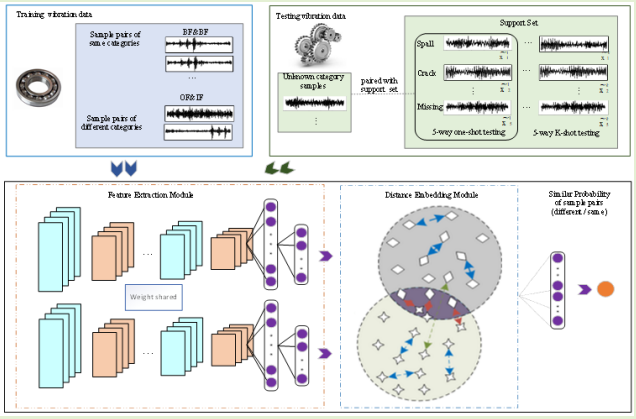
Cross-Category Mechanical Fault Diagnosis Based on Deep Few-Shot Learning

Juan Xu, *Member, IEEE*, Yongfang Shi, Xiaohui Yuan, *Senior Member, IEEE*, and Siliang Lu, *Senior Member, IEEE*

***Abstract*—Industrial fault diagnosis often faces challenges from insufficient examples. Methods leveraging Generative Adversarial Network or transfer learning address this problem. However, the model trained by the labeled examples of one component is not applicable to classify the new fault categories of other components. This problem aggravates when there exist very few examples. In this paper, we propose a cross-category fault diagnosis method (CFDM) based on few-shot learning. Our method constructs a convolutional Siamese neural network to extract fault features from example pairs. A cross-entropy based loss function is used that includes parameters for feature discrepancy to maximize the inter-category distances and minimize the intra-category distances. This enables the proposed method to learn the accurate classification boundaries between fault features of the example pairs. We conduct experiments on two public benchmark datasets and one lab-built dataset. Our evaluation includes analysis of the proposed method to classify fault types with one or five examples in each category of the target component. Our results demonstrate that the proposed method improves the fault diagnosis accuracy and robustness in comparison to the state-of-the-art methods.**



***Index Terms*— Fault diagnosis, few-shot learning, cross-category, Siamese neural network.**

I. Introduction

MECHANICAL fault diagnosis is important to ensure the safety of equipment and personnel. In recent years, deep neural network models have been developed to predict fault and assist diagnosis [1]–[4]. The success of deep learning-based methods heavily depends on a large number of training examples and the training and testing data follow identical data distribution and fault category [5]–[8]. In the actual industrial scenarios, the data distributions of the training and testing sets are usually different due to the variable working conditions and numerous mechanical components.

Manuscript received October 14, 2021; accepted October 25, 2021. Date of publication October 28, 2021; date of current version December 14, 2021. This work was supported in part by the Key Research and Development Plan of Anhui Province under Grant 202104a04020003, in part by the National Nature Science Foundation of China under Grant 61806067, and in part by the Fundamental Research Funds for the Central Universities under Grant PA2021KCPY0045. The associate editor coordinating the review of this article and approving it for publication was Dr. Andre E. Lazzaretti. *(Corresponding author: Xiaohui Yuan.)*

Juan Xu is with the Key Laboratory of Knowledge Engineering with Big Data, Hefei University of Technology, Hefei, Anhui 230000, China.

Yongfang Shi is with the School of Computer Science and Information, Hefei University of Technology, Hefei, Anhui 230000, China.

Xiaohui Yuan is with the Department Computer Science and Engineering, University of North Texas, Denton, TX 76207 USA (e-mail: xiaohui.yuan@unt.edu).

Siliang Lu is with the School of Electrical Engineering and Automation, Anhui University, Hefei, Anhui 230601, China.

Digital Object Identifier 10.1109/JSEN.2021.3123807

Collecting a large number of faulty examples model training and deriving a general model for all operating conditions and components are challenging if not impractical [9]. This situation poses an obstacle to the direct application of the existing deep learning methods.

To circumvent this problem, transfer learning has been employed that inherits the knowledge of the source domain (or problem) and refines the model with additional examples of the target domain to accommodate the discrepancy between data distributions. Transfer learning refers to a class of machine learning methods, which obtain some additional data or an existing model and apply it to new and relevant tasks. We can divide the data for transfer learning into two categories: source data and target data. Source data refers to additional data, which is not directly related to the task to be solved, while target data is directly related to the task. In a typical transfer learning model, source data is often huge, while target data is often small. Fig. 1(a) shows the existing transfer learning-based fault diagnosis problem, in which data are of the same or similar mechanical component. Hence, the training and testing cases share the same fault categories.

However, a more practical industrial scenario is that we can only obtain a large amount of labeled fault data of one component to train a model and need a model to identify the fault types of another different component with much fewer labeled examples. Due to the different fault categories in target

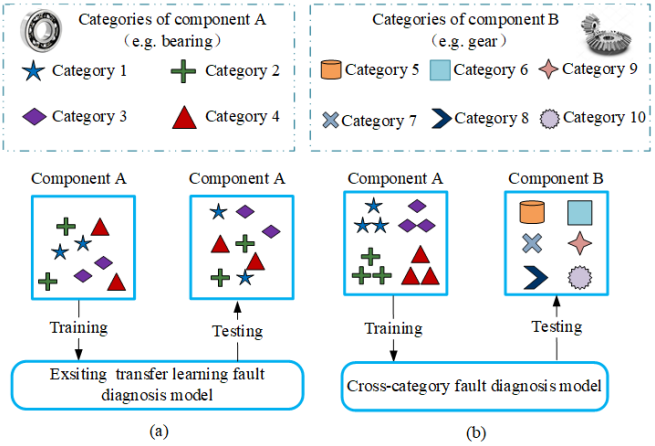


Fig. 1. Intelligent fault diagnosis: (a) existing transfer learning-based fault diagnosis model; (b) cross-category fault diagnosis model.

components, the well-developed transfer learning methods are less capable of addressing this issue. We call such learning tasks that develop models for datasets that include drastic different fault categories the *cross-category fault diagnosis* problems. A graphical illustration is shown in Fig. 1(b). The training data and testing data are of different mechanical components. The fault categories in these two datasets are different. To our best knowledge, limited studies can be found on the cross-category fault diagnosis.

Few-shot learning methods model new categories from few labeled examples, which is a promising approach to solve the cross-category fault diagnosis problem. Few-shot learning is committed to learning from a limited number of examples with supervised information. Usually, one considers the N-way-K-shot classification, in which the training set contains $I = K*N$ examples from N classes each with k examples, k is usually less than 20. Transfer learning is one of the implementation methods to solve the problem of few-shot learning. In recent years, studies begin to apply few-shot learning methods for fault diagnosis. However, the current fault diagnosis methods based on few-shot learning deal with fault diagnosis of the same component with insufficient examples [10]–[14]. Thereby it is urgent to explore an effective cross-category fault diagnosis approach.

This paper proposes a few-shot learning-based cross-category fault diagnosis method (CFDM). We model the metric space as an embedding function with a convolutional Siamese neural network. The fault features from sample pairs are extracted through the network, then the distance distribution between example pairs is derived through learning, rather than the data distribution of the training examples. Thereby the output of the model is a probability of the inputs belonging to the same class. Specifically, the training example pairs are drawn from a mechanical component, e.g., a bearing, whereas the testing example pairs are from a different mechanical component, e.g., a gear. Then the model compares the high-dimensional feature distance between the testing data and all the samples in the support set to get the most similar sample in the support set. Then the model predicts the fault category of the testing data by comparing the distance to the paired fault examples of the support set.

The main contributions of this study are summarized as follows:

1) Different from the existing methods, the proposed method leverages training examples of one mechanical component and achieves fault classification of a different mechanical component. The data distribution of the training and testing sets differs and the classes (i.e., fault categories) could be disjoint. We name such problems as cross-category fault diagnosis, which is common in real-world applications but has not yet been studied.
2) The proposed method extends the Siamese network and builds upon the deep convolutional neural network to capture invariances to transformation in the input space and use a large number of training examples to prevent overfitting. The training of the network derives a model of the metric space of temporal data (e.g., vibration signals), which enables few-shot learning for fault diagnosis of mechanical components.
3) A novel loss function is designed using parameter weighting factor and distance factor. By reducing the intra-category distance and increasing the inter-category distance, an accurate optimization classification boundary is obtained.

The rest of the paper is organized as follows. Section II includes the related works. Section III presents our proposed method. Section IV discusses the experimental results. Section V concludes this paper with a summary.

II. Related Work

The prevalent approaches to deal with the intelligent mechanical fault diagnosis with the insufficient labeled sample are based on Generative Adversarial Network (GAN) and transfer learning mechanism.

GAN-based fault diagnosis approaches lie in augmentations of the available data by GANs [15]. The most published works can be divided into two directions. One approach is using unlabeled data for model-assisted training to improve the fault diagnosis accuracy by semi-supervised GAN [16]–[18]. Alternatively, GANs are used to create synthetic examples to enrich the training set [19], [19], [20]. Wang *et al.* [21] proposed a conditional Variational auto-encoder generative adversarial network to solve the imbalanced fault diagnosis. Wang *et al.* [22] proposed a fault diagnosis method that combined GAN and Stacked Auto Encoder (SAE).

These GAN-based methods have achieved satisfactory performance since training and testing data follow a similar or even identical data distribution. Otherwise, transfer learning offers an alternative method to transfer source domain knowledge to solve fault classification tasks in target domains with the different data distribution [23]–[25].

There are two general strategies of transfer learning-based fault diagnosis methods: pre-trained transfer methods and domain adaptation methods. The pre-trained transfer methods fine-tune the parameters of the specified layer in the model trained with source domain data, and the model

identifies the faults of the target domain [26]. The domain adaptation method minimizes the probability distribution distance of the two domains to realize cross-domain fault recognition [27]–[29]. Li *et al.* [30] used a deep convolutional neural to extract the fault features from the bearing under different operating conditions and then minimize the maximum mean difference to learn transferable feature distribution. Lei *et al.* [31] transfer the fault knowledge of the laboratory bearings to locomotive bearings by using multi-layer domain adaptation and pseudo-label learning to reduce the distance of probability distribution of different domains.

While a number of studies can be found on diagnostic knowledge transformation across different domains, they still focus on the shift problem with respect to different operating conditions of the same components, i.e., rely on the hypothesis that the data of the two domains must be from the same category space.

The current aforementioned two studies both deal with cross-domain fault diagnosis of the same component with insufficient examples. That is, testing set D_{test} is small, meanwhile training set D_{train} and testing set D_{test} are different data distributions with the same set of fault categories. Formally, $I(p(D_{train}); p(D_{test})) = 0$, and $Y_{train} = Y_{test}$. Y_{train} and Y_{test} denote the class sets of and training set and testing set, respectively.

However, our work deals with a completely different problem called Cross-category fault diagnosis, that is, the target problem has a different set of fault categories. Formally, $I(p(D_{train}); p(D_{test})) = 0$, and $Y_{train} \cap Y_{test} = \emptyset$.

To our best knowledge, limited studies have been conducted and reported so far. Thereby it is urgent to explore an effective cross-category fault diagnosis approach.

Few-shot learning is committed to understanding new categories from tiny labeled samples, which has demonstrated success in the field of image classification [32]–[34]. Some implementation approaches include metric-based [35]–[37], model-based [38], optimization-based [39], and Graph Neural Network (GNN) methods [40], [41].

In the field of fault diagnosis, few-shot learning-based methods are newly developed. Hu *et al.* [42] proposed a data augmentation algorithm based on the Order Tracking and present a self-adaptive convolutional network for fault diagnosis. The few-shot experiments using two bearing databases validated the generalization ability of the model. Ren *et al.* [43] proposed a capsule auto-encoder model (CaAE) for intelligent fault diagnosis. A bearing dataset is utilized to validate the performance of the proposed CaAE. Wu *et al.* [44] considered two transfer situations of rotating machinery intelligent diagnosis named conditions transfer and artificial-to-natural transfer and construct seven few-shot transfer learning methods based on a unified 1D convolution network for fault diagnosis of three bearing datasets.

Although these studies have proposed new application scenarios, they still focus on the cross-domain fault diagnosis problem with insufficient label data. Thereby the existing few-shot learning-based methods cannot be well adapted in the cross-category fault diagnosis. Nonetheless, it is still promising to exploit the relationships of different components and extract more generalized fault knowledge, where tiny data of the target component are available.

III. Cross-Category Fault Diagnosis Network

Without loss of generality, the problem of cross-category fault diagnosis concerns two mechanical components A and B that are of different types, e.g., bearings and gears. Abundant samples of component A are acquired and labeled, which are denoted with set $D = \{x_i, y_i\}_{i=1}^{M}$ where x_i and y_i denote an acquired fault vibration data and its fault category (i.e., label), respectively. The fault categories form a category set $Y = \{y^1, y^2, \ldots, y^L\}$. Fault vibration data of the mechanical component B are also acquired, among which limited samples are labeled. The data set of component B is denoted with $\tilde{D} = \{\tilde{x}_i, \tilde{y}_i\}_{i=1}^{N}$, and its category set is $\tilde{Y} = \{\tilde{y}^1, \tilde{y}^2, \ldots, \tilde{y}^K\}$. The data sets and category sets satisfy the following conditions:

$$\begin{cases} I(p(D); p(\tilde{D})) = 0 \\ Y \cap \tilde{Y} = \emptyset, \end{cases} \tag{1}$$

where $I(\cdot)$ computes the mutual information of two distributions $p(D)$ and $p(\tilde{D})$. That is, the data distributions of D and $\tilde{D}$ are different and the category sets Y and $\tilde{Y}$ are disjoint.

The cross-category fault diagnosis can be formulated as a few-shot learning problem (a.k.a., N-shot, K-way classification). Following the idea of the Siamese neural network that models the distance distribution between sample pairs, we model the metric space as an embedding function with a deep network. The output of this network is a probability of the two inputs belonging to the same class.

Let's randomly divide dataset D into training sample pair set T. A model $f(x; \theta)$ trained with sample pairs T, i.e., the training sample pairs maps the input pair of examples, x and x', to a target label $y \in Y$ by maximizing the pairwise similarity $p(x, x')$:

$$f(x; \theta) : (x, x') \rightarrow y, \tag{2}$$

where $x', x \in T$, and

$$\theta = \arg\max_{\theta} \sum_{i=1}^{M} p(x_i, x_i').$$

When identifying the fault categories of the component B, we divide the dataset $\tilde{D}$ into a support set $\tilde{S}$ and a testing set $\tilde{T}$. $\tilde{S}$ consists of few labelled samples, and the testing set $\tilde{T}$ is the rest unlabelled samples, where $\tilde{D} = \tilde{S} \cup \tilde{T}$ and $\tilde{S} \cap \tilde{T} = \emptyset$. In this paper, there are at most five labeled fault samples in each category to construct support set $\tilde{S}$. In the Five-shot, K-way test, the support set is $\tilde{S}_5 = \{\tilde{S}^1, \tilde{S}^2, \ldots, \tilde{S}^5\}$, and more specifically,

$$\tilde{S}_5 = \begin{bmatrix} (\tilde{x}_1^1, \tilde{y}^1), (\tilde{x}_1^2, \tilde{y}^2), \ldots, (\tilde{x}_1^K, \tilde{y}^K) \\ (\tilde{x}_2^1, \tilde{y}^1), (\tilde{x}_2^2, \tilde{y}^2), \ldots, (\tilde{x}_2^K, \tilde{y}^K) \\ \vdots \\ (\tilde{x}_5^1, \tilde{y}^1), (\tilde{x}_5^2, \tilde{y}^2), \ldots, (\tilde{x}_5^K, \tilde{y}^K) \end{bmatrix}, \tag{3}$$

a randomly selected sample from the testing set is paired with an example in the support set of component B and used as

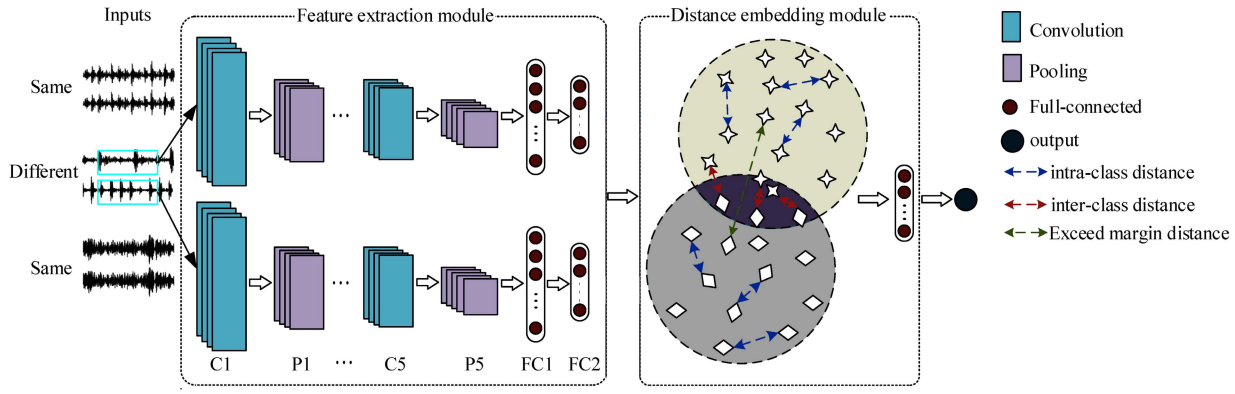


Fig. 2. The architecture of the cross-category fault diagnosis network.

the input to the model. For an instance $\tilde{x}^i$, its classification is decided by comparing the embedding distance against the support set, then the formal representation of the training sample pair constructed by the i-th test data is as follows:

$$\left\{ \begin{array}{l} \left(\tilde{x}^i, \tilde{x}_1^1\right), \left(\tilde{x}^i, \tilde{x}_1^2\right), \ldots, \left(\tilde{x}^i, \tilde{x}_1^K\right) \\ \left(\tilde{x}^i, \tilde{x}_2^1\right), \left(\tilde{x}^i, \tilde{x}_2^2\right), \ldots, \left(\tilde{x}^i, \tilde{x}_2^K\right) \\ \vdots \\ \left(\tilde{x}^i, \tilde{x}_5^1\right), \left(\tilde{x}^i, \tilde{x}_5^2\right), \ldots, \left(\tilde{x}^i, \tilde{x}_5^K\right) \end{array} \right\}, \tag{4}$$

among them, the probability of the $\left(\tilde{x}^i, \tilde{x}_n^k\right)$ model output is the largest, then the fault category of the data to be tested is the label corresponding to $\tilde{x}_n^k$, where $1 \leq n \leq 5$ and $1 \leq k \leq K$. When $N = 1$, it is One-shot, K-way test. The support set $\tilde{S}$ contains K categories, and each category contains one sample $\tilde{x}_k$, $\tilde{y}_k$ is the label:

$$\tilde{S}_1 = \{(\tilde{x}^1, \tilde{y}^1), (\tilde{x}^2, \tilde{y}^2), \ldots, (\tilde{x}^K, \tilde{y}^K)\}. \tag{5}$$

The structure of the training example pairs from the i-th test case is as follows:

$$\left\{\left(\tilde{x}^i, \tilde{x}_1^1\right), \left(\tilde{x}^i, \tilde{x}_1^2\right), \ldots, \left(\tilde{x}^i, \tilde{x}_1^K\right)\right\}. \tag{6}$$

Model $f(x;\theta)$ is employed to predict the similarity of samples against the limited data set in $\tilde{D}$

$$f(x;\theta) : (x, \tilde{x}) \rightarrow \tilde{y}, \tag{7}$$

where $x \in \tilde{T}$ and $\tilde{x} \in \tilde{S}$. The predicted fault type of component B is hence decided by the label of $\tilde{x}$ that yields the greatest similarity.

A. Network Architecture

Our proposed Cross-category Fault Diagnosis Model (CFDM) consists of a feature extraction module and a distance embedding module. The feature extraction module extends the convolution Siamese neural network to learn fault features. The distance embedding module facilitates the learning of distinguishable features via distance embedding. Fig. 2 illustrates the architecture of the network.

The network consists of normalized convolutional layers, pooling layers, and normalized fully connected layers. For convenience, we use C, P, and FC to denote normalized convolutional layers, pooling layers, and normalized fully connected layers, respectively. The input of the network is the sample pair, which is processed with five consecutive convolutional and pooling operations. The output of pooling layer five (P5) is flattened into a vector and fed into the fully connected layers. Note that each input example is processed separately through this network. The output of the fully connected layers is the key fault features extracted, which is processed in the distance embedding module to calculate the distance between the features of the two inputs. If this distance is significant enough, the sample pair belongs to different categories of faults; otherwise, it belongs to the same fault category. Table I lists the key parameters of the CFDM.

TABLE I
CFDM LAYERS AND PARAMETERS

Layer	Kernel Size	Stride	Kernel Number	Padding
Input	-	-	-	-
C1	64x1	16	16	same
P1	2x1	2	-	-
C2	3x1	1	32	same
P2	2x1	2	-	-
C3	3x1	1	64	same
P3	2x1	2	-	-
C4	3x1	1	64	same
P4	2x1	2	-	-
C5	3x1	1	64	valid
P5	2x1	2	-	-
FC1	100	-	1	-
FC2	50	-	1	-

B. Loss Function

We present a weighted feature discrepancy metrics loss function that integrates a weighting factor and the distance factor to reduce the intra-category distance and increase the inter-category distance. This loss function highlights the impact of the small difference of fault vibration data on the training process to get a more accurate fault classification boundary.

In the training phase, example pairs are randomly drawn from data set D. Let χ denote a pair of examples x_u and x_v that are of the same fault category: $\chi = \{x_u, x_v, 1\}$, where $x_u \in D$, $x_v \in D$. Let $\hat{\chi}$ denote a pair of examples that are of different

fault categories: $\hat{\chi} = \{x_u, x_v, 0\}$, where $x_u \in D$, $x_v \in D$. The difference between χ and $\hat{\chi}$ is the label that indicates the same or different category. The training set T consists of a balanced number of pairs of examples of the same and different categories: $T = \{\chi_1, \chi_2, \ldots, \chi_{\frac{M}{2}}, \hat{\chi}_1, \hat{\chi}_2, \ldots, \hat{\chi}_{\frac{M}{2}}\}$. This training set is used to build a model of pairwise distance.

In the testing phase, a randomly selected sample is paired with the example in the support set of component B and used as the input to the model. For an instance $\tilde{x}$, its classification is decided by comparing the embedding distance against the support set.

The distance of the output of the feature vectors from the Siamese neural network is computed as follows:

$$\Phi(x_u, x_v, \theta) = \text{sigmoid}(FC(\rho(f(x_u), f(x_v)), \theta)) \quad (8)$$

where $\rho(f(x_u), f(x_v))$ denotes the distance metric between the fault features of sample pairs. There are many discrepancy measure methods for feature vectors, including Euclidean Distance, Cosine Distance, Mean Relative Error, and Pearson Correlation Coefficient. Variable θ denotes the parameters of the network.

To highlight the impact of the gentle difference of fault vibration data, we introduce a weighting factor α for fault samples of the same category to reduce the intra-category distance and a minimum margin distance factor m_r to increase the inter-category distance. The range of α is $[0, 1]$. m_r is applied to the default samples of the different categories and its range is $[0, 1]$ to ensure the distance is at least m_r. The loss function of our model is as follows:

$$\begin{aligned}\mathcal{L}(x_u, x_v, \theta) = \lambda^T \Sigma_i^l \theta_i^2 - [y \log(\Phi(x_u, x_v, \alpha\theta)) \\ + (1-y) \log(1-\min(m_r, \Phi(x_u, x_v, \alpha\theta)))]\end{aligned} \quad (9)$$

where θ is a vector of the set of model parameters and λ is a hyperparameter of the model for the regularization, l is the number of the network parameters and y is the label.

When the two samples are of the same class, i.e., $y = 1$, the third term of the loss function becomes zero. Hence, the loss depends on the probability of χ. When the two samples are of different classes, i.e., $y = 0$, the second term of the loss function becomes zero and the loss depends on the probability of $\hat{\chi}$. By introducing α ($\alpha < 1$), the range of network outputs is suppressed, which essentially increases the normalized, quantized outputs (i.e., probability). Hence, reduced intra-class distance is favored in the training process. The scalar m_r increases the penalty to the examples of the 'dissimilar' class (i.e., $y = 0$). As the cross-entropy based loss function decreases, the probability with which the inter-class distance is greater than the intra-class distance increases [45].

C. Training and Application of CFDM

Training and application of our proposed CFDM include three phases: constructing the training set from component A, building and training the model, and testing the fault samples from component B. The flowchart is shown in Fig. 3.

1) Constructing the training data set: Following aforementioned the sample pairwise method, we select the fault data of the same category or different categories. We enumerate all the sample pairs to construct the training set.
2) Training the CFDM model: We enumerate all pairwise comparisons about the feature distance discrepancy between sample pairs of component A to enrich the supervised information of categories in the training experience. By calculating the distance between the high-dimensional features extracted from the sample pair, the model learns whether the sample pair belongs to the same fault category or not according to the feature distance. In this paper, an adaptive moments algorithm Adam [46] is adopted to optimize the network parameters.
3) Applying the trained model to the fault samples: the samples in testing set $\tilde{T}$ are paired with the samples in support set $\tilde{S}$ as input sample pairs. The model compares the high-dimensional feature distance between the testing data and all the samples in the support set to get the most similar sample in the support set. Then the testing data belongs to the fault category of the most similar sample in the support set, i.e. the label $\tilde{Y}$.

IV. Experimental Results

A. Datasets

The performance of our proposed CFDM is evaluated with three datasets: the bearing dataset from the Lab-built testbed, the bearing datasets from Case Western Reserve University (CWRU) [31] and the gearing dataset [47], as shown in Fig. 4.

1) Bearing Dataset From Lab-Built Testbed: The lab-built testbed is shown in Fig. 4(a). The three-phase motor through a flexible coupling controls the speed of bearing, and an acceleration sensor is used to collect vibration signals as the training dataset. The experimental bearing contains four faults: rolling body fault (BF), inner ring fault (IF), outer ring fault (OF), and normal status under this condition. We simulate the actual working conditions on our testbed and acquire the vibration signals from four categories of fault bearings. The sampling frequency is 128kHz.

2) CWRU Bearing Dataset: The testbed consists of a 2 HP motor (left), a torque sensor (middle), a dynamometer (right), electronic control equipment, and acceleration sensors to collect vibration signal, as shown in Fig. 4(b). This dataset is one of the most commonly used benchmark datasets in the field of fault diagnosis. Single point pitting faults are arranged on the bearings using EDM technology with fault diameters of 0.007, 0.014, and 0.021 inches, respectively. Each fault diameter contains three faults: rolling body fault, inner ring fault, outer ring fault, and normal status.

3) Gearing Dataset: The gearing testbed is shown in Fig. 4(c). A 32-tooth pinion and an 80-tooth gear are mounted on the first stage input shaft. The second stage consists of a 48-tooth pinion and a 64-tooth gear. The gear speed is controlled by a motor. The torque is provided by an electromagnetic brake, which can be adjusted by changing its input voltage. The speed of the input shaft is measured by a tachometer, and the vibration signal was collected by

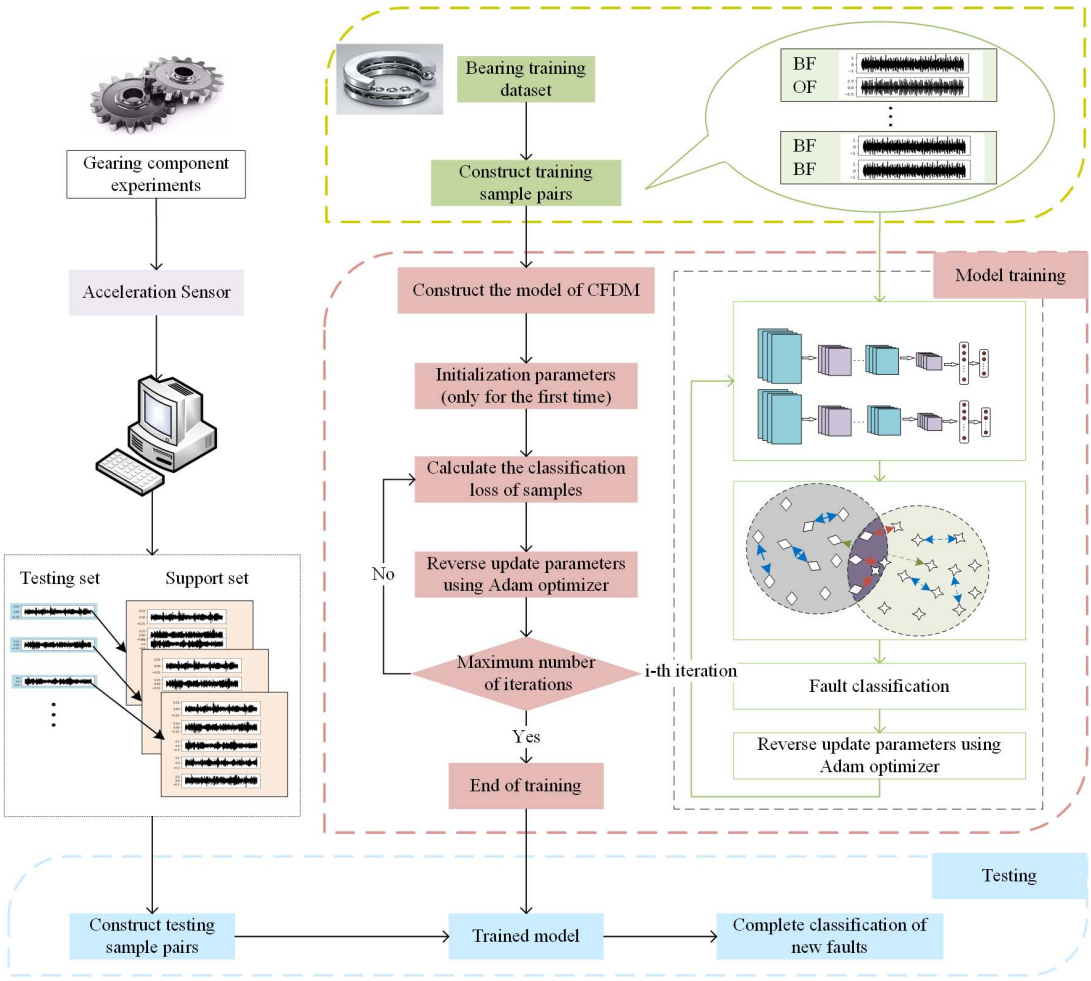


Fig. 3. Flowchart of the training and testing of the proposed method.

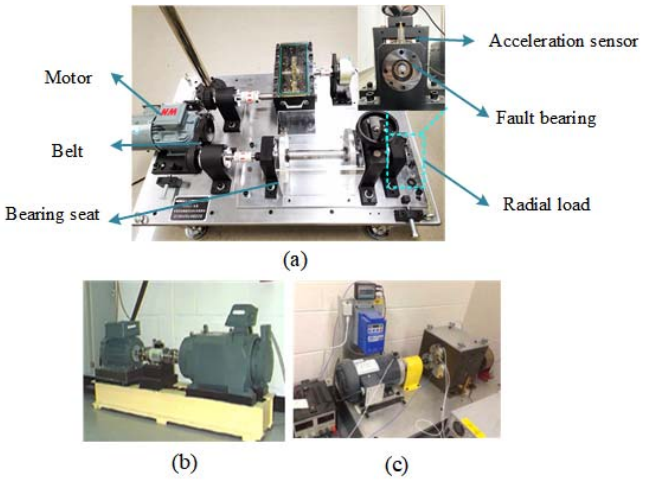


Fig. 4. Bearing and gearing platforms: (a) Lab-built experimental bearing platform, (b) CWRU bearing platform, and (c) Gearing experimental platform.

an accelerometer. The signals were recorded by the dSPACE (DS1006 processor board, dSPACE Inc.) The sampling frequency is 20 kHz. The Gearing Dataset contains five categories: Health, Missing, Spall, Chip, and Crack.

B. Fault Classification of Different Diagnosis Task

We conduct three cross-category fault diagnosis experiments between the bearing and gearing. The three diagnosis tasks are shown in Table II. For example, the training set of Task A is the CWRU bearing dataset, which contains vibration samples of 10 different fault categories under three kinds of fault size, 0.007, 0.014, and 0.021 inches. The testing set is a gear dataset, which contains vibration samples of 5 different fault categories.

We set up eight cross-category fault diagnosis experiments for each task to verify the fault classification accuracy of the proposed method. For each series of experiments, we repeat the experiment ten times and report the average performance.

The number of sample pairs of the training and testing sets of the experiments is listed in Table III. One-shot90 indicates that the training set contains 90 labeled sample pairs of component A with a length of 2048 time-domain data points. In the training sample pairs, half of them are the same fault category, and the rest are randomly matched with different kinds of faults.

The testing set of Task A and Task B contains 9000 vibration samples of component B, while the testing set of Task C contains 6600 vibration samples of component B. The support set of One-shot from the testing set consists

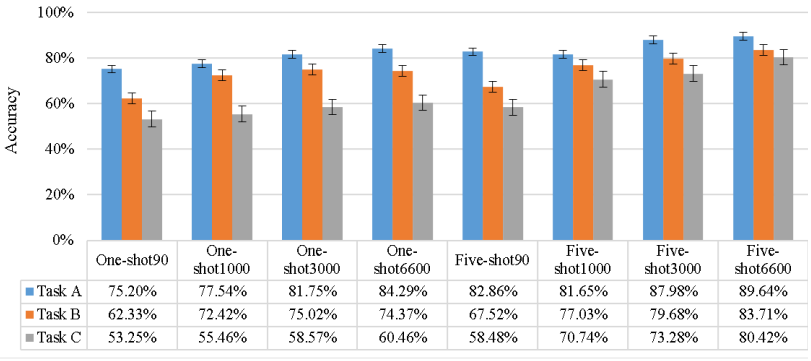


Fig. 5. Accuracy of fault classification on different diagnosis tasks.

TABLE II
DESCRIPTION OF THREE CROSS-CATEGORY FAULT DIAGNOSIS TASK

Cross-category diagnosis task		Fault category	
		Component A	Component B
Task A	CWRU bearing to Gearing	0.007OF, 0.007BF, 0.007IF, 0.014OF, 0.014BF, 0.014IF, 0.021OF, 0.021BF, 0.021IF, Normal	Health Missing Spall Chip Crack
Task B	Lab-build bearing to Gearing	OF BF IF Normal	Healthy Missing Spall Chip Crack
Task C	Gearing to CWRU bearing	Healthy Missing Spall Chip Crack	0.007OF, 0.007BF, 0.007IF, 0.014OF, 0.014BF, 0.014IF, 0.021OF, 0.021BF, 0.021IF, Normal

TABLE III
NUMBER OF EXAMPLES IN THE TRAINING (trn) AND TESTING (tst) DATASETS

Strategy	Task A		Task B		Task C	
	trn	tst	trn	tst	trn	tst
One-shot90	90	9000	90	9000	90	6600
One-shot1000	1000	9000	1000	9000	1000	6600
One-shot3000	3000	9000	3000	9000	3000	6600
One-shot6600	6600	9000	6600	9000	6600	6600
Five-shot90	90	9000	90	9000	90	6600
Five-shot1000	1000	9000	1000	9000	1000	6600
Five-shot3000	3000	9000	3000	9000	3000	6600
Five-shot6600	6600	9000	6600	9000	6600	6600

of one labeled example for each fault category. The support set of Five-shot only has five labeled samples for each fault category.

Fig. 5 illustrates the average classification accuracy and the standard deviation of the three tasks. It can be seen that the classification accuracy of all cross-category fault diagnosis experiments is more than 53% with respect to one or five labeled samples for each fault category, which shows that the proposed CFDM is effective. The fault classification accuracy of Task A is higher than that of Task B and Task C. The highest accuracy of 8 experiments in Task A is as high as 85%. The classification accuracy in Task C is the lowest. The lowest classification accuracy of 8 experiments in Task C is 53.25%. The reason is that the fault categories of the training set are less than that of the testing set. When the sample numbers of the training set are insufficient, the model learned by the training set is not robust enough for fault classification in the testing set.

We can also see that experiment results of Five-shot are generally about 5% higher than that of One-shot in the three tasks. For example, in the One-shot experiments of Task A, the fault classification accuracy increases from 75.2% to 84.29%, while in the Five-shot experiments of Task A, the fault classification accuracy increases from 82.86% to 89.64%. These results are also reasonable, because as the number of labeled samples increases, the classification boundary between categories becomes more obvious, which leads to higher classification accuracy.

In addition, the capacity influence of the training set on the fault classification accuracy is also discussed. As can be seen from Fig. 5 that the fault classification results are more accurate with the increase in the number of training samples.

Fig. 6 depicts the confusion matrix of the fault classification results of Task A, Task B, and Task C on Five-shot learning using 6600 training example pairs. The testing set of Task A and Task B contains 9000 samples of component B and the testing set of Task C contains 6600 samples of component B. The support set of Five-shot has five labeled examples for each category. The abscissa of the confusion matrix is the predicted label and the ordinate is the ground truth. For tasks A and B, the accuracy for each class is mostly greater than 70% with some in the 90% range as shown in Fig. 6(a) and Fig. 6(b). The accuracy of detecting 'Spall' and 'Crack' faults are superior. In task A, the accuracy of detecting 'Spall' achieved 100%. The majority of confusion is between 'missing' and 'chip' and 'missing' and 'health', which drive the accuracy of 'missing' to 68%. In Fig. 6(c), the classification accuracy of seven out of ten classes is greater than 72% and the accuracy of four classes is 100% or very close to 100%.

Noted that three fault types 007BF, 014BF, and 021BF in task C have relatively low accuracy, rather than other faults. The reason is partly that the number of categories differentiable in the training phase is less than that in the testing phase. In addition, the fault data with different diameters are similar in the CWRU bearing dataset, which makes correct classification much challenging. Specifically, Fig. 7 shows the time domain vibration signal of the CWRU bearing.

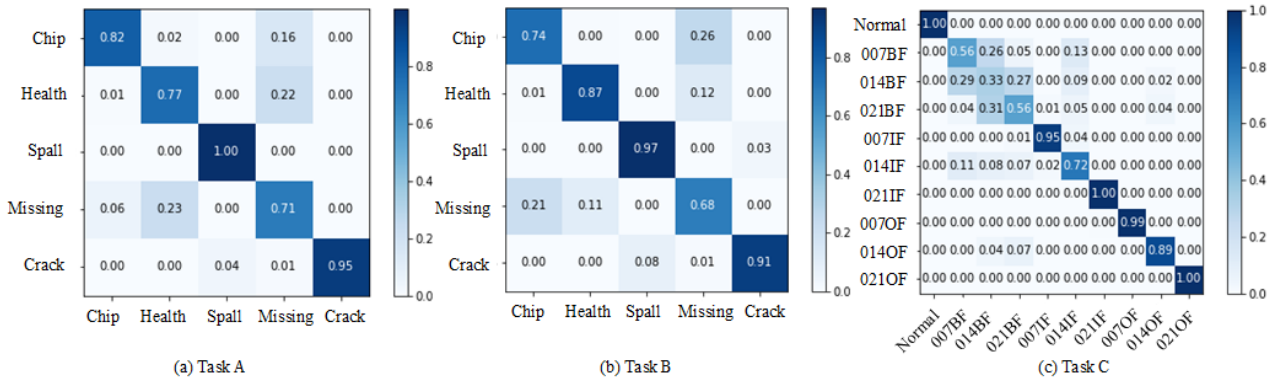


Fig. 6. The confusion matrix of fault diagnosis of the case of five-shot6600.

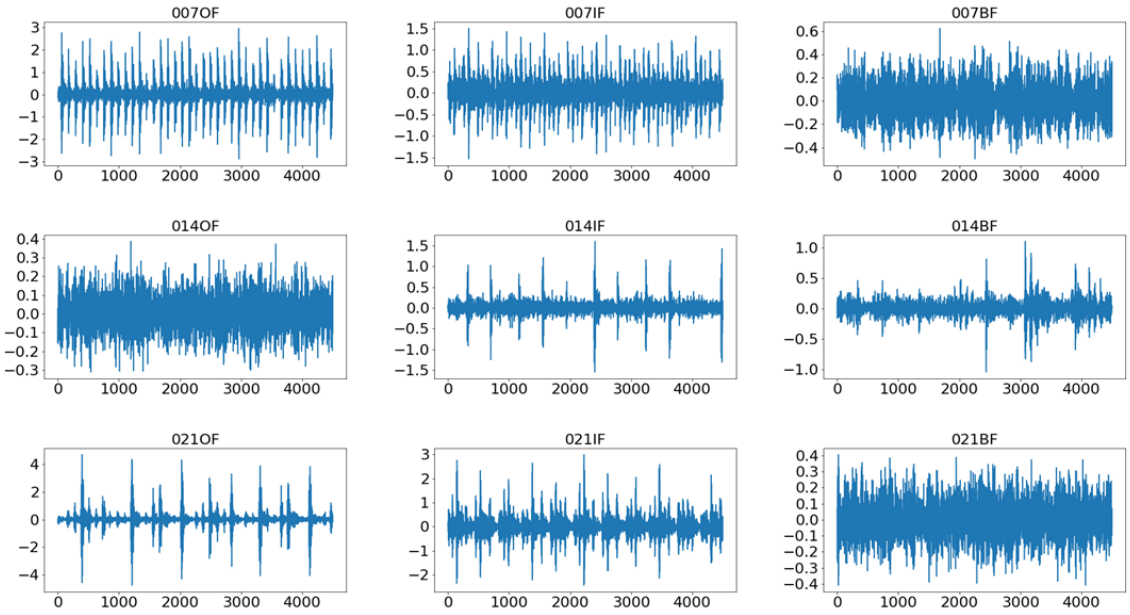


Fig. 7. The original time-domain vibration signals of CWRU bearing.

The amplitude range of the signals of BF is significantly smaller than that of IF and OF. For instance, the maximal amplitude range of 0F is $(-4, 4)$, while the maximal amplitude range of 0F is $(-1, 1)$. Noted that the amplitude range of BF with different diameters is similar. Thus, the difference in data distribution of BF with different diameters is weak. It is difficult for the model to distinguish between 007BF, 014BF, and 021BF, rather than other faults. For instance, the probabilities of misclassification of 014BF to 007BF and 021BF are 0.29 and 0.27, respectively.

C. Parameter Analysis

The weight factor α and the distance factor m_r of our method are empirically determined. In this section, we evaluate the impact of various choices on the accuracy by varying α in the range of $[0.05, 1]$ and m_r in the range of $[0, 1]$. Without loss of generality, we conducted experiments with task A. Fig. 8 illustrates the box plot of accuracy with respect to various m_r. At each m_r, we vary α and the data distribution is depicted.

Compared with the model without the distance and parameter weight factors, as shown in Fig. 5, the fault classification accuracy improves when the distance and parameter weight factors are reasonably adopted. For example, in the case of one-shot 90, when $m_r = 0.5$ the average accuracy of fault classification is over 80%, which is improved by about 5% than that of the model integrating the two factors. This demonstrates that measuring the inter-category and intra-category distances is important for fault classification. When we combine five different distance factors and seven parameter weight factors for multi-scale analysis, the network achieved a stronger fault classification ability.

In the four cases of one-shot testing, when $m_r = 0.7$, with the change of α, the fault classification accuracy is higher than other values of m_r, and the change range of the accuracy is the least. Therefore, the model is most stable when $m_r = 0.7$. In the four cases of five-shot testing, when $m_r = 0.5$, with the change of α, the fault classification accuracy is higher than m_r is set to other values, and the change range of the accuracy is the least. Therefore, the model is most stable when $m_r = 0.5$.

TABLE IV
THE DESCRIPTION OF THE COMPARISON METHODS

Methods	Parameters setting
SAE	Qi et al. [48], proposed a stacked sparse autoencoder-based machine fault diagnosis method. In our comparative experiment, we use two layers of a fully connected encoder and decoder to form two stacks. The first autoencoder uses a 4096-dimension input layer and 400-dimension hidden layer, and then takes the output of the first encoder as the input of the second encoder, the second autoencoder uses a 100-dimensional hidden layer and finally outputs a 4-dimensional fault feature for classification.
CNN-SVM	The CNN-SVM system is applied in bearing fault diagnosis by Han et al. [49]. The features are extracted by CNN and realize the final bearing state recognition by SVM. We use CNN of three-layer convolution and full connection layer as feature extraction. The Dropout layer avoids overfitting. Then the model is mapped to 4-dimensional feature space. Finally, SVM was used for fault classification.
WDCNN	A wide Deep Convolutional Neural Network (WDCNN) is proposed by Zhang et al. [50]. It is a Siamese CNN structure with a large convolution kernel at the first layer. It is proved for the first time that the fault diagnosis model based on few-shot learning can make full use of the same or different sample pairs to identify test samples from the classes with only one or a few samples, to improve the performance of fault diagnosis. More details about WDCNN can be found in [50].
SSGAN	SSGAN is a semi-supervised generated adversarial network proposed by Odena in 2016 [51]. The main idea of SSGAN lies in the design of a discriminator, which can not only play the role of classifier for the image classification task but also distinguish generated samples and real data generated by the generator. Verstraete et al. [52] proposed a semi-supervised generative adversarial network-based methodology for fault diagnostics. The layer configuration and hyper-parameters are presented in [52].
Transfer learning	Zhang et al. [53] proposed a transfer learning-convolutional neural network (TLCNN) based on AlexNet for bearing fault diagnosis. We take the classical convolutional neural network AlexNet as our network backbone. The labeled data of the source domain dataset (i.e., CWRU bearing dataset) is used to pre-train the network. Then we freeze the hyper-parameters of the first five layers of the network, and then use the target domain data, i.e., 5 or 25 labeled examples of gear dataset to fine-tune the hyper-parameters of the last three layers so that the model enables to generalize to the target domain.

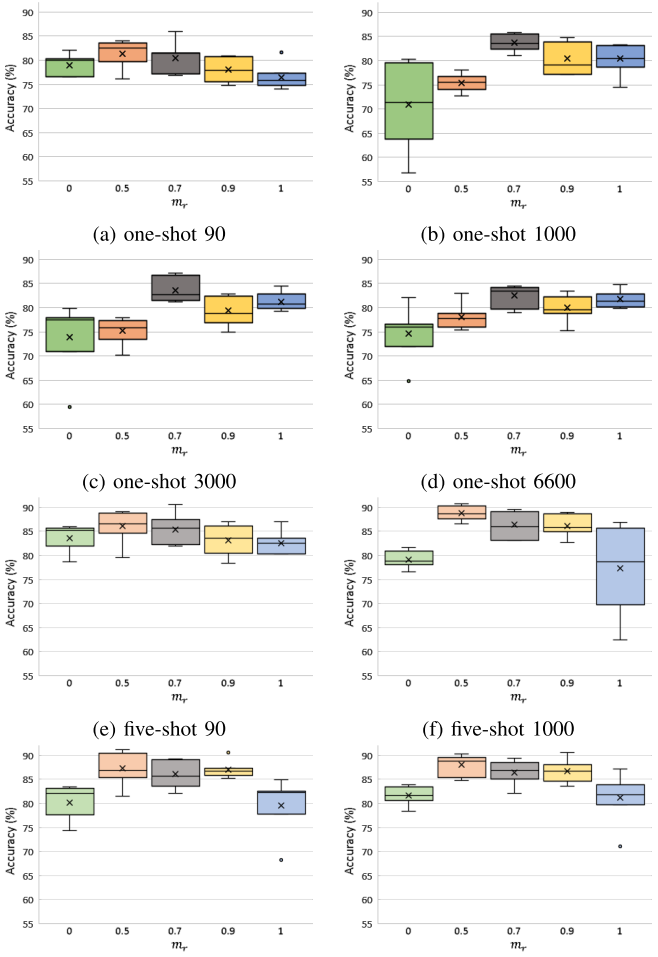


Fig. 8. Influence of distance and weighting factor.

When $m_r = 0$ and 1, the fault classification accuracy in eight experiments is lower than other values of m_r, and the change range of classification accuracy is also the largest. In addition, outliers tend to appear when $m_r = 0$ or 1, which implies unstable performance.

According to Eq.9, when $m_r = 0$, the loss function of our model becomes $\mathcal{L}(x_u, x_v, \theta) = \lambda^T \Sigma_i^l \theta_i^2 - y \log(\Phi(x_u, x_v, \alpha\theta))$. Likewise, when $m_r = 1$, the loss function becomes $\mathcal{L}(x_u, x_v, \theta) = \lambda^T \Sigma_i^l \theta_i^2 - [y \log(\Phi(x_u, x_v, \alpha\theta)) + (1 - y) \log(1 - \min(m_r, \Phi(x_u, x_v, \alpha\theta)))]$. In both cases, m_r does not work to restrain the samples of different classes, thereby the loss functions are unable to maximize the inter-category distance, which degrades the performance of the model. If the support set has only one labeled data for each category of fault, the larger m_r can increase the distance between different classes as much as possible, thereby the classification accuracy of the model is superior. If the support set has five or more labeled data for each category of fault, the loss function needs to strengthen the constraints of the same class; thereby the smaller m_r will make the performance of the model superior.

In summary, when $m_r = 0$ and 1, the output of the model has singular values and exhibits poor accuracy and stability. If the support set has only one labeled data for each category of fault, $m_r > 0.5$ is suitable.

D. Fault Classification Using Different Models

In this section, we compare the overall accuracy of our proposed CFDM method with state-of-the-art methods on task A. The compared models include SAE, CNN_SVM, WDCNN, and SSGAN. In addition, a transfer learning method is implemented based on AlexNet. In our evaluation, the training examples were randomly selected from the gearing dataset. The support set of the One-shot case has five labeled instances, and the support set of the Five-shot case has 25 labeled data. The description of the comparison methods is shown in Table IV.

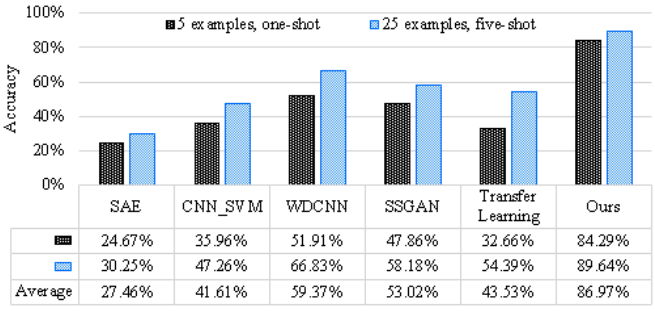


Fig. 9. The overall accuracy of the compared methods for fault classification.

TABLE V
THE FAULT CLASSIFICATION RESULTS OF DIFFERENT DISCREPANCY METRIC METHOD

Case	Mean Relative Error	Cosine similarity	Euclidean	Pearson Corr. Coef.
One-shot 90	64.8 (6.6)	74.0 (3.8)	**75.6 (19.7)**	75.2 (9.8)
One-shot 1000	73.4 (10.8)	74.6 (8.6)	76.5 (4.1)	**77.5 (3.6)**
One-shot 3000	75.8 (9.8)	77.46 (10.4)	80.25 (5.1)	**81.8 (8.9)**
One-shot 6600	79.6 (7.4)	77.56 (11.8)	80.5 (12.8)	**84.3 (7.4)**
Five-shot 90	69.9 (6.9)	79.16 (3.6)	80.9) (20.7)	**82.9 (8.3)**
Five-shot 1000	79.2 (10.0)	80.6 (7.1)	81.2) (4.2)	**81.6 (4.9)**
Five-shot 3000	82.4 (10.0)	83.8 (9.9)	86.9) (3.8)	**87.9 (8.3)**
Five-shot 6600	85.4 (5.2)	84.8 (10.7)	87.5) (9.2)	**89.6 (5.2)**

The overall accuracy of fault classification using the aforementioned methods is depicted in Fig. 9. For each method, two bars depict the accuracy of one-shot and five-shot results. The accuracies of each case as well as the average are also listed under the plot.

It is clear that by using more examples (i.e., Five-shot learning), greater accuracy is achieved by all methods. The improvement is as much as 66.5% (transfer learning based on AlexNet). In contrast, our method exhibits the minimum difference between one-shot and five-shot learning. This is partly due to the much superior performance and a smaller margin for improvement. Comparing across all methods, our method achieved the best accuracy in all cases. The overall accuracies for one-shot learning and five-shot learning are 84.29% and 89.64%, respectively. The improvement with respect to the second-best (i.e., WDCNN) is 62.4% in one-shot learning and 34.1% in five-shot learning. This demonstrates that the metric model derived with our proposed method from learning bearing samples is effective for the diagnosis of fault types of gearing.

The compared methods achieved inferior performance due to the small training set. Note that for a five-class classification problem, 20% is the average performance of random guess. Clearly, the figure illustrates that more training examples help. For neural networks, it is necessary to train with a large amount of labeled data to have a satisfactory accuracy. When there are only 5 or 25 labeled examples to train a model, it is difficult for the model to learn the data distribution of the target domain. Thereby the classification accuracy of the compared methods is unsatisfactory.

The important message of Fig. 9 is that learning data distribution is unsuitable for the cross-category classification problem, which is evident that the much lower accuracy was achieved by the compared methods.

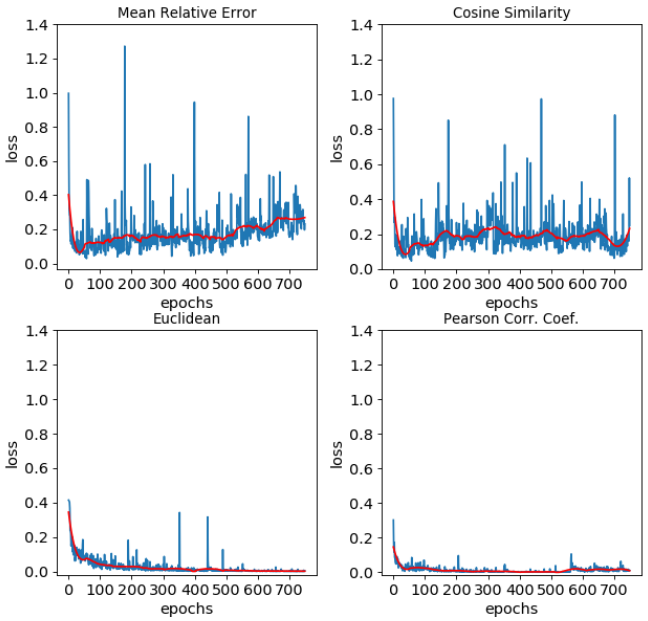


Fig. 10. The loss of different discrepancy metrics.

E. Fault Classification With Different Discrepancy Metrics

We conducted multiple sets of experiments on task A to evaluate the impact of loss functions using different discrepancy metric methods on fault classification accuracy. We used four discrepancy metric methods to construct different loss functions in our model. The training set and testing set are the same as those in Table III. The average accuracy for each method is shown in Table V. The numbers in parenthesis are the standard deviation. Among almost all cases, the fault classification results using the Pearson correlation coefficient yielded higher accuracy than the other three discrepancy metric methods. In the case of one-shot 90, Euclidean distance achieved the best accuracy at 75.6% that is slightly better than that of the Pearson correlation coefficient (75.2%).

Fig. 10 shows the loss using different discrepancy metrics during the training iterations. Loss values at every 20 epochs are plotted, which are illustrated with a curve in blue. The red curve presents the moving average of the loss, which greatly suppresses the spikes and visualizes the trend over time. Different discrepancy metrics have a great impact on the loss of the model.

Cosine similarity evaluates the similarity of two vectors by calculating the cosine of the angle between them. It is insensitive to the absolute value of the feature. The loss of cosine similarity fluctuates greatly, which performs similarly to Mean Feature Error.

Both Euclidean distance and Pearson correlation coefficient exhibited much better performance by converging to a smaller loss. In contrast to Euclidean distance, the Pearson correlation coefficient is more stable and its convergence appears to be faster.

The sample pairs of the fault vibration signals from different components are fine-grained recognition. The range of values between different categories of fault feature variables is independent of the distance, which makes the Pearson correlation coefficient is more suitable for the fault diagnosis model. From Fig. 10, we can find that the loss fluctuation using the Pearson correlation coefficient is the smallest.

V. Conclusion

This paper proposes a few-shot learning based cross-category fault diagnosis method. Different from the existing methods, the proposed method leverages training examples of one mechanical component and achieves fault classification of a different mechanical component. The data distribution of the training and testing sets differs and the classes (i.e., fault categories) could be disjoint. The proposed method extends the Siamese network to extract fault features from sample pairs. By training the deep network with example pairs, the method learns the distribution of distances between example pairs, rather than the data distribution of the training examples. The training of the network derives a model of the metric space of temporal data (e.g., vibration signals). When this model is applied to the testing cases, it predicts the fault types by comparing the distance to the paired fault examples of the support set. Experiments were conducted with two public datasets and a lab-built dataset. The results cross-category demonstrate that even if there was only one or five labeled examples in each category for the target component, the model achieved a superior generalization performance without model retaining or refinement. The proposed method provides a new idea for cross-category fault diagnosis with few labeled examples.

References

[1] S. Shao, R. Yan, Y. Lu, P. Wang, and R. X. Gao, "DCNN-based multi-signal induction motor fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 2658–2669, Jul. 2020.

[2] J. Xiong, Q. Liang, J. Wan, Q. Zhang, X. Chen, and R. Ma, "The order statistics correlation coefficient and PPMCC fuse non-dimension in fault diagnosis of rotating petrochemical unit," *IEEE Sensors J.*, vol. 18, no. 11, pp. 4704–4714, Mar. 2018.

[3] A. Choudhary, D. Goyal, and S. S. Letha, "Infrared thermography-based fault diagnosis of induction motor bearings using machine learning," *IEEE Sensors J.*, vol. 21, no. 2, pp. 1727–1734, Aug. 2021.

[4] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.

[5] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, Jan. 2016.

[6] Q. Hu, A. Qin, Q. Zhang, J. He, and G. Sun, "Fault diagnosis based on weighted extreme learning machine with wavelet packet decomposition and KPCA," *IEEE Sensors J.*, vol. 18, no. 20, pp. 8472–8483, Aug. 2018.

[7] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106587.

[8] Z. Wang, Q. Zhang, J. Xiong, M. Xiao, G. Sun, and J. He, "Fault diagnosis of a rolling bearing using wavelet packet denoising and random forests," *IEEE Sensors J.*, vol. 17, no. 17, pp. 5581–5588, Sep. 2017.

[9] Y. Fan, G. Shen, X. Xu, J. Xu, and X. Yuan, "Flight track pattern recognition based on few labeled data with outliers," *J. Electron. Imag.*, vol. 30, no. 3, Jan. 2021, Art. no. 031204.

[10] S. Wang, D. Wang, D. Kong, J. Wang, W. Li, and S. Zhou, "Few-shot rolling bearing fault diagnosis with metric-based meta learning," *Sensors*, vol. 20, no. 22, pp. 1–15, 2020.

[11] H. Yin, Z. Li, J. Zuo, H. Liu, K. Yang, and F. Li, "Wasserstein generative adversarial network and convolutional neural network (WG-CNN) for bearing fault diagnosis," *Math. Problems Eng.*, vol. 2020, no. 6, pp. 1–16, May 2020.

[12] C. Yu, Y. Ning, Y. Qin, W. Su, and X. Zhao, "Multi-label fault diagnosis of rolling bearing based on meta-learning," *Neural Comput. Appl.*, vol. 3, pp. 1–15, May 2020.

[13] H. Wang, X. Bai, J. Tan, and J. Yang, "Deep prototypical networks based domain adaptation for fault diagnosis," *J. Intell. Manuf.*, pp. 1–11, Nov. 2020.

[14] Z. Qiao, X. Yuan, C. Zhuang, and A. Meyarian, "Attention pyramid module for scene recognition," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 7521–7528.

[15] F. Zhou, S. Yang, H. Fujita, D. Chen, and C. Wen, "Deep learning fault diagnosis method based on global optimization GAN for unbalanced data," *Knowl.-Based Syst.*, vol. 187, Jan. 2020, Art. no. 104837.

[16] N. Huang, Q. Chen, G. Cai, D. Xu, L. Zhang, and W. Zhao, "Fault diagnosis of bearing in wind turbine gearbox under actual operating conditions driven by limited data with noise labels," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.

[17] Y. Ding, L. Ma, J. Ma, C. Wang, and C. Lu, "A generative adversarial network-based intelligent fault diagnosis method for rotating machinery under small sample size conditions," *IEEE Access*, vol. 7, pp. 149736–149749, 2019.

[18] Z. Li, T. Zheng, Y. Wang, Z. Cao, Z. Guo, and H. Fu, "A novel method for imbalanced fault diagnosis of rotating machinery based on generative adversarial networks," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–17, 2021.

[19] S. Shao, P. Wang, and R. Yan, "Generative adversarial networks for data augmentation in machine fault diagnosis," *Comput. Ind.*, vol. 106, pp. 85–93, Apr. 2019.

[20] Y. Gao, X. Liu, and J. Xiang, "FEM simulation-based generative adversarial networks to detect bearing faults," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4961–4971, Jul. 2020.

[21] Y.-R. Wang, G.-D. Sun, and Q. Jin, "Imbalanced sample fault diagnosis of rotating machinery using conditional variational auto-encoder generative adversarial network," *Appl. Soft Comput.*, vol. 92, Jul. 2020, Art. no. 106333.

[22] Z. Wang, J. Wang, and Y. Wang, "An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition," *Neurocomputing*, vol. 310, pp. 213–222, Oct. 2018.

[23] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," *Artif. Neural Netw. Mach. Learn.*, vol. 11141, pp. 270–279, Oct. 2018.

[24] Z. Chen, G. He, J. Li, Y. Liao, K. Gryllias, and W. Li, "Domain adversarial transfer network for cross-domain fault diagnosis of rotary machinery," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 8702–8712, Nov. 2020.

[25] F. Zhuang *et al.*, "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 1–34, 2020.

[26] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446–2455, Apr. 2019.

[27] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7316–7325, Sep. 2019.

[28] X. Wang, H. He, and L. Li, "A hierarchical deep domain adaptation approach for fault diagnosis of power plant thermal system," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5139–5148, Sep. 2019.

[29] X. Li, W. Zhang, Q. Ding, and J.-Q. Sun, "Multi-layer domain adaptation method for rolling bearing fault diagnosis," *Signal Process.*, vol. 157, pp. 180–197, Apr. 2019.

[30] X. Li, W. Zhang, and Q. Ding, "A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning," *Neurocomputing*, vol. 310, pp. 77–95, Oct. 2018.

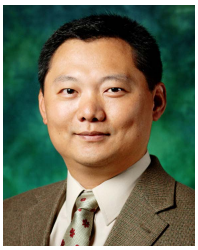
[31] B. Yang, Y. Lei, F. Jia, and S. Xing, "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," *Mech. Syst. Signal Process.*, vol. 122, pp. 692–706, May 2019.
[32] L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 594–611, Apr. 2006.
[33] D. Das and C. S. G. Lee, "A two-stage approach to few-shot learning for image recognition," *IEEE/ACM Trans. Image Process.*, vol. 29, pp. 3336–3350, 2020.
[34] Y. Zhu, W. Min, and S. Jiang, "Attribute-guided feature learning for few-shot image recognition," *IEEE Trans. Multimedia*, vol. 23, pp. 1200–1209, 2020.
[35] G. Koch *et al.*, "Siamese neural networks for one-shot image recognition," in *Proc. ICML Deep Learn. Workshop*, vol. 2, 2015, pp. 1–8.
[36] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 3630–3638.
[37] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, Dec. 2017, pp. 4077–4087.
[38] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, "Meta-learning with memory-augmented neural networks," in *Proc. 33rd Int. Conf. Mach. Learn.*, vol. 4, 2016, pp. 1842–1850.
[39] A. Srinivasan, A. Bharadwaj, M. Sathyan, and S. Natarajan, "Optimization of image embeddings for few shot learning," in *Proc. 10th Int. Conf. Pattern Recognit. Appl. Methods*, 2021, pp. 236–242.
[40] H. Yao *et al.*, "Graph few-shot learning via knowledge transfer," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, 2020, pp. 6656–6663.
[41] D. Zhang, E. Stewart, M. Entezami, C. Roberts, and D. Yu, "Intelligent acoustic-based fault diagnosis of roller bearings using a deep graph convolutional network," *Measurement*, vol. 156, May 2020, Art. no. 107585.
[42] T. Hu, T. Tang, R. Lin, M. Chen, S. Han, and J. Wu, "A simple data augmentation algorithm and a self-adaptive convolutional architecture for few-shot fault diagnosis under different working conditions," *Measurement*, vol. 156, May 2020, Art. no. 107539.
[43] Z. Ren *et al.*, "A novel model with the ability of few-shot learning and quick updating for intelligent fault diagnosis," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106608.
[44] J. Wu, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Few-shot transfer learning for intelligent fault diagnosis of machine," *Measurement*, vol. 166, Dec. 2020, Art. no. 108202.
[45] R. Das and S. Chaudhuri, "On the separability of classes with the cross-entropy loss function," 2019, *arXiv:1909.06930*.
[46] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, May 2015, pp. 1–13.
[47] P. Cao, S. Zhang, and J. Tang, "Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning," *IEEE Access*, vol. 6, pp. 26241–26253, 2018.
[48] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu, "Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery," *IEEE Access*, vol. 5, pp. 15066–15079, 2017.
[49] T. Han, L. Zhang, Z. Yin, and A. C. C. Tan, "Rolling bearing fault diagnosis with combined convolutional neural networks and support vector machine," *Measurement*, vol. 177, Jun. 2021, Art. no. 109022.
[50] A. Zhang, S. Li, Y. Cui, W. Yang, R. Dong, and J. Hu, "Limited data rolling bearing fault diagnosis with few-shot learning," *IEEE Access*, vol. 7, pp. 110895–110904, 2019.
[51] A. Odena, "Semi-supervised learning with generative adversarial networks," 2016, *arXiv:1606.01583*.
[52] D. B. Verstraete, E. L. Droguett, V. Meruane, M. Modarres, and A. Ferrada, "Deep semi-supervised generative adversarial fault diagnostics of rolling element bearings," *Structural Health Monitor.*, vol. 19, no. 2, pp. 390–411, Dec. 2020.
[53] P. Ma, H. Zhang, W. Fan, C. Wang, G. Wen, and X. Zhang, "A novel bearing fault diagnosis method based on 2D image representation and transfer learning-convolutional neural network," *Meas. Sci. Technol.*, vol. 30, no. 5, May 2019, Art. no. 055402.

Juan Xu (Member, IEEE) was born in 1982. She received the Ph.D. degree from the Hefei University of Technology, Hefei, China, in 2012.

Now, she is working with the Hefei University of Technology. She is also working with the Institute of Distributed Intelligence and Internet of Things. Her research interests include the Internet of Things and intelligent computing, deep learning, and fault diagnosis.

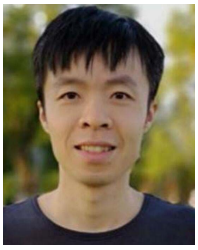
Yongfang Shi was born in Huainan, China, in 1996. He received the M.Sc. degree in software engineering from the Hefei University of Technology, Hefei, China, in 2021.

His current research interests include deep learning and transfer learning methods for intelligent fault diagnostics and prognostics.



Xiaohui Yuan (Senior Member, IEEE) received the B.S. degree in electrical engineering from the Hefei University of Technology, China, in 1996, and the Ph.D. degree in computer science from Tulane University in 2004.

He is currently an Associate Professor with the University of North Texas. His research findings have been published in more than 180 peer-reviewed papers. His research interests include computer vision, artificial intelligence, data mining, and machine learning. He was a recipient of the Ralph E. Powe Junior Faculty Enhancement Award in 2008. He is also the Editor-in-Chief of the *International Journal of Smart Sensor Technologies and Applications*, serves on the editorial board for several international journals, and as chairs in several international conferences.



Siliang Lu (Senior Member, IEEE) received the B.S. and Ph.D. degrees in mechanical engineering from the University of Science and Technology of China, Hefei, China, in 2010 and 2015, respectively.

He is currently an Associate Professor with the College of Electrical Engineering and Automation, Anhui University, Hefei, and also a Visiting Scholar with the Traction Power State Key Laboratory, South west Jiaotong University, Chengdu, and the State Key Laboratory of Mechanical Transmissions, Chongqing University, Chongqing, China. His current research interests include machinery condition-based monitoring and fault diagnosis, signal processing, the IoT and edge computing, and robotics.