2024 IEEE 20th International Conference on Automation
Science and Engineering (CASE)
August 28 - September 1, 2024. Bari, Italy

2024 IEEE 20th International Conference on Automation Science and Engineering (CASE) | 979-8-3503-5851-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/CASE59546.2024.10711424

Open-Set Fault Diagnosis based on Prototype Learning with Dual Category-Classifier

Shenqiang Ke, Liang Gao, *IEEE Senior Member*, Xinyu Li, *IEEE Member*, Yiping Gao, *IEEE Member*

Abstract— **Fault diagnosis is one of the most essential parts of industry quality inspection. Recent years have seen great success in Deep-Learning (DL) based fault diagnosis methods. However, traditional supervised DL methods only aim at reducing the empirical risk but ignore the open space risk. Thus, when applied in the open-set scenario, these methods cannot recognize the unknown faults that have not occurred yet. To tackle this issue, this paper develops a novel dual category-classifier open-set framework to reduce both the empirical classification risk and the open space risk. In this framework, a reciprocal point and a prototypical point are deployed for each known category in the training phase. The reciprocal points are optimized by the extra-class space to enlarge the between-class distance, while the prototypical points are optimized to compress the within-class distance. Moreover, a margin constraint term is added for further restricting the distribution range. Finally, a novel-categories detector named Kernel Null Foley-Sammon Transform (KNFST) is adopted to reject the unknown fault modes. Computational experiments conducted on CWRU dataset, and a practical bearing dataset shows that the proposed method can successfully deal with the open-set fault diagnosis problem and achieve a remarkable improvement compared with most state-of-the-art methods.**

I. Introduction

Intelligent fault diagnosis methods have been successfully developed and extensively applied to practical scenario over the years. Among them, the data-driven methods, especially deep-learning (DL) based, have gained increasing attention. DL methods, including convolutional neural networks [1] (CNNs), recurrent neural networks [2] (RNNs), and graph neural networks [3] (GNNs), automatically extract high-level feature representations from original data and perform highly accurate fault diagnoses.

However, the excellent performance of DL-based methods depends on the restricting assumption that the training data and the testing data must share an identical label space, which deviates from most real-world scenarios. For instance, fault modes are unpredictable and dynamic for variations in working conditions and equipment degradation. It is incredibly costly or even impossible to consider all the fault modes in the training stage. In this case, the testing categories unseen in the training phase will be misclassified as the known categories, leading to detrimental effects of the diagnosis methods [4]. Hence, intelligent fault diagnosis systems need to classify known fault modes and identify unknown fault modes simultaneously, which is viewed as an open-set problem. And Fig. 1 illustrates the open-set fault diagnosis (OSFD) scenario.

Fig. 1. Illustration of OSFD. The left subfigure shows two cases of label incompleteness in the training phase. The right subfigure presents the misclassification of unknown labels in the testing phase.

There are some existing works about open-set fault diagnosis. Tian *et al.* [5] designed a novelty detector to recognize unknown fault modes by kernel null Foley-Sammon transform with a self-adaptive threshold based on subspace learning. Razavi-Far *et al.* [6] proposed an open-set diagnosis framework based on Extreme Verification Latency that contains a double-stage detector to detect abrupt changes and determine the novel category. Wang *et al.* [7] introduced metric learning into open-set fault diagnosis and developed a novel loss function with an adaptive angle margin to learn and recognize new faults. Peng *et al.* [8] proposed an open-set fault diagnosis framework using supervised contrastive learning to generate the negative out-of-distribution data as unknown faults by Soft Brownian Offset. Moreover, Zhang et al. [9] utilized adversarial learning and an instance-level weighted mechanism to the open-set fault diagnosis combined with the domain-shift problem. Yu *et al.* [10] proposed a bilateral weighted adversarial network (BWAN) model to reweight the shared and outlier classes as well as reject unknown-class samples based on Extreme Value Theory. Zhao and Shen [11] adopted an auxiliary domain discriminator to further assign similarity weights and restrict the overlap of the known and unknown classes, learning accurate hyperplanes. Compared to closed-set methods, these methods can recognize unknown classes but have considerable room for improvement in feature extraction and fusion.

Although the research in current literature can alleviate the open-set fault diagnosis challenges to some extent, big research gaps still exist. The performances of related works highly depend on the feature representations learned for known classes in open-set recognition. On the one hand, it is a challenging step to design neural models for learning a good representation that can separate different known classes. On

This work was supported by the National Key R&D Program of China under Grant No.2022YFB3302700.

Shenqiang Ke, Liang Gao, Xinyu Li, Yiping Gao are with the State Key Laboratory of Digital Manufacturing Equipment and Technology, School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan, China. Their emails are: ksq@hust.edu.cn; gaoliang@mail.hust.edu.cn; lixinyu@mail.hust.edu.cn; gaoyiping@hust.edu.cn; Corresponding author: *Yiping Gao*.

the other hand, the learned representations based on the softmax function will force embedding features to fall into several half-open spaces divided by hyperplanes in the whole feature space, leaving no space for unknown class samples [4]. The key to open-set fault diagnosis is to make the distribution of unknown classes discriminative enough from that of known classes, avoiding the overlap of embedding features between known and unknown classes [12]. In other words, it is decisive to simultaneously reduce the empirical classification risk on labeled-known data and the open risk on labeled-unknown data.

In this paper, a novel open-set fault diagnosis framework called Dual Category-Classifier Network (DCCN) is proposed to achieve the reduction mentioned above. The DCCN framework is mainly based on prototype learning and designs a reciprocal point and a prototypical point for each known category in the training phase. The embedding features of the training data should be far away from the corresponding reciprocal point to reduce the empirical classification risk and nearby the corresponding prototypical point to reduce the open risk on labeled-unknown data. A margin constraint term is also added for further restricting the distribution range. Finally, a novel-categories detector named Kernel Null Foley-Sammon Transform (KNFST) [13] is adopted to distinguish the known and unknown fault categories. Experiments have been conducted on CWRU dataset, and a practical bearing dataset and results show that the proposed method can successfully recognize the known and unknown fault categories and achieve a remarkable improvement compared with most state-of-the-art methods.

II. Preliminaries

A. Problem Definition

In real scenarios, new fault modes may occur owing to the uncertainty of dynamic working conditions and equipment degradation, contributing to the incompleteness of the label space of the training data. Given a set of training data $D_{train} = \{(x_1^{tr}, y_1^{tr}), (x_2^{tr}, y_2^{tr}), \cdots, (x_n^{tr}, y_n^{tr})\}$ with N known categories. $y_i^{tr} \in \{1, 2, \cdots, N\}$ is the label of x_i^{tr} which is a k-dimensional vector. If the potential category doesn't belong to any known categories, it's considered to $N+1$ category. It's likely that the data of $N+1$ category may come from different classes in fact, but their specific classes are not significant to open-set recognition. The testing data D_{test} is denoted as $D_{test} = \{(x_1^{te}, y_1^{te}), (x_2^{te}, y_2^{te}), \cdots, (x_n^{te}, y_n^{te})\}$ and $y_i^{te} \in \{1,2, \cdots, N+1\}$ is the label of x_i^{te}. It's assumed that the testing data contains U unknown categories to describe the category gap between the training and testing data clearly. An indicator called openness is defined as follows:

$$\mathcal{O}_{openness} = 1 - \sqrt{\frac{2N}{N + (N+U)}} \tag{1}$$

It's noticed that larger openness leads to more difficult open problems. The problem is completed close to the zero-shot problem when the openness equals 1 and near to the closed-set problem when the openness equals 0. The goal of open-set recognition is to classify the known categories and reject the unknown categories correctly. In the training phase, it can be regarded as to determine a recognition function $f \in \mathfrak{H}$ to minimize the empirical classification risk and open space risk simultaneously. Hence, it can be formulated as

$$\arg\min_f \{\mathcal{R}_\mathcal{E}(D_{train}, f) + \mathcal{R}_O(D_{test}, f)\} \tag{2}$$

where $\mathcal{R}_\mathcal{E}$ denotes the empirical risk and $\mathcal{R}_O$ denotes the open risk.

B. Prototype Learning for Open-set Recognition

Prototype learning usually uses the prototype to refer to one or more points that can represent the cluster. The most known prototype model is learning vector quantization (LVQ). The model [14] learned one or more prototypes for each class to gain a better representation and make different classes more distinguishable. Numerous works derived from LVQ have been proposed for OSR tasks. Chen *et al.* [15] proposed an adversarial reciprocal point learning (ARPL) framework to generate confusing training samples and restrict the overlap of known and unknown distributions. Yang *et al.* [16] developed several discriminative losses and a generative loss for the training phase, whereas two rejection rules were used to detect the unknown classes. These methods have improved the effectiveness of open-set recognition because the prototype is learned with the objective of minimizing the within-class distance largely.

As for a prototype center O^k for a certain category k, the distance between original data x and the prototype center O^k can be measured in the following two metrics:

$$d_e(\Theta(x), O^k) = \frac{1}{m}\|\Theta(x) - O^k\|_2^2 \tag{3}$$

$$d_d(\Theta(x), O^k) = \Theta(x) \cdot O^k \tag{4}$$

where $\Theta(x)$ means the embedding feature of original data x and m denotes the dimension of $\Theta(x)$ and O^k. $d_e(\Theta(x), O^k)$ is the Euclidean distance that gauge the distance in the feature space, while $d_d(\Theta(x), O^k)$ mainly measures the angle distance.

III. Proposed DCCN Method for OSFD

In this section, the proposed OSFD framework DCCN is described in detail. Reciprocal and prototypical points are utilized to learn a more discriminative representation that ensures within-class compactness and between-class dispersion in the feature space. Furthermore, a novel-category detector is proposed to reject the unknown classes by comparing the novelty scores and the self-adaptive threshold.

A. General Framework of The Proposed Method

The proposed method consists of a feature extractor *E*, a known-categories classifier *C*, and a novel-categories detector *D*, with the parameters of θ, θ_C and θ_D respectively as shown in Fig. 2.

The feature extractor *E* employs a one-dimensional neural network ResNet18 [17] to automatically learn a better and deeper feature representation from original data.

The known-categories classifier *C* aims to maximize the between-class distance and minimize the within-class distance with the guarantee of correct recognition of known categories.

The novel-categories detector *D* is utilized to distinguish known categories and unknown categories by KNFST model. The minimum Euclidean distance between the point mapped

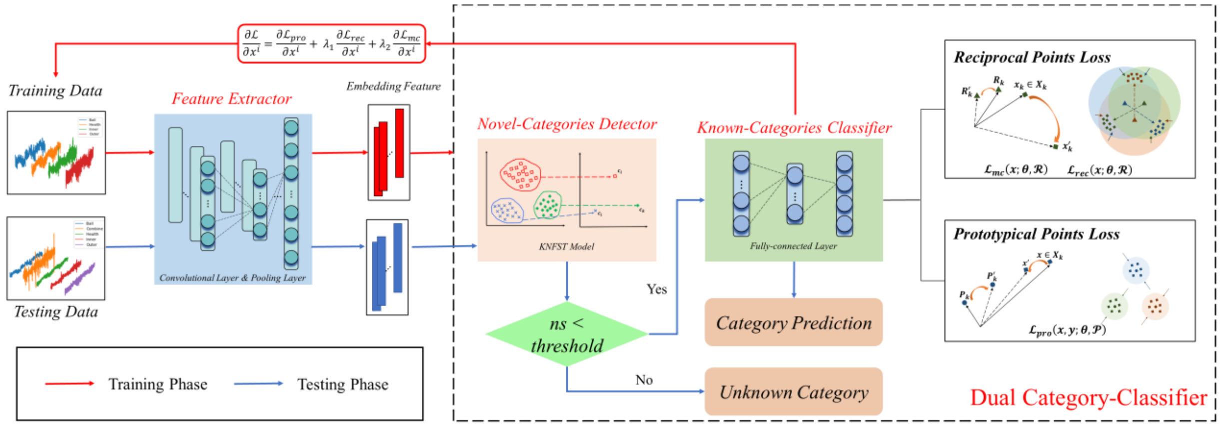

Fig. 2. Main Framework of the proposed Dual Category-Classifier Network (DCCN) for Open-Set Fault Diagnosis problem. The proposed DCCN method utilizes a dual-classifier learning paradigm. Prototype learning is introduced to the known classifier to ensure within-class compactness and between-class dispersion. Actually, the novelty detector KNFST can be recognized as prototype learning to some extent. The points in the null-space are the prototypes for the data in the feature space.

with testing data and known-categories points can be regarded as the probability it belongs to unknown categories, which is defined as novelty score.

B. The Known-Categories Classifier

Embedding features of original data are gained by the trained feature extractor E and are supposed to be deeper and better representations with good category separability. Then features are input into the known-categories class C, which mainly comprises fully-connected layers.

1) Prototypical Points Loss

The prototypical points, $\mathcal{P} = \{P^k, k = 1,2,\cdots N\}$, are learnable and initialized randomly by Gaussian distribution, aiming to minimize the within-class distance by letting samples surround the corresponding prototypical points. It's used as a criterion for classifier training. Loss function can be defined as

$$\mathcal{L}_{pro}(x, y; \theta_{\mathcal{P}}) = d_e(x, P^y) \tag{5}$$

where x and y denote the embedding feature and label of original data respectively. The optimization is converged only if the embedding feature x is very close to the corresponding prototypical point P^y through minimizing (5). In this case, within-class distances are reduced notably.

2) Reciprocal Points Loss

However, since the Euclidean distance has a limited effect on reducing the empirical classification risk that it only focuses on the samples from same category, reciprocal points, defined as $\mathcal{R} = \{R^k, k = 1,2,\cdots N\}$, are utilized to further reduce the empirical classification risk and the open space risk.

Assume that S^k denotes the subspace that represents the data belonging to category k and its corresponding open space is defined as $S^{\neq k}$. For each training sample, it is supposed to be further away from the corresponding reciprocal point than the samples from different categories, which can be formulated as

$$\max(\xi(S^{\neq k}, R^k)) \leq \min(\xi(S^k, R^k)) \tag{6}$$

where $\xi(\cdot)$ is a distance function that calculates the distance between embedding features and reciprocal points. Given a sample x and reciprocal point R^k, $\xi(y = k|x, R^k)$ is defined as

$$\xi(y = k|x, R^k) = d_e(x, R^k) - d_d(x, R^k) \tag{7}$$

From the nature of reciprocal points, the distance between sample x and reciprocal point R^k is proportional to the probability that sample x belongs to category k. A greater distance leads to a larger probability that assigns the sample to label k. Drawing on the idea of cross-entropy loss, reciprocal points loss is defined as

$$\mathcal{L}_{rec}(x, y; \theta, \mathcal{R}) = -\log \frac{e^{\xi(x, R^y)}}{\sum_{i=1}^{N} e^{\xi(x, R^i)}} \tag{8}$$

3) Margin Constraint Loss

Equation (6) only maximizes the interval between the subspace S^k and the corresponding open space $S^{\neq k}$ but ignores the constraint of $S^{\neq k}$. Hence, there will be an estimable overlap between S^k and $S^{\neq k}$. Margin constraint term is proposed to further reduce the open space risk by separating S^k and $S^{\neq k}$ as much as possible, compressing the distribution of $S^{\neq k}$ in a limited range by (7); thus, the total open space risk is restricted.

$$\max(\xi(S^{\neq k}, R^k)) \leq O \tag{9}$$

It is almost impossible to limit all the samples in the open space in a certain range cause $S^{\neq k}$ has many unknown categories samples. However, the open space risk can be restricted by constraining the distance between the subspace S^k and R^k, which can be expressed as

$$\mathcal{L}_{mc}(x, y; \theta, O) = \max(d_e(x, R^k) - O, 0) \tag{10}$$

where O is a learnable parameter and initialized with a value of 0.

C. The Novel-Categories Detector

The novel-categories detector D is adopted to distinguish the known and unknown categories via KNFST model. Numerous works [18] have used KNFST model as novelty detector for the open-set problem and the outlier detection problem.

KNFST is a modified version of Foley-Sammon Transform whose goal is to determine a discriminant space where data from the same categories will be mapped into one point and data from different categories will be mapped into separated points.

Assume that X_j is the set of data belonging to class j and N_j is the sample number of class j. Let μ_j and μ denote the mean of class j and all samples, respectively. A kernel function $\Phi: x \longrightarrow \Phi(x)$ is utilized to map X into a feature space $\mathcal{H}$ to extend the linear model to the nonlinear case. The goal can be defined as

$$\arg\max_{\psi^T S_w^\Phi \psi = 0} \psi^T S_b^\Phi \psi \tag{11}$$

$$s.t. \begin{cases} S_b^\Phi = \sum_{j=1}^{N} N_j (\mu_j^\Phi - \mu^\Phi)(\mu_j^\Phi - \mu^\Phi)^T \\ S_w^\Phi = \sum_{j=1}^{N} \sum_{x \in X_j} (\Phi(x) - \mu_j^\Phi)(\Phi(x) - \mu_j^\Phi)^T \\ \mu_j^\Phi = \frac{1}{N_j} \sum_{x \in X_j} \Phi(x), \mu^\Phi = \frac{1}{N} \sum_{j=1}^{N} \sum_{x \in X_j} \Phi(x) \end{cases}$$

D. Training Procedure and Open-Set Fault Diagnosis

Based on the description above, a dual category-classifier framework is utilized for category prediction and novelty detection. In the known-categories classifier, feature extractor *E* and classifier *C* are trained by the training data with the joint losses:

$$\mathcal{L} = \mathcal{L}_{pro}(x, y; \theta_{\mathcal{P}}) + \lambda_1 \mathcal{L}_{rec}(x, y; \theta_{\mathcal{R}}) + \lambda_2 \ell_{mc}(x, y; \theta_O) \tag{12}$$

where $\lambda_1 \in (0,1)$ and $\lambda_2 \in (0,1)$ are the trade-off coefficients to balance the loss terms.

For each testing data, if it belongs to the known categories, it will ideally be mapped into a point of target points. Otherwise, it will be mapped into a point far away from all the target points. So, the minimum distance in the null space between the point mapped with testing data and target points can be regarded as the novelty score. If a sample gains a high novelty score, it will most likely belong to unknown categories. However, data are variable from others even though they are from the same categories, which leads to samples from known categories having novelty scores greater than 0 as well.

In this case, a threshold is needed to distinguish the samples from known or unknown categories. Once the novelty score exceeds the threshold, the sample will be detected as an unknown category, or it will be predicted as one of the known categories by the known-categories classifier. As the magnitude of Euclidean distance varies significantly with extracted feature variation, a static threshold can only play a limited effect. This paper divides the training data into two datasets to address this problem. One is utilized to establish and train the KNFST model, and the threshold is set as the maximal novelty score among the others.

IV. Experimental Study

A. Dataset Description

1) CWRU Dataset

CWRU bearing dataset was obtained from the electrical engineering laboratory of Case Western Reserve University, USA. In this paper, vibration signals of various healthy states with the sampling frequency of 12 kHz at the fan end are adopted. Each fault type contains different damage diameters, such as 0.007, 0.014, and 0.021 inches by the electric spark. Healthy states include healthy bearing, inner race faulty bearing, outer race faulty bearing, and rolling faulty bearing.

2) HUST Dataset

HUST dataset is gained from our own bearing experimental platform, which contains bearing faults and gearbox faults. The platform consists of AC motor, AC drive, bearings, gearbox, loader, and sensors. Bearings contain five different health states: (1) healthy; (2) inner race fault; (3) outer race fault; (4) ball fault; (5) combined fault. Moreover, the gearboxes have three condition states: (1) Normal; (2) Missing Gear; (3) Broken Gear. Vibration signals are sampled at 25.6 kHz.

Each health state contains 300 samples for the two datasets, 250 of which are used for training and the others for testing. And each sample consists of 1024 data points. Considering the two cases of label incompleteness, eight tasks are designed to demonstrate the effectiveness of the proposed method. The detailed information and diagnosis tasks of the two datasets are presented in TABLE III.

TABLE III. Information and Diagnosis Tasks of Two Datasets

Dataset	Class	0	1	2	3	4	5	6	7	8
CWRU	Type	H	IF	BF	OF	IF	BF	OF	IF	BF
	Size		7	7	7	14	14	14	21	21
HUST	Type	H	IF	BF	OF	CF	NG	MG	BG	
	Part	B	B	B	B	B	G	G	G	

Dataset	Task	Speed	Training Label Set	Testing Label Set
CWRU	C_0	1797*rpm*	1,2,3,4,5,6,7,8,9	0,1,2,3,4,5,6,7,8,9
	C_1	1797*rpm*	0,1,2,3,4,5,6	0,1,2,3,4,5,6,7,8,9
	C_2	1772*rpm*	0,1,2,3,4,5	0,1,2,3,4,7,8
	C_3	1772*rpm*	0,1,3,5,7,9	0,1,2,3,4,5
HUST	H_0	70*HZ*	1,2,3,4,5,6,7	0,1,2,3,4,5,6,7
	H_1	70*HZ*	0,1,2,3,4	0,1,2,3,4,5,6,7
	H_2	80*HZ*	0,2,4,6	0,1,2,3,4,5
	H_3	80*HZ*	0,1,6,7	0,1,3,5,7

The unit of size in CWRU Dataset is *mil*. B and G denotes the fault is from bearing and gearbox. Due to layout constraints, CWRU Dataset included nine fault modes, with the 9th class representing inner race faults (fault size of 21 *mils*).

B. Experimental Details and Evaluation Metrics

The code of the DCCN method is implemented with PyTorch 1.10. All the experiments are performed on NVIDIA RTX 3090Ti GPU (with 24G memory) under CUDA 11.6 on Ubuntu 16.04 system. The backbone network used in experiments is 1-D ResNet18 without pre-trained. For the training phase, the batch size is set as 32, the number of total training iterations is set as 100, and the learning rate is set as 0.005 with a decay of 0.001.

To demonstrate the proposed framework thoroughly, three evaluation metrics are utilized by previous works. Metrics KNO and UNK are defined to evaluate the accuracy of the known and unknown classes in the testing data, respectively. $H-score$ is used to comprehensively evaluate the methods that its value will be high if and only if both KNO and UNK are high. The formulation is expressed as

$$H-score = 2 \times \frac{KNO \times UNK}{KNO + UNK} \tag{13}$$

C. Comparison with State-of-the-art Methods

This section presents the comparison results between the proposed methods and four state-of-the-art methods as well as a close-set baseline in the open-set recognition, including Softmax with a threshold [19], EVT [10], CROSR [20], Center Loss [12] and GCPLoss [16]. The latter two methods is also based on prototype learning that have not yet been utilized in the field of OSFD. All methods share the same network architecture and hyper-parameters. The average results are evaluated after five trials.

The experimental results on two datasets are presented in TABLE IV and TABLE V. It can be seen that the conventional close-set method based on softmax loss has a limited performance on open-set recognition problem, which can only recognize the known faults but cannot detect the unknown faults. For each diagnosis task, the metric *UNK* of the proposed method is more than 98%, which can largely demonstrate that our proposed method has extremely good efficiency in recognizing and detecting new unknown classes. In addition, it can be seen obviously the average KNO of the proposed method is also the highest score 98.60%, which is 1.39% higher than that of the next best method. Comparing the result of ResNet18 with EVT, the average *H-score* on all tasks of the proposed is 99.15%, achieving an average improvement of 4.52% on the all tasks and an improvement of 6.09% on the task H_3. These results strongly demonstrate that the proposed method overperforms most state-of-the-art methods in both classifying known classes and rejecting unknown classes.

D. Visualization of Learned Feature

Fig. 3 shows the distribution of each class in the feature space learned by different open-set methods on task C_1. From Fig. 3(a)-(b), we can see that conventional method without prototype learning have a minimal effect on the unknown classes that the samples from the unknown classes are not well clustered together as those from the known classes. The methods based on prototype learning can reduce the overlap between the known and unknown classes by compressing the distribution range of each class shown in Fig. 3(c)-(d). However, their performance can be further improved, for within-class distance can be still reduced and between-class distance can be enlarged. By contrast, the feature distribution of each class learned by the proposed method is restricted largely that the overlaps between different classes are eliminated. In this way, the empirical classification and the open-space risk are simultaneously reduced, leading to excellent performance of the proposed method. Moreover, we further visualize the distribution of each class in both training and testing data on task H_3 as shown in Fig. 3(f)-(j).

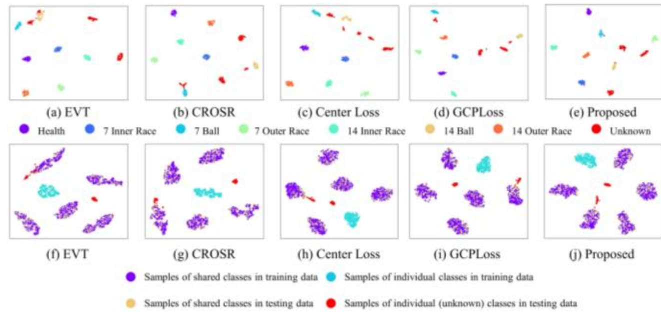

Fig. 3. Visualization. (a)-(e) Visualization of learned features by different open-set methods via t-SNE technology on task C_1. (f)-(j) Visualization of learned features by different open-set methods via t-SNE technology on task C_1.

Compared to the other methods, the samples from the shared-classes are clustered compactly and the samples from the individual-classes are separated away from the samples from the other classes. In addition, for the proposed method, the distribution range of the share-classes in the training data is limited to a maximum extent. These results strongly prove that the proposed method can maximize the between-class distance as well as minimize the within-class distance.

E. Ablation Study

This section explores the contributions from each component of the proposed method, consists of three loss functions and the novelty detector.

1) The Influence of the Joint Losses

$\mathcal{L}_{pro}$, $\mathcal{L}_{rec}$ and $\mathcal{L}_{mc}$ play different and crucial roles in the classifier. As shown in TABLE VI, without the joint losses, the benchmark method can only reach 96.37% and 92.86% respectively on metric *H-score*. When only applying $\mathcal{L}_{pro}$, *H-score* is improved by 1.49% and 2.51%. For the reciprocal points loss, the separate work of $\mathcal{L}_{rec}$ achieves 99.04% and

TABLE VI. Results of the Ablation Study

$\mathcal{L}_{rec}$	$\mathcal{L}_{mc}$	$\mathcal{L}_{pro}$	C_0 KNO	C_0 UNK	C_0 H-score	H_3 KNO	H_3 UNK	H_3 H-score
			100	92.66	96.19	**99.66**	85.87	92.25
		✓	98.00	97.80	97.90	92.33	98.80	95.46
✓		✓	98.73	**100**	99.36	95.39	99.20	97.26
✓	✓		99.24	98.00	98.62	96.80	98.84	97.81
✓	✓	✓	99.87	**100**	**99.93**	96.17	**100**	**98.05**

TABLE IV. The Diagnosis Result of CWRU Dataset

Method		H_0 KNO	H_0 UNK	H_0 H-score	H_1 KNO	H_1 UNK	H_1 H-score	H_2 KNO	H_2 UNK	H_2 H-score	H_3 KNO	H_3 UNK	H_3 H-score
Close-set	Softmax	99.29	76.80	86.81	100	63.87	77.95	100	29.20	45.20	100	7.00	13.08
Non-prototype Learning	EVT	97.64	91.60	94.52	97.89	91.47	94.57	97.28	93.00	95.09	97.33	98.80	98.06
	CROSR	96.17	96.95	96.56	96.78	96.40	96.59	96.42	97.81	97.11	96.80	97.64	97.22
Prototype Learning	Center Loss	97.80	**100**	98.89	**99.20**	**100**	**99.60**	97.84	96.80	97.32	98.33	99.00	98.66
	GCPLoss	98.13	**100**	99.06	96.84	99.73	98.26	98.00	97.80	97.90	99.73	**100**	99.86
	Ours (DCCN)	**99.87**	**100**	**99.93**	98.46	**100**	99.22	**99.52**	**98.20**	**98.86**	**99.87**	**100**	**99.98**

TABLE V. The Diagnosis Result of HUST Dataset

Method		H_0 KNO	H_0 UNK	H_0 H-score	H_1 KNO	H_1 UNK	H_1 H-score	H_2 KNO	H_2 UNK	H_2 H-score	H_3 KNO	H_3 UNK	H_3 H-score
Close-set	Softmax	100	73.20	84.53	100	42.66	59.81	100	53.73	69.90	100	60.20	75.16
Non-prototype Learning	EVT	95.78	96.80	96.29	97.20	90.67	93.82	97.52	89.80	93.50	94.40	88.62	91.42
	CROSR	98.67	96.04	97.34	96.55	93.28	94.89	96.86	94.18	95.50	94.66	93.08	93.86
Prototype Learning	Center Loss	96.80	**100**	98.37	**97.42**	98.69	98.05	95.33	98.80	97.03	91.80	97.52	94.57
	GCPLoss	97.00	98.80	97.89	95.30	97.26	96.27	96.26	98.50	97.37	**96.40**	99.00	97.68
	Ours (DCCN)	**99.73**	**100**	**99.86**	97.33	**99.00**	**98.16**	**97.86**	**100**	**98.92**	96.17	**100**	**98.05**

97.26% respectively on metric *H-score*. While adding all three losses, improvements can be seen on all the metric, obtaining a highest score of 99.93% and 98.04% on metric *H-score*, which indicates that a weighting combination of three losses brings a better result than a single addition of either one.

2) The Influence of the Threshold in KNFST

A threshold is set to distinguish the known and unknown classes. The threshold is learned as the maximal novelty score among the validation data. Let α denotes the proportion that the threshold is set as the novelty score of the top- α sample. If $\alpha > 1$, it means the times of the maximal novelty score.

Fig. 4 shows the results on task H_0 with varying α. It can be observed that the metric KNO decreases and UNK increases when using larger α. This is due to that larger α is more prone to rejecting the samples as the unknown

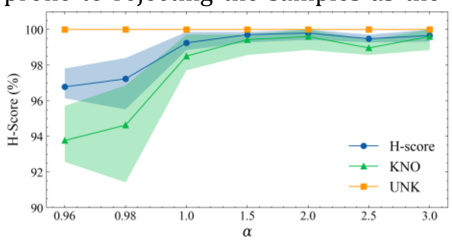

Fig.4. The performance of different thresholds for the novel-categories detector on task H_0. The shaded parts indicate the variances of the proposed method.

classes. All the metrics are high when α is set to 2.0 which is also recommended within online applications.

V. Conclusion

This paper proposes a dual category-classifier framework to address open-set fault diagnosis by using prototype learning. Feature representations are learned by the joint work of prototypical points loss and reciprocal points loss, which minimizes the within-class distance and maximizes the between-class distance. A margin constraint term is utilized to further compress the distribution range of the known classes in the feature space. Moreover, an extra novelty detector is used to recognize and reject the unknown classes. As demonstrated in the experiments, the proposed method overperforms the four state-of-the-art methods.

The limitations of the proposed DCCN method include the following aspects. First, the proposed method is based on the stable signals, but there are more time-varying signals in the realistic scenarios. Second, this paper only focus on the label incompleteness but ignores the data distribution shift due to the varying working conditions. Thus, two directions can be explored to take a step further towards the practicality. One is to develop more accurate algorithms aiming to complex scenarios. Another one is to apply transfer learning methods such as domain generalization to open-set recognition tasks.

References

[1] Y. Gao, L. Gao, X. Li and S. Cao, "A Hierarchical Training-Convolutional Neural Network for Imbalanced Fault Diagnosis in Complex Equipment," *IEEE Trans. Industr. Inform*, vol. 18, no. 11, pp. 8138-8145, Nov. 2022, doi: 10.1109/TII.2022.3177662.

[2] M. Xia, H. Shao, X. Ma, and C. W. de Silva, "A Stacked GRU-RNN-Based Approach for Predicting Renewable Energy and Electricity Load for Smart Grid Operation," *IEEE Trans. Industr. Inform.*, vol. 17, no. 10, pp. 7050–7059, Oct. 2021, doi: 10.1109/TII.2021.3056867.

[3] Y. Han, Q. Li, C. Wang, and Q. Zhao, "A novel knowledge enhanced graph neural networks for fault diagnosis with application to blast furnace process safety," *Process Safety Environ. Protection*, vol. 166, pp. 143–157, Oct. 2022, doi: 10.1016/j.psep.2022.08.014..

[4] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward Open Set Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1757–1772, Jul. 2013, doi: 10.1109/TPAMI.2012.256.

[5] Y. Tian, Z. Wang, L. Zhang, C. Lu, and J. Ma, "A subspace learning-based feature fusion and open-set fault diagnosis approach for machinery components," *Adv. Eng. Inf.*, vol. 36, pp. 194–206, Apr. 2018, doi: 10.1016/j.aei.2018.04.006.

[6] R. Razavi-Far, E. Hallaji, M. Saif, and G. Ditzler, "A Novelty Detector and Extreme Verification Latency Model for Nonstationary Environments," *IEEE Trans. Ind. Electron.*, vol. 66, no. 1, pp. 561–570, Jan. 2019, doi: 10.1109/TIE.2018.2826477.

[7] C. Wang, C. Xin, and Z. Xu, "A novel deep metric learning model for imbalanced fault diagnosis and toward open-set classification," *Knowl Based Syst*, vol. 220, p. 106925, May 2021, doi: 10.1016/j.knosys.2021.106925.

[8] P. Peng, J. Lu, T. Xie, S. Tao, H. Wang, and H. Zhang, "Open-Set Fault Diagnosis via Supervised Contrastive Learning with Negative Out-of-Distribution Data Augmentation," *IEEE Trans. Industr. Inform.*, pp. 1–1, 2022, doi: 10.1109/TII.2022.3149935.

[9] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Open-Set Domain Adaptation in Machinery Fault Diagnostics Using Instance-Level Weighted Adversarial Learning," *IEEE Trans. Industr. Inform.*, vol. 17, no. 11, pp. 7445–7455, Nov. 2021, doi: 10.1109/TII.2021.3054651..

[10] X. Yu et al., "Deep-Learning-Based Open Set Fault Diagnosis by Extreme Value Theory," *IEEE Trans. Industr. Inform.*, vol. 18, no. 1, pp. 185–196, Jan. 2022, doi: 10.1109/TII.2021.3070324.

[11] C. Zhao and W. Shen, "Dual adversarial network for cross-domain open set fault diagnosis," *Reliab. Eng. Syst. Saf.*, vol. 221, p. 108358, May 2022, doi: 10.1016/j.ress.2022.108358.

[12] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A Discriminative Feature Learning Approach for Deep Face Recognition," in *Proc. Eur. Conf. Comput. Vis.*, Cham, 2016, pp. 499–515. doi: 10.1007/978-3-319-46478-7_31.

[13] Y. Lin, G. Gu, H. Liu, and J. Shen, "Kernel Null Foley-Sammon Transform," in 2008 International Conference on Computer Science and Software Engineering, Dec. 2008, vol. 1, pp. 981–984. doi: 10.1109/CSSE.2008.621.

[14] A. Sato and K. Yamada, "A formulation of learning vector quantization using a new misclassification measure," in *Proc. Fourteenth Int. Conf. Pattern Recognit.*, Aug. 1998, vol. 1, pp. 322–325 vol.1. doi: 10.1109/ICPR.1998.711145.

[15] G. Chen, P. Peng, X. Wang, and Y. Tian, "Adversarial Reciprocal Points Learning for Open Set Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2021, doi: 10.1109/TPAMI.2021.3106743.

[16] H.-M. Yang, X.-Y. Zhang, F. Yin, Q. Yang, and C.-L. Liu, "Convolutional Prototype Network for Open Set Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 5, pp. 2358–2370, May 2022, doi: 10.1109/TPAMI.2020.3045079.

[17] 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.

[18] F. Dufrenois and D. Hamad, "Sparse and online null proximal discriminant analysis for one class learning in large-scale datasets," in *Proc. Int. Jt. Conf. Neural Netw.*, Jul. 2019, pp. 1–8. doi: 10.1109/IJCNN.2019.8852312.

[19] A. Bendale and T. E. Boult, "Towards Open Set Deep Networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1563-1572

[20] R. Yoshihashi, W. Shao, R. Kawakami, S. You, M. Iida, and T. Naemura, "Classification-reconstruction learning for open-set recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4016–4025.