Knowledge-Based Systems 251 (2022) 109272

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Cross-domain intelligent bearing fault diagnosis under class imbalanced samples via transfer residual network augmented with explicit weight self-assignment strategy based on meta data

Xuan Liu [a], Jinglong Chen [a,*], Kaiyu Zhang [a], Shen Liu [a], Shuilong He [b,*], Zitong Zhou [c]

[a] State Key Laboratory for Manufacturing and Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[b] School of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin 541004, China
[c] ShaanXi Fast Gear Co., Ltd., Xi'an 710119, China

ARTICLE INFO

Article history:
Received 22 March 2022
Received in revised form 11 June 2022
Accepted 11 June 2022
Available online 17 June 2022

Keywords:
Fault diagnosis
Transfer learning
Imbalanced data
ResNet
Meta data

ABSTRACT

Intelligent fault diagnosis methods are significant to mitigate the dependency on expert knowledge and the cost. For the limited faulty data and variational working conditions of real scenarios, cross-domain diagnosis using existing diagnosis models is widely discussed. Especially, methods based on cross-domain transfer learning show great potentiality. However, the class imbalanced data of actual working conditions make it difficult to learn the actual fault feature distribution. To this end, a transfer residual network augmented with explicit weight self-assignment strategy based on meta data(TRN-EWM) is proposed. Specifically, we use a domain-shared ResNet to extract depth features of the data, which effectively avoid gradient disappearance and improve classification performance. Then, to lessen diagnosis difficulties in cross-domain and fully mine the actual feature distribution of the samples, a class imbalanced cross-domain transfer method is carried out. Ultimately, we creatively construct an explicit weight self-assignment strategy based on meta data for sample weight rebalancing, which prevents the dominance of major classes and the overfitting of minor classes. Two transfer experiments are conducted, and average cross-domain diagnosis accuracy of 99.60% is achieved by the proposed method, which shows the effectiveness in bearing fault diagnosis.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

With the development of the digital process, more large-scale intelligent mechanical equipment is deployed in diverse fields. Once failure occurs, it will cause serious consequences. Hence, mechanical fault diagnosis technology is significant for ensuring the safe operation of these equipment. Previously, diagnostic techniques relied on expert experience and required vast manpower and material resources. Recently, aided by artificial intelligence methods such as deep neural networks, intelligent fault diagnosis has been development rapidly [1,2]. Low-cost, fast, and effective intelligent fault diagnosis is crucial for ensuring the operation of mechanical equipment and avoiding economic losses [3].

The existing fault diagnosis methods are basically based on the analysis of time series vibration data of transmission components such as bearings. Traditional signal processing methods (time-domain characteristic Analysis [4], Fourier transform, wavelet transform [5], empirical mode decomposition [6], and etc.) and machine learning-based methods (support vector machines (SVM) [7], nearest neighbor (NN) [8], and etc.) have been shown to extract the fault feature used for classification. These methods analyze the statistical features from the raw vibration data, which are not efficient for complex conditions. To cope with the diagnosis in different working conditions, researchers need to construct appropriate features to analyze the vibration data, which means numerous experiments are needed for an application. However, the high cost of manpower and time resources does not guarantee an optimal design of diagnostic features. To optimize the process of diagnosis, researchers constantly proposed new fault diagnosis methods [8,9]. Recently, due to the excellent performance and low priority cost of deep networks, more intelligent fault diagnosis schemes based on deep learning have been proposed [10–12]. Various deep neural networks such as Convolution Neural Network (CNN) [13], ResNet [14], Autoencoder [15], Generative Adversarial Networks (GAN) [16], Graph Neural Networks (GNN) [17], and so on have been used to extract fault features. These intelligent diagnosis methods based on artificial intelligence and big data make the diagnosis process no longer dependent on expert knowledge and manual labor [16,17].

* Corresponding authors.
E-mail addresses: jlstrive2008@mail.xjtu.edu.cn (J. Chen), xiaofeilonghe@guet.edu.cn (S. He).

https://doi.org/10.1016/j.knosys.2022.109272
0950-7051/© 2022 Elsevier B.V. All rights reserved.

Generally, the premise for the effective application of deep learning methods is the data used for training and testing satisfy the same distribution [18]. However, in the real industrial scene, due to the alterations of working conditions, the distribution of training and testing data is always different. This discrepancy represents an impediment to the application of deep learning methods in bearing fault diagnosis. On one hand, due to the constant evolution of speed and load, and the enormous cost of data collection under different working conditions, labeled data for the deep diagnosis model is hard to collect from all conditions. Meanwhile, faults in different working conditions are not completely consistent. The discrepancy in data distribution makes the already trained deep model have a poor generalization ability in testing data. On the other hand, the health status of the equipment lasts far longer than the fault status. Only fault data under a particular condition can be collected when failure occurs, which means the fault data from the real industrial scene is extremely limited. whereas, a fault diagnosis method based on deep learning with satisfying performance often requires adequate data for training. Consequently, a fault diagnosis model with great performance and generalization ability is hard to be trained in a real scene. Overfitting and decreased accuracy can be easily caused by the limited fault samples and the discrepancies of work conditions. This challenge can be referred to as cross-domain fault diagnosis, which restricts the use of deep learning methods in real diagnosis scenarios.

To solve cross-domain problems, transfer learning methods have received increasing attention, which plays an important role by utilizing existing prior knowledge. By learning the unknown domain data distribution through the existing domain data distribution, transfer learning has shown potentiality in actual industrial diagnosis scenarios. Recently, researchers have discussed the application of methods based on deep transfer learning in the field of cross-domain fault diagnosis [19]. For example, Xie et al. [20] proposed to use Transfer Component Analysis (TCA) to fusion from time-domain and frequency-domain to extract the cross-domain feature for gearbox fault diagnosis. Han et al. [21] proposed extending the marginal distribution adaptation to Joint Distribution Adaptation (JDA) to accomplish accurate distribution matching in cross-domain transfer. Yang et al. [22] proposed a Feature-based transfer neural network (FTNN) to reduce the distribution discrepancy of deep transferable features in the transfer from laboratory bearings to locomotive bearings.

In intelligent fault diagnosis methods based on transfer learning, it satisfies a premise assumption: The training data is plentiful class-balanced labeled data. However, for the actual industrial mechanical equipment, most of the data that can be collected is the data under the normal state, and the fault data is tiny. Hence, it is difficult to obtain sufficient class-balanced fault data for the training of the diagnosis model. In the actual industrial diagnosis scene, it can be regarded as a class imbalanced classification issues with limited data [23]. If the limited failure data is directly used for training, it will result in a poor generalization ability and domination of major classes [24]. Accordingly, there have been some researches paying attention to the problems of class imbalance, such as data resampling methods [25] and data enhancement [24,25]. Data resampling methods oversample the minority classes and under-sample the major classes to balance the classes, which lead to the overfitting of minority classes and the underfitting of major classes. Meanwhile, data enhancement methods are complicated to calculate the generated formula and the generated samples are difficult to assess. Specifically, these methods are less effective when directly dealing with the diagnosis problems in complex and changeable real scenarios.

Priori experience has a heuristic effect on solving problems in most fields. To make full use of healthy sample information and priori experience from existing diagnostic models, we use the transfer learning method to deal with cross-domain imbalanced sequential vibration data. Meanwhile, to overcome the overfitting of minor class bring by imbalance training, we focus more on samples located in the marginal distribution. Considering that bearing fault sample data is limited and imbalanced in real scenarios, a transfer residual network augmented with explicit weight self-assignment strategy based on meta data (TRN-EWM) is proposed for the cross-domain intelligent bearing fault diagnosis under class imbalanced samples. In TRN-EWM, besides the basic cross-domain transfer network, optimized meta-network and weight assignment network are also built to process imbalanced data effectively. Firstly, the residual neural network is constructed as a fault feature extractor. By improving the depth of the network, the residual network extracts easily transferable fault features in depth, avoiding the possible gradient disappearance. Secondly, the target domain fault samples are limited and imbalanced, which makes it difficult to extract fault features directly. In comparison, the source domain has abundant class-balanced fault samples. Through transferring learning method, using the prior knowledge of the source domain fairly lessen diagnosis difficulties. Thirdly, imbalanced data in the target domain cause great interference to the process of transfer and easily leads to over-fitting of major class. In order to fully mine the fault feature distribution, the imbalanced data is re-weighted by explicit weight self-assignment network based on meta data. After reweighting, indistinguishable samples get high weight factor, and the model gets more suitable for actual imbalanced diagnosis tasks.

The contributions of this paper can be briefly summarized as follows:

(1) A novel intelligent diagnosis method suitable for imbalanced cross-domain fault diagnosis is proposed. With the transfer of residual network, source domain prior knowledge is used to improve the fault classification performance. By augmenting with explicit weight self-assignment strategy, it avoids the overfitting of imbalanced train samples and mine the actual distribution of fault feature.

(2) To obtain the deep transferable features of fault, improve the network depth, avoid possible gradient disappearance, and realize the intelligent fault diagnosis process of bearings, a fault feature extractor based on deep residual network is constructed.

(3) To avoid over-fitting of limited fault samples in fault diagnosis of target domain, cross-domain transfer is carried out. The prior knowledge of source domain samples obtained by fault feature extractor is transferred to target domain to lessen diagnosis difficulties.

(4) Besides cross-domain transfer residual network, an explicit weight self-assignment strategy based on meta data is adopted to construct a meta-model and weight network to optimize the sample weighting process.

The remainder of this paper is organized as follows. The preliminaries of this study are introduced in Section 2. Section 3 details the proposed method. In Section 4, transfer tasks are carried out to evaluate the proposed method. In Section 5, this paper makes some discussion about the proposed method. Finally, the conclusion is provided in Section 6.

2. Preliminaries

2.1. Deep residual network

Convolutional neural network (CNN) is a supervised learning neural network with a multi-layer convolutional structure, commonly used as feature extractor. Due to its excellent adaptability, CNN achieved great success in numerous areas such as computer vision. It has also been widely designed as signal feature

extractor in intelligent fault diagnosis schemes. Generally, the convolutional layer f_{conv} in CNN enhances the signal features, and the pooling layer f_{pool} down samples to improve computational efficiency and avoid overfitting. Besides, batch normalization (BN) layer f_{BN} and activation function f_{ReLU} are constructed to accelerate the process of training. During training, the error gradient computed by the loss function is back-propagated to make the network converge. Mathematically, the function of the CNN is constructed by fundamental formulated as follows:

$$x_i^{(n)} = f_{pool}\left\{f_{ReLU}\left[f_{BN}\left(f_{conv}(x_i^{(n-1)}), f_{conv}(X^{(n-1)})\right)\right]\right\}, x_i \in R^L \quad (1)$$

where represents the input data space with length and the input can be denoted as $X = \{x_1, x_2, \ldots, x_N\} \in R^{N\times L}$. The operation of BN reduces the displacement of internal covariates caused by the different distribution of the input of each layer, which is formulated as:

$$f_{BN}(x_i, X) = \frac{x_i - \mu(X)}{\sqrt{\sigma(X)^2 + \varepsilon}} \quad (2)$$

where x_i denotes the sample i of the input $X.\mu(X)$ and $\sigma(\mathrm{x})$ in Eq. (2) denotes the mean and variance, which are defined as:

$$\mu(X) = \frac{1}{K}\sum_{i=1}^{K} x_i \quad (3)$$

$$\sigma(X) = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(x_i{}^2 - \mu(X))^2} \quad (4)$$

The activation function f_{ReLU} is described as:

$$f_{ReLU}(x) = \max\{0, x\} \quad (5)$$

To extract deeper signal features, researchers try CNN with more network layers. However, as the layers of the network increase, the degree of training difficulty improves rapidly. Meanwhile, the gradient easily disappears during backpropagation, and the model accuracy does not increase any more. Hence, K. He et al. [26] proposed the deep residual network. As shown in Eq. (5), deep residual network designed the identity mappings x to ensure that gradients do not disappear. The residual network realizes residual learning by constructing multiple residual blocks as shown in Fig. 1, which is easily optimized. Generally, the skip structure of residual network ensures that the gradient will not disappear when backpropagation, and the accuracy can be improved by increasing the depth.

$$H(x) = F(x) + x \quad (6)$$

2.2. Transfer learning

Traditional intelligent fault diagnosis methods usually rely on plentiful labeled fault data to provide feature knowledge [19]. However, it is difficult to obtain these labeled fault samples in the actual scenarios. Diagnosis models from labeled laboratory fault data have a poor generalization performance when facing a different working condition. By using the prior knowledge from source domain to assist target domain, the transfer learning method is fairly promising for actual industrial scenarios where sufficient fault data cannot be obtained. After giving the source domain $\mathcal{D}^s$ the source domain learning task $\mathcal{T}^s$, the target domain $\mathcal{D}^t$, and the target domain learning task $\mathcal{T}^s$. The goal of transfer learning is to obtain prior knowledge from the source domain and learning tasks to help improve the learning of learning tasks in the target domain, i.e. $\mathcal{D}^s \neq \mathcal{D}^t$ and $\mathcal{T}^s \neq \mathcal{T}^t$.

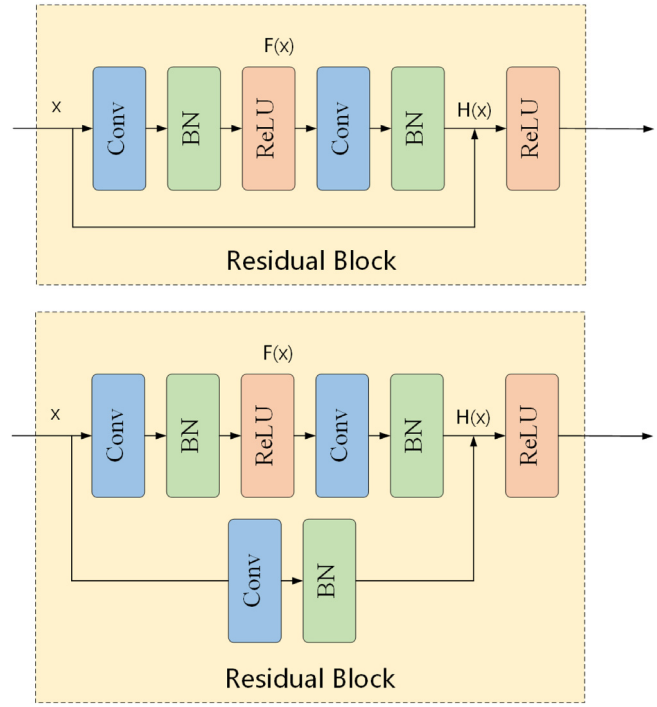

Fig. 1. Basic structure of the residual block.

Transfer learning methods can be divided as instance-based [26,27], model-based [28,29] and feature-based [20,30,31]. Transfer learning-based fault diagnosis methods have great performance under similar working conditions and balanced train samples. Once imbalanced datasets are used as target domains, the transferred prior knowledge may easily lead to overfitting, which is shown in the following Fig. 2. Finding a transfer learning method that can effectively deal with imbalanced data will guide the practical application of mechanical fault diagnosis.

2.3. Class imbalance and meta learning

Class imbalance refers to the situation where consists in a disproportion of the number of different classes in the classification task [24]. Zhang et al. [23] pointed out that for industrial machinery and equipment, although multiple sensors are arranged to collect vibration data, most of the data is in a healthy state, and fault data is limited. Intelligent fault diagnosis in actual engineering scenarios can be regarded as an imbalance problem. Compared with the long-tailed imbalanced dataset that often appears in the image field, it is difficult to evaluate the probability of various failures in the field of fault diagnosis. Consequently, we assume that the probability of occurrence of various types of failures is close, and specialize the imbalanced dataset as a polarization distribution which is shown in Fig. 3. Fault samples class form imbalanced target domain are sampled the same number. Meanwhile, using imbalance rate (IBR) to characterize the degree of sample imbalance, which is defined as

$$IBR = \frac{\min\{N\}}{\max\{N\}} \quad (7)$$

where $N = \{n_1, n_2, \ldots, n_m\}$ is the class number of imbalanced datasets, $\max\{N\}$ is the maximum of imbalance class number and $\min\{N\}$ is the minimum of imbalance class number.

Meanwhile, for imbalanced dataset, the accuracy is no longer proper measure of the class imbalanced model. Therefore, the F_1 score (the harmonic mean of the precision rate and the recall

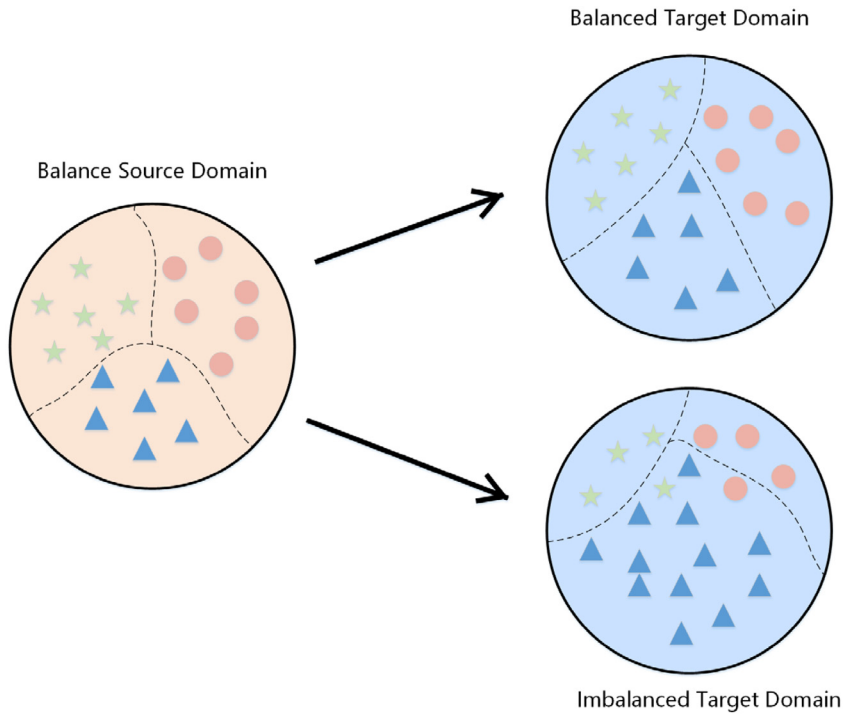

Fig. 2. Overfitting of minority when transfer from balanced source domain to imbalanced target domain.

domain.

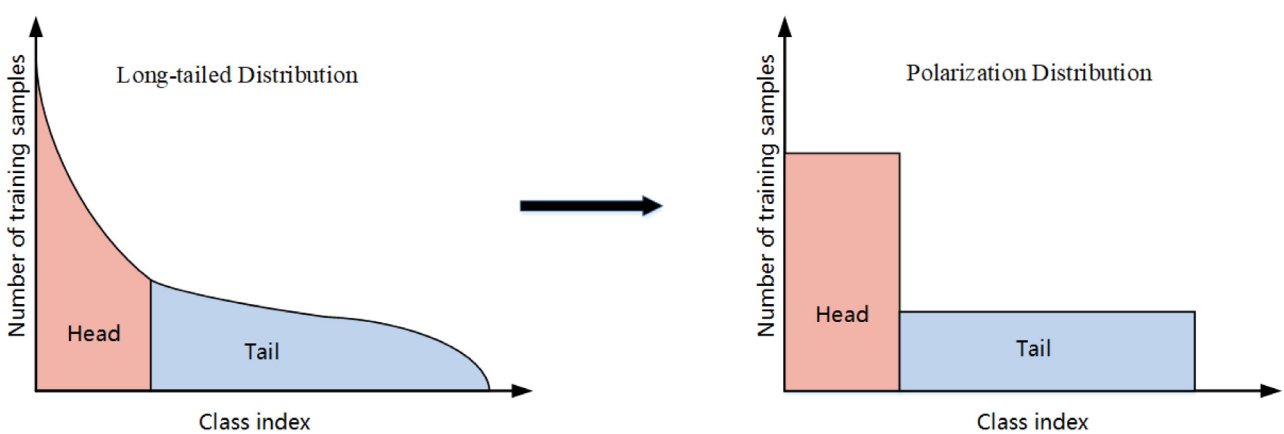

Fig. 3. Datasets with long-tail distributions and polarization distributions.

rate) is introduced to reflect the category balance problem, and the formula shown in Eqs. (7)–(9)

$$P = \frac{1}{n}\sum_{i=1}^{n}\frac{TP_i}{TP_i + FP_i} \tag{8}$$

$$R = \frac{1}{n}\sum_{i=1}^{n}\frac{TP_i}{TP_i + FN_i} \tag{9}$$

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{10}$$

where TP_i, FP_i and FN_i represent the amount of True Positive (Actually positive and predicted to be positive), False Positive (Actually negative but predicted to be positive), and False Negative (Actually positive but predicted to be negative) in classi respectively.P and R represent the amount of Precision (The proportion of correct classified instances among the ones classified as positive), Recall (The proportion of total positive instances correctly classified as positive) in class i respectively. n is the number of classes.

Generally, the higher level of class imbalance, the more difficult to classify, which causes poor generalization and performance. Sample rebalance achieved through weight factor alleviate this problem. Xiao et al. [32] proposed a data re-weight method based on TrAdaBoost to distribute the weights of training samples. However, this method only processes the data, and requires the source domain have a high similarity with the target domain. Once there is a large deviation between the data distribution, it may lead to a negative transfer. Cao et al. [33] used the loss function to implicitly assign the weights, which alleviates

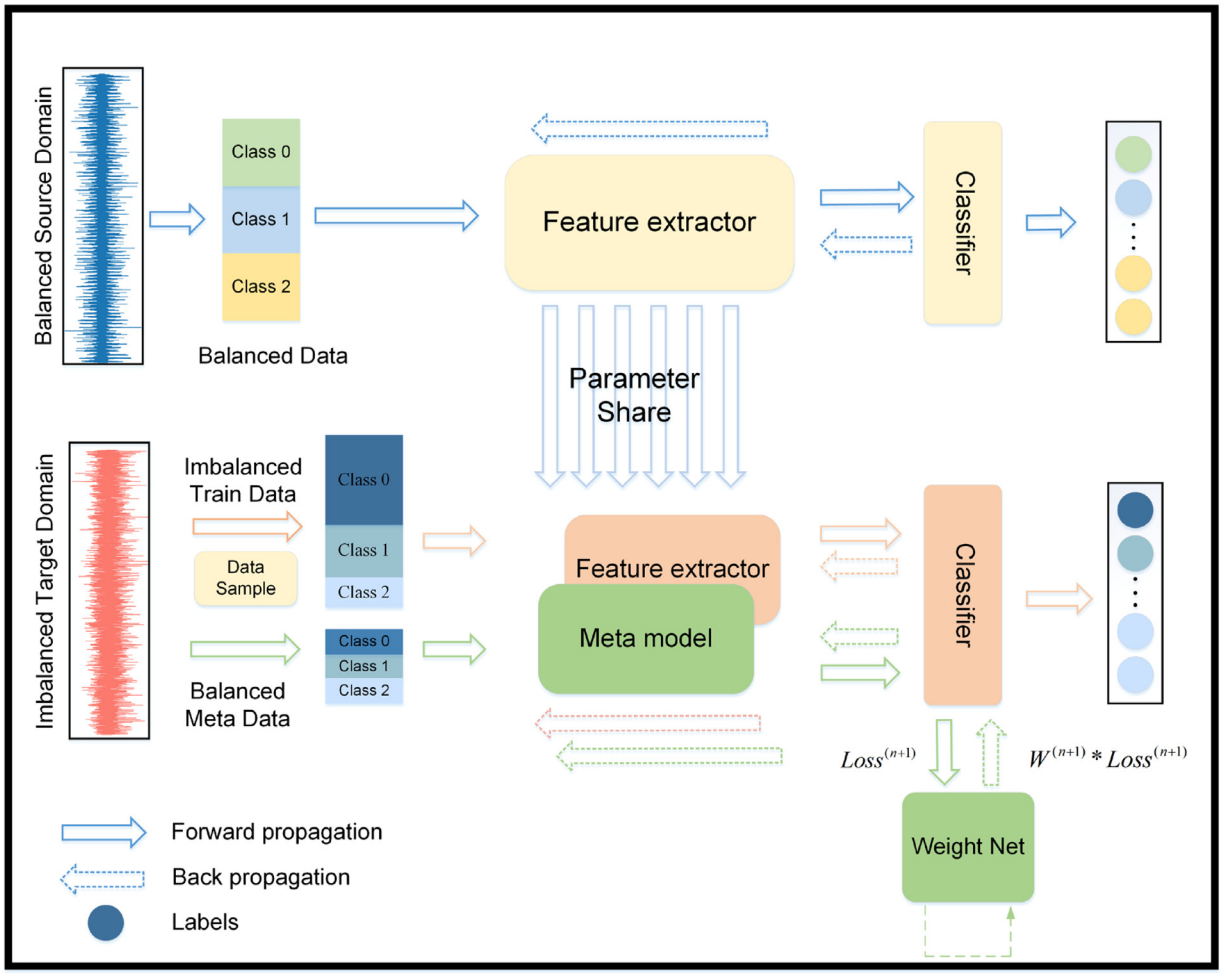

Fig. 4. Framework of TRN-EWM.

the dominant situation of major class, but there are a few types of overfitting. Obviously, these methods still face the overfitting of minority. Hence, effectively alleviate the minority overfitting is a key issue to improve the utilization of imbalanced data for learning in transfer learning.

Meta learning is a thought of "learn to learn" [34]. Inspired by the idea of meta learning, some ideas of adaptively learning sample weights are adopted to solve class imbalance [35]. Deep network with attention mechanism [36], bidirectional LSTM network [37] and etc. are used in the field of fault diagnosis. These methods generally require input of complex sample feature information, which makes it difficult to obtain the same performance as traditional intelligent diagnosis methods. Additionally, its unstable weighting method leads to the decline of generalization performance. Explicit weight network learns from meta-data and focuses on uncertain samples with greater loss which more likely locate at the classification boundary. The weight self-assignment strategy is added to the network to realize the mapping from training loss to sample weight.

3. Proposed method

In this section, framework of the proposed method is detailed. As shown in Fig. 4, the framework of the TRN-EWM is given, which can be divided into three parts: domain shared feature extractor, explicit weight self-assignment module and meta model. Each part details are as follows.

3.1. Domain-shared feature extractor

The domain-shared feature extractor F_e are independent networks on source–target domain, and the target domain use the balanced samples training parameters w^s from source domain. As shown in Fig. 5, we choose the Resnet18 which contains four residual blocks as feature extractor F_e. A total of 18 convolution operations f_{conv} were performed in the input layer of the source domain. The sample data is processed into balanced class samples, number of each class is $N^s = \{n_1^s, n_2^s, \ldots, n_c^s\}$, where $n_1^s = n_2^s =, \ldots, = n_c^s$.

After processed by the convolutional layer and batch normalized, feature extractor obtain the shallow features, and then perform the average pooling process f_{avg_pool} to reduce the dimensionality of the features while maintaining the invariance of the maximum value feature. f_{avg_pool} is described as

$$f_{avg_pool} = \frac{1}{|R_m|} \sum_{i \in R_m} \mathbf{x}_i^k \tag{11}$$

where $\mathbf{x}_i^k$ is the kth feature map of element x_i in pooling region R_m.

Only with the input layer processed, the extracted shallow transferable features are difficult to generalize. As a consequence, by superimposing the residual block, the deep transferable features are extracted. In the residual block, the depth of the network is improved through identity mapping, which avoid the emergence of gradient explosion and gradient disappearance. The

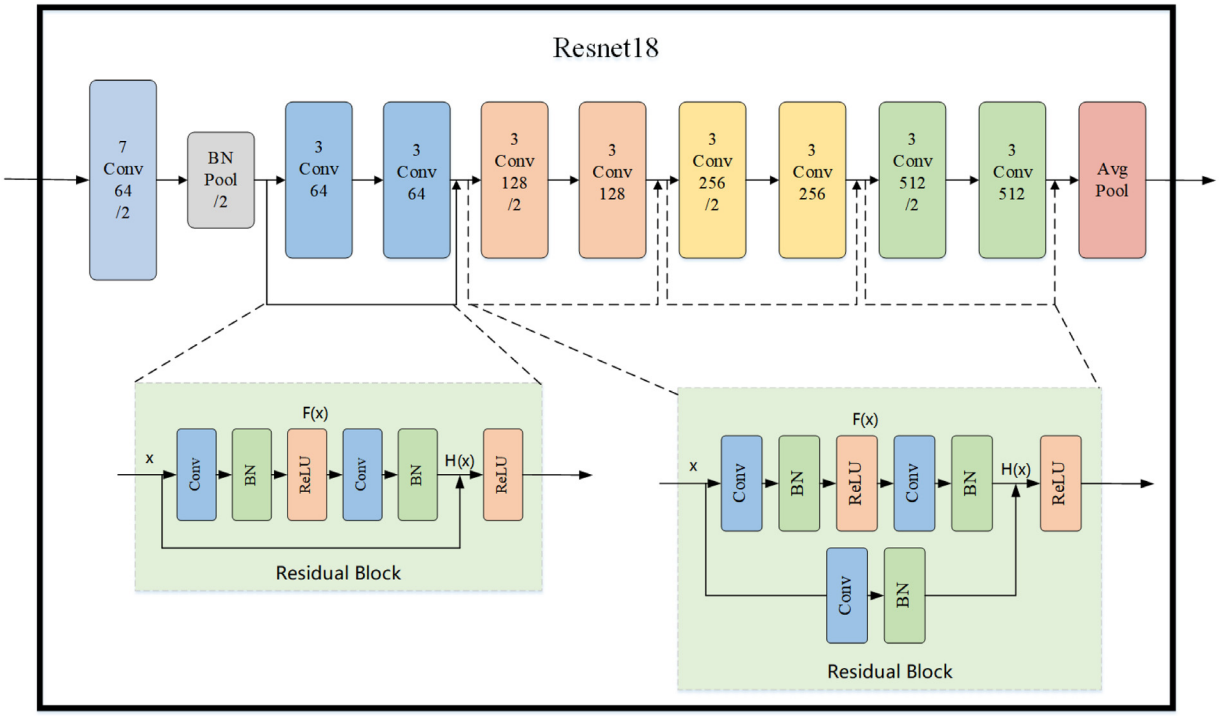

Fig. 5. Construction of domain-shared feature extractor.

classifier F_c maps directly from the identical dimension of the feature extractor output to the dimension which is same with the sample classes. Finally, the cross-entropy loss function is used to calculate loss, back propagate and iterative optimization, which is expressed as follow:

$$L^n(y_k, \hat{y}_k) = -\sum_{k=1}^{b} y_k \log(\hat{y}_k) \tag{12}$$

where n is the iteration factor, and b is the sample number of batches. y_k and $\hat{y}_k$ represent the true and predicted labels, which are designed as one-hot vectors.

3.2. Explicit weight self-assignment module

The explicit weight self-assignment module F_{ew} learns how to optimize the initial weights w^t through the meta model. After reweighting, target domain feature extractor F_e^t learns to deal with indistinguishable samples, and avoid the guiding of normal easy identified samples. Explicit weight self-assignment module F_{ew} increases the weight of the indistinguishable samples, so that the target domain classification F_c^t can realize a more possibly estimation strategy. The explicit weight self-assignment module uses a three-layers fully connected neural network (Multi-layer Perceptron) to realize the mapping from the meta data loss *Meta_Loss* to updated weights *Weight*, which is defined as

$$Meta_Loss = \begin{bmatrix} meta_Loss_1 \\ meta_Loss_2 \\ \vdots \\ meta_Loss_n \end{bmatrix} \tag{13}$$

$$Weight = \begin{bmatrix} weight_1 \\ weight_2 \\ \vdots \\ weight_n \end{bmatrix} \tag{14}$$

$$Weight^{(k)} = V(Meta_Loss^{(k)}; \Theta) \tag{15}$$

where V is the explicit weight self-assignment module and Θ is the parameters of weight self-assignment module. $meta_Loss_i$ and $weight_i$ respectively represent the ith loss calculated by meta model and weight for ith samples.

As shown in Fig. 6, each hidden layer node is composed with activation function f_{ReLU} and sigmoid activation function $f_{sigmoid}$. $f_{sigmoid}$ guarantee the weight locates in the interval of [0,1], which is described as:

$$f_{sigmoid} = \frac{1}{1 + e^{-x}} \tag{16}$$

The weight module can approximate all continuous functions in the real number domain. Through the iteration of parameters from the meta-model F_e^m and the classifier model F_c^t, the self-assignment of the sample weights is realized.

3.3. Meta model

Plentiful indistinguishable samples of normal classes in imbalance training will mislead explicit weight self-assignment module and cause the overfitting. To make the model focus on minor fault class, a meta-model is constructed to provide balanced meta data as input of explicit weight self-assignment module. The explicit weight self-assignment module learns to assign the weight equitable for class and adaptively adjusts for different transfer tasks. Meta model F_e^m have the same construction as domain-shared feature extractor F_e. Its parameter shared from the imbalanced target domain feature extractor F_e^t, which can be thought as a loss calculated network from limited sample transfer network. Each iteration period, meta model recalculate the meta data loss with the updated parameters W^t, hence the classifier model F_c^t is robust against to imbalanced dataset.

Imbalanced samples directly as training data will easily cause overfitting of the minority class and dominance of the majority class. In comparison, with only transfer, the model is easily misled by limited balanced samples. To facilitate the extraction of effective features from imbalanced samples, model should focus on the indistinguishable sample classes. Therefore, an explicit weight

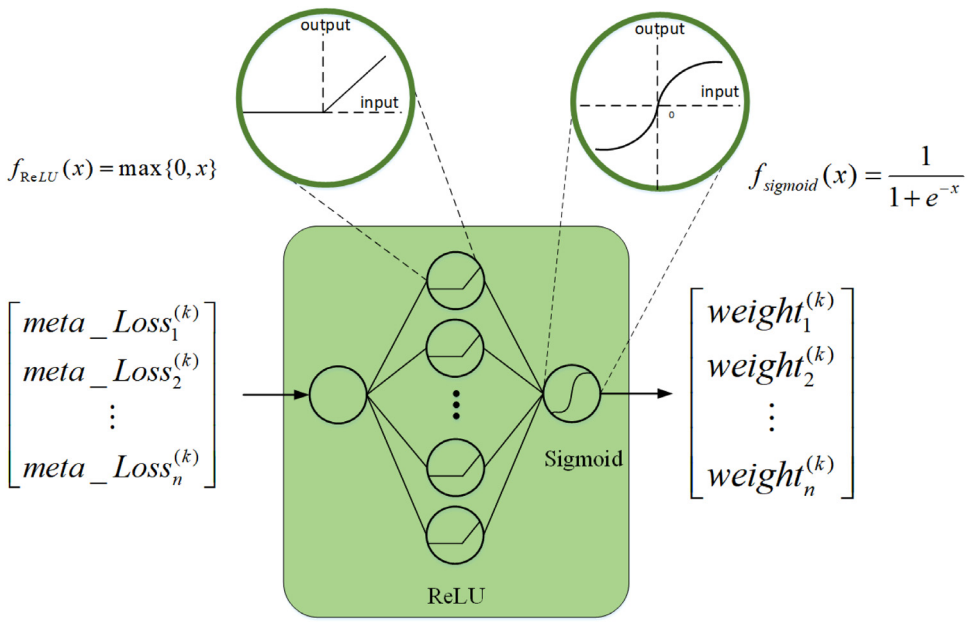

Fig. 6. Illustration for explicit weight self-assignment module.

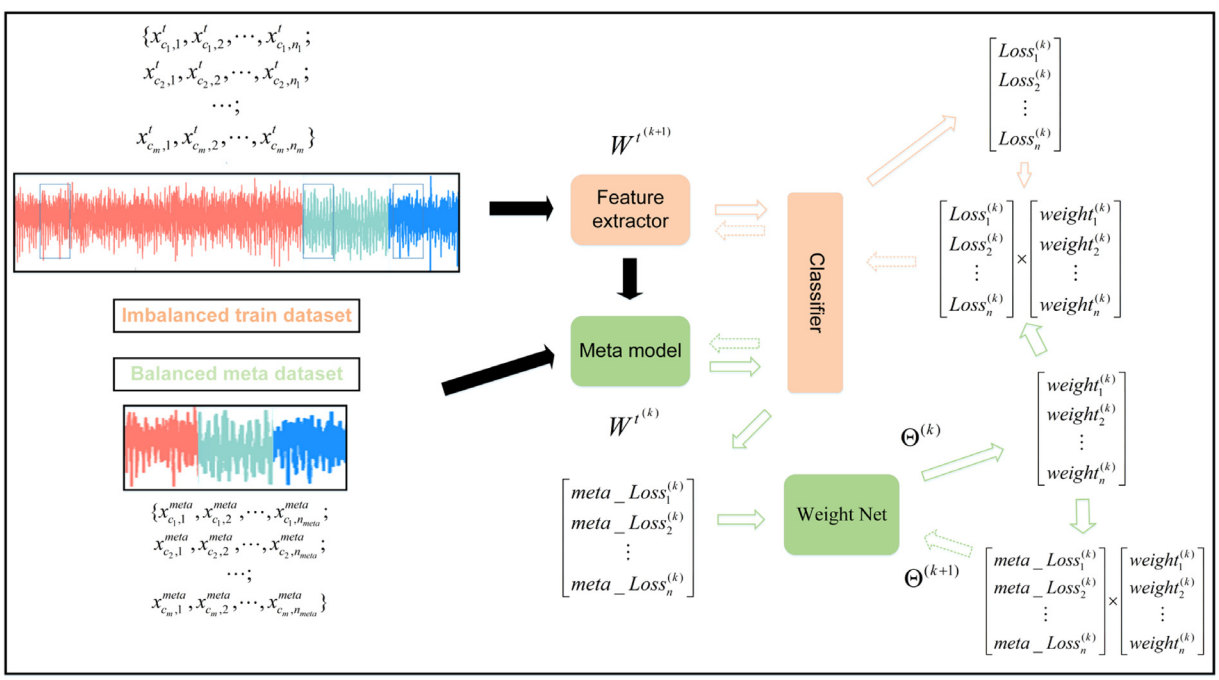

Fig. 7. Data processing flow of meta model and explicit weight self-assignment module.

self-assignment strategy based on meta data is proposed, which data processing flow is shown in Fig. 7. Through inconsistent sampling, built imbalanced train dataset and balanced meta dataset. Imbalanced dataset is used to fully mine the fault distinguishing feature information of the samples, while balanced meta data is used to adjust the weight of the samples to avoid overfitting of the minority class and the dominance of the majority class. The proposed strategy uses meta-learning thought to learn and adjust the weight module parameters Θ, and adopts meta model loss as input to adaptively assign the weights. Meta model adopts the same network structure as the classification feature extractor, but with independent parameters. SGD optimization on few labeled balanced meta samples modifies the gradient propagation direction. The update direction of the classifier model parameters is weighted to avoid the deviation bring by plentiful normal data that is easily distinguishable. While dealing with imbalanced target domain training data, model parameters can be updated to the direction which achieves effective division.

3.4. Training strategy

Since cross-domain bearing fault diagnosis is commonly facing with the end-to-end situations, which hardly achieves simultaneous training of domains. Therefore, it is not convenient to evaluate the similarity between domains. A shared pre-train model is more suitable for the rapid diagnosis of such industrial equipment. Meanwhile, due to the bearing failure mechanisms of

domains are similar, we assume that the generalization of optimal model obtained in the source domain training phase is locally optimal. The train model with the highest fault identification accuracy is selected as shared pre-train model.

The implementation steps of this training strategy are as follow:

(1) Input: The input includes the samples from the balanced source domain and imbalanced target domain. Each sample contains 1024 sequent data points sliced from a time-series signal, i.e. $x_i^s, x_i^t \in R^{1\times 1024}$. Hence, the source–target domain input data can be denoted as $X_i^s = [x_1^s, x_2^s, \ldots, x_{N_s}^s] \in R^{N_s\times 1\times 1024}$, $X_i^t = [x_1^t, x_2^t, \ldots, x_{N_t}^t] \in R^{N_t\times 1\times 1024}$. Before sending into the network, samples were normalized as follow:

$$x_i = \frac{\dot{x}_i}{\sum_{k=1}^{1024} \dot{x}_i^2(k)} \tag{17}$$

(2) Pre-train: Randomly select labeled class balanced source domain data (x_i^s, y_i^s) as the input. After fully iteration training, the optimal extractor model parameters w^{s*} verify fully with test sets. Then, the feature extractor parameters w^{s*} are transferred to the target domain extractor as the initialized network parameters $w^{t^{(0)}}$.

(3) Target domain data description: Through the cross iteration in target domain, weight self-assignment of imbalanced train datasets is realized. Imbalanced train datasets are combined by adequate normal samples and slight faulty samples, which can be defined as

$$X^{train} = \{x_{c_1,1}^t, x_{c_1,2}^t, \ldots, x_{c_1,n_1}^t; x_{c_2,1}^t, x_{c_2,2}^t, \ldots, x_{c_2,n_2}^t; \ldots; x_{c_m,1}^t, x_{c_m,2}^t, \ldots, x_{c_m,n_m}^t\} \tag{18}$$

where c_m is the class of samples and n_m is the number of class c_m. c_1 is the normal class. Therefore, $n_1 \gg n_2 = n_3 =, \ldots, = n_m$ and $IBR = \frac{n_m}{n_1}, m \neq 1$. Besides, meta samples X^{meta} are randomly extracted from X^{train}, the number of each class is n_{meta}.X^{meta} can be defined as:

$$X^{meta} = \{x_{c_1,1}^{meta}, x_{c_1,2}^{meta}, \ldots, x_{c_1,n_{meta}}^{meta}; x_{c_2,1}^{meta}, x_{c_2,2}^{meta}, \ldots, x_{c_2,n_{meta}}^{meta}; \ldots; x_{c_m,1}^{meta}, x_{c_m,2}^{meta}, \ldots, x_{c_m,n_{meta}}^{meta}\} \tag{19}$$

(4) Weight self-assignment: Target domain feature extractor F_e^t parameter w^t use the weight loss as the input of gradient back propagation. By minimizing loss to get the optimal parameter of a single iteration, which can be denoted as:

$$w^t(\Theta) = \underset{w^t}{argmin}(L_i^{train}(w^t); \Theta) \triangleq \frac{1}{N}\sum_{i=1}^{N} V(L_i^{train}(w^t); \Theta)\nabla_{(w^t)} L_i^{train}(w^t) \tag{20}$$

$$\Theta^* = \underset{\Theta}{argmin} L_i^{meta}(w^t(\Theta)) \triangleq \frac{1}{M}\sum_{i=1}^{M} L_i^{meta}(w^t(\Theta)) \tag{21}$$

The SGD optimizer was used to update the parameters, and the iteration format of the feature extractor parameters was obtained as Eq. (20).

$$\hat{w}^{t^{(k)}}(\Theta) = w^{t^{(k)}} - \alpha\frac{1}{n} \times \sum_{i=1}^{n} V(L_i^{train}(w^{t^{(k)}}); \Theta)\nabla_{(w)} L_i^{train}(w^{t(k)})|_{w^{t^{(k)}}} \tag{22}$$

where α is the learning rate and $L_i^{train}(w) = L(y_i^{train}, F_e^t(y_i^{train}, w))$ is the loss calculated by the cross entropy function.

Shared parameters ensure the meta-model parameter F_e^{meta} are consistent with the feature extractor F_e^t. Update the weight self-assignment module parameters Θ, which can be denoted as:

$$\Theta^{(k+1)} = \Theta^{(k)} - \beta\frac{1}{m}\sum_{i=1}^{m} \nabla_\Theta L_i^{meta}(\hat{w}^{t^{(k)}}(\Theta))|_{\Theta^{(k)}} \tag{23}$$

where β is the learning rate and $Meta_Loss = L_i^{meta}(\hat{w}^{t^{(k)}}(\Theta))$ is the loss calculated on meta data. Then, recalculate with updated parameters and complete a parameter iteration. The parameters will be updated as:

$$w^{t^{(k+1)}}(\Theta) = w^{t^{(k)}} - \gamma\frac{1}{n} \times \sum_{i=1}^{n} V(L_i^{train}(w^{t^{(k)}}); \Theta^{(k+1)})\nabla_{(w)} \times L_i^{train}(w^t)^{(k)}|_{w^{t^{(k)}}} \tag{24}$$

where γ is the learning rate. Finally, save the best feature extractor parameters, and test in target domain balanced test dataset.

Table 1
Working condition of modular test rig.

Number	Rotational speed [rpm]	Load torque [N m]	Radial force [N]
PU0	1500	0.7	1000
PU1	900	0.7	1000
PU2	1500	0.1	1000
PU3	1500	0.7	400

Table 2
Details of the used bearings of PU dataset.

Bearing code	Fault element	Damaged way	Damaged extent
K004	Health		
KI03	Inner ring	Electric Engraver (EE)	1
KI07	Inner ring	Electric Engraver (EE)	2
KA03	Outer ring	Electric Engraver (EE)	2
KA05	Outer ring	Electric Engraver (EE)	1

4. Case study

In this section, TRN-EWM is evaluated through model validation. By compared with typical mechanical transfer learning model, the performance of TRN-EWM is explored. And proposed method is apply to PU datasets [38] and EB datasets.

4.1. Datasets

4.1.1. PU bearing datasets

PU dataset is a widely used public bearing dataset from Paderborn University [38]. As shown in Fig. 8, the test rig is a modular system, composed of the drive motor, a torque measurement shaft, bearing module, flywheel and a load motor. Current, rotation speed, torque, and vibration were collected synchronously. This dataset collected include three kinds of bearing: artificial damages, accelerated-failure real damages, healthy.

In the transfer learning case study, four kinds of working condition which shown in Table 1 bearing vibration signals are used for transfer task. The vibration signals with 64 kHz sampling frequency were divided into sample length of 1024. Bearings used in this experiment are 6203 type. Details of the used bearings are shown in Table 2.

4.1.2. EB bearing datasets

Due to the difference between the laboratory and in the actual scenario, in order to verify the performance of the proposed method in the actual scenario, a high-precision simulation Escalator Bearing(EB) rig is applied to simulate the bearing vibration dataset under actual working conditions. The EB dataset is a low rotation speed vibration dataset with only 0.15 Hz rotating

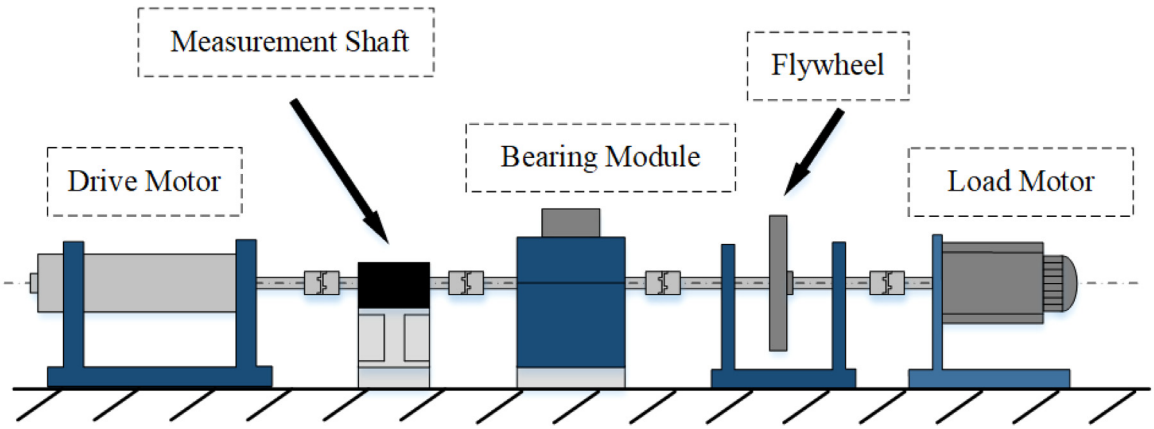

Fig. 8. Modular test rig of PU dataset.

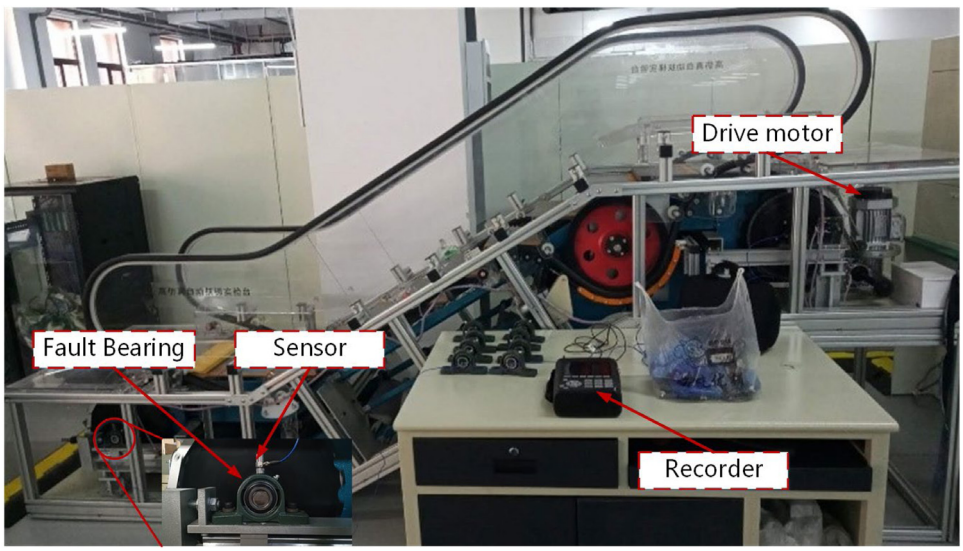

Fig. 9. High-precision simulation escalator bearing test rig.

frequency. The EB test rig is composed of a drive motor, a fault bearing, high-simulation escalator body. The vibration signals were collected by sensor and recorder. The structure of EB test rig is shown in Fig. 9.

EB bearing datasets include three kinds of bearing: normal (NC), inner ring damage (IR) and outer ring damage (OR), which is shown in Fig. 10. Test faulty bearings were artificial manufactured and had three degrees of failure. Bearings used in this experiment are 6205 type. The vibration signals with 12.8 kHz sampling frequency were divided into sample length of 1024. Due to the low speed of the test rig, single sample has faint signal characteristics, which is close to the hardly transferable operating state in actual scenarios.

4.2. Data preprocessing

When the source domain data is used to train the pre-training model, 200 samples are extracted balanced for each category, and each sample contains 1024 sequent vibration signal points. The sample ratio of the train datasets to the test datasets is 3:1, which avoids possible confusion of datasets. After 400 iterations, pre-trained feature extractor parameters with best accuracy are saved as the transfer parameters.

As regards the target domain data, training datasets are imbalanced, with limited labeled fault data and plentiful normal data. Meanwhile, test datasets are abundant labeled class balanced data which evaluate learned characteristics of data distribution. There is no duplication of data between two stages. 100 samples are selected from the normal bearing class during each generation of training, and the imbalance rate set as from 0.1 to 0.3 which refer to 10 to 30 samples of each type of faulty bearing class are selected. 200 samples are selected from each class during the test stage. Each sample also contains 1024 sequent vibration signal points. Extract meta samples number is $n_{meta} = 5$ of the same length from each class of training datasets for updating meta model parameters.

4.3. Transfer tasks and performance comparison

In order to assess the performance of the proposed transfer algorithm, this section establishes three transfer tasks to demonstrate the effect of the algorithm under variable operating conditions and cross-domain conditions. Concomitantly, proposed method is compared with several classic transfer learning algorithms (TCA, JDA, ResNet, FTNN) in different imbalanced transfer tasks. The implementation details of compared methods are listed as follows:

- TCA: feature-based method, maps data together into a high-dimensional regenerated kernel Hilbert space and minimizes the domain data distance by MMD.

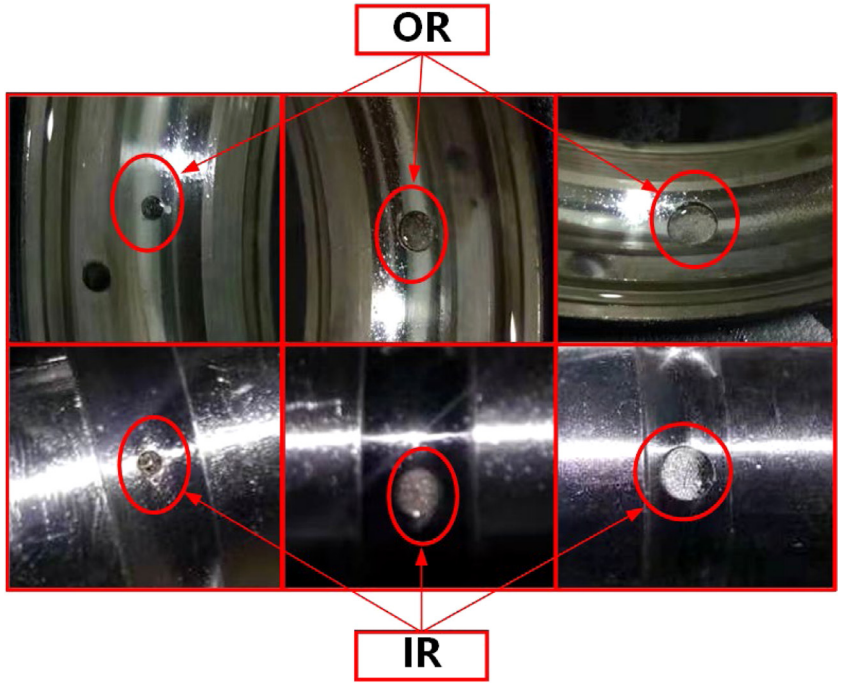

Fig. 10. Inner ring damage (IR) and outer ring damage (OR) bearing of EB dataset.

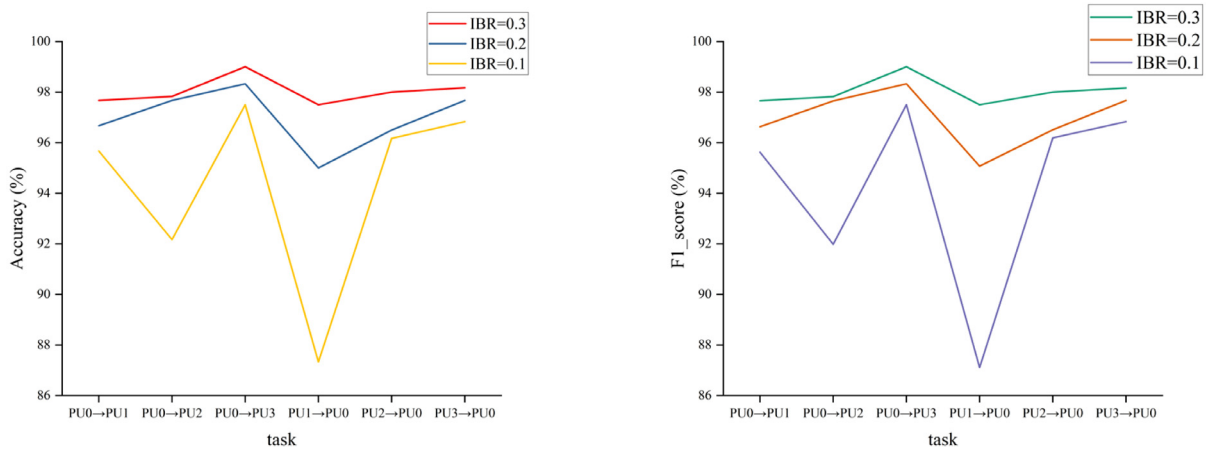

Fig. 11. Result of transfer tasks under single working condition change.

- JDA: feature-based method, realizes the alignment of marginal distribution and conditional distribution by minimizing class conditional probability base on TCA.
- FTNN: feature-based method uses the same transfer network architecture with TRN-EWM without the meta model.
- ResNet: same network architecture with TRN-EWM without share parameter from source domain.

4.3.1. Transfer tasks under variable conditions

The PU bearing datasets contain four kinds bearing working state, including different rotational speeds, load torques, and radial forces. Firstly, PU0 datasets are taken as benchmark, and we build PU0 → PU1, PU0 → PU2, PU0 → PU3, PU1 → PU0, PU2 → PU0, PU3 → PU0 six single working condition change transfer tasks. Next, six transfer tasks with two working condition change: PU1→PU2, PU1→PU3, PU2→PU1, PU2→PU3, PU3→PU1, PU3→PU2. In each transfer task, three imbalance rates are assigned for experimentation: 0.1, 0.2, 0.3 and meta samples number n_{meta} is set to 5. The results of proposed method in the transfer tasks are shown in Figs. 11 and 12.

Figures shows that the TRN-EWM has good performance in the transfer tasks both single working conditions and two working conditions change. In the case of the imbalance rate is 0.2 or 0.3, the classification accuracy rate basically over 95%. When the imbalance rate is 0.1, the classification accuracy fluctuates greatly due to the definitely discrepancy between the majority and the minority class. As for the classification accuracy and F1 score, it can be seen that there is fairly consistent, which means that the weight self-assignment of TRN-EWM is effective.

4.3.2. Transfer tasks under different datasets

Only the laboratory datasets under different working conditions are involved, the bearing types of the faulty bearing are identical. The fault characteristics between source and target datasets is similar, which lessen the difficulties of transfer tasks. While in actual scenarios, the target datasets no longer have the similar source datasets. Absolutely, most of the transfer diagnosis tasks are facing a cross-domain situation which generally has different bearing types or datasets. In order to make the transfer

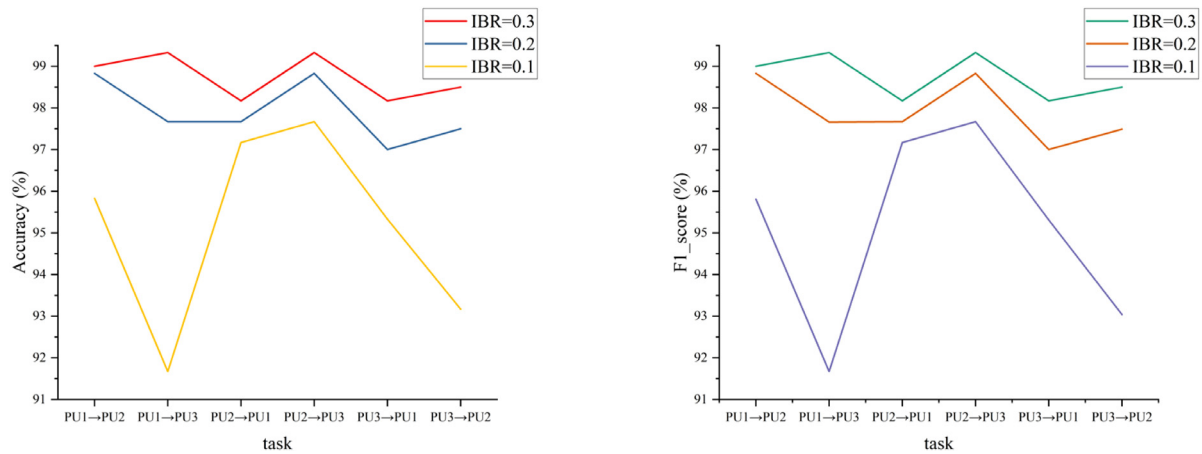

Fig. 12. Result of transfer tasks under two working condition change.

Table 3
The description for compared methods.

Methods	Imbalance rate	Accuracy (%)	F1_score (%)	Precision (%)	Recall (%)
TCA	0.3	48.67 ± 1.47	31.89 ± 3.48	30.15 ± 3.90	31.15 ± 5.37
	0.2	41.26 ± 3.38	27.07 ± 3.12	29.32 ± 5.09	29.41 ± 6.73
	0.1	31.05 ± 3.80	18.20 ± 3.96	19.47 ± 8.69	29.53 ± 6.49
JDA	0.3	53.90 ± 2.14	35.47 ± 4.20	35.59 ± 4.08	34.35 ± 3.25
	0.2	42.30 ± 4.67	30.17 ± 3.41	31.42 ± 4.70	32.35 ± 4.93
	0.1	33.03 ± 1.66	26.53 ± 4.14	29.244 ± 5.85	33.03 ± 1.66
FTNN	0.3	86.26 ± 0.39	86.09 ± 0.43	86.43 ± 0.46	86.26 ± 0.39
	0.2	84.5 ± 0.33	84.28 ± 0.35	84.58 ± 0.42	84.5 ± 0.33
	0.1	77.63 ± 0.61	76.96 ± 0.70	77.59 ± 0.47	77.63 ± 0.61
ResNet	0.3	84.53 ± 0.25	84.61 ± 0.40	85.69 ± 073	84.55 ± 0.28
	0.2	82.87 ± 0.12	82.94 ± 0.10	83.41 ± 0.57	82.87 ± 0.11
	0.1	74.63 ± 0.26	74.64 ± 0.27	78.70 ± 0.28	74.63 ± 0.26
TRN-EWM	0.3	**99.6 ± 0.24**	**99.7 ± 0.12**	**99.6 ± 0.25**	**99.6 ± 0.25**
	0.2	**98.23 ± 0.78**	**98.28 ± 0.73**	**98.24 ± 0.78**	**98.23 ± 0.78**
	0.1	**90.97 ± 1.29**	**91.27 ± 1.66**	**90.97 ± 1.26**	**90.96 ± 1.28**

tasks as close to the actual transfer scenario as possible, we set the cross-domain transfer tasks from the PU datasets to the EB datasets.

Since the EB datasets collects from escalator bearing, its rotation frequency is only 0.15 Hz. We chose the PU1 datasets with a lowest rotation speed as the source domain datasets. For each transfer task, three imbalance rates are assigned for experimentation: 0.1, 0.2, 0.3 and meta samples number is set as 5. The results of comparison methods in the transfer tasks are shown in the Table 3. Visualization of the comparison results is presented in Figs. 13 and 14 [39].

As shown in the comparison results and visualization of cross-domain transfer tasks, TRN-EWM has achieved the best classification accuracy as 99.67%,98.17%,91.67% and F1 score as 99.17%, 98.17%, 91.61% under three imbalance rate. When TCA and JDA deal with imbalanced data, the classification accuracy is far over the F1 score. It can be considered that these methods have poor classification on minor fault categories and induce overfitting of the normal category. The classification accuracy of FTNN and ResNet is basically about 80% due to the extraction of deep features. However, the F1 score is slightly lower than the classification accuracy. It can be seen that the normal category is still dominated, and part of the fault category is mistakenly classified as the normal category. Since TRN-EWM realizes the rebalance of samples, it can normally distinguish between normal and faulty classes, which has less probability of misclassification.

5. Discussion

In this section, some discussion about proposed methods are given. In order to analyze the influence of several key parameters on the proposed method, we repeatedly trained and tested 5 times to obtain more reliable results.

5.1. Number of meta samples

Considering that the loss of meta model directly affects the weight self-assignment of the weight network. Specifically, the number of meta samples n_{meta} input to the meta model is a key factor affecting the training loss which is the input of the weight network. Discuss and analyze its impact by comparing the performance of network classification under different numbers of meta-samples. As shown in Fig. 15, we recorded the transfer performance in the number of meta-samples of 5 categories when the training set fault class contains 30 samples in the transfer task PU1 → EB. Each result is the average of five repetitions.

As shown in the test results, since the n_{meta} is changed, number of meta samples actually affect the transfer performance. It can be seen from the results that the distribution error of recognition accuracy is larger when the number is smaller. Except $n_{meta} = 4$, it can be considered that the higher the number of meta-samples, the better the network recognizes. When the number reaches 5, the accuracy basically comes up to 100%. Arguably, n_{meta} affect the transfer performance possibly by influencing the

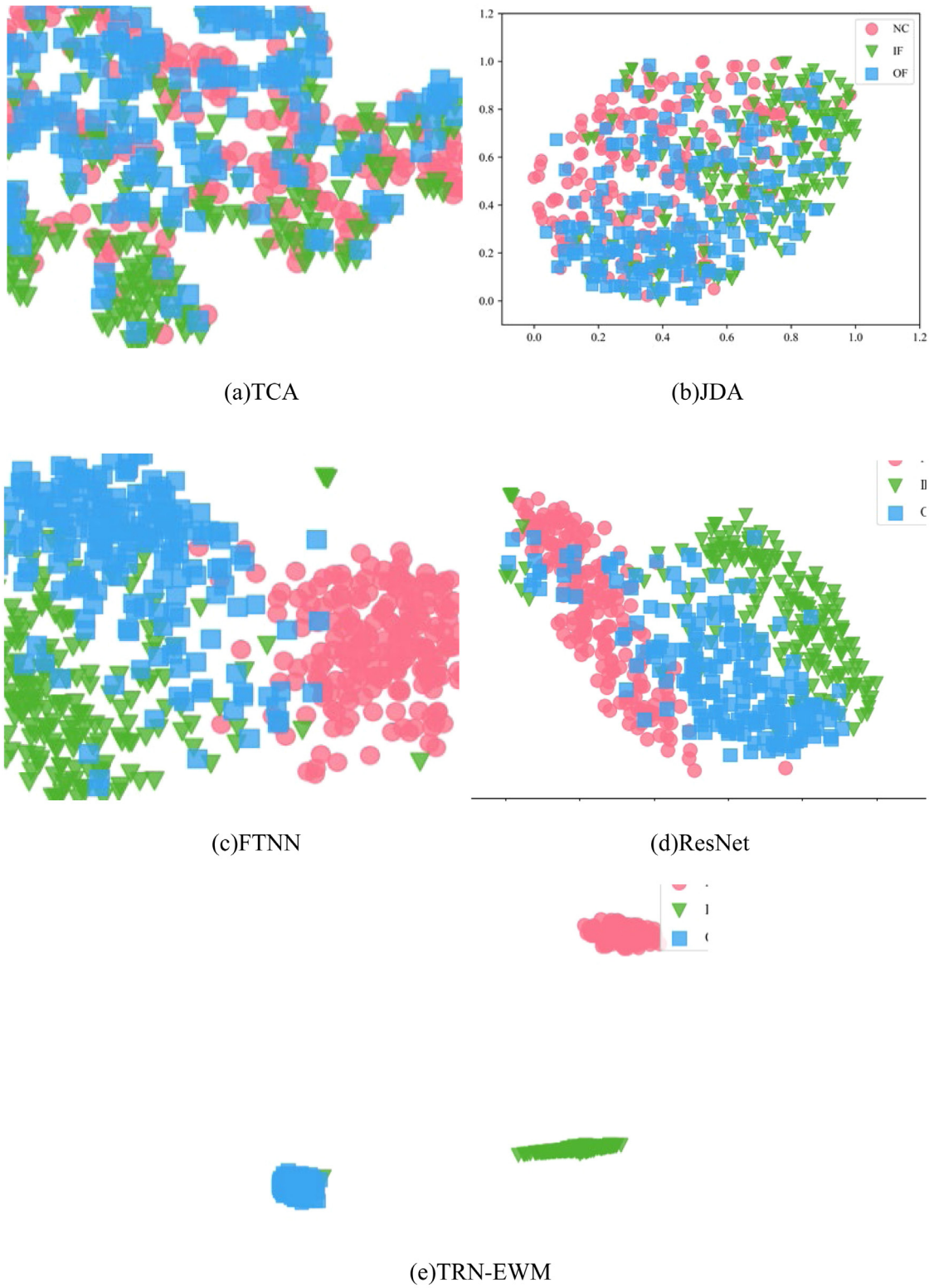

Fig. 13. T-SNE visualization of the comparison results in transfer task PU1→EB.

back propagation process of the weight self-assignment module gradient.

5.2. Architecture of weight module

Due to explicit weight self-assignment module is designed as an MLP with one hidden layer. Its structure may affect the weight self-assignment. As shown in Table 4, we depict representative results under 3 different structures with different widths.

As shown in the result, different MLP architecture have little effect on the final result. This may be owing to the function of the weight module achieves function fitting from loss to weight, which influenced teeny in fault recognition. On the contrary,

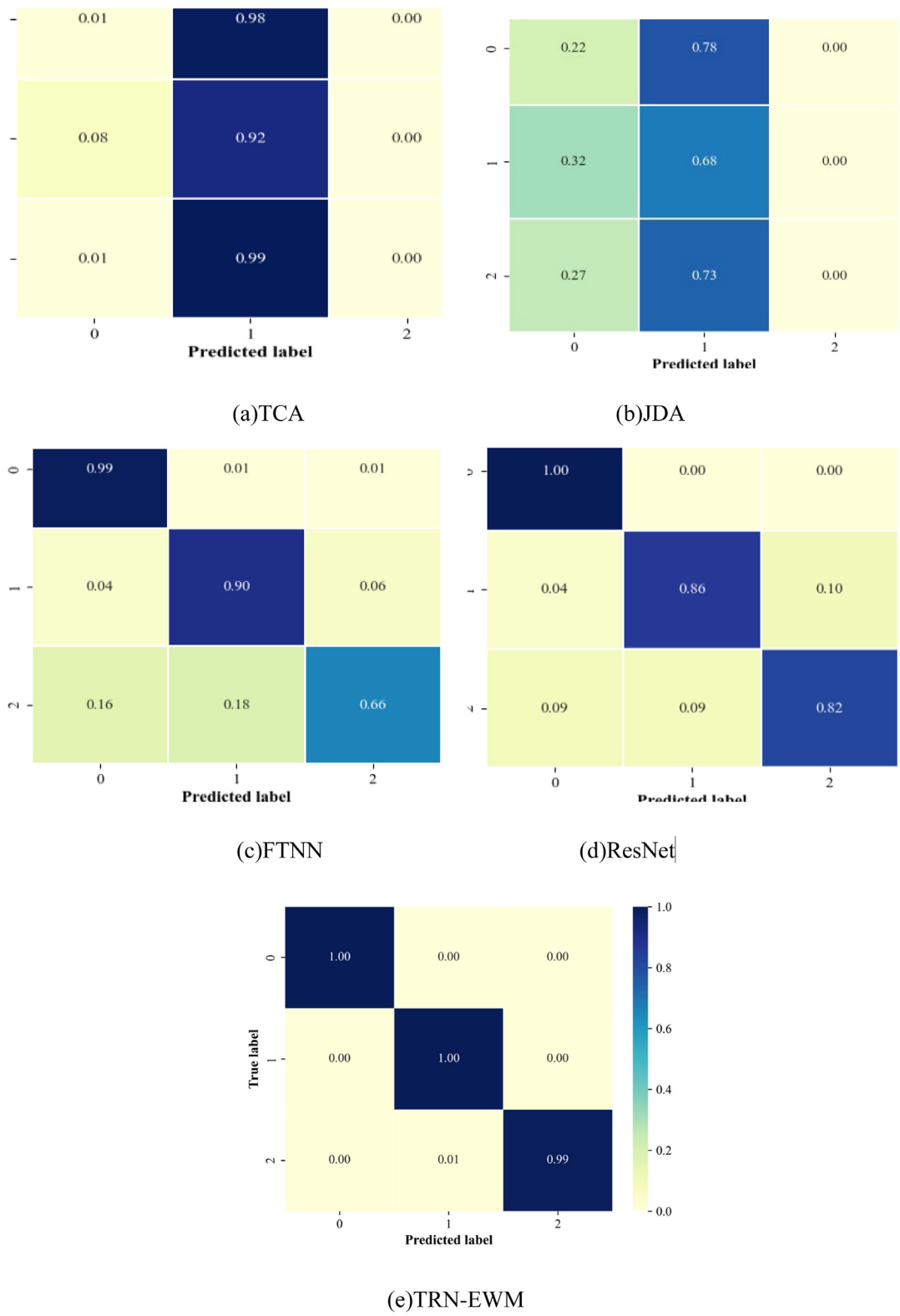

Fig. 14. Comparison methods confusion matrix in the transfer task PU1→EB.

the change of MLP architecture more likely to affect the weight fitting, which is not a major factor.

5.3. Influence of ResNet

Generally, increasing of the network depth can improve the feature extraction ability. However, network with more layers means the more requirement of computational resources, which easily cause the waste and more training time. For actual diagnostic scenarios, it should try to use networks that satisfy performance requirements and consume fewer resources. In the proposed methods, we choose the deep residual network as feature extractor. In comparison, we change the multilayer CNN as

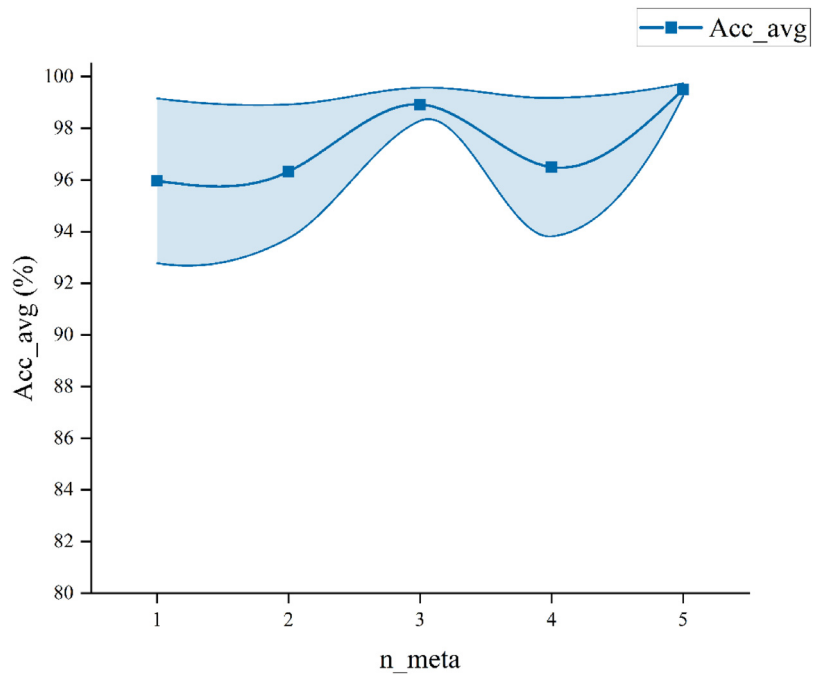

Fig. 15. The influence with different number of meta-samples.

Table 4
The influence of explicit weight self-assignment module.

Architecture	1-50-1	1-100-1	1-200-1
Accuracy (%)	99.27 ± 0.56	99.50 ± 0.24	99.13 ± 0.78
F1_score (%)	99.27 ± 0.56	99.50 ± 0.24	99.13 ± 0.79

feature extractor to determine the necessity of residual connections. Each fault category contained 30 samples, and imbalanced training data with an imbalance rate of 0.3 were set in this experiment. Five repetitions in the transfer task PU4→EB. Fig. 16 records the result of comparison.

As the network depth increases, the accuracy of feature network become lower. It seems becomes overfitting as the number of network layers' increases. Since the CNN as the feature extractor, it achieves the accuracy average as 89.57% in the transfer task. In comparison, the accuracy average is 99.87% in the condition of ResNet18 as domain-shared feature extractor. Obviously, the residual connections actually increase the performance of feature extractor and avoid an overfitting.

5.4. Optimization learning rate

For the proposed method, SGD is chosen as optimization algorithm. In the target domain, three network parameters are optimizable, which means the different combinations of learning rate α, β, γ may affect the direction and speed of gradient back propagation. Hence, we change the learning rate α, β, γ as 0.001, 0.0001, 0.00001 to get a better combinations of learning rate. Likewise, 30 samples of each fault class and a 0.3 imbalance rate are set in this experiment. Five repetitions in the transfer task PU4 → EB. Table 5 records the result of combinations of different learning rate.

As shown in the results, different learning rate have limited influence on feature extraction. The proposed network can basically obtain a global optimal parameter after iterations. Meanwhile, the proposed method achieve a better fault recognition as the learning rate satisfy the equation $\alpha = \gamma$ in particular. The same network structure of meta model and target domain feature extractor may influence, which makes the feature extractor easily come to the optimal as a same learning rates are chosen.

Table 5
The result of combinations of different learning rate.

A. $\alpha = 0.001$				
$\alpha = 0.001$		γ		
		0.001	0.0001	0.00001
β	0.001	99.27 ± 1.47	98.67 ± 2.41	97.97 ± 2.49
	0.0001	99.83 ± 0.10	99.60 ± 0.55	98.89 ± 2.19
	0.00001	99.96 ± 0.06	99.96 ± 0.06	99.90 ± 0.07
B. $\alpha = 0.0001$				
$\alpha = 0.0001$		γ		
		0.001	0.0001	0.00001
β	0.001	98.60 ± 2.80	99.94 ± 0.07	98.60 ± 2.72
	0.0001	99.80 ± 0.19	99.93 ± 0.08	98.47 ± 2.98
	0.00001	99.53 ± 0.65	99.33 ± 1.17	99.27 ± 1.14
C. $\alpha = 0.00001$				
$\alpha = 0.00001$		γ		
		0.001	0.0001	0.00001
β	0.001	97.97 ± 3.90	98.33 ± 3.65	99.93 ± 0.08
	0.0001	99.90 ± 0.13	99.93 ± 0.08	99.96 ± 0.07
	0.00001	99.80 ± 0.32	99.93 ± 0.08	99.96 ± 0.07

6. Conclusion

In this paper, we propose a cross-domain transfer residual network augmented with explicit weight self-assignment strategy based on meta data. The proposed method takes into account the limited samples and class imbalanced real industrial scenes in the cross-domain transfer. By explicit weight self-assignment strategy based on meta data, it makes the network highly applicable to imbalanced cross-domain bearings fault scenarios. Deep residual network is utilized as base domain-shared model, which demonstrates strong ability in deep features extraction and lessen the difficulties of diagnosis. Datasets PU, and EB are utilized to evaluate the effectiveness of proposed method. Compared

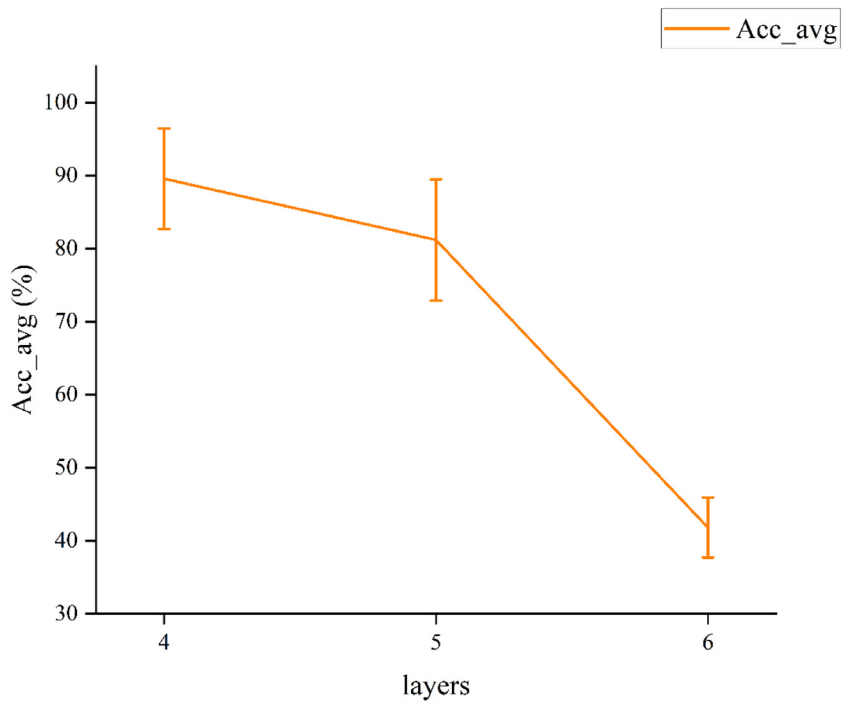

Fig. 16. Accuracy of CNN with different numbers of convolution layers as feature extractor.

with several transfer learning methods, the proposed method shows good recognition ability and effectiveness in the transfer of imbalanced data and achieves best overall performance. The subsequent discussion in this paper shows the proposed method has great potentiality and stability in the mechanical transfer diagnosis tasks from laboratory scenes to actual industrial scenes.

CRediT authorship contribution statement

Xuan Liu: Conceived and designed the experiments, Performed the experiments, Implemented methods and analyzed the data, Wrote the paper. **Jinglong Chen:** Conceived and designed the experiments, Implemented methods and analyzed the data. **Kaiyu Zhang:** Performed the experiments, Implemented methods and analyzed the data. **Shen Liu:** Performed the experiments, Implemented methods and analyzed the data. **Shuilong He:** Modified the language of the paper. **Zitong Zhou:** Performed the experiments, Implemented methods and analyzed the data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to sincerely thank all the anonymous reviewers for the valuable comments that greatly helped to improve the manuscript.

This research is supported financially by the National Natural Science Foundation of China (No. 51875436), Guangxi Science and Technology Major Project, China (No. Guike AA22068001), National Key Research and Development Program of China (No. 2019YFF0302204), Scientific research and technology development in Liuzhou, China (No. 2021AAA0112), Open Fund of State Key Laboratory (No. sklms2022005), Guangxi Natural Science Foundation, China Program (No. 2020GXNSFAA159081) and Fundamental Research Funds for the Central Universities, China (No. XZY022020007, No. XZY022021006).

References

[1] T. Pan, J. Chen, Z. Zhou, C. Wang, S. He, A novel deep learning network via multiscale inner product with locally connected feature extraction for intelligent fault detection, IEEE Trans. Ind. Inf. 15 (9) (2019) 5119–5128, http://dx.doi.org/10.1109/TII.2019.2896665.

[2] J. Pan, Y. Zi, J. Chen, Z. Zhou, B. Wang, LiftingNet: A novel deep learning network with layerwise feature learning from noisy mechanical data for fault classification, IEEE Trans. Ind. Electron. 65 (6) (2017) 4973–4982.

[3] R. Huang, J. Li, S. Wang, G. Li, W. Li, A robust weight-shared capsule network for intelligent machinery fault diagnosis, IEEE Trans. Ind. Inf. 16 (10) (2020) 6466–6475, http://dx.doi.org/10.1109/TII.2020.2964117.

[4] L. Cui, W. Li, X. Wang, D. Zhao, H. Wang, Comprehensive remaining useful life prediction for rolling element bearings based on time-varying particle filtering, IEEE Trans. Instrum. Meas. 71 (2022) 1–10, http://dx.doi.org/10.1109/TIM.2022.3163167.

[5] A. Bouzida, O. Touhami, R. Ibtiouen, A. Belouchrani, M. Fadel, A. Rezzoug, Fault diagnosis in industrial induction machines through discrete wavelet transform, IEEE Trans. Ind. Electron. 58 (9) (2010) 4385–4395.

[6] D. Yu, J. Cheng, Y. Yang, Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings, Mech. Syst. Signal Process. 19 (2) (2005) 259–270.

[7] Q. Hu, Z. He, Z. Zhang, Y. Zi, Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble, Mech. Syst. Signal Process. 21 (2) (2007) 688–705.

[8] B. Bagheri, H. Ahmadi, R. Labbafi, Application of data mining and feature extraction on intelligent fault diagnosis by artificial neural network and k-nearest neighbor, in: The XIX International Conference on Electrical Machines-ICEM 2010, 2010, pp. 1–7.

[9] F. Lv, C. Wen, Z. Bao, M. Liu, Fault diagnosis based on deep learning, in: 2016 American Control Conference (ACC), 2016, pp. 6851–6856.

[10] R. Zhang, H. Tao, L. Wu, Y. Guan, Transfer learning with neural networks for bearing fault diagnosis in changing working conditions, IEEE Access 5 (2017) 14347–14357.

[11] M.J. Hasan, M.M. Islam, J.-M. Kim, Acoustic spectral imaging and transfer learning for reliable bearing fault diagnosis under variable speed conditions, Measurement 138 (2019) 620–631.

[12] D. Liu, L. Cui, W. Cheng, D. Zhao, W. Wen, Rolling bearing fault severity recognition via data mining integrated with convolutional neural network, IEEE Sens. J. 22 (6) (2022) 5768–5777, http://dx.doi.org/10.1109/JSEN.2022.3146151.

[13] L. Eren, T. Ince, S. Kiranyaz, A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier, J. Signal Process. Syst. 91 (2) (2019) 179–189.
[14] L. Wen, X. Li, L. Gao, A transfer convolutional neural network for fault diagnosis based on ResNet-50, Neural Comput. Appl. 32 (10) (2020).
[15] C. Lu, Z.-Y. Wang, W.-L. Qin, J. Ma, Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification, Signal Process. 130 (2017) 377–388.
[16] Z. Wang, J. Wang, Y. Wang, An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition, Neurocomputing 310 (2018) 213–222.
[17] K. Zhang, J. Chen, S. He, F. Li, Y. Feng, Z. Zhou, Triplet metric driven multi-head GNN augmented with decoupling adversarial learning for intelligent fault diagnosis of machines under varying working condition, J. Manuf. Syst. 62 (2022) 1–16.
[18] L. Guo, Y. Lei, S. Xing, T. Yan, N. Li, Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data, IEEE Trans. Ind. Electron. 66 (9) (2018) 7316–7325.
[19] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, A.K. Nandi, Applications of machine learning to machine fault diagnosis: A review and roadmap, Mech. Syst. Signal Process. 138 (2020) 106587, http://dx.doi.org/10.1016/j.ymssp.2019.106587.
[20] J. Xie, L. Zhang, L. Duan, J. Wang, On cross-domain feature fusion in gearbox fault diagnosis under various operating conditions based on transfer component analysis, in: 2016 Ieee International Conference on Prognostics and Health Management (Icphm), 2016, pp. 1–6.
[21] T. Han, C. Liu, W. Yang, D. Jiang, Deep transfer network with joint distribution adaptation: A new intelligent fault diagnosis framework for industry application, ISA Trans. 97 (2020) 269–281.
[22] B. Yang, Y. Lei, F. Jia, S. Xing, An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings, Mech. Syst. Signal Process. 122 (2019) 692–706, http://dx.doi.org/10.1016/j.ymssp.2018.12.051.
[23] T. Zhang, et al., Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions, ISA Trans. 119 (2022) 152–171.
[24] J. Wu, Z. Zhao, C. Sun, R. Yan, X. Chen, Learning from class-imbalanced data with a model-agnostic framework for machine intelligent diagnosis, Reliab. Eng. Syst. Saf. 216 (2021) 107934, http://dx.doi.org/10.1016/j.ress.2021.107934.
[25] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: Synthetic minority over-sampling technique, J. Artificial Intelligence Res. 16 (2002) 321–357.
[26] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[27] T. Pan, J. Chen, J. Xie, Z. Zhou, S. He, Deep feature generating network: A new method for intelligent fault detection of mechanical systems under class imbalance, IEEE Trans. Ind. Inf. 17 (9) (2021) 6282–6293, http://dx.doi.org/10.1109/TII.2020.3030967.
[28] Z. Zhang, H. Chen, S. Li, Z. An, Unsupervised domain adaptation via enhanced transfer joint matching for bearing fault diagnosis, Measurement 165 (2020) 108071.
[29] F. Shen, R. Langari, R. Yan, Exploring sample/feature hybrid transfer for gear fault diagnosis under varying working conditions, J. Comput. Inf. Sci. Eng. 20 (4) (2020).
[30] S. Shao, S. McAleer, R. Yan, P. Baldi, Highly accurate machine fault diagnosis using deep transfer learning, IEEE Trans. Ind. Inf. 15 (4) (2018) 2446–2455.
[31] J. Zhou, X. Yang, L. Zhang, S. Shao, G. Bian, Multisignal VGG19 network with transposed convolution for rotating machinery fault diagnosis based on deep transfer learning., Shock Vib. (2020).
[32] D. Xiao, Y. Huang, C. Qin, Z. Liu, Y. Li, C. Liu, Transfer learning with convolutional neural networks for small sample size problem in machinery fault diagnosis, Proc. Inst. Mech. Eng. C 233 (14) (2019) 5131–5143.
[33] K. Cao, C. Wei, A. Gaidon, N. Arechiga, T. Ma, Learning Imbalanced Datasets with Label-Distribution-Aware Margin Loss.
[34] J. Wu, Z. Zhao, C. Sun, R. Yan, X. Chen, Few-shot transfer learning for intelligent fault diagnosis of machine, Measurement 166 (2020) 108202.
[35] M. Ren, W. Zeng, B. Yang, R. Urtasun, Learning to reweight examples for robust deep learning, in: International Conference on Machine Learning, 2018, pp. 4334–4343.
[36] Y. Feng, J. Chen, T. Zhang, S. He, E. Xu, Z. Zhou, Semi-supervised meta-learning networks with squeeze-and-excitation attention for few-shot fault diagnosis, ISA Trans. (2021) S0019057821001543, http://dx.doi.org/10.1016/j.isatra.2021.03.013.
[37] Q. Chen, Q. Xie, Q. Yuan, H. Huang, Y. Li, Research on a real-time monitoring method for the wear state of a tool based on a convolutional bidirectional LSTM model, Symmetry 11 (10) (2019) 1233.
[38] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, T. Zhang, Deep model based domain adaptation for fault diagnosis, IEEE Trans. Ind. Electron. 64 (3) (2016) 2296–2305.
[39] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (11) (2008).