IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 22 May 2023, accepted 27 May 2023, date of publication 31 May 2023, date of current version 7 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3281559

RESEARCH ARTICLE

Discriminative Angle Feature Learning for Open-Set Deep Fault Classification

JIE MEI[1,2,3], WEI LIU[1], (Member, IEEE), MING ZHU[1,2,3], YONGKA QI[1,2,3], MING FU[1,2,3], YUSHI LI[1,2,3], AND QUAN YUAN[4]

[1]School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China
[2]National Engineering Research Center of Fire and Emergency Rescue, Wuhan 430074, China
[3]Hubei Province Internet Technology and Engineering Research and Development Center, Wuhan 430074, China
[4]School of Mechanical Engineering and Electronic Information, China University of Geosciences, Wuhan 430074, China

Corresponding author: Ming Zhu (zhuming@mail.hust.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62071189.

ABSTRACT Unknown faults may occur in practical applications, necessitating an open-set classifier that can classify known classes as well as recognize unknown faults. The current deep open-set classification methods are implicit in optimizing the intra- or inter-class distances, which may result in performance degradation when the number of unknown classes far exceeds that of the known. In this study, the discriminative angle deep features for vibration signals are investigated. A novel normalized one-versus-all classification loss with center and contrastive regularization is proposed. The trained network can explicitly optimize deep features to ensure intra-class compactness and inter-class divergence. In this case, such discriminative features can be used for open-set fault classification. Furthermore, the effectiveness of the proposed method is verified using field-measured motor bearing and gear vibration signals. The results demonstrate the evident advantages of our proposed method over other approaches in practical fault diagnosis scenarios.

INDEX TERMS Classification algorithms, convolutional neural networks, deep learning, detection algorithms, fault diagnosis, feature extraction, pattern recognition, rotating machines, supervised learning, vibration measurement.

I. INTRODUCTION

With the development of deep learning, the data-driven fault diagnosis of industrial equipment has attracted significant research attention in recent years [1], [2], [3]. Fault diagnosis based on vibration signals is particularly favored by researchers and engineers owing to the ease of signal collection on the surfaces of equipment without affecting operations [4], [5]. Recently, fault diagnosis techniques based on vibration signals and deep learning have been extensively studied for a variety of industrial equipment, including compressors [6], rotating machinery [7], and high-speed train [8]. Moreover, various intelligent diagnostic methods have been proposed recently. Wang et al. [9] presented a transformation method to convert vibration signals into images, and further adopted a two-dimensional convolutional neural network with multi-head attention to classify faults. Ding et al. [10] fed the time-frequency representation of vibration signals into a Transformer network for fault diagnosis. Tang et al. [11] developed a bi-directional deep belief network to improve the accuracy and generalization performance of the fault diagnosis model.

However, most intelligent fault diagnosis models assume a closed-set classification scenario. Under this assumption, the labels of all test samples are previously seen by the model during the training phase. The classification model divides the entire feature space into K subspaces, where K denotes the number of known classes, as depicted in Figure 1(a). Therefore, when a sample from an unknown class is fed into the model, it is erroneously classified as a specific known class. This does not match real application scenarios. Typically, an intelligent diagnosis model deployed in practical application scenarios should be capable of open-set recognition, as shown in Figure 1(b). Open-set classification models

The associate editor coordinating the review of this manuscript and approving it for publication was Gerard-Andre Capolino.

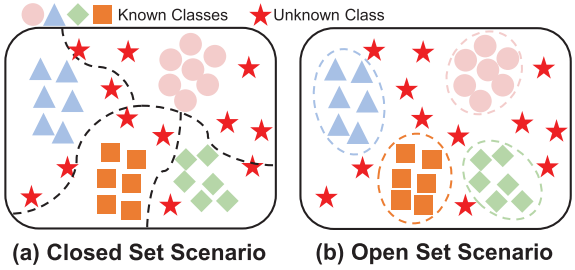


FIGURE 1. Comparison of decision boundaries in (a) closed-set and (b) open-set classification scenarios. The dashed lines indicate the decision boundaries and different shapes represent different classes. (a) Unknown samples are erroneously classified as a specific known class in the closed-set scenario. (b) Samples from known and unknown classes are correctly classified in the open-set scenario.

narrow down the decision boundaries of known classes so that the model can simultaneously classify samples from known classes and recognize unknown faults. However, actual industrial applications pose a challenging problem for the deployed fault diagnosis model, which is only trained on K known classes but outputs $K + 1$ classes, where $K + 1$ represents the extra unseen fault classes. In this study, the focus lies on establishing an open-set deep fault classification model for practical applications.

The open-set fault classification problem is primarily related to open-set recognition (OSR) [12] and open-set domain adaptation (OSDA) [13]. Generally, OSDA methods learn domain-invariant features between the source and target domains to realize the classification of shared classes while rejecting samples from unknown classes. Currently, most OSDA methods adopt the domain adversarial learning technique [14], which learns domain-invariant features through adversarial training of the feature extractor and domain classifier. Zhang et al. [15] proposed an instance-level weighted mechanism to determine the similarities of test samples with known classes for weighted domain adversarial learning. Mao et al. [16] designed an interactive dual adversarial neural network with a closed-set and weighted open-set domain adversarial network. Yu et al. [17] presented a bilateral weighted adversarial network that utilized the output of an additional domain classifier to obtain the similarities between the test samples and the known classes. Zhao and Shen [18] proposed a dual adversarial network with weighted and separated adversarial learning for cross-domain open-set fault diagnosis. However, these OSDA methods require sufficient test samples to participate in training, which does not satisfy the real-time diagnosis requirements of industrial applications.

The OSR model, which does not require any test samples to participate in training, is more suitable for field applications. Existing OSR methods enable the deep network to learn discriminative features that can be used for both the classification of known classes and the detection of unknown faults. A predefined feature distance is further used to determine the category of the input sample. Bendale et al. [19] proposed an OpenMax layer to generate extra unknown class activation and used the Euclidean distances for unknown detection. Chen et al. [20], [21] extracted Euclidean discriminative features using adversarial reciprocal point learning, where each reciprocal point represents the extra-class space corresponding to a known class. Similarly, Yu et al. [17] utilized the Euclidean distance between the output logit vector and class center to determine whether an input sample is an unknown fault. The variational auto-encoder (VAE) network can also be improved for OSR. He et al. [22] simultaneously trained a VAE network and deep latent feature classifier, utilizing Bhattacharyya distances of latent features for unknown detection. Chen et al. [23] directly replaced the decoder of the VAE network with a classifier and trained an encoder-classifier network for discriminative feature extraction. Zhang et al. [24] employed a flow-based model to estimate the density of known-class features and treated its output density as a special distance for unknown detection. Small intra-class variance and large inter-class separation are two important aspects of discriminative features. However, these methods are implicit in optimizing the intra- or inter-class distances, which may result in performance degradation when the number of unknown classes far exceeds that of the known.

Inspired by intelligent face recognition which has demonstrated considerable progress, discriminative angle feature learning is considered in this study. Wang et al. [25] added a cosine margin term to the L2 normalized SoftMax classification loss to force the network to learn deep features with large inter-class variance. Deng et al. [26] proposed an additive angle margin SoftMax loss for discriminative feature learning. Nevertheless, these face recognition methods do not explicitly optimize the intra-class variance of deep features. In addition, vibration signals are non-stationary signals with strong periodicity and high noise. Face recognition methods typically focus on images, which are non-serial and completely different from vibration signals.

Compared with previous OSR methods, the discriminative angle features for vibration signals are investigated in this study. We propose a novel normalized one-versus-all classification loss with center and contrastive regularization, which is called the discriminative angle feature learning for open-set classification of vibration signals (OSDAF). The normalized one-versus-all classification loss can effectively extract the deep features due to its low information interaction of each known class. In addition, the center and contrastive regularization terms explicitly optimize deep features to ensure intra-class compactness and inter-class divergence, while the existing method is implicit optimization.

The key contributions of our study are as follows:

1) We improve the deep learning-based industrial equipment fault diagnosis system by extending it from a closed-set to an open-set scenario that is closer to practical applications.
2) We propose a novel normalized one-versus-all classification loss with center and contrastive regularization. The trained network sufficiently learns discriminative

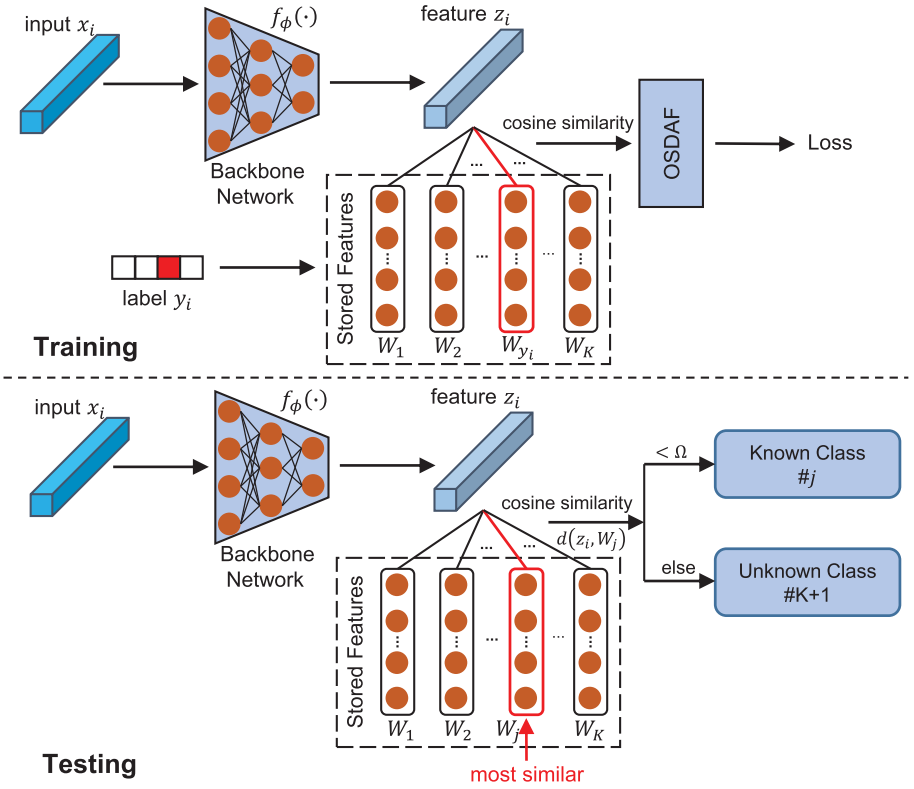


FIGURE 2. Outline of the proposed method in the training and testing phases. The model is trained end-to-end by the OSDAF loss and classifies samples based on the cosine distances between the learned and stored features.

angular features that are intra-class compact and inter-class divergent.

3) We verify the effectiveness of our proposed method using field-measured industrial equipment vibration signals. The results demonstrate the apparent advantages of the proposed method over other approaches in practical open-set fault diagnosis scenarios.

The rest of this paper is organized as follows. In Section II, the proposed approach is detailed. Comparative and ablation experiments are carried out in Section III. Finally, Section IV provides the conclusion and suggestions for future research.

II. PROPOSED METHOD

We mathematically formalize the open-set fault classification problem in Section II-A. Section II-B presents a detailed derivation of our proposed loss function $\mathcal{L}$. The pipeline of the training and testing phases is introduced in Section II-C based on the proposed loss.

A. PROBLEM STATEMENT

The label set of a training dataset $\mathcal{D}_{tr} = \{x_i, y_i\}_{i=1}^N$ comprising N samples and K known classes is $Y_{tr} = \{1, \ldots, K\}$. A test set $\mathcal{D}_{te}$, which contains unknown faults that have not occurred in the training set $\mathcal{D}_{tr}$, is collected in practical industrial applications. The training label set is a subset of the test label set $Y_{tr} \subset Y_{te}$. A learned model g that has been trained on $\mathcal{D}_{tr}$ attempts to make an open-set classification $g : x \to \{1, \ldots, K, K+1\}$, where $K+1$ denotes the label of unknown fault and $x \in \mathcal{D}_{te}$.

B. LEARNING OBJECTIVE

The objective of our approach is to learn the discriminative angle features. Two properties of these features are considered: intra-class compact and inter-class divergent. The two properties can reserve sufficient space between known classes for samples from unknown classes to occupy. Consequently, both the classification of known classes and the detection of unknown faults could be more effectively performed. Furthermore, this motivation is similar to that of the Fisher discrimination criteria. The purpose of a Fisher discriminant is to learn a linear projection that maximizes the inter-class distance and minimizes the intra-class distance. In this study, we employ a deep network as a non-linear projection to learn these features.

The proposed model consists of two components: a backbone network $f_\phi(\cdot)$ with parameters ϕ and stored weights $W \in \mathbb{R}^{d \times K}$, where d represents the dimension of deep feature $z = f_\phi(x)$. The outline of the proposed method in the training and testing phases is illustrated in Figure 2. The detailed derivation of the loss function is as follows.

1) NORMALIZED ONE-VERSUS-ALL CLASSIFICATION LOSS

The one-versus-all training strategy [27] includes two aspects: increasing the probability that sample x_i belongs to its correct class y_i, as well as the probability that x_i does not belong to any other class. The one-versus-all classification loss can be expressed as follows:

$$\mathcal{L}_{oc} = -\frac{1}{N}\sum_{i=1}^{N}\{\log p_{i,y_i} + \sum_{j=1,\neq y_i}^{K}\log(1-p_{i,j})\}, \quad (1)$$

where $p_{i,j}$ represents the output probability from the model that the sample x_i belongs to the j-th class, K is the number of classes, and N is the number of samples.

A deep backbone network $f_\phi(\cdot)$ is used to learn a d-dimensional high-level feature $z_i = f_\phi(x_i) \in \mathbb{R}^d$. By using a fully connected layer $W \in \mathbb{R}^{d\times K}$ and a Sigmoid activation function $\sigma(\cdot)$, the probability $p_{i,j}$ is calculated as follows:

$$p_{i,j} = \sigma(\|W_j^T\|\|z_i\|\cos\theta_{i,j}), \quad (2)$$

where $\theta_{i,j}$ is the angle between feature z_i and weight W_j. Thus, equation (1) can be rewritten as follows:

$$\mathcal{L}_{oc} = -\frac{1}{N}\sum_{i=1}^{N}\{\log\sigma(\|W_{y_i}^T\|\|z_i\|\cos\theta_{i,y_i}) + \sum_{j=1,\neq y_i}^{K}\log(1-\sigma(\|W_j^T\|\|z_i\|\cos\theta_{i,j}))\}. \quad (3)$$

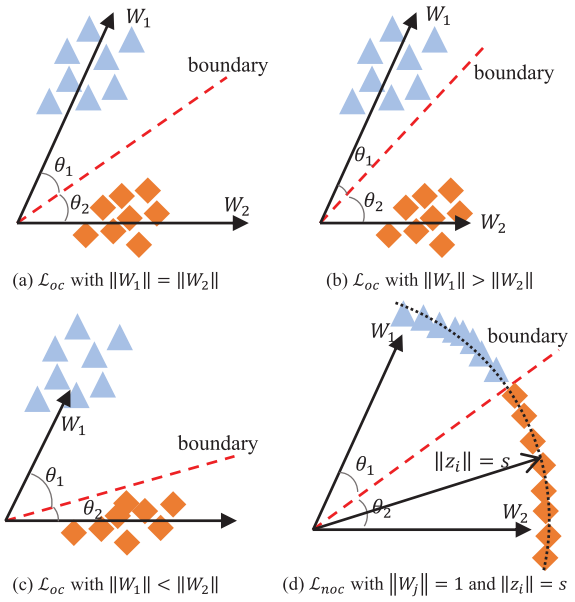


FIGURE 3. Geometrical interpretation of (a)-(c) $\mathcal{L}_{oc}$ and (d) $\mathcal{L}_{noc}$. The decision boundary is positively correlated with the magnitude of the weight vector W_j. $\|W_j\| = 1$ and $\|z_i\| = s$ are fixed to ensure reasonable angle decision boundary in the $\mathcal{L}_{noc}$.

We consider a binary scenario with classes $C1$ and $C2$ to simplify the analysis, as presented in Figure 3. Let W_1 and W_2 represent the weight vectors of $C1$ and $C2$, respectively. The decision boundary is defined by

$$\|W_1\|\cos\theta_1 = \|W_2\|\cos\theta_2, \quad (4)$$

where θ_1 is the angle between the feature vector z_i and W_1, and θ_2 is the angle between the feature vector z_i and W_2. Thus, its boundary depends on both the magnitudes of the weight vectors and the cosine of the angles. When $\|W_1\| = \|W_2\|$, the decision boundary is the angular bisector of W_1 and W_2, as depicted in Figure 3 (a). Otherwise, the decision boundary for C_j is positively correlated with the magnitude of weight vector W_j when $\|W_1\| \neq \|W_2\|$, as shown in Figure 3 (b)-(c).

To ensure reasonable angle decision boundaries, $\|W_j\| = 1$ and $\|z_i\| = s$ are fixed, where s is a predefined norm of the feature z_i [25], [28]. Therefore, the decision boundaries merely depend on the angle $\theta_{i,j}$, as shown in Figure 3 (d). The normalized one-versus-all classification loss is defined as follows:

$$\mathcal{L}_{noc} = -\frac{1}{N}\sum_{i=1}^{N}\{\log\sigma(s\cos\theta_{i,y_i}) + \sum_{j=1,\neq y_i}^{K}\log(1-\sigma(s\cos\theta_{i,j}))\}, \quad (5)$$

subject to

$$z_i = \frac{f_\phi(x_i)*}{\|f_\phi(x_i)*\|}, \quad w_j = \frac{w_j^*}{\|w_j^*\|}, \quad \cos\theta_{i,j} = w_j^T z_i. \quad (6)$$

2) NARROW INTRA-CLASS VARIANCE WITH CENTER LOSS

Typically, loss $\mathcal{L}_{noc}$ can only complete the classification task of known classes, which leads to the reservation of insufficient space for unknown fault samples in the feature space, as shown in Figure 4 (a). Compact intra-class features make the decision boundaries of known classes more robust for unknown fault detection, as shown in Figure 4 (b). We first define the cosine distance $d(z_i, W_j)$ between feature z_i and weight vector W_j as follows:

$$d(z_i, W_j) = \frac{1}{2}(1-\cos\theta_{i,j}). \quad (7)$$

We propose a cosine center loss [29] to reduce the cosine distance between feature z_i and its corresponding prototype W_{y_i}, which is expressed as follows:

$$\mathcal{L}_{cent} = \frac{1}{N}\sum_{i=1}^{N} d(z_i, W_{y_i}). \quad (8)$$

3) EXPAND INTER-CLASS SEPARATION WITH CONTRASTIVE LOSS

The cosine center loss $\mathcal{L}_{cent}$ minimizes the distance between the feature vectors z_i and corresponding weight vector W_{y_i} to

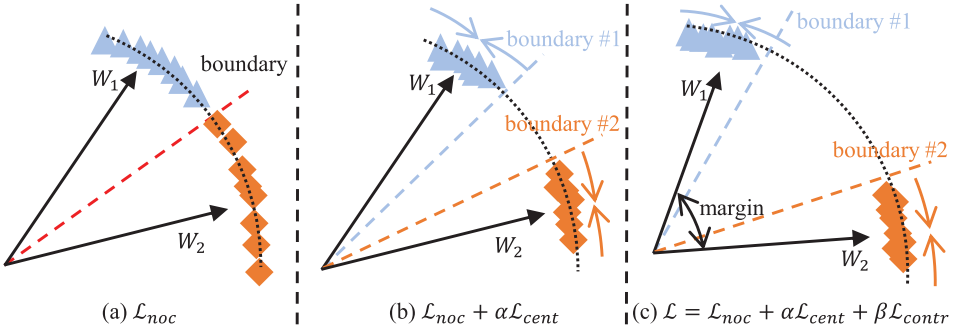


FIGURE 4. Illustrative comparison among (a) $\mathcal{L}_{noc}$, (b) $\mathcal{L}_{noc} + \alpha\mathcal{L}_{cent}$, and (c) $\mathcal{L} = \mathcal{L}_{noc} + \alpha\mathcal{L}_{cent} + \beta\mathcal{L}_{contr}$. $\mathcal{L}_{noc}$ is only used for classification. $\mathcal{L}_{cent}$ and $\mathcal{L}_{contr}$ are used to compress the intra-class distance and enlarge the inter-class margin, respectively.

reserve sufficient feature space for unknown-class samples. Intuitively, enlarging the gap between the decision boundaries of different categories is another approach for improving the unknown fault detection performance, as shown in Figure 4 (c). A contrastive cosine loss is proposed to push the weight vectors $W_j (j \neq y_i)$ that is not related to their corresponding labels y_i far away from the feature z_i with a large margin m, which is expressed as follows:

$$\mathcal{L}_{contr} = \frac{1}{N(K-1)} \sum_{i=1}^{N} \sum_{j=1, \neq y_i}^{K} [m - d(z_i, w_j)]^+, \quad (9)$$

where $[\cdot]^+$ is a function that returns the positive part of the argument.

4) OVERALL LEARNING OBJECTIVE

The final incorporative normalized one-versus-all loss with center and contrastive regularization, called the open-set discriminative angle feature learning loss (OSDAF), can be expressed as

$$\mathcal{L} = \mathcal{L}_{noc} + \alpha\mathcal{L}_{cent} + \beta\mathcal{L}_{contr}, \quad (10)$$

where α and β are the trade-off hyperparameters.

C. TRAINING AND TESTING PIPELINES

We summarize the training and testing procedures in Algorithm 1. During the end-to-end training procedure, parameters ϕ and W are jointly optimized using the loss $\mathcal{L}$ in equation (10), as shown in lines 1-10 of Algorithm 1. After training, the threshold Ω_k for the k-th known class is set to accept the $\eta\%$ of the correctly classified validation samples, as shown in lines 11-14 of Algorithm 1.

For the testing procedure, the model classifies samples based on the cosine distance $d(z_i, W_j)$ in (7). The cosine distance d_j between the output feature $f_\phi(x)$ and j-th weight W_j is first calculated for a test sample x. Subsequently, the closed-set prediction label $\hat{y} \in \{1, \ldots, K\}$ and its corresponding distance $d_{\hat{y}}$ are obtained. The label of test sample x is determined to be an unknown class $K+1$ if $d_{\hat{y}}$ exceeds the preset threshold $\Omega_{\hat{y}}$. Otherwise, the label of test sample x is predicted as $\hat{y}$.

III. EXPERIMENTAL VALIDATION

A. DATASET DESCRIPTIONS

1) GEAR AND BEARING DATASET

The gear and bearing dataset [30], [31] were created by the Southeast University (SEU) in China. Four types of bearing faults and four types of gear faults were included in this dataset, as listed in Table 1. Data collected under working conditions with a rotating speed-load configuration of 20 Hz–0 V were selected in this study.

2) BEARING DATASET

The bearing dataset [32], [33] was provided by the Paderborn University (PU) in Germany. Real damaged bearing data were chosen to form the PU dataset, and these included KA04, KA15, KA16, KA22, KA30, KB23, KB24, KB27, KI14, KI16, KI17, KI18, and KI22, as listed in Table 2. KA, KB, and KI indicate that the damaged positions were on the outer ring, both the outer and inner rings, and the inner ring, respectively. The test rig ran at 1,500 rpm with a load torque of 0.7 Nm and a radial force of 1,000 N on the bearing.

B. FAULT DIAGNOSIS TASKS AND IMPLEMENTATION DETAILS

Several OSR fault diagnosis tasks were established for each dataset to verify the effectiveness of our proposed method, as shown in Table 3. Each task had different numbers of known and unknown classes. Openness $\mathcal{O} = 1 - \sqrt{K/(K+U)}$ was used to denote the difficulty of a particular open-set recognition task, following the definition in [12]. Here, K and U denote the number of known and unknown classes, respectively. The higher the degree of openness $\mathcal{O}$, the greater the difficulty associated with the OSR task. The established tasks had a wide range of openness. For example, the openness of the tasks on the PU dataset varied from 0.5196 to 0.0801.

We divided the samples from the raw vibration signals in a partially overlapping or non-overlapping manner [34], with 1024 points for each sample and 3000 samples for each class. The dataset was constructed according to the known

Algorithm 1 Training and Testing Process of Proposed Model

Input: Training set $\{(x_i, y_i)\}_{i=1}^{N}$ and validation set. Test sample x. A backbone network $f_\phi(\cdot)$ with parameters ϕ, and stored features W. Hyperparameters α, β, m, and s. Learning rate γ, and acceptance rate $\eta\%$. Number of iterations $t \leftarrow 0$.

Output: The predicted test sample label.

```
// training procedure
```

while *not converge* **do**

- $t \leftarrow t + 1$;
- Normalize feature $z_i = \frac{f_\phi(x_i)*}{\|f_\phi(x_i)*\|}$;
- Normalize weight $W_j = \frac{W_j^*}{\|W_j^*\|}$;
- Compute cosine $\cos\theta_{i,j} = W_j^T z_i$;
- Compute the loss $\mathcal{L} = \mathcal{L}_{noc} + \alpha\mathcal{L}_{cent} + \beta\mathcal{L}_{contr}$ using (10);
- Compute the backpropagation error $\frac{\partial\mathcal{L}^t}{\partial x_i^t}$ for each i using $\frac{\partial\mathcal{L}^t}{\partial x_i^t} = \frac{\partial\mathcal{L}_{noc}^t}{\partial x_i^t} + \alpha\frac{\partial\mathcal{L}_{cent}^t}{\partial x_i^t} + \beta\frac{\partial\mathcal{L}_{contr}^t}{\partial x_i^t}$;
- Update the parameters ϕ using $\phi^{t+1} = \phi^t - \gamma^t\frac{\partial\mathcal{L}^t}{\partial x_i^t}$;
- Update the weight w using $W^{t+1} = W^t - \gamma^t\frac{\partial\mathcal{L}^t}{\partial W^t}$;

```
// set threshold
```

for $j = 1, \ldots, K$ **do**

- Compute the cosine distance $d_{i,j}$ using (7) of the correctly classified validation sample x_i;
- Set the threshold Ω_k to accept $\eta\%$ of $d_{i,j}$;

```
// test procedure
```

for $j = 1, \ldots, K$ **do**

- Compute the cosine distance $d_j = \frac{1}{2}(1 - \frac{w_j^T}{\|w_j\|}\frac{f_\phi(x)}{\|f_\phi(x)\|})$ of the test sample x;
- Compute the closed-set prediction label $\hat{y}$ and its corresponding cosine distance $d_{\hat{y}}$;

if $d_{\hat{y}} > \Omega_{\hat{y}}$ **then**

- Predict the label as $K + 1$;

else

- Predict the label as $\hat{y} \in \{1, \ldots, K\}$.

TABLE 1. Detailed description of the SEU dataset. Nine types of health states were considered.

Label	Fault Code	Description
0	N	Health State
1	IR	Inner Ring Fault
2	OR	Outer Ring fault
3	IOR	Inner + Outer Ring Fault
4	RE	Rolling Element Fault
5	C	Chipped Tooth Fault
6	M	Missing Tooth Fault
7	R	Root Fault
8	S	Surface Fault

and unknown label sets of each task (Table 3). Note that the labels of all unknown samples were set to $K + 1$.

TABLE 2. Detailed description of the PU bearing dataset. Thirteen types of health states were considered.

Label	Fault Code	Fault Mode[1]	Fault Location[2]	Description[3]
0	KA04	SP+S+1	O	FP
1	KA15	SP+S+1	O	PDI
2	KA16	SP+R+2	O	FP
3	KA22	SP+S+1	O	FP
4	KA30	D+R+1	O	PDI
5	KB23	SP+M+2	OI	FP
6	KB24	D+M+3	OI	FP
7	KB27	D+M+1	OI	PDI
8	KI14	SP+M+1	I	FP
9	KI16	SP+S+3	I	FP
10	KI17	SP+R+1	I	FP
11	KI18	SP+S+2	I	FP
12	KI21	SP+S+1	I	FP

[1] For the first term, SP and D represent single-point and distributed damage, respectively. For the second term, S, R, and M denote single, repetitive, and multiple damage, respectively. The third term represents the damage level.
[2] O, I, and OI indicate that the damage is on the outer ring, inner ring, and both rings, respectively.
[3] FP represents damage caused by fatigue and pitting, and PDI represent damage caused by plastic deformation and indentation.

TABLE 3. Open-set fault diagnosis tasks on two datasets. Eighteen OSR fault diagnosis tasks were randomly established for each dataset to verify the effectiveness of our proposed method.

Dataset	Task	Known Classes	Unknown Classes	Openness
SEU	S1	0,1,2	3,4,5,6,7,8	0.4226
	S2	0,5,6	1,2,3,4,7,8	0.4226
	S3	0,1,5	2,3,4,6,7,8	0.4226
	S4	0,1,2,3	4,5,6,7,8	0.3333
	S5	0,5,6,7	1,2,3,4,8	0.3333
	S6	0,1,2,3,4	5,6,7,8	0.2546
	S7	0,5,6,7,8	1,2,3,4	0.2546
	S8	0,1,2,3,4,5	6,7,8	0.1835
	S9	0,1,2,3,5,6,7	4,8	0.1181
	S10	0,1,2,3,4,5,6	7,8	0.1181
PU	P1	0,1,2	3,4,5,6,7,8,9,10,11,12	0.5196
	P2	5,6,7	0,1,2,3,4,8,9,10,11,12	0.5196
	P3	0,5,8	1,2,3,4,6,7,9,10,11,12	0.5196
	P4	0,1,2,3	4,5,6,7,8,9,10,11,12	0.4453
	P5	0,1,2,3,4	5,6,7,8,9,10,11,12	0.3798
	P6	0,1,5,6,8,9	2,3,4,7,10,11,12	0.3206
	P7	0,1,2,5,6,7,8,9,10	3,4,11,12	0.1679
	P8	0,1,2,3,5,6,7,8,9,10,11	4,12	0.0801

In addition, the one-dimensional ResNet-18 network was employed as the backbone feature extractor $f_\phi(\cdot)$, owing to its powerful feature extraction ability [35]. The deep ResNet-18 network mapped the input of 1024-dimensional raw vibration data into 64-dimensional features and was trained using the Adam optimizer at a learning rate of 0.001. The unknown detection threshold was set to 99% of the acceptance rate of known samples to ensure a high known-class classification objective, i.e., $\eta = 99$. We set the fixed length s of the feature to 30, following [25] and [26]. The hyperparameters α, β, and m were set to 0.01, 0.01, and 0.8, respectively. Each task was run independently five times to reduce the risk of randomization.

C. EVALUATION METRICS

In addition to the classification of known classes, the detection of unknown-class samples is a key but difficult-to-achieve goal of open-set fault classification. We used the unknown detection accuracy (UDA) to evaluate the ability of

TABLE 4. Results of different methods on tasks of the SEU dataset. The results were the averages of five random dataset splits. The best results are highlighted in bold font.

Metric	Task	Method									
		CNNevt	OSSCevt	OSSCentropy	OpenHybrid	RPL	RPL++	OpenHybrid	CosFace	ArcFace	OSDAF(ours)
UDA	S1	0.6869	0.0496	0.2179	0.1988	0.0862	0.4377	0.5679	0.5470	0.3941	**0.8507**
	S2	0.3088	0.0493	0.1438	0.0466	0.2927	0.2040	0.2157	0.2126	0.2173	**0.4421**
	S3	0.4461	0.0477	0.1951	0.1228	0.0963	0.4169	0.4232	0.3925	0.3260	**0.7402**
	S4	0.8646	0.0481	0.2858	0.2581	0.0043	0.6673	0.7566	0.6855	0.6515	**0.9889**
	S5	0.4805	0.0499	0.1967	0.0718	0.2453	0.3278	0.3295	0.4177	0.3279	**0.7585**
	S6	0.9892	0.0540	0.3389	0.3479	0.0072	0.6790	0.7588	0.8274	0.7255	**0.9921**
	S7	0.6062	0.0511	0.1748	0.0957	0.1855	0.5253	0.5317	0.5358	0.4742	**0.6632**
	S8	0.9818	0.0561	0.3956	0.3053	0.0101	0.8475	0.8780	0.8860	0.8603	**0.9958**
	S9	0.5235	0.0463	0.3718	0.1018	0.0058	0.4387	0.4842	0.8572	0.8063	**0.9940**
	S10	0.1713	0.0503	0.2997	0.0130	0.0028	0.0962	0.0662	0.7268	0.6955	**0.9906**
	Avg.	0.6059	0.0502	0.2620	0.1562	0.0936	0.4640	0.5012	0.6089	0.5479	**0.8416**
F1	S1	0.8010	0.3561	0.4577	0.5372	0.3519	0.6728	0.7349	0.7165	0.6311	**0.8965**
	S2	0.6092	0.3679	0.4296	0.4639	0.4180	0.5573	0.5629	0.5589	0.5559	**0.6756**
	S3	0.7317	0.3744	0.4699	0.5498	0.4436	0.6771	0.6888	0.6709	0.6459	**0.8495**
	S4	0.9296	0.4579	0.5760	0.6635	0.5379	0.8539	0.8852	0.8355	0.8176	**0.9877**
	S5	0.7658	0.4661	0.5440	0.5781	0.5250	0.7152	0.7178	0.7439	0.7113	**0.8782**
	S6	0.9906	0.5309	0.6502	0.7551	0.6210	0.8957	0.9159	0.9275	0.8939	**0.9913**
	S7	0.8712	0.5501	0.6084	0.6879	0.7028	0.8496	0.8528	0.8454	0.8251	**0.8794**
	S8	0.9881	0.6040	0.7141	0.8118	0.7088	0.9543	0.9602	0.9622	0.9549	**0.9918**
	S9	0.9137	0.6663	0.7489	0.8171	0.7921	0.8965	0.9057	0.9674	0.9578	**0.9917**
	S10	0.8486	0.6695	0.7361	0.8119	0.8005	0.8300	0.8226	0.9459	0.9416	**0.9908**
	Avg.	0.8449	0.5043	0.5935	0.6676	0.5902	0.7903	0.8047	0.8174	0.7935	**0.9133**

the model to detect unknown samples, which is defined as

$$UDA = M_u/N_u, \tag{11}$$

where M_u and N_u respectively represent the number of correctly classified unknown-class test samples and all unknown-class test samples. A higher UDA indicates better unknown-class detection performance.

In addition, the F1 score is commonly used to evaluate the open-set classification performance of a model that is trained with K classes but outputs $K + 1$ classes [22], [23]. The F1 score is defined as

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN}, \tag{12}$$

where TP, FP, and FN represent the number of true positives, false positives, and false negatives, respectively. A high F1 score indicates high known-class classification accuracy and strong unknown-class detection ability.

D. COMPARISON WITH OTHER METHODS

The proposed method was compared with the following methods.

1) The three most current open-set fault classification methods. CNNevt [17] utilized the Euclidean distance between the output logit vector and class center to determine whether an input sample is an unknown fault. OSSC [23] replaced the decoder of the VAE network with a classifier and trained an encoder-classifier network for discriminative feature extraction. The extreme value theory and entropy, which are termed as OSSCevt and OSSCentropy, respectively, were further used to detect the unknown fault.
2) Four mainstream OSR baselines in computer vision. OpenMax [19] proposed a new layer to generate extra unknown class activation and used the Euclidean distances for unknown detection. OpenHybrid [24] employed a flow-based model to estimate the density of known-class features and treated its output density as a special distance for unknown detection. RPL [20] and RPL++ [20] extracted Euclidean discriminative features using adversarial reciprocal point learning, where each reciprocal point represents the extra-class space corresponding to a known class.
3) Two face recognition methods based on discriminative angle feature learning. CosFace [25] added a cosine margin term to the L2 normalized SoftMax classification loss to force the network to learn deep features with large inter-class variance. ArcFace [26] proposed an additive angle margin SoftMax loss for discriminative feature learning.

For a fair comparison, all tasks were run on the same computer, which was equipped with an i9-12900KF CPU, 64 GB RAM, and an RTX 3090Ti GPU. All the baselines adopted the same backbone network, and trainable parameter initialization and dataset division were conducted using the same random seed. The results for the SEU and PU datasets were the averages of five random dataset splits, as shown in Tables 4 and 5, respectively. The best results are highlighted in bold font.

The open-set fault classification results for the SEU dataset are presented in Table 4. Our method achieved the highest UDA on each task of the SEU dataset. The average UDA for all tasks of the proposed method is 84.16%, which is 23.27% higher than that of the next best method. Although the face recognition methods CosFace and ArcFace also adopt discriminative angle feature learning, the UDA of the proposed method OSDAF is still 23.27% and 29.37% higher than the UDAs of the foregoing methods, respectively. This is owing to the explicit optimization of the intra- and inter-class cosine distances of the depth angle features by the proposed method.

TABLE 5. Results of different methods on tasks of the PU dataset. The results were the averages of five random dataset splits. The best results are highlighted in bold font.

Metric	Task	Method									
		CNNevt	OSSCevt	OSSCentropy	OpenHybrid	RPL	RPL++	OpenHybrid	CosFace	ArcFace	OSDAF(ours)
UDA	P1	0.6000	0.0500	0.1806	0.0660	0.2032	0.3488	0.4189	0.4049	0.2817	**0.6525**
	P2	0.3531	0.0449	0.1184	0.0549	0.0156	0.2610	0.2567	0.2822	0.1472	**0.4271**
	P3	0.3530	0.0488	0.1384	0.0356	0.0970	0.2430	0.2527	0.4940	0.4207	**0.5342**
	P4	**0.3367**	0.0525	0.1996	0.0455	0.2070	0.2091	0.1911	0.2321	0.2206	0.3300
	P5	0.3886	0.0551	0.1879	0.0449	0.1950	0.2277	0.2362	**0.4155**	0.3174	0.3965
	P6	0.2742	0.0523	0.2335	0.0273	0.0000	0.2839	0.3302	0.4261	0.3726	**0.5628**
	P7	0.3753	0.0494	0.2280	0.0304	0.0003	0.1881	0.1891	0.3765	0.4003	**0.4743**
	P8	0.2363	0.0458	0.1927	0.0080	0.0003	0.0622	0.0703	0.2642	0.2950	**0.5232**
	Avg.	0.3647	0.0499	0.1849	0.0391	0.0898	0.2280	0.2432	0.3619	0.3069	**0.4876**
F1	P1	0.6493	0.2809	0.3595	0.3310	0.5275	0.5067	0.5533	0.5813	0.5180	**0.7160**
	P2	0.5885	0.3279	0.3827	0.4313	0.4089	0.5367	0.5331	0.5836	0.4797	**0.6327**
	P3	0.5451	0.2765	0.3312	0.3269	0.4516	0.4619	0.4622	0.6338	0.5825	**0.6591**
	P4	0.6211	0.3634	0.4463	0.4604	0.6121	0.5527	0.5357	0.5823	0.5651	**0.6536**
	P5	0.7342	0.4293	0.5013	0.5450	0.6495	0.6418	0.6584	0.7356	0.6790	**0.7496**
	P6	0.6883	0.4862	0.5652	0.5844	0.5613	0.6937	0.7071	0.7496	0.7271	**0.8073**
	P7	0.8459	0.6207	0.6719	0.7556	0.7367	0.8024	0.8032	0.8526	0.8562	**0.8729**
	P8	0.8897	0.6922	0.7294	0.8524	0.8190	0.8587	0.8615	0.8972	0.9016	**0.9305**
	Avg.	0.6953	0.4346	0.4984	0.5359	0.5958	0.6318	0.6393	0.7020	0.6636	**0.7527**

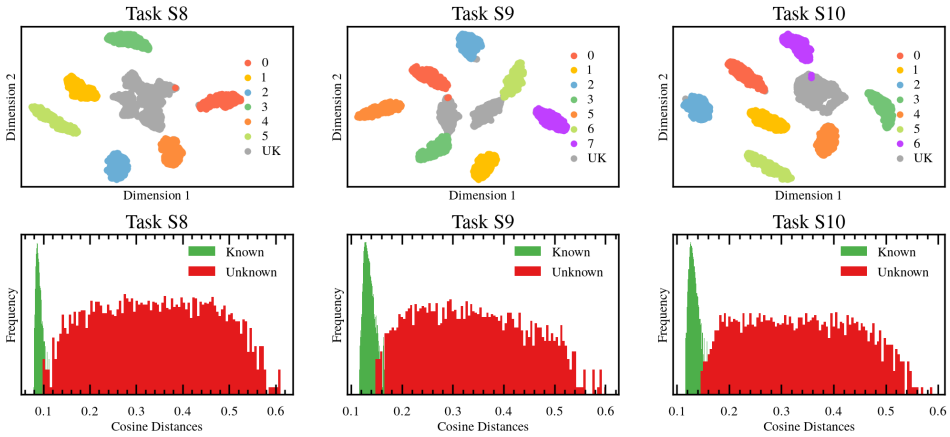


FIGURE 5. Visualizations of extracted features on the S8, S9, and S10 tasks. Top: visualizations of extracted features; bottom: known and unknown statistical distributions.

Moreover, the unknown detection threshold was set to 99% of the acceptance rate of known samples to ensure a high known-class classification objective, which is another goal of the open-set classification task. Table 4 indicates that the proposed method achieves promising F1 scores for all tasks, which further confirms its effectiveness and superiority in addressing the open-set fault classification problem.

The open-set fault classification results for the PU dataset are presented in Table 5. A pattern similar to that for the SEU dataset can be observed. The average UDA and F1 score of the proposed method are 48.76% and 0.7527, respectively, indicating that it has a strong ability for known-class classification and unknown-class detection. Compared to other baselines, the average UDA and F1 score of the proposed method were greater by at least +12.29% and 0.0574, respectively. Method CNNevt on the P4 task and method CosFace on the P5 task obtained slightly higher UDAs but lower F1 scores than the proposed method. This may be because they use unreasonable threshold settings to detect samples from unknown classes. Although this improves the detection accuracy of unknown classes, it impairs the classification performance of known classes. The high F1 scores listed in Table 5 indicate that the training and testing strategy of the OSDAF method is reasonable and superior in the field of open-set fault classification.

To intuitively demonstrate the effectiveness of the proposed approach, we visualized the normalized output features on the S8, S9, and S10 tasks, as shown in Figure 5. The features of samples from known and unknown classes are represented by colored and gray dots, respectively. The features from known classes form clusters according to their labels, indicating that the proposed model can ensure a high classification accuracy of known classes. Moreover, we also made statistics on the cosine distance from the output feature to the nearest stored feature, which is defined in equation (7). The cosine distance distributions between known and unknown classes presented an evident margin. Therefore, this cosine distance can be used to detect unknown-class samples.

E. ABLATION STUDY

1) A TOY EXAMPLE

The proposed OSDAF loss includes three components: the normalized one-versus-all classification loss $\mathcal{L}_{noc}$, center regularization $\mathcal{L}_{cent}$, and contrastive regularization $\mathcal{L}_{contr}$.

We studied a toy example to better visualize the learned features and validate the effectiveness of each component. We extracted the two-dimensional features of each known sample on the S4 task for simplicity. As presented in Figure 6 (a), we can observe that the features learned by only $\mathcal{L}_{noc}$ are distributed at angles but are not necessarily more discriminative. Under the regularization of $\mathcal{L}_{cent}$, the intra-class angular distribution is compact, as illustrated in Figure 6 (b). The contrastive regularization $\mathcal{L}_{contr}$ helps to enlarge the inter-class margin, especially the yellow and red classes in Figure 6 (c). This toy experiment validates our motivation of Figure 4.

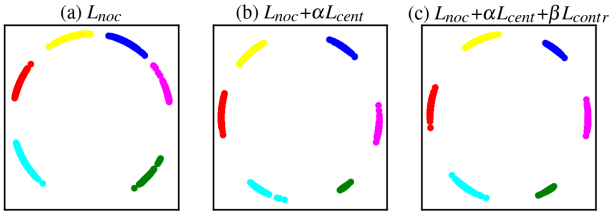


FIGURE 6. Toy experiment of different loss functions on the S4 task with two-dimensional features. Different colors represent different known classes. This experiment validates our motivation of Figure 4.

2) ABLATION EXPERIMENTS

Ablation experiments were conducted on the SEU dataset to verify the contribution of each part to the proposed loss. The following ablation experiments on the SEU dataset were conducted:

1) w/o $\mathcal{L}_{noc}$: The normalized one-versus-all classification loss $\mathcal{L}_{noc}$ was replaced by the commonly used normalized SoftMax (nsm) loss $\mathcal{L}_{nsm} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp(s\cos\theta_{i,y_i})}{\sum_j \exp(s\cos\theta_{i,j})}$ [25], [26], [36]. The training loss can be expressed as $\mathcal{L}_1 = \mathcal{L}_{nsm} + \alpha\mathcal{L}_{cent} + \beta\mathcal{L}_{contr}$.
2) w/o $\mathcal{L}_{cent}$: The center regularization $\mathcal{L}_{cent}$ was removed, leaving only the contrastive cosine regularization. The training loss can be expressed as $\mathcal{L}_2 = \mathcal{L}_{noc} + \beta\mathcal{L}_{contr}$.
3) w/o $\mathcal{L}_{contr}$: The contrastive regularization $\mathcal{L}_{contr}$ was removed, leaving only the center regularization. The training loss can be expressed as $\mathcal{L}_3 = \mathcal{L}_{noc} + \alpha\mathcal{L}_{cent}$.

TABLE 6. Averaged results of the ablation study on tasks of the SEU dataset. The results were the averages of five random dataset splits. The best results are highlighted in bold font.

Method	classification loss	regularization		averaged result	
	nsm or noc	contrastive	center	UDA	F1 score
w/o $\mathcal{L}_{noc}$	nsm	✔	✔	0.7337	0.8718
w/o $\mathcal{L}_{cent}$	noc	✔	✘	0.7042	0.8564
w/o $\mathcal{L}_{contr}$	noc	✘	✔	0.8207	0.9041
OSDAF	noc	✔	✔	**0.8416**	**0.9133**

The results in Table 6 indicate that the OSDAF method achieves the best average F1 score and UDA. The center regularization $\mathcal{L}_{cent}$ plays the most important role in the overall loss according to the averaged results of the ablation studies. In the absence of $\mathcal{L}_{cent}$, the average UDA and F1 scores decrease by 13.74% and 0.0569, respectively. Moreover, the normalized one-versus-all classification loss $\mathcal{L}_{noc}$ is another key factor that improves the performance of open-set fault classification. When $\mathcal{L}_{noc}$ was used instead of the normalized SoftMax classification loss, the average UDA and F1 scores of the OSDAF method increased by approximately 11% and 0.04, respectively. Similarly, the average UDA and F1 scores of the model were improved by approximately 2% with the help of contrastive regularization $\mathcal{L}_{contr}$.

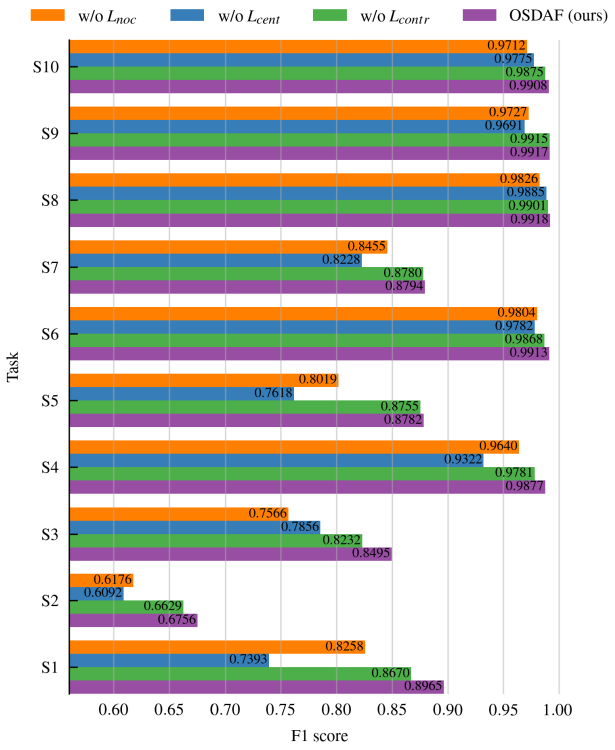


FIGURE 7. F1 score comparisons of different ablation methods for each task of the SEU dataset. The proposed OSDAF method has evident advantages, especially in high-openness tasks.

In addition, the proposed OSDAF method has evident advantages in high-openness tasks, as shown in the F1 scores of each task in Figure 7. For example, OSDAF achieves an F1 score of at least 0.0295 higher than those of other methods for the S1 task with an openness of 0.4226. We also plotted the confusion matrix of the S1 task to observe the accuracy of each class in detail, as shown in Figure 8. The accuracy of unknown detection of OSDAF was at least 6% higher than those of the other three ablation methods. In conclusion, the proposed normalized one-versus-all classification loss $\mathcal{L}_{noc}$, center regularization $\mathcal{L}_{cent}$, and contrastive regularization $\mathcal{L}_{contr}$ all contribute to the open-set fault classification performance in practical application scenarios.

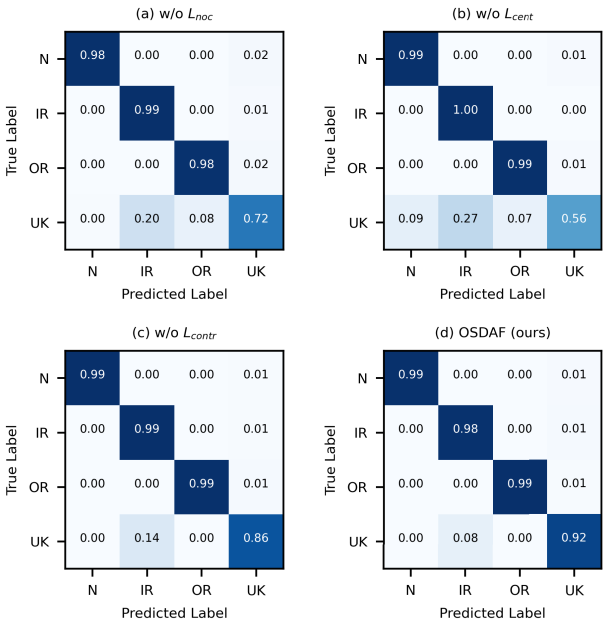


FIGURE 8. Confusion matrices on task S1 for three ablation methods and the proposed method: (a) w/o $\mathcal{L}_{noc}$, (b) w/o $\mathcal{L}_{cent}$, (c) w/o $\mathcal{L}_{contr}$, and (d) proposed OSDAF method, where "UK" refers to the unknown class.

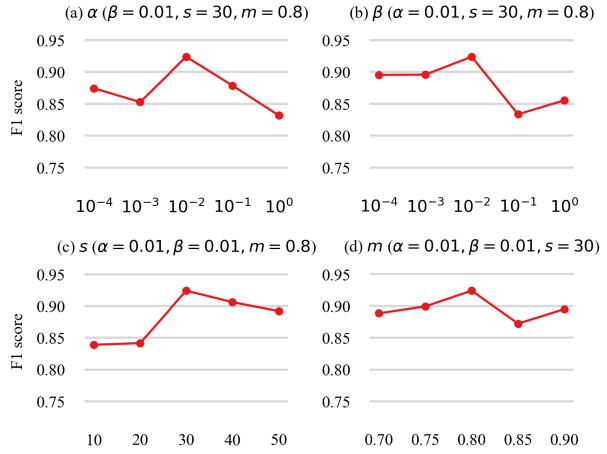


FIGURE 9. F1 scores on task S6 with varying (a) trade-off hyperparameter α, (b) trade-off hyperparameter β, (c) fixed feature length s, and (d) large margin m. The recommended hyperparameter values were $\alpha = \beta = 0.01$, $s = 30$, and $m = 0.8$ because the model received the highest F1 score.

F. PARAMETER ANALYSIS

The unknown detection threshold was set as a 99% acceptance rate of known samples to ensure a high known-class classification objective, i.e., $\eta = 99$. We analyzed the impact of other hyperparameters on the performance of open-set classification on the S6 task, as shown in Figure 9. The hyperparameters α and β were the trade-off coefficients of the regularization terms $\mathcal{L}_{cent}$ and $\mathcal{L}_{contr}$ in the loss function, respectively. The values of α and β were selected as $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}\}$. The hyperparameter s was expressed as the length of the fixed feature, whose value was chosen as $\{10, 20, 30, 40, 50\}$. The hyperparameter m represented the large angle margin, whose value was selected as $\{0.70, 0.75, 0.80, 0.85, 0.90\}$. As shown in Figure 9, the recommended hyperparameter values were $\alpha = \beta = 0.01$, $s = 30$, $m = 0.8$ because the model received the highest F1 score on the S6 task under this setting.

IV. CONCLUSION

This paper proposes a novel normalized one-versus-all classification loss with center and contrastive regularization to address the open-set fault classification scenarios in practical applications. This loss enables the network to sufficiently learn discriminative angular features that are intra-class compact and inter-class divergent. The effectiveness of the proposed method is verified through a comparison with other mainstream algorithms, using field-measured industrial equipment vibration signals obtained from motor bearing and gear datasets. The results demonstrate the apparent advantages of the proposed method over other approaches. Consequently, the trained model can be used for unseen fault recognition and fault classification based on cosine distances in the angle feature space.

Despite the promising results, some limitations exist and need to be further investigated. Considering that the field-deployed model requires real-time fault diagnosis of vibration signals, it is necessary to design a lightweight model to reduce deployment costs. In addition, this study focuses on the open-set fault classification problem where the training label set is a subset of the test label space. Concerning the OSDA, the invariant angle feature extraction method between the source and target domains will be investigated in future research.

REFERENCES

[1] S. Tang, S. Yuan, and Y. Zhu, "Deep learning-based intelligent fault diagnosis methods toward rotating machinery," *IEEE Access*, vol. 8, pp. 9335–9346, 2020.

[2] M. T. Pham, J.-M. Kim, and C. H. Kim, "Efficient fault diagnosis of rolling bearings using neural network architecture search and sharing weights," *IEEE Access*, vol. 9, pp. 98800–98811, 2021.

[3] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.

[4] W. Zhang, T. Zhang, G. Cui, and Y. Pan, "Intelligent machine fault diagnosis using convolutional neural networks and transfer learning," *IEEE Access*, vol. 10, pp. 50959–50973, 2022.

[5] H. Fang, J. Deng, Y. Bai, B. Feng, S. Li, S. Shao, and D. Chen, "CLFormer: A lightweight transformer based on convolutional embedding and linear self-attention with strong robustness for bearing fault diagnosis under limited sample conditions," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–8, 2022.

[6] Q. Lv, X. Yu, H. Ma, J. Ye, W. Wu, and X. Wang, "Applications of machine learning to reciprocating compressor fault diagnosis: A review," *Processes*, vol. 9, no. 6, p. 909, May 2021.

[7] R. Huang, Y. Liao, S. Zhang, and W. Li, "Deep decoupling convolutional neural network for intelligent compound fault diagnosis," *IEEE Access*, vol. 7, pp. 1848–1858, 2019.

[8] H. Chen, B. Jiang, S. X. Ding, and B. Huang, "Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 1700–1716, Mar. 2022.

[9] H. Wang, J. Xu, R. Yan, C. Sun, and X. Chen, "Intelligent bearing fault diagnosis using multi-head attention-based CNN," *Proc. Manuf.*, vol. 49, pp. 112–118, Jan. 2020.

[10] Y. Ding, M. Jia, Q. Miao, and Y. Cao, "A novel time–frequency transformer based on self–attention mechanism and its application in fault diagnosis of rolling bearings," *Mech. Syst. Signal Process.*, vol. 168, Apr. 2022, Art. no. 108616.

[11] J. Tang, J. Wu, B. Hu, and J. Liu, "Towards a fault diagnosis method for rolling bearing with bi-directional deep belief network," *Appl. Acoust.*, vol. 192, Apr. 2022, Art. no. 108727.

[12] W. J. Scheirer, A. D. R. 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.

[13] P. P. Busto and J. Gall, "Open set domain adaptation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 754–763.

[14] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1180–1189.

[15] 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. Ind. Informat.*, vol. 17, no. 11, pp. 7445–7455, Nov. 2021.

[16] G. Mao, Y. Li, S. Jia, and K. Noman, "Interactive dual adversarial neural network framework: An open-set domain adaptation intelligent fault diagnosis method of rotating machinery," *Measurement*, vol. 195, Apr. 2022, Art. no. 111125.

[17] X. Yu, Z. Zhao, X. Zhang, Q. Zhang, Y. Liu, C. Sun, and X. Chen, "Deep-learning-based open set fault diagnosis by extreme value theory," *IEEE Trans. Ind. Informat.*, vol. 18, no. 1, pp. 185–196, Jan. 2022.

[18] C. Zhao and W. Shen, "Dual adversarial network for cross-domain open set fault diagnosis," *Rel. Eng. Syst. Saf.*, vol. 221, May 2022, Art. no. 108358.

[19] A. Bendale and T. E. Boult, "Towards open set deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1563–1572.

[20] G. Chen, L. Qiao, Y. Shi, P. Peng, J. Li, T. Huang, S. Pu, and Y. Tian, "Learning open set network with discriminative reciprocal points," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 507–522.

[21] G. Chen, P. Peng, X. Wang, and Y. Tian, "Adversarial reciprocal points learning for open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 8065–8081, Nov. 2022.

[22] A. He and X. Jin, "Deep variational autoencoder classifier for intelligent fault diagnosis adaptive to unseen fault categories," *IEEE Trans. Rel.*, vol. 70, no. 4, pp. 1581–1595, Dec. 2021.

[23] J. Chen, G. Wang, J. Lv, Z. He, T. Yang, and C. Tang, "Open-set classification for signal diagnosis of machinery sensor in industrial environment," *IEEE Trans. Ind. Informat.*, vol. 19, no. 3, pp. 2574–2584, Mar. 2023.

[24] H. Zhang, A. Li, J. Guo, and Y. Guo, "Hybrid models for open set recognition," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 102–117.

[25] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, "CosFace: Large margin cosine loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5265–5274.

[26] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4685–4694.

[27] L. Shu, H. Xu, and B. Liu, "DOC: Deep open classification of text documents," 2017, *arXiv:1709.08716*.

[28] 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, May 2021, Art. no. 106925.

[29] 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, Switzerland: Springer, 2016, pp. 499–515.

[30] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446–2455, Apr. 2019.

[31] S. Shao, S. McAleer, R. Yan, and P. Baldi. (2018). SEU dataset. Southeast University in China. [Online]. Available: https://github.com/cathysiyu/Mechanical-datasets

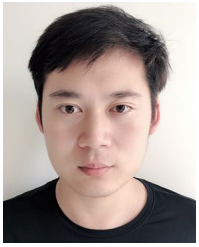
[32] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, "Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification," in *Proc. PHM Soc. Eur. Conf.*, 2016, vol. 3, no. 1, pp. 5–8.

[33] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro. (2016). *PU Bearing Dataset*. [Online]. Available: https://mb.uni-paderborn.de/kat/forschung/datacenter/bearing-datacenter/

[34] X. Pei, X. Zheng, and J. Wu, "Rotating machinery fault diagnosis through a transformer convolution network subjected to transfer learning," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

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

[36] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-margin softmax loss for convolutional neural networks," in *Proc. 33rd Int. Conf. Mach. Learn. (ICML)*, vol. 48, New York, NY, USA, Jun. 2016, pp. 507–516.



JIE MEI received the B.S. degree in electronic information and engineering from the Changsha University of Science and Technology, Changsha, China, in 2012. He is currently pursuing the Ph.D. degree in information and communications engineering with the Huazhong University of Science and Technology, Wuhan, China. He was an Operation and Maintenance Engineer at a nuclear power plant, from 2012 to 2018. His research interests include open-set recognition and domain adaptation in deep learning.

WEI LIU (Member, IEEE) received the B.S. degree in telecommunication engineering and the Ph.D. degree in electronics and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 1999 and 2004, respectively. He is currently a Professor with the School of Electronic Information and Communications, Huazhong University of Science and Technology. His research interests include information processing, network measurement, and learning behavior analysis.

MING ZHU received the Ph.D. degree in electronic information and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008. He was a Visiting Scholar with Northwestern University, USA, from 2014 to 2015. He is currently a Professor with the School of Electronic Information and Communications, Huazhong University of Science and Technology. His current research interests include aerosol measurement, acoustic gas detection, intelligent fault diagnosis, and deep learning.

YONGKA QI received the B.S. degree in electronic information and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2020, where he is currently pursuing the M.S. degree in information and communications engineering. His current research interests include intelligent fault diagnosis and deep learning.

MING FU received the B.S. degree in electronic information and engineering from Central China Normal University, Wuhan, China, in 2020. She is currently pursuing the M.S. degree in information and communications engineering with the Huazhong University of Science and Technology, Wuhan. Her current research interests include intelligent fault diagnosis and deep learning.

YUSHI LI received the B.S. degree in information engineering from the Wuhan University of Technology, Wuhan, China, in 2021. She is currently pursuing the Ph.D. degree in information and communications engineering with the Huazhong University of Science and Technology, Wuhan. Her current research interests include time series prediction and deep learning.

QUAN YUAN received the Ph.D. degree in electrical circuit and system from the Huazhong University of Science and Technology, Wuhan, China, in 2010. He is currently an Associate Professor with the School of Mechanical Engineering and Electronic Information, China University of Geosciences, Wuhan. His current research interests include nonlinear dynamics, neural networks, and wireless communications.