Task-Sequencing Meta Learning for Intelligent Few-Shot Fault Diagnosis With Limited Data

Yidan Hu, Ruonan Liu, *Member, IEEE*, Xianling Li, Dongyue Chen, and Qinghua Hu, *Senior Member, IEEE*

***Abstract*—Recently, deep learning-based intelligent fault diagnosis methods have been developed rapidly, which rely on massive data to train the diagnosis model. However, it is usually difficult to collect sufficient failure data in practical industrial production, thus limits the application of intelligent diagnosis methods. To address the few-shot fault diagnosis problem, a task-sequencing meta-learning method is proposed in this article. First, the meta-learning model is trained over a series of learning tasks to obtain knowledge about how to diagnosis. Thus, the learned knowledge can help adapt and generalize with a few examples when dealing with new tasks that have never been encountered. Then, considering the difference and connection between different failures and diagnosis tasks, a task-sequencing algorithm is proposed to sort meta training tasks from easy to difficult, which followed the way human acquire knowledge. After evaluating the difficulty of each task, the proposed method learns simple tasks first and generalizes the learned knowledge to complex tasks. Better knowledge adaptability is obtained by gradually increasing the task difficulty. Finally, utilizing gradient-based meta learning, the initialization parameters are trained by a small number of gradient steps. The effectiveness of the proposed method is validated by a practice rolling bearing dataset and a power system dataset. The experiment results illustrate that the proposed method can identify new categories with only several samples. In addition, it also shows advantages in fault diagnosis when the categories are fine-grained according to the working conditions. Therefore, the proposed method is suitable for solving the few-shot problem in practice and complicated fault diagnosis.**

***Index Terms*—Fault diagnosis, few-shot problem, initialization algorithm, meta learning, task-sequencing.**

I. Introduction

As AN effective tool to improve the reliable operation of equipment, fault diagnosis methods have been widely used in the industrial field and play a key role in the industrial Internet of Things (IIOT) [1]. In modern industrial manufacturing systems, data-driven intelligent fault diagnosis methods have been developed rapidly due to the development of sensor techniques and the accumulation of industrial big data. Through the analysis of the status data collected in IIOT environment, maintenance measures can be scheduled to ensure the safety in production. Specifically, fault identification is an important aspect in fault diagnosis.

In recent literature, deep learning-based intelligent fault diagnosis methods have led to a series of breakthroughs due to its attractive characteristic that directly learns the high-level and hierarchical representations from massive raw data [2], [3]. But in practical industry, the data available for training is insufficient for deep learning model training. On the one hand, machines operate under the normal condition in most of their life time while faults seldom happen. That causes that the monitoring data in the fault state is far less than the data obtained in the normal state. On the other hand, only a tiny fraction of monitoring data and the corresponding state is known. Most data need to be manually labeled, which increases the labor costs. What is more, when certain sudden catastrophic failures come, the system will be immediately shut down for maintenance. Therefore, it is hard to collect enough failure data to train the deep network. The scarcity of data often leads to overfitting problems [4] and obstructs the high-precision diagnostic model training. How to solve fault diagnosis problems with only a small amount of data [5] and obtain satisfied results need to be paid attention to. Therefore, few-shot fault diagnosis has become a critical problem to be tackled in modern industrial systems.

In the literature, the solutions to few-shot learning include data enhancement [6], synthetic data generation, transfer learning [7], self-supervised learning and meta learning [8]. After the application of these methods in other fields, including object detection [9], image segmentation [10], and image classification [11], there have been some successful examples of few-shot learning methods [12], [13] in fault diagnosis.

Sampling techniques and data generation techniques have been utilized mostly for fault data augmentation. Martin-Diaz *et al.* [14] adopt a supervised classification approach for induction motors faults based on the adaptive boosting algorithm with an optimized sampling technique, which aims at dealing with the imbalanced experimental dataset. Mathew *et al.* [15] propose a weighted kernel-based synthetic minority oversampling technique by oversampling in the feature space of support

Manuscript received May 14, 2021; revised August 7, 2021; accepted September 3, 2021. Date of publication September 14, 2021; date of current version February 18, 2022. Paper no. TII-21-2037. *(Corresponding author: Ruonan Liu.)*

Yidan Hu, Ruonan Liu, Dongyue Chen, and Qinghua Hu are with the College of Intelligence and Computing, Tianjin University, Tianjin 300072, China (e-mail: yidan_hu@tju.edu.cn; ruonan.liu@tju.edu.cn; dyue_chen@163.com; huqinghua@tju.edu.cn).

Xianling Li is with the Science and Technology on Thermal Energy and Power Laboratory, Wuhan 430205, China (e-mail: lixl_csic719@163.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2021.3112504.

Digital Object Identifier 10.1109/TII.2021.3112504

1551-3203 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on June 01,2022 at 09:45:51 UTC from IEEE Xplore. Restrictions apply.

vector machine (SVM) classifier that overcomes the limitation for nonlinear problems. There are also ways to change the model architecture directly. The model based on a stacked sparse autoencoder [16] is utilized to deal with limited sample data. The bidirectional gated recurrent unit [17] using cost sensitive active learning is developed for fault diagnosis. The siamese neural network [18] learns by exploiting sample pairs of the same or different categories to solve the few-shot problem. In addition, transfer learning is used to solve few-shot problems. Zhong *et al.* [19] reuse the internal layers of CNN trained on the normal dataset to extract the feature representations for fault dataset, and apply SVM to fault diagnosis. Wu *et al.* [20] construct few-shot transfer learning method based on a unified 1D convolution network for few-shot diagnosis. Two transfer situations, named conditions transfer and artificial-to-natural transfer, are considered.

Nevertheless, the whole network of the abovementioned methods needs to be trained from scratch to ensure the effect when facing new tasks, which limits their adaptability in practical industrial production. What is more, in the practical engineering, the changeable operational conditions will also decrease the performance of the diagnosis methods under new conditions. Therefore, the identification of the fault categories under different working conditions also needs to be paid attention to.

As a highly adaptable learning strategy for few-shot problem, meta learning focuses on how to acquire learning ability instead of learning itself [21]–[23]. With the help of learning ability, only simple adjustments are needed to adapt to new tasks in practical industrial scenarios. Specifically, meta learning does not directly learn a mathematical model used for prediction, but learns how to learn a generalized mathematical model. If the feature extraction is regarded as the process of learning directly from data, the meta learner obtains the learning experience by evaluating the process and the target task can be completed via a few samples. As an optimization-based meta-learning method [24]–[26], model-agnostic meta-learning (MAML) [27] hopes to find an initial parameter set that is sensitive to new tasks, so that the model can improve the performance after gradient update on a small amount of data of new tasks.

In industrial systems, the data used for fault diagnosis is often a complex time-series signal that is collected from the equipment under different working conditions. When fault samples are insufficient, different working conditions will cause interference to identify complex fault information. This requires meta learning to have better knowledge adaptability. Therefore, to solve the problem of few-shot fault identification and enhance the adaptability under different working conditions, a task-sequencing meta-learning (TSML) approach is proposed in this article.

The main contributions of this article are summarized as follows.

1) An intelligent fault diagnosis approach, TSML, is proposed to avoid overfitting in few-shot scenario. By finding sensitive initialization parameters with robust knowledge adaptability, TSML shows advantages in dealing with few-show problems under variable operational conditions.

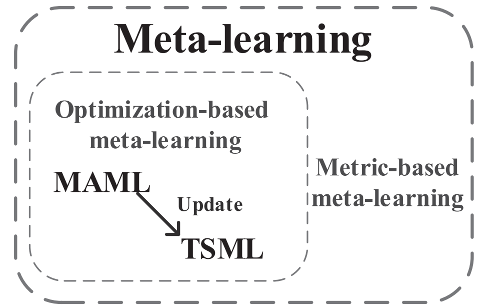


Fig. 1. Relationship between meta learning, MAML, TSML.

2) To increasing the stability of the diagnosis performance under the variable operational conditions, a task-sequencing strategy is proposed to adapt to new categories quickly and perform well in different working conditions with the learning experience. Via designing a curriculum from easy to difficult for task learning based on curriculum learning, the tasks in meta learning are arranged in order. Then, the adaptation from task to task can be more stable through the stepped learning process. Therefore, a more general knowledge representation can be learned for better convergence within few samples.

3) A rolling bearing fault dataset and a power system fault dataset are analyzed to verify the effectiveness of the proposed method. The experimental results show that the TSML can adapt to the few-shot scenario and fine-grained working conditions. The effectiveness and superiority of the proposed method are verified by the comparison with state-of-the-art methods.

The rest of this article is organized as follows. The proposed TSML method is introduced in Section II. Two case studies are conducted in Section III to verify the effectiveness of the proposed TSML. Section IV concludes this article.

II. Proposed Method

This section presents the proposed meta-learning fault diagnosis methods. First, the concept and the task-based settings of meta learning is explained to help better understanding. Then, MAML algorithm which is an optimization-based meta-learning method is presented. Moreover, the proposed TSML algorithm, which updates MAML is detailed. The relationship between the three part is introduced in Fig. 1.

A. Meta-Learning Problem

Deep learning requires massive data to train a good model. In contrast, human learn new concepts in a more efficient way. For example, a kid can distinguish cats from dogs after seeing them a few times. Meta learning expects to design such a model that can learn new skills with a few training examples. It is hoped that the machine can observe how machine learning approaches perform on tasks [28], and then learn the experience. Finally, it can use past experience to solve problems on new tasks instead

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on June 01,2022 at 09:45:51 UTC from IEEE Xplore. Restrictions apply.

of learning from scratch. That is why meta learning aims at learning to learn [29].

In machine learning, suppose the dataset of the classification task is $\mathcal{D} = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, where x is the input sample and y is the sample label. We need to optimize the parameter θ to predict a model $y = f_\theta(x)$

$$\theta^* = \arg\min_\theta \mathcal{L}(\mathcal{D}; \theta, \omega) \tag{1}$$

where $\mathcal{L}$ is loss function, ω is learning strategy. Similar to machine learning, whose inputs are labeled samples, the inputs of meta learning are a series of tasks $D_{\text{train}} = \{(\mathcal{D}_1^{\text{train}}, \mathcal{D}_1^{\text{test}}), \ldots, (\mathcal{D}_{\text{n}}^{\text{train}}, \mathcal{D}_{\text{n}}^{\text{test}})\}$ and $D_{\text{test}} = \{(\mathcal{D}_1^{\text{train}}, \mathcal{D}_1^{\text{test}}), \ldots, (\mathcal{D}_{\text{m}}^{\text{train}}, \mathcal{D}_{\text{m}}^{\text{test}})\}$. These tasks are divided into training tasks and test tasks. On these tasks, we need to predict a model

$$\min_\omega \mathop{\mathbb{E}}_{\mathcal{T} \sim p(\mathcal{T})} \mathcal{L}(\mathcal{D}; \omega) \tag{2}$$

where all tasks are in $p(\mathcal{T})$ and ω is meta knowledge. By learning a general knowledge representation ω, the loss of different tasks can all be small enough.

In the meta-learning process, a two-layer learning process is constructed. The meta-level outer layer learns the general knowledge representation, which continues to evolve through learning from one task to another. The task-level inner layer updates the model like traditional machine learning, which concentrates on only one task with training and testing. Whether it is a training task or a testing task, each task utilizes the meta learner to perform a learning process, which uses training data $\mathcal{D}^{\text{train}}$ and testing data $\mathcal{D}^{\text{test}}$. To distinguish the data D used for the whole learning process and data $\mathcal{D}$ used for meta task, the training data of meta task is called *support set*, and the test data of meta task is called *query set*. Each meta task trains with a support set and tests with a query set. The meta learner needs to classify samples into N categories. In addition, N-way represents that a meta-classification task needs to classify N categories, and K-shot represents the number of support set of each category, so each meta task has $N \cdot K$ samples for training. Because randomly selecting samples to form tasks, data in support set of one task and data in query set of another task may be the same. The repetition of samples will not affect the generalization ability of the model, because meta learning is a fast-fitting process through training a large number of tasks. As long as there are specific differences in the training samples of each task, there is no need for each task to be completely independent.

B. Model-Agnostic Meta-Learning

Uphold the idea of meta learning, MAML algorithm [27] expects to train the initialization parameters of the model in the outer layer. With initialization parameters that are universal enough, the model can obtain a good performance efficiently after only a few steps of gradient update when focused on the new tasks with a few data. From the perspective of representation learning, the MAML algorithm tries to find general internal representations that are easier to transfer to other tasks. Thus, only fine-tuning of parameters is required when dealing with new tasks. In order to find a predominant general internal representation, the sensitivity of the parameters when facing with new tasks needs to be maximized, so that small changes in the parameters can lead to a considerable increase in task loss.

MAML trains the parameters explicitly, and the model is represented by a function f_θ, which is determined by the parameter θ. The whole model is divided into the inner loop and the outer loop, and θ is shared by both of them. The inner loop calculates the loss function of the subtask, and then updates the parameters of the new task. That is, the parameter θ of the model will be updated as θ_i' after gradient descent

$$\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta) \tag{3}$$

where α is the inner learning rate.

The outer loop is called meta optimization. According to the optimized parameter θ_i' of the inner loop, the loss can be recalculated on the new task, and the gradient of the initial parameter θ can be calculated and updated

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}\left(f_{\theta_i'}\right) \tag{4}$$

where β is the learning rate. The optimal parameters of the meta-learning model for the task distribution $p(\mathcal{T})$ can be obtained by the alternating optimization of the inner and outer loops. Thus, the goal of meta optimization is to minimum the loss function of the task

$$\min_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}\left(f_{\theta_i'}\right) = \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}\left(f_{\theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)}\right). \tag{5}$$

C. Task-Sequencing Meta-Learning

A good initialization parameter representation, which plays a key role in optimization-based meta learning [30], can be obtained for each fault diagnosis task by learning multiple classification tasks [31]. During this process, the samples of each task are randomly selected, and the order of tasks is not taken into account. The selection and arrangement of this random mode will bring great randomness. What is more, the categories in some tasks may have a direct correlation, while the categories in other tasks can be scarcely relevant. The uncertainty and diversity among tasks [32], [33] can lead to different difficulty of tasks. To thoroughly learn good parameter representation and obtain more superior generalization performance when classifying new categories, the learning order of tasks is measured to get a better model because the previous tasks will inevitably affect the later tasks. Considering the human learning process generally follows the order from easy to difficult, curriculum learning [34] advocates that the model should start learning from easy samples and gradually transition to complex samples. In nonconvex problems, with judgment on the difficulty of samples [35], this learning strategy brings a vast performance improvement by accelerating the convergence and finding a better local optimum. Moreover, curriculum learning finds that when a model learns the simple knowledge more sufficiently, it can obtain a better generalization performance.

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on June 01,2022 at 09:45:51 UTC from IEEE Xplore. Restrictions apply.

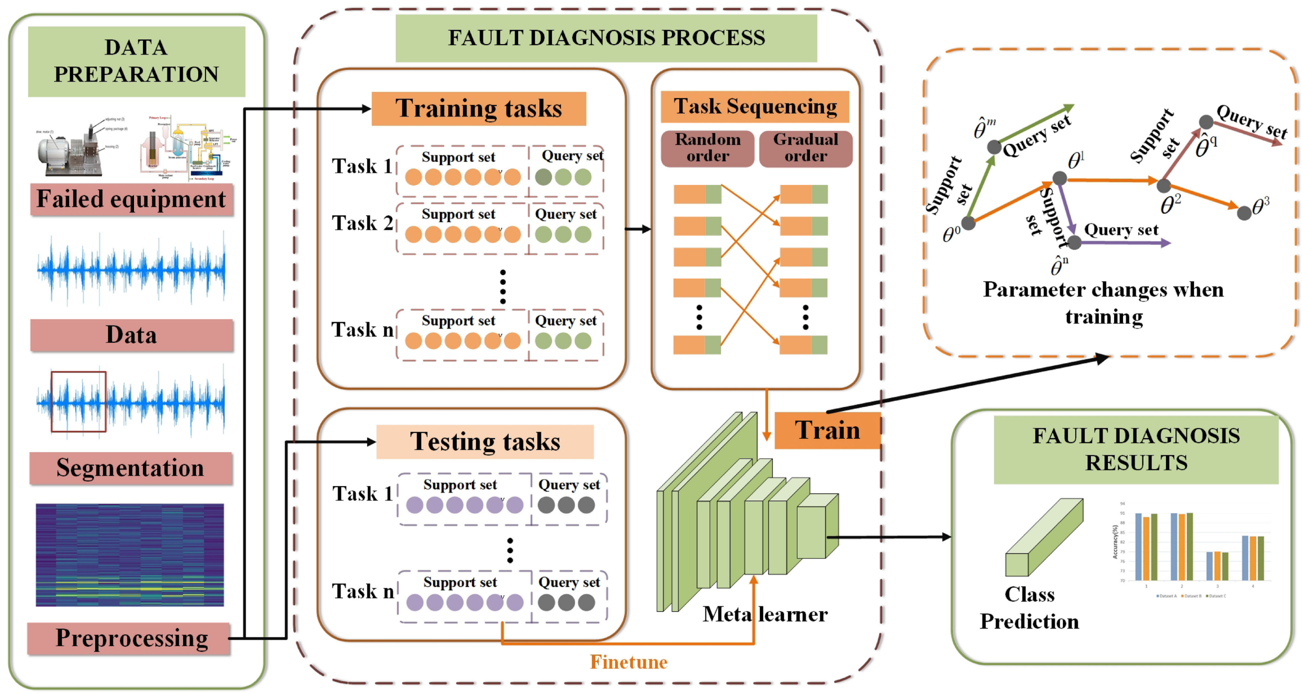


Fig. 2. Basic idea of the TSML for intelligent few-shot fault diagnosis.

Like the curriculum learning which learns samples from easy to difficult, TSML also adopts the way from easy to difficult to learn each task. Because learning on different tasks will affect the same initialization parameters θ, the different interactions between tasks will lead to different learning results. Therefore, the order of tasks will have an impact on how well the model learns. In an orderly way, the trained general knowledge representation of the outer layer can benefit the later learning of the more difficult task. To sort the tasks mentioned above, it is necessary to evaluate what kind of task is simple and what kind of task is difficult. In curriculum learning, the definition of curriculum is open. Different evaluation criteria can be set for different problems. In this article, TSML firstly implements the sampling of each task and conducts cluster analysis in each task. Then, the clustering accuracy can be evaluated, and the task with a high score in the clustering task is defined as a simple task. As an unsupervised learning method, clustering is often more difficult than supervised learning without the help of labels. The task that gets better effect in unsupervised clustering is easier to classify supervisedly. Therefore, the clustering result can be used as a preliminary index to judge the task's difficulty.

K-means clustering is one of the most commonly used clustering algorithms based on Euclidean distance. The closer samples have a greater similarity. Unlike the fact that the neural network with tens of thousands of parameters is prone to overfitting in few-shot learning, the nonparametric method does not need to optimize the parameters. Therefore, the K-means clustering is a reasonable method when samples are scarce. In this article, the cluster centroids of N clusters are randomly selected as $\mu_1, \mu_2, \ldots, \mu_N \in \mathbb{R}^n$, where N represents N categories in the N-way K-shot classification task. For each sample $\mathbf{x}^{(m)}$, the category $c^{(m)}$, which is the closest one between sample $\mathbf{x}^{(m)}$ and all categories can be calculated as

$$c^{(m)} = \arg\min_{n} \left\| \mathbf{x}^{(m)} - \mu_n \right\|^2. \tag{6}$$

For each category n, calculate the cluster centroid μ_n, where the distance from sample m to $c^{(m)}$ is the shortest

$$\mu_n = \frac{\sum_{m=1}^{k} 1\left\{c^{(m)} = n\right\} \mathbf{x}^{(m)}}{\sum_{m=1}^{k} 1\left\{c^{(m)} = n\right\}}. \tag{7}$$

The abovementioned two formulas need to be repeated until converging, i.e., centroid locations are stable. In each task, the K-means clustering method is used to model the distance distribution between examples. The clustering accuracy of each task is used as the evaluation index. After sorting, the task is learned in order for the update process of the inner layer and outer layer. In the classification task, the cross-entropy is used as the loss function uses, where $\mathbf{x}^{(j)}, \mathbf{y}^{(j)}$ are an input/output pair sampled from task $\mathcal{T}_i$, f_ϕ represents the model, and the parameters ϕ are organized according to (3) and (4)

$$\mathcal{L}_{\mathcal{T}_i}(f_\phi) = \sum_{\mathbf{x}^{(j)},\mathbf{y}^{(j)}} \mathbf{y}^{(j)} \log f_\phi\left(\mathbf{x}^{(j)}\right) + \left(1 - \mathbf{y}^{(j)}\right) \log\left(1 - f_\phi\left(\mathbf{x}^{(j)}\right)\right). \tag{8}$$

Fig. 2 shows the framework of applying the TSML model to industrial process fault diagnosis. Moreover, as shown in Algorithm 1, the learning process of TSML is as follows.

1) First, the tasks in meta learning are sampled. There are NK samples in a N-way K-shot classification task,

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on June 01,2022 at 09:45:51 UTC from IEEE Xplore. Restrictions apply.

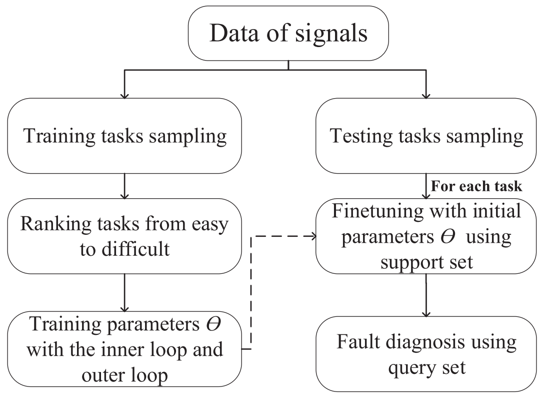


Fig. 3. Flowchart of the TSML for intelligent few-shot diagnosis.

Algorithm 1: TSML.

Require: $p(\mathcal{T})$:distribution over tasks
Require: α, β:step size hyperparameters
random initialize θ
while not done **do**
  Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$
  for all $\mathcal{T}_i$ **do**
    Sample NK datapoints $\mathcal{D}=\{\mathbf{x}^{(j)}, \mathbf{y}^{(j)}\}$ from $\mathcal{T}_i$
    Randomly select N cluster centroids:
    $\mu_1, \mu_2, \mu_3, \ldots, \mu_N \in \mathbb{R}^n$
    Compute K-means accuracy $\mathcal{A}_{\mathcal{T}_i}$ in Eqn.(x) and
    (x)
    Sample datapoints $\mathcal{D}_i' = \{\mathbf{x}^{(j)}, \mathbf{y}^{(j)}\}$ from $\mathcal{T}_i$ as
    new task
  end for
  sort $\mathcal{T}_i$ according to $\mathcal{A}_{\mathcal{T}_i}$
  for all $\mathcal{T}_i$ **do**
    Evaluate $\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$ using $\mathcal{D}$ and $\mathcal{L}_{\mathcal{T}_i}$ in Eqn.(x)
    Compute parameters with gradient descent:
    $\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$
  end for
  Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i'})$
end while

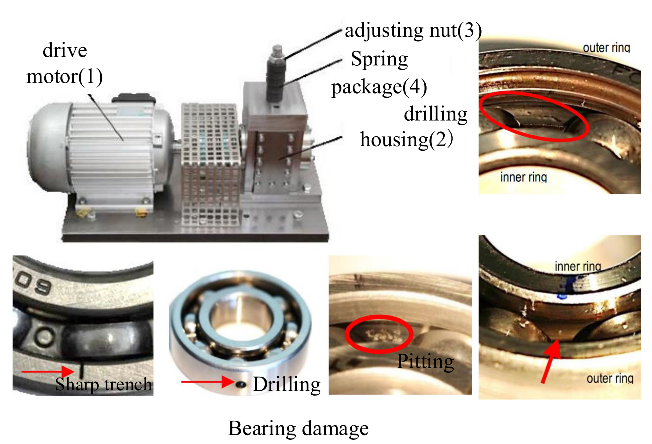


Fig. 4. Bearing damages and apparatus for accelerated life time test.

and each sample X corresponds to a label Y: $\mathcal{D} = \{\mathbf{x}^{(j)}, \mathbf{y}^{(j)}\}$.

2) K-means clustering is performed on the support set in each task. It is an uncomplicated algorithm and will not bring much computational burden. The clustering accuracy is obtained by comparing the clustering results with the actual label Y.
3) For each task, the clustering accuracy can be regarded as the task's score to help rank tasks from easy to difficult.
4) Cross entropy loss is set to be the loss function in each classification task. The parameters of the task in the inner loop are optimized according to (3). The optimizer is ADAM, which has the ability of AdaGrad to deal with sparse gradients and the ability of RMSProp to deal with nonstationary objectives.
5) After the model parameters of tasks in a batch in the inner loop are optimized, the parameters of the outer loop can be updated according to (4).
6) Repeat steps (4) and (5) to get the optimal parameter initialization for task distribution $p(\mathcal{T})$ of the TSML model. The flowchart is shown in Fig. 3.

III. Case Study

To comprehensively evaluate the proposed TSML method, it is used to analyze a bearing fault dataset and a system failure dataset. Experiments on these datasets are conducted to verify the effectiveness in a few-shot scenario and fine-grained working conditions, which shows that it has a high accuracy in few-shot fault diagnosis. It also achieves a good performance in fault diagnosis scene migration.

A. PU Bearings Fault Dataset

1) Datasets Description: PU bearing dataset [36] contributed by Paderborn University, is a fault diagnosis dataset of rolling bearings. It collects current signals and vibration signals of rolling bearings, which are artificially damaged, naturally damaged, and in a healthy state. The artificial damages are manually caused by electric discharge machining, drilling, and manual electric engraving. The real bearing damages are generated by accelerated lifetime tests, which are shown in Fig. 4. Fig. 4 also shows some faulty bearings, and the faults of these bearings will be diagnosed by the proposed method. Both artificial faults and real faults exist at the inner and outer ring, and the extent of damages has two levels. There are 32 types of different faults in total. All faults of bearings are tested under four different working conditions. To imitate the complex working conditions in the practical industry and accurately obtain fine-grained diagnosis results, the faults under different working conditions are regarded as different categories, which means that 128 fine-grained categories are counted. A total of 100 categories are

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on June 01,2022 at 09:45:51 UTC from IEEE Xplore. Restrictions apply.

TABLE I
OPERATING PARAMETERS

No.	Rotational Speed [rpm]	Load Torque [Nm]	Radial Force [N]
0	1500	0.7	1000
1	900	0.7	1000
2	1500	0.1	1000
3	1500	0.7	400

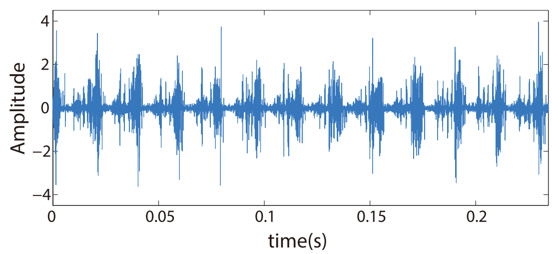


Fig. 5. Visualization of vibration signal from PU dataset.

TABLE II
PARAMETERS OF EACH MODULE

Layer Number	Module1	Module2
L1	Conv(3×3×32)	Conv(3×3×32)
L2	Relu	Relu
L3	BN	BN
L4	Maxpool(2×2)	-

TABLE III
STRUCTURE OF THE BASE LEARNER

Module Number	PU Dataset	Farop Dataset
1	Module 1	Module 2
2	Module 1	Module 1
3	Module 1	Module 2
4	Module 1	Module 1
5	FC	FC

randomly selected for training, while the other 28 categories are utilized for testing. The detailed working conditions are shown in Table I. The visualization of the vibration signal is shown in Fig. 5. Besides vibration analysis, acoustic analysis is the signal processing method based on acoustic emission (AE) signals. The main challenges of acoustic analysis are as follows: first, the fault signals are weak under slow rotation speed conditions; second, the AE signal denoising is difficult when compared with the vibration analysis. This article mainly analyzes the vibration signals.

In the following experiments, vibration signals of time series are utilized. To obtain more information in the few-shot scenario, the time-variant nonstationary signal is processed with time-frequency analysis before put into the network. That can help acquire joint distribution information of the time domain and the frequency domain. Fast Fourier transform (FFT) is adopted to get better locality analysis in both time domain and frequency domain, which uses a time window to slide in the time direction to perform Fourier transform. And Hanning window function is utilized. After FFT, each category of vibration signal generates 80 time-frequency maps.

2) Experimental Setup: The N-way K-shot experimental protocol is a general experimental setting for the few-shot fault diagnosis. It is set as follows: When constructing a classification task, N categories are selected randomly first. For each category, K examples and Q examples are randomly sampling as support set and query set, respectively. Following the abovementioned settings of a task, the entire experimental process samples 10 000 training tasks and 100 testing tasks. The whole learning process has two layers, the task-level inner layer and the meta-level outer layer. The task-level inner update learning α rate and meta-lever outer learning rate β are 0.01 and 0.001, respectively. The task-level inner update steps are set as 5, and the update steps for finetuning are set as 10.

All tasks are learned using the same base network. And each task will contribute to the initialize parameters of this learner. In the following experiments, the architecture of the base learner has four modules. The first layer in each module is a 3×3 convolutions with 32 filters. Then, the following part consists of RELU nonlinearity, batch normalization, and 2×2 max-pooling. For all base learners, the last layer of the whole structure is fed into a softmax function. The loss function, we used is a cross-entropy error between the true and predicted category. The parameters in each layer of each module are presented in Table II, where RELU is the Rectified Linear Units activation function, and BN is batch normalization. The whole architecture of the inner loop is presented in Table III, where FC represents the fully connected layers.

3) Compared With Some State-of-the-Art Network for Few-Shot Learning: In practical fault diagnosis, there have been cases of using transfer learning for the few-shot analysis. To compare with the transfer learning methods, two deep neural networks with excellent learning capabilities, i.e., VGG-11 [37] and Resnet-18 [38], were selected as backbone networks for knowledge transfer. In the process of knowledge transfer, the fault data of the source domain is used to pretrain the basic network, and then the fault data of the target domain is used to fine-tune the network. After the preliminary experiment, we found that fine-tuning the whole network can get better accuracy compared with only fine-tuning the classifier. Therefore, the experimental results of transfer learning are all based on the fine-tuning of the entire network. What is more, the N-way K-shot setting is used for these comparative methods for a fair comparison.

The experimental results of few-shot fault diagnosis are shown in Table IV. The accuracy of TSML under different tasks are 90.97%, 91.05%, 78.96%, 84.03%, which is significantly higher than the transfer learning methods. It can be seen that when there are fewer training samples, the performance improvement of TSML is more obvious. It proves the initial aims

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on June 01,2022 at 09:45:51 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
FAULT CLASSIFICATION RESULTS OF THE THREE METHODS

Methods	5-way Accuracy 5-shot	5-way Accuracy 6-shot	10-way Accuracy 5-shot	10-way Accuracy 6shot
Transfer VGG-11	79.2	86.8	71.1	74.78
Transfer Resnet-18	82.55	87.85	71.61	76.02
TSML	90.97	91.05	78.96	84.03

TABLE V
FAULT CLASSIFICATION RESULTS OF TSML AND MAML ON PU DATASET

Methods	5-way 5-shot	5-way 6-shot	10-way 5-shot	10-way 6-shot
MAML Accuracy	89.65	90.9	74.17	83.25
TSML Accuracy	90.97	91.05	78.96	84.03
MAML F1	89.06	89.78	75.21	83.44
TSML F1	90.77	90.95	76.48	83.12
MAML precision	87.9	88.92	71.39	80.60
TSML precision	89.26	90.23	75.22	82.56

of the TSML method: it can diagnose and classify faults better in the few-shot scenarios. Another significant performance improvement occurs when a task needs to classify more categories. Compared with 5-way tasks, TSML gets more significant accuracy improvement on 10-way tasks, which directly illustrates the practicability of the proposed method when dealing with more complex industrial situations. The reason why TSML is better than transfer learning is that the initialization parameters learned in the pretraining stage in transfer learning are directly updated. The original intention of transfer learning is to minimize the sum of losses of all tasks, which only focuses on the performance of the current task but cannot guarantee that all tasks can be trained well. In contrast, the parameters updating process of the meta learning has two gradient descent update steps. As shown in Fig. 2, the direction of the second one is the final direction of the initialization parameter update. After getting the parameters $\widehat{\theta}^m$ specific to task m, the loss of task m is computed based $\widehat{\theta}^m$, and the gradient of the loss function on the parameters $\widehat{\theta}^m$ is calculated. Then, the initialization parameters θ^0 will be optimized by the gradient. In this way, TSML does not care about the current task but aims to find more potential initialization parameters that can quickly adapt to new tasks. Therefore, TSML can find more general knowledge representation among tasks by a few gradient descent update steps.

4) Effectiveness of the Task-Sequencing: Task sequencing is to further increase the adaptability of knowledge through a more orderly and reasonable way of task learning. To evaluate the effectiveness of task sequencing, TSML is compared with the baseline algorithm MAML without task sequencing. The inner network architecture of TSML is the same as that of MAML, and their training strategy and data set partition are also consistent. The only difference between the two methods is whether there is a consideration of task sequencing in the outer layer. The results of fault diagnosis are shown in Table V. The number of categories N and the number of samples in each category K are changed to evaluate the effectiveness in different task settings. It can be seen that the classification accuracy of TSML is better than that of MAML under different task settings. Significantly, in the 10 way 5-shot task, the accuracy of TSML is improved more obviously, which shows its superiority in the task with more categories and fewer samples. When the number of task categories is large and the number of samples is small, the task will be more challenging to learn.

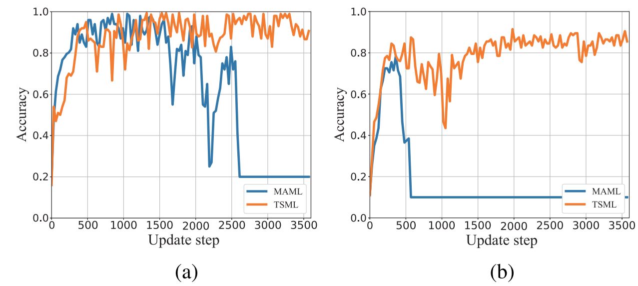


Fig. 6. Accuracy in the outer loop, (a) 5-way 6-shot and (b) 10-way 10-shot.

Besides accuracy, precision, and $F1$ score were also introduced for a more comprehensive evaluation. The precision is the ratio of true positives to the sum of true positives and false positives. It measures the proportion of examples classified as positive that are actually positive. $F1$ score can be regarded as a weighted average of model precision and recall. It takes into account both the precision and recall of the classification model. In these two indicators, TSML is also superior to the baseline method. Except in the case of 10-way 6-Shot, the performance of TSML is slightly inferior, which may be due to the increase in the number of samples in each task. The $F1$ score and accuracy of TSML are still excellent when dealing with 10way 6-shot tasks. The performance of TSML reflects its adaptability in complex few-shot scenarios. TSML improves the performance because ranking tasks from easy to difficult is more in line with the acquisition process of meta knowledge. Standardizing the learning process of meta learning will bring better learning outcomes, such as making an orderly task learning step. This step is conducive to the general knowledge representation learned from the previous tasks to adapt better to the later tasks, which helps the model obtain a better generalization performance.

In addition, in the training process, MAML was more prone to degenerate when tuning hyperparameters. After training to a certain extent, the accuracy rate saturates and then drops rapidly, but the optimal training effect is not achieved. Under the same hyperparameters settings, the training process of TSML is stable until the end of training. The training processes are proposed in Fig. 6. Because the training process of TSML is stair-likely progressive, there will be no huge differences in the difficulty of adjacent tasks, and the task learned first is not easy to bring greater adverse effects to the subsequent tasks.

5) Effectiveness in Different Scenario Settings: To identify the faults in few-shot practical industrial scenarios, it is not only

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on June 01,2022 at 09:45:51 UTC from IEEE Xplore. Restrictions apply.

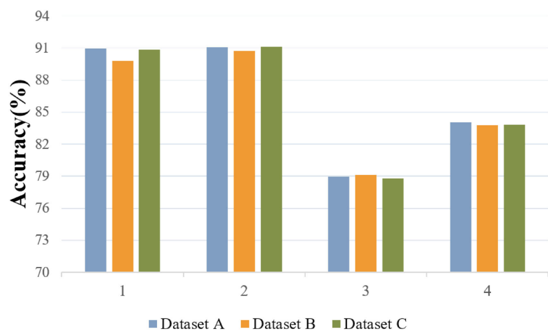


Fig. 7. Visualization of the experiment results in different scenarios.

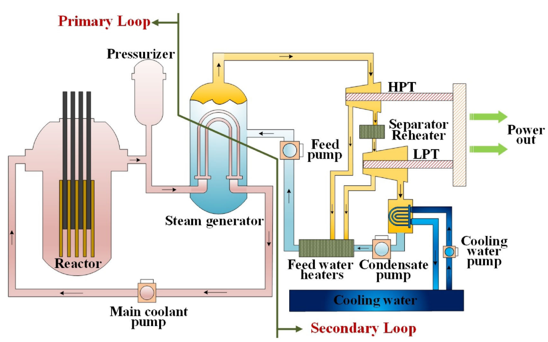


Fig. 8. Simplified schematic of a typical power system.

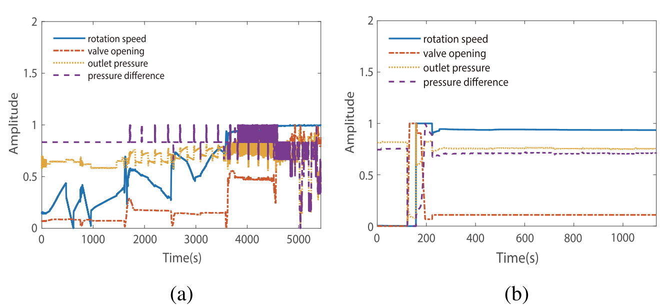


Fig. 9. Visualization of the partial signals of the Farop dataset.

necessary to overcome the problem of insufficient fault samples, but also need to have scene migration adaptability. When the training set and the testing set are of different categories, and the categories are subdivided according to the working conditions, TSML has achieved high accuracy. This reflects the adaptability of TSML to different failure scenarios under different working conditions. However, since the categories of the training set and the testing set are randomly divided, it is impossible to judge whether the high accuracy comes from that the faults in the testing set are easy to identify. Therefore, to eliminate the influence of the randomness of the fault categories, we randomly selected the fault categories in the training set and the testing set many times. Then, experiments are conducted under these different data divisions. The experimental results are shown in Table VI and Fig. 7. It can be seen that TSML can get good results under different divisions, which proves that the learning ability of TSML is stable, and the general knowledge learned from some faults can be used to express the effect on other faults.

TABLE VI
FAULT CLASSIFICATION RESULTS IN DIFFERENT SCENARIOS

Dataset	5-way Accuracy 5-shot	5-way Accuracy 6-shot	10-way Accuracy 5-shot	10-way Accuracy 6shot
Dataset A	90.97	91.05	78.96	84.03
Dataset B	89.82	90.74	79.11	83.79
Dataset C	90.85	91.1	78.77	83.82

B. Farop Dataset

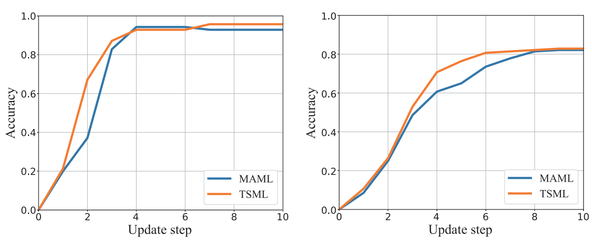
1) Datasets Description: The Farop dataset, which is used for FAult Recognition Of Power system, consists of a large number of operational data with large-scale fault categories and senor measurements. As shown in Fig. 8, the power system consists of the following main parts: reactor, steam generator, coolant pump, pressurizer, feedwater heaters, feed pump, condensate pump, turbines, and reheater. The original simulation data, collected during the simulation process of six different working conditions, has 121 sensor measurements including pressure, flow rate, temperature, and so on. The whole Farop dataset contains 252 860 samples of 66 categories, of which 53 484 are healthy samples and 199 376 are faulty samples. While collecting fault data, failures are introduced in all of the simulation processes of power system, and the faults of main components such as the feedwater pump, circulating pump, condenser, and relevant valves were simulated to comprehensively analyze the fault situation. We segment these samples according to time sequence, and segmentation with 20 s of information is considered as an input sample. Therefore, the whole dataset has 12 643 samples. In the training process, we randomly selected 46 categories of data are randomly selected as the training set, the other 20 categories as the testing set, and all samples were randomly scrambled. The visualization of some sensor signals is shown in Fig. 9.

2) Experimental Setup: The N-way K-shot experimental protocol of few-shot learning is still utilized. The entire experimental process samples 10 000 training tasks and 100 testing tasks. The task-level inner update learning α rate and meta-lever outer learning rate β are 0.01 and 0.001, respectively. The task-level inner update steps are set as 5, and the update steps for finetuning are set as 10. Like the base learner of PU, the base learner of Farop still has four modules. However, since the data of Farop is more straightforward than the image data of PU, certain changes have been made to the structure of the neural network. The 3 input channels are replaced by 1 input channel, and the 2×2 max-pooling layer in the first and third module are removed. Following the four modules is a fully connected layer, and the last layer is fed into softmax. The loss function is the cross-entropy function. The settings of each layer in the module are presented in Table II, and the whole architecture of the inner loop is shown in Table III.

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on June 01,2022 at 09:45:51 UTC from IEEE Xplore. Restrictions apply.

TABLE VII
FAULT CLASSIFICATION RESULTS OF TSML AND MAML ON FAROP DATASET

Methods	5-way 5-shot	5-way 6-shot	10-way 5-shot	10-way 6shot
MAML Accuracy	85.3	86.13	74.66	76.07
TSML Accuracy	87.3	87.65	76.4	77.1
MAML Precision	88.19	86.25	75.68	77.81
TSML Precision	89.2	88.23	77.24	77.34
MAML F1	82.12	83.45	73.6	74.72
TSML F1	83.51	84.01	74.2	75.25



(a) (b)

Fig. 10. Accuracy of a task in the inner loop, (a) 5-way 6-shot and (b) 10-way 6-shot.

3) Effectiveness of the Task-Sequencing: To verify the effectiveness of task sequencing, TSML with task-sequencing is compared with the baseline algorithm MAML. The inner network architecture, the dataset's partition of the two algorithms, and the training strategy are consistent. The results of fault diagnosis are shown in Table VII. Under four different task settings, the accuracy rates of the proposed TSML are 87.3%, 87.65%, 76.4%, 77.1%, respectively. It can be seen that the meta-learning strategy with task sequencing can obtain better results than the one without sequencing. Moreover, when the number of samples in the support set is smaller, the accuracy improves more obviously. It verifies the effectiveness of TSML in few-shot scenarios. Besides accuracy, precision, and $F1$ score were also introduced as evaluation indicators. As shown in Table VII, the $F1$ score and precision of TSML are also excellent. To further explain the training process of TSML, Fig. 10 shows how the accuracy in a task changes as the update step increases. To optimize the efficiency of the experiment, the task-level update step in our experiments is set to 5. But here to show more intuitively, the task-level update step in Fig. 10 is 10. It can be seen that the accuracy of the proposed TSML method is improved more obviously and quickly.

4) Effectiveness in Different Scenario Settings: The above-mentioned experimental results are all obtained on the basis that testing categories and training categories are different. It shows that after learning, TSML can identify faults by using initialization parameters and a few steps of adjustment when faced with new categories. The excellent adaptation to the new categories reflects the credibility of the method in practical industrial scenarios. What is more, to eliminate the impact of random division of data categories, we also conduct three different random partitions on the Farop dataset. And experiments are conducted under these divisions. The experimental results are shown in Table VIII, which demonstrates the good adaptability of TSML on the Farop dataset. Under the three different divisions, the results of the proposed method are balanced, which reflects that no matter how to divide the fault categories kind of fault classification is made, the learning ability of TSML can be steadily exerted.

TABLE VIII
FAULT CLASSIFICATION RESULTS IN DIFFERENT SCENARIOS ON FAROP DATASET

Dataset	5-way Accuracy 5-shot	5-way Accuracy 6-shot	10-way Accuracy 5-shot	10-way Accuracy 6shot
Dataset A	87.3	87.65	76.4	77.1
Dataset B	87.11	87.53	76.25	76.97
Dataset C	87.35	87.48	76.31	77.18

IV. CONCLUSION

In this article, a novel meta-learning-based approach was proposed for the few-shot fault diagnosis, which is named TSML. The proposed method can be effective in the few-shot scenario of fault diagnosis, and has strong work condition transfer adaptability to adapt to practical industrial. It constructed a meta-learning strategy by fixing tasks in an orderly arrangement and finding suitable initialization parameters by a small number of gradient steps. The practicality is verified with a publicly available rolling bearing dataset and a power system dataset. The experiment results prove that it has high accuracy in the few-shot fault diagnosis scene. And because its ability to learn to learn, the proposed method could adapt to new categories using a small number of gradient steps. What is more, through a more orderly and reasonable way of task learning, the proposed method obtains more robust knowledge adaptability. It can help the model be sensitive enough to identify faults under fine-grained working conditions.

Considering that the adaptation of cross-condition has been verified, meta learning may also deal with cross-modal tasks. Meta-knowledge can be acquired from tasks in one modality and be applied to new tasks in another modality. And there are many works that can be developed and improved upon in future study. For example, first, the convolution network was utilized as the base learner to extract meta knowledge, and more methods such as semisupervised and self-supervised methods based on gradient descent optimization can also be the base learner to conduct the few-shot task; second, a more interpretable and robust structure of parameters, such as a hierarchical structure, can be considered when extracting meta knowledge through many tasks.

REFERENCES

[1] M. Huang, Z. Liu, and Y. Tao, "Mechanical fault diagnosis and prediction in IoT based on multi-source sensing data fusion," *Simul. Modelling Pract. Theory*, vol. 102, 2020, Art. no. 101981.

[2] R. Liu, F. Wang, B. Yang, and S. J. Qin, "Multiscale kernel based residual convolutional neural network for motor fault diagnosis under nonstationary conditions," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 3797–3806, Jun. 2020.

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on June 01,2022 at 09:45:51 UTC from IEEE Xplore. Restrictions apply.

[3] F. Wang, R. Liu, Q. Hu, and X. Chen, "Cascade convolutional neural network with progressive optimization for motor fault diagnosis under nonstationary conditions," *IEEE Trans. Ind. Informat.*, vol. 17, no. 4, pp. 2511–2521, Apr. 2021.

[4] D. Wang, M. Zhang, Y. Xu, W. Lu, J. Yang, and T. Zhang, "Metric-based meta-learning model for few-shot fault diagnosis under multiple limited data conditions," *Mech. Syst. Signal Process.*, vol. 155, 2021, Art. no. 107510.

[5] S. Xing, Y. Lei, B. Yang, and N. Lu, "Adaptive knowledge transfer by continual weighted updating of filter kernels for few-shot fault diagnosis of machines," *IEEE Trans. Ind. Electron.*, 2021.

[6] H. Zhang, J. Zhang, and P. Koniusz, "Few-shot learning via saliency-guided hallucination of samples," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2770–2779.

[7] H. Yao *et al.*, "Graph few-shot learning via knowledge transfer," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 6656–6663.

[8] T. M. Hospedales, A. Antoniou, P. Micaelli, and A. J. Storkey, "Meta-learning in neural networks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020.

[9] T. Wang, X. Zhang, L. Yuan, and J. Feng, "Few-shot adaptive faster R-CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 7173–7182.

[10] K. Wang, J. H. Liew, Y. Zou, D. Zhou, and J. Feng, "Panet: Few-shot image semantic segmentation with prototype alignment," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 9197–9206.

[11] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 3637–3645.

[12] Q. Zhou, Y. Li, Y. Tian, and L. Jiang, "A novel method based on nonlinear auto-regression neural network and convolutional neural network for imbalanced fault diagnosis of rotating machinery," *Measurement*, vol. 161, 2020, Art. no. 107880.

[13] Z. He, H. Shao, P. Wang, J. J. Lin, J. Cheng, and Y. Yang, "Deep transfer multi-wavelet auto-encoder for intelligent fault diagnosis of gearbox with few target training samples," *Knowl.-Based Syst.*, vol. 191, 2020, Art. no. 105313.

[14] I. Martin-Diaz, D. Morinigo-Sotelo, O. Duque-Perez, and R. D. J. Romero-Troncoso, "Early fault detection in induction motors using adaboost with imbalanced small data and optimized sampling," *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 3066–3075, May/Jun. 2017.

[15] J. Mathew, C. K. Pang, M. Luo, and W. H. Leong, "Classification of imbalanced data by oversampling in kernel space of support vector machines," *IEEE Trans Neural Netw Learn. Syst.*, vol. 29, no. 9, pp. 4065–4076, Sep. 2018.

[16] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, "Gearbox fault diagnosis using a deep learning model with limited data sample," *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6263–6271, Oct. 2020.

[17] P. Peng, W. Zhang, Y. Zhang, Y. Xu, H. Wang, and H. Zhang, "Cost sensitive active learning using bidirectional gated recurrent neural networks for imbalanced fault diagnosis," *Neurocomputing*, vol. 407, pp. 232–245, 2020.

[18] A. Zhang, S. Li, Y. Cui, W. Yang, R. Dong, and J. Hu, "Limited data rolling bearing fault diagnosis with few-shot learning," *IEEE Access*, vol. 7, pp. 110895–110904, 2019.

[19] S.-S. Zhong, S. Fu, and L. Lin, "A novel gas turbine fault diagnosis method based on transfer learning with CNN," *Measurement*, vol. 137, pp. 435–453, 2019.

[20] J. Wu, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Few-shot transfer learning for intelligent fault diagnosis of machine," *Measurement*, vol. 166, 2020, Art. no. 108202.

[21] S. Bengio, Y. Bengio, J. Cloutier, and J. Gecsei, "On the optimization of a synaptic learning rule," in *Optimality in Biological and Artificial Networks?* Routledge, 2013, pp. 281–303.

[22] D. K. Naik and R. J. Mammone, "Meta-neural networks that learn by learning," in *Proc. IJCNN Int. Joint Conf. Neural Netw.*, 1992, pp. 437–442.

[23] S. Ravi and H. Larochelle, "Optimization as a model for few-shot learning," in *ICLR*, 2017.

[24] Y. Lee and S. Choi, "Gradient-based meta-learning with learned layerwise metric and subspace," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 2927–2936.

[25] E. Grant, C. Finn, S. Levine, T. Darrell, and T. Griffiths, "Recasting gradient-based meta-learning as hierarchical Bayes," in *Proc. 6th Int. Conf. Learn. Representations*, 2018.

[26] R. Zhang, T. Che, Z. Ghahramani, Y. Bengio, and Y. Song, "Metagan: an adversarial approach to few-shot learning," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 2371–2380.

[27] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1126–1135.

[28] J. Vanschoren, "Meta-learning: A survey," 2018, *arXiv:1810.03548*.

[29] D. A. Braun, C. Mehring, and D. M. Wolpert, "Structure learning in action," *Behav. Brain Res.*, vol. 206, no. 2, pp. 157–165, 2010.

[30] Y. Tian, Y. Wang, D. Krishnan, J. B. Tenenbaum, and P. Isola, "Rethinking few-shot image classification: A good embedding is all you need?," in *Proc. Comput. Vis.-ECCV 16th Eur. Conf.*, 2020, pp. 266–282.

[31] J. Bronskill, J. Gordon, J. Requeima, S. Nowozin, and R. Turner, "Tasknorm: Rethinking batch normalization for meta-learning," in *Proc. Int. Conf. Mach. Learn.,* 2020, pp. 1153–1164.

[32] H. Yao, Y. Wei, J. Huang, and Z. Li, "Hierarchically structured meta-learning," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 7045–7054.

[33] Q. Sun, Y. Liu, T.-S. Chua, and B. Schiele, "Meta-transfer learning for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 403–412.

[34] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 41–48.

[35] Y. Ge, F. Zhu, D. Chen, R. Zhao, and H. Li, "Self-paced contrastive learning with hybrid memory for domain adaptive object re-ID," 2020, *arXiv:2006.02713*.

[36] C. Lessmeier, J. K. Kimotho, D. Zimmer, and 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 *PHM Soc. Eur. Conf.*, vol. 3, no. 1, 2016.

[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, abs/1409.1556, 2014.

[38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

Yidan Hu received the B.S. degree in energy and power engineering in 2019 from Tianjin University, Tianjin, China, where she is currently working toward the M.S. degree in computer applications and techniques with the College of Intelligence and Computing.

Her research interests include deep learning and meta learning.

Ruonan Liu (M'19) received the B.S., M.S., and Ph.D. degrees in mechanical engineering from the School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China, in 2013, 2015, and 2019, respectively.

She was a Postdoctoral Researcher with the School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA, in 2019. She is currently an Associate Professor with the College of Intelligence and Computing, Tianjin University, Tianjin, China. Her research interests include machine learning and artificial intelligence.

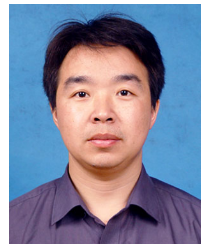
Xianling Li received the B.S., M.S., and Ph.D. degrees in power machinery and engineering from the School of Energy Science and Engineering, Harbin Institute of Technology, Harbin, China, in 2005, 2007, and 2011, respectively.

He is currently a Senior Engineer with the Science and Technology on Thermal Energy and Power Laboratory, Wuhan, China. His interests are focused on big data, fault diagnosis of complex power system, intelligent control.

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on June 01,2022 at 09:45:51 UTC from IEEE Xplore. Restrictions apply.

Dongyue Chen received the B.S. degree in transportation equipment information engineering and the M.S. degree in precision instruments and machinery from Southwest Jiaotong University, Chengdu, China, in 2015 and 2018, respectively. She is currently working toward the Ph.D. degree in computer applications and techniques with the College of Intelligence and Computing, Tianjin University, Tianjin, China.

Her research interests include deep learning, condition monitoring, and fault diagnosis of mechanical system.



Qinghua Hu (Senior Member, IEEE) received the B.S. and M.S. degrees in energy science and engineering from the Harbin Institute of Technology, Harbin, China, in 1999 and 2002, respectively, and the Ph.D. degree in aerospace from the Harbin Institute of Technology, Harbin, China, in 2008.

He was a Postdoctoral Fellow with the Department of Computing, Hong Kong Polytechnic University, Hong Kong, from 2009 to 2011. He is currently the Dean of the School of Artificial Intelligence, the Vice Chairman of the Tianjin Branch of China Computer Federation, the Vice Director of the SIG Granular Computing and Knowledge Discovery, and the Chinese Association of Artificial Intelligence. He is currently supported by the Key Program, National Natural Science Foundation of China. He has authored or coauthored more than 200 peer-reviewed papers. His current research is focused on uncertainty modeling in big data, machine learning with multimodality data, intelligent unmanned systems.

Dr. Hu is currently an Associate Editor for the IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Acta Automatica Sinica*, and *Energies*.

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on June 01,2022 at 09:45:51 UTC from IEEE Xplore. Restrictions apply.