Machine Intelligence Research
www.mi-research.net

20(4), August 2023, 583-594
DOI: 10.1007/s11633-022-1363-y

A New Diagnosis Method with Few-shot Learning Based on a Class-rebalance Strategy for Scarce Faults in Industrial Processes

Xinyao Xu[1,2] De Xu[1,2] Fangbo Qin[1]

[1] Research Center of Precision Sensing and Control, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

[2] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: For industrial processes, new scarce faults are usually judged by experts. The lack of instances for these faults causes a severe data imbalance problem for a diagnosis model and leads to low performance. In this article, a new diagnosis method with few-shot learning based on a class-rebalance strategy is proposed to handle the problem. The proposed method is designed to transform instances of the different faults into a feature embedding space. In this way, the fault features can be transformed into separate feature clusters. The fault representations are calculated as the centers of feature clusters. The representations of new faults can also be effectively calculated with few support instances. Therefore, fault diagnosis can be achieved by estimating feature similarity between instances and faults. A cluster loss function is designed to enhance the feature clustering performance. Also, a class-rebalance strategy with data augmentation is designed to imitate potential faults with different reasons and degrees of severity to improve the model's generalizability. It improves the diagnosis performance of the proposed method. Simulations of fault diagnosis with the proposed method were performed on the Tennessee-Eastman benchmark. The proposed method achieved average diagnosis accuracies ranging from 81.8% to 94.7% for the eight selected faults for the simulation settings of support instances ranging from 3 to 50. The simulation results verify the effectiveness of the proposed method.

Keywords: Data augmentation, feature clustering, class-rebalance strategy, few-shot learning, fault diagnosis.

Citation: X. Xu, D. Xu, F. Qin. A new diagnosis method with few-shot learning based on a class-rebalance strategy for scarce faults in industrial processes. *Machine Intelligence Research*, vol.20, no.4, pp.583–594, 2023. http://doi.org/10.1007/s11633-022-1363-y

1 Introduction

The data-driven fault diagnosis methods for various industrial applications have been investigated for years, which rely on sufficient fault instances. They can be classified into two categories: statistic-based methods[1, 2] and learning-based methods[3–7]. In general, the new scare faults for industry processes are identified by experts. However, these fault instances are too few to affect the models in data-driven methods, which results in a reduction of precision in the aforementioned diagnosis methods.

In order to diagnose faults under the condition of imbalanced fault instances, researchers use the strategies at different levels such as the data-level, model-level, and feature-level to process the distribution of faults, as shown in Fig. 1.

The data-level strategies equalize the number of instances for different faults in the original data space. Further sampling and data augmentation are two popular approaches. From the data point of view, downsampling[8] is a traditional but efficient sampling strategy that selects a small amount of typical data from the majority classes to balance the majority and minority classes. However, downsampling strategies will damage the original data distributions. The synthetic minority oversampling technique (SMOTE)[9, 10] and adaptive synthetic (ADASYN)[11, 12] sampling are also typical sampling strategies. However, the performance of the oversampling strategies is severely influenced by instances' distribu-

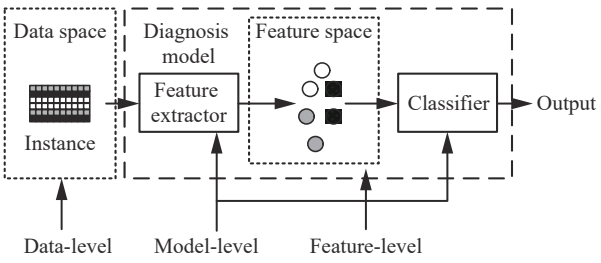


Fig. 1 Illustration of the strategies at different levels to process the distribution of faults

Research Article
Manuscript received on April 22, 2022; accepted on August 1, 2022; published online on February 18, 2023
Recommended by Associate Editor Jin-Jun Chen
Colored figures are available in the online version at https://link.springer.com/journal/11633

tions of the minority classes. Data augmentation is another popular group of data-level strategies. The generative adversarial network (GAN) based models[13, 14] perform well in generating fault instances. Since instances of new faults are too scarce to support the reliable generation of GAN, Zhuo and Ge[14] designed a generative adversarial network model using fault attributes (FAGAN) based on auxiliary classifier GAN[15]. The model uses fault attributes as prior knowledge to assist the generation of fault instances. However, when the generative network of FAGAN is fine-tuned with scarce fault instances of new faults, it still has the risk of overfitting to these support instances.

The model-level strategies optimize model structures or training procedures to alleviate the problem caused by data imbalance. The classifier ensemble strategy[16] and the cost-sensitive learning strategy[17] are two typical strategies. However, many of these strategies essentially need to modify conventional algorithms, including the discrimination threshold, while it is difficult to define a cost function that ensures performance stability[18]. The new faults can also be diagnosed with incremental models. Yu and Zhao[19] designed an incremental neural network called the broad convolutional neural network that allows the model to diagnose new faults with few samples. However, the incremental networks induce the catastrophic forgetting problem[20], which is also problematic in the machine learning field.

In real industrial processes, common features exist for many faults. However, many strategies above treat the new classes as individual ones and neglect the features learned from historical records that can be highly related to the new faults. The methods with such strategies usually retrain entire diagnosis models, which is usually unnecessary. In contrast, methods with feature-level strategies extract universal features from fault instances. The new faults can be easily identified with these robust features.

Within feature-based strategies, many researchers have designed few-shot learning strategies to handle the problem in recent years, which require limited adjustments for the diagnosis models. The diagnosis models are trained on the source data sets with sufficient instances and then transferred into the target data sets with few support instances. For image-based tasks, the source data sets can be ImageNet[21] and other universal data sets. For tasks with specialized knowledge, the source data sets can be collected from similar objects[22] or the same object under different working conditions[23, 24]. For example, Lu et al.[25] proposed a transfer relation network to accomplish the few-shot transfer learning task in rotation machinery. During the training procedure, the meta-learning strategies[23, 26, 27] are used to enhance the generalizability of the models. Wang et al.[27] proposed a feature space metric-based meta-learning model (FSM3) based on the prototypical network[28] and the matching network[29] to diagnose the faults of bearings and gearboxes under limited data conditions. The model casts fault instances into a feature embedding space and diagnoses the faults by comparing the feature similarity between the tested instances and the support instances of faults.

It is promising to diagnose new rare faults with few-shot learning methods. However, the limitation is that the source data sets with sufficient instances are always required for model training of the few-shot diagnosis methods above. In real industrial processes, it is difficult to collect source data sets that can cover all working states. The restriction of the source data set containing limited fault categories reduces the performance of these methods.

In this paper, a new diagnosis method with few-shot learning is designed to diagnose new rare faults in industrial processes. Strategies are designed to handle the restriction above. In industrial processes, many faults are caused by deviations from the normal states in representative variables or latent features. Such prior knowledge can be used to augment the source data set. A class-rebalance strategy is designed to construct class-balanced batches from the initial data set. Data augmentation is used on these data batches to generate new fault batches. The proposed model is trained via feature clustering with these generated fault batches to take full advantage of these unlabeled fault instances. In addition, a cluster loss function is designed for the unsupervised feature clustering tasks during model training. With the strategies above, the proposed model can effectively identify the new rare faults with few support instances. The main contributions are as follows:

1) A new diagnosis method is designed based on the prototypical network to diagnose new rare faults in industrial processes. It comprises a standardization module, a data segmentation module, a class-rebalance module, a data augmentation module, a feature extractor, a feature mapping module, and a similarity calculation module. The simulation results on the Tennessee-Eastman benchmark verify the effectiveness of the proposed method.

2) A class-rebalance strategy is designed to handle the restriction of source data sets with limited fault categories. During the model training process, instances with an equal number for each fault are selected to construct class-balanced batches. Then, data augmentation with prior knowledge is used on these class-balanced batches to generate new fault batches. The model can be trained on these generated batches by feature clustering. The strategy expands the variety of the source data set and improves the diagnosis performance of the proposed model effectively.

3) A cluster loss function is designed to emphasize the differences between interclass instances and the similarit-

ies of intraclass instances in feature embedding space. The loss function can effectively improve the proposed model's performance for feature clustering and fault diagnosis.

2 Problem statement and background

2.1 Problem definition

The diagnosis problem of the new rare faults with few support instances is investigated. Sufficient records for normal working conditions and several historical faults are provided as the initial data set. The initial data set $D_{ini} = \{S_n, S_{fl}, \cdots, S_{fn}\}$, where S_n denotes normal records, S_{fl} to S_{fn} denote the records of n historical faults, respectively. An initial diagnosis model $f(\cdot)$ is supposed to be trained with D_{ini}. Then, a new data set $D_{new} = \{S_{fn+1}, \cdots, S_{fm}\}$ including $(m-n)$ new faults is added to the initial data set. The support instances of new faults are much fewer than the support instances of historical faults. The new diagnosis model $f^n(\cdot)$ is supposed to be trained with the mixture data set D_{ini}, D_{new}. The purpose of this paper is to achieve an effective diagnosis of new faults with few support instances while keeping the diagnosis performance on historical faults.

2.2 Classification model with metric-learning

Framework of the model with metric-learning. Generally, models with metric-learning strategies cast instances of different classes into separate feature clusters in feature embedding spaces. New instances are usually identified by comparing the feature similarity between these instances and class representations, as described in Fig. 2. The representations can be the support instances of data classes directly. The matching network and the Siamese neural network[30] identify instances by comparing the feature similarity between these tested instances and the labeled support instances in feature embedding spaces. As another choice, the representations can also be the feature centers of corresponding support instances. The prototypical network in [28] is a popular metric-based model for few-shot learning tasks. The single representation for each data class is calculated as the mean value of the features of the corresponding support instances. The tested instances are identified by comparing the feature similarity between these instances and the data classes. The proposed method is designed based on the prototypical network model.

Generally, the feature transformation networks should be trained on sufficient instances. However, considering the limited instances for few-shot learning tasks, the feature transformation networks are usually trained on similar data sets with abundant instances.

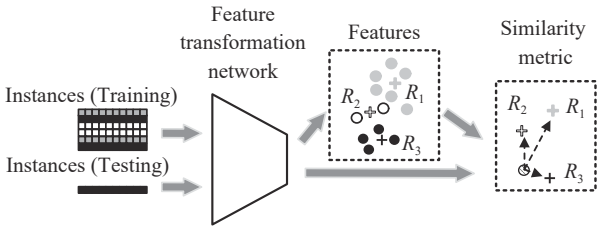


Fig. 2 Framework of models with metric learning. R_1, R_2 and R_3 denote the representations of feature clusters.

Triple Loss. In order to project the features of different faults into separate clusters, the training loss functions should be designed for the feature transformation networks. The nature of metric-learning is to learn a data-driven metric to make the distances between intraclass features smaller and the distances between interclass features larger. A general formula of the training losses is[31]

$$L = \lambda_1 L_1(d^w) + \lambda_2 L_2(d^b) + \lambda_3 L_3(d^w, d^b) \tag{1}$$

where d^w and d^b denote the intraclass distance and interclass distance, respectively. λ_1, λ_2, and λ_3 are the trade-offs. The three terms represent the intraclass constraint, interclass constraint, and relative constraint, respectively.

The triple loss is one of the widely applied loss functions for metric-learning. It is first proposed in face recognition tasks[32], as given in (2). The triple loss involves three instances, named anchor instance, positive instance, and negative instance, respectively. The positive instance belongs to the same class as the anchor instance, but the negative instance belongs to a different class. The intention of triple loss is to draw the distance between features of the anchor instance and the positive instance closer than the distance between the features of the anchor instance and the negative instance.

$$L_t = \sum_i^N \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+ \tag{2}$$

where x_i^a is the anchor instance and x_i^p is the positive instance. x_i^n is the negative instance. α denotes the predefined margin of each data cluster. $f(\cdot)$ denotes the transformation function of the feature transformation network. $[\cdot]_+$ is the operation of $\max(\cdot, 0)$.

3 Fault diagnosis method with few-shot learning based on a class-rebalance strategy

This section describes the proposed method. Fig. 3 shows the framework of the proposed method with few-shot learning based on a class-rebalance strategy, which

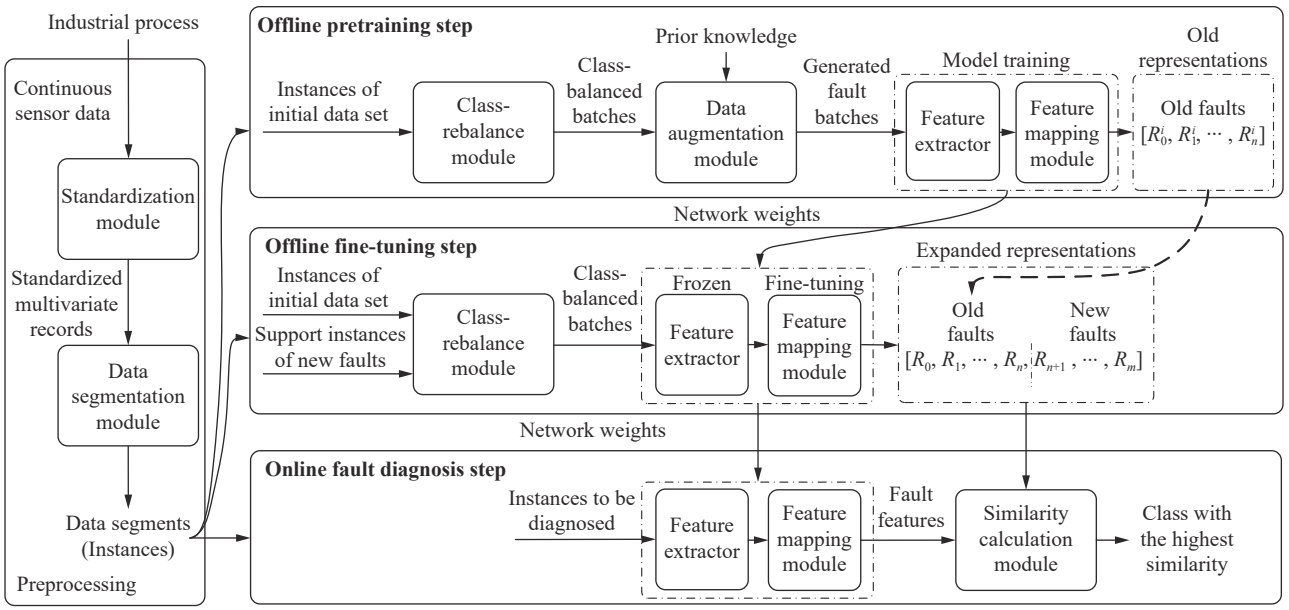


Fig. 3 Framework of the proposed method with few-shot learning based on a class-rebalance strategy. $R_1^i, \cdots, R_n^i$ are the representations of the n faults in the initial data set. R_0^i is the representation of normal working states. $R_0, R_1, \cdots R_m$ are the expanded representations that are calculated by the fine-tuned model. R_{n+1} to R_m are the representations of new faults.

mainly contains three parts:

Preprocessing of the historical records. The consecutive working states of industrial processes are usually monitored by sensor data that are recorded as floating point numbers. Considering the system with k sensors, data vector X_t is sampled at time t with k sensor data ($X_t = [x_1, x_2, \cdots, x_k]$). The standardization module first standardizes the continuous sensor records $[X_0, X_1, \cdots, X_t]$ along each variable, as $[\hat{X}_0, \hat{X}_1, \cdots, \hat{X}_t]$. Then, in order to analyze the dynamic features of the standardized multivariate records, the data segmentation module divides the standardized records into individual data segments using a sliding window with fixed sliding steps s. Data segment S_t collected at time t includes i consecutive data vectors $[\hat{X}_{t-i+1}, \hat{X}_{t-i+2}, \cdots, \hat{X}_t]$. These data segments are the model's input instances.

Training of the diagnosis model. The training includes the pretraining and fine-tuning steps.

In the pretraining step, a robust feature extractor is supposed to be trained with the initial data set. The class-rebalance module is first executed to select instances with equal numbers of normal states and historical faults in the initial data set to construct class-balanced batches. Then, the data augmentation module described in Section 3.2 is used on these data batches to generate new fault instances that are different from the real faults. Instances with an equal number are selected from each generated fault to construct generated fault batches in a similar way to the class-balanced batches. The model is trained via feature clustering with these generated fault batches. With the pretrained model, instances of different faults can be projected into separate feature clusters in the feature embedding space. The centers of the normal states and n real feature clusters are defined as the fault representations R_0^i to R_n^i.

The model needs to be adjusted to identify new rare faults during the fine-tuning step. The class-rebalance module is first executed to construct class-balanced batches with the instances of the initial data set and the support instances of new faults. The parameters of the feature extractor are fixed. Then, the feature mapping module is fine-tuned with those class-balanced batches directly. Since the parameters of the feature mapping module have been changed, the representations of normal states and m faults should be recalculated with the corresponding fault instances.

Online diagnosis. During the online diagnosis, the features of instances to be diagnosed are first calculated by the model. Then, the similarity calculation module is executed to diagnose the instances with the feature similarity between instances and the faults. The cosine similarity is used as a similarity metric of the proposed method. The fault with the most similar representation is diagnosed as the output.

3.1 Architecture of the proposed model

Model structure. Fig. 4 shows the basic architecture of the proposed model. The basic structure of the feature extractor is a 1-D convolutional neural network (1-D CNN) with a channel-wise attention mechanism. With a two-layer 1-D CNN as an example, the Conv1d 1 and Conv1d 2 are 1-D convolution layers with kernels sized m, whose input and output dimensions are d_1, d_2, and d_2, d_3, respectively. The attention vectors are calculated by auto-encoder modules. The feature maps are av-

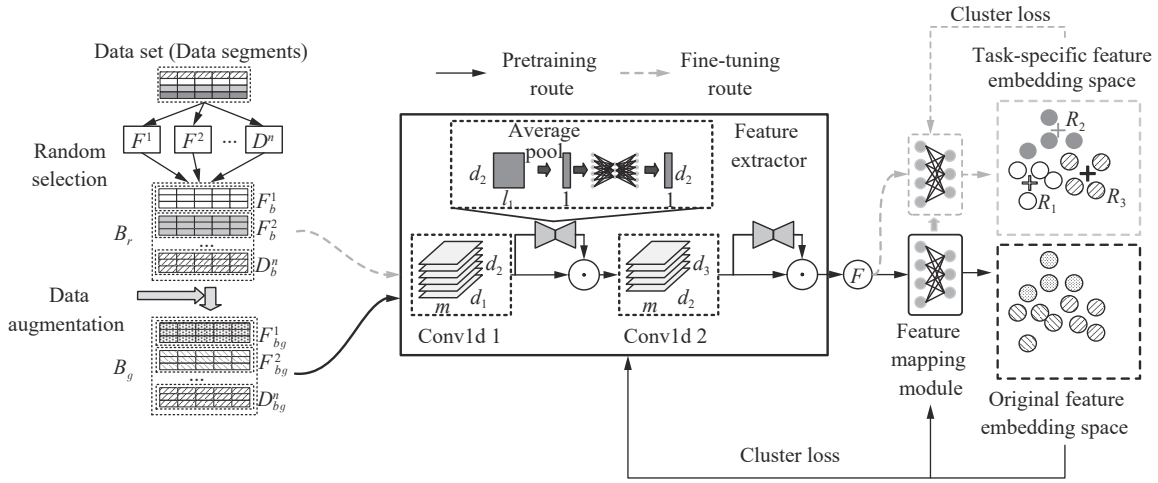


Fig. 4 The model architecture of the proposed method. F^i in the blank square in the left part denotes the sub-dataset for the fault i, and D^n denotes the sub-dataset for normal working states. F_b and D_b denote the instances of faults and normal working states in the class-balanced batches B_r formed with real instances obtained using the class-rebalance strategy. F_{bg} and D_{bg} denote the generated fault instances and instances of normal working states in generated fault batch B_g, respectively. m, d_1, d_2, d_3 are the parameters of the convolution kernels. R_1, R_2, and R_3 are the fault representations.

eraged in each channel, and the averaged feature maps are taken as the inputs of the auto-encoder module. The dot in a circle after the Conv1d, as shown in the middle part of Fig. 4, means the operation of multiplying along channels. After the feature extractor, the extracted features are cast into the final feature embedding space with the feature mapping module. The feature mapping module can be a fully connected layer only. The features of different faults and normal states are projected into separate feature clusters.

Generally, the task of fault diagnosis is to identify the deviations from normal working states. Because of some large feature deviations, the minor feature deviations might be suppressed to small values after the normalization operation, which are too small to be detected. To highlight those slight deviations, a designed transformation function is added after the feature extractor, shown as the circled symbol F in Fig. 4. The function is

$$F(x) = \begin{cases} \ln(x+1), & \text{if } x > 0 \\ 0, & \text{if } x = 0 \\ -\ln(|x|+1), & \text{if } x < 0. \end{cases} \tag{3}$$

Cluster loss. The proposed method is designed to be trained on abundant generated faults rather than real limited faults. With the pretrained model, the instances of different faults can be cast into separate feature clusters in the feature embedding space in Fig. 4. To make the distances between interclass features larger and the distances between intraclass features smaller, cluster loss is proposed, which is shown in Fig. 5(b). Compared to the triple loss in Fig. 5(a), the intraclass instances shrink together; meanwhile, the interclass instances are pushed away from each other. Since the cosine similarity is used as the model's distance metric, the cluster loss L can be calculated by (4) to (6). $D(\cdot)$ denotes the distance based on cosine similarity. C denotes the categories within a single training batch. n_C denotes the total category number. c_l denotes the l-th class. γ is a predefined parameter. x_i, x_j, and x_k are the input instances.

$$D(x_i, x_j) = \frac{1}{2}\left(1 - \frac{f(x_i)^{\mathrm{T}} f(x_j)}{|f(x_i)||f(x_j)|}\right). \tag{4}$$

$$L'_t = \frac{1}{n_C}\sum_{c^l}^{C}\left(\arg\max_{x_i, x_j \in c^l}(D(x_i, x_j)) - \arg\min_{x_i \in c^l, x_k \notin c^l}(D(x_i, x_k)) + \alpha\right)_+ \tag{5}$$

$$L = L'_t + \frac{1}{n_C}\sum_{c^l}^{C}\left(\arg\max_{x_i, x_j \in c^l}(D(x_i, x_j)) + \frac{\gamma}{\arg\min_{x_i \in c^l, x_k \notin c^l}(D(x_i, x_k))}\right). \tag{6}$$

3.2 Training procedure with a class-rebalance strategy

In order to train the model on a limited source data set, a class-rebalance strategy is proposed, as shown in Fig. 6. It is prior knowledge that many faults in industrial processes are caused by deviations from normal working states in representative variables or latent features. Therefore, this strategy is designed to generate potential faults according to such prior knowledge.

Firstly, the instances with equal numbers from differ-

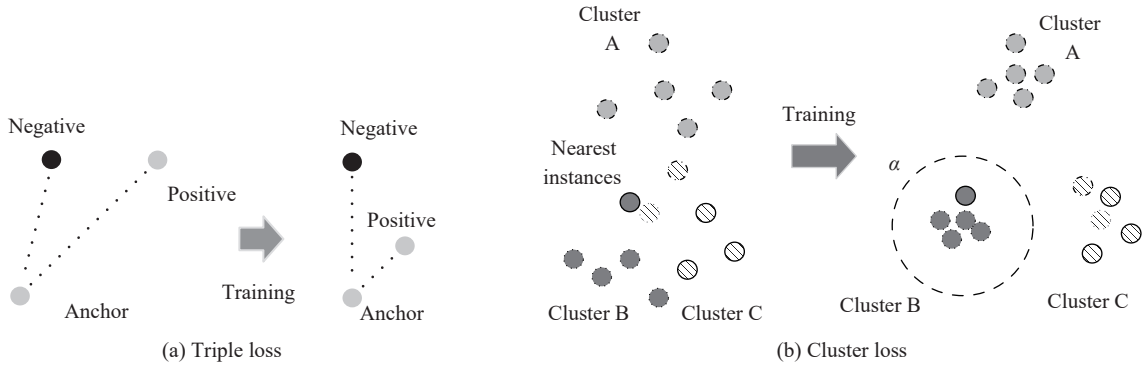


Fig. 5 Triple loss and the cluster loss. The circles with different shading represent feature vectors belonging to different classes. The circles with margins in different styles denote the nearest intercluster instances to the clusters with the same style margins. α denotes the predefined margin of each data cluster.

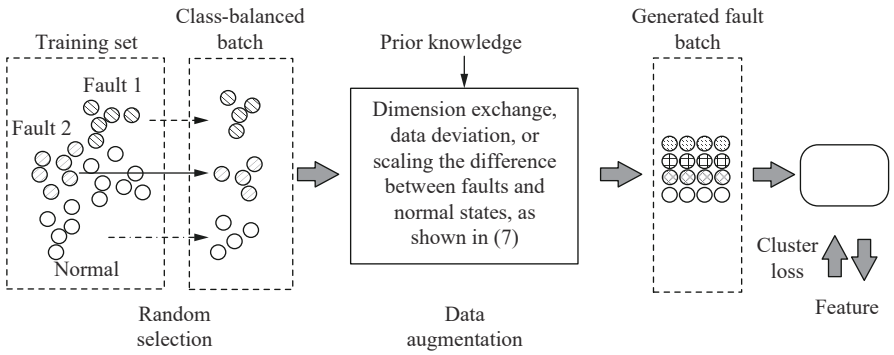


Fig. 6 Generation details of the training batch in the pretraining step. Solid circles with different shading represent the instances belonging to different classes.

ent faults and normal states are randomly selected from the training set to construct class-balanced batches. Then, data augmentation is used on these class-balanced batches to generate potential fault instances for different reasons and severity degrees. Specifically, the differences between the real fault instances and the center of normal instances are randomly scaled. Then, operations such as dimension exchange and scaling along the time are performed on these fault instances. The operations can be expressed as

$$x_g = g\left(\mu_n^x + \gamma(x_f - \mu_n^x)\right) \tag{7}$$

where x_g is the generated fault instance and x_f is the instance belonging to the real fault type in a class-balanced batch. μ_n^x denotes the statistical mean of normal instances. γ is a random amplitude, which can be set to either constant 1 or follow the standard distribution $U(0, 1)$. $g(\cdot)$ is the data augmentation operation, which can be the operation of dimension exchange and scaling over time. Since the strategy does not augment the real faults with limited support instances, it alleviates the overfitting problem suffered by traditional data augmentation strategies.

During the fine-tuning step, the feature mapping module is directly fine-tuned with the class-balanced batches that are selected from the real data set.

3.3 Online diagnosis with representations

The diagnosis result is decided by the average of the feature similarity between the new instance and the support instances of different faults in the feature embedding space. Since the features in the feature embedding space are normalized, the similarity can be calculated by (8).

$$s^l = \frac{1}{n^l} \sum_{x_i \in c^l} f(x)^{\mathrm{T}} f(x_i) = f(x)^{\mathrm{T}} \mu^l \tag{8}$$

where s^l denotes the average of the similarity from instance x to the support instances of fault c^l in feature embedding space. n^l denotes the number of the support instances belonging to c^l. μ^l denotes the average feature of support instances belonging to c^l, and it can be referred to as the representation of c^l.

The diagnosis result c is the fault c^l with the highest similarity $s_{\max}$.

$$\left\{ c = c^l | s_{\max} = \arg\max_{c^l \in C} \left(f(x)^{\mathrm{T}} \mu^l \right) \right\}. \tag{9}$$

4 Simulation on Tennessee-Eastman process

4.1 Data set

In order to verify the effectiveness of the proposed method in the fault diagnosis for new scarce faults, the Tennessee-Eastman process[33] is used in simulation, which is a typical benchmark for evaluating fault detection and diagnosis methods for industrial processes. The process includes five units (a condenser, a compressor, a reactor, a cooler, and a striper). There are 53 monitored variables involved, and 21 faults can be simulated with the program[34]. Simulations are conducted in mode one and the simulation for each fault lasts for 600 hours (fault 6 ends at 7.1 hours). The sampling period is three minutes. The simulations involve 52 variables used for monitoring the working states of the chemical process[33] (except the agitator setting, which is a constant value of 100) and eight types of faults. The descriptions of the faults are shown in Table 1. A total of 52 sensor data can be sampled as the data vector X_t ($X_t = [x_1, x_2, \cdots, x_{52}]$) at the time t. All the records are segmented by the sliding window with a length of 72 minutes and a sliding step of six minutes. Hence, the instance S_t collected at time t contains 24 consecutive data vectors $[X_{t-23}, X_{t-22}, \cdots, X_t]$. The initial data set includes fault 1 to fault 3, as well as the normal states. Each class has the first 3 000 consecutive instances. Faults 4 to 8 are the new faults. A test set including 1 120 normal consecutive instances and 6 309 fault instances is built. There are 900 consecutive instances for each fault except fault 6. There are nine instances for fault 6. Since the durations of the faults are different, two sampling settings are considered here. One setting is the sparse random sampling within long faults existing procedures, which imitates multiple faults records with short recording procedures. The other is consecutive dense sampling within a single fault record. The two sampling settings are called sparse sampling and dense sampling in the following sections.

Table 1 Descriptions of chosen faults

No.	Fault state	Disturb
0	Normal	/
1	A/C feed ratio, B composition constant (Stream 4)	Step change
2	B composition, A/C ratio constant (Stream 4)	Step change
3	D feed temperature (Stream 2)	Step change
4	Reactor cooling water inlet temperature	Step change
5	Condenser cooling water inlet temperature (Stream 2)	Step change
6	A feed loss (Stream 1)	Step change
7	C header pressure loss – Reduced availability	Step change
8	A, B, C feed composition (Stream 4)	Random variants

The Tennessee-Eastman process produces two products from reactants A, C, D, and another reactant E. B is inert. The details of the process are described in [33].

4.2 Parameters and details

Table 2 shows the settings of the proposed method and the settings of its training procedure. For the proposed model, the feature extractor is a 1-D CNN with three convolution layers. Conv and Attention denote the convolution module and the attention module in the feature extractor, respectively. Conv1d denotes a 1-D convolution layer with kernels sized three and stride two. FC denotes the fully connected layer. (a, b) attached behind the Conv1d and FC are the layers′ input dimension a and output dimension b, respectively. The LReLU represents the leaky rectified linear unit activation function. The feature mapping module is a fully connected layer with input dimension 128 and output dimension 32. The pre-training step lasts 900 batches. For each batch, 80 random instances for the normal states and n types of real faults are collected to generate a training batch with a size of $80\times(n+1)$. For the fine-tuning step, the instance number for each fault is the same as the number of support instances for each minority class. The fine-tuning step only lasts for one training batch.

Table 2 Network parameters and training settings

Preprocessing	
Data segmentation: Sliding window with sliding Step 2 Signal denoise: Wavelet filtering	
Model parameters	
Feature extractor	Conv 1: {Conv1d (52, 64) LReLU} Attention 1: {FC (64, 16) LReLU FC (16, 64) Sigmoid} Conv 2: {Conv1d (64, 96) LReLU} Attention 2: {FC (96, 12) LReLU FC (12, 96) Sigmoid} Conv 3: {Conv1d (96,128) LReLU} Attention 3: {FC (128,16) LReLU FC (16,128) Sigmoid}
Feature mapping module	{FC(128, 32)}
Training settings	
Adam optimizer, initial training step: 0.001, sample number: 80, max iteration: 900, α in cluster loss: 0.01, γ in cluster loss: 0.001, fine-tuning training step: 0.000 1, fine-tuning iteration: 1	

4.3 Comparison study

In this section, comparison simulations are conducted for various few-shot settings. The support instance number N_s of new faults for both sampling settings above varies from 3 to 50. The data augmentation method

FAGAN[14] and two methods based on feature similarity named FSM3[27] and PNC[28] are selected as the comparison methods. The PNC denotes the prototypical network with the cosine similarity metric. The simulations for each setting are repeated 40 times, and the average diagnosis accuracies for the eight faults are given in Table 3. Fig. 7 provides the distributions of diagnosis accuracies of the methods above for dense sampling settings. Moreover, the average diagnosis accuracies of each fault with five support instances for dense sampling settings are given in Table 4. The results under faults 1 to 3 are the average diagnosis accuracies. The 1-D CNN denotes a basic 1-D CNN classification network.

For FAGAN, the model's generative network is established on a 1-D CNN rather than a multilayer perceptron network. Moreover, its fault attributes refer to the original paper[27]. To achieve a reasonable comparison, the feature extractors of 1-D CNN, PNC, FSM3, and the discriminator in FAGAN have the same architectures as the proposed method.

Since the instances of normal states and faults 1 to 3 are sufficient, the diagnosis accuracies for the three faults are generally high for all methods, ranging from 87.8% to 100.0% in Table 4. Moreover, compared with the support instances for sparse sampling settings, the support instances for dense sampling settings are more similar to each other. Therefore, similar support instances lead to models' lower diagnosis performances than the diagnosis performances for sparse sampling settings, as given in Table 3.

Tables 3 and 4 show that 1-D CNN has low diagnosis accuracies for the five new faults. Similar results are shown in Fig. 7(a). The results indicate an invalid learning of the new faults for 1-D CNN. FAGAN performs well when the support instances are sufficient, as shown in Fig. 7(b). However, FAGAN fails to learn the patterns of fault 5 and fault 8 with few support instances. Hence, it has extremely low diagnosis performance on the two

Table 3 Average diagnosis accuracies for different sampling settings (%)

Type	Sparse sampling					Dense sampling				
N_s	3	5	10	20	50	3	5	10	20	50
1-D CNN	53.8	57.0	65.4	71.0	75.2	52.2	51.2	58.9	62.1	68.9
FAGAN[14]	71.7	73.7	78.6	88.6	**95.7**	71.1	72.0	74.4	79.4	84.8
FSM3[27]	65.6	64.3	64.4	63.5	62.7	65.6	63.5	63.8	64.0	65.0
PNC[28]	75.5	76.4	77.9	77.7	79.1	72.9	75.1	75.0	75.5	75.0
Proposed	**86.5**	**90.5**	**91.9**	**93.9**	94.7	**81.8**	**82.8**	**83.1**	**85.0**	**88.9**

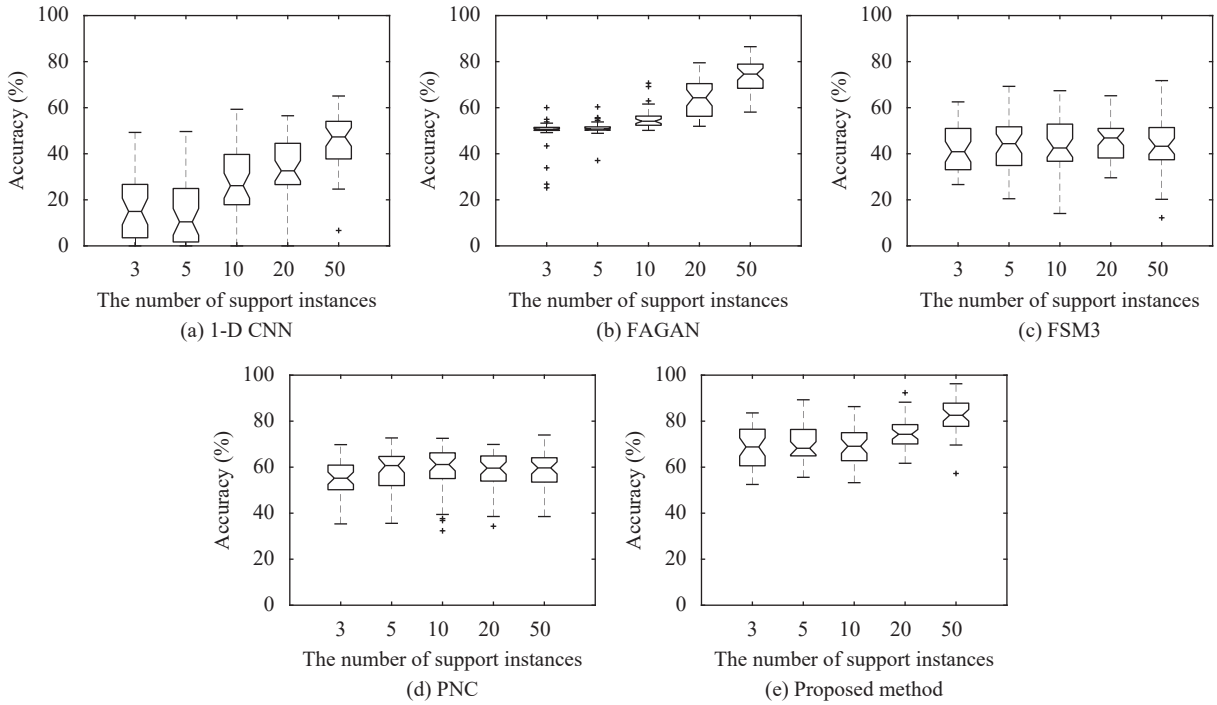


Fig. 7 Distributions of diagnosis accuracies for the five new faults for dense sampling settings in comparison simulations with different methods

Table 4 Diagnosis details with five support instances for a dense sampling setting (%)

Faults	0	1 to 3	4	5	6	7	8
1-D CNN	99.7	**100.0**	40.0	0.3	20.0	20.3	9.0
FAGAN [14]	**99.9**	**100.0**	**99.9**	1.1	90.0	98.4	5.0
FSM3 [27]	72.7	87.8	85.9	30.0	59.4	41.6	12.3
PNC [28]	78.1	97.2	97.4	30.1	88.6	76.1	14.9
Proposed	93.5	99.6	99.2	**41.8**	**92.5**	**100.0**	**39.0**

faults, as given in Table 4. When the support instances are similar, the instances generated by FAGAN overemphasize the biased distributions of the few support instances, which worsens the overfitting problem. When the number of support instances is 50, the average diagnosis accuracy drops from 95.7% to 84.8% for FAGAN. The PNC has good performance when the support instances are extremely rare. However, since the networks are trained on the initial data set only, the performances of both metric-based methods FSM3 and PNC have few improvements when the support instances increase, as shown in Figs. 7(c) and 7(d). The proposed method performs the best for most simulation settings.

The t-SNE (t-distributed stochastic neighbor embedding)[35] is one of the most widely used feature visualization algorithms. With t-SNE, the distributions of features in feature embedding space can be visualized with the 2-D vectors projected from the features with 32 dimensions. PNC, FSM3, and the proposed method have similar classification mechanisms. They identify new instances by comparing feature similarity between these instances and existing data classes in feature embedding spaces. The fault features calculated by the three feature-based methods are investigated by t-SNE. Fig. 8 provides visualized feature distributions of the tested instances of faults 0 to 3 and fault 5, which are calculated by PNC, FSM3, and the proposed method with the same five support instances of fault 5. The visualized feature distributions of fault 5 in Figs. 8(b) and 8(c) partly overlap with the visualized feature distributions of fault 0. For PNC and FSM3, the overlapped distributions of fault 0 and fault 5 confuse these models' identification for both fault classes and lead to lower diagnosis performance, as shown in Table 4.

Besides, support instances located at the border of the distributions can lead to serious misdiagnosis of the instances. Compared with Figs. 8(b) and 8(c), the proposed method can provide a clearer visualized feature distribution of fault 5, as shown in Fig. 8(a). Despite the random nature of locations for the support instances, the proposed method can provide more distinguishable representations with clearer feature distributions. As a result, the proposed method performs best in most cases.

Then, ablation simulations are conducted for the dense sampling settings to compare the effectiveness of each module. Simulations for each setting are repeated 15 times, and the average results are given in Table 5. The CNN in Table 5 denotes the 1-D CNN trained with the cluster loss function. Compared with the diagnosis accuracies of 1-D CNN with the traditional cross-entropy classification loss function in Table 3, the 1-D CNN with cluster loss function performs much better. The diagnosis accuracies of 1-D CNN rise from the range of 51.2% to 68.9% to the range of 74.2% to 79.1% for the dense sampling settings. As given in Table 5, the data augmentation module can effectively improve the diagnosis performance of the model. The diagnosis accuracies are further raised to the range of 78.0% to 85.5%. The class-rebalance module selects instances with an equal number for each fault in class-balanced batches. The operation highlights the instances of new faults, whereas it has a more severe overfitting problem when the model's generalizability is insufficient. Although the individual class-rebalance module is not effective with cluster loss, it improves the model's diagnosis performances on the basis of data augmentation. Since the proposed model with the class-rebalance module cannot cluster the feature of each fault with a single instance, the average diagnosis accuracy of the model is slightly inferior to the model with the data augmentation module only for the set with a single support instance. However, it outperforms the rest of the simulation settings.

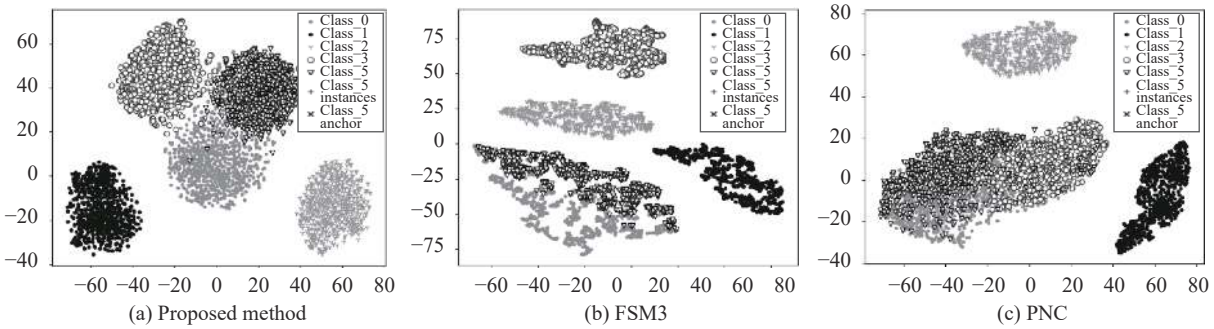


Fig. 8 Visualized feature distributions of faults 0 to 3 and fault 5 for dense sampling settings in comparison simulations with different methods. The visualized distributions are calculated with t-SNE. Class i denotes the instances of fault i in the test set. Class_5_anchor is the representation of fault 5. Class_5_instances are the five support instances of fault 5.

Table 5 Average diagnosis accuracies for the ablation test (%)

N_s	1	5	10	20
CNN	74.2	78.7	79.1	79.1
CNN + Data-rebalance	73.4	77.3	77.1	77.1
CNN + Data augmentation	**78.0**	83.0	85.5	85.0
CNN + Data-rebalance + Data augmentation (proposed)	77.7	**84.9**	**86.0**	**86.6**

4.4 Generalizability of the feature extractor

The generalizability of the proposed method is further investigated in this section. 1-D CNN, PNC, FSM3, and the proposed method are only trained on the initial data set. The four models have feature extractors with the same network architecture as well. The tested instances of all faults are projected into the feature embedding spaces of the four models. Fig. 9 shows the visualized feature distributions, which are calculated by t-SNE. It can be seen from Fig. 9(a) that the new fault features extracted by 1-D CNN mix with other feature clusters except for the features of fault 4. The feature distributions show that the initial diagnosis model for 1-D CNN has low generalizability for the new faults. In comparison, the proposed method has much clearer feature distributions for the new faults. Compared with Figs. 9(b) and 9(c), the proposed method has the clearest feature distributions for the new faults, as shown in Fig. 9(d). Therefore, the proposed method can diagnose the new faults more efficiently.

4.5 Discussions

For real industrial processes, the initial data set can hardly cover all possible working states. Moreover, since the feature extractor′ parameters are fixed during the fine-tuning step, it is more difficult for the model to learn the features of new faults with new fault mechanisms. Therefore, the balance strategy to deal with the "flexible-stable" dilemma with limited instances will be developed in the future.

Another limitation lies in the fine-tuning step for new faults. Since the instances of new faults are extremely rare, it is difficult to design a proper indicator to stop the fine-tuning early.

5 Conclusions

In this article, a fault diagnosis method with few-shot learning based on a class-rebalance strategy is proposed

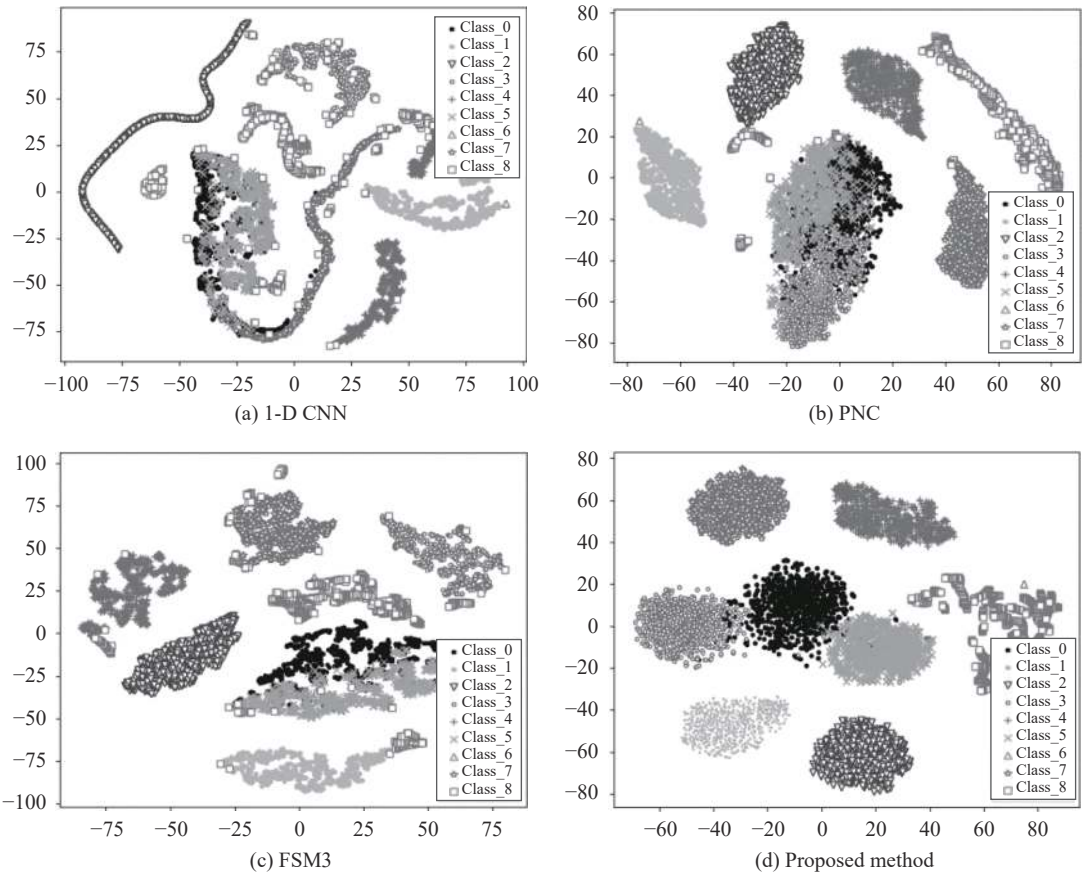


Fig. 9 Visualized feature distributions with t-SNE, class 0 denotes the normal working state, Class_1 to Class_8 denote fault 1 to fault 8.

to diagnose the new scarce faults identified by experts for industrial processes. The proposed model projects instances of different faults into separate feature clusters in a feature embedding space. The average feature of the support instances for each fault is the fault representation. During the online diagnosis procedure, the diagnosis of new instances is decided by feature similarity between instances and faults. A cluster loss is designed to enhance the proposed model's clustering performance in feature embedding space. As well, a class-rebalance strategy with data augmentation is designed to improve the diagnosis performance of the proposed model. The simulations of fault diagnosis on the Tennessee-Eastman benchmark were performed. The simulation results verify the effectiveness of the proposed method.

In the future, the parameters' updating strategy of the feature extractor will be focused on enabling the feature extractor to diagnose the faults with new fault mechanisms while maintaining good diagnosis performance for existing faults.

Acknowledgements

This work was partly supported by National Natural Science Foundation of China (Nos. 61733004, 62103413), the National Key Research and Development Program of China (No. 2018YFD0400902)。

Declarations of conflict of interest

The authors declared that they have no conflicts of interest to this work.

References

[1] C. H. Hu, J. Y. Luo, X. Y. Kong, X. W. Feng. Novel fault subspace extraction methods for the reconstruction-based fault diagnosis. *Journal of Process Control*, vol. 105, pp. 129–140, 2021. DOI: 10.1016/j.jprocont.2021.07.008.

[2] P. Zhou, R. Y. Zhang, J. Xie, J. P. Liu, H. Wang, T. Y. Chai. Data-driven monitoring and diagnosing of abnormal furnace conditions in blast furnace ironmaking: An integrated PCA-ICA method. *IEEE Transactions on Industrial Electronics*, vol. 68, no. 1, pp. 622–631, 2021. DOI: 10.1109/TIE.2020.2967708.

[3] L. Wen, X. Y. Li, L. Gao, Y. Y. Zhang. A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, 2018. DOI: 10.1109/TIE.2017.2774777.

[4] H. Liu, J. Z. Zhou, Y. Zheng, W. Jiang, Y. C. Zhang. Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Transactions*, vol. 77, pp. 167–178, 2018. DOI: 10.1016/j.isatra.2018.04.005.

[5] S. Yin, S. X. Ding, X. C. Xie, H. Luo. A review on basic data-driven approaches for industrial process monitoring. *IEEE Transactions on Industrial Electronics*, vol. 61, no. 11, pp. 6418–6428, 2014. DOI: 10.1109/TIE.2014.2301773.

[6] N. Laouti, S. Othman, M. Alamir, N. Sheibat-Othman. Combination of model-based observer and support vector machines for fault detection of wind turbines. *International Journal of Automation and Computing*, vol. 11, no. 3, pp. 274–287, 2014. DOI: 10.1007/s11633-014-0790-9.

[7] Y. Zhang, C. Bingham, M. Garlick, M. Gallimore. Applied fault detection and diagnosis for industrial gas turbine systems. *International Journal of Automation and Computing*, vol. 14, no. 4, pp. 463–473, 2017. DOI: 10.1007/s11633-016-0967-5.

[8] H. Q. Wang, Y. L. Ke, G. G. Luo, G. Tang. Compressed sensing of roller bearing fault based on multiple down-sampling strategy. *Measurement Science and Technology*, vol. 27, no. 2, Article number 025009, 2016. DOI: 10.1088/0957-0233/27/2/025009.

[9] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002. DOI: 10.1613/jair.953.

[10] J. Mathew, C. K. Pang, M. Luo, W. H. Leong. Classification of imbalanced data by oversampling in kernel space of support vector machines. *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 9, pp. 4065–4076, 2018. DOI: 10.1109/TNNLS.2017.2751612.

[11] H. B. He, Y. Bai, E. A. Garcia, S. T. Li. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In *Proceedings of IEEE International Joint Conference on Neural Networks*, Hong Kong, China, pp. 1322-1328, 2008. DOI: 10.1109/IJCNN.2008.4633969.

[12] X. X. Wu, Y. G. He, J. J. Duan. A deep parallel diagnostic method for transformer dissolved gas analysis. *Applied Sciences*, vol. 10, no. 4, Article number 1329, 2020. DOI: 10.3390/app10041329.

[13] Q. W. Guo, Y. B. Li, Y. Song, D. C. Wang, W. Chen. Intelligent fault diagnosis method based on full 1-D convolutional generative adversarial network. *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 2044–2053, 2020. DOI: 10.1109/TII.2019.2934901.

[14] Y. Zhuo, Z. Q. Ge. Auxiliary information-guided industrial data augmentation for any-shot fault learning and diagnosis. *IEEE Transactions on Industrial Informatics*, vol. 17, no. 11, pp. 7535–7545, 2021. DOI: 10.1109/TII.2021.3053106.

[15] A. Odena, C. Olah, J. Shlens. Conditional image synthesis with auxiliary classifier GANs. In *Proceedings of the 34th International Conference on Machine Learning*, Sydney, Australia, pp. 2642–2651, 2017.

[16] Z. Y. Wu, W. F. Lin, Y. Ji. An integrated ensemble learning model for imbalanced fault diagnostics and prognostics. *IEEE Access*, vol. 6, pp. 8394–8402, 2018. DOI: 10.1109/ACCESS.2018.2807121.

[17] L. J. Zhou, J. W. Dang, Z. H. Zhang. Fault classification for on-board equipment of high-speed railway based on attention capsule network. *International Journal of Automation and Computing*, vol. 18, no. 5, pp. 814–825, 2021. DOI: 10.1007/s11633-021-1291-2.

[18] Z. X. Hu, P. Jiang. An imbalance modified deep neural network with dynamical incremental learning for chemical fault diagnosis. *IEEE Transactions on Industrial Electronics*, vol. 66, no. 1, pp. 540–550, 2019. DOI: 10.1109/TIE.2018.2798633.

[19] W. K. Yu, C. H. Zhao. Broad convolutional neural network based industrial process fault diagnosis with incremental learning capability. *IEEE Transactions on Indus-*

trial Electronics, vol. 67, no. 6, pp. 5081–5091, 2020. DOI: 10.1109/TIE.2019.2931255.

[20] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, S. Wermter. Continual lifelong learning with neural networks: A review. *Neural Networks*, vol. 113, pp. 54–71, 2019. DOI: 10.1016/j.neunet.2019.01.012.

[21] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, F. F. Li, ImageNet: A large-scale hierarchical image database. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Miami, USA, pp. 248–255, 2009. DOI: 10.1109/CVPR.2009.5206848.

[22] J. Xu, P. F. Xu, Z. C. Wei, X. Ding, L. Shi. DC-NNMN: Across components fault diagnosis based on deep few-shot learning. *Shock and Vibration*, vol. 2020, Article number 3152174, 2020. DOI: 10.1155/2020/3152174.

[23] C. J. Li, S. B. Li, A. S. Zhang, Q. He, Z. H. Liao, J. J. Hu. Meta-learning for few-shot bearing fault diagnosis under complex working conditions. *Neurocomputing*, vol. 439, pp. 197–211, 2021. DOI: 10.1016/j.neucom.2021.01.099.

[24] A. S. Zhang, S. B. Li, Y. X. Cui, W. L. Yang, R. Z. Dong, J. J. Hu. Limited data rolling bearing fault diagnosis with few-shot learning. *IEEE Access*, vol. 7, pp. 110895–110904, 2019. DOI: 10.1109/ACCESS.2019.2934233.

[25] N. Lu, H. Y. Hu, T. Yin, Y. G. Lei, S. H. Wang. Transfer relation network for fault diagnosis of rotating machinery with small data. *IEEE Transactions on Cybernetics*, to be published. DOI: 10.1109/TCYB.2021.3085476.

[26] C. Finn, P. Abbeel, S. Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning*, Sydney, Australia, pp. 1126–1135, 2017.

[27] D. Wang, M. Zhang, Y. C. Xu, W. N. Lu, J. Yang, T. Zhang. Metric-based meta-learning model for few-shot fault diagnosis under multiple limited data conditions. *Mechanical Systems and Signal Processing*, vol. 155, Article number 107510, 2021. DOI: 10.1016/j.ymssp.2020.107510.

[28] J. Snell, K. Swersky, R. Zemel. Prototypical networks for few-shot learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, USA, pp. 4080–4090, 2017.

[29] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, D. Wierstra. Matching networks for one shot learning. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, Barcelona, Spain, pp. 3637–3645, 2016.

[30] G. Koch, R. Zemel, R. Salakhutdinov. Siamese neural networks for one-shot image recognition. In *Proceedings of the 32nd International Conference on Machine Learning*, Lille, France, 2015.

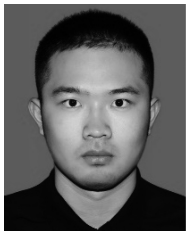
[31] D. W. Li, Y. J. Tian. Survey and experimental study on metric learning methods. *Neural Networks*, vol. 105, pp. 447–462, 2018. DOI: 10.1016/j.neunet.2018.06.003.

[32] F. Schroff, D. Kalenichenko, J. Philbin. FaceNet: A unified embedding for face recognition and clustering. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Boston, USA, pp. 815-823, 2015. DOI: 10.1109/CVPR.2015.7298682.

[33] J. J. Downs, E. F. Vogel. A plant-wide industrial process control problem. *Computers & Chemical Engineering*, vol. 17, no. 3, pp. 245–255, 1993. DOI: 10.1016/0098-1354(93)80018-I.

[34] A. Bathelt, N. L. Ricker, M. Jelali. Revision of the Tennessee Eastman process model. *IFAC-PapersOnLine*, vol. 48, no. 8, pp. 309–314, 2015. DOI: 10.1016/j.ifacol.2015.08.199.

[35] L. Van der Maaten, G. Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008.



Xinyao Xu received the B. Sc. degree in automation from Tianjin University, China in 2018. He is currently a Ph. D. degree candidate in control science and engineering at Research Center of Precision Sensing and Control, Institute of Automation, Chinese Academy of Sciences (CASIA), and School of Artificial Intelligence, University of Chinese Academy of Sciences (UCAS), China.

His research interests include deep learning, industrial fault detection, and fault diagnosis.

E-mail: xuxinyao2018@ia.ac.cn

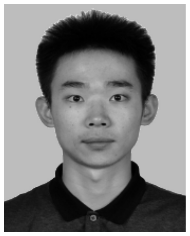
ORCID iD: 0000-0002-6371-1948

De Xu received the B. Sc. and M. Sc. degrees from Shandong University of Technology, China in 1985 and 1990, respectively, and the Ph. D. degree from Zhejiang University, China in 2001, all in control science and engineering. He has been with CASIA since 2001. He is currently a professor with Research Center of Precision Sensing and Control, CASIA, China. He is also with School of Artificial Intelligence, UCAS, China.

His research interests include robotics and automation such as visual measurement, visual control, intelligent control, visual positioning, microscopic vision, micro-assembly, and skill learning.

E-mail: de.xu@ia.ac.cn (Corresponding author)

ORCID iD: 0000-0002-7221-1654



Fangbo Qin received the B. Sc. degree in automation from Beijing Jiaotong University, China in 2013, the Ph. D. degree in control science and engineering from CASIA and UCAS, China in 2019. He is currently an associate professor with Research Center of Precision Sensing and Control, CASIA, China.

His research interests include robot vision, robot manipulation, and deep learning.

E-mail: qinfangbo2013@ia.ac.cn

ORCID iD: 0000-0002-4085-0857