A Source-Free Universal Domain Adaptation Method for Fault Diagnosis Based on Dual-View Disentanglement and Orthogonal Decomposition

Huijuan Hao[1,2], Feifei Zhang[1,2], Jinqiang Bai[1,2*], Qingyan Ding[1,2], and Huanqing Xu[1,2]
[1]*Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputing Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan, China*
[2]*Shandong Provincial Key Laboratory of Industrial Network and Information System Security, Shandong Fundamental Research Center for Computer Science, Jinan, China*
Email:183520498@qq.com, 13188932594@163.com, baijq@sdas.org*, dingqy@sdas.org,3848682717@qq.com

***Abstract*—Cross-operating-condition fault diagnosis in real industrial scenarios typically encounters two major challenges. First, source-domain data are often inaccessible due to privacy and compliance constraints, which hinders the deployment of conventional domain adaptation methods that rely on joint training with both source and target data. Second, the relationship between the source and target label spaces is unknown a priori, which often leads to misalignment between target samples and source categories. To address these issues, this paper focuses on Source-Free Universal Domain Adaptation (SF-UniDA) for fault diagnosis and proposes a two-stage target adaptation framework. In the source training stage, we leverage RandMix-based data augmentation to construct a second view and impose a supervised contrastive learning objective to enforce cross-view consistency among samples of the same class, thereby encouraging domain-invariant representation learning and feature disentanglement. In the target adaptation stage, we first construct three criteria to determine whether unknown classes exist in the target domain; we then propose a pseudo-label construction method that integrates DBSCAN clustering and orthogonal feature decomposition by fusing cluster-level and instance-level information. Experiments on the Case Western Reserve University (CWRU) and Paderborn University (PU) datasets under universal-domain settings demonstrate the effectiveness of the proposed fault diagnosis model.**

***Index Terms*—fault diagnosis, universal domain adaptation, transfer learning, feature decomposition**

I. Introduction

In modern intelligent manufacturing, conducting highly reliable fault diagnosis research for rotating machinery is of practical significance for ensuring industrial safety and reducing operation and maintenance costs. Over the past decade, the introduction of deep learning has reshaped the fault diagnosis paradigm, promoting a shift from traditional approaches that rely on handcrafted feature engineering and heuristic rules to data-driven, end-to-end intelligent diagnosis. A variety of deep neural networks have been applied to fault diagnosis, such as convolutional neural networks (CNNs) [1], recurrent neural networks (RNNs) [2], and Transformer models [3]. However, data distributions under different operating conditions are often inconsistent, resulting in distribution discrepancies between the source and target domains (i.e., domain shift) in cross-domain fault diagnosis, which in turn degrades the performance of deep learning-based methods.

To address the domain shift issue, domain adaptation (DA) [4] has been widely adopted. DA transfers diagnostic knowledge learned from labeled data in one operating condition (source domain) to unlabeled data in other conditions (target domain). Most existing DA methods assume that the source and target domains share the same label set. For example, Li et al. [5] integrated the Domain-Adversarial Neural Network (DANN) with a GAN to simultaneously generate few-shot samples and perform cross-operating-condition fault diagnosis. For the open-set scenario, Zhao et al. [6] used an auxiliary domain discriminator to identify suspicious unknown samples and then jointly employed weighted domain adversarial learning and separative adversarial learning to enlarge the margin between known and unknown classes. Although these methods can be effective, they typically assume that the label-space relationship between the source and target domains is known in advance and are designed for a specific label-shift scenario; therefore, they are difficult to transfer directly to other label-shift settings.

To cope with complex real-world industrial environments, Universal Domain Adaptation (UniDA) has been proposed. UniDA requires no prior knowledge about the relationship between the source and target label spaces; that is, neither the shared classes nor the numbers of domain-private classes in the source and target domains are known in advance. However, industrial data are subject to strict privacy and compliance constraints, and conventional unsupervised domain adaptation methods often expose practical bottlenecks in real deployments. Motivated by these issues, this paper investigates the more challenging Source-free Universal Domain Adaptation (SF-UniDA) setting for bearing fault diagnosis and proposes a two-stage target adaptation framework. In the source training stage, we train a source model with strong generalization capability; in the target adaptation stage, we perform target-

*Corresponding author.

domain adaptation using the trained source model, thereby achieving isolation between source-domain data and target-domain adaptation. The main contributions of this paper are summarized as follows.

(1) At the shallow layers of the network, we propose a dual-view contrastive disentanglement module based on data augmentation. Without relying on any target-domain data, it disentangles domain-invariant and domain-specific features, enabling the training of a model with strong generalization ability.

(2) During target-domain adaptation, we construct three criteria to determine whether unknown classes exist, namely unknown-subspace projection strength, predictive uncertainty, and similarity to source anchors.

(3) Based on DBSCAN clustering and orthogonal feature decomposition, we propose a pseudo-label construction method that fuses cluster-level and instance-level information.

II. Related Work

A. Universal Domain Adaptation

According to the set relationship between the source and target label spaces, DA methods can be categorized into four types: (1) Closed-set DA (CDA), where the target label set is identical to that of the source domain; (2) Partial-set DA (PDA), where the target label set is a proper subset of the source label set; (3) Open-set DA (OSDA), where the source label set is a proper subset of the target label set, indicating the presence of unknown classes in the target domain; and (4) Open-partial DA (OPDA), which assumes that both source and target domains contain their own private classes.

In real-world industrial fault diagnosis, the key challenge of cross-domain fault diagnosis under UniDA lies in that the diagnostic model must simultaneously cope with distribution shift and the above four class-transfer scenarios. Qian et al. [7] proposed a multi-source transfer fault diagnosis method via Adaptive Intermediate Class-Wise Distribution Alignment (AICDA), which improves diagnostic accuracy; however, its performance in open-set transfer tasks remains limited. Zhang et al. [8] identified target-private samples based on the prediction consistency of two classifiers, but the alignment performance may degrade under class imbalance. Lin et al. [9] proposed a UniDA method that combines self-supervised clustering-based matching with extreme value theory, improving diagnosis under imbalanced data, but it is sensitive to hyperparameters. Xu et al. [10] introduced a channel-level feature transfer criterion together with a meta-recognition calibration mechanism to reduce channel-wise feature drift between shared-class and private-class samples; nevertheless, it relies on multiple hyperparameters and threshold calibration. Liu et al. [11] leveraged the self-supervised clustering structure of target-domain data to learn discriminative representations and used source class centers to guide alignment; however, when the target domain is noisy, unknown-class discrimination may become unstable.

B. Source-Free Universal Domain Adaptation

However, although universal domain adaptation has achieved notable progress, traditional approaches that require simultaneous access to both source and target data for training become impractical in modern intelligent industrial systems with strict privacy constraints. Consequently, SF-UniDA has attracted growing attention to address this challenge [12]. Liu et al. [13] proposed a source-free UniDA method that generates pseudo-labels via global and local clustering and further incorporates contrastive learning; however, it is sensitive to the initial pseudo-label quality and may degrade when the target-domain distribution is highly complex. Li et al. [14] developed a neighborhood-search-based strategy to construct reliable sample pairs, together with an entropy-based unknown-fault detection mechanism, but it fails to effectively disentangle compound faults. To address these limitations, this paper proposes a fault diagnosis method that leverages the intrinsic structural information of the target domain and exhibits strong generalization capability.

III. Method

A. Data Augmentation

We introduce style-perturbed RandMix2D to construct a second-view pseudo-source domain. Specifically, given an in-batch sample x, we randomly shuffle the batch to obtain a style sample x_s, and compute its channel-wise mean and standard deviation (μ_s, σ_s). We then apply Adaptive Instance Normalization (AdaIN) [15] to the original sample as:

$$\mathrm{AdaIN}(x; \mu_s, \sigma_s) = \left(\frac{x - \mu_x}{\sigma_x}\right) \odot \sigma_s + \mu_s, \tag{1}$$

where (μ_x, σ_x) denote the channel-wise statistics of x. We introduce a mild random jitter to the style standard deviation as $\sigma_s \leftarrow \sigma_s \odot (1 + \epsilon_{\mathrm{jit}})$ with $\epsilon_{\mathrm{jit}} \sim \mathcal{N}(0, \delta^2)$. After style replacement, RandMix2D further injects weak Gaussian noise. We sample a mixing coefficient λ from a Beta distribution and linearly mix the original image with the stylized image as $\tilde{x} = \lambda x + (1 - \lambda)\,\mathrm{AdaIN}(x; \mu_s, \sigma_s)$, where $\lambda \sim \mathrm{Beta}(\alpha, \alpha)$. In implementation, we clip λ as $\lambda \leftarrow \mathrm{clip}(\lambda, 0.2, 0.8)$.

B. Dual-view Contrastive Disentanglement Module

To enable the network to learn more stable and transferable features during source-domain training without accessing any target-domain data, we introduce a dual-view contrastive mechanism on shallow features, as illustrated in Fig. 1. We concatenate the original and augmented samples to form a batch of size $2B$. We define a learnable parameter $M \in \mathbb{R}^{2\times C\times 1\times 1}$ and apply a softmax with temperature τ to obtain two sets of channel weights: $W = \mathrm{softmax}(M/\tau),\ W \in \mathbb{R}^{2\times C\times 1\times 1}$. Accordingly, we perform channel-wise weighted decomposition of the shallow feature F: $F_{\mathrm{inv}} = F \odot W_{\mathrm{inv}}, \quad F_{\mathrm{spc}} = F \odot W_{\mathrm{spc}}$, where $\odot$ denotes element-wise multiplication, and inv and spc represent the two shallow-feature components after channel-wise routing. We expect the domain-invariant representations of the same class to remain consistent across the two views. Therefore, we

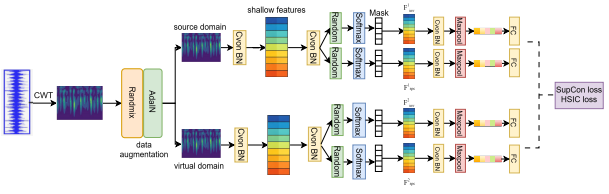

Fig. 1. Dual-view contrastive disentanglement module in the shallow feature layers.

feed the domain-invariant features from the two views into a projector to obtain z^1_{inv} and z^2_{inv}, and employ a supervised contrastive loss to enforce cross-view consistency. Taking view 1 as the anchor and view 2 as the contrast set, we define the positive set $P(i)$ as samples belonging to the same class. The loss is formulated as:

$$L^{\text{inv}}_{\text{nce}} = -\frac{1}{B}\sum_{i=1}^{B}\frac{1}{|P(i)|}\sum_{p\in P(i)} \log\frac{\exp\left(\text{sim}\left(z^1_{\text{inv},i}, z^2_{\text{inv},p}\right)/\tau_{\text{nce}}\right)}{\sum_{j=1}^{B}\exp\left(\text{sim}\left(z^1_{\text{inv},i}, z^2_{\text{inv},j}\right)/\tau_{\text{nce}}\right)}. \tag{2}$$

where $\text{sim}(\cdot,\cdot)$ denotes cosine similarity and τ_{nce} is the temperature parameter. To further reduce the coupling between the two components, we regularize the statistical dependence between F_{inv} and F_{spc} within each view using the normalized Hilbert–Schmidt Independence Criterion (HSIC) with a multi-scale RBF kernel, and average it over the two views:

$$L_{\text{hsic}} = \frac{1}{2}\Big(\text{nHSIC}(v^1_{\text{inv}}, v^1_{\text{spc}}) + \text{nHSIC}(v^2_{\text{inv}}, v^2_{\text{spc}})\Big), \tag{3}$$

where v_{inv} and v_{spc} are the vectors obtained by applying global average pooling. This regularization encourages a clearer division of labor between the two components and reduces information entanglement. We adopt label-smoothed cross-entropy as the classification loss:

$$L_{\text{ce}} = -\frac{1}{2B}\sum_{i=1}^{2B}\sum_{k=1}^{K}\tilde{y}_{ik}\log p_{ik}, \quad \tilde{y}_{ik} = (1-\epsilon_{\text{ls}})y_{ik} + \epsilon_{\text{ls}}/K. \tag{4}$$

where p denotes the softmax probabilities, y is the one-hot label, and ϵ_{ls} is the smoothing factor. Overall, when considering only the above three constraints, the objective for the source training stage is formulated as:

$$L_{\text{total}} = L_{\text{ce}} + \lambda_{\text{nce}}L^{\text{inv}}_{\text{nce}} + \lambda_{\text{hsic}}L_{\text{hsic}}, \tag{5}$$

where λ_{nce} and λ_{hsic} are the weighting coefficients of the corresponding losses, which are set to 0.6 and 0.5 in our experiments, respectively.

C. Unknown-Class Existence Detection in the Adaptation Stage

In SF-UniDA, we freeze the classifier to preserve the source-specific decision function by setting $h^t_\theta = h^s_\theta$, and only learn a target-specific feature extractor g^t_θ, as illustrated in Fig. 2. We then construct three criteria to determine whether unknown classes exist. The first criterion is the similarity between target samples and source known classes. We treat the weight vectors of the linear classification head trained on the source domain as source class prototypes $\{s_c\}_{c=1}^{C}$, and define the source-similarity score for the i-th target sample as

$$q^{\text{src}}_i = 1 - \max_{c\in\{1,\dots,C\}}\left\langle z^t_i, s_c\right\rangle, \tag{6}$$

where z^t_i denotes the normalized feature of the target sample, s_c is the prototype of the c-th source class, and $\langle\cdot,\cdot\rangle$ denotes cosine similarity. The second criterion comes from predictive uncertainty. We adopt the normalized entropy:

$$q^{\text{ent}}_i = \frac{H(p_i)}{\log C}, \qquad H(p_i) = -\sum_{c=1}^{C} p_{ic}\log p_{ic}, \tag{7}$$

where p_i denotes the predicted probability over the C known classes and $H(\cdot)$ is the information entropy. The third criterion is the unknown-subspace energy induced by orthogonal decomposition. Consider the source classifier weight matrix $W_{\text{cls}} \in \mathbb{R}^{C\times D}$, where the subspace spanned by its row vectors can be viewed as the set of discriminative directions for source known classes. To obtain a stable orthogonal basis, we perform Singular Value Decomposition (SVD) on W_{cls}:

$$W_{\text{cls}} = U\Sigma V^\top, \tag{8}$$

where U and V are orthogonal matrices and Σ is a diagonal matrix. Accordingly, the target feature z^t_i can be decomposed into the sum of two orthogonal components in the known and unknown subspaces:

$$z^t_i = \underbrace{\sum_{n=1}^{C}\ell_n v_n}_{z^t_{i,\text{knw}}} + \underbrace{\sum_{n=C+1}^{D}\ell_n v_n}_{z^t_{i,\text{unk}}}, \qquad \left\|z^t_{i,\text{knw}}\right\|_2^2 + \left\|z^t_{i,\text{unk}}\right\|_2^2 = 1. \tag{9}$$

where ℓ_n denotes the projection coefficient of z^t_i onto the basis vector v_n, and $z^t_{i,\text{knw}}$ and $z^t_{i,\text{unk}}$ are the components in the known and unknown subspaces, respectively. We use the ℓ_2 norm of the unknown component as the third criterion:

$$q^{\text{upn}}_i = \left\|z^t_{i,\text{unk}}\right\|_2. \tag{10}$$

After obtaining the three criteria q^{src}_i, q^{ent}_i, and q^{upn}_i, we perform a bimodality test (via two-component GMM model selection) on each evidence sequence. If the high-mean component has a non-negligible proportion, we regard the corresponding evidence as supporting the existence of unknown classes.

D. Cluster-Level Pseudo-Label Construction for Unknown Classes

When unknown classes are detected in the target domain, we first perform DBSCAN clustering on the target features to mine suspected unknown clusters. The DBSCAN parameters ε and `min_samples` are selected via grid search. After clustering, we identify unknown clusters solely based on their similarity to the source prototypes. Specifically, for each target

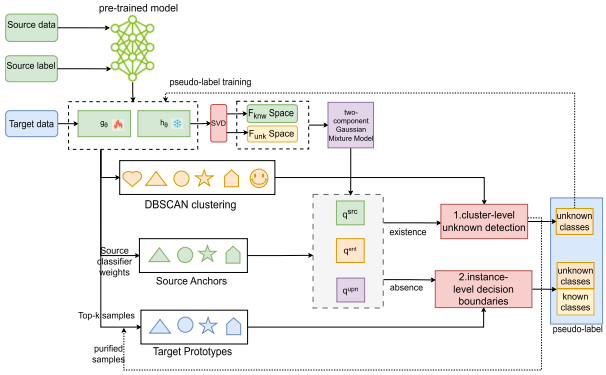

Fig. 2. Implementation details of target-domain adaptation

sample, we compute the maximum similarity to source class prototypes $\{s_c\}_{c=1}^{C}$ as $s_i^{\mathrm{src}} = \max_{c\in\{1,\dots,C\}} \max(0, \langle z_i^t, s_c\rangle)$, and then set a threshold using a quantile of $\{s_i^{\mathrm{src}}\}_{i=1}^{N}$ as $\tau_{\mathrm{src}} = \mathrm{Quantile}\big(\{s_i^{\mathrm{src}}\}_{i=1}^{N}, q\big)$. For any non-noise cluster C_k, we compute the average within-cluster similarity $\bar{s}_k^{\mathrm{src}} = \frac{1}{|C_k|}\sum_{i\in C_k} s_i^{\mathrm{src}}$. If $\bar{s}_k^{\mathrm{src}} < \tau_{\mathrm{src}}$, we mark the entire cluster as a suspected unknown cluster and assign the cluster-level unknown indicator $I_{\mathrm{clu}}(i) = 1$. When unknown classes are not detected, we skip this cluster-level unknown mining.

E. Instance-Level Pseudo-Label Construction

Afterwards, for each known class c, we select the Top-K target samples with the highest predicted probability for class c to form a Top-K set Ω_c, and compute the target prototype t_c as the mean feature of Ω_c. During the Top-K selection, we explicitly exclude samples that are marked as unknown clusters (i.e., $I_{\mathrm{clu}}(i) = 1$), thereby reducing the interference of unknown samples on the target prototypes.

Following prior work [16], after constructing the target prototypes, we compute the target similarity and source similarity and define a fused score: $u_i(c) = \sqrt{(1 - e^{-s_i^{\mathrm{tar}}(c)}) \cdot e^{\,s_i^{\mathrm{src}}(c)-1}}$,where $s_i^{\mathrm{tar}}(c) = \max(0, \langle z_i^t, t_c\rangle)$ and $s_i^{\mathrm{src}}(c) = \max(0, \langle z_i^t, s_c\rangle)$ denote the similarity to the target prototype and the similarity to the source prototype, respectively. We assign the initial pseudo-label as $\hat{y}_i = \arg\max_c u_i(c)$. We fit a two-component Gaussian Mixture Model (2-GMM) to $\left\|z_{i,\mathrm{unk}}^t\right\|_2$ and obtain the larger-mean component μ_2. For each class c, we compute the mean unknown energy of its Top-K samples as a prior π_c, which yields a class-conditional threshold: $\tau_{i,c} = \pi_c + u_i(c) \cdot \max(0, \mu_2 - \pi_c)$,where μ_2 denotes the larger mean in the 2-GMM, π_c is the unknown-energy prior for class c, and $\tau_{i,c}$ is the threshold for sample i under class c. The final unknown-class decision is given by

$$\tilde{y}_i = \begin{cases} C+1, & I_{\mathrm{clu}}(i) = 1 \text{ or } \left\|z_{i,\mathrm{unk}}^t\right\|_2 \geq \tau_{i,\hat{y}_i}, \\ \hat{y}_i, & \text{otherwise}, \end{cases} \tag{11}$$

where $I_{\mathrm{clu}}(i)$ is the cluster-level unknown indicator and $C+1$ denotes the index assigned to the unknown class.

The loss function in the target-domain adaptation stage consists of three terms. The first term is the weighted pseudo-label cross-entropy:

$$L_{\mathrm{psd}} = \frac{1}{B}\sum_{i=1}^{B}\left[-w_i \sum_{c=1}^{C} \tilde{y}_{ic} \log(p_{ic} + \epsilon_{\mathrm{num}})\right], \tag{12}$$

where $\tilde{y}_{ic} = \mathbb{I}[\tilde{y}_i = c]$ for $c \in \{1, \dots, C\}$, p_i is the current predicted probability, ϵ_{num} is a numerical stabilizer, and w_i is the sample weight. The second term is a subspace-regularization loss:

$$L_{\mathrm{reg}} = \frac{1}{B}\sum_{i=1}^{B}\left[-w_i \sum_{j\in\{\mathrm{knw},\mathrm{unk}\}} u_{ij} \log(r_{ij} + \epsilon_{\mathrm{num}})\right], \tag{13}$$

where u_i is a binary one-hot vector derived from the pseudo-unknown indicator and r_{ij} denotes the probability that sample i belongs to the known or unknown subspace. The third term is a KNN consistency loss. For each sample, we retrieve its Top-K nearest neighbors from a feature memory bank, use the mean prediction of these neighbors as a soft target q_i, and minimize

$$L_{\mathrm{knn}} = \frac{1}{B}\sum_{i=1}^{B}\left[-\sum_{c=1}^{C} q_{ic} \log(p_{ic} + \epsilon_{\mathrm{num}})\right]. \tag{14}$$

The overall loss is

$$L = \lambda_1 L_{\mathrm{psd}} + \lambda_2 L_{\mathrm{reg}} + \lambda_3 L_{\mathrm{knn}}, \tag{15}$$

where λ_1, λ_2, and λ_3 are weighting coefficients.

IV. Experiments

A. Datasets

In this study, we adopt two widely used public benchmark datasets, the CWRU bearing fault dataset and the PU bearing dataset. The CWRU dataset includes signals collected under normal conditions as well as faults on the inner race, outer race, and rolling elements. In our experiments, we select one normal class and nine fault classes, as listed in Table I. We treat different operating conditions as different domains to construct transfer tasks. Specifically, the CWRU dataset is divided into four domains according to motor load and speed: Domain A (0 hp / 1797 rpm), Domain B (1 hp / 1772 rpm), Domain C (2 hp / 1750 rpm), and Domain D (3 hp / 1730 rpm). Based on these domains, we construct eight adaptation tasks covering CDA, PDA, OSDA, and OPDA settings to evaluate performance under the universal adaptation scenario, as detailed in Table II. The Paderborn University (PU) dataset contains two types of bearing damage data: artificially induced defects and naturally occurring real damage. In this study, we select five artificial fault classes and five real fault classes, as summarized in Table III. Similarly, we construct eight domain adaptation tasks on PU. For each task, the source/target label-space setting follows the same eight configurations as those used for CWRU, as listed in Table II.

TABLE I
INFORMATION ABOUT CWRU DATASET.

label	0	1	2	3	4	5	6	7	8	9
types	normal	IR007	B007	OR007	IR021	B021	OR021	IR014	B014	OR014
location	–	IR	Ball	OR	IR	Ball	OR	IR	Ball	OR

TABLE II
CONFIGURATIONS OF UNIDA TASKS FOR CWRU DATASET.

Task	Source	Target	Source label	Target label	Scenario
T1	A	B	0,1,2,3,4,5,6,7,8,9	0,1,2,3,4,5,6,7,8,9	CDA
T2	B	D	0,1,2,3,4,5,6,7,8,9	0,1,2,3,4,5,6,7,8,9	CDA
T3	A	B	0,1,2,3,4,5,6,7,8,9	0,1,2,3,4,5,6	PDA
T4	B	C	0,1,2,3,4,5,6,7,8,9	0,1,2,3,4,5,6	PDA
T5	C	B	0,1,2,3,4,5,6	0,1,2,3,4,5,6,7,8,9	OSDA
T6	C	D	0,1,2,3,4,5,6	0,1,2,3,4,5,6,7,8,9	OSDA
T7	A	B	0,1,2,3,4,5,6	0,1,2,3,7,8,9	OPDA
T8	C	A	0,1,2,3,4,5,6	0,1,2,3,7,8,9	OPDA

TABLE III
INFORMATION ABOUT PU DATASET.

Class label	0	1	2	3	4	5	6	7	8	9
Fault types	KA01	KA03	KA04	KA07	KA16	KI01	KI03	KI04	KI07	KI16
Fault location	OR	OR	OR	OR	OR	IR	IR	IR	IR	IR

B. Experimental Setup

Experiments are conducted on an NVIDIA RTX 5070 GPU with Python 3.12, PyTorch 2.7.1, and CUDA 12.8. We set the batch size to 64, initialize the learning rate to 5×10^{-4}, and optimize the model using stochastic gradient descent (SGD). The loss weights are set to $\lambda_1 = 0.6$ and $\lambda_2 = \lambda_3 = 1$. We split the training and test sets chronologically on the raw time series, and then segment the signals using a sliding window of 1024 points. Each segment is transformed by the continuous wavelet transform (CWT), mapping the 1D vibration signal into a 2D time–frequency energy image. We employ a five-layer feature extractor, where the contrastive disentanglement module is inserted after the third convolutional layer. The fourth and fifth layers are implemented using separable multi-scale convolution (SMC) and a broadcast self-attention (BSA) block [17], with architectural details summarized in Table IV. For CDA and PDA tasks, we report the per-class accuracy, whereas for OSDA and OPDA tasks, both the per-class accuracy and the H-score are reported.

To validate the effectiveness of our method, we compare it with several approaches: (1) DANN [18]; (2) WATN [19]; (3) UAN [20]; (4) THC-DAN [8]; and (5) SFUDA-HCS [13].

TABLE IV
ARCHITECTURE DETAILS OF THE FEATURE EXTRACTOR.

Layers	Output size	Main settings
Input	$3 \times 256 \times 256$	–
Conv1	$32 \times 128 \times 128$	Conv 3×3 (s= 1, p= 1), BN, ReLU, MaxPool 2×2
Conv2	$64 \times 64 \times 64$	Conv 3×3 (s= 1, p= 1), BN, ReLU, MaxPool 2×2
Conv3	$128 \times 32 \times 32$	Conv 3×3 (s= 1, p= 1), BN, ReLU, MaxPool 2×2
C4: SMC	$256 \times 16 \times 16$	SepConv2d $k = 3/5/7$ (s= 1, p= $1/2/3$) $\times3 \rightarrow$ Concat (768ch) $\rightarrow$ Conv 1×1 (768$\rightarrow$256), BN, ReLU, MaxPool 2×2
C5: BSA	$256 \times 8 \times 8$	Proj $k/v/q$: Conv 1×1, q via GAP, heads= 4, out_proj: Conv 1×1, residual, BN, ReLU, MaxPool 2×2

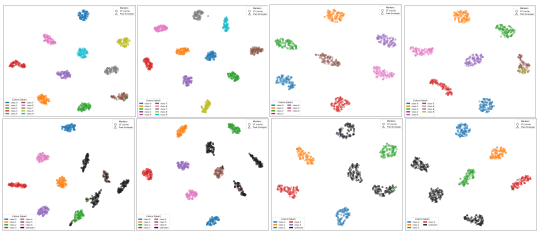

Fig. 3. Visualization of the UniDA scenario on the CWRU dataset

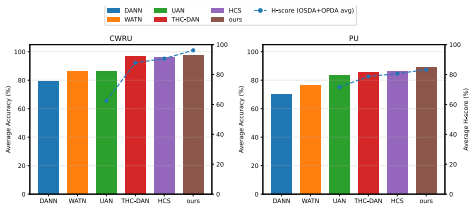

Fig. 4. Comparison of different methods on the CWRU and PU datasets

C. Results of Comparative Experiments

Fig. 3 visualizes the eight universal adaptation tasks on the CWRU dataset. Combined with the results in Table V and Fig. 4, the average performance shows that our method consistently outperforms the best competing approach. On CWRU, it improves the average accuracy by 1.17% and the average H-score by 5.57%. On PU, it increases the average accuracy by 2.26% and the average H-score by 2.64%. These gains mainly come from the OSDA and OPDA scenarios that involve unknown classes, which aligns with our design: we first assess whether unknown classes truly exist in the target domain, and then perform cluster-level exclusion and instance-level filtering when unknowns are present. This strategy effectively reduces the risk that unknown samples contaminate Top-K prototypes, improves pseudo-label quality, and ultimately yields a substantial increase in H-score.

D. Ablation Study

From the ablation results in Table VI, M1 yields the weakest performance. This is mainly because when the target domain contains unknown-class samples, they may be mistakenly absorbed into the Top-K set of known classes, thereby contaminating the corresponding known-class prototypes. After

TABLE V
THE AVERAGE CLASS ACCURACY AND H-SCORE(%) ON THE CWRU DATASET.

Task	DANN	WATN	UAN		THC-DAN		HCS		ours	
	Acc	Acc	Acc	H-Score	Acc	H-Score	Acc	H-Score	Acc	H-Score
T1	95.90	96.60	97.20	—	98.23	—	98.35	—	99.03	—
T2	95.31	95.54	98.70	—	98.90	—	98.56	—	98.90	—
T3	91.21	98.65	91.21	—	98.82	—	99.03	—	99.90	—
T4	91.50	97.54	93.25	—	99.48	—	98.56	—	98.82	—
T5	73.80	70.74	72.11	50.70	95.39	83.63	90.68	90.78	89.30	92.10
T6	71.76	87.13	82.36	70.40	94.42	85.56	93.76	89.54	98.10	94.60
T7	56.75	62.68	76.11	62.81	93.55	91.30	92.67	90.45	99.30	99.60
T8	57.08	80.74	79.60	65.81	93.00	90.95	94.87	91.87	97.80	98.60
Average	79.16	86.20	86.32	62.43	96.47	87.86	95.81	90.66	97.64	96.23

TABLE VI
ABLATION STUDY RESULTS ON THE CWRU DATASET.

Method	Randmix	Contrastive disentanglement module	Unknown detection cluster pse-label	Instance pse-label	Acc	H-score
M1	✗	✗	✗	✓	83.34	80.91
M2	✗	✗	✓	✓	86.37	85.04
M3	✓	✗	✗	✓	90.21	88.51
M4	✓	✗	✓	✓	95.45	94.72
M5	✓	✓	✗	✓	95.23	94.57
M6	✓	✓	✓	✓	97.64	96.23

incorporating unknown-class existence detection and cluster-level unknown pseudo-label construction in M2, the prototype contamination issue is significantly alleviated, leading to improved accuracy and H-score. With RandMix augmentation introduced in M3, both accuracy and H-score are further improved, indicating that the style-perturbed second view is beneficial for feature transfer. Comparing M4 and M5 shows that adding either the contrastive disentanglement module or the unknown-class detection mechanism consistently improves model performance: the former emphasizes cross-view consistency for samples of the same class, while the latter enhances the reliability of pseudo-labels. Finally, M6 further boosts the accuracy by 2.19% and the H-score by 1.51% compared with M4. Overall, these components jointly contribute to the overall performance gain.

V. CONCLUSION

This paper addresses cross-condition fault diagnosis in real industrial scenarios, where the source–target label-space relationship is unknown and industrial data are constrained by privacy requirements. We propose a two-stage target-domain adaptation framework. A dual-view contrastive disentanglement module is introduced at shallow feature layers to learn robust domain-invariant features. During target-domain adaptation, we first detect the existence of unknown classes based on three evidence signals, and then construct high-quality pseudo-labels by combining DBSCAN-based cluster-level exclusion with orthogonal-decomposition-based instance-level filtering. Experimental results show that the proposed method achieves clear improvements in accuracy and H-score under the universal adaptation setting. Future work will focus on more adaptive unknown-class decision criteria and more stable pseudo-label calibration mechanisms to further improve generalization when classes are highly overlapped.

ACKNOWLEDGMENT

This work was supported by the following projects: Shandong Provincial Key Research and Development Program (2024CXGC010905,2024TSGC0603), Shandong Provincial Technology Innovation Guidance Program (YDZX2024121),and the Public Service Platform Project for Solution Promotion and Application(TC200802C).

REFERENCES

[1] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, 2018.

[2] Y. Zhang, T. Zhou, X. Huang, L. Cao, and Q. Zhou, "Fault diagnosis of rotating machinery based on recurrent neural networks," *Measurement*, vol. 171, p. 108774, 2021.

[3] Y. Hou, J. Wang, Z. Chen, J. Ma, and T. Li, "Diagnosisformer: An efficient rolling bearing fault diagnosis method based on improved transformer," *Engineering Applications of Artificial Intelligence*, vol. 124, p. 106507, 2023.

[4] W. Li, R. Huang, J. Li, Y. Liao, Z. Chen, G. He, R. Yan, and K. Gryllias, "A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges," *Mechanical Systems and Signal Processing*, vol. 167, p. 108487, 2022.

[5] D. Li, X. Nie, C. Wu, J. Song, L. Ma, and J. Yang, "Bearing cross-domain fault diagnosis based on domain adversarial network," in *13th International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE 2023)*, vol. 2023. IET, 2023, pp. 572–578.

[6] C. Zhao and W. Shen, "Dual adversarial network for cross-domain open set fault diagnosis," *Reliability Engineering & System Safety*, vol. 221, p. 108358, 2022.

[7] Q. Qian, J. Luo, and Y. Qin, "Adaptive intermediate class-wise distribution alignment: A universal domain adaptation and generalization method for machine fault diagnosis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 3, pp. 4296–4310, 2025.

[8] J. Zhang, X. Wang, Z. Su, P. Lian, H. Xu, J. Zou, and S. Fan, "Two-head classifier guided domain adversarial learning for universal domain adaptation in intelligent fault diagnosis," *Reliability Engineering & System Safety*, vol. 256, p. 110708, 2025.

[9] F. Lin, C. Guo, Z. Zhao, X. Zhang, X. Chen, and Z. Tao, "A universal domain adaptation method with cluster matching for machinery fault diagnosis," *IEEE Internet of Things Journal*, vol. 12, no. 6, pp. 7487–7502, 2025.

[10] M. Xu, Y. Zhang, B. Lu, Z. Liu, and Q. Sun, "A novel domain-private-suppress meta-recognition network based universal domain generalization for machinery fault diagnosis," *Knowledge-Based Systems*, vol. 309, p. 112775, 2025.

[11] Y. Liu, A. Deng, G. Chen, Y. Shi, and Q. Hu, "Universal domain adaptation in rotating machinery fault diagnosis: A self-supervised orthogonal clustering approach," *Reliability Engineering & System Safety*, vol. 257, p. 110828, 2025.

[12] J. Liang, D. Hu, J. Feng, and R. He, "Umad: Universal model adaptation under domain and category shift," *arXiv preprint arXiv:2112.08553*, 2021.

[13] J. Liu, Z. Liu, Z. Jia, and K. Zhao, "Source-free universal domain adaptation for compressor component fault diagnosis guided by hybrid clustering strategy," *Mechanical Systems and Signal Processing*, vol. 232, p. 112771, 2025.

[14] J. Li, K. Yue, Z. Wu, F. Jiang, Z. Zhong, W. Li, and S. Zhang, "Source-free progressive domain adaptation network for universal cross-domain fault diagnosis of industrial equipment," *IEEE Sensors Journal*, vol. 25, no. 5, pp. 8067–8078, 2025.

[15] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[16] S. Qu, T. Zou, L. He, F. Röhrbein, A. Knoll, G. Chen, and C. Jiang, "Lead: Learning decomposition for source-free universal domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 23 334–23 343.

[17] S. Yan, H. Shao, J. Wang, X. Zheng, and B. Liu, "Liconvformer: A lightweight fault diagnosis framework using separable multiscale convolution and broadcast self-attention," *Expert Systems with Applications*, vol. 237, p. 121338, 2024.

[18] J. Jiao, M. Zhao, and J. Lin, "Unsupervised adversarial adaptation network for intelligent fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 11, pp. 9904–9913, 2020.

[19] W. Li, Z. Chen, and G. He, "A novel weighted adversarial transfer network for partial domain fault diagnosis of machinery," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 1753–1762, 2020.

[20] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Universal domain adaptation in fault diagnostics with hybrid weighted deep adversarial learning," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 7957–7967, 2021.