Measurement 166 (2020) 108202

Contents lists available at ScienceDirect

Measurement

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

Few-shot transfer learning for intelligent fault diagnosis of machine

Jingyao Wu, Zhibin Zhao, Chuang Sun *, Ruqiang Yan, Xuefeng Chen

School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

A R T I C L E I N F O

Article history:
Received 6 May 2020
Received in revised form 9 June 2020
Accepted 5 July 2020
Available online 11 July 2020

Keywords:
Few-shot learning
Intelligent diagnosis
Transfer learning
Meta-learning
Rotating machinery

A B S T R A C T

Rotating machinery intelligent diagnosis with large data has been researched comprehensively, while there is still a gap between the existing diagnostic model and the practical application, due to the variability of working conditions and the scarcity of fault samples. To address this problem, few-shot transfer learning method is constructed utilizing meta-learning for few-shot samples diagnosis in variable conditions in this paper. We consider 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 few-shot diagnosis of three datasets. Baseline accuracy under different sample capacity and transfer situations are provided for comprehensive comparison and guidelines. What is more, data dependency, transferability, and task plasticity of various methods in the few-shot scenario are discussed in detail, the data analysis result shows meta-learning holds the advantage for machine fault diagnosis with extremely few-shot instances on the relatively simple transfer task. Our code is available at https://github.com/a1018680161/Few-shot-Transfer-Learning.

1. Introduction

1.1. Why few-shot transfer important

Diagnosis and prognosis of rotating machinery [1–3], such as aero-engine, high-speed train motor, and wind turbine generator, plays a core role in its safe operation and efficient work. Various signal processing methods based on sparse decomposition, manifold learning, and Minimum entropy deconvolution have been introduced to extract fault-attention features and diagnose fault modes from vibration signals in recent years [4–6]. Depending on the powerful feature extraction capabilities of deep learning, intelligent diagnostic methods greatly surpass traditional signal processing methods and quickly attract the interest of research [7,8]. Most of the famous machine learning methods such as Support Vector Machine (SVM) [9], Deep Boltzmann Machine (DBM) [10], Convolution Neural Networks (CNN) [11], autoencoders [12], Recurrent Neural Networks (RNN) [13], Generative Adversarial Networks (GAN) [14] have been utilized in the field of intelligent diagnosis. Comprehensive open source benchmark for intelligent fault diagnosis with supervised deep learning methods and unsupervised deep transfer learning methods were also presented recently [15,16].

The success of deep-learning-based intelligence diagnosis, however, relies on big data collected by sensors installed on the rotating machinery. Although the effectiveness of intelligence diagnosis methods has been demonstrated by a huge amount of experiments on open-source datasets [17,18], few-shot learning is still a challenge for rotating machinery intelligent diagnosis. In the few-shot application scenario, most of these methods would fail because deep learning is data-greedy, which has not been technically considered in the field of intelligent diagnosis. To explore the data sensitivity of deep learning methods with limited data, the experiment results of a bearing, spur gear, and helical gear datasets with various sample sizes are performed here. The experiment results are shown in Fig. 1. Limited by the extremely small training instances, the accuracy of training from scratch on the target domain will significantly decrease as the number of training instances decreases. A deep nonnegativity-constraint sparse autoencoder based method was proposed in [19] for the rolling bearing fault diagnosis with limited labeled data, in which the accuracy decrease with limited samples is also significant. It should be noted that although a few-shot training strategy was used by Zhang et al. [20] for bearing fault diagnosis, its essence is still the large instance training because there is no class splitting between training and testing sets.

* Corresponding author at: No. 28, Xianning West Road, Xi'an, Shaanxi 710049, China.

E-mail address: ch.sun@xjtu.edu.cn (C. Sun).

https://doi.org/10.1016/j.measurement.2020.108202
0263-2241/

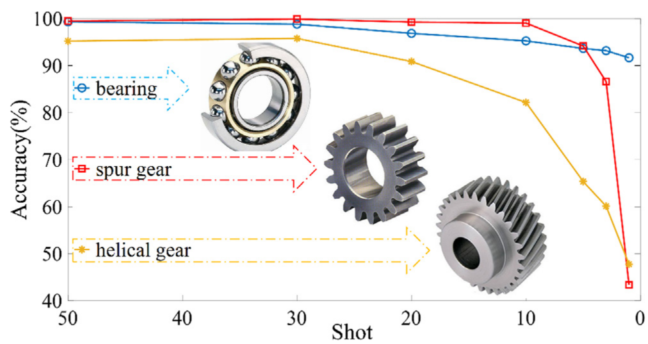

Fig. 1. The dependency relationship between diagnosis accuracy and shot.

1.2. How to transfer for intelligent diagnosis

To address performance degradation in the few-shot scenario, leveraging knowledge or data from relevant tasks is an intuitive way, also called transfer learning. We can transfer knowledge from the source domain to the target domain by training a classification network with large data on the source domain (such as data on one working condition) and selectively fine-tuning it on the target domain (such as data on another working condition). Meta-learning based on the episodic training strategy was also proposed in [21] and has flourished in various fields [22–24]. The training stage is conducted in the few-shot scenario by sampling a large number of meta tasks from data in the source domain. It means that the model is trained to learn from few-shot instances. As a result, in the testing stage, the trained network can adapt to the few-shot task in the target domain effectively.

According to the application background and research status of fault diagnosis, transferring between different working conditions (conditions transfer) and transferring from artificially simulated faults to natural faults (artificial-to-natural transfer) are two major transfer situations that can effectively solve the problem of insufficient samples.

Firstly, numerous rotating machinery operates under different working conditions in practice [25]. For example, the rotating speed of machines always increases from 0 to a fixed or ever-changing value and loads of machines vary according to different requirements. The changes in both rotating speeds, loads, and other parameters of machines will significantly influence the working conditions. To get a good generalization of the fault detector, adequate data on almost every working condition are necessary, which is not practical because of the high cost of fault simulation experiments. To address this issue, a few-shot learning strategy, which can transfer knowledge from one or multiple conditions to other new conditions, is required.

Secondly, for expensive large scale equipment, such as aero-engine, every single fault would cause huge economic cost and unacceptable casualty, so it is impossible to get enough natural damage data to train the deep neural network for fault classification. In practice, artificially simulated faults were made on rotating machinery to represent faults of the rotating machinery damaged in reality [26–28]. But there is a big gap between artificially simulated faults and natural faults. The artificially simulated fault usually made by electrical discharge machining (EDM) [29], drilling [30], wire-electrode cutting, manual electric engraving, etc. As shown in Fig. 2, bearings with artificially simulated faults have obvious marks of processing and are different from the natural fault in both appearance and mechanism, which results in different vibration modes, so the model trained with artificially simulated faults cannot be effectively utilized for the classification of natural faults. Although transferring knowledge from artificial bearing damages to natural bearing damages has been performed in [31], the influence of limited samples is not studied in detail. To train the fault detector with a small amount of natural damaged instances while utilizing abundant data with artificially simulated faults, it is indispensable to develop few-shot transfer learning methods for rotating machinery intelligent diagnosis.

1.3. What matters in few-shot transfer

From the few-shot perspective, three characteristics of different methods need to be considered during the transfer from the source domain to the target domain to reasonably employ few-shot learning strategy under different sample capacity and different task difficulty. (1). What is the data dependency of various methods in the few-shot scenario? (2). How does the model capacity to be fine-tuned affect the transferability of methods in the few-shot scenario? (3). How can we select appropriate models for transfer situations with different difficulty levels? These can be summarized as data dependency, transferability, and task plasticity.

In this paper, to explore those three different characteristics in-depth, seven models based on fine-tuning and meta-learning with a unified 1D convolution networks are implemented for the few-shot transfer learning in three datasets. According to the results of performance comparisons, three qualitative conclusions are made corresponding to data dependency, transferability, and task plasticity.

The contributions of this paper can be summarized as follow:

- A unified few-shot transfer learning framework based on 1D convolution network is designed for machine fault diagnosis under two typical transfer situations named working conditions transfer and artificial-to-natural transfer.
- Comparison analysis among designed feature transfer method, fine-tuning based method and meta-learning based method are performed to provide baseline accuracy under different sample capacity and transfer situations.
- Data dependency, transferability, and task plasticity analysis are discussed to conduct the reasonable methods selection under different transferring tasks and provide guidelines for further research.

2. Background

2.1. Few-shot learning

Few-shot learning was initially motivated by the human learning procedure. The human can learn to recognize a new object or master a new concept with only one or few-shot instances, so machine learning and deep learning models are also being placed great expectations to learn from few-shot instances. Few-shot methods in current research can be roughly classified into three threads [32], that is, data augmentation, data/model transfer learning, and meta-learning.

Data augmentation which does not rely on the additional datasets is a simple approach to perform during the training procedure. Generative adversarial networks (GANs) are widely used to augment data in recent years benefiting from its superior performance. For example, an auto-augment approach was proposed in [33] to find the best image augmentation policy from the designed search space. However, deep-learning-based data augmentation strategies can get into trouble when data are extremely scarce.

Learning from related data and transferring it to the target domain is the most intuitive solution for few-shot tasks. Representative methods include feature transfer methods and fine-tuning based methods. There are also some derivates such as fine-tuning based methods with partly frozen layers. Moreover, knowledge

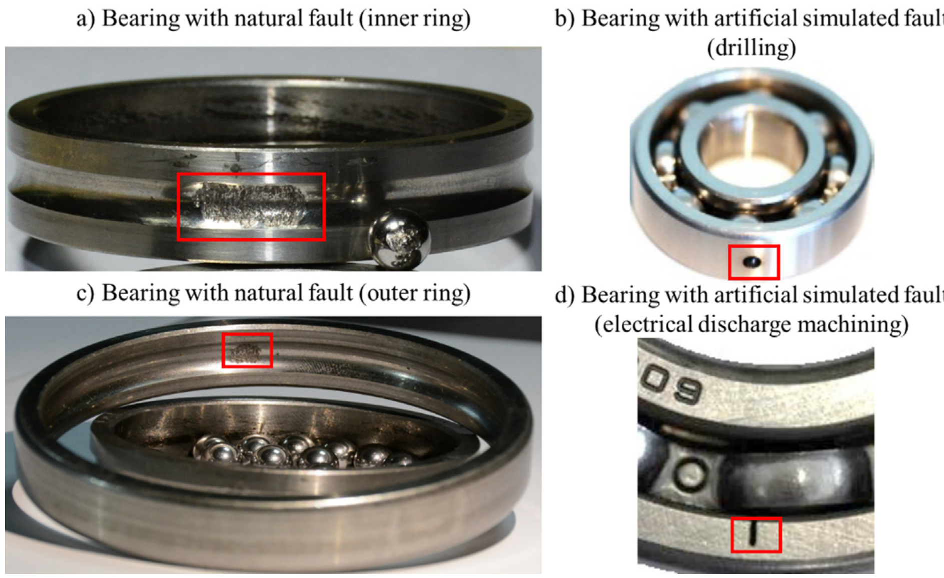

Fig. 2. Differences between bearings with natural faults and bearings with artificially simulated faults.

distillation was proposed in [34] with a teacher and student learning mode. For feature transfer methods, models trained on other datasets are directly used as feature encoders and a new classifier is subsequently trained on few-shot instances together with frozen feature encoders. For feature transfer methods, models can be pre-trained on related datasets and fine-tuned on the target data.

The core of meta-learning based method is "learn to learn". By imitating the learning procedure of human, meta-learning tries to improve the learning ability of networks for higher-level tasks rather than pure classification tasks. Model-Agnostic Meta-Learning [35] (MAML) and its derivates learn to adapt quickly with few-shot instances. Prototypical networks [36] and its derivates learn to assess the similarity between support instances and testing instances.

All the above few-shot learning methods have been widely applied in the field of computer vision, reinforcement learning, and so on. Various baseline datasets [37] and the benchmark study [38] are also conducted for the meaningful comparison and improvement of these methods on computer vision. Besides, in the field of medical diagnosis, few-shot learning methods, such as few-shot decision tree, was utilized for diagnosing the ultrasound breast tumor [39]. A three-headed siamese network was presented to diagnose Colon, Lung, and Breast tissue [40], and a predictive model based on the matching neural network architecture was used for glaucoma diagnosis [41].

Similar to the medical diagnosis field, preparing extensive and well-labeled datasets is also not feasible for certain tasks in rotating machinery intelligent diagnosis. To emphasize the importance of few-shot learning for intelligent diagnosis and provide guidance for further research, it is necessary to perform a baseline study of few-shot rotating machinery intelligent diagnosis and explore the data dependency, transferability, and task plasticity of different methods.

2.2. Few-shot transfer learning

Based on the assumption that source and target domains have different distributions, the main goal of transfer learning is to transfer knowledge learned from the source domain into the target domain. So the strategy of transfer learning methods is focused on aligning distributions between two domains. According to different expressive forms to be transferred, transfer learning methods can be divided into four categories [42], instances based, feature representation based, parameters based, and relational knowledge-based methods. When the number of support instances decreases sharply, the transfer learning task turns into few-shot transfer learning and the influence of limited instances is as noteworthy as the influence of different domains. Although not emphasized before, most of the few-shot learning methods mentioned above are based on transfer scenarios.

3. Few-shot transfer learning framework

3.1. Preprocessing

As the input length increasing, the parameters size of neural networks would increase rapidly, which means more training data are required to match the model capacity. But too small input length will lead to low frequency resolution and is not conducive to effective fault identification. So after the trade-off, the input length of models is set as 1024 uniformly. During the case study, all vibration signals in three datasets are firstly truncated into instances with the length of 2048 firstly. Although signals in the time domain retain all the information of original signals, it is widely recognized that signals in the frequency domain are more fault-sensitive than those in the time domain. So in this paper, Fast Fourier Transform (FFT) is applied to every instance and the length of instance is reduced to half of the original one.

3.2. Unified base network

For the fairness of comparison, we have designed one unified 1D convolution network, including a feature encoder network$f_{\theta}(\ )$and a classifier network$g_{\kappa}(\ )$, as shown in Fig. 3(a) and (b). The feature encoder network consists of four 1D convolutional blocks and one 1D adaptive-max-pooling layer which is suitable for processing 1D sequence signals. Instead of using a small convolutional kernel in image processing to capture the edge information in the image, we designed big convolutional kernel with kernel size 10 for the first convolutional layer to extract the rich shallow information representation of vibration signals. Because the vibration signal is more sensitive to the relevance of the whole in both time domain and frequency domain, and too small kernel size in first layer will

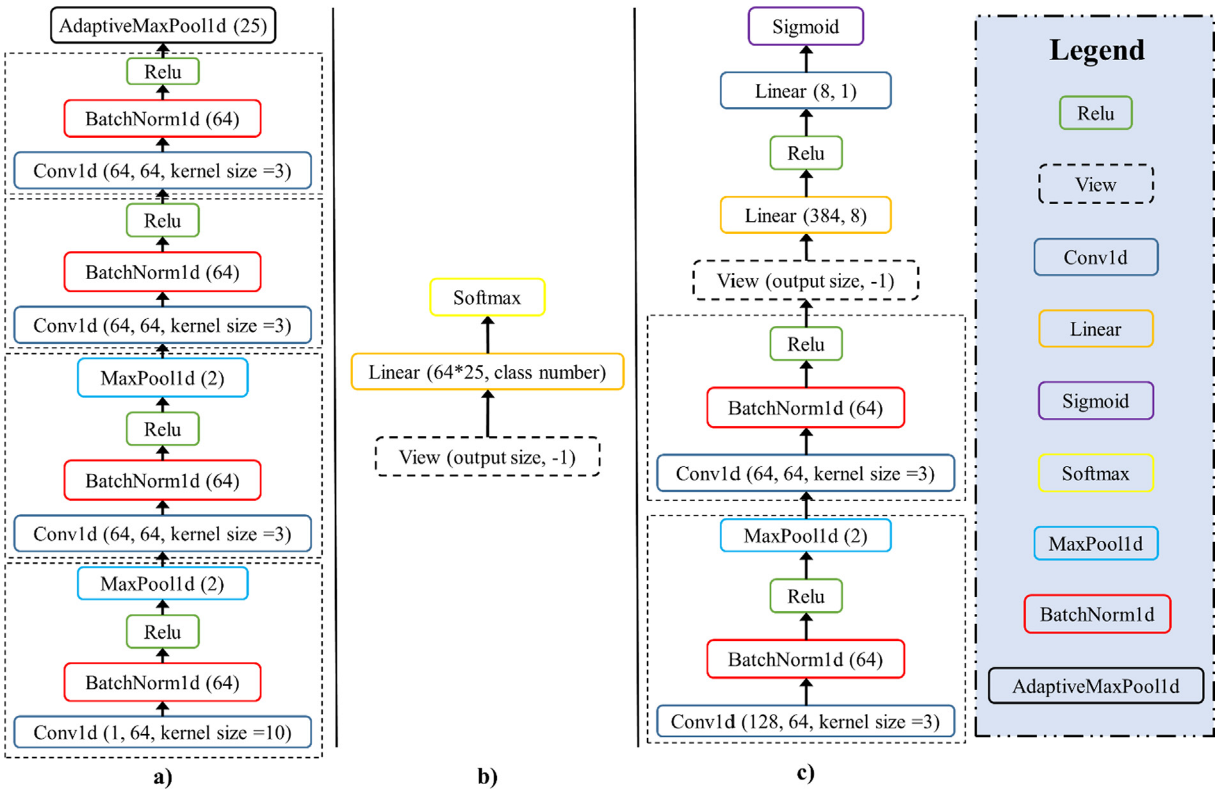

Fig. 3. Network structures of 1D convolution network (a) feature encoder, (b) classifier, (c) metric learner.

lose the useful information in the signal. The output of the feature encoder is a 1600 dimensional feature tensor with the shape of 64 channels * 25 widths. This feature encoder $f_\theta(\,)$ is implemented in all methods introduced in Fig. 4. The classifier network consists of a simple fully connected layer and a softmax activation function. The classifier network $g_\kappa(\,)$ is implemented in all methods except for the meta-learning based method.

3.3. Methods

In the few-shot transfer learning task, the data from source domain and target domain are all included, and the knowledge learned from source domain would be transferred for the diagnosis of target domain. In target domain, N types of classes are included and K instances are available for every class. So in the following, we refer the training data in target domain as N way * K shot for brevity. In meta-learning, data from source domain or target domain are divided into support set, which serves as labelled sample to produce Prototypical features for models and query set, which serves as the training samples to update the models.

3.3.1. Direct training method

The direct training method is indeed a simple way to directly classify the target category without any pre-training or few-shot strategy. Although it is not fair to compare it with other methods which have learned features from the source domain, it is

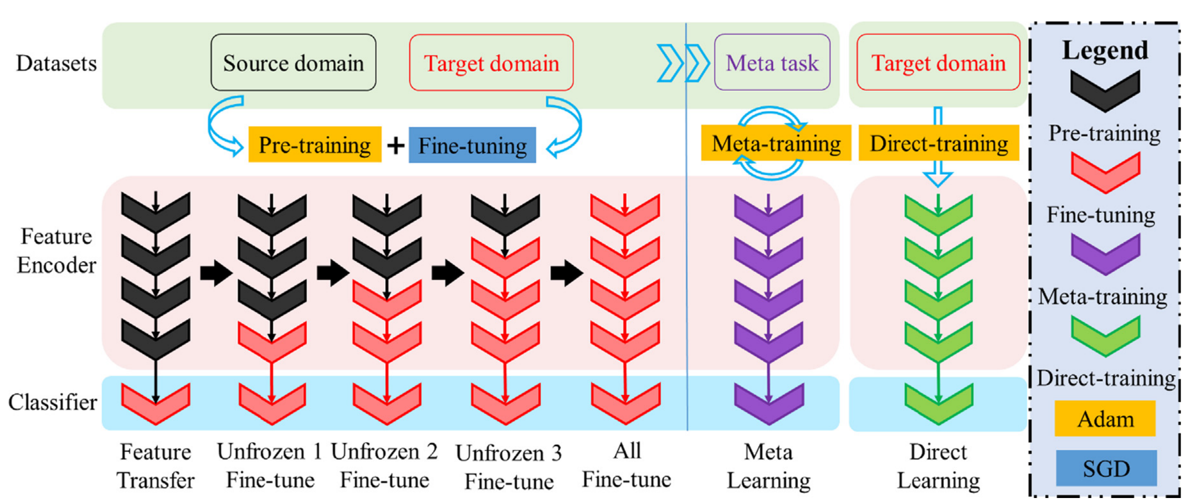

Fig. 4. Learning process of all methods.

necessary to demonstrate the capacity of this simple method to provide a lower bound of few-shot learning.

The learning process of direct training methods is to train the feature encoder and the classifier directly with the data of the target domain as shown in Fig. 4. Adam with the learning rate of 0.001 is utilized as the optimizer. All N way * K shot data are divided into mini-batches and trained with 50 epochs. The mean accuracy of the last 10 epochs is regarded as the final accuracy of this method. In the following, the direct training method is uniformly indicated as Direct Training Net (DTN).

3.3.2. Feature transfer method

A way to transfer is learning a good feature representation in the source domain and the learned feature encoder is regarded as a fixed feature extractor [43]. The learning process of feature transfer method can be divided into two stages as shown in Fig. 4. Firstly, a feature encoder is learned in the source domain by Adam optimizer with the learning rate of 0.001 and the unified training epochs of 200. Then the feature encoder learned before is fixed and a new classifier is trained in the target domain with the momentum-accelerated Stochastic Gradient Descent (SGD) optimizer [44] with the learning rate of 0.01. All N way * K shot data are divided into mini-batches and trained with 50 epochs. The mean accuracy of the last 10 epochs is regarded as the final accuracy of this method. When the source and target domains are similar enough, a large amount of source domain knowledge can be utilized to learn a unified representation which is suitable for the target domain. So during the training of the target classifier, dependence on large instances can be greatly reduced. In the following, the feature transfer method is uniformly indicated as Feature Transfer Net (FTN).

3.3.3. Fine-tuning based method

Another way to transfer the knowledge learned from the source domain is training a feature encoder as a good initialization and fine-tuning the pre-trained feature encoder with the data in the target domain [45]. The effectiveness of fine-tuning based methods has been demonstrated by the success of Bidirectional Encoder Representations from Transformers (BERT) [46] in natural language processing. So, in the field of fault diagnosis, pre-trained models based on vibration signals may also make sense to a certain extent. In this paper, four different levels of fine-tuning corresponding to the four convolutional blocks of the feature encoder are considered. As shown in Fig. 4, the number of unfrozen layers gradually increases from 1 to 4. For simplicity, these four models are indicated as Unfrozen m Fine-tuning Net (UmFN). The learning process of the fine-tuning based method can also be divided into two stages. Firstly, a feature encoder is learned in the source domain by Adam optimizer with the learning rate of 0.001 and the unified training epochs of 200. After that, a new classifier and the feature encoder with some layers fixed are fine-tuned in the target domain with SGD optimizer with a learning rate of 0.01. All N way * K shot data are divided into mini-batches and trained with 50 epochs. The mean accuracy of the last 10 epochs is regarded as the final accuracy of these methods.

3.3.4. Meta-learning based method

3.3.4.1. Meta relation net. Based on the Match Net in [21], Relation Net [47] was proposed by replacing the traditional distance metric function (such as Euclidean distance) with a metric learner, which can adaptively learn a distance metric function according to characteristics of features encoded by the feature encoder. Relation Net has made state-of-art records in various few-shot and zero-shot tasks for image classification. 1D convolutional Meta Relation Net is modified from the Relation Net and consists of a feature encoder f_θ in Fig. 3(a) and a metric learner r_ϕ in Fig. 3(c) with 2 convolutional blocks and 2 fully connected layers. As shown in Fig. 5, both instances of the support and query sets are fed into the feature encoder and projected into the unified feature space. Then extracted features in double branches would be concatenated into a relation feature and fed into the metric learner. Finally, a similar score between the instances in double branches can be calculated and the query instance would be classified as the category with the largest similar score.

3.3.4.2. Episodic training strategy. In traditional classification strategy, the data of all classes $\mathbb{C}$ are divided into training set $\mathrm{D}^{\mathrm{train}}\{\mathbf{x}\}$ and testing set $\mathrm{D}^{\mathrm{test}}\left\{\widehat{\mathbf{x}}\right\}$. The objective is learning the optimal projection function$f_{\theta^*}(\mathbf{x})$ between signals $\mathbf{x} \in \mathrm{D}^{\mathrm{train}}$and class labels $c \in \mathbb{C}$, which can be formulized as follow:

$$\kappa^*, \theta^* = \underset{\theta}{argmin}\left(\sum \mathrm{Loss}(g_\kappa(f_\theta(x)), c)\right);\ c \in \mathbb{C}, \mathbf{x} \in \mathrm{D}^{\mathrm{train}} \tag{1}$$

where Cross Entropy Loss is always chosen as the loss function and $g_\kappa(f_\theta(x))$ is tested on signals $\widehat{\mathbf{x}} \in \mathrm{D}^{\mathrm{test}}$.

Different from the traditional classification strategy, the episodic training strategy for meta-learning based method can be modeled as a task learning process. Firstly the training data from source domain is divided into training support set $\mathrm{D}_s^{\mathrm{source}}\{\mathbf{x}\}$, including K instances for each training class, and training query set $\mathrm{D}_q^{\mathrm{source}}\{\mathbf{x}\}$. Similarly, the testing data from target domain is divided into testing support set $\mathrm{D}_s^{\mathrm{target}}\left\{\widehat{\mathbf{x}}\right\}$, including K instances for each testing class, and testing query set $\mathrm{D}_q^{\mathrm{target}}\left\{\widehat{\mathbf{x}}\right\}$. The objective is learning the projection function $r_\varphi(f_\theta(x))$between signals $\mathbf{x}$and class labels c, which can be formulized as follow:

$$\mathbf{r}^*, \theta^* = \underset{\theta}{argmin}\left(\sum Loss(f_\theta(\mathbf{x}), c)\right); \mathbf{x} \in \mathrm{D}_s^{\mathrm{source}} \tag{2}$$

During the training stage, K instances from random N classes in source domain are selected as the support set and the rest data of these N classes are selected as the query set to update the networks. During the testing stage, K instances from N classes in target domain are selected as the support set and the rest data of these N classes are selected as the query set to test the classification accuracy of the model.

The learning procedure of Meta Relation Net with the episodic training strategy is shown in Fig. 6, including preparation, meta-learning stage, and transfer stage. And the detailed pseudo-code of the method is listed in Table 1. In the preparation, hyper-parameter is set up and data in source and target domains are organized according to the demand of different tasks. In the meta-learning stage, the feature encoder and metric learner are trained in the source domain with the episodic training strategy. Mean square error (MSE) is utilized as the loss function to calculate the distance between predicted relation scores and labels. In the transfer stage, query instances are fed into the learned network together with N way * K shot support instances in the target domain to validate the testing accuracy of the method. Adam with the learning rate of 0.001 is utilized as the optimizer. The mean accuracy of 100 episodes is regarded as the final accuracy. In the following, 1D convolutional Meta Relation Net is uniformly indicated as MRN.

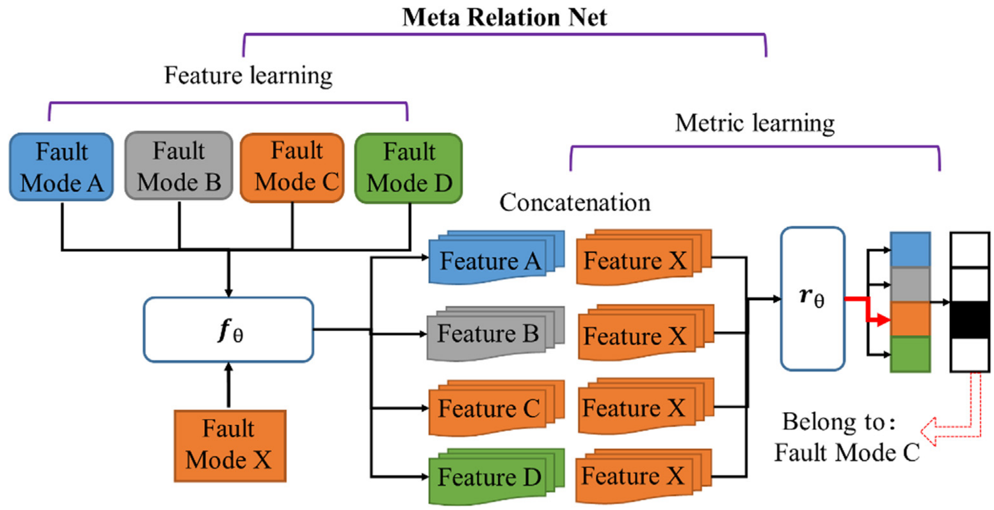

Fig. 5. 1D Meta Relation Net with double input branches.

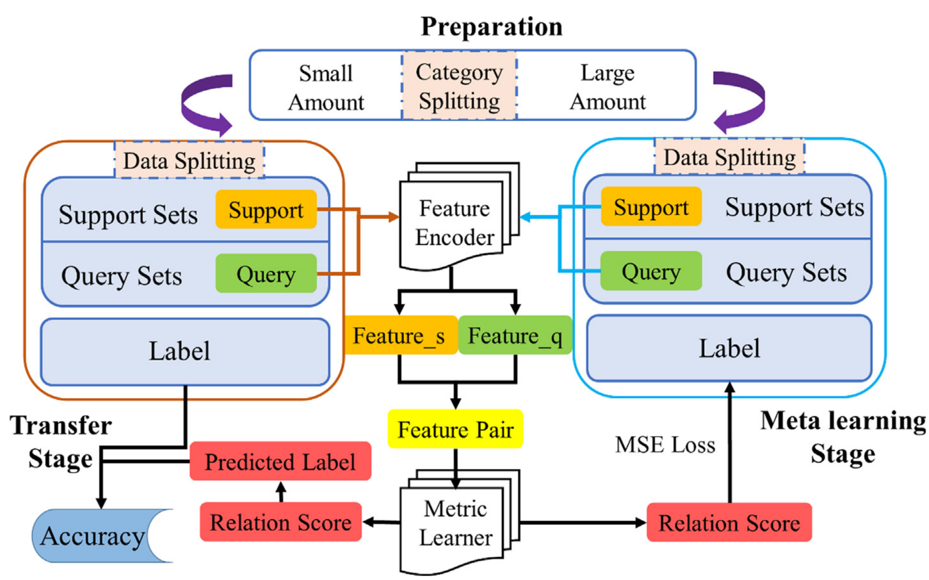

Fig. 6. Learning and transfer procedure of Meta Relation Net with the episodic training strategy.

4. Case study

4.1. Datasets

4.1.1. PU bearings fault datasets

PU bearing datasets [48] were provided by Lessmeier et al for condition monitoring and diagnosis of bearings. 32 bearings of type 6203 including 12 artificial damaged bearings, 14 natural fault bearings by accelerated lifetime tests, and 6 healthy bearings. All of them were tested on 4 different working conditions listed in Table 2. Meanwhile, fault severity is classified into two levels according to the length of damage, where levels 1 and 2 represent the damage length less than and greater than 2 mm respectively. Three processing modes were considered for artificially simulated faults. Bearings with natural faults were produced by an accelerated run-to-failure test. After that, all bearings were installed in the modular test rig for a unified test. Motor currents and vibration signals together with three parameters (the radial force, the load torque, and oil temperature) were collected during experiments.

In the following case study, only the data of 1 healthy bearing, 8 artificial damaged bearings, and 5 natural fault bearings in N09_M07_F10 working condition are used to perform artificial-to-natural few-shot transfer learning. The details of these 14 bearings are shown in Table 3. The visualization of signal in time domain and frequency domain is shown in (1) and (4) of Fig. 7.

4.1.2. SQI gears fault datasets

Conditions transfer experiments were carried out on SQI drive system diagnosis integrated testbed. As shown in Fig. 8(a), the testbed consists of the motor, the controller, a two-stage planetary gearbox, a two-stage parallel shaft gearbox supported by rolling bearings, a radial loading device, and a magnetic powder brake. It can simulate the fault of spur gears, helical gears, rolling bearings, and the coupling fault of them.

In this experiment, only gears faults in the first stage of parallel shaft gearbox were considered and one healthy spur gear and three faulty spur gears were tested with three different rotating modes. First mode is a cycle of acceleration and deceleration, accelerating from the rotating speed of 0 rpm to 1800 rpm and decelerating from the rotating speed of 1800 rpm to 0 rpm, abbreviated as '0–30–0 Hz'. Second mode is smoothly operating at a constant speed of 1800 rpm, abbreviated as '30 Hz'. The third mode is smoothly

Table 1
The detailed pseudo-code of Meta Relation Net with the episodic training strategy.

Preparation

1. Set up and **initialize** feature encoder f_θ and metric learner r_ϕ;
K, N, M → way, shot, query instances number for each class
2. Splitting support and query
$D^{source}\{\mathbf{x}\},C^{source}\{c\}\rightarrow D_s^{source}\{\mathbf{x}\},D_q^{source}\{\mathbf{x}\},C_s^{source}\{c\},C_q^{source}\{c\}$
$D^{target}\{\mathbf{x}\}\rightarrow D_s^{target}\left\{\widehat{\mathbf{x}}\right\},D_q^{target}\left\{\widehat{\mathbf{x}}\right\}$

Meta-learning Stage

3. Input: $D_s^{source}\{\mathbf{x}\},D_q^{source}\{\mathbf{x}\},C_s^{source}\{c\},C_q^{source}\{c\}$
For i in train epochs
(1) N*K instances$\{\mathbf{x}^s\}$ from $D_s^{source}\{\mathbf{x}\}$M*K instances$\{\mathbf{x}^q\}$from $D_q^{source}\{x\}$
(2) $\{\mathbf{z}^s\},\{\mathbf{z}^q\}=f_\theta(\{\mathbf{x}^s\},\{\mathbf{x}^q\})$
(3) $\{\mathbf{z}\}$ = concat($\{\mathbf{z}^s\},\{\mathbf{z}^q\}$)
(4) $\{\mathbf{y}\}$ = softmax($r_\phi(\{\mathbf{z}\})$)
(5) $\{d\}$ **= Match(**$C_s^{source}\{c\},C_q^{source}\{c\}$**)**
(6) **Backpropagation** Loss = MSE ($\{y\},\{d\}$)
End For
Output: f_θ,r_ϕ

Transfer Stage

4. Input: $D_s^{target}\{\mathbf{x}\},D_q^{target}\{\mathbf{x}\},C_s^{target}\{c\}$
For i in test epochs
(1) N*K instances $\{\mathbf{x}^s\}$ from $D_s^{target}\left\{\widehat{\mathbf{x}}\right\}$M*K instances $\{\mathbf{x}^q\}$ from $D_q^{target}\left\{\widehat{\mathbf{x}}\right\}$
(2) $\{\mathbf{z}^s\},\{\mathbf{z}^q\}=f_\theta(\{\mathbf{x}^s\},\{\mathbf{x}^q\})$
(3) $\{\mathbf{z}\}$ = concat($\{\mathbf{z}^s\},\{\mathbf{z}^q\}$)
(4) $\{\mathbf{y}\}$ = softmax($r_\phi(\{\mathbf{z}\})$)
End For
Output:$\{\mathbf{y}\}$

Table 2
Description of 4 work conditions in PU bearing fault datasets.

Setting name	Load torque [Nm]	Radial force [N]	Rotational speed [rpm]
N15_M07_F10	1500	0.7	1500
N09_M07_F10	900	0.7	900
N15_M01_F10	1500	0.1	1500
N15_M07_F04	1500	0.7	1500

Table 3
Details of bearings used in artificial-to-natural transfer few-shot learning.

Name	Fault location	Manufacturing way	Severity
KA01	Outer ring	electrical discharge machining	1
KA03	Outer ring	electric engraver	2
KA05	Outer ring	electric engraver	1
KA07	Outer ring	Drilling	1
KA08	Outer ring	Drilling	2
KI01	Inner ring	electrical discharge machining	1
KI03	Inner ring	electric engraver	1
KI05	Inner ring	electric engraver	2
Name	Fault location	Fault description	Severity
K001	Healthy		
KA04	Outer ring	fatigue: pitting	1
KB23	Outer + inner	fatigue: pitting	2
KB27	Outer + inner	Plastic deform.: Indentations	1
KI04	Inner ring	fatigue: pitting	1

operating at a constant speed of 2400 rpm, abbreviated as '40 Hz'. So both of the constant rotation and variable rotation are included in this experiment. The details of the gears are shown in Fig. 8 b) including healthy gear, missing teeth fault, root cracks fault, and surface wear fault. The location of every fault is marked with a red rectangle. Vibration signals were collected by acceleration sensors with a sampling frequency of 20480 Hz. The visualization of signal in time domain and frequency domain is shown in (2) and (5) of Fig. 7.

4.1.3. Qianpeng gears fault datasets

Another conditions transfer experiments are carried out on Qianpeng diagnosis integrated testbed. As shown in Fig. 9(a), the testbed consists of the motor, a gearbox fault simulation, a dynamic balance simulation, a bearings fault simulation, a radial loading device, and a magnetic powder brake. It can simulate the fault of the gearbox, shaft dynamic balance, bearings, and the coupling fault of them.

In this experiment, only gearbox faults were considered and one healthy helical gear and six faulty helical gears were tested with 21 different rotating speed modes. All of these rotating speed modes are smoothly operating at a constant speed of 1000 rpm, 1020 rpm, 1040 rpm, ..., 1400 rpm respectively. The details of the gears are shown in Fig. 9(b) including healthy gear, missing teeth fault, four gears with different sizes of root cracks (1 mm, 2 mm, 3 mm, and 4 mm), one gear with one tooth surface wear fault, and one gear with double teeth surface wear fault. The thickness of the surface wear is 0.5 mm. Vibration signals were collected by acceleration sensors with a sampling frequency of 20480 Hz. The visualization of signal in time domain and frequency domain is shown in (3) and (6) of Fig. 7.

4.2. Conditions transfer

Conditions transfer is commonly researched for rotating machinery fault diagnosis. For rotating machinery, working conditions (such as rotating speed and load) are always time-varying. Traditional intelligence diagnosis methods always focus on the fault classification of single working conditions or certain conditions. Once the machine operates on an unseen working condition, the model learned before will fail. To be adaptable to few-shot instances of new conditions, the few-shot transfer study between various conditions of rotating machinery should be implemented. Experiments on two datasets (SQI gears fault datasets and Qianpeng gears fault datasets) are carried out for this case study. Seven methods, including Direct Training Net (DTN), Feature Transfer Net (FTN), Unfrozen 1 Fine-tuning Net (U1FN), Unfrozen 2 Fine-tuning Net (U2FN), Unfrozen 3 Fine-tuning Net (U3FN), Unfrozen 4 Fine-tuning Net (U4FN), and Meta Relation Net (MRN) are utilized for the few-shot transfer of those two datasets.

Four transfer situations representing different transfer difficulty are considered on SQI gears fault datasets, including 30 → 40 (transfer from '30 Hz' to '40 Hz'), 40 → 30 (transfer from '40 Hz' to '30 Hz'), 0–30–0 → 40 (transfer from '0–30–0 Hz' to '40 Hz'), and 40 → 0–30–0(transfer from '40 Hz' to '0–30–0 Hz'). Four shot sets from 1 shot to 10 shot are considered for each transfer situation. Results of final accuracy on SQI gears fault datasets are listed in Table 4, in which the best method in each transfer situation is in bold and the worst one is underlined.

It is shown that FTN and MRN achieve the best performance in all transfer situations. Unexpectedly, fine-tuning based methods achieve the worst performance in most of the situations, even lower than the "low-bound" (DTN). This is probably because that fine-tuning the feature encoder unreasonably would destroy the learned feature representation and backfire, especially on extremely few-shot situations. All the best accuracies in 30 → 40, 40 → 30, 0–30–0 → 40 situations are close to or above 98% even with only 1-shot instances, which shows that the signals between two working conditions are similar enough. While the results in 40 → 0–30–0 are all below 90%. Because transferring from single to variable working conditions is obviously more difficult than transference from variable to single working conditions or transferring between single working conditions.

In Qianpeng gears fault datasets, data on 21 sets of rotating speed were collected. To carry out few-shot transfer experiments with different transfer difficulty, four kinds of train-test splitting

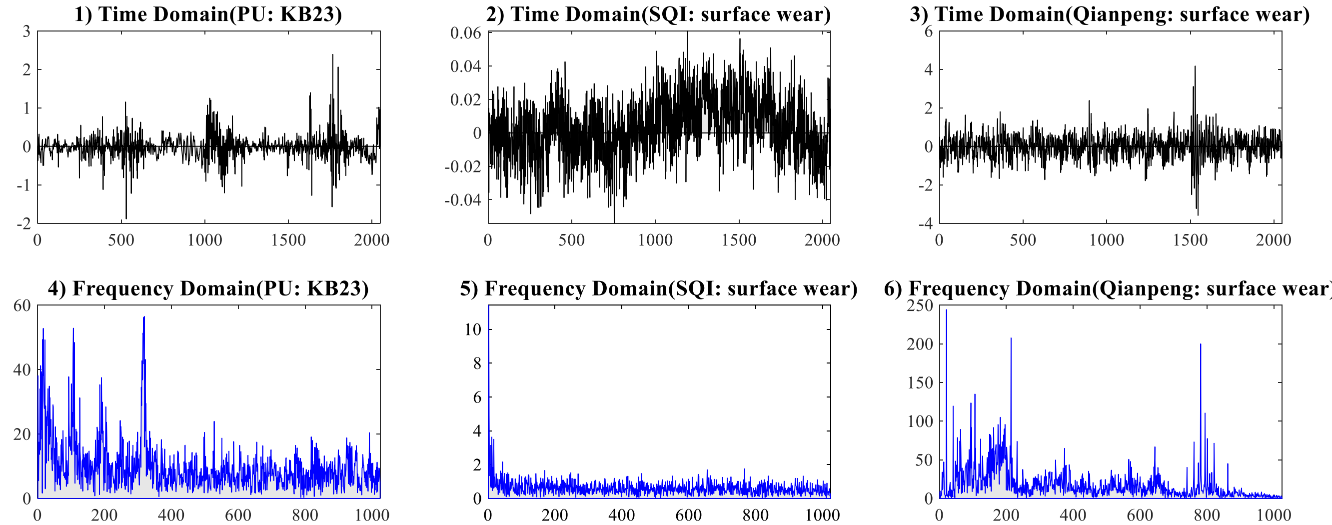

Fig. 7. Visualization of signal from three datasets in time domain and frequency domain.

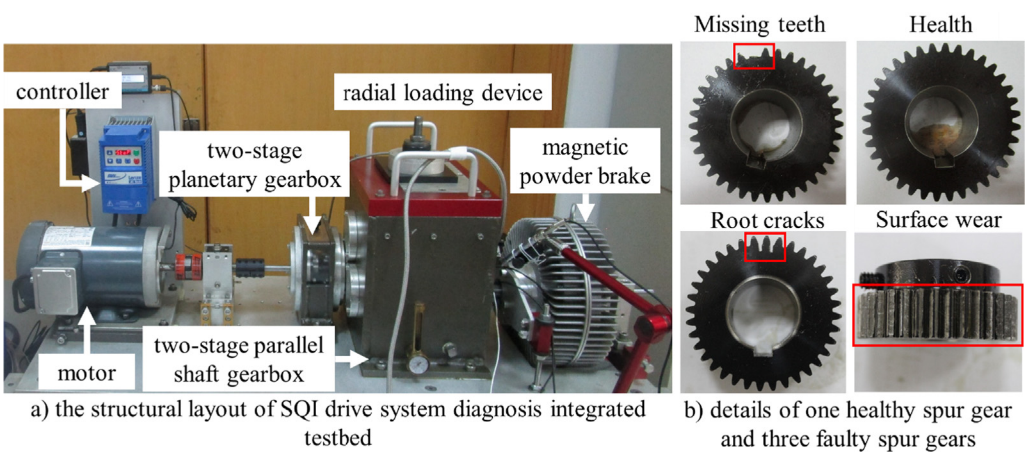

Fig. 8. SQI datasets.

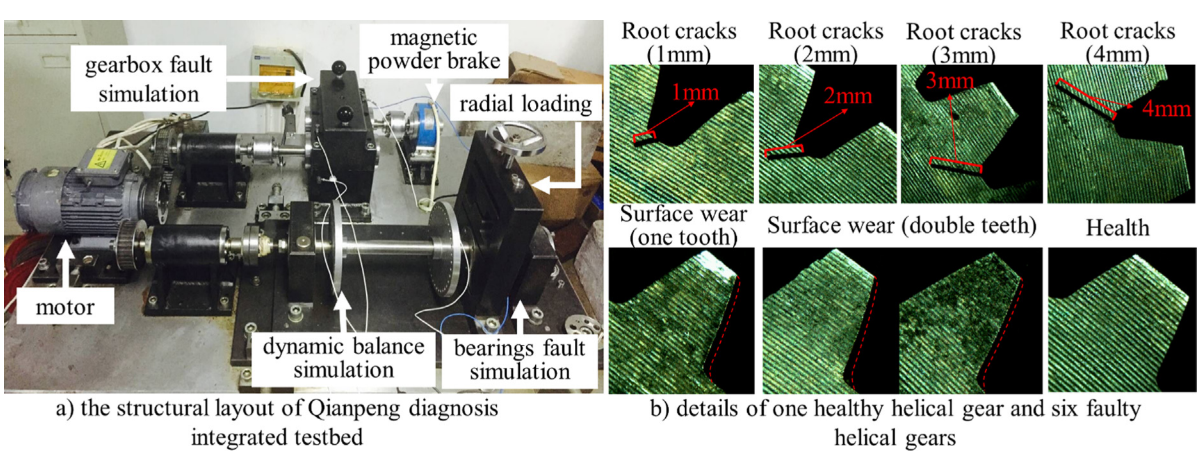

Fig. 9. Qianpeng datasets.

situations are considered, including 1 → 2–21 (transfer from the speed 1000 rpm to the speed 1020, . . ., –1400 rpm), 1–5 → 6–21 (transfer from the speed 1000, . . ., –1080 rpm to the speed 1100, . . ., –1400 rpm), 1–10 → 11–21 (transfer from the speed 1000, . . ., –1180 rpm to the speed 1200, . . ., –1400 rpm) and 1–20 → 21 (transfer from the speed 1000, . . ., –1380 rpm to the speed 1400 rpm). Four shot sets from 1 shot to 10 shot are also considered for each transfer situation. Results of the final accuracy on

Table 4
Accuracy comparison on SQI gears fault datasets (percentage).

Conditions	Shot	DTN	FTN	U1FN	U2FN	U3FN	U4FN	MRN
30 → 40	1-shot	75.31	90.85	91.75	93.12	72.75	66.57	**97.99**
	3-shot	97.73	98.93	98.75	77.83	92.35	73.64	**99.59**
	5-shot	94.25	**99.86**	97.11	96.25	88.46	94.21	99.44
	10-shot	99.13	**99.84**	98.55	96.34	97.29	92.32	99.69
40 → 30	1-shot	43.39	87.51	68.60	73.35	55.41	57.20	**98.62**
	3-shot	86.51	97.00	78.84	**99.68**	93.89	81.03	99.61
	5-shot	94.17	99.66	99.79	98.99	92.63	85.42	**99.91**
	10-shot	99.05	**100.00**	90.52	98.81	94.14	93.12	99.63
0–30–0 → 40	1-shot	72.55	93.32	91.62	82.23	56.28	74.40	**98.72**
	3-shot	89.36	96.69	94.10	72.04	90.89	69.57	**98.91**
	5-shot	92.84	98.68	97.90	95.82	83.28	86.97	**99.24**
	10-shot	94.88	**99.86**	98.44	94.48	90.23	94.51	99.10
40 → 0–30–0	1-shot	51.66	67.78	47.68	45.94	44.84	35.63	**70.38**
	3-shot	60.22	**86.70**	84.27	58.12	69.78	50.91	82.92
	5-shot	81.92	**87.31**	69.93	74.24	56.57	51.62	85.07
	10-shot	88.33	**88.56**	84.88	84.75	83.76	70.57	85.99

SQI gears fault datasets are listed in Table 5, in which the best method in each transfer situation is in bold and the worst one is underlined.

Same as in SQI datasets, FTN and MRN achieve the best performance in all transfer situations and fine-tuning based methods achieve worst performance in most of the situations, even lower than the "low-bound" (DTN). Unexpectedly, MRN achieves the worst accuracy in 1 → 2–21, which means that MRN does not work at all in this situation. The reason may be that transferring from one working condition to multiple working conditions deviates from the definition of meta-learning. The knowledge learned from the source domain does not meet the requirements of transferability. In other words, the performance gains of MRN depends on the abundant information in the source domain.

4.3. Artificial-to-natural transfer

In fact, the existing machine fault diagnosis methods rely heavily on the data produced by fault simulation in laboratory, which means there is still a gap between research and industrial application. Transferring from artificially manufactured faults to natural faults is a considerable way in the application of rotating machinery intelligence diagnosis to address the predicament that only small samples available for natural faults. In other words, we can replace a large number of costly natural fault samples with relatively low-cost artificial fault samples and get relatively high accuracy. Inspired by this, a case study is conducted on PU bearings fault datasets, including both artificially simulated faults data and natural faults data, using the same seven methods.

Table 5
Accuracy comparison on Qianpeng gears fault datasets (percentage).

Train-test splitting	Shot	DTN	FTN	U1FN	U2FN	U3FN	U4FN	MRN
1 → 2–21	1-shot	34.57	**52.23**	48.52	48.37	38.93	48.01	42.76
	3-shot	56.78	**67.31**	63.54	57.59	58.30	54.58	43.64
	5-shot	64.16	**77.20**	66.81	73.12	66.13	62.80	43.45
	10-shot	77.43	**82.82**	74.12	79.17	63.16	76.22	55.69
1–5 → 6–21	1-shot	47.78	53.92	40.29	49.90	48.95	39.01	**67.28**
	3-shot	60.12	71.99	65.08	70.12	66.74	69.16	**86.24**
	5-shot	65.35	82.50	71.57	68.45	62.13	60.68	**93.88**
	10-shot	82.16	91.16	80.91	76.26	73.99	78.16	**97.39**
1–10 → 11–21	1-shot	41.54	59.50	58.78	55.36	54.45	50.92	**88.43**
	3-shot	67.79	73.58	78.90	70.62	68.71	68.49	**93.28**
	5-shot	73.32	88.76	75.32	81.35	78.09	70.83	**96.57**
	10-shot	85.28	90.47	81.71	86.95	76.93	79.95	**94.95**
1–20 → 21	1-shot	79.63	77.97	69.14	85.48	58.14	77.35	**97.84**
	3-shot	91.15	95.48	89.79	88.46	87.54	80.77	**99.91**
	5-shot	94.34	95.54	92.30	81.43	95.46	86.55	**99.38**
	10-shot	96.23	97.64	95.57	89.50	92.21	90.22	**99.43**

Table 6
Accuracy comparison on PU bearings fault datasets (percentage).

Shot	DTN	FTN	U1FN	U2FN	U3FN	U4FN	MRN
1-shot	91.67	87.80	**94.27**	93.86	91.20	84.48	94.10
3-shot	93.10	**98.62**	93.10	96.61	92.88	89.57	94.85
5-shot	93.67	**98.45**	96.84	98.36	94.03	91.89	95.69
10-shot	95.26	**97.79**	97.00	96.82	92.04	97.14	95.87

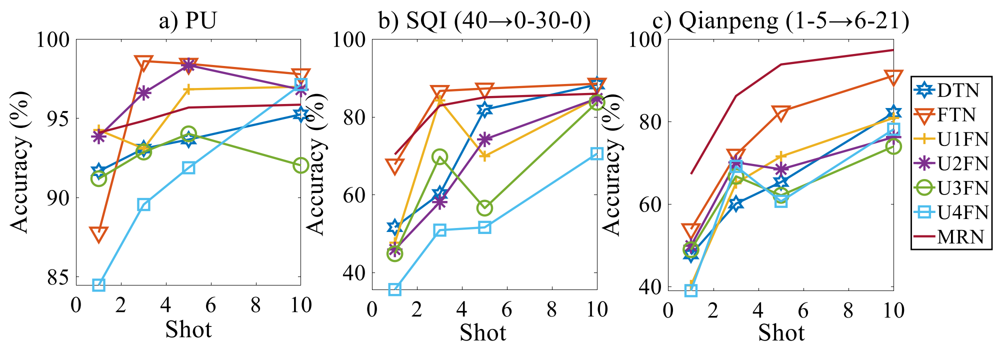

Fig. 10. Data dependency of seven methods on three datasets.

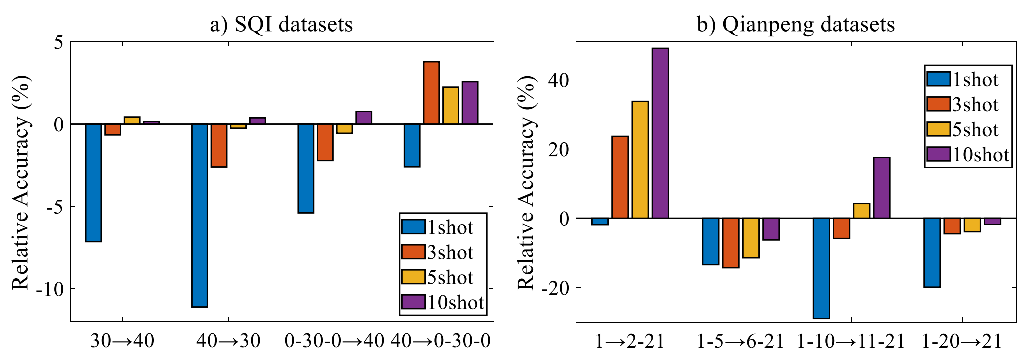

Fig. 11. Data dependency of FTN and MRN on SQI and Qianpeng datasets. (relative accuracy means subtracting the accuracy of FTN to the accuracy of MRN).

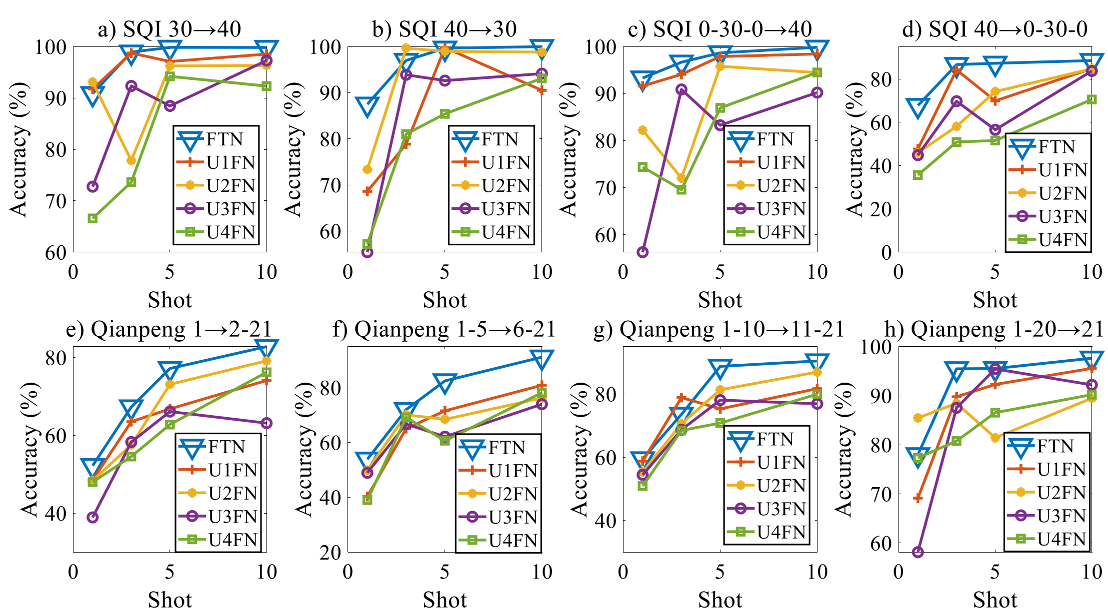

Fig. 12. Unfrozen study on SQI and Qianpeng datasets.

Transfer situations in this case study are set as transferring from eight bearings with artificially manufactured faults to four bearings with natural faults and one healthy bearing as listed in Table 3. Four shot sets from 1 to 10 are also considered for each transfer situation. The results of the final accuracy are listed in Table 6, in which the best method in each transfer situation is in bold and the worst one is underlined.

It is shown that FTN achieves the best accuracy in 3-shot, 5-shot and 10-shot situations. And U1FN achieves the best accuracy in the 1-shot situation. Meanwhile, in the 1-shot situation, MRN achieves the accuracy close to the first place. The worst accuracy is also made by U3FN and U4FN. It can be concluded that knowledge learned from artificially simulated faults data can be effectively transferred into natural faults data and improve the classification accuracy. This provides a feasible way for models learned from laboratory datasets to improve the generalization performance in reality.

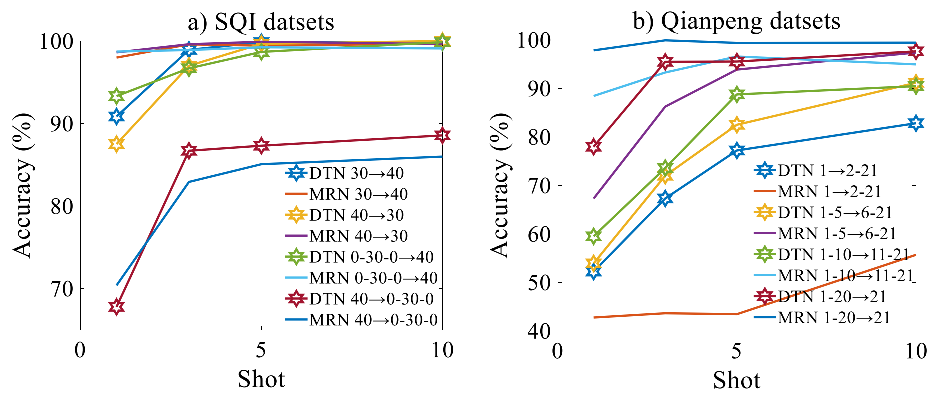

Fig. 13. Transfer difficulty study on SQI and Qianpeng datasets.

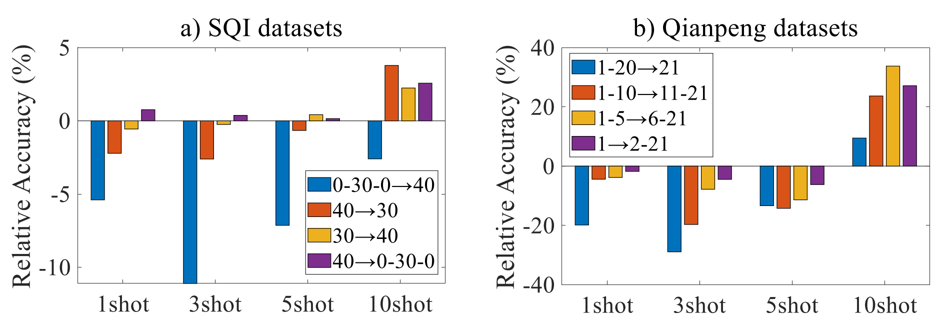

Fig. 14. Task plasticity study of FTN and MRN on SQI and Qianpeng datasets (relative accuracy means subtracting the accuracy of FTN to the accuracy of MRN).

5. Results analysis

Case study of conditions transfer and artificial-to-natural transfer shows that none of the methods dominates in all transfer situations. So it is necessary to employ different methods in different transfer situations. To be specific, data dependency analysis is proposed to study the influence of sample capacity on diagnostic accuracy. The goal of transferability analysis is to explore how does the model capacity to be fine-tuned affect the transferability of methods in the few-shot scenario. By changing the relationship between the source domain and target domain, task plasticity (difficulty) analysis can be employed to guide the selection of the methods under various transfer situations.

5.1. Data dependency study

Regardless of the influence of randomness, it can be expected that the test accuracy increases with the increase of the shot number as shown in Fig. 10. Meanwhile, there is some degree of random fluctuation on the results of fine-tuning based methods, because of the instability derived from too many unfrozen layers in the fine-tuning.

It is also notable that neither FTN nor MRN can dominate in all shots. Furthermore, it is observed in Fig. 11 that as the shot number increases, the performance of FTN gradually surpasses that of MRN. Specifically, MRN dominates only in extremely few-shot situations and when the shot number reaches a certain level, FTN will outperform MRN. The reason is that FTN is data-greedy and can get more benefit from the data growth, while MRN is specifically designed for few-shot situations and less sensitive to data growth. Therefore, the data dependency of two methods leads to the crossover of performance.

5.2. Transferability study

Results of fine-tuning based methods and FTN are shown in Fig. 12. It can be observed that FTN outperforms all of the fine-tuning based methods. No model is dominating in fine-tuning based methods and the performance of them relies on different transfer situations. Obviously, the unreasonable unfreezing strategy would be detrimental to the classification performance and even backfire. This phenomenon demonstrates that the capacity of fine-tuned layers should be compatible with the shot number. Too many unfrozen layers will lead to the disorder of fine-tuning in few-shot situations. So in the extremely few-shot set, merely unfreezing the classifier makes more sense for these datasets.

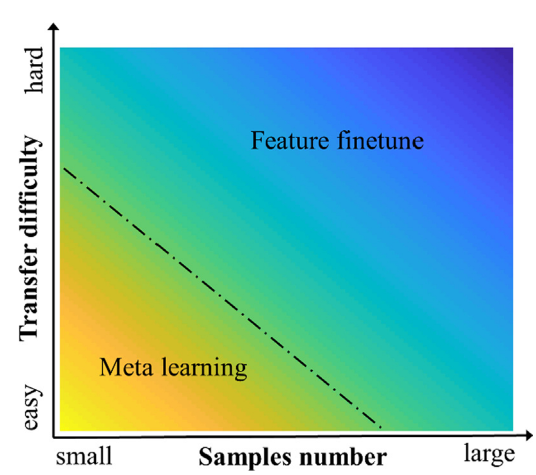

Fig. 15. Task plasticity and data dependency of FTN and MRN in few-shot situations (yellow and blue in background respectively indicate MRN and FTN dominating).

Table 7
Pros and Cons of different methods.

Methods	Pros	Cons
Direct	training	Without source domain data required
Low	accuracy	
Feature	transfer	Suitable for easy transfer task Suitable for medium-shot data
Require	related source domain data	
Fine-tuning	Low computer resources required	Low accuracy
Meta-	learning	Suitable for extremely low-shot dataSuitable for difficult transfer task
Require	sufficient source domain data	

5.3. Task plasticity (difficulty) study

In the case study, different transfer situations represent different transfer difficulty levels. For SQI datasets, the rank of transfer difficulty (from easy to difficult) is about 30 → 40 (single to single) = 40 → 30 (single to single) ≈ 0–30–0 → 40 (varying to single) < (single to varying). For Qianpeng datasets, the rank of transfer difficulty is about 1–20 → 21 (multiple to single) < 1–10 → 11–21 (multiple to multiple) < 1–5 → 6–21 (multiple to multiple) < 1 → 2–21 (single to multiple). Just as expected, it is shown that in Fig. 13, the more difficult the transfer situation is, the lower the accuracy for both DTN and MRN is. The accuracy gap between the easiest situation and the most difficult situation is huge and can reach 50%. So it is necessary to realize the degree of difficulty of situations before performing the few-shot transfer learning task. And it is also notable that in the field of rotating machinery intelligence diagnosis, difficult transfer situations, such as transferring from single to varying, transferring from single to varying or even cross-domain transfer, are more considerable.

It can be shown in Fig. 14 that neither FTN nor MRN can dominate in all difficulty levels. As the difficulty increases, FTN gradually surpasses MRN in performance. Specifically, MRN dominates only in relatively easy situations and when the difficulty reaches a certain level, FTN will outperform MRN. This suggests that when the two domains are similar enough, meta-learning strategies can more stably transfer source domain knowledge to the target domain while avoiding overfitting. But as the difference between two domains becomes larger, fine-tuning with samples from the target domain becomes more necessary.

5.4. Comprehensive analysis

Combining the analysis above, the combined effect of shot number and task difficulty on FTN and MRN can be expressed more clearly as in Fig. 15. MRN can dominate in extremely few-shot and relatively easy transfer situations and the feature transfer method can dominate as the shot number or the transfer difficulty increasing, which would guide the development of few-shot transfer learning in the future by taking advantage of both methods. Additionally, we summary the pros and cons of different methods constructed in this paper in Table 7

6. Conclusion

In this paper, the few-shot transfer learning problem in rotating machinery intelligence diagnosis is researched considering transfer situations of conditions transfer and artificial-to-natural transfer. Seven models, including DTN, FTN, U1FN, U2FN, U3FN, U4FN, and MRN, are implemented with a unified 1D convolution base network specially designed for the vibration signals. Case studies on PU bearings fault datasets, SQI gears fault datasets, and Qianpeng gears fault datasets are carried out. A baseline accuracy for detailed research under different sample capacity in rotating machinery intelligence diagnosis is provided.

Conclusions of transfer learning in few-shot situations are summarized as (1) The testing accuracy increases with the increase of the shot number. Meta Relation Net tends to dominate in extremely few-shot (<5) and as the shot number increasing, Feature Transfer Net dominates gradually. (2) The testing accuracy increases with the decrease in transfer difficulty. Meta Relation Net is more specific to simple transfer tasks. When the similarity between the source and target domains decreases, Feature Transfer Net outperforms others. (3) Too many unfrozen layers will lead to the disorder of fine-tuning or even backfire and weaken the transferability of models, and hence, in the extremely few-shot set, merely unfreezing the classifier makes more sense. These conclusions would raise more concerns beyond diagnostic accuracy and provide guidelines for the further research of few-shot transfer learning in rotating machinery intelligence diagnosis, e.g. extremely few-shot specific methods, cross-domain few-shot transfer methods, and hybrid-model few-shot transfer methods.

CRediT authorship contribution statement

Jingyao Wu: Conceptualization, Methodology, Visualization, Writing - original draft, Writing - review & editing, Formal analysis. **Zhibin Zhao:** Conceptualization, Data curation, Investigation. **Chuang Sun:** Conceptualization, Funding acquisition, Project administration, Writing - original draft, Writing - review & editing. **Ruqiang Yan:** Funding acquisition, Supervision, Data curation. **Xuefeng Chen:** Funding acquisition, Supervision, Resources.

Declaration of Competing Interest

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

Acknowledgments

This work was supported in part by The National Key Research and Development Program of China [No. 2018YFB1702400], Natural Science Foundation of China [No. 51705398], Shaanxi Province 2020 Natural Science Basic Research Plan [No. 2020JQ-042], and Aeronautical Science Foundation [No. 2019ZB070001].

References

[1] L. Cui, X. Wang, Y. Xu, H. Jiang, J. Zhou, A novel Switching Unscented Kalman Filter method for remaining useful life prediction of rolling bearing, Measurement 135 (2019) 678–684, https://doi.org/10.1016/j.measurement.2018.12.028.

[2] L. Song, H. Wang, P. Chen, Step-by-step fuzzy diagnosis method for equipment based on symptom extraction and trivalent logic fuzzy diagnosis theory, IEEE Trans. Fuzzy Syst. (2018), https://doi.org/10.1109/TFUZZ.2018.2833820.

[3] L. Cui, J. Huang, F. Zhang, F. Chu, HVSRMS localization formula and localization law: localization diagnosis of a ball bearing outer ring fault, Mech. Syst. Signal Process. (2019), https://doi.org/10.1016/j.ymssp.2018.09.043.

[4] Y. Li, K. Ding, G. He, X. Jiao, Non-stationary vibration feature extraction method based on sparse decomposition and order tracking for gearbox fault diagnosis, Measurement 124 (2018) 453–469. https://doi.org/10.1016/j.measurement.2018.04.063.

[5] X. Jiang, X. Cheng, J. Shi, W. Huang, C. Shen, Z. Zhu, A new l_0-norm embedded MED method for roller element bearing fault diagnosis at early stage of damage, Measurement 127 (2018) 414–424. https://doi.org/10.1016/j.measurement.2018.06.016.

[6] X. Ding, Q. Li, L. Lin, Q. He, Y. Shao, Fast time-frequency manifold learning and its reconstruction for transient feature extraction in rotating machinery fault diagnosis, Measurement 141 (2019) 380–395. https://doi.org/10.1016/j.measurement.2019.04.030.
[7] H. Wang, S. Li, L. Song, L. Cui, A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals, Comput. Ind. (2019), https://doi.org/10.1016/j.compind.2018.12.013.
[8] X. Guo, L. Chen, C. Shen, Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis, Measurement 93 (2016) 490–502. https://doi.org/10.1016/j.measurement.2016.07.054.
[9] A. Widodo, B.S. Yang, Support vector machine in machine condition monitoring and fault diagnosis, Mech. Syst. Sig. Process. (2007), https://doi.org/10.1016/j.ymssp.2006.12.007.
[10] G. Hu, H. Li, Y. Xia, L. Luo, A deep Boltzmann machine and multi-grained scanning forest ensemble collaborative method and its application to industrial fault diagnosis, Comput. Ind. (2018), https://doi.org/10.1016/j.compind.2018.04.002.
[11] X. Li, W. Zhang, Q. Ding, J.Q. Sun, Multi-Layer domain adaptation method for rolling bearing fault diagnosis, Signal Process. (2019), https://doi.org/10.1016/j.sigpro.2018.12.005.
[12] H. Shao, H. Jiang, H. Zhao, F. Wang, A novel deep autoencoder feature learning method for rotating machinery fault diagnosis, Mech. Syst. Signal Process. (2017), https://doi.org/10.1016/j.ymssp.2017.03.034.
[13] H. Liu, J. Zhou, Y. Zheng, W. Jiang, Y. Zhang, Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders, ISA Trans. (2018), https://doi.org/10.1016/j.isatra.2018.04.005.
[14] X. Li, W. Zhang, Q. Ding, Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks, IEEE Trans. Ind. Electron. (2019), https://doi.org/10.1109/TIE.2018.2868023.
[15] Z. Zhao et al., Unsupervised Deep Transfer Learning for Intelligent Fault Diagnosis: An Open Source and Comparative Study, Dec. 2019.
[16] Z. Zhao et al., Deep Learning Algorithms for Rotating Machinery Intelligent Diagnosis: An Open Source Benchmark Study, Mar. 2020.
[17] Z. Chen, K. Gryllias, W. Li, Intelligent fault diagnosis for rotary machinery using transferable convolutional neural network, IEEE Trans. Ind. Informatics 16 (1) (2020) 339–349, https://doi.org/10.1109/TII.2019.2917233.
[18] Z. Wu, H. Jiang, K. Zhao, X. Li, An adaptive deep transfer learning method for bearing fault diagnosis, Measurement 151 (2020) 107227, https://doi.org/10.1016/j.measurement.2019.107227.
[19] X. Li, H. Jiang, K. Zhao, R. Wang, A deep transfer nonnegativity-constraint sparse autoencoder for rolling bearing fault diagnosis with few labeled data, IEEE Access 7 (2019) 91216–91224, https://doi.org/10.1109/ACCESS.2019.2926234.
[20] 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, https://doi.org/10.1109/access.2019.2934233.
[21] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, D. Wierstra, Matching networks for one shot learning, Adv. Neural Inf. Process. Syst. (2016) 3637–3645.
[22] J. Nie, N. Xu, M. Zhou, G. Yan, Z. Wei, 3D Model classification based on few-shot learning, Neurocomputing (2019), https://doi.org/10.1016/j.neucom.2019.03.105.
[23] G.J. Aguiar, R.G. Mantovani, S.M. Mastelini, A.C.P.F.L. de Carvalho, G.F.C. Campos, S.B. Junior, A meta-learning approach for selecting image segmentation algorithm, Pattern Recogn. Lett. 128 (Dec. 2019) 480–487, https://doi.org/10.1016/j.patrec.2019.10.018.
[24] I. Olier et al., Meta-QSAR: a large-scale application of meta-learning to drug design and discovery, Mach. Learn. 107 (1) (Jan. 2018) 285–311, https://doi.org/10.1007/s10994-017-5685-x.
[25] 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 (January) (2020), https://doi.org/10.1016/j.ymssp.2019.106587.
[26] M. Blodt, P. Granjon, B. Raison, G. Rostaing, Models for bearing damage detection in induction motors using stator current monitoring, IEEE Trans. Ind. Electron. 55 (4) (2008) 1813–1822, https://doi.org/10.1109/TIE.2008.917108.
[27] S. Nandi, H.A. Toliyat, X. Li, Condition monitoring and fault diagnosis of electrical motors – a review, IEEE Trans. Energy Convers. 20 (4) (2005) 719–729, https://doi.org/10.1109/TEC.2005.847955.
[28] J. Zarei, J. Poshtan, An advanced Park's vectors approach for bearing fault detection, Tribol. Int. 42 (2) (2009) 213–219, https://doi.org/10.1016/j.triboint.2008.06.002.
[29] S.A. Niknam, T. Thomas, J. Wesley Hines, R. Sawhney, Analysis of acoustic emission data for bearings subject to unbalance, Int. J. Progn. Heal. Manag. 4 (Special issue 2) (2013).
[30] Y. Amirat, V. Choqueuse, M. Benbouzid, EEMD-based wind turbine bearing failure detection using the generator stator current homopolar component, Mech. Syst. Signal Process. 41 (1–2) (2013) 667–678, https://doi.org/10.1016/j.ymssp.2013.06.012.
[31] Y. Chen, G. Peng, C. Xie, W. Zhang, C. Li, S. Liu, ACDIN: bridging the gap between artificial and real bearing damages for bearing fault diagnosis, Neurocomputing 294 (2018) 61–71, https://doi.org/10.1016/j.neucom.2018.03.014.
[32] J. Shu, Z. Xu, D. Meng, Small Sample Learning in Big Data Era, 2018, pp. 1–76.
[33] E.D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, Q.V. Le, Autoaugment: learning augmentation strategies from data. In: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2019-June, no. Section 3, pp. 113–123, 2019. https://doi.org/10.1109/CVPR.2019.00020.
[34] N. Frosst, G. rey Hinton, Distilling a neural network into a soft decision tree, in: CEUR Workshop Proceedings, vol. 2071, 2018.
[35] C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in: 34th Int. Conf. Mach. Learn. ICML 2017, vol. 3, 2017, pp. 1856–1868.
[36] J. Snell, K. Swersky, R. Zemel, Prototypical networks for few-shot learning, in: Adv. Neural Inf. Process. Syst., vol. 2017-Decem, no. Nips, 2017, pp. 4078–4088.
[37] E. Triantafillou et al., Meta-Dataset: A Dataset of Datasets for Learning to Learn from Few Examples, 2019, pp. 1–22.
[38] Y. Guo, N.C.F. Codella, L. Karlinsky, J.R. Smith, T. Rosing, R. Feris, A New Benchmark for Evaluation of Cross-Domain Few-Shot Learning, 2019.
[39] Q. Huang, F. Zhang, X. Li, Few-shot decision tree for diagnosis of ultrasound breast tumor using BI-RADS features, Multimed. Tools Appl. 77 (22) (2018) 29905–29918, https://doi.org/10.1007/s11042-018-6026-1.
[40] A. Medela et al., Few shot learning in histopathological images: reducing the need of labeled data on biological datasets, in: Proc. – Int. Symp. Biomed. Imaging, vol. 2019-April, no. Isbi, 2019, pp. 1860–1864. https://doi.org/10.1109/ISBI.2019.8759182.
[41] M. Kim, J. Zuallaert, W. De Neve, Few-shot learning using a small-sized dataset of high-resolution FUNDUS images for glaucoma diagnosis, in: MMHealth 2017 – Proc. 2nd Int. Work. Multimed. Pers. Heal. Heal. Care, co-located with MM 2017, 2017, pp. 89–92. https://doi.org/10.1145/3132635.3132650.
[42] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359, https://doi.org/10.1109/TKDE.2009.191.
[43] R. Raina, A. Battle, H. Lee, B. Packer, A.Y. Ng, Self-taught learning: Transfer learning from unlabeled data, ACM Int. Conf. Proc. Ser. 227 (2007) 759–766, https://doi.org/10.1145/1273496.1273592.
[44] I. Sutskever, J. Martens, G. Dahl, G. Hinton, On the importance of initialization and momentum in deep learning, in: 30th International Conference on Machine Learning, ICML 2013, 2013, no. PART 3, pp. 2176–2184.
[45] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks?, Adv. Neural Inf. Process. Syst. 4 (January) (2014) 3320–3328.
[46] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North, 2018, pp. 4171–4186. https://doi.org/10.18653/v1/N19-1423.
[47] F. Sung, Y. Yang, L. Zhang, T. Xiang, P.H.S. Torr, T.M. Hospedales, Learning to compare: relation network for few-shot learning, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018, pp. 1199–1208, https://doi.org/10.1109/CVPR.2018.00131.
[48] C. Lessmeier, J.K. Kimotho, D. Zimmer, W. Sextro, Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification, in: Third European Conference of the Prognostics and Health Management Society 2016, 2016, no. Cm, pp. 152–156.