Mechanical Systems and Signal Processing 155 (2021) 107510

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Metric-based meta-learning model for few-shot fault diagnosis under multiple limited data conditions

Duo Wang [a,1], Ming Zhang [b,d,1], Yuchun Xu [d], Weining Lu [c], Jun Yang [a,*], Tao Zhang [a,c,*]

[a] Department of Automation, Tsinghua University, Beijing 100084, China
[b] Department of information science and technology, Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China
[c] Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China
[d] School of Engineering and Applied Science, Aston University, Birmingham B4 7ET, UK

ARTICLE INFO

Article history:
Received 25 April 2020
Received in revised form 15 November 2020
Accepted 30 November 2020
Available online 30 January 2021

Keywords:
Metric-based meta-learning
Few-shot learning
Feature space
Fault diagnosis
Limited data conditions

ABSTRACT

The real-world large industry has gradually become a data-rich environment with the development of information and sensor technology, making the technology of data-driven fault diagnosis acquire a thriving development and application. The success of these advanced methods depends on the assumption that enough labeled samples for each fault type are available. However, in some practical situations, it is extremely difficult to collect enough data, e.g., when the sudden catastrophic failure happens, only a few samples can be acquired before the system shuts down. This phenomenon leads to the few-shot fault diagnosis aiming at distinguishing the failure attribution accurately under very limited data conditions. In this paper, we propose a new approach, called Feature Space Metric-based Meta-learning Model (FSM3), to overcome the challenge of the few-shot fault diagnosis under multiple limited data conditions. Our method is a mixture of general supervised learning and episodic metric meta-learning, which will exploit both the attribute information from individual samples and the similarity information from sample groups. The experiment results demonstrate that our method outperforms a series of baseline methods on the 1-shot and 5-shot learning tasks of bearing and gearbox fault diagnosis across various limited data conditions. The time complexity and implementation difficulty have been analyzed to show that our method has relatively high feasibility. The feature embedding is visualized by t-SNE to investigate the effectiveness of our proposed model.

1. Introduction

Fault diagnosis has become an indispensable technology in the large industrial complex systems, due to the increasing development of high-speed and complexity of machinery. These industrial systems are gradually accumulating massive data, which causes unprecedented research and development of data-driven fault diagnosis methods in recent years [1–4]. However, the deep models perform well only when enough labeled data are available for training. Otherwise, the performance of the data-driven deep models will be significantly decreased. With the in-depth research, the few-shot learning problem has been revealed, which aims at training the deep model with very limited data. Specifically, industrial equipment generally

* Corresponding authors at: Department of Automation, Tsinghua University, Beijing 100084, China (T. Zhang).
E-mail addresses: d-wang15@mails.tsinghua.edu.cn (D. Wang), m.zhang21@aston.ac.uk (M. Zhang), y.xu16@aston.ac.uk (Y. Xu), luwn@tsinghua.edu.cn (W. Lu), yangjun603@tsinghua.edu.cn (J. Yang), taozhang@tsinghua.edu.cn (T. Zhang).
[1] Duo Wang and Ming Zhang are equally contributed to this work.

https://doi.org/10.1016/j.ymssp.2020.107510
0888-3270/

operates under normal status. When certain sudden catastrophic failures come, the system will be immediately shut down for maintenance. Hence, the data coming from these failures should be scarce, and in contrast, the normal data is abundant. Based on these very few fault samples, the typical data-driven supervised learning strategy will train an overfitting model that cannot generalize very well. An obvious way to solve such a problem is to recollect data for all tasks or scenarios, which will incur high costs or even be unfeasible.

We try to tackle the fault diagnosis problem with limited data from the few-shot learning prospective. Few-shot learning is currently a very popular topic in computer vision area, especially in image classification [5–17]. Despite the different design of methods, one common characteristic among these works is that they exploit a large and fully-annotated auxiliary set from some disjoint source domain, where a series of few-shot learning tasks are randomly sampled to simulate the few-shot learning situation and extract general knowledge as additional information to facilitate few-shot learning tasks in the target domain, which forms the idea of meta-learning. The source and target domains in image classification field are constructed by randomly splitting a large dataset by categories. In the fault diagnosis field, the data are collected with clear discrimination by their working conditions and failure attributes. Intuitively, machinery does not often operate at high speed and heavy load unless there is an emergency production requirement, and normally they are not allowed to work with worrying fault. In such cases, only a few valid samples will likely be collected. Meanwhile, there are amounts of data samples coming from other situations, which could provide transferable knowledge for supporting the limited data tasks. Therefore, the application of few-shot learning with meta-learning to fault diagnosis problem is reasonable and promising.

Based on these analyses, we propose a novel few-shot learning method named Feature Space Metric-based Meta-learning Model (FSM3) for fault diagnosis under multiple limited data conditions. Our method is based on two popular and effective metric-based meta-learning models for few-shot learning, i.e., Matching Network (MN)[5] and Prototypical Network (PN)[7]. However, we argue that mere metric-based training only teaches the model to focus on the relative similarity information from sample groups, thus the attribute information of each specific category is ignored, which means that the provided labeled source data are not fully exploited. To tackle this problem, we design a hybrid method that combines the merit of general supervised learning and metric meta-learning. Specifically, the first several layers of the model are trained to recognize the fault types of source data in a global supervised manner. Then these layers are fixed as Feature Extractor to transform raw data into basic feature space. Finally, the rest of the model is trained by metric meta-learning with the extracted features. In this way, our model can exploit not only the relative information between data pairs but also supervision information from individual samples. To the best of our knowledge, our proposed method is the first attempt to address the few-shot learning problem in fault diagnosis by utilizing the metric meta-learning based on deep neural networks.

The contributions of this paper are:

1. A novel FSM3 model has been proposed for the few-shot fault diagnosis issue under various limited data conditions. The core of our method is the creative combination of the relative similarity information from sample groups and the supervision information in each specific category from the annotated source data. A hybrid training strategy with global supervised training and episodic training in the learned feature space is designed to support this combination.
2. To tackle the few-shot fault diagnosis problem under limited data conditions using metric-based meta-learning, the clear detail of explanation about its interpretability and feasibility has been discussed and analyzed.
3. The effectiveness of the FSM3 model has been verified with experiments on the bearing dataset and gearbox dataset under different fault types, speed, and load conditions. The results illustrate that our method outperforms other state-of-the-art methods and presents great robustness.

The rest of the paper is organized as follows. In Section 2 we give some background knowledge about few-shot learning and deep learning models for fault diagnosis. In Section 3 we introduce the technical details of the proposed FSM3. In Section 4 we introduce experiments setup and present results and analysis. In Section 5 we draw conclusion from this paper.

2. Background

2.1. Few-shot learning

A supervised classification task $\mathcal{T}$ usually consists of a training set (support set, denoted as $\mathcal{S}$) with labeled data to train the model and a testing set (query set, denoted as $\mathcal{Q}$) with unlabeled data from the same domain to evaluate the performance of training, see Fig. 1(a). When the amount of data in the training set is small, the task is termed as a few-shot learning task. Recently proposed methods for few-shot learning mostly exploit an auxiliary set from some source domain to extract knowledge to help the model training with the given few-shot support set in target domain, shown in Fig. 1(b). Note that the auxiliary set contains a large amount of labeled data and its label space is disjointed to that of target domain. One way to exploit the source domain data is to randomly sample a series of few-shot learning tasks. Transferable knowledge is extracted from the interaction procedure between these tasks and the classification model to facilitate the tasks of the target domain, which forms the episodic training mechanism [5], see Fig. 1(c). Here, each few-shot learning task is considered as an episode. The whole procedure can also be viewed as meta-learning, as the learning is performed at the task level other than the data level. In this paper, different data domains can be considered as different working conditions or fault categories.

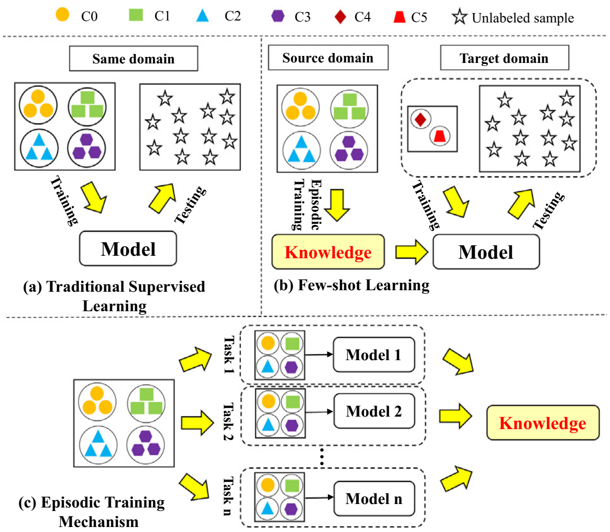

Fig. 1. Illustration of different learning strategies. (a) Traditional supervised learning. (b) General procedure of few-shot learning. (c) Episodic training mechanism for few-shot learning.

According to different forms of knowledge extracted from auxiliary tasks, there are two main branches of the recent few-shot learning area, i.e., metric-based meta-learning and optimization-based meta-learning. Metric-based meta-learning models[5–12] try to learn a unified, category-independent feature space that the intra-class distance of samples is smaller than the inter-class distance. The query samples are classified by their distance to each support sample in the learned space. Optimization-based meta-learning models[13–17] exploit an additional trainable model (meta model) to perform the parameters update of the classification model and the meta model is trained to generate suitable classification parameters by the limited support set that works well on the query set. In this paper, we follow the idea of metric-based meta-learning and propose a novel FSM3 for few-shot fault diagnosis problem.

2.2. Deep learning models for fault diagnosis

Fault diagnosis with deep learning is the typical data-driven method, which provides an end-to-end diagnosis model directly establishing the relationship between increasing monitored data and fault categories [3,4]. To solve the cross-domain problem in the deep learning fault diagnosis, transfer learning theories and methods have been well researched in recent years, attempting to utilize the knowledge from the different but related diagnosis tasks for making its wide application in the actual industrial situation [1]. The fundamental idea is to learn the shared feature on the high-dimensional data space by minimizing the distribution discrepancy between the source and target domain, which is divided by Maximum Mean Discrepancy (MMD) based method [18–23] and Generative Adversarial Networks (GAN) based method [24–27]. Although these methods significantly improve the adaptability of deep learning based fault diagnosis, they still must meet the hypothesis that enough data is available. However, the sudden catastrophic failure data samples are much less than normal condition data samples or other slight fault samples. There may be only one or a few samples in a fault dataset. This kind of problem is the few-shot fault diagnosis problem, one of the critical challenges hindering the deep models' wide application in the actual industrial situation. This problem began to attract the attention of the research in the fault diagnosis field [28–30]. Nevertheless, these methods have not utilized the knowledge of the different but related domains as an auxiliary for supporting the few-shot fault diagnosis task. In this paper, for the first time, a novel Metric-based Meta-learning model is proposed for the Few-shot fault diagnosis problem, called FSM3, which can rely on learning the transferable knowledge of the source domain to overcome the few-shot learning problem in the target domain.

2.3. Definition of few-shot fault diagnosis problem

We follow [13] to define the few-shot fault diagnosis problem. Let $\mathcal{T}^T$ denote a C^T-way, K-shot, M-test few-shot learning task of fault diagnosis from target domain, which consists of a labeled few-shot support set $\mathcal{S}^T$ and unlabeled query set $\mathcal{Q}^T$. $\mathcal{S}^T$ contains K samples per class and $\mathcal{Q}^T$ contains M samples per class. Samples from the support set and query set are $(x^T_{Sa}, y^T_{Sa})_{a=1}^{N^T_S}$ and $(x^T_{Qa})_{a=1}^{N^T_Q}$ respectively, where N^T_S equals to $C^T \times K$ and N^T_Q equals to $M \times K$. K is a small value, and M is not limited. For the fault diagnosis problem, x here denotes the vibration signal wave of length L from some mechanical device, and y denotes its fault type. For the auxiliary set from fully annotated source domain, samples are defined as $X^S = (x^S_a, y^S_a)_{a=1}^{N^S}$. Here, N^S is a relatively large number, which means the source domain data is sufficient. We assume that the source domain contains C^S different fault types. For episodic training mechanism, we randomly sample a series of C^T-way, K-shot, M-test few-shot learning tasks, denoted as $\mathcal{T}^S$s, that have a similar data structure with target task $\mathcal{T}^T$. The only difference is that labels of the query set are also available. The goal of episodic meta-learning is to train a metric model with the source tasks $\mathcal{T}^S$s so that it can identify the fault types of target query set $\mathcal{Q}^T$ well based on the limited annotated data in support set $\mathcal{S}^T$. In this paper, we set K to 1 or 5 following the standard protocol of few-shot image classification problem [13]. 5 denotes a normal few-shot learning situation, and 1 denotes the extreme situation where only one support sample per class is available. In [13], M is much larger than K and set to 15 in during episodic training due to the limitation of GPU memory. During testing, M can be larger since testing saves much more computational resources than training. However, M is also set to 15 for efficient evaluation. Since in the fault diagnosis scenario, the input samples are 1-d signals, which take up less memory than 2-d images, we increase the M to 25. We clarify that the setting of K and M is just for the unified and fair comparison for academic research. They can be set flexibly based on different working conditions or practical demands.

3. Method

In this paper, we propose Feature Space Metric-based Meta-learning Model (FSM3) for few-shot fault diagnosis under multiple limited data conditions. Our method is extended from the popular and effective few-shot learning models, Matching Network (MN) and Prototypical Network (PN), in several aspects.

3.1. Architecture of feature space metric-based meta-learning model

Our proposed FSM3 consists of three modules, including a Feature Extractor (FE), to which the Global Classifier (GC) and Metric Embedding (ME) module are connected for different training steps, as shown in Fig. 2. For the fault diagnosis tasks, the input samples are mechanical vibration waves, which are typical one-dimensional signals. Accordingly, we use one-dimensional convolution layers in our model. Following[29,31], the Feature Extractor contains five convolution layers, each of which is followed by a ReLU function. We use max-pooling in the first four convolution layers to down-sample features. We also set the kernel size of the first layer to be large. The Global Classifier consists of a flatten layer and a fully connected layer with output size equal to the number of categories in the source data domain. The Metric Embedding module contains two convolution layers, followed by a flatten layer and a fully-connected layer with the output size of 100, which converts fault data into 100-dimension features for metric learning. The flatten layers are omitted in Fig. 2 for simplicity. Architecture details are shown in Table 1.

3.2. Learning procedure

3.2.1. Global supervised training

In our FSM3, we first train the Feature Extractor(FE) in a global supervised way with the fully labeled dataset X^S from the source domain. Denote the FE as function $f_{FE}(\cdot)$ with parameters θ_{FE}, which maps the input fault data $x_a^S \in \mathbb{R}^L$ to convolution

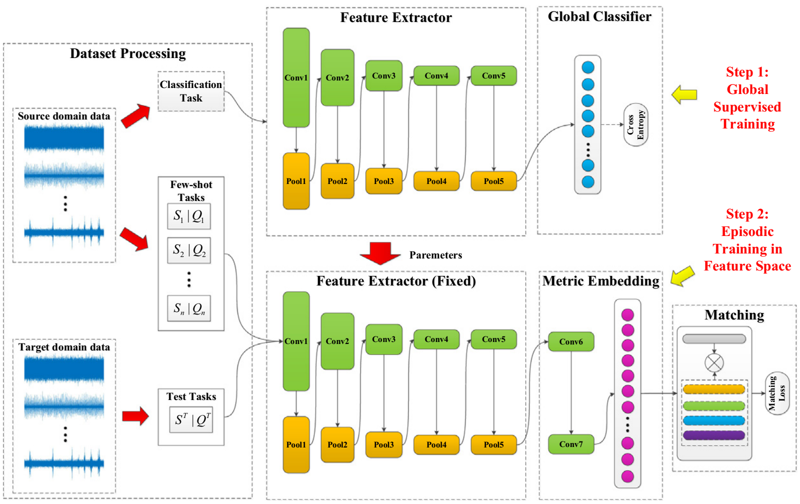

Fig. 2. Training and evaluating procedure of the proposed FSM3. We first train a classification model with source domain data following traditional supervised learning (upper branch). Then we fix the Feature Extractor and train the Metric Embedding module in an episodic manner with a series of few-shot tasks sampled from the source domain. Finally, the Feature Extractor and Metric Embedding are used for target test tasks (lower branch).

feature $b_a \in \mathbb{R}^{c \times L_f}$, where c is the number of channels and L_f is the length of feature. Then the feature is input to the Global Classifier (GC, with parameters θ_{GC}) followed by softmax activation function to get the possibility vector $p_a \in \mathbb{R}^{\mathbb{C}^S}$ indicating how likely the input data belongs to each fault type, see the upper branch of Fig. 2. The objective function L_B of the global supervised training is the cross-entropy loss defined as follows:

$$L_B(X^S; \theta_{FE}, \theta_{GC}) = -\frac{1}{N^S} \sum_{a=1}^{N^S} \sum_{i=1}^{C^S} \mathbb{I}[y_a^S = i] log \frac{e^{p_a(i)}}{\sum_j e^{p_a(j)}} \tag{1}$$

where $\mathbb{I}[\cdot]$ is the indicator function, $p_a(i)$ is the i_{th} element of p_a, y_a^S is the label of fault types of input data, C^S and N^S are the number of categories and samples in source domain respectively. After training, we remove the Global Classifier and fix the Feature Extractor for later use.

3.2.2. Episodic training in feature space

We then train the Metric Embedding module denoted as $f_M(\cdot)$ with parameters θ_M in episodic training manner. For this purpose, we first randomly sample a series of few-shot fault diagnosis tasks from source domain, denoted as $\mathcal{T}^S$, that have similar data structure with the target C^T-way, K-shot, M-test task $\mathcal{T}^T$. Let $\mathcal{S}^S = (x_{\mathcal{S}a}^S, y_{\mathcal{S}a}^S)_{a=1}^{N_{\mathcal{S}}^S}$ and $\mathcal{Q}^S = (x_{\mathcal{Q}a}^S, y_{\mathcal{Q}a}^S)_{a=1}^{N_{\mathcal{Q}}^S}$ be the support set and query set of $\mathcal{T}^S$ respectively, where $N_{\mathcal{S}}^S$ equals to $C^T \times K$ and $N_{\mathcal{Q}}^S$ equals to M. Then we extract basic features (denote as b) from all the raw fault data with the pre-trained Feature Extractor in the last step. The extracted basic features are further processed by the Metric Embedding module to the metric features. After that, we perform classification of the query samples by matching their metric features to the support ones, see the lower branch of Fig. 2. Specifically, for the nth query

Table 1
Details of Network Architecture of FSM3.

Components	Layer type	Kernel	Stride	Channels	Padding
Feature Extractor	Convolution 1	64×1	16×1	16	No
	ReLU 1				
	Max Pooling 1	2×1	2×1	16	No
	Convolution 2	3×1	1×1	32	Yes
	ReLU 2				
	Max Pooling 2	2×1	2×1	32	No
	Convolution 3	3×1	1×1	64	Yes
	ReLU 3				
	Max Pooling 3	2×1	2×1	64	No
	Convolution 4	3×1	1×1	64	Yes
	ReLU 4				
	Max Pooling 4	2×1	2×1	64	No
	Convolution 5	3×1	1×1	64	Yes
	ReLU 5				
Global Classifier	Flatten				
	Fully Connected 1			C_S	
Metric Embedding	Convolution 6	3×1	1×1	64	Yes
	ReLU 6				
	Convolution 7	3×1	1×1	64	Yes
	ReLU 7				
	Flatten				
	Fully Connected 1			100	

sample in one task $\mathcal{T}^S$ from source domain, the predicted label is calculated by the weighted sum of support labels, which is (note that we omit the superscripts S for simplicity):

$$\hat{y}_{\mathcal{Q}n} = \sum_{a=1}^{N_{\mathcal{S}}} w[f_M(b_{\mathcal{Q}n}), f_M(b_{\mathcal{S}a})] \cdot y_{\mathcal{S}a} \tag{2}$$

where the weight w is calculated by softmax normalization of the distance between the query metric feature and every support metric feature:

$$w[f_M(b_{\mathcal{Q}n}), f_M(b_{\mathcal{S}a})] = \frac{exp(-\tau * d[f_M(b_{\mathcal{Q}n}), f_M(b_{\mathcal{S}a})])}{\sum_{j=1}^{N_{\mathcal{S}}} exp(-\tau * d[f_M(b_{\mathcal{Q}n}), f_M(b_{\mathcal{S}j})])} \tag{3}$$

Here τ is a scale factor for fast convergence of training. We choose $d[\cdot]$ to be the cosine distance following recent representative few-shot learning works [5,32,33], which is:

$$d[x_i, x_j] = \frac{x_i \cdot x_j^T}{|x_i||x_j|} \tag{4}$$

The training objective of the Metric Embedding module is the cross-entropy loss over all sampled tasks $\mathcal{T}^S$ from source domain as follows:

$$L_M(\theta_M) = \sum_{\mathcal{T}^S} L_M(\mathcal{T}^S; \theta_M) = \sum_{\mathcal{T}^S} [-\frac{1}{N_Q^S} \sum_{n=1}^{N_Q^S} \sum_{i=1}^{C^S} \mathbb{I}[y_{Qn} = i] \cdot log(\hat{y}_{Qn}(i))] \tag{5}$$

Detailed learning procedure is shown in Alg. 1.

The matching procedure described above follows the idea of Matching Network (MN). One alternative is matching between category prototype as Prototypical Network (PN), where the fault type of query sample is predicted by:

$$\hat{y}_{Qn} = \sum_{a=1}^{C^S} w[f_M(b_{Qn}), P_a^S] \cdot y_a \tag{6}$$

where P_a^S is the mean of metric features of the ath category:

$$P_a^S = \frac{1}{K} \sum_{j \in a} f_M(b_{Sj}) \tag{7}$$

The calculation of weights and distance and training objective are similar to Eqs. (3)–(5). In this paper, both of the two versions are implemented and compared.

Our model belongs to the metric-based meta-learning family, but training is performed in feature space pre-trained by global supervision other than raw data space, so we term it as Feature Space Metric-based Meta-learning Model (FSM3).

Algorithm 1 Feature Space Metric-based Meta-learning Model (FSM3) Learning Procedure

---

Require source data X^S, mini-batch size for global training m, global training steps n_B, learning rate α_B, ME training step n_M, ME learning rate α_M.

1: Initialize the parameters θ_{FE} and θ_{GC}.

*********Global Supervised Training*********

2: **for** $t = 1, \ldots, n_B$ **do**

3: Sample mini-batch $X_m^S = (x_a^S, y_a^S)_{a=1}^m$ from X^S

4: $\theta_{FE}, \theta_{GC} \leftarrow \theta_{FE}, \theta_{GC} - \alpha_B \nabla_\theta L_B(X_m^S)$

5: **end for**

6: Fix the parameters θ_{FE} and initialize the parameters θ_M.

*******Episodic Training in Feature Space*******

7: **for** $i = 1, \ldots, n_M$ **do**

8: Sample a few-shot task $\mathcal{T}^S$ from source data X^S.

9: Transform raw data in $\mathcal{T}^S$ into feature space with trained θ_{FE}

10: $\theta_M \leftarrow \theta_M - \alpha_M \nabla_{\theta_M} L_M(\mathcal{T}^S)$

11: **end for**

3.2.3. Evaluation of target tasks

Once the training is finished, the Feature Extractor and the Metric Embedding module are used for target fault diagnosis tasks. All the samples from target tasks are transferred into basic feature space, and a similar matching operation is conducted to predict the fault types of query data based on the provided limited support data, as shown in the lower branch of Fig. 2 with target domain data.

4. Experiments

4.1. Experiment setup

4.1.1. Few-shot setup for fault diagnosis

In this work, we assume the source domain contains sufficient labeled data, which is used to support the few-shot fault diagnosis tasks with very limited training data in the target domain. We consider two kinds of few-shot tasks, 1-shot and 5-shot tasks, which means the training sets contain only one or five samples. All experiments are implemented in the following scenarios:

(1) The source and target domain are drawn from the different working conditions, which in this paper are load and speed, for 1-shot learning fault diagnosis.
(2) The source and target domain are drawn from the different categories under the same working condition for 1-shot learning fault diagnosis.
(3) 5-shot learning fault diagnosis for the tasks which are challenging to address in the 1-shot learning situation.

4.1.2. Compared methods

To better evaluate our proposed method, we compare our FSM3 with several baseline few-shot learning methods for all types of few-shot tasks described above, which are detailed as follows:

(1) Finetune Last;
(2) Finetune Whole;
(3) Feature Knn;
(4) Feature Knn Proto;
(5) Data Space Matching Network (DSMN);
(6) Data Space Matching Network with Pre-train (DSMN-Pre);
(7) Feature Space Matching Network (FSM3-MN, ours);
(8) Data Space Prototypical Network (DSPN);
(9) Data Space Prototypical Network with Pre-train (DSPN-Pre);
(10) Feature Space Prototypical Network (FSM3-PN, ours).

(1) to (4) are based on pre-training FE + ME with source domain data under supervised learning. (1) and (2) then attach a new classifier after FE + ME and use the few-shot support data from target domain to finetune the last layer (the classifier) or the whole model respectively. (3) and (4) classify target data by matching the extracted features from FE + ME backbone to those of support samples or support class prototypes. (5) and (8) are the original Matching Network and Prototypical Network model with FE + ME as backbone. The whole model is completely trained in an episodic way in raw data space. (6) and (9) are similar to (5) and (8), but the backbone is pre-trained with source data. (7) and (10) are our models of MN and PN version, where FE is trained with supervised learning and then fixed, ME is trained in an episodic way.

4.1.3. Implementation Details

We use Adam optimization[34] to train all the models. For the supervised pre-training with source domain data, we set the learning rate as 0.001, batch size as 16, maximum number of training iterations (epochs) as 80. We stop the pre-training if training loss stops decreasing for 15 epochs and load the model with the lowest loss for later use. For finetuning-based models (Finetune Last and Finetune Whole), the number of finetuning steps is set to 100. For KNN based methods, cosine distance is exploited. For the episodic metric training, we set learning rate as 0.0001, number of query samples (M) in few-shot learning tasks as 25, τ as 100 for fast convergence, maximum number of training iterations (epochs) as 100. For each epoch, we randomly sample 100 few-shot learning tasks from source domain to perform metric learning. For evaluation of all methods, we sample 600 tasks from target domain and the mean accuracy of classification is recorded as final results. All experiments are implemented on the computer with one Nvidia GeForce GTX 1080 Ti GPU, one Intel Core i7-6850 K CPU of 3.60 GHz and 64 GB memory. A detailed list of the experiment settings is provided in Table 2.

4.2. Case Study 1: Bearing dataset

4.2.1. Data preparation and diagnosis scenarios

The bearing center of Case Western Reserve University (CWRU) [35] provides this bearing dataset, which was collected by the accelerometer fixed on the motor housing at the 12 o'clock position. The testbed including a motor, accelerometer, torque transducer, and dynamometer, shown in Fig. 3. There are ten categories in the bearing dataset which consist of Normal, 3 various fault sizes (0.007, 0.014 and 0.021 in.) for each of 3 fault locations (inner race, outer race, and ball), respectively. All these categories are gathered on 4 different loads (0, 1, 2, 3 hp) and the sampling frequency is set to 12 kHz. Each category has 500 samples and each sample is a vibration signal with 2048 points. The data samples of bearing dataset in both normal and fault status under different conditions are shown in Fig. 4.

Table 3 shows the few-shot diagnosis scenarios of the bearing dataset, including Different Loads and Different Categories. The Different Loads contains 4 scenarios, each of which has 10 categories in both the source domain and target domain. The Different Categories also contains 4 scenarios including "Ball", "Inner Race", "Outer Race", and "Worst IOB" as the target domain, each of which has 7 categories in the source domain and 3 categories in the target domain under the same load condition. The "Worst IOB" denotes that the target domain contains the worst fault of Inner race, Outer race, and Ball with the defect size of 0.021 in., and the source domain contains other fault types including normal, IOB with 0,07 in. and IOB with 0,014 in.. Moreover, the "Inner Race", "Outer Race" and "Ball" scenarios denote that the target domain contains 3 fault types (3 different sizes) of the particular position respectively and the source domain contains the rest types.

Table 2
Detailed experiment settings.

Description		Value
	Optimizer	Adam
Pre-training	Learning rate	0.001
	Batch size	16
	Maximum epochs	80
	Early stop duration epochs	15
Finetune	Finetune steps	100
Episodic training	Distance metric	Cosine
	Learning rate	0.0001
	Scale factor τ	100
	Maximum epochs	100
	Tasks per epoch	100
	Support samples per class (K)	1 or 5
	Query samples per class (M)	25
	Evaluation tasks	600

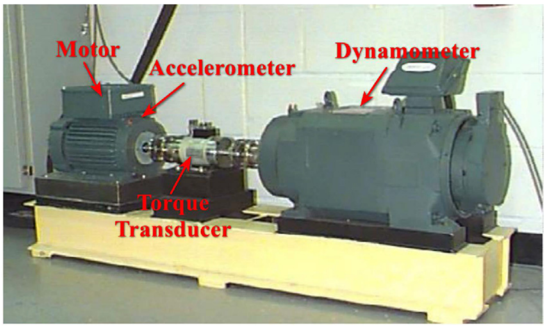

Fig. 3. CWRU testbed.

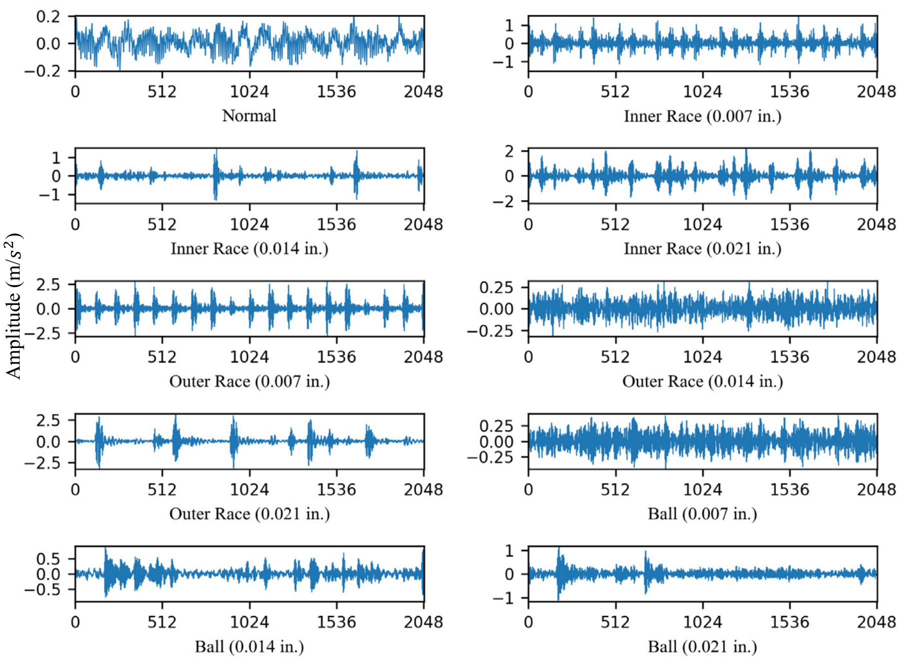

Fig. 4. Data samples of bearing under one normal condition and nine failure conditions.

Table 3
Few-shot fault diagnosis scenarios of bearing dataset.

Different Loads		Different Categories	
Source Domain	Target Domain	Source Domain	Target Domain
Load 0	Load 3	Normal, Inner Race, Outer Race	Ball
Load 1	Load 3	Normal, Outer Race, Ball	Inner Race
Load 2	Load 3	Normal, Inner Race, Ball	Outer Race
Load 0	Load 2	Normal, IOB (0.007), IOB (0.014)	Worst IOB (0.021)

4.2.2. Results and analysis

Based on the few-shot setup, we first analyze the 1-shot learning fault diagnosis problem between different working conditions. The results are shown in Table 4, which present that all the MN methods and Feature Knn perform very well in these tasks, and there are no obvious differences between the MN methods. These results indicate that the similarity between data pairs of different domains caused by changing working conditions is relatively close, thus 1-shot learning tasks, in this case, are not difficult to solve.

We then conduct experiments in the Different Categories scenarios. Table 5 shows the results of the 1-shot task of "Worst IOB", "Ball", and "Inner Race", which illustrates that all the MN-based methods perform great in these three 1-shot tasks on 4 different working loads, with even more than 99% accuracy.

Since the accuracy of "Outer Race" scenario is generally lower than that of other 3 scenarios, we test it on 1-shot and 5-shot setup, and the results are shown in Table 6. These results verify that our FSM3 performs the best compared with other baseline methods and the performance can be improved greatly by increasing the number of support samples. It is obvious that the accuracy of Load 2 and 3 is higher than that of low load conditions. We believe that this phenomenon is related to the difficulty degree of distinguishing fault categories in the source domain. The difficult task of the high load in the source domain will offer a well-trained model, which will significantly help the model used in the target domain. Otherwise, the model is easier to reach the expected performance when the source domain is Load 0 or 1. Such a model cannot provide sufficient support for the few-shot task in the target domain. Limited by the mechanism [7], we can only compare between MN-based and PN-based method at the 5-shot situation, and the results confirm that the performance of these two methods is very close, MN-based method is better at certain conditions.

4.3. Case Study 2: Gearbox dataset

4.3.1. Data preparation and diagnosis scenarios

The second dataset is collected from a gearbox stand under various conditions [36]. Fig. 5 shows the testbed, where an accelerometer is located on the output of the gearbox. There are three categories of gear failure: Normal, Chipped Tooth, and Missing Tooth. Each category is split by different speeds (30, 40, and 50 Hz) and different loads (Low and High), and contains 500 samples of each condition. The sampling frequency is 66.67 kHz, and each sample has 6600 points. Fig. 6 displays the data samples under different conditions in the gearbox dataset, which are time-domain vibration waveform.

The few-shot fault diagnosis scenarios of the gearbox dataset are shown in Table 7, which incorporate different working conditions (loads and speeds) and fault categories. Both the source domain and target domain have 3 categories, including Normal, Chipped Tooth, and Missing Tooth in the different working conditions situation. There are two kinds of tasks, including CT and MT in the few-shot diagnosis scenarios of different categories. CT denotes the data of Normal and Chipped Tooth as the target domain, while the data of Normal and Missing Tooth as the source domain. MT is the task with the data of Normal and Missing Tooth as the target domain, and the data of Normal and Chipped Tooth as the source domain.

4.3.2. Results and analysis

To further verify the performance of our proposed FSM3 method, a more difficult evaluation on the gearbox dataset has been conducted and analyzed in this part. Firstly, our approach has been evaluated on the 1-shot and 5-shot learning tasks under different load and speed conditions. The results are shown in Table 8, which illustrate that our FSM3 has performed the best in most of the tasks. However, due to the improvement of task difficulty degree, the overall effect is inferior to that of the bearing dataset.

Next, according to the above experimental settings, we test different advanced methods including our FSM3 on the few-shot learning tasks with different gear fault types under the same condition. There are two kinds of cases in this mission, including CT and MT. Table 9 displays the results of the 1-shot learning task. Our FSM3 performs the best at the majority of cases compared with other approaches. These results present an interesting phenomenon that the accuracy of CT is all lower than that of MT. We believe the reason is the same as discussed in the bearing section, which is the hard task in the source domain will offer more support for the few-shot fault diagnosis task of the target domain.

Finally, the experiments of 5-shot learning tasks for CT fault under the same conditions have been carried out, and the results are shown in Table 10. Although increasing the number of data samples does not make these tasks perfectly solved, it does notably improve the accuracy of these methods. In a word, our proposed FSM3 still achieves the best results in most

Table 4
Accuracy (%) on 1-shot learning tasks for bearing fault under different conditions

	Load 0→3	Load 1→3	Load 2→3	Load 0→2
Finetune Last	58.04	58.39	57.97	58.82
Finetune Whole	88.21	89.51	90.62	90.15
Feature Knn	98.78	99.51	99.30	98.98
DSMN	99.08	***99.58***	99.88	99.38
DSMN-Pre	99.38	***99.58***	99.91	99.39
FSM3-MN(ours)	***99.42***	99.48	***99.96***	***99.86***

Table 5
Accuracy (%) on 1-shot learning tasks for Bearing fault with different types under same working conditions

	Load 0			Load 1		
	Worst IOB	Ball	Inner Race	Worst IOB	Ball	Inner Race
Finetune Last	88.85	85.83	84.49	89.05	92.78	88.79
Finetune Whole	97.42	97.35	97.37	97.39	97.86	97.88
Feature Knn	98.12	94.53	99.71	98.96	98.91	99.97
DSMN	99.36	99.88	***100.00***	99.77	99.94	99.95
DSMN-Pre	***99.48***	***99.91***	***100.00***	99.87	***99.96***	***100.00***
FSM3-MN(ours)	99.24	99.84	***100.00***	***99.95***	***99.96***	99.88
	Load 2			Load 3		
	Worst IOB	Ball	Inner Race	Worst IOB	Ball	Inner Race
Finetune Last	90.92	87.43	84.31	89.72	88.80	79.30
Finetune Whole	96.05	94.93	96.81	97.20	94.05	95.04
Feature Knn	99.66	99.21	97.83	97.77	93.51	98.29
DSMN	99.62	***99.99***	99.76	99.77	99.04	99.88
DSMN-Pre	99.78	***99.99***	99.82	99.80	99.58	99.40
FSM3-MN(ours)	***100.00***	99.94	***100.00***	***99.98***	***99.94***	***99.99***

Table 6
Accuracy (%) on 1-shot and 5-shot learning tasks for Outer Race fault under same working conditions.

	Load 0		Load 1		Load 2		Load 3	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Finetune Last	56.38	64.98	65.26	67.07	83.14	85.18	76.60	77.73
Finetune Whole	65.08	85.75	75.11	89.15	93.93	95.86	92.52	97.68
Feature Knn	67.72	80.06	79.42	90.72	88.50	98.73	91.78	93.48
DSMN	66.97	87.81	81.87	86.50	90.26	97.29	84.21	95.15
DSMN-Pre	68.24	85.28	85.50	88.07	94.56	98.72	93.86	97.04
FSM3-MN(ours)	***72.94***	***90.74***	***88.07***	91.22	***97.78***	***99.10***	***97.65***	***97.93***
Feature Knn Proto	*	78.47	*	87.90	*	94.13	*	93.55
DSPN	*	73.74	*	84.06	*	93.14	*	93.22
DSPN-Pre	*	75.76	*	87.13	*	93.26	*	93.76
FSM3-PN(ours)	*	83.69	*	***92.73***	*	98.80	*	94.52

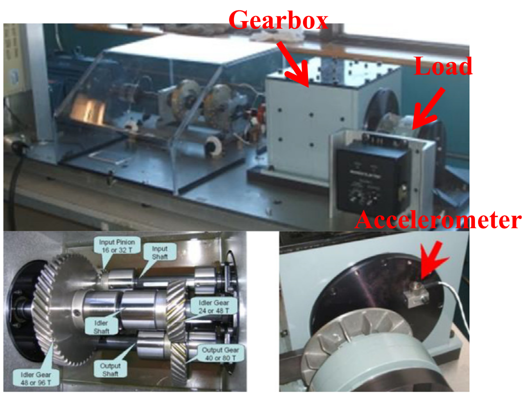

Fig. 5. Gearbox testbed.

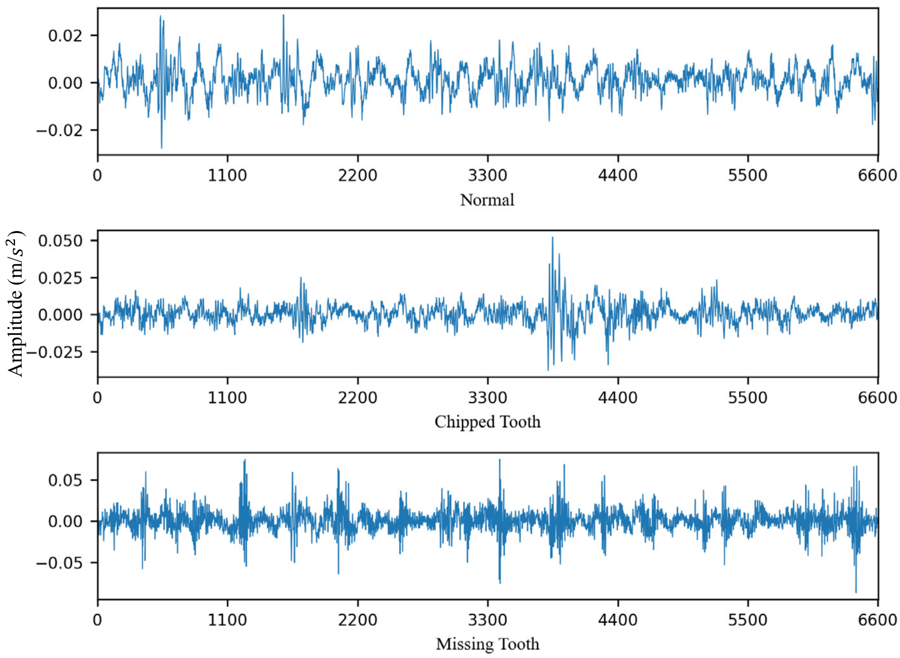

Fig. 6. Data samples of gearbox under one normal condition and two failure conditions.

Table 7
Few-shot fault diagnosis scenarios of gearbox dataset.

Different Loads and Speeds		Different Categories	
Source Domain	Target Domain	Source Domain	Target Domain
30L	30H	30H CT/MT	30H MT/CT
40L	40H	40H CT/MT	40H MT/CT
50L	50H	50H CT/MT	50H MT/CT
40H	50H	*	*

Table 8
Accuracy (%) on 1-shot and 5-shot learning tasks for gearbox fault under different conditions.

	30L →30H		40L →40H		50L →50H		40H →50H	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Finetune Last	66.62	67.46	66.19	68.74	64.45	66.82	73.82	75.54
Finetune Whole	66.50	68.21	67.30	69.64	67.16	68.12	74.77	83.71
Feature Knn	68.71	69.71	69.12	73.77	72.00	73.89	***91.83***	98.45
DSMN	72.02	88.38	77.37	84.80	72.73	75.66	91.53	99.50
DSMN-Pre	74.27	83.28	74.79	79.51	74.41	76.77	89.84	98.74
FSM3-MN(ours)	***76.54***	***88.56***	***78.42***	***86.38***	***75.32***	77.18	91.43	99.08
Feature Knn Proto	*	68.80	*	72.14	*	74.00	*	97.63
DSPN	*	82.81	*	82.94	*	***78.00***	*	***99.65***
DSPN-Pre	*	81.94	*	78.77	*	76.12	*	98.32
FSM3-PN(ours)	*	83.66	*	82.48	*	76.92	*	98.60

of the cases, which verifies the effectiveness of our method again. When comparing the results of 50H CT task with the tasks from other datasets, we find a phenomenon that the accuracy of our proposed Feature Space methods (FSMN/FSPN) is lower than Data Space ones (DSMN/DSPN) by an obvious margin. The possible reason is: the data samples from the 50H condition cover more rotation periods thus contain more diagnosis information. Exploiting the data of MT type as the source to pretrain the feature extractor in traditional supervised way will make it too biased towards the MT task to generalize well to the target CT task. However, DSMN and DSPN are initially trained in the episodic way, which mainly focuses on the similarity between data samples rather than features of the individual sample. As a result, DSMN and DSPN perform better than our proposed methods in this case. Despite this, the metric-based meta-learning models still outperform the traditional finetune-based baselines.

Table 9
Accuracy (%) on 1-shot learning tasks for gearbox fault with different categories under same conditions

	30H		40H		50H	
	CT	MT	CT	MT	CT	MT
Finetune Last	51.89	88.78	51.78	96.00	51.16	81.74
Finetune Whole	52.92	89.13	59.36	97.56	55.81	97.11
Feature Knn	55.54	89.92	61.82	92.40	64.20	96.62
DSMN	61.23	89.13	62.59	94.21	***70.82***	***99.98***
DSMN-Pre	61.48	90.84	62.73	95.31	66.35	99.60
FSM3-MN(ours)	***62.28***	***96.31***	***64.54***	***98.44***	66.71	99.47

Table 10
Accuracy (%) on 5-shot learning tasks for Chipped Tooth (CT) fault under same conditions

	30H	40H	50H
Finetune Last	53.23	51.96	53.59
Finetune Whole	54.52	68.09	68.45
Feature Knn	56.33	68.64	73.44
DSMN	61.92	69.94	***83.76***
DSMN-Pre	62.64	68.47	76.54
FSM3-MN(ours)	62.76	***74.04***	76.48
Feature Knn Proto	57.95	70.73	75.45
DSPN	61.92	69.05	80.56
DSPN-Pre	62.49	68.57	77.88
FSM3-PN(ours)	***63.73***	70.99	75.17

4.4. Time complexity and feasibility analysis

In Table 11, we provide the computational time of all the models on the Bearing dataset, Outer Race task with Load 3 and the Gearbox dataset, CT task with 30 Hz High Load. Both the time of training with the source domain (denote by Train) and evaluating with 600 few-shot diagnosis tasks from the target domain (denote by Eval) are given. For a fair comparison, we remove the early stop mechanism and train all the models for the maximum epochs listed in Table 2. A series of conclusions can be drawn. First, the training time of Finetune Last, Finetune Whole, Feature KNN, and Feature KNN Proto is very close because the training procedure is identical, i.e., supervised training with source domain data. Second, training of Gearbox data is much faster than that of Bearing data because the Gearbox dataset is smaller and contains fewer class types. Third, DSMN-Pre and DSPN-Pre have the longest training time because the whole architecture is trained by two stages. Fourth, our proposed method saves lots of time during training compared with DSMN-Pre and DSPN-Pre although it also requires two-stage training, because our method only trains a part of the model in each stage. Fifth, Finetune Last and Finetune Whole spend much more time evaluating than other methods because they need additional training with support data when dealing with target tasks while other methods don't. This indicates that their feasibility of evaluating is very low. Sixth, the evaluating time of other methods does not vary too much. Based on the running time from Table 11 and the implementation difficulty, we give a feasibility assessment of all the methods in Table 12. The training feasibility of Finetune Last, Finetune Whole, Feature KNN, and Feature KNN Proto are high as the standard supervised training is very easy to implement. The training feasibility of DSMN-Pre and DSPN-Pre is low because the two-stage training of the whole model is complex and costs the most training time. By contrast, the DSMN, DSPN, and our method have medium training feasibility. The evaluating feasibility of all the metric-based methods, including ours, is high because no additional finetuning is required. We can see from Table 12 that our method has relatively high feasibility. Based on the results from the previous subsection, our method provides the best accuracy in most cases. Taking accuracy and feasibility into consideration, we can say that our method presents the best overall performance.

4.5. Visualization analysis

We visualize the feature embedding of source and target domain from our FSM3 with t-SNE for both two datasets, shown in Fig. 7 and Fig. 8. Different colors represent different fault types. For each experiment setting, we select one to visualize. For the bearing dataset, we consider the Load 0→3 task and the Outer Race task under Load 3. For the gearbox dataset, we consider the 40H →50H task and CT fault task with 30 Hz high load. FSM3-MN trained for the 1-shot and 5-shot task and FSM3-PN trained for the 5-shot task are implemented. From these visualization results, we can see that the embeddings from the same category are close to each other, and those from different categories are separated, indicating the effectiveness of our method. There exists some intersection between the features of different types in the CT task, which means this task is a little difficult to tackle. One interesting phenomenon is that the embeddings from the model trained with 5-shot tasks are more scattered than those trained with 1-shot tasks. We believe that this is because, for the 5-shot scenario, the model is trained to

Table 11
Computational time (s) of all the models with different datasets and settings.

	Bearing, Outer Race Load 3				Gearbox, Clipped Tooth 30H			
	1-shot		5-shot		1-shot		5-shot	
	Train	Eval	Train	Eval	Train	Eval	Train	Eval
Finetune Last	112.25	72.58	113.00	73.86	19.74	73.25	19.38	74.52
Finetune Whole	112.34	243.78	111.09	248.42	19.80	253.77	20.22	261.02
Feature Knn	113.21	2.05	114.77	2.26	19.83	2.35	19.93	2.60
DSMN	121.83	2.17	135.19	2.56	100.73	2.25	112.51	2.77
DSMN-Pre	233.66	2.32	245.86	2.66	117.87	2.56	130.47	2.73
FSM3-MN(ours)	149.91	2.33	155.66	2.45	67.95	2.66	69.84	2.66
Feature Knn Proto	*	*	111.74	2.34	*	*	19.76	2.62
DSPN	*	*	137.22	2.55	*	*	112.02	2.75
DSPN-Pre	*	*	245.35	2.41	*	*	128.35	2.78
FSM3-PN(ours)	*	*	154.32	2.37	*	*	68.90	2.66

Table 12
Feasibility Assessment of Methods.

	Train	Eval
Finetune Last	High	Low
Finetune Whole	High	Low
Feature KNN (Proto)	High	High
DSMN/PN	Medium	High
DSMN/PN-Pre	Low	High
FSM3-MN/PN(ours)	Medium	High

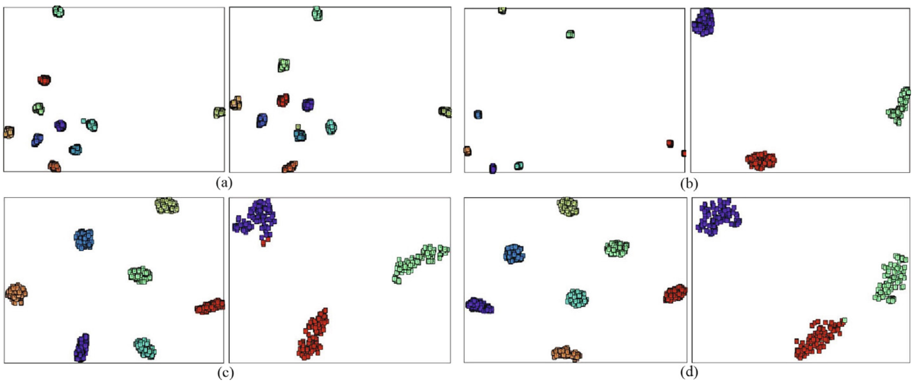

Fig. 7. t-SNE visualization of bearing data feature embedding derived from our FSM3. (a) denotes results of the Load 0→3 task with the same 10 categories. (b), (c) and (d) denote results of the Outer Race fault task under Load 3, which are FSM3-MN 1-shot, FSM3-MN 5-shot and FSM3-PN 5-shot, respectively. For each sub-figure, the left figure denotes the data feature from source domain and the right denotes the data feature from target domain.

match query samples to one of the 5 support samples, and different query samples will be matched to different support samples, thus the embedding distribution will come to a multi-center mode. However, for the 1-shot scenario, all the query samples will be matched to the same support sample, making the distribution of feature embedding more concentrated.

5. Conclusion

In this paper, we introduce the few-shot learning into the data-driven fault diagnosis field and proposed a novel method, FSM3, for the few-shot fault diagnosis with multiple limited data conditions. The performance of our method has been evaluated on bearing and gearbox datasets, where 1-shot and 5-shot tasks are set up in the target domain. There are four conclusions that can be drawn from these experiments: 1) Compared with traditional finetune-based methods, metric-based meta-learning methods achieve higher accuracy on both datasets; 2) More difficult tasks in the source domain can provide more transferable knowledge for the deep model of the target domain, which leads to more effective model; 3) Our proposed FSM3 performs better than a series of baseline methods on the 1-shot and 5-shot learning of bearing and gearbox fault diagnosis under various limited data conditions, while the FSM3-MN is usually better than FSM3-PN; 4) The feasibility of our proposed FSM3 is relatively high.

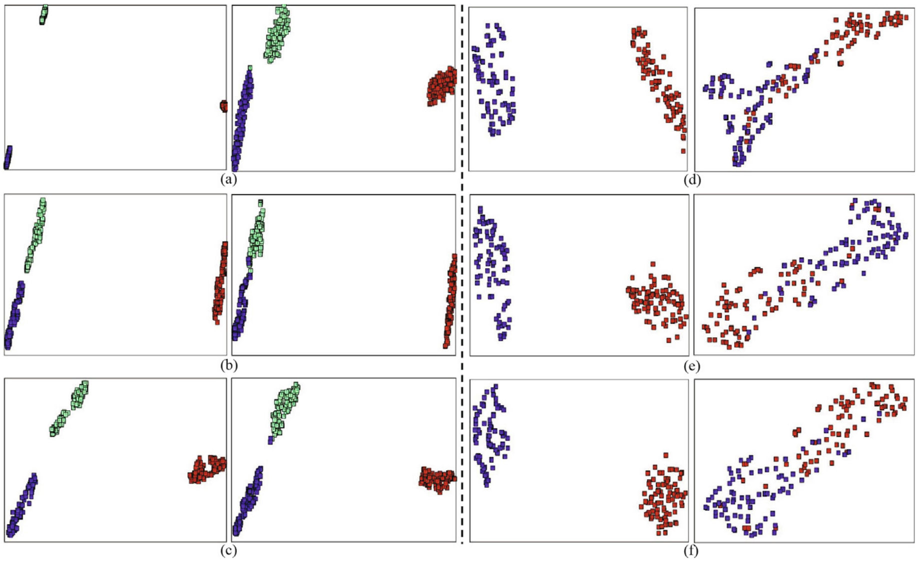

Fig. 8. t-SNE visualization of gearbox data feature embedding derived from our FSM3. (a), (b) and (c) denote results of FSM3-MN 1-shot, FSM3-MN 5-shot and FSM3-PN 5-shot respectively for the 40H→50H task. (d), (e) and (f) denote results of FSM3-MN 1-shot, FSM3-MN 5-shot and FSM3-PN 5-shot respectively for the Chipped Tooth (CT) fault task under condition of 30 Hz high load. For each sub-figure, the left figure denotes the data feature from source domain and the right denotes the data feature from target domain.

CRediT authorship contribution statement

Duo Wang: Conceptualization, Methodology, Software, Writing - original draft. **Ming Zhang:** Conceptualization, Methodology, Writing - original draft. **Yuchun Xu:** Writing - review & editing. **Weining Lu:** Data curation. **Jun Yang:** Project administration, Supervision. **Tao Zhang:** Writing - review & editing, Supervision.

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.

Acknowledgment

The authors acknowledge the financial supported by the National Key Research and Development Program of China (No.2017YFB0602700).

References

[1] 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.
[2] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, Deep learning and its applications to machine health monitoring, Mech. Syst. Signal Process. 115 (2019) 213–237.
[3] L. Wen, X. Li, L. Gao, Y. Zhang, A new convolutional neural network-based data-driven fault diagnosis method, IEEE Trans. Industr. Electron. 65 (7) (2017) 5990–5998.
[4] G. Jiang, H. He, J. Yan, P. Xie, Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox, IEEE Trans. Industr. Electron. 66 (4) (2018) 3196–3207.
[5] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, D. Wierstra, Matching networks for one shot learning, in: Advances in Neural Information Processing Systems, 2016, pp. 3630–3638.
[6] G. Koch, R. Zemel, R. Salakhutdinov, Siamese neural networks for one-shot image recognition, in: ICML Deep Learning Workshop, Vol. 2, 2015.
[7] J. Snell, K. Swersky, R. Zemel, Prototypical networks for few-shot learning, in: Advances in Neural Information Processing Systems, 2017, pp. 4077–4087.
[8] F. Sung, Y. Yang, L. Zhang, T. Xiang, P.H. Torr, T.M. Hospedales, Learning to compare: Relation network for few-shot learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1199–1208.
[9] B. Oreshkin, P. Rodríguez López, A. Lacoste, Tadam: Task dependent adaptive metric for improved few-shot learning, in: Advances in Neural Information Processing Systems, 2018, pp. 719–729.
[10] D. Wang, Y. Cheng, M. Yu, X. Guo, T. Zhang, A hybrid approach with optimization-based and metric-based meta-learner for few-shot learning, Neurocomputing 349 (2019) 202–211.
[11] Y. Lifchitz, Y. Avrithis, S. Picard, A. Bursuc, Dense classification and implanting for few-shot learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 9258–9267.

[12] W. Li, L. Wang, J. Xu, J. Huo, Y. Gao, J. Luo, Revisiting local descriptor based image-to-class measure for few-shot learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 7260–7268.
[13] S. Ravi, H. Larochelle, Optimization as a model for few-shot learning, in: International Conference on Learning Representations, Vol. 1, 2017.
[14] T. Munkhdalai, H. Yu, Meta networks, arXiv e-prints (2017) arXiv:1703.00837 arXiv:1703.00837.
[15] C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in: Proceedings of the 34th International Conference on Machine Learning, Vol. 70, 2017, pp. 1126–1135.
[16] Z. Li, F. Zhou, F. Chen, H. Li, Meta-SGD: Learning to learn quickly for few-shot learning, arXiv e-prints (2017) arXiv:1707.09835 arXiv:1707.09835.
[17] Q. Sun, Y. Liu, T.-S. Chua, B. Schiele, Meta-transfer learning for few-shot learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 403–412.
[18] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, T. Zhang, Deep model based domain adaptation for fault diagnosis, IEEE Trans. Industr. Electron. 64 (3) (2016) 2296–2305.
[19] 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. Industr. Electron. 66 (9) (2018) 7316–7325.
[20] X. Li, W. Zhang, Q. Ding, A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning, Neurocomputing 310 (2018) 77–95.
[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] X. Li, W. Zhang, Q. Ding, J.-Q. Sun, Multi-layer domain adaptation method for rolling bearing fault diagnosis, Signal Processing 157 (2019) 180–197.
[23] B. Yang, Y. Lei, F. Jia, N. Li, Z. Du, A polynomial kernel induced distance metric to improve deep transfer learning for fault diagnosis of machines, IEEE Trans. Ind. Electron.
[24] X. Li, W. Zhang, Q. Ding, Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks, IEEE Trans. Industr. Electron. 66 (7) (2018) 5525–5534.
[25] M. Zhang, D. Wang, W. Lu, J. Yang, Z. Li, B. Liang, A deep transfer model with wasserstein distance guided multi-adversarial networks for bearing fault diagnosis under different working conditions, IEEE Access 7 (2019) 65303–65318.
[26] X. Wang, F. Liu, Triplet loss guided adversarial domain adaptation for bearing fault diagnosis, Sensors 20 (1) (2020) 320.
[27] X. Li, W. Zhang, N.-X. Xu, Q. Ding, Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places, IEEE Trans. Ind. Electron.
[28] Z. Ren, Y. Zhu, K. Yan, K. Chen, W. Kang, Y. Yue, D. Gao, A novel model with the ability of few-shot learning and quick updating for intelligent fault diagnosis, Mech. Syst. Signal Process. 138 (2020) 106608.
[29] A. Zhang, S. Li, Y. Cui, W. Yang, R. Dong, J. Hu, Limited data rolling bearing fault diagnosis with few-shot learning, IEEE Access 7 (2019) 110895–110904.
[30] T. Hu, T. Tang, R. Lin, M. Chen, S. Han, J. Wu, A simple data augmentation algorithm and a self-adaptive convolutional architecture for few-shot fault diagnosis under different working conditions, Measurement 156 (2020) 107539.
[31] W. Zhang, G. Peng, C. Li, Y. Chen, Z. Zhang, A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals, Sensors 17 (2) (2017) 425.
[32] W. Li, L. Wang, J. Xu, J. Huo, Y. Gao, J. Luo, Revisiting local descriptor based image-to-class measure for few-shot learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7260–7268.
[33] R. Hou, H. Chang, B. MA, S. Shan, X. Chen, Cross attention network for few-shot classification, in: Advances in Neural Information Processing Systems 32, 2019, pp. 4003–4014.
[34] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.
[35] K. Loparo, Case western reserve university bearing data center (2012).
[36] M. Zhang, W. Lu, J. Yang, D. Wang, L. Bin, Domain adaptation with multilayer adversarial learning for fault diagnosis of gearbox under multiple operating conditions, in: 2019 Prognostics and System Health Management Conference (PHM-Qingdao), IEEE, 2019, pp. 1–6.