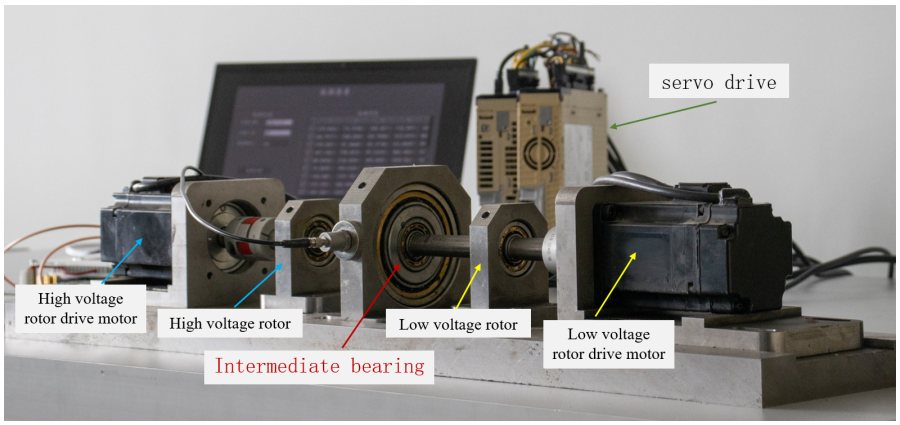


Figure 6. Variable Mechanical Structure Bearing Experimental Platform.

4.2. Implementation Details and Tasks

To make the experimental tasks conform to actual working conditions and validate our proposed method better, we created different tasks for the three adopted datasets. For the CWRU dataset, 12 tasks were designed considering load changes and fault types. The source and target domains were set to include different fault types, covering various situations in actual working conditions. Also, different load conditions were set for the two domains to reflect load fluctuations in actual operation. The specific task settings are shown in Table 3. We can evaluate the model's performance in different tasks, and the difference between adjacent tasks lies only in the load to further verify the impact of load changes on the method's performance.

The change in the bearing speed will also have a certain impact on the fault characteristics of the bearing. In order to verify the generalization ability of our model, we verify the performance of our proposed method under variable speed conditions on the JNU dataset. The specific task settings are shown in Table 3, which is divided into 10 tasks. The fault types contained in the source domain and the target domain are different, and the speed is also different. Through the performance comparison on the task with the same fault type setting, the influence of speed change on the performance of the method can be analyzed.

At the same time, to further verify the detection ability of bearings in machinery of different structures, we chose the support bearings on the VMS experimental platform in the lab for further experimental verification. The specific task settings are similar to the JNU dataset, but with larger speed differences we set to further verify the performance under variable speed conditions. The specific task settings are shown in Table 3.

4.3. Comparison Methods

In order to reflect the advancement of our proposed method, we chose DANN [30], OSDAFD [31], BWAN [32], UDA [33], EVT-OSFD [32], DAN-COSFD [34] and other methods for comparison. These methods are based on transfer learning methods, and a brief introduction to each method is as follows:

1. DANN: A classical transfer learning approach, based on adversarial neural networks, and oriented to domain adaptation problems.
2. BWAN: A representative method of open set domain adaptation, using extreme value theory to reject unknown class samples.
3. UDA: Universal domain adaptation is a classical open set data mining method. The domain similarity and prediction uncertainty of each sample are used to achieve shared class adaptation and unknown class detection.

4. EVT-OSFD: This method uses the output of additional domain classifiers and proposes a bilateral weighted adversarial network model, which assigns large weights to shared classes and small weights to outlier classes in the feature alignment process.
5. OSDAFD: An open set domain adaptive method for rotating machine fault diagnosis, which uses a state classifier to identify known machine health states and a binary classifier to identify unknown machine health states.
6. DAN-COSFD: An adaptation network based on dual adversarial learning. The auxiliary domain searcher assigns similarity weights to each target sample to distinguish between known and unknown fault modes, and employs weighted adversarial learning to selectively adapt the domain distribution.

Table 3. Task information of the datasets.

Dataset	Task	Source→Target (Motor Load) (hp)	Source→Target (Motor Speed) (rpm)	Source Label Set	Target Label Set
CWRU	C01	0→1	1797→1772	0, 1, 2, 3, 4, 5, 6	0, 1, 2, 3, 4, 7, 8, 9
	C02	0→2	1797→1750	0, 1, 2, 3, 4, 5, 6	0, 1, 2, 3, 4, 7, 8, 9
	C03	0→3	1797→1730	0, 1, 2, 3, 4, 5	0, 1, 2, 3, 4, 7, 8
	C10	1→0	1772→1797	0, 1, 2, 3, 4, 5	0, 1, 2, 3, 4, 7, 8
	C12	1→2	1772→1750	0, 1, 2, 3, 4, 5	0, 1, 6, 7
	C13	1→3	1772→1730	0, 1, 2, 3, 4, 5	0, 1, 6, 7
	C20	2→0	1750→1797	0, 1, 3, 5, 7, 9	0, 1, 2, 3, 4, 5
	C21	2→1	1750→1772	0, 1, 3, 5, 7, 9	0, 1, 2, 3, 4, 5
	C23	2→3	1750→1730	0, 1, 3, 5, 7	0, 1, 2, 3, 4
	C30	3→0	1730→1797	0, 1, 3, 5, 7	0, 1, 2, 3, 4
	C31	3→1	1730→1772	0, 2, 4, 6, 8	0, 2, 5
	C32	3→2	1730→1730	0, 2, 4, 6, 8	0, 2, 5
JNU	J01	-	600→1000	0, 1, 2, 3	0, 1, 2, 3
	J02			0, 1, 2	
	J03			0, 1, 3	
	J04			0, 1	
	J05			0, 2	
	J06			0, 1, 2	
	J07		1000→600	0, 2, 3	
	J08			0, 1	
	J09			0, 2	
	J10			0, 3	
VMS	V01	-	677→1797	0, 1, 2, 3	0, 1, 2, 3
	V02			0, 1, 2	
	V03			0, 1, 3	
	V04			0, 2, 3	
	V05			0, 1, 2, 3	
	V06		1797→677	0, 1, 2	
	V07			0, 1, 3	
	V08			0, 2, 3	

4.4. Experimental Performance and Analysis

We show the experimental results on three datasets in Tables 4–6. First, we analyze the experimental results on the CWRU dataset. On the whole, the accuracy of our proposed method is significantly higher than that of the existing methods, which shows that our method is superior to the existing methods in performance. Below we analyze the performance on specific tasks in detail. Existing methods perform relatively poorly on the three tasks C03, C20, and C30, which we first focus on analyzing. In terms of fault type

distribution, the main characteristics of these three tasks are as follows: (1) C03 is mainly characterized by containing different fault categories; (2) C20 and C30 contain different fault sizes of the same fault category. As we have introduced in the previous chapter, for the open-domain adaptation problem, the unknown type faults existing in the target domain should be classified into the same additional type. Therefore, we can analyze that in the C03 task, the existing methods fail to accurately identify the two unknown fault categories of label 7 and 8. And it is confused with the existing fault type. In the two tasks C20 and C30, the existing method incorrectly identifies the unknown fault size, which is confused with the known type. However, our proposed method achieves significant improvement in these three tasks. Therefore, our proposed method can accurately extract the characteristics of faults, especially the characteristics of different fault sizes, which is of great significance for judging the danger degree of faults. Further, we compare these three tasks with the corresponding tasks C10, C21, and C23 of the same fault type, and it can be seen that the accuracy still lags behind. The main difference in comparison is the load situation. In the case of a large load difference, the accuracy will be significantly reduced, and the accuracy of our proposed method will also be reduced. However, the reduction is very little, which confirms that our proposed method has better generalization ability for load changes in different working conditions.

Table 4. The classification accuracy (%) of the CWRU dataset.

Task	DANN	BWAN	UDA	EVT-OSFD	Ours
C01	69.4	70.8	79.9	83.8	97.3
C02	66.8	70.2	78.1	81.5	97.2
C03	73.8	72.8	70.8	78.6	97.8
C10	72.5	73.0	80.7	84.1	96.5
C12	68.2	71.0	84.6	90.7	99.8
C13	65.7	69.0	79.0	84.4	98.6
C20	65.3	65.4	70.5	82.9	97.9
C21	74.2	70.6	78.4	83.9	98.6
C23	74.6	74.6	78.4	87.3	97.6
C30	56.4	50.3	68.6	74.4	95.5
C31	76.3	82.4	71.3	94.9	99.8
C32	75.9	83.1	74.7	96.3	98.4
Avg	69.9	71.1	77.3	85.2	97.9

In the experiments on the JNU dataset, we focus on verifying the generalization ability of the proposed method for bearing fault diagnosis in the variable speed environment. It can be seen that in the vast majority of tasks, our proposed method still has the best effect, which fully proves the advancement of our method. Further analysis shows that the performance of our method on the last five tasks is better than for the first five tasks; the difference between the two types of tasks is mainly reflected in the difference between the speed change of the source domain and the target domain. The first five tasks are changed from a low speed in the source domain to a high speed in the target domain. The situation is opposite for the latter five tasks, so the bearing fault features in the target domain are enhanced compared with those in the source domain, which is more conducive to extracting the fault features. The performance of our proposed method is more stable while maintaining overall high accuracy, which confirms that our proposed method has good generalization ability in the variable speed environment.

Table 5. The classification accuracy (%) of the JNU dataset.

Task	DANN	OSDAFD	BWAN	DAN-COSFD	Ours
J01	95.96	75.22	74.15	79.52	90.52
J02	70.65	64.92	69.06	72.30	87.66
J03	64.22	67.58	57.74	82.00	89.24
J04	64.31	70.39	61.09	55.35	86.76
J05	65.76	74.55	84.44	93.02	87.40
J06	69.38	56.87	72.39	85.32	94.28
J07	70.60	57.85	65.98	87.59	95.52
J08	53.85	64.63	73.95	75.46	91.42
J09	66.00	68.43	77.53	90.73	94.19
J10	55.16	67.95	69.17	85.71	95.18
Avg	67.56	66.84	70.55	80.70	91.22

To prove the advancement and generalization ability of our proposed method, we verified it on the bearing test bed in the laboratory. Our method shows superior and stable performance. Compared with the JNU dataset, the speed variation is larger, but the accuracy fluctuation is smaller. First, our test bed's structure differs. Its drive shaft is directly connected to the motor, while the JNU test bed is indirectly driven, making our vibration signal more obvious. Second, differences in the selected bearings and man-made faults also affect fault characteristic performance. Overall, our model shows excellent generalization ability on different structures.

Table 6. The classification accuracy (%) of the VMS dataset.

Task	DANN	Proposed
V01	97.7	98.3
V02	72.2	96.4
V03	75.0	99.0
V04	71.5	98.5
V05	96.8	97.7
V06	70.5	98.1
V07	74.6	98.3
V08	71.4	96.8
Avg	78.7	97.9

Figures 7–9 show the performance superiority of our proposed method more clearly. In particular, it can be clearly seen that the performance fluctuation of our proposed method on multiple tasks is very small, which reflects the good adaptability of our proposed method to different task types.

At the same time, to present the experimental results of our proposed method more vividly, we show the results of running on the JNU dataset in the form of t-SNE plots in Figure 10. We can see that our method performs well in all 10 tasks. Further analysis shows that the faults labeled 2 and 3 are easily confused, and problems of identification errors or unclear boundaries of these two types of faults occur in multiple tasks, indicating that there is still room for improvement in the recognition of these two faults.

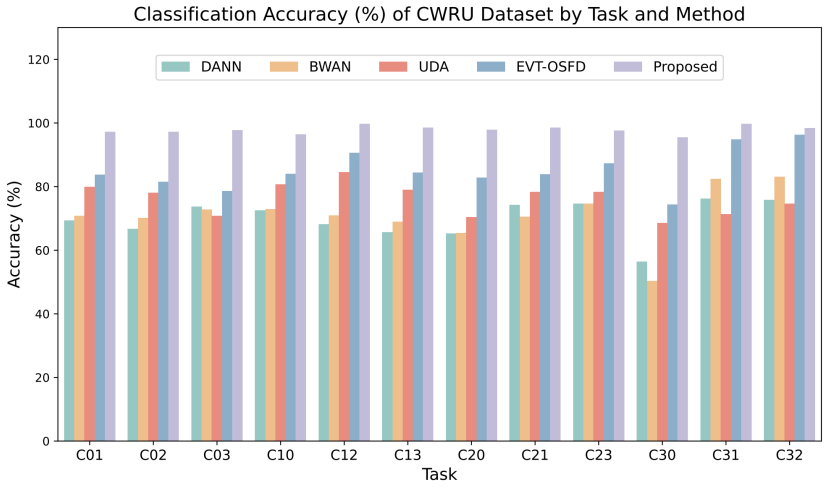


Figure 7. The accuracy of various methods applied to tasks in the CWRU dataset.

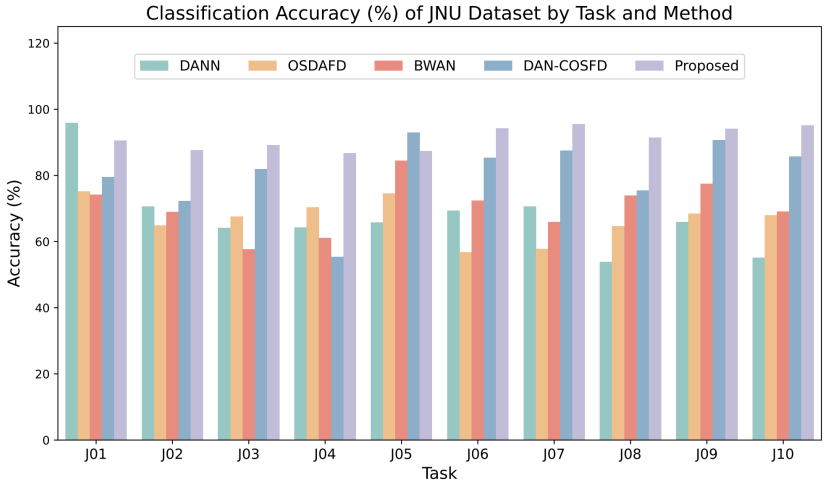


Figure 8. The accuracy of various methods applied to tasks in the JNU dataset.

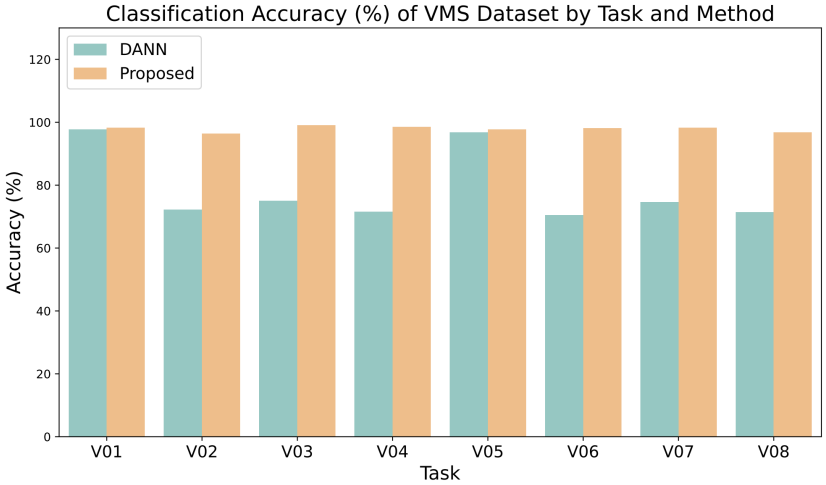


Figure 9. The accuracy of various methods applied to tasks in the VMS dataset.

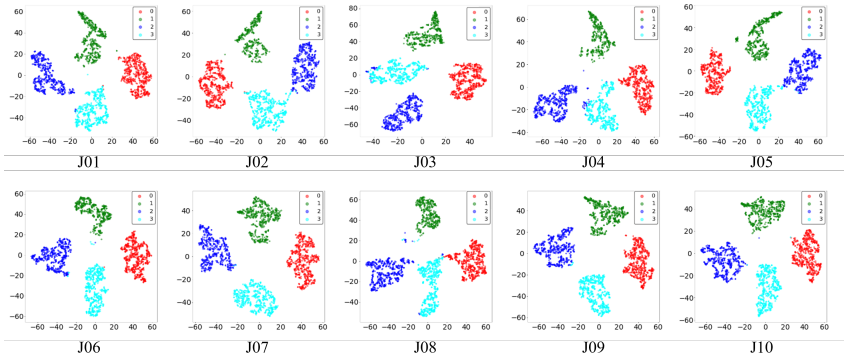


Figure 10. t-SNE plots for each task on the JNU dataset using the proposed method.

5. Conclusions

This paper proposes an unknown bearing fault diagnosis method based on transfer learning, which realizes the diagnosis of unknown faults by defining the problem as an open set domain adaptation problem. We introduce the multi-scale convolution module and AcConv module to design the MCRCNet feature extraction network, which realizes the extraction of key features of bearing faults at multiple scales and the elimination of redundant features, so as to better meet the high requirements of transfer learning for feature extraction. At the same time, in order to improve the performance of the method, we introduce an improved loss function, which not only improves the stability of the model under the condition of data imbalance, but also improves the efficiency of adversarial training, which significantly improves the robustness and generalization ability of the model. We validate our proposed method on the CWRU dataset, the JNU dataset, and the VMS dataset from our laboratory. Experimental results show that our proposed method achieves high accuracy in a variety of task conditions and is advanced compared with existing methods. In conclusion, this paper proposes a new and feasible research method for the unknown bearing fault field, which has a good application prospect.

Although the method proposed in this paper has been advanced, the problem we study concerning the fault diagnosis of a unified dataset or the same device remains, and we have not been able to extend the method to the field of unknown fault diagnosis across devices. Therefore, in future work, we will further optimize the transfer learning method and our proposed feature extraction network to achieve wider application in the field of bearing fault diagnosis. In addition, in the field of fault diagnosis, there is an extreme case where only normal bearing data can be obtained. In future work, we will also try to study how to achieve fault diagnosis in this extreme case.

Author Contributions: Conceptualization, G.X. and F.Z.; methodology, G.X. and R.G.; software, G.X. and F.Z.; validation, G.X.; formal analysis, G.X. and Z.Y.; investigation, G.X. and Z.Y.; resources, Z.Y. and R.G.; data curation, F.Z.; writing—original draft preparation, G.X.; writing—review and editing, G.X., R.G. and Z.Y.; visualization, G.X.; supervision, Z.Y. and F.Z.; project administration, Z.Y.; funding acquisition, Z.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Major Project of Liaoning Province "Key Technologies and Application Development of Intelligent Design".

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: https://engineering.case.edu/bearingdatacenter/apparatus-and-procedures, http://www.52phm.cn/datasets/bear/Bearing-data-set-of-Jiangnan-University.html (accessed on 1 January 2025).

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

References

1. Henao, H.; Capolino, G.A.; Fernandez-Cabanas, M.; Filippetti, F.; Bruzzese, C.; Strangas, E.; Pusca, R.; Estima, J.; Riera-Guasp, M.; Hedayati-Kia, S. Trends in fault diagnosis for electrical machines: A review of diagnostic techniques. *IEEE Ind. Electron. Mag.* **2014**, *8*, 31–42. [CrossRef]
2. Antonino-Daviu, J.A.; Martínez-Giménez, F.; Peris, A.; Ramezanzadeh, N.; Rodenas, F. Accurate Criteria for Broken Bar Detection in Induction Motors Based on the Wavelet (Packet) Transform. *Mathematics* **2024**, *12*, 1057. [CrossRef]
3. Guo, J.; Zhang, H.; Zhen, D.; Shi, Z.; Gu, F.; Ball, A.D. An enhanced modulation signal bispectrum analysis for bearing fault detection based on non-Gaussian noise suppression. *Measurement* **2020**, *151*, 107240. [CrossRef]
4. Tang, Y.; Zhang, C.; Wu, J.; Xie, Y.; Shen, W.; Wu, J. Deep Learning-based Bearing Fault Diagnosis Using a Trusted Multi-scale Quadratic Attention-embedded Convolutional Neural Network. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–15.
5. Qin, Y.; Liu, H.; Wang, Y.; Mao, Y. Inverse physics–informed neural networks for digital twin–based bearing fault diagnosis under imbalanced samples. *Knowl. -Based Syst.* **2024**, *292*, 111641. [CrossRef]
6. Fu, Z.; Liu, Z.; Ping, S.; Li, W.; Liu, J. TRA-ACGAN: A motor bearing fault diagnosis model based on an auxiliary classifier generative adversarial network and transformer network. *ISA Trans.* **2024**, *149*, 381–393. [CrossRef]
7. Wen, C.; Xue, Y.; Liu, W.; Chen, G.; Liu, X. Bearing fault diagnosis via fusing small samples and training multi-state siamese neural networks. *Neurocomputing* **2024**, *576*, 127355. [CrossRef]
8. Peng, C.; Sheng, Y.; Gui, W.; Tang, Z.; Li, C. A Rolling Bearing Fault Diagnosis Method Based on Multimodal Knowledge Graph. *IEEE Trans. Ind. Inform.* **2024**, *20*, 13047–13057. [CrossRef]
9. An, Y.; Zhang, K.; Liu, Q.; Chai, Y.; Huang, X. Rolling bearing fault diagnosis method base on periodic sparse attention and LSTM. *IEEE Sens. J.* **2022**, *22*, 12044–12053. [CrossRef]
10. Tang, H.; Tang, Y.; Su, Y.; Feng, W.; Wang, B.; Chen, P.; Zuo, D. Feature extraction of multi-sensors for early bearing fault diagnosis using deep learning based on minimum unscented kalman filter. *Eng. Appl. Artif. Intell.* **2024**, *127*, 107138. [CrossRef]
11. Tang, H.H.; Zhang, K.; Wang, B.; Zu, X.J.; Li, Y.Y.; Feng, W.W.; Jiang, X.; Chen, P.; Li, Q.A. Early bearing fault diagnosis for imbalanced data in offshore wind turbine using improved deep learning based on scaled minimum unscented kalman filter. *Ocean Eng.* **2024**, *300*, 117392. [CrossRef]
12. Zhang, F.; Yin, Z.; Xu, F.; Li, Y.; Xu, G. MAB-DrNet: Bearing fault diagnosis method based on an improved dilated convolutional neural network. *Sensors* **2023**, *23*, 5532. [CrossRef] [PubMed]
13. Deng, J.; Liu, H.; Fang, H.; Shao, S.; Wang, D.; Hou, Y.; Chen, D.; Tang, M. MgNet: A fault diagnosis approach for multi-bearing system based on auxiliary bearing and multi-granularity information fusion. *Mech. Syst. Signal Process.* **2023**, *193*, 110253. [CrossRef]
14. Wang, R.; Jiang, H.; Zhu, K.; Wang, Y.; Liu, C. A deep feature enhanced reinforcement learning method for rolling bearing fault diagnosis. *Adv. Eng. Inform.* **2022**, *54*, 101750. [CrossRef]
15. Tang, J.; Wu, J.; Hu, B.; Liu, J. An intelligent diagnosis method using fault feature regions for untrained compound faults of rolling bearings. *Measurement* **2022**, *204*, 112100. [CrossRef]
16. Peng, C.; Xiao, H.; Gui, W.; Tang, Z. Fault Classification of Rolling Bearings Based on Multi-Source Heterogeneous Data. *IEEE Sens. J.* **2024**, *24*, 24189–24199. [CrossRef]
17. Liu, Y.; Li, X.; Zhang, X.; Fan, L.; Chen, X.; Gong, B. Imbalanced deep transfer network for fault diagnosis of high-speed train traction motor bearings. *Knowl.-Based Syst.* **2024**, *293*, 111682. [CrossRef]
18. Tian, S.; Zhen, D.; Sun, G.; Liu, X.; Feng, G.; Gu, F. Few-shot condition diagnosis of rolling bearing using adversarial transfer network with class aggregation-guided. *Meas. Sci. Technol.* **2024**, *35*, 066120. [CrossRef]
19. Li, J.; Ye, Z.; Gao, J.; Meng, Z.; Tong, K.; Yu, S. Fault transfer diagnosis of rolling bearings across different devices via multi-domain information fusion and multi-kernel maximum mean discrepancy. *Appl. Soft Comput.* **2024**, *159*, 111620. [CrossRef]
20. Zhong, Z.; Zhang, Z.; Cui, Y.; Xie, X.; Hao, W. Failure Mechanism Information-Assisted Multi-Domain Adversarial Transfer Fault Diagnosis Model for Rolling Bearings under Variable Operating Conditions. *Electronics* **2024**, *13*, 2133. [CrossRef]
21. Saha, D.; Hoque, M.E.; Chowdhury, M.E. Enhancing bearing fault diagnosis using transfer learning and random forest classification: A comparative study on variable working conditions. *IEEE Access* **2023**, *12*, 5986–6000. [CrossRef]
22. Xing, S.; Wang, J.; Ma, J.; Han, B.; Zhang, Z.; Bao, H.; Ma, H.; Jiang, X.; Shen, Y. A New Dual-domain Signal Collaborative Transfer Network for Bearing Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–11. [CrossRef]
23. Liu, C.; Dong, F.; Ge, K.; Tian, Y. A New Bearing Fault Diagnosis Method Based on Deep Transfer Network and Supervised Joint Matching. *IEEE Photonics J.* **2024**, *16*, 1–17. [CrossRef]
24. Panareda Busto, P.; Gall, J. Open set domain adaptation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 754–763.
25. Saito, K.; Yamamoto, S.; Ushiku, Y.; Harada, T. Open set domain adaptation by backpropagation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 153–168.

26. Fu, J.; Wu, X.; Zhang, S.; Yan, J. Improved open set domain adaptation with backpropagation. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 2506–2510.
27. Wang, X.; Shi, Z.; Sun, S.; Li, L. Open-set domain adaptation via feature clustering and separation for fault diagnosis. *IEEE Sens. J.* **2024**, *24*, 16347–16361. [CrossRef]
28. Smith, W.A.; Randall, R.B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mech. Syst. Signal Process.* **2015**, *64*, 100–131. [CrossRef]
29. Li, K.; Ping, X.; Wang, H.; Chen, P.; Cao, Y. Sequential fuzzy diagnosis method for motor roller bearing in variable operating conditions based on vibration analysis. *Sensors* **2013**, *13*, 8013–8041. [CrossRef] [PubMed]
30. Ganin, Y.; Lempitsky, V. Unsupervised domain adaptation by backpropagation. In Proceedings of the International Conference on Machine Learning, Lille, France, 7–9 July 2015; PMLR; pp. 1180–1189.
31. Zhang, W.; Li, X.; Ma, H.; Luo, Z.; Li, X. Open-set domain adaptation in machinery fault diagnostics using instance-level weighted adversarial learning. *IEEE Trans. Ind. Inform.* **2021**, *17*, 7445–7455. [CrossRef]
32. Yu, X.; Zhao, Z.; Zhang, X.; Zhang, Q.; Liu, Y.; Sun, C.; Chen, X. Deep-learning-based open set fault diagnosis by extreme value theory. *IEEE Trans. Ind. Inform.* **2021**, *18*, 185–196. [CrossRef]
33. You, K.; Long, M.; Cao, Z.; Wang, J.; Jordan, M.I. Universal domain adaptation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2720–2729.
34. Zhao, C.; Shen, W. Dual adversarial network for cross-domain open set fault diagnosis. *Reliab. Eng. Syst. Saf.* **2022**, *221*, 108358. [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.