Mechanical Systems and Signal Processing 228 (2025) 112460

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Full Length Article

Dual disentanglement domain generalization method for rotating Machinery fault diagnosis

Guowei Zhang [a,b], Xianguang Kong [a,b], Hongbo Ma [a,b,*], Qibin Wang [a,b], Jingli Du [a,b], Jinrui Wang [c]

[a] School of Mechano-Electronic Engineering, Xidian University, Xi'an 710071, PR China
[b] State Key Laboratory of Electromechanical Integrated Manufacturing of High-performance Electronic Equipments, Xidian University, Xi'an 710071, PR China
[c] College of Mechanical and Electronic Engineering, Shandong University of Science and Technology, Qingdao 266590, PR China

ARTICLE INFO

Keywords:
Rotating machinery
Domain generalization
Fault diagnosis
Feature disentanglement

ABSTRACT

The objective of domain generalization fault diagnosis is to develop a robust model that can generalize to unseen domains. This makes it a highly ambitious and challenging task. However, most current methods rely on domain labels to extract domain-invariant features and do not consider the negative impact of the presence of class-irrelevant features in domain-invariant features on generalization. Therefore, this paper proposes a dual disentanglement domain generalization method for rotating machinery fault diagnosis that does not depend on domain labels. Based on the analysis of the potential features between domains and class labels, a dual contrastive disentanglement module and an adversarial mask disentanglement module are proposed to disentangle the domain-invariant and class-relevant features, respectively. Specifically, in the dual contrastive disentanglement module, the concept of contrasting is employed to train the network shallow features of the source data and the style-enhanced data to produce domain-aware mask decoupled domain-specific and domain-invariant representations. The adversarial mask disentanglement module uses an adversarial classifier to update the class-aware mask and further accurately separate class-relevant and class-irrelevant features. Concurrently, the KLD loss is devised to guarantee that the class-relevant features encompass sufficient labeling information. Finally, the efficacy of the method is substantiated by comprehensive experimental findings on both public and private datasets. The code will be available at: https://github.com/GuoweiaaZhang/DDDG.

1. Introduction

In the wave of rapid development of information technology, the industrial world is witnessing a radical change. This change not only requires industrial machinery and equipment to enhance their level of intelligence, but also poses smarter challenges for operation and maintenance [1–3]. In this context, rotating components as the core of industrial equipment are particularly critical, and they must meet extremely high standards of reliability and safety. At the same time, the rise of artificial intelligence technology has inspired great enthusiasm among researchers and innovators. As a result, a series of innovative deep learning-based fault diagnosis

* Corresponding author.
E-mail address: mhb@mail.xidian.edu.cn (H. Ma).

https://doi.org/10.1016/j.ymssp.2025.112460
Received 20 August 2024; Received in revised form 7 January 2025; Accepted 9 February 2025
Available online 14 February 2025

techniques have sprung up, opening a new chapter in intelligent monitoring and fault diagnosis of rotating machinery [4–6].

Fault diagnosis methods enabled by deep learning technology can flexibly and automatically extract equipment state information from the signals of rotating machinery, demonstrating their adaptive capabilities [7,8]. This means that intelligent fault diagnosis methods are no longer strictly dependent on expert knowledge, which helps reduce the dependence on human resources and lower operating costs [9]. Although these methods have achieved significant results in fault diagnosis, they generally have a limitation: the training and testing data used are all derived from stable operating conditions. However, in real-world engineering applications, the operating state of equipment is often variable, which leads to distribution differences between training data and actual test data [10–12]. Therefore, domain-adaptation intelligent fault diagnosis technology has attracted much attention in the academic community because it can effectively overcome the data distribution inconsistency caused by changes in working conditions [13–15]. Currently, most domain adaptation fault diagnosis methods mainly rely on adversarial training [16,17] and distance metric techniques [18,19] to reduce the differences between the source and target domains. However, these domain adaptation methods are designed for scenarios where target domain data is accessible during training, which makes them unsuitable for domain generalization (DG) scenarios where no target domain data is available.

DG fault diagnosis, an emerging method in the field of fault diagnosis, can achieve efficient diagnosis across different domains without relying on target domain data during the model training stage, which has attracted extensive academic research and attention [20–23]. Currently, research on fault diagnosis based on DG can be mainly summarized into the following three categories: (1) The first category is data manipulation methods, which help generalization mainly through data augmentation or generating diversity samples. Fan et al. [24] introduced a data augmentation technique that extends mixup to class and domain dimensions, aiming to bridge the gap between categories and domains while enhancing sample diversity. Zhao et al. [25] proposed an innovative data augmentation perspective aimed at capturing stable inter-class relationships across domain environments to improve robustness. (2) The second category is the learning strategies approach, which focuses on the use of general learning strategies to improve generalization skills. Ren et al. [26] adopted a *meta*-learning framework and integrated a gradient alignment algorithm to learn a domain-invariant and robust prediction strategy under unpredictable working conditions. Wang et al. [27] adopted a model-independent learning strategy aimed at establishing a shared optimization trajectory for each source domain, thereby facilitating the learning of domain-invariant features. (3) The third major category of DG methods is the representation learning that has received the most attention from scholars, Chen et al [28] effectively utilized multiple domain data through an adversarial learning mechanism between feature extraction and domain classifiers in order to explore and utilize cross-domain invariant knowledge, and Shi et al [29] proposed a comparative learning framework to minimize discrepancies between the same failure modes across multiple domains. It is well known that extracting domain-independent representations using domain adversarial learning [30] and feature alignment [31] is a common strategy to improve model generalization.

Recent research based on disentanglement learning further promotes the development of DG fault diagnosis. Domain decomposition-based methods [32–35] aim to partition features into domain-specific and domain-invariant components, thereby enhancing generalization across varying working conditions. Wang et al. [32] first designed multiple domain-specific auxiliary classifiers to learn domain-specific features and then constructed a convolutional autoencoder module to remove domain-specific features. An et al. [33] proposed a model-decoupled autoencoder to extract fault information and eliminate the working condition information as much as possible. Ren et al. [34] introduced a novel semi-supervised domain generalization approach, the domain-invariant feature fusion network, which incorporates two distinct branches to capture both inter-domain and intra-domain invariant features. Xie et al. [35] developed a novel framework called the domain-specific invariant adversarial network, which integrates domain-invariant representation learning with feature disentanglement techniques to tackle the issue of imbalanced distribution in fault diagnosis data. Furthermore, structural causal model-based methods [36–39] leverage causal reasoning to identify and isolate factors that influence fault-related and domain-related features. Jia et al. [36] decomposed features into causal factors (fault related representation) and non-causal factors (domain related representation) based on causal mechanism to enhance the model generalization ability. Jia et al. [37] introduced a deep causal decomposition network designed for cross-machine bearing diagnosis without relying on target domain data. Leveraging a structural causal model derived from bearing fault signals, the approach identifies the cross-machine generalized fault representation as a causal factor and the domain-specific representation as a non-causal factor. Guo et al. [38] introduced causally independent and sparse shift networks for fault diagnosis, leveraging a structural causal model that captures the causal relationships of features involved in the diagnostic process. Ma et al. [39] developed a causality-inspired multi-source DG method for fault diagnosis, incorporating feature diversity activation and the suppression of non-causal features to enable diagnosis under unseen operating conditions.

In summary, the current DG fault diagnosis methods based on representation learning have been widely studied for their effectiveness, but there are still some key issues that have been neglected:

(1) Most of the work [28–32,34] focuses on extracting domain invariant features, while neglecting further decomposition of class-relevant and class-irrelevant features, which may exist in the domain invariant features and affect the model's generalization ability.
(2) Many methods [28–32,34,36,37,39] rely on domain labels to learn or disentanglement domain-invariant features, however, domain labels are often difficult to obtain in real situations, in addition, the source domain data may be formed by a mixture of single or multiple domains.
(3) Current disentanglement methods [32,33,35,36] mainly rely on multiple encoders for disentanglement, and the encoder-decoder network causes additional computational overhead.

To address the above issues, in this paper, we propose a dual disentanglement domain generalization fault diagnosis method decouples class-relevant domain-invariant features for generalization to unseen domains without relying on domain labels and decoders. Specifically, the style enhancement method Random Mixup (RandMix) is used to simulate domain shifts. Then a dual contrastive disentanglement loss is designed to decouple channel domain-invariant and domain-specific features at the shallow level of the network and to eliminate channels with domain-specific features using a masking mechanism. Finally, adversarial mask disentanglement loss is designed in the deeper layers of the network to generate class-aware masks for accurate detection of class-relevant and class-independent features, while a Kullback-Leibler Divergence (KLD) loss is designed to ensure that class-relevant features contain sufficient labeling information. The main contributions in this work are summarized as follows:

(1) A dual disentanglement domain generalization method is proposed for rotating machinery fault diagnosis, based on the analysis of potential features between domains and class labels. Dual contrastive disentanglement module and adversarial mask disentanglement module are proposed to decouple the domain-invariant and class-relevant features respectively, in addition, the method can complete the generalization task without relying on domain labels.
(2) Learnable domain-aware masks and class-aware masks are used instead of the traditional multiple encoders for disentanglement, which is not only more efficient, but also ensures that the components for disentanglement are naturally complementary.
(3) Public and private datasets are employed to design multiple domain generalization scenarios, including single and mixed domains, respectively, and extensive experiments are conducted to show the superiority of the proposed approach.

The paper is structured as follows: Section 2 gives the problem definition and the Venn diagram analysis of the potential features. Section 3 details the dual disentanglement domain generalized fault diagnosis method. Section 4 describes the details of the experiments and analyzes the performance of the proposed method through experiments. Finally, Section 5 concludes.

2. Problem Formulation

Formulate DG from input sample $x \in \mathscr{X}$ to prediction label $y \in \mathscr{Y}$ in fault diagnosis settings, where x can come from a single domain $D^s = \left\{ \left(x_i^s, y_i^s\right)|_{i=1}^{N_s} \right\}$, or mixed domain $D^{s,m}$, $D^{s,m} = \left\{ \left(x_i^{s,m}, y_i^{s,m}\right)|_{i=1}^{N_{s,m}} \right\}$ represents the complete label space $y_i^{s,m} \in \mathscr{Y}$ formed by mixing multi-source domains. Given a family Θ of models and training data x sampled from one of the above distributions P^s, the goal is to find a model $\theta \in \Theta$ that can generalize well to the unseen target distribution P^t.

Consider a learning model composed of a feature extractor $F : x \to z$ and a classifier $C : z \to y$, where z is the feature embedding space. Divide the latent feature space z into three distinct parts based on its association with the domain and class labels, as illustrated in the Venn diagram in Fig. 1(a). Considering that it is not possible to distinguish domains without relying on domain labels, data augmentation is used to model domain shifts, therefore, only two domains are considered, namely the source domain S and the augmented source domain S^a, and z and z^a are the corresponding latent features. Specifically, the domain-specific features A and A^* denote, respectively, the specific features of each domain that are exclusive to its domain, and thus the domain-specific features are not the same for different domains. Domain-invariant features B and C represent common features across domains that do not change with domain shift. Further considering the relationship between features and labels, B and C can be more specifically defined as domain-invariant features unrelated to the class and domain-invariant features related to the class, respectively.

Previous studies [28–32,34] mainly focused on decoupling domain-specific and domain-invariant features, as shown in Fig. 1(b), and proved that extracting domain-invariant features can improve the generalization ability of the model, but they did not further consider separating class-relevant and class-irrelevant features. Some studies [40,41] indicate that class-irrelevant features B within domain-invariant representations negatively impact generalization. In addition, many methods rely on domain labels to guide the extraction of domain-invariant features, however, the acquisition of domain labels is a time-consuming and labor-intensive process in industrial applications. Therefore, the focus of this paper is on extracting class-relevant domain-invariant features C without relying on domain labels, as illustrated in Fig. 1(c). To address this issue, we introduce our proposed dual disentanglement domain generalization framework in the following section.

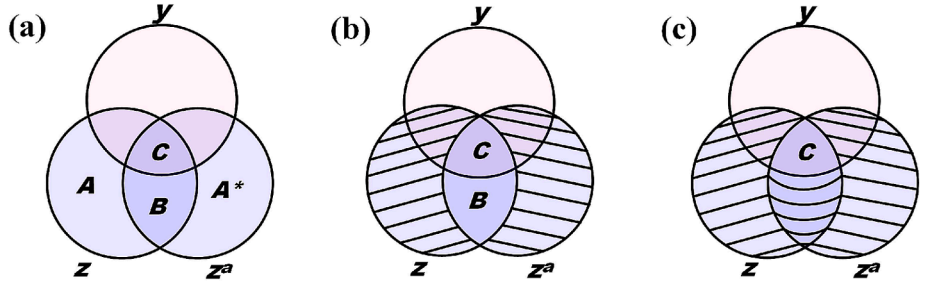

Fig. 1. The Venn diagram shows the relationship between the class label y and two different source domains z and z^a.

3. Proposed method

In this section, the three modules of the dual disentanglement domain generalized fault diagnosis approach are presented in detail, as shown in Fig. 2. Specifically, Section 3.1 presents a randomized mixed data enhancement module RandMix to generate domain shifted data. In Section 3.2, the dual contrastive disentanglement module is proposed, which allows the decoupling of domain-invariant and domain-specific features between source and augmented domains without relying on domain labels. Section 3.3 introduces the adversarial mask disentanglement module, which is capable of further decoupling class-relevant and class-irrelevant features based on domain-invariant features, thereby ensuring the adequate extraction of class-relevant domain-invariant features.

3.1. Random mixup

Considering that it is difficult to extract domain invariant features from source domain data without domain labels, the style enhancement method RandMix is employed to simulate domain shifts, and RandMix not only increases the diversity of the data to enhance the model generalization ability, but also provides the possibility of decoupling the domain invariant features for the dual contrastive disentanglement module. The specific flow for implementing RandMix is shown in stage1 of Fig. 2, where RandMix consists of several simple style feature generation modules, where each style feature generator consists of a converter (l_c) and an inverse converter (l_i). To ensure that the style feature generation module generates reasonable augmented data, inspired by Wang et al. [42] each converter and inverse converter is implemented by convolutional and transpositional convolutional layers at different scales, while considering the randomness of the unknown domain, AdaIN [43] is applied to inject noise into the style feature generation modules to simulate random style transforms. Specifically, AdaIN contains two linear layers l_{l1} and l_{l2} that are capable of generating two corresponding noise outputs when these layers receive noise inputs from normalized distributions ($n \sim \mathscr{P}_{\mathrm{N}(0,1)}$), and considering that pushing the style-enhanced data too far away from the original data may result in a decrease in the model's generalization ability [44], the input of noise with smaller variance values is considered, and then the two noise outputs are injected as multiplicative and additive noise, respectively, into the autoencoder representation:

$$R(x) = l_i(l_{l1}(n) \times l_{IN}(f_c(x)) + l_{l2}(n)) \tag{1}$$

where $l_{IN}(\cdot)$ denotes the instance normalization layer, and f_c denotes the CNN feature extractor. The training data are fed into all the style feature generation modules, and multiple style-enhanced data are mixed with the original data with random weights, which are randomly drawn from the normalized distribution ($w_i \sim Normal(0,1)$). Finally, the blended results are scaled by a sigmoid function, which is expressed as $\sigma(x) = 1/(1+e^{-x})$:

$$\mathbf{R}(x) = \sigma\left(\frac{1}{\sum_{i=1}^{N} w_i}\left[w_0 x + \sum_{i=1}^{N}(w_i R_i(x))\right]\right) \tag{2}$$

where N denotes the number of samples. Using RandMix it is possible to generate augmented data with the same labels that correspond to all labeled samples in the source domain.

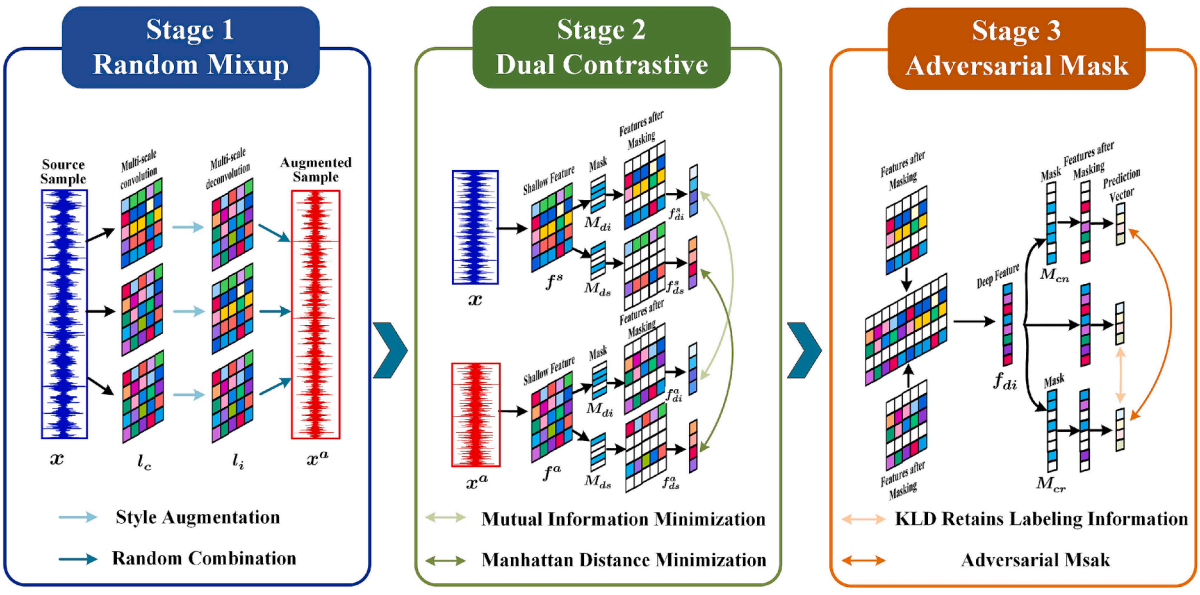

Fig. 2. Main flow of dual disentanglement domain generalization methods for fault diagnosis.

3.2. Dual contrastive disentanglement

Based on the analysis of the Venn diagram (Fig. 1) of the domain-class relationship between source domain data and augmented data, for a given source domain data and style augmented data, their domain-invariant features should be similar and domain-specific features should be different. In order to be able to disentangle and discard the domain specific feature representations, the idea of contrast learning is considered for disentangling the similarities and differences.

Considering that the shallow features of convolutional networks are believed to contain more domain-sensitive information [45,46], to decouple domain-invariant feature representations from domain-specific feature representations, single-domain data is used to train a fault diagnosis model, i.e., a baseline, and to evaluate the impact of the features extracted from the first convolutional layer on the fault diagnosis generalization performance. A typical example is given in Fig. 3(a) and (b), where the blue line is the diagnostic accuracy obtained on the target domain versus the index of the deleted feature channel. First, the figure indicates that nearly half of the channels do not significantly affect the generalization performance of the model after being removed, which suggests that the removed features are redundant, which coincides with the lottery ticket theory [47]. It can also be found that the diagnostic accuracy of the generalization is significantly improved after removing the features of some channels, such as the 3 channels in Fig. 3 (a) and the 1 channel in Fig. 3(b), indicating that these channels contain domain-specific features, which have a large discrepancy with the feature representations of the target domain and are difficult to generalize to the target domain. On the contrary, removing some channel features results in a significant decrease in the generalization accuracy, such as channel 8 in Fig. 3(a) and channels 6, 7 in Fig. 3 (b), suggesting that the channel features contain domain-invariant features that are extremely critical to the generalization task.

Based on the above phenomenon, we consider decoupling the channel features at the shallow level of the network using the idea of comparison to identify and exclude the channels with domain-specific features from the extracted vibration signal features. The detailed process of the dual contrastive disentanglement module is shown in Fig. 4. For source domain data and style-augmentation data, shallow features f^s and f^a are learned using convolutional networks, respectively, and a learnable domain-aware mask M is proposed to explicitly decompose domain-specific features f_{di} and domain-invariant f_{ds} features in the shallow features.

$$f_{di} = f \times softmax(M_{di}/\tau) \tag{3}$$

$$f_{ds} = f \times softmax(M_{ds}/\tau) \tag{4}$$

where f denotes f^s or f^a, M_{di} and M_{ds} denote domain-invariant and domain-specific masks, respectively. τ denotes the temperature parameter that regulates the binary vector of the domain-aware mask.

The computation of dual contrastive disentanglement follows the following idea: domain-invariant features of source f^s_{di} and augmentation domains f^a_{di} must be the same after decoupling, since they don't change with the domain. Nevertheless, domain-specific features of source f^s_{ds} and augmentation domains f^a_{ds} must be different, due to the style change induced by RandMix. In this case, to ensure that the f^s_{ds} and f^a_{ds} feature representations are different, the correlation between f^s_{ds} and f^a_{ds} is minimized, inspired by mutual information [48]. First define the mutual correlation metric I between f^s_{ds} and f^a_{ds}:

$$I(f^s_{ds};f^a_{ds}) = \mathbb{E}_{p(f^s_{ds},f^a_{ds})}\left[\log\frac{p(f^a_{ds}|f^s_{ds})}{p(f^a_{ds})}\right] \tag{5}$$

Next minimize the mutual information between f^s_{ds} and f^a_{ds}. The mutual information upper bound is defined in [49]:

$$I(f^s_{ds};f^a_{ds}) \le \mathbb{E}_{p(f^s_{ds},f^a_{ds})}\left[\log p(f^a_{ds}|f^s_{ds})\right] - \mathbb{E}_{p(f^s_{ds})p(f^a_{ds})}\left[\log p(f^a_{ds}|f^s_{ds})\right] \tag{6}$$

Since the conditional distribution p is difficult to handle, it is not possible to minimize the upper bound of $I(f^s_{ds};f^a_{ds})$ directly. Therefore, the variational distribution $q_\theta(f^a_{ds}|f^s_{ds})$ is employed to approximate the upper limit of mutual information $\widehat{I}(f^s_{ds};f^a_{ds})$ by a neural network parameterized by θ_q:

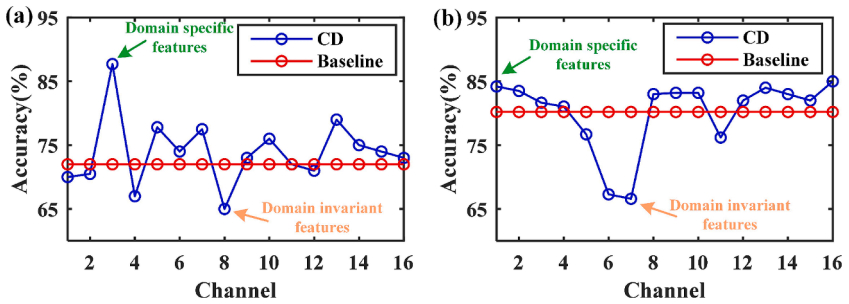

Fig. 3. The relationship between the diagnostic accuracy of the target domain and the index of deleted feature channels. (a) The JNU dataset. (b) The SDUST dataset.

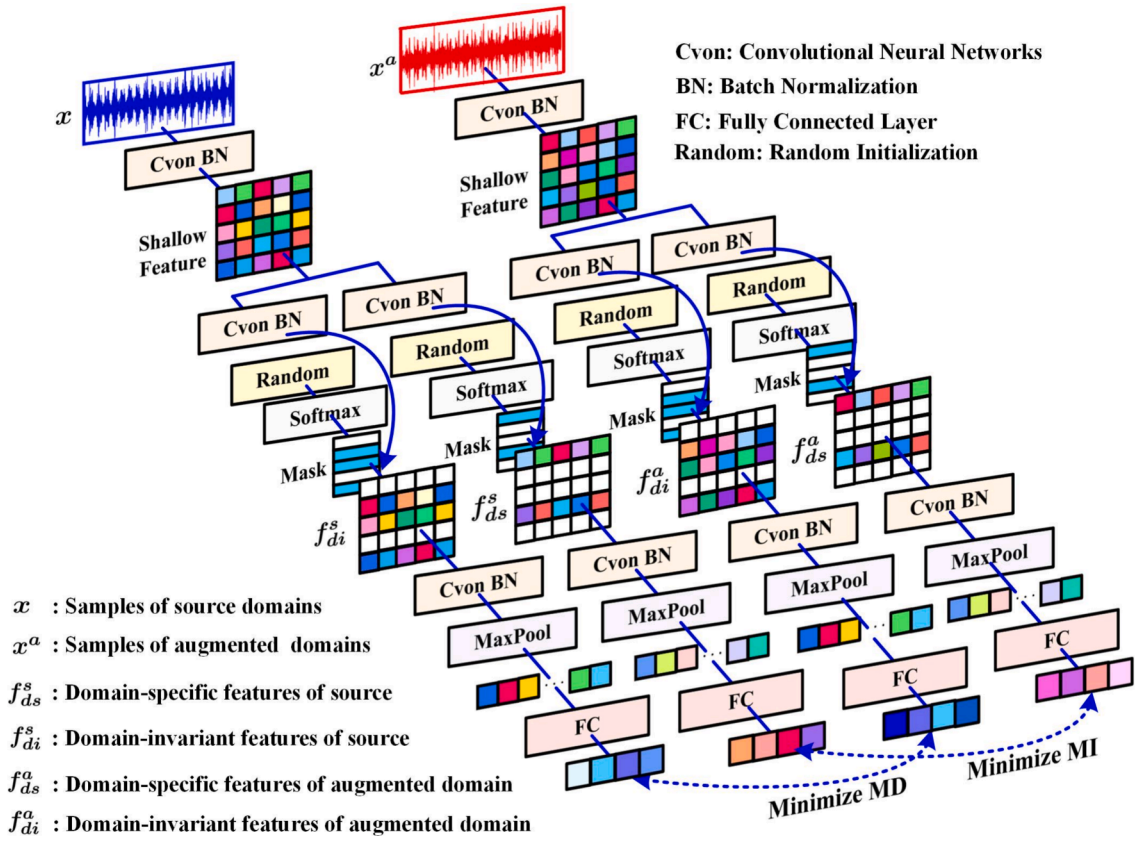

Fig. 4. The detailed process of the dual contrastive disentanglement module.

$$\widehat{I}(f_{ds}^{s};f_{ds}^{a}) = \frac{1}{N}\sum_{i=1}^{N}\left[\log q_{\theta}(f_{ds}^{a}|f_{ds}^{s}) - \frac{1}{N}\sum_{j=1}^{N}\log q_{\theta}(f_{ds}^{a}|f_{ds}^{s})\right] \tag{7}$$

The conditional distribution $p(f_{ds}^{a}|f_{ds}^{s})$ during minimization of equation (7) may be replaced by the variational approximation $q_{\theta}(f_{ds}^{a}|f_{ds}^{s})$ causing $\widehat{I}(f_{ds}^{s};f_{ds}^{a})$ to no longer be an upper bound on the mutual information. If the discrepancy between the two distributions is small $\widehat{I}(f_{ds}^{s};f_{ds}^{a})$ can still be considered reliable. Specifically, the KLD is used to estimate the difference Δ between $\widehat{I}(f_{ds}^{s};f_{ds}^{a})$ and $I(f_{ds}^{s};f_{ds}^{a})$:

$$\begin{aligned}\Delta &= KLD(p(f_{ds}^{a}|f_{ds}^{s}) \parallel q_{\theta}(f_{ds}^{a}|f_{ds}^{s}))\\ &= \mathbb{E}_{p(f_{ds}^{s};f_{ds}^{a})}\left[\log\left(p(f_{ds}^{a}|f_{ds}^{s})p(f_{ds}^{s})\right) - \log\left(q_{\theta}(f_{ds}^{a}|f_{ds}^{s})p(f_{ds}^{s})\right)\right] = \mathbb{E}_{p(f_{ds}^{s};f_{ds}^{a})}\left[\log p(f_{ds}^{a}|f_{ds}^{s})\right] - \mathbb{E}_{p(f_{ds}^{s})p(f_{ds}^{a})}\left[\log p(f_{ds}^{a}|f_{ds}^{s})\right]\end{aligned} \tag{8}$$

The above equation shows that the difference Δ is affected by two terms. Since the first term of Equation (8) is independent of θ_q, we minimize the negative log likelihood L_{li} between f_{ds}^{a} and f_{ds}^{s} rather than minimizing Δ directly:

$$L_{li} = -\frac{1}{N}\sum_{i=1}^{N}\log q_{\theta}(f_{ds}^{a}|f_{ds}^{s}) \tag{9}$$

In addition, to ensure that f_{di}^{a} and f_{di}^{s} feature representations are the same, the Manhattan distance (MD) is considered to measure their discrepancy and minimize it:

$$L_{di} = \frac{1}{N}\sum_{i=1}^{N}|f_{di}^{a} - f_{di}^{s}| \tag{10}$$

The final dual contrastive disentanglement loss L_{DC} can be expressed as:

$$L_{DC} = \widehat{I}(f_{ds}^{s};f_{ds}^{a}) + L_{li} + L_{di} \tag{11}$$

3.3. Adversarial mask disentanglement

The dual contrastive disentanglement module decouples the shallow features of the network into domain-specific features and domain-invariant features, based on the analysis of Venn diagrams Fig. 1, in this section the adversarial mask disentanglement module is proposed to further disentangle the domain-invariant features into class-relevant domain-invariant features and class-irrelevant domain-invariant features. The specific process is shown in Fig. 5. First, domain-invariant features in the source f_{di}^{s} and style-augmentation domains f_{di}^{a} are spliced together along the feature dimensions to form new domain-invariant features f_{di} after convolutional network. Second, the adversarial mask is designed to accurately detect class-relevant and class-irrelevant features, and finally, the KLD loss is proved to ensure that the class-relevant features contain sufficient labeling information.

For the fault diagnosis task, in order to extract the class-relevant feature information to support the diagnostic task, the domain-invariant features are fed into the classifier through a convolutional network for supervised training using labels y:

$$L_C = l_{ce}\left(l_f(f_{di}), y\right) \tag{12}$$

where l_{ce} denotes the cross-entropy loss, l_f denotes the fully connected network. However, relying on the cross-entropy loss alone does not guarantee that the learned features are all strongly correlated with the fault categories, and there may be class-irrelevant features that make a smaller contribution to the classification, thus affecting the generalization of the model. Therefore, we consider decoupling the deep network features into class-relevant features and class-irrelevant features, and eliminating the class-irrelevant features to give full play to the class-relevant features to further improve the generalization ability. Based on the idea of adversarial [16], the adversarial mask module is designed to decouple class-relevant and class-irrelevant features, assuming $z = w(f_{di})$ as a probability vector, where w denotes the learned contribution of each dimension, and $\sum_j z_j = 1$. The class-aware mask is updated using the commonly derivable Gumbel-Softmax trick [50] to learn the contribution of each feature dimension to the diagnostic task:

$$M_j = \max_{l\in\{1,\cdots,c\}} \frac{\exp\left(\left(\log z_j + \xi_j^l\right)/\tau_d\right)}{\sum_{j'=1}^{N}\exp\left(\left(\log z_{j'} + \xi_{j'}^l\right)/\tau_d\right)} \tag{13}$$

$$\xi_j^l = -\log\left(-\log u_j^l\right), u_j^l \sim \text{Uniform}(0,1) \tag{14}$$

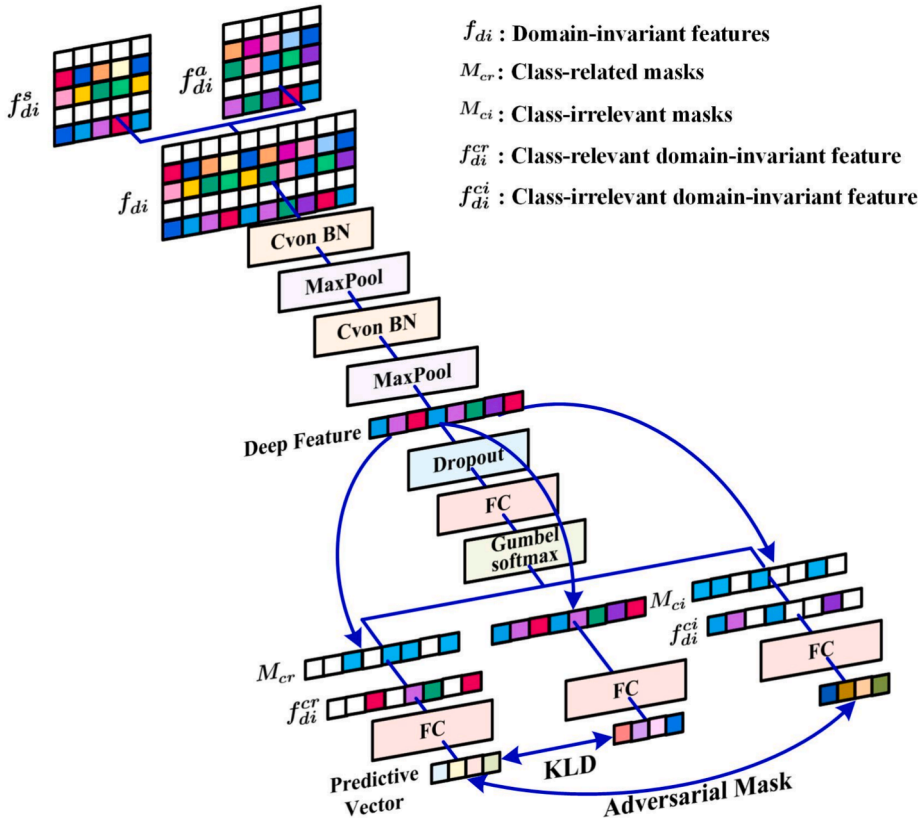

Fig. 5. The detailed process of the adversarial mask disentanglement module.

where $j \in \{1, \cdots N\}$. By multiplying the learned domain-invariant representation f_{di} by the updated class-relevant mask M_{cr}, the class-relevant domain invariant feature f_{di}^{cr} is obtained, which is inputted to the class-relevant feature classifier:

$$L_{CR} = l_{ce}\left(l_f\left(f_{di} \odot M_{cr}\right), y\right) \tag{15}$$

Instead, the class-irrelevant domain-invariant feature f_{di}^{ci} is obtained by multiplying the learned domain-invariant representation f_{di} by the class-irrelevant mask M_{ci}, where $M_{ci} = 1 - M_{cr}$, which is then fed into the class-irrelevant feature classifier:

$$L_{CN} = l_{ce}\left(l_f\left(f_{di} \odot M_{ci}\right), y\right) \tag{16}$$

To further ensure that the optimized class-relevant features retain sufficient diagnostic information, i.e., to ensure that $I(f_{di}; y) = I\left(f_{di}^{cr}; y\right)$, however, direct estimation of the mutual information is challenging, considering that the mutual information $I(f_{di}; y)$ and $I\left(f_{di}^{cr}; y\right)$ can be expressed as:

$$I(f_{di}; y) = H(y) - H(y|f_{di}) \tag{17}$$

$$I\left(f_{di}^{cr}; y\right) = H(y) - H\left(y|f_{di}^{cr}\right) \tag{18}$$

Then estimate the discrepancy between the two mutual information:

$$I(f_{di}; y) - I\left(f_{di}^{cr}; y\right) = H(y|f_{di}) - H\left(y|f_{di}^{cr}\right) = l_{KLD}\left[P_{f_{di}} \middle\| P_{f_{di}^{cr}}\right] \tag{19}$$

Through the above equation, it can be found that minimizing the mutual information between $I(f_{di}; y)$ nd $I\left(f_{di}^{cr}; y\right)$ is equivalent to minimizing $l_{KLD}\left[P_{f_{di}} \middle\| P_{f_{di}^{cr}}\right]$, and therefore, the KLD loss is designed to ensure that the class-relevant features do not lose important diagnosis information:

$$L_{KLD} = l_{KLD}\left[l_f\left(f_{di}\right) \middle\| l_f\left(f_{di}^{cr}\right)\right] \tag{20}$$

3.4. Training and inference

The overall training process of the proposed method can be divided into two phases, the first phase of the dual contrastive disentanglement loss decouples the shallow features of the network into domain-invariant and domain-specific features:

$$L_{step1} = L_{DC} + L_{CR} + L_{CN} \tag{21}$$

The second stage of adversarial mask disentanglement loss further decouples the domain-invariant features of the deeper layers of the network into class-relevant domain-invariant features and class-irrelevant domain-invariant features:

$$L_{step2} = L_{KLD} + L_{CR} - L_{CN} \tag{22}$$

In the inference phase, the final diagnosis is obtained using only domain-invariant class-relevant features f_{di}^{cr} input to the class-relevant feature classifier.

4. Experimental verifications

4.1. Dataset introduce

To validate the validity of the dual disentanglement DG method for fault diagnosis, two rotating machinery test benches with different operating conditions were used for comprehensive experimental analysis, which are shown in Fig. 6. First, to facilitate the replicability of the results, the public bearing dataset of Jiangnan University (JNU), which contains four bearing health states at three

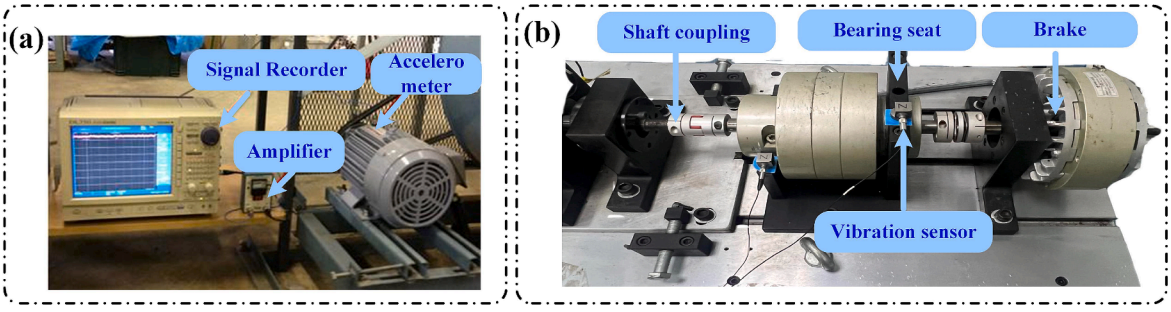

Fig. 6. Composition diagram of two test benches, (a) JNU, and (b) SDUST.

different rotational speeds, is used for validation. Meanwhile, the Shandong University of Science and Technology (SDUST) private dataset, which contains ten bearing health states at three different rotational speeds, was also adopted for validation. The details of the dataset are described below.

4.1.1. JNU public dataset

The JNU dataset uses a centrifugal fan system at three speeds (600, 800 and 1000 rpm) to collect bearing vibration signals for 4 health states (healthy, inner ring failure, outer ring failure and roller failure). The vibration acceleration signals were collected at a sampling frequency of 5 kHz, and each health state was divided into 200 samples, where each sample was 2048 in length, with specific parameters shown in Table 1.

4.1.2. SDUST private dataset

The SDUST private lab bench simulated three failure locations (inner ring, outer ring, and roller) and each failure type was categorized into three failure levels (0.2 mm, 0.4 mm, and 0.6 mm) for a total of 10 health states Vibration data were collected at three different speeds (2000, 2500 and 3000 rpm) and the sampling frequency of the sensors was set to 12.8 kHz. each class has 400 sample datasets of length 2048 are described in detail in Table 2.

4.2. Experimental detail setup

4.2.1. Domain generalization diagnostic task setup

Six domain generalization fault diagnosis tasks are designed for public and private datasets respectively, including three single DG tasks and three mixed-domain generalization tasks. In which the data of different working conditions are combined into a complete failure mode called hybrid domain, and the specific composition details are shown in Table 3. The domain generalization fault diagnosis task setup is shown in Table 4, where data from one domain is used to train the model and test the performance on the unknown domain.

4.2.2. Parameter details

The base feature extraction model in this paper is a five-layer simple convolutional neural network, and the classifier consists of a three-layer fully connected network. The model learning rate and iteration number parameters are 0.0005 and 100. The dual contrastive disentanglement module is set in the third layer of the network with the domain-aware masks temperature parameter set to 0.01, and the adversarial mask disentanglement module is set in the last layer of the convolutional network with the class-aware masks temperature parameter set to 0.01.

4.2.3. Comparative methods

The base model and some state-of-the-art domain generalization methods are chosen for comparison with the method's proposed in this paper.

(1) Base model: the CNN used in this paper without any added strategies.

(2) Advanced domain generalization methods:

L2D [42] ensures semantic consistency while generating more style features to solve the single domain generalization problem.

AMINet [51] uses adversarial mutual information between domain enhancement and classification tasks to obtain generalization features.

MSG-ACN [46] combines a multi-scale style generation strategy with inverse contrast learning to increase feature diversity.

ACL [22] combines anti-causal inference domain shift causes with data augmentation for domain-specific adaptation to improve domain generalization.

HmmSeNet [52] synthesizes data with different distributions of the same semantic information to extract domain invariant features using trainable linear dimensions.

4.3. Results and discussion

4.3.1. Main results

The results of comparing the generalized diagnosis accuracy of this paper's method with other methods are shown in Tables 5 and 6, where the JN public dataset and the SDUST private dataset contain a total of 12 DG diagnosis tasks. Firstly, it can be found that the base model performs the worst, with only 87.87 % and 58.82 % average diagnostic accuracies on the two datasets due to the absence of

Table 1
Introduction of the JNU public dataset.

Working condition	Health condition	Training/test sample	Class label
C1: 600 rpm	Normal	200/200	1
C2: 800 rpm	Inner ring failure	200/200	2
C3: 1000 rpm	Outer ring failure	200/200	3
	Roller failure	200/200	4

Table 2
Introduction of the SDUST private dataset.

Working condition	Health condition	Training/test sample	Class label
D1: 2000 rpm	Normal	200/200	1
D2: 2500 rpm	Inner ring failure 0.2 mm	200/200	2
D3: 3000 rpm	Inner ring failure 0.4 mm	200/200	3
	Inner ring failure 0.6 mm	200/200	4
	Outer ring failure 0.2 mm	200/200	5
	Outer ring failure 0.4 mm	200/200	6
	Outer ring failure 0.6 mm	200/200	7
	Roller failure 0.2 mm	200/200	8
	Roller failure 0.4 mm	200/200	9
	Roller failure 0.6 mm	200/200	10

Table 3
The specific composition details of the mixed domain.

Dataset	Mixed domain	Working condition	Class
JNU dataset	C4	600 rpm	1, 2
		800 rpm	3, 4
	C5	800 rpm	1, 2
		1000 rpm	3, 4
	C6	600 rpm	1, 2
		1000 rpm	3, 4
SDUST dataset	D4	2000 rpm	1, 2, 3, 4, 5
		2500 rpm	6, 7, 8, 9, 10
	D5	2500 rpm	1, 2, 3, 4, 5
		3000 rpm	6, 7, 8, 9, 10
	D6	2000 rpm	1, 2, 3, 4, 5
		3000 rpm	6, 7, 8, 9, 10

Table 4
Domain generalization fault diagnosis task.

Dataset	Task	Source	Target	Dataset	Task	Source	Target
JNU	T1	C1	C2, C3	SDUST	T7	D1	D5, D6
dataset	T2	C2	C1, C3	dataset	T8	D2	D4, D6
	T3	C3	C1, C2		T9	D3	D4, D5
	T4	C4	C3		T10	D4	D3
	T5	C5	C1		T11	D5	D1
	T6	C6	C2		T12	D6	D2

Table 5
Comparison of the results (%) of different methods for six generalization tasks on the JNU dataset.

Method	T1	T2	T3	T4	T5	T6	Average
CNN	86.74 ± 4.85	82.52 ± 5.12	92.33 ± 3.15	86.62 ± 4.56	88.75 ± 4.14	90.24 ± 3.81	87.87
L2D	94.49 ± 2.03	88.85 ± 3.40	95.40 ± 2.24	89.71 ± 2.96	93.68 ± 3.88	92.28 ± 3.34	92.40
AMINet	94.70 ± 2.45	90.12 ± 2.56	95.05 ± 1.57	88.44 ± 3.15	94.33 ± 3.42	93.71 ± 2.25	92.73
MSG-ACN	95.49 ± 1.97	93.45 ± 1.74	94.70 ± 1.85	91.68 ± 2.43	93.91 ± 3.17	94.15 ± 1.78	93.90
ACL	96.26 ± 2.17	94.24 ± 1.62	95.26 ± 1.63	89.45 ± 2.31	94.23 ± 2.74	93.55 ± 2.02	93.83
HmmSeNet	97.29 ± 1.48	95.21 ± 1.83	96.37 ± 1.51	95.98 ± 1.78	94.81 ± 2.85	94.04 ± 1.93	95.62
Our method	**98.73 ± 1.32**	**96.85 ± 1.35**	**98.16 ± 1.21**	**98.37 ± 0.72**	**97.25 ± 1.71**	**98.62 ± 1.05**	**98.00**

any added strategies. Secondly, the current state-of-the-art generalization methods show a more significant improvement in accuracy compared to the base model, which is attributed to the fact that these methods all perform data augmentation or feature diversification combined with a generalization strategy to enhance the model representation. Finally, the dual disentanglement network was able to decouple the class-relevant domain invariant features to achieve better generalization results, achieving the highest average accuracies of 98.00 % and 88.63 % on both datasets.

To intuitively demonstrate the capability of the proposed method and related approaches in handling domain discrepancies, kernel density estimation (KDE) plots of feature distributions for the source and target domains in the JNU and SDUST datasets were generated, as shown in Figs. 7 and 8. The KDE plots provide insights into the overlap and alignment between the two domains under different models. The degree of overlap is critical for evaluating the effectiveness of each model in transferring knowledge from the

Table 6
Comparison of the results (%) of different methods for six generalization tasks on the SDUST dataset.

Method	T7	T8	T9	T10	T11	T12	Average
CNN	49.37 ± 6.45	66.23 ± 5.37	67.75 ± 4.75	59.70 ± 5.87	33.97 ± 8.65	75.89 ± 4.85	58.82
L2D	89.05 ± 3.52	83.82 ± 3.52	74.93 ± 4.33	81.62 ± 3.12	64.55 ± 6.59	84.08 ± 3.15	79.68
AMINet	90.57 ± 3.14	86.34 ± 3.28	74.66 ± 3.61	83.69 ± 3.51	62.78 ± 5.85	85.13 ± 3.54	80.53
MSG-ACN	91.24 ± 2.78	86.13 ± 2.86	78.19 ± 3.15	88.24 ± 2.84	67.05 ± 4.21	85.91 ± 3.82	82.79
ACL	91.81 ± 2.89	85.95 ± 2.93	80.58 ± 3.42	87.12 ± 2.77	66.49 ± 4.73	87.79 ± 3.11	83.29
HmmSeNet	91.01 ± 2.65	88.78 ± 2.19	82.82 ± 3.75	89.87 ± 2.45	70.09 ± 5.13	86.40 ± 3.48	84.83
Our method	**93.14 ± 1.58**	**91.36 ± 2.01**	**83.75 ± 3.24**	**93.12 ± 1.98**	**75.97 ± 4.26**	**94.41 ± 2.73**	**88.63**

source domain to the target domain, with higher overlap generally indicating better generalization performance. In subfigure (a), the overlap is less pronounced, and the prominent peaks in the target domain are not well aligned with those in the source domain. This indicates a larger discrepancy between the two domains, which may adversely affect the model's performance on the target data. In contrast, in subfigures (b)–(f), the overlap becomes more apparent compared to the baseline method (CNN), with improved consistency between the prominent peaks in the source and target domains. Finally, in subfigure (g), a high degree of overlap is observed, indicating that the feature distributions of the source and target domains are closely aligned. This highlights the effective generalization capability of the proposed method in transferring knowledge from the source domain to the target domain.

Fig. 9 presents the comparison of testing accuracy and two-stage loss curves throughout the training process for (a) the JNU dataset and (b) the SDUST dataset. In both datasets, the green curves represent the testing accuracy, while the red and blue curves correspond to the first-stage loss (Lstep1) and second-stage loss (Lstep2), respectively. For the JNU dataset (Fig. 9a), the testing accuracy increases rapidly during the initial stages, stabilizing at around 90 %, while both losses (Lstep1 and Lstep2) decrease steadily. Lstep1 declines rapidly and then maintains a more stable value, whereas Lstep2 remains lower and gradually levels off. Similarly, for the SDUST dataset (Fig. 9b), the testing accuracy rises sharply during the early stages, approaching 95 %, while both loss curves decrease over time, exhibiting a trend similar to Fig. 9a. These results indicate that the proposed method demonstrates a consistent pattern across both datasets, with increasing accuracy and decreasing losses as training progresses. This highlights the effectiveness and stability of the proposed method in learning and optimizing the model.

4.3.2. Ablation Study

4.3.2.1. Comparison of accuracy of ablation methods. Ablation experiments are designed to verify the effectiveness of the three components of this paper's methodology, where the RandMix module refers to style enhancement of the source domain data and training using both the source domain and style-enhanced data, the dual contrastive disentanglement module refers to decoupling the

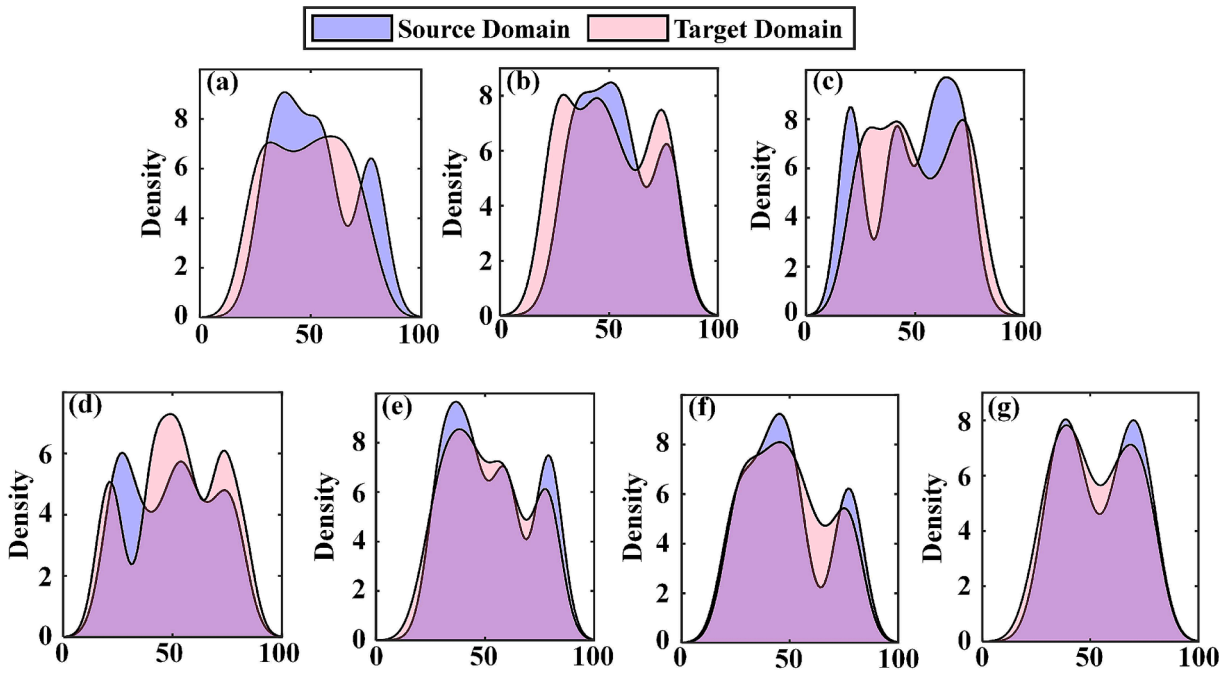

Fig. 7. KDE of feature distributions in the source and target domains of a generalization task in the JNU dataset. (a) CNN, (b) L2D, (c) AMINet, (d) MSG-CAN, (e) ACL, (f) HmmSeNet and (g) Our method.

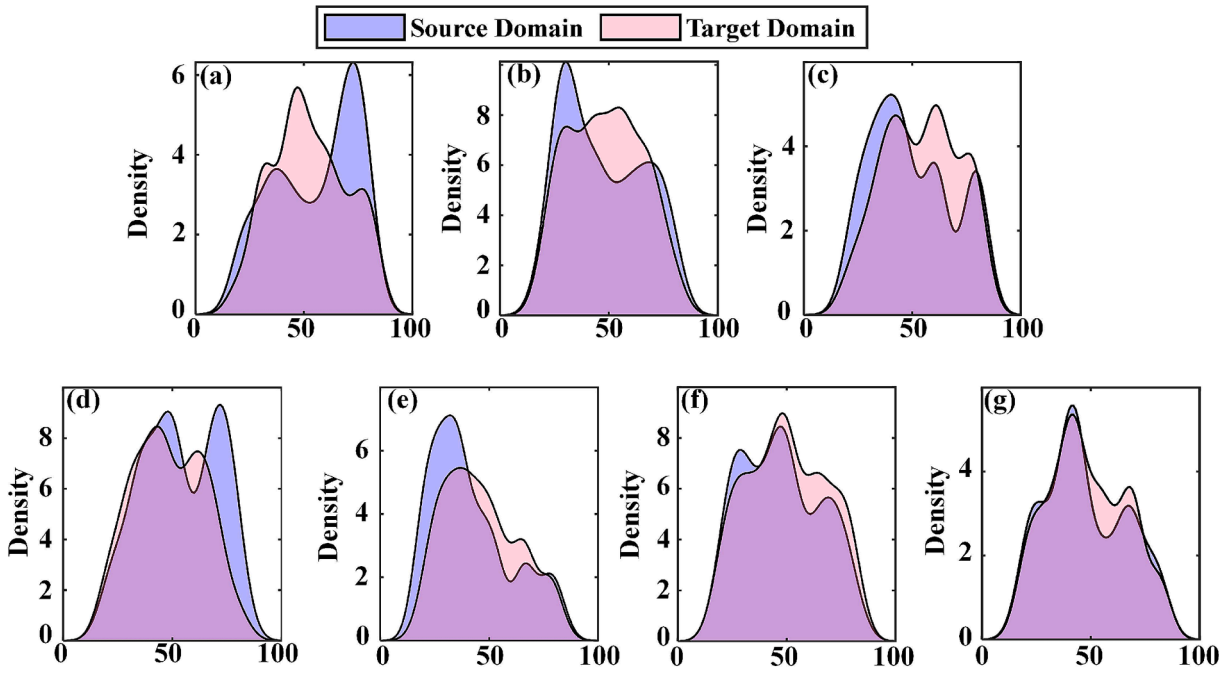

Fig. 8. KDE of feature distributions in the source and target domains of a generalization task in the SDUST dataset. (a) CNN, (b) L2D, (c) AMINet, (d) MSG-CAN, (e) ACL, (f) HmmSeNet and (g) Our method.

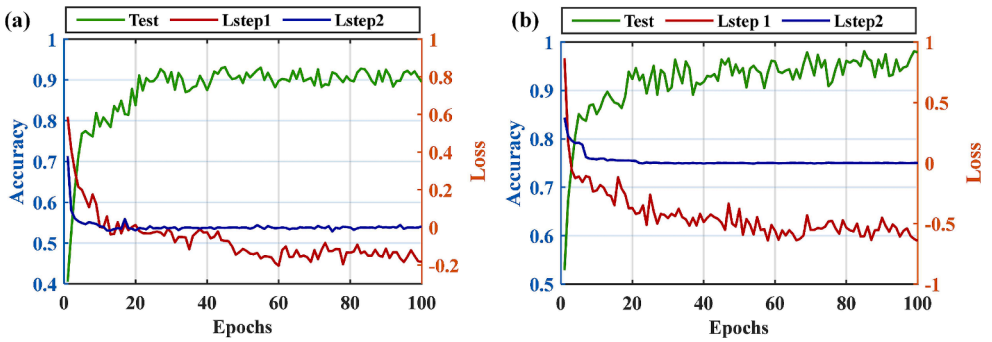

Fig. 9. Comparison of testing accuracy and two stage loss curves during training process: (a) the JNU dataset, (b) the SDUST dataset.

shallow features of the network into domain-specific and domain-invariant features and training the model using the domain-invariant features, and the adversarial mask disentanglement module refers to decoupling the deep features of the network into class-relevant features and class-irrelevant features. Table 7 illustrates the diagnostic performance of the ablation experiments conducted on the two datasets. It is evident that the style enhancement module, RandMix, has significantly enhanced the model's generalization performance. This is attributed to the fact that the generated diverse data can effectively simulate domain shifts. Secondly, the dual contrastive disentanglement module can decouple domain invariant features without relying on domain labels, further enhancing the model's generalization ability. The phenomenon that −extracting domain invariant features can improve domain generalization performance has been widely verified. Similarly, the decoupling of class-relevant features by the adversarial mask disentanglement

Table 7
Comparison of results (%) of different ablation methods on two datasets.

Method	RandMix	Dual Contrastive	Adversarial Mask	JNU dataset	SDUST dataset
Basic Model	–	–	–	87.87	58.82
Method1 (M1)	√	–	–	91.34	76.11
Method2 (M2)	√	√	–	94.12	84.49
Method2 (M3)	√	–	√	96.58	85.87
Our Method	√	√	√	**98.00**	**88.63**

module further enhances the model's generalization ability, suggesting that the elimination of class-irrelevant features is crucial for effective generalization. Therefore, the dual disentanglement network proposed in this paper further decouples class-relevant features based on the extraction of domain-invariant features to achieve the highest diagnostic accuracy

4.3.2.2. Visual Comparison of ablation methods. To illustrate the efficacy of each strategy in addressing the domain generalization fault diagnosis task, the features acquired through the various ablation methods on the two generalization tasks are depicted in two dimensions using the t-SNE technique, as illustrated in Figs. 10 and 11. The colors and shapes in the figure serve to indicate the respective health state and domain. First, the t-SNE clustering diagrams of the features of the source and target domains of the base model are shown in Fig. 10(a) and 11(a). It is evident that the features of the same class in different domains are not clustered together, indicating that there are significant domain discrepancies between the source and target domains. This observation suggests that the ordinary model is unable to accomplish the desired generalization task. An examination of the t-SNE clustering in Fig. 10(b) and 11(b) of method 2 reveals that the distance between features of the same class in different domains is reduced, yet some degree of overlap between different classes persists. This suggests that the decoupled domain-invariant features mitigate the domain discrepancy, yet there remain instances where classes are confounded due to the lack of focus on class information. Through Fig. 10(c) and 11(c), it is found that the adversarial mask module removes the class-irrelevant information so that different classes are clearly distinguished from each other, and due to the presence of domain-specific information, it results in some features of the same class from different domains not being clustered together. Finally, Fig. 10(d) and 11(d) illustrate the t-SNE features extracted by the dual disentanglement network. The features of the same class in different domains are clustered together, and the different classes are clearly distinguished. This indicates that the dual disentanglement network fully combines the advantages of the dual contrastive disentanglement and adversarial mask disentanglement modules. Furthermore, the extracted class-relevant domain invariant features achieve a superior domain generalization effect

4.3.2.3. Research on the Decoupling Performance of Dual Contrastive Disentanglement Modules. To illustrate the efficacy of the dual contrastive disentanglement module in a more intuitive manner, two generalized scenarios are selected for each dataset to plot the kernel density estimation of domain-invariant features versus domain-specific features after decoupling. These are presented in Figs. 12 and 13. The domain-invariant features and domain-specific features of the source and target domains are reduced to a single dimension through the use of t-SNE for kernel density estimation. As evidenced by the plots, the discrepancy in the distribution of the domain-specific features after decoupling between the source and target domains is more pronounced, while the similarity in the distribution between the domain-invariant features of the source and target domains is more pronounced. This fully demonstrates that dual contrastive disentanglement can disentangle similarities and discrepancies. Based on the above phenomenon, it verifies the rationality of this paper to use the idea of contrast to decouple the channel features at the shallow level of the network, and to identify and exclude channels with domain-specific features from the extracted vibration signal features.

To intuitively observe domain-invariant features and domain-specific features, we extract the features from the first layer of the convolutional network and compare them with the signal features. Meanwhile, we removed the activation function and pooling layer from the first layer of the convolutional network to better observe the differences between the signals and the disentangled features. The comparison results of the signals and disentangled features on the two datasets are shown in Figs. 14 and 15. For the JN dataset, we visualized the time domain signals, enhanced signals, and the features learned from the first convolutional layer, specifically channel 1 (domain-invariant features) and channel 16 (domain-specific features). The results show that while the augmented signals differ from the original signals due to simulated condition changes, channel 1 maintains a consistent trend across both, highlighting its robustness to operating condition variations. In contrast, channel 16 exhibits significant fluctuations that fail to align with the signal trends, indicating sensitivity to condition changes. Similarly, for the SDUST dataset, we visualized channel 3 (domain-invariant features) and channel 6 (domain-specific features) after disentanglement. The findings confirm that domain-invariant features are highly stable under varying operating conditions, further validating the robustness of the proposed approach.

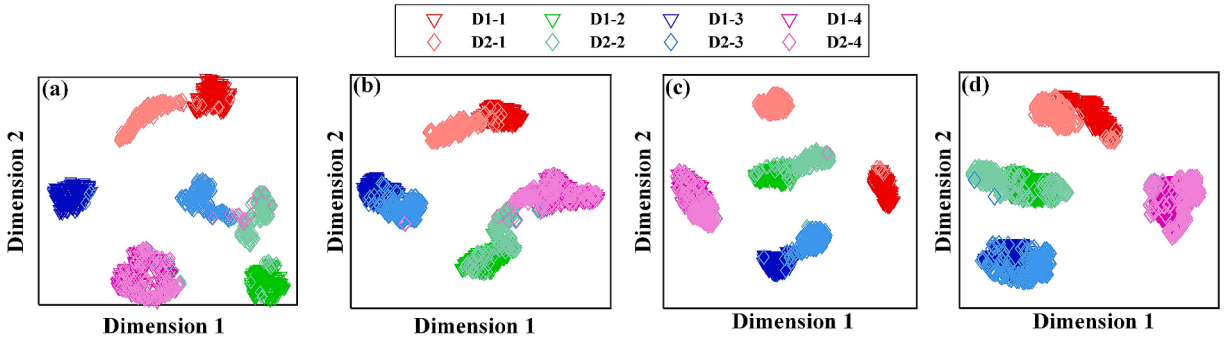

Fig. 10. A comparative 2D visualization of feature representations from the JNU dataset using different ablation techniques: (a) the standard model, (b) the M2 variant, (c) the M3 variant, and (d) our proposed approach.

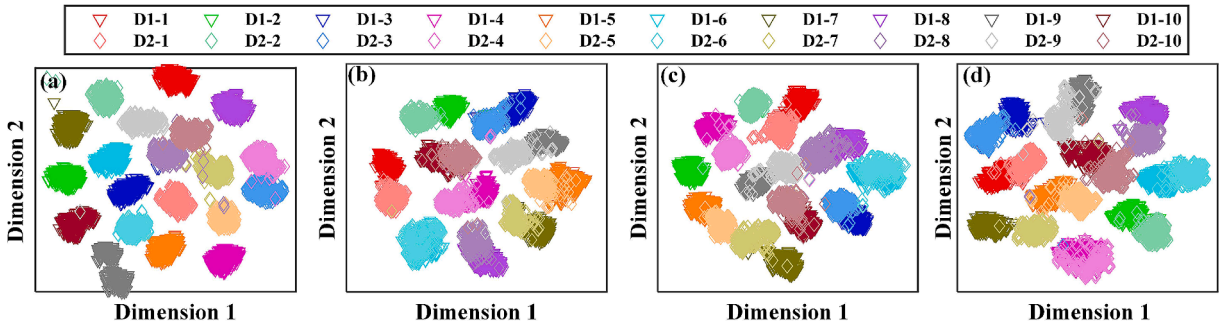

Fig. 11. A comparative 2D visualization of feature representations from the SDUST dataset using different ablation techniques: (a) the standard model, (b) the M2 variant, (c) the M3 variant, and (d) our proposed approach.

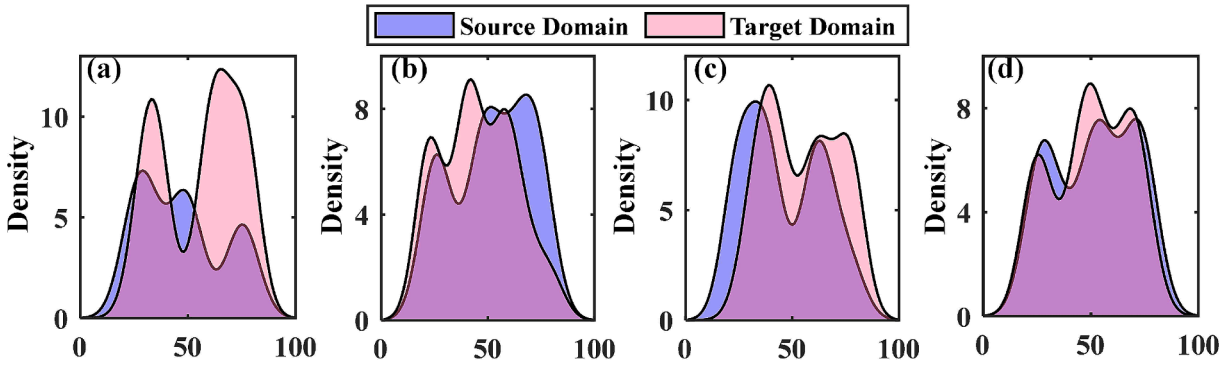

Fig. 12. Kernel density estimation is performed on the feature distributions for two generalization tasks in the JNU dataset. (a) Domain specific features under task T1, (b) domain invariant features under task T1, (c) domain specific features under task T2, and (d) domain invariant features under task T2.

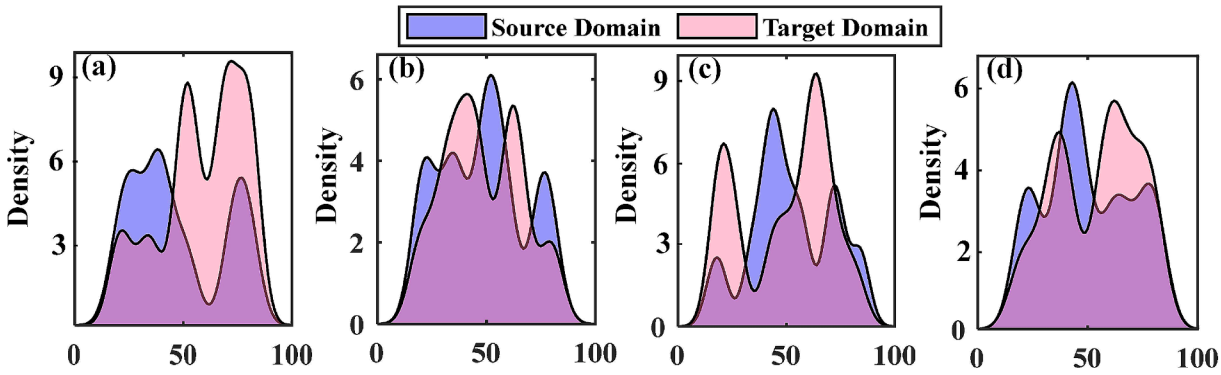

Fig. 13. Kernel density estimation is performed on the feature distributions for two generalization tasks in the SDUST dataset. (a) Domain specific features under task T7, (b) domain invariant features under task T7, (c) domain specific features under task T8, and (d) domain invariant features under task T8.

4.3.2.4. Research on the Decoupling Performance of Adversarial Mask Disentanglement Modules. Further, also to visualize that the adversarial mask disentanglement module is able to disentangle class-relevant versus class-irrelevant features, t-SNE plots of class-relevant versus class-irrelevant features of the target domain after adversarial mask disentanglement are drawn for the two generalization tasks, as shown in Figs. 16 and 17. From the Fig. 16(a) and 17(a), it can be found that the features of the same class of the target domain features of the non-disentangled are clustered together, but there is an overlap between individual classes, which may be due to the fact that the inclusion of class-irrelevant features in the non-disentangled features affects the differentiation between faults. While observing the downscaled visualization of the decoupled class-relevant features in Fig. 16(b) and 17(b) it can be observed that there are clear boundaries between different classes and fewer classes are misclassified. Finally, in Fig. 16(c) and 17(c), it can be

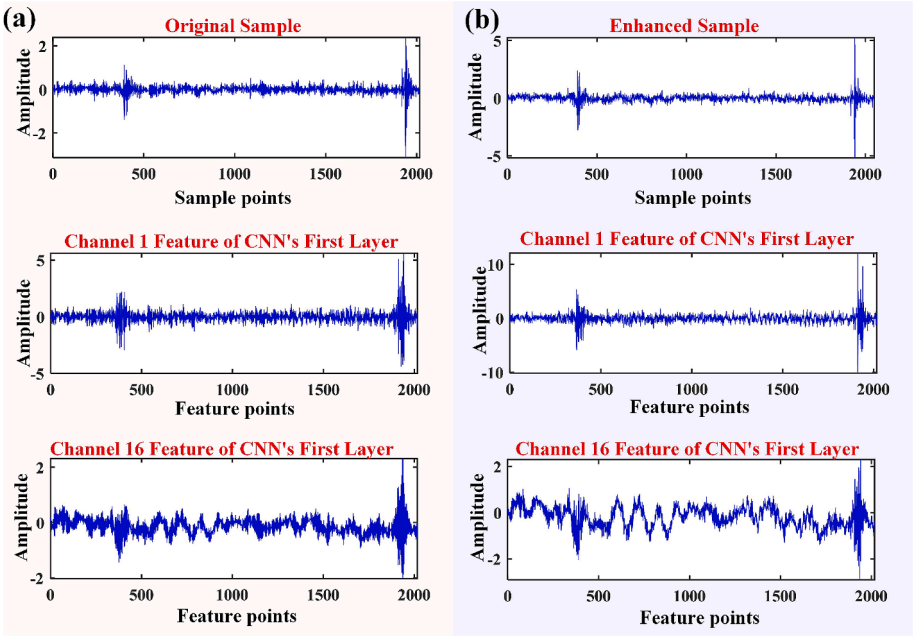

Fig. 14. Comparison of samples, Channel 1 features, and Channel 16 features in the JN dataset: (a) Original sample and disentangled features, (b) Enhanced sample and disentangled features.

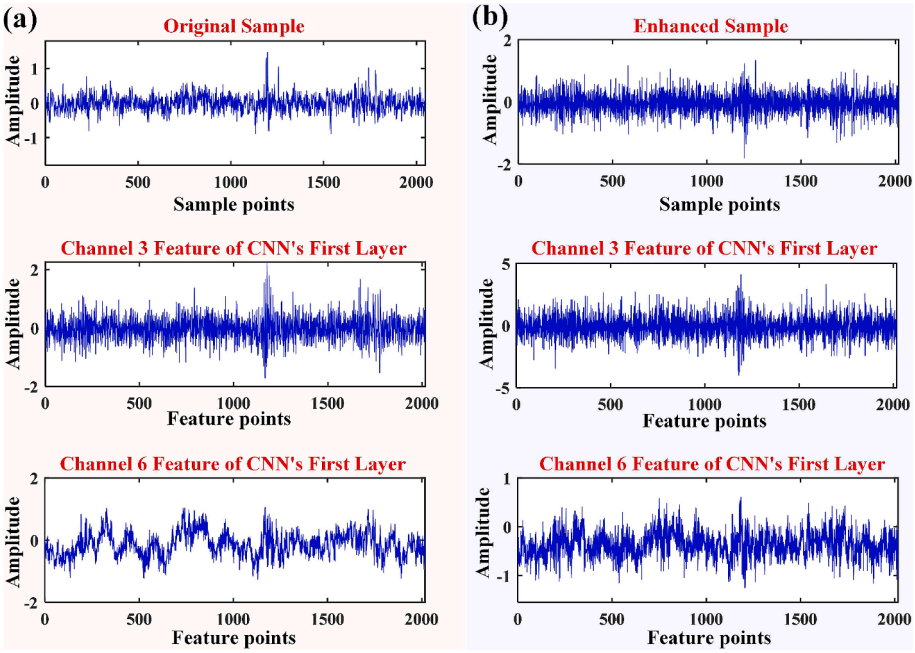

Fig. 15. Comparison of samples, Channel 3 features, and Channel 6 features in the SDUST dataset: (a) Original sample and disentangled features, (b) Enhanced sample and disentangled features.

observed that the clustering of class-irrelevant features is very poor, and features of different classes are mixed together, indicating that there is no class information embedded in the decoupled class-irrelevant features. This indicates that adversarial mask disentanglement can clearly disentangle class-relevant features from class-irrelevant features and avoid the effect of class-irrelevant features on model generalization.

Since the deep features input into the classifier (including class-relevant and class-irrelevant features) are abstract, it is difficult to directly observe the differences and roles between class-relevant and class-irrelevant features. Therefore, we employed Class

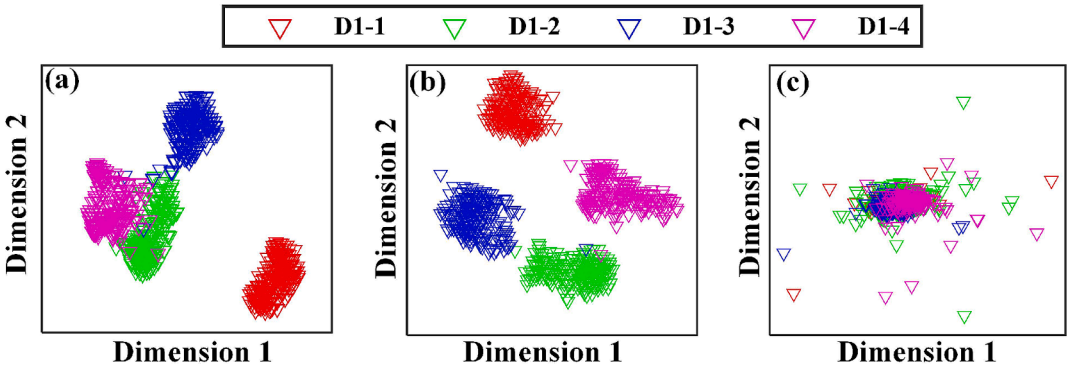

Fig. 16. 2D visualization of adversarial mask decoupling features in JNU dataset generalization task. (a) Non-disentangled features, (b) disentanglement class-relevant features, (c) disentanglement class-irrelevant features.

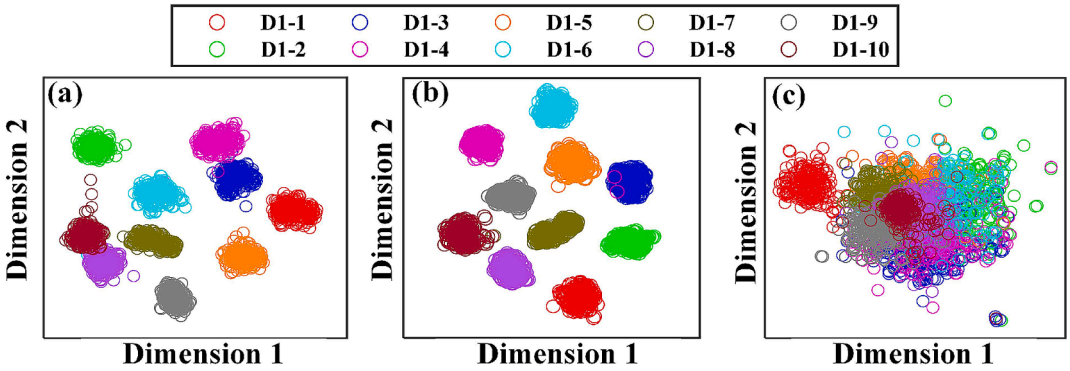

Fig. 17. 2D visualization of adversarial mask decoupling features in SDUST dataset generalization task. (a) Non-disentangled features, (b) disentanglement class-relevant features, (c) disentanglement class-irrelevant features.

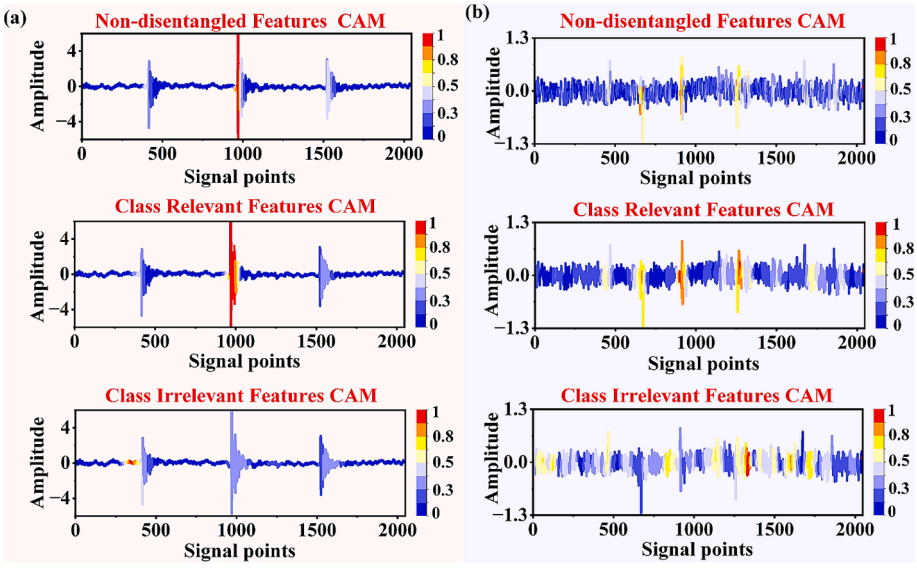

Fig. 18. CAM comparison of non-disentangled features, class-relevant features, and class-irrelevant features: (a) Samples from the JN dataset, (b) Samples from the SDUST dataset.

Activation Mapping (CAM) to visualize the attention regions of the deep features (Fig. 18), by visualizing the activation area of the vibration signal, it can be clarified whether the model has paid attention to the features related to the fault. In the visualization, red areas represent high attention, and blue areas represent low attention. The observations are as follows: (1) The model with non-disentangled deep features can moderately focus on the impact features in vibration signals. (2) Compared to the model with non-disentangled features, the model with class-relevant features demonstrates a clearer focus on the impact regions, emphasizing critical task-related information. The activation locations of the fault signals are mainly concentrated near the impact region, indicating that the class-relevant deep features can focus on more fault features, and it can also be found that these sections with larger activation levels have a certain periodicity. (3) The model with class-irrelevant features does not focus on fault-related impact components but instead pays excessive attention to the stable parts of the signal, which is detrimental to fault classification. This indicates that our method can effectively decouple class related and class-irrelevant features, and removing class-irrelevant features allows the model to concentrate on task-relevant information, improving generalization classification performance.

4.3.3. Mask parameter analysis

The setting of the hyperparameters of the respective masks in the dual contrastive disentanglement module and the adversarial mask disentanglement module relates to the degree of decoupling and thus affects the final generalization accuracy. First, the mask temperature coefficients of both the dual contrast disentanglement module and the adversarial mask disentanglement module are generated to varying degrees, and three masks with different temperature coefficients are plotted as shown in Figs. 19 and 20. It can be found that the mask is a binary form with information complementarity between the masks, at the same time as the temperature coefficient increases the degree of the mask gradually decreases, taking into account that smaller temperature values can better ensure the complementarity between the features, a larger temperature value leads to a reduction in the degree of the mask may result in the leakage of feature information cannot be completely decoupled features. Next, the effect of different temperature coefficients on the generalization results is explored on the two datasets, and the results are shown in Fig. 21, from which it can be seen that the sensitivity of the generalization results to the temperature coefficients is lower, and when the two temperature hyperparameters are between 0.005–0.1, the model has a high level of generalization accuracy, and this result proves that the above empirical analysis, which provides a reference for the adjustment of the parameters of the model.

5. Conclusion

This paper proposes a dual disentanglement network-based domain generalization method for rotating machinery fault diagnosis, which does not rely on domain labels. The dual contrastive disentanglement module and the adversarial mask disentanglement module are designed to decouple domain-invariant features and class-relevant features, respectively. The effectiveness of this method in generalization tasks has been validated through extensive experiments conducted on both public and private datasets. The proposed method achieved average accuracies of 98.00 % and 88.63 % on the two datasets, representing improvements of 10.13 % and 29.81 %, respectively, compared to baseline models. Moreover, the proposed method outperformed other state-of-the-art domain generalization diagnostic methods in terms of both accuracy and standard deviation. The experimental results demonstrate the following: (1) By performing dual contrastive training on source data and style-enhanced data through shallow features in the network, domain-specific representations and domain-invariant representations can be effectively disentangled without relying on domain labels. This approach improved the average generalization accuracy by 2.4 % and 8.38 % on the two datasets, respectively. (2) The adversarial mask disentanglement module can accurately identify class-relevant and class-irrelevant features. Eliminating class-irrelevant features enhances the model's generalization ability, leading to average generalization accuracy improvements of 5.24 % and 9.76 % on the two datasets, respectively. (3) A higher masking degree in the dual disentanglement network helps to ensure complementarity between features, whereas a lower masking degree may result in feature information leakage. These findings not only highlight the effectiveness of the proposed approach in tackling domain generalization challenges but also demonstrate its potential for practical applications in real-world rotating machinery fault diagnosis tasks. By eliminating reliance on domain labels, this method addresses a significant barrier in industrial diagnostic systems where domain-specific information is often unavailable or difficult to obtain. Additionally, the proposed framework can serve as a foundation for exploring other industrial fault diagnosis tasks.

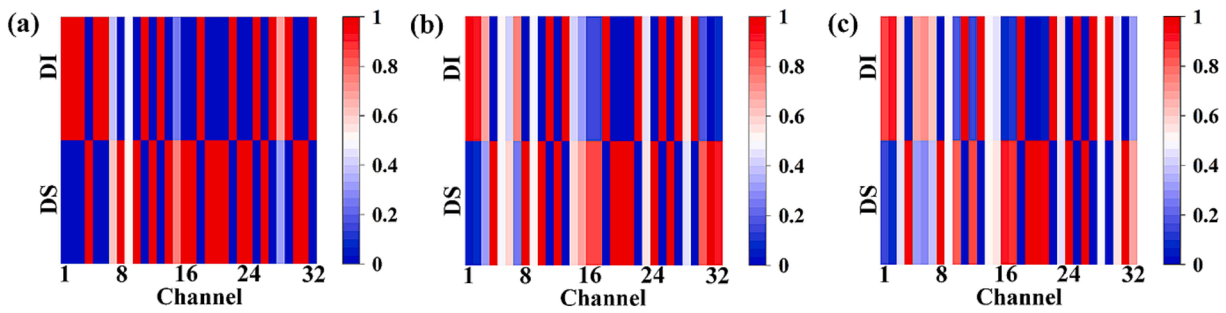

Fig. 19. The influence of different temperature parameters of the dual contrastive disentanglement module on the mask. (a) $\tau = 0.01$, (b) $\tau = 0.1$, (c) $\tau = 0.5$.

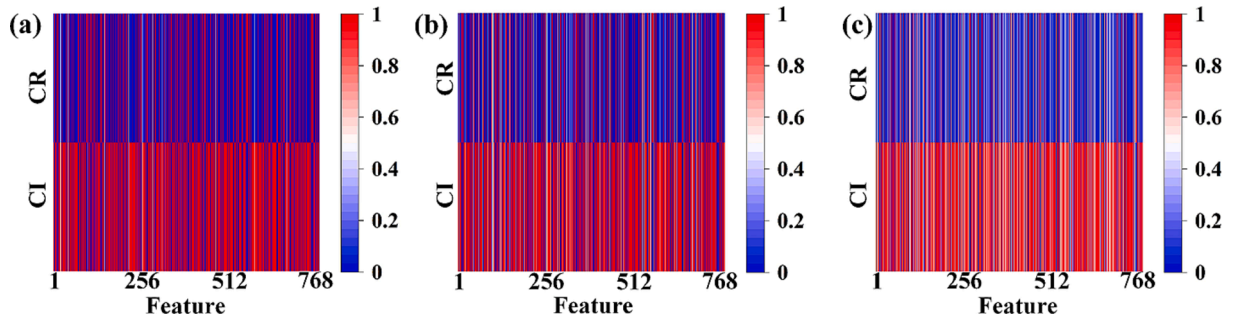

Fig. 20. The influence of different temperature parameters of the adversarial mask disentanglement module on the mask. (a) $\tau_d = 0.01$, (b) $\tau_d = 0.1$, (c) $\tau_d = 0.5$.

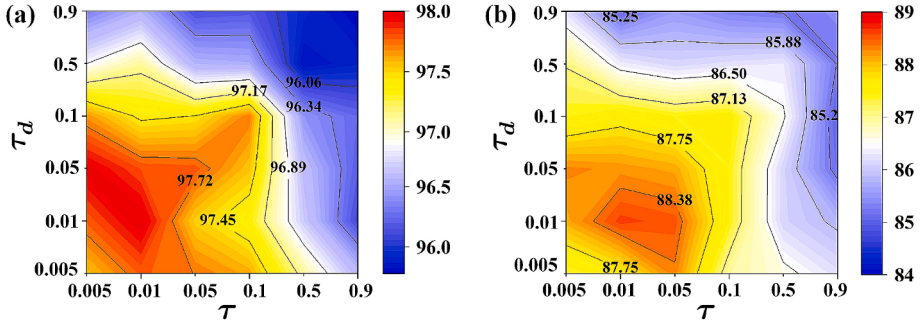

Fig. 21. The impact of different mask hyperparameters on the generalization accuracy of two datasets, (a) JN dataset and (b) SDUST dataset.

In future research, we plan to explore the following directions: (1) Further explore the impact of decoupling domain-specific class-relevant information on generalization performance by leveraging domain-specific class-relevant features to assist in domain generalization for fault diagnosis tasks. (2) Investigate the relationship between domain features, class features, and signal characteristics in-depth by incorporating the physical properties of rotating machinery signals to assist the model in learning domain and class knowledge.

CRediT authorship contribution statement

Guowei Zhang: Writing – original draft, Validation, Software, Methodology, Conceptualization. **Xianguang Kong:** Supervision, Project administration, Funding acquisition. **Hongbo Ma:** Validation, Resources, Investigation, Funding acquisition. **Qibin Wang:** Validation, Conceptualization. **Jingli Du:** Supervision, Data curation. **Jinrui Wang:** Data curation, Conceptualization.

Declaration of competing interest

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

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2022YFB3706803), and the National Natural Science Foundation of China [grant number 52335002, 52375120].

Data availability

Data will be made available on request.

References

[1] L. Pinciroli, P. Baraldi, E. Zio, Maintenance optimization in industry 4.0[J], Reliab. Eng. Syst. Saf. 234 (2023) 109204.
[2] S. Gawde, S. Patil, S. Kumar, et al., Multi-fault diagnosis of Industrial Rotating Machines using Data-driven approach: A review of two decades of research[J], Eng. Appl. Artif. Intel. 123 (2023) 106139.

[3] P. Zhou, S. Chen, Q. He, et al., Rotating machinery fault-induced vibration signal modulation effects: A review with mechanisms, extraction methods and applications for diagnosis[J], Mech. Syst. Sig. Process. 200 (2023) 110489.
[4] F. Perez-Sanjines, C. Peeters, T. Verstraeten, et al., Fleet-based early fault detection of wind turbine gearboxes using physics-informed deep learning based on cyclic spectral coherence[J], Mech. Syst. Sig. Process. 185 (2023) 109760.
[5] G. Zhang, X. Kong, Q. Wang, et al., Multi-source partial domain adaptation method based on pseudo-balanced target domain for fault diagnosis[J], Knowl.-Based Syst. 284 (2024) 111255.
[6] G. Zhang, X. Kong, J. Du, et al., Adaptive multispace adjustable sparse filtering: A sparse feature learning method for intelligent fault diagnosis of rotating machinery[J], Eng. Appl. Artif. Intel. 120 (2023) 105847.
[7] D. Liu, L. Cui, H. Wang, Rotating machinery fault diagnosis under time-varying speeds: A review[J], IEEE Sens. J. (2023).
[8] Y. Xiao, H. Shao, J. Wang, et al., Bayesian variational transformer: A generalizable model for rotating machinery fault diagnosis[J], Mech. Syst. Sig. Process. 207 (2024) 110936.
[9] X. Yu, S. Wang, H. Xu, et al., Intelligent fault diagnosis of rotating machinery under variable working conditions based on deep transfer learning with fusion of local and global time–frequency features[J], Struct. Health Monit. 23 (4) (2024) 2238–2254.
[10] K. Zhao, Z. Liu, J. Li, et al., Self-paced decentralized federated transfer framework for rotating machinery fault diagnosis with multiple domains[J], Mech. Syst. Sig. Process. 211 (2024) 111258.
[11] B. Zheng, J. Huang, X. Ma, et al., An unsupervised transfer learning method based on SOCNN and FBNN and its application on bearing fault diagnosis[J], Mech. Syst. Sig. Process. 208 (2024) 111047.
[12] Z. Li, J. Ma, J. Wu, et al., A Gated Recurrent Generative Transfer Learning Network for Fault Diagnostics Considering Imbalanced Data and Variable Working Conditions[J], IEEE Trans. Neural Networks Learn. Syst. (2024).
[13] H. Wang, M. Li, Z. Liu, et al., Rotary Machinery Fault Diagnosis Based on Split Attention Mechanism and Graph Convolutional Domain Adaptive Adversarial Network[J], IEEE Sens. J. (2024).
[14] S. Zhang, S.U. Lei, G.U. Jiefei, et al., Rotating machinery fault detection and diagnosis based on deep domain adaptation: A survey[J], Chin. J. Aeronaut. 36 (1) (2023) 45–74.
[15] X. Wang, B. She, Z. Shi, et al., Partial adversarial domain adaptation by dual-domain alignment for fault diagnosis of rotating machines[J], ISA Trans. 136 (2023) 455–467.
[16] X. Chen, H. Shao, Y. Xiao, et al., Collaborative fault diagnosis of rotating machinery via dual adversarial guided unsupervised multi-domain adaptation network [J], Mech. Syst. Sig. Process. 198 (2023) 110427.
[17] J. Jiao, H. Li, Inter-to Intradomain: A Progressive Adaptation Method for Machine Fault Diagnosis[J], IEEE Trans. Ind. Inf. (2023).
[18] J. Jiao, H. Li, T. Zhang, et al., Source-free adaptation diagnosis for rotating machinery[J], IEEE Trans. Ind. Inf. 19 (9) (2022) 9586–9595.
[19] B. Han, X. Zhang, J. Wang, et al., Hybrid distance-guided adversarial network for intelligent fault diagnosis under different working conditions[J], Measurement 176 (2021) 109197.
[20] H. Li, J. Jiao, Z. Liu, et al., Trustworthy Bayesian deep learning framework for uncertainty quantification and confidence calibration: Application in machinery fault diagnosis[J], Reliab. Eng. Syst. Saf. 255 (2025) 110657.
[21] Q. Qian, J. Zhou, Y. Qin, Relationship transfer domain generalization network for rotating machinery fault diagnosis under different working conditions[J], IEEE Trans. Ind. Inf. 19 (9) (2023) 9898–9908.
[22] G. Zhang, X. Kong, Q. Wang, et al., Single domain generalization method based on anti-causal learning for rotating machinery fault diagnosis[J], Reliab. Eng. Syst. Saf. 250 (2024) 110252.
[23] Y. Wang, J. Gao, W. Wang, et al., Curriculum learning-based domain generalization for cross-domain fault diagnosis with category shift[J], Mech. Syst. Sig. Process. 212 (2024) 111295.
[24] Z. Fan, Q. Xu, C. Jiang, et al., Deep mixed domain generalization network for intelligent fault diagnosis under unseen conditions[J], IEEE Trans. Ind. Electron. 71 (1) (2023) 965–974.
[25] C. Zhao, E. Zio, W. Shen, Multi-domain Class-imbalance Generalization with Fault Relationship-induced Augmentation for Intelligent Fault Diagnosis[J], IEEE Trans. Instrum. Meas. (2024).
[26] L. Ren, T. Mo, X. Cheng, Meta-learning based domain generalization framework for fault diagnosis with gradient aligning and semantic matching[J], IEEE Trans. Ind. Inf. 20 (1) (2023) 754–764.
[27] H. Wang, X. Bai, S. Wang, et al., Generalization on unseen domains via model-agnostic learning for intelligent fault diagnosis[J], IEEE Trans. Instrum. Meas. 71 (2022) 1–11.
[28] L. Chen, Q. Li, C. Shen, et al., Adversarial domain-invariant generalization: A generic domain-regressive framework for bearing fault diagnosis under unseen conditions[J], IEEE Trans. Ind. Inf. 18 (3) (2021) 1790–1800.
[29] Z. Shi, J. Chen, X. Zhang, et al., A reliable feature-assisted contrastive generalization net for intelligent fault diagnosis under unseen machines and working conditions[J], Mech. Syst. Sig. Process. 188 (2023) 110011.
[30] J. Li, C. Shen, L. Kong, et al., A new adversarial domain generalization network based on class boundary feature detection for bearing fault diagnosis[J], IEEE Trans. Instrum. Meas. 71 (2022) 1–9.
[31] M. Ragab, Z. Chen, W. Zhang, et al., Conditional contrastive domain generalization for fault diagnosis[J], IEEE Trans. Instrum. Meas. 71 (2022) 1–12.
[32] R. Wang, W. Huang, Y. Lu, et al., A novel domain generalization network with multidomain specific auxiliary classifiers for machinery fault diagnosis under unseen working conditions[J], Reliab. Eng. Syst. Saf. 238 (2023) 109463.
[33] Z. An, X. Jiang, J. Liu, Mode-decoupling auto-encoder for machinery fault diagnosis under unknown working conditions[J], IEEE Trans. Ind. Inf. (2023).
[34] H. Ren, J. Wang, W. Huang, et al., Domain-invariant feature fusion networks for semi-supervised generalization fault diagnosis[J], Eng. Appl. Artif. Intel. 126 (2023) 107117.
[35] S. Xie, P. Xia, H. Zhang, Domain adaptation with domain specific information and feature disentanglement for bearing fault diagnosis, Meas. Sci. Technol. 35 (5) (2024) 056101.
[36] L. Jia, T.W.S. Chow, Y. Yuan, Causal disentanglement domain generalization for time-series signal fault diagnosis[J], Neural Netw. 172 (2024) 106099.
[37] S. Jia, Y. Li, X. Wang, et al., Deep causal factorization network: A novel domain generalization method for cross-machine bearing fault diagnosis[J], Mech. Syst. Sig. Process. 192 (2023) 110228.
[38] C. Guo, Z. Shang, J. Ren, et al., CIS2N: Causal independence and sparse shift network for rotating machinery fault diagnosis in unseen domains[J], Reliab. Eng. Syst. Saf. 251 (2024) 110381.
[39] H. Ma, J. Wei, G. Zhang, et al., Causality-inspired multi-source domain generalization method for intelligent fault diagnosis under unknown operating conditions[J], Reliab. Eng. Syst. Saf. 252 (2024) 110439.
[40] Y. Wei, J. Ye, Z. Huang, et al., Online prototype learning for online continual learning[C]//Proceedings of the IEEE/CVF, International Conference on Computer Vision. (2023:) 18764–18774.
[41] X. Zhou, X. Deng, Z. Liu, et al., Domain generalized open-set intelligent fault diagnosis based on feature disentanglement meta-learning[J], Meas. Sci. Technol. 35 (11) (2024) 115001.
[42] Z. Wang, Y. Luo, R. Qiu, et al., Learning to diversify for single domain generalization[C]//Proceedings of the IEEE/CVF, International Conference on Computer Vision. (2021:) 834–843.
[43] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4401-4410.
[44] Liu C, Wang L, Lyu L, et al. Deja vu: Continual model generalization for unseen domains[J]. arXiv preprint arXiv:2301.10418, 2023.
[45] Liu S, Jin X, Yang X, et al. StyDeSty: Min-Max Stylization and Destylization for Single Domain Generalization[J]. arXiv preprint arXiv:2406.00275, 2024.

[46] J. Wang, H. Ren, C. Shen, et al., Multi-scale style generative and adversarial contrastive networks for single domain generalization fault diagnosis[J], Reliab. Eng. Syst. Saf. 243 (2024) 109879.
[47] H. Zhou, J. Lan, R. Liu, et al., Deconstructing lottery tickets: Zeros, signs, and the supermask[J], Adv. Neural Inf. Proces. Syst. 32 (2019).
[48] M.I. Belghazi, A. Baratin, S. Rajeshwar, et al., Mutual information neural estimation[C]//International conference on machine learning, PMLR (2018) 531–540.
[49] P. Cheng, W. Hao, S. Dai, et al., Club: A contrastive log-ratio upper bound of mutual information[C]//International conference on machine learning, PMLR (2020) 1779–1788.
[50] Jang E, Gu S, Poole B. Categorical reparameterization with gumbel-softmax[J]. arXiv preprint arXiv:1611.01144, 2016.
[51] C. Zhao, W. Shen, Adversarial mutual information-guided single domain generalization network for intelligent fault diagnosis[J], IEEE Trans. Ind. Inf. 19 (3) (2022) 2909–2918.
[52] J. Tang, X. Ding, C. Wei, et al., HmmSeNet: A Novel Single Domain Generalization Equipment Fault Diagnosis Under Unknown Working Speed Using Histogram Matching Mixup[J], IEEE Trans. Ind. Inf. (2024).