2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Few-shot Metric Adversarial Adaptation for Cross-machine Fault Diagnosis

Qitong Chen, Hong Zhuang, Yueyuan Zhang, Liang Chen* and Qi Li

Abstract— **Research interest in the area of fault diagnosis is shifting from cross-domain to cross-machine, which is crucial for industrial applications with variable operation conditions and different machine configurations. This paper proposes a method named Few-shot Metric Adversarial Adaptation (FMAA) for cross-machine diagnosis of industrial machinery. Firstly, FMAA reduces the data distribution differences between few-shots belonging to the same category in the source domain and the target domain through metric adversarial learning, while increasing the feature distances among different categories. Secondly, the Label Self-Correcting Maximum Mean Discrepancy (LSMMD) method is proposed to correct misclassifications of the model while reducing the conditional distribution differences between the source and target domains. Furthermore, a lightweight attention mechanism-based diagnosis model is proposed to perform cross-machine fault classification tasks. The robustness, universality, and superiority of the proposed method are verified through comprehensive experiments on two platforms for industrial robots and bearings. The code is available on: https://github.com/CCSLab425/FMAA.**

I. INTRODUCTION

In the era of intelligent manufacturing, real-time condition monitoring and intelligent maintenance of industrial devices are crucial to improving production efficiency and product quality. While the research interests for real industrial applications of fault diagnosis are emerging, the shifting of diagnosis tasks from cross-domain to cross-machine is still challenging. The industrial production line encompasses various types of machines with variable operation conditions and different configurations, and there are significant differences in data distribution among different machines. This makes it difficult to transfer diagnostic models developed from one machine to other machines. Therefore, it is necessary to research a method to perform cross-machine diagnostic tasks.

Currently, pioneering scholars have developed intelligent methods for cross-machine diagnosis. For example, Zhu et al. proposed a multi-adversarial learning method to extract domain-invariant features among different machines [1]. Guo et al. [2] proposed a deep convolutional network based on conditional recognition and domain adaptation for performing cross-machine diagnosis of bearings, which achieved approximately 10 % higher accuracy compared to Deep Domain Confusion (DDC [3]) and Deep Adversarial Neural Network (DANN [4]). Qian et al. proposed an enhanced joint distribution adaptation technique to perform cross-machine diagnosis [5]. This method achieved a diagnostic accuracy of 90.52 % in the transfer tasks across three bearing datasets.

The above pioneering work is remarkable except for complex and cost-sensitive industrial scenarios where the target domain (TD) data is generally unknown or the amount of labeled data collected is limited. Therefore, some researchers have proposed few-shot approaches for cross-machine diagnosis tasks in such scenarios, which only require a limited number of labeled samples from the TD to achieve good diagnostic results, aligning more with practical industrial needs. Zhang et al. developed a few-shot algorithm based on asymmetric distribution measurement for cross-machine diagnosis of bearings [6]. This method utilizes meta-models to learn domain-invariant knowledge and transfers the knowledge to another machine to complete the diagnostic task. Yue et al. introduced a multi-scale wavelet convolution and meta-learner strategy for cross-machine classification tasks, achieving satisfactory results in three bearing experiments [7]. Additionally, Wu et al. proposed few-shot methods that accomplished cross-machine diagnosis between bearings and gearboxes [8].

This paper can be regarded as an inheritance and development of the above research. The motivation is to establish a method based on few-shot learning and adversarial adaptation to promote the robustness and universality of cross-machine diagnosis. The main idea is to map the few-shot features in the TD to the vicinity of the feature space of the same category in the source domain (SD) while increasing the feature distance among different categories. In this way, cross-machine diagnosis can be used in more complex scenarios and on a wider range of industrial devices [9]. In this paper, the proposed method is called FMAA for cross-machine diagnosis. FMAA can achieve outstanding performance with just one labeled sample from the TD. In order to tackle the challenge of model misclassification caused by the scarcity of labeled samples in the TD, we propose the LSMMD strategy to enhance the diagnostic performance of FMAA. LSMMD utilizes CMMD (Class-specific MMD [10]) to reduce the conditional distribution difference between the SD and the TD, while also correcting model misclassifications by reducing the feature distance between the small samples from the TD and the overall dataset. Furthermore, a lightweight attention module is proposed to be integrated into the fault classification model, which enhances diagnostic performance by focusing on high-frequency and low-frequency information in high-

This work is financially supported by the National Natural Science Foundation of China under Grant 52375114.

Qitong Chen, Hong Zhuang, Yueyuan Zhang and Liang Chen are with the School of Mechanical and Electrical Engineering, Soochow University, Suzhou 215000, China (e-mail: qtchen0730@163.com; hzhuang@suda.edu.cn; zhangyueyuan@suda.edu.cn; ChenL@suda.edu.cn)

Qi Li is with the Department of Mechanical Engineering, Tsinghua University, Beijing 100000, China (e-mail: liq22@tsinghua.org.cn)

dimensional features. The potential contributions made in this paper are as follows:

1) FMAA is proposed for conducting cross-machine diagnosis tasks on rotating machinery. FMAA employs a metric adversarial approach to aggregate samples belonging to the same category from the SD and TD, while simultaneously increasing the feature distance among different categories.
2) LSMMD is proposed to correct misclassifications of the model while reducing the conditional distribution differences between the SD and TD.
3) A lightweight attention module is proposed, which enhances the classification performance of the model by extracting high-frequency and low-frequency information from the diagnostic signals.
4) We conducted cross-machine experiments using ball screw and bearing datasets, and the proposed model performed remarkably well in the diagnostic task.

II. PRELIMINARIES

A. Few-shot Adversarial Domain Adaptation

Few-shot Adversarial Domain Adaptation (FADA) maps the features of few-shots belonging to the same category from the TD to the SD through adversarial learning, which reduces the data distribution differences between the two domains [11]. Specifically, FADA randomly samples few-shots from the SD and TD and divides the few-shots into four groups, denoted as $\mathcal{G}_{i=1}^{4}$. Each group contains two samples, X_1 and X_2. FADA inputs samples from four groups into the model for adversarial training. The feature extractor and discriminator are optimized using (1) and (3), respectively.

$$\begin{aligned} L_{\mathrm{FADA}} = &-\alpha \mathrm{E}[\mathrm{Y}_{\mathcal{G}_1}\log(\mathbf{D}(\mathbf{F}(\mathcal{G}_2))) + \mathrm{Y}_{\mathcal{G}_3}\log(\mathbf{D}(\mathbf{F}(\mathcal{G}_4)))] \\ &+ \mathrm{E}[l_{ce}((\mathbf{F}_{\mathrm{c}}(\mathrm{X}_s^{shot}), \mathrm{Y}_s^{shot})] + \mathrm{E}[l_{ce}(\mathbf{F}_{\mathrm{c}}(\mathrm{X}_t^{shot}), \mathrm{Y}_t^{shot})] \end{aligned} \tag{1}$$

$$l_{ce} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \tag{2}$$

$$L_{\mathrm{D}} = -\mathrm{E}\left[\sum_{i=1}^{4} \mathrm{Y}_{\mathcal{G}_i}\log\left(\mathbf{D}\left(\mathcal{C}(\mathcal{G}_i)\right)\right)\right] \tag{3}$$

where α is a weight parameter, E is the expectation, and $\mathrm{Y}_{\mathcal{G}_i}$ is the label of $\mathcal{G}_i$. $\mathbf{F}$ and $\mathbf{D}$ are denoted as feature extractor and discriminator, respectively. l_{ce} is the cross-entropy loss [4], y_i and $\hat{y}_i$ are the true and predicted labels, respectively. $\mathbf{F}_{\mathrm{c}}$ is the combination of $\mathbf{F}$ and classifier. X_s^{shot} and X_t^{shot} denote few-shots from the SD and TD, respectively. Y_s^{shot} and Y_t^{shot} denote the labels of X_s^{shot} and X_t^{shot}, respectively. $\mathcal{C}$ is the channel concatenation.

B. Class-specific MMD

CMMD can use pseudo-labels $\hat{y}_t$ to reduce the conditional distribution differences between SD and TD [10]. This can decrease the feature distance among samples of the same category in both domains, thereby improving the diagnostic performance of the model. CMMD is computed as follows :

$$L_{\mathrm{CMMD}} = \sum_{c=1}^{C} \left\| \frac{1}{M_c}\sum_{y_i^s=c}\Phi\left(x_i^s\right) - \frac{1}{N_c}\sum_{\hat{y}_j^t=c}\Phi\left(x_j^t\right) \right\|_H \tag{4}$$

where C is the quantity of the health status of rotating machinery, and Φ is the mapping function that maps fault features to the Hilbert space. M_c and N_c represent the number of categories c in the SD and TD, respectively.

III. PROPOSED METHOD

A. Grouping Settings for Few-shot

Step 1: A small number of samples are randomly selected from the SD and the TD datasets, and then combined to form the sets $\mathcal{D}_s^{shot} = \{x_i^s\}_{i=1}^{C}$ and $\mathcal{D}_t^{shot} = \{x_i^t\}_{i=1}^{C}$. Here, C represents the number of healthy states of the ball screw or bearing.

Step 2: Samples are respectively taken from $\mathcal{D}_s^{shot}$ and $\mathcal{D}_t^{shot}$ and combined to form positive pairs P_p and negative pairs P_n. In P_p, the labels of the two samples are the same, while in P_n, the labels of the two samples are different. The representation of P_p^{s-t} indicates that one sample in the positive pair comes from the SD (s) and the other sample comes from the TD (t). Specifically, P_p^{s-s}, P_p^{s-t}, P_n^{s-s}, and P_n^{s-t} can be represented as follows:

$$\begin{aligned} P_p^{s-s} &= \{\mathcal{G}_g\}_{g=1}^{C} = \{(x_i^s, x_i^s)\}_{i=1}^{C}, \\ P_p^{s-t} &= \{\mathcal{G}_g\}_{g=C+1}^{2C} = \{(x_i^s, x_i^t)\}_{i=1}^{C}, \\ P_n^{s-s} &= \{\mathcal{G}_g\}_{g=2C+1}^{3C} = \{(x_i^s, x_{i+1}^s)\}_{i=1}^{C}, \\ P_n^{s-t} &= \{\mathcal{G}_g\}_{g=3C+1}^{4C} = \{(x_i^s, x_{i+1}^t)\}_{i=1}^{C}, \end{aligned}$$

where $1 < i < C$, and $x_i = x_C$ when $i+1 = C+1$.

Step 3: According to metric learning theory, the few-shot adversarial process is divided into two processes: positive adversarial process and negative adversarial process. The positive adversarial process reduces the feature distance between the same categories through adversarial interactions within the positive group, while the negative adversarial process increases the feature distance among different categories through adversarial interactions within the negative group, as shown in Fig.1.

Step 4: Positive adversarial groups are created. P_p^{s-s} and P_p^{s-t} contain the same fault categories, with P_p^{s-t} containing a TD sample. Through positive adversarial interactions, the features of the TD sample are embedded into the feature space of the same category in the SD, continuously reducing the data distribution difference between the SD and TD. Similarly, P_n^{s-s} and P_n^{s-t} reduce distribution differences through positive adversarial interactions.

Step 5: Negative adversarial groups are created. Groups $\mathcal{G}_q$ are selected from P_p^{s-s} and P_n^{s-s}, and groups $\mathcal{G}_v$ are selected from P_p^{s-t} and P_n^{s-t}. $\mathcal{G}_q$ and $\mathcal{G}_v$ contain completely different fault categories, forming negative adversarial pairs. Through negative adversarial interactions, $\mathcal{G}_q$ and $\mathcal{G}_v$ gradually increase the feature distance among different categories to avoid confusion among different categories.

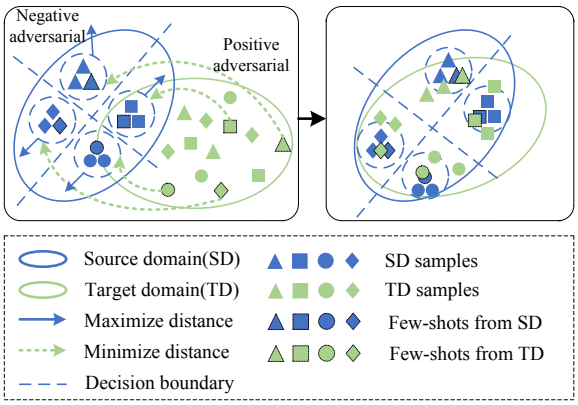


Fig. 1. The principle of few-shot adversarial domain adaptation based on metric learning.

B. Structural Design of the Model

The cross-machine diagnostic model mainly consists of the feature extractor **F**, discriminator **D**, and attention module, as shown in Fig. 2. The feature extractor consists of standard convolutions (Conv1 and Conv2), feature extraction modules, GC (Group convolution), and the attention module. The detailed structure of the model and network parameters are shown in Table I. The design of the model draws inspiration from the residual idea and feature fusion strategy to enhance the feature extraction capability. The number of feature extraction blocks is determined by the input size of the samples. As the sample size increases, the number of feature extraction blocks should be increased. PW (Pointwise convolution), DW (Depthwise convolution), and GC are lightweight convolutions, which reduce the model's parameters and computational complexity [12]. PConv (Partial convolution) operates convolution only on partial features, reducing the model's computational load while preserving some original features [13]. We use PConv instead of standard convolution in the attention module to reduce computational complexity. The attention module enhances the diagnostic performance of the model by extracting global and local features of the signal to obtain low-frequency or high-frequency information [14]. Due to the lightweight design of the network structure and attention module, the number of parameters and computational cost of the model has been effectively reduced. According to Table I, the model has 67,312 parameters (Params), a computational cost of 2.47 MFlops, and requires 0.51 MB of memory. The design of the lightweight model makes training more efficient while meeting the requirements for deployment in enterprise terminals.

C. The Method of FMAA

Firstly, the first sample in $\{\mathcal{G}_g\}_{g=1}^{4C}$ is combined into a vector X_1, and the second sample is combined into X_2. X_1 and X_2 are respectively input into **F** and obtain features f_1 and f_2, as shown in Fig. 2. Next, f_1 and f_2 are fed into the Softmax function to predict the label of X_s^{shot}. The cross-entropy function is utilized to calculate the classification loss L_s^{shot} for X_s^{shot} :

TABLE I

MODEL STRUCTURE AND NETWORK PARAMETERS.

Strcuture	Kernel size	Stride	Padding	Output channels	Output size
Signals	-/-	-/-	-/-	1	32×32
Conv1	3×3	2	1	48	16×16
PW	1×1	1	0	48	16×16
DW	3×3	2	1	48	8×8
PW	1×1	1	0	48	8×8
Avgpool	3×3	2	1	48	8×8
Block 1 channel concatenating				96	8×8
PW	1×1	1	0	96	8×8
DW	3×3	2	1	96	4×4
PW	1×1	1	0	96	4×4
Avgpool	3×3	2	1	96	4×4
Block 2 channel concatenating				192	4×4
GC	3×3	1	1	96	4×4
Attention	-/-	-/-	-/-	96	4×4
DW	3×3	2	1	96	2×2
PW	1	2	0	4	1×1
Params:	67,312	Flops:	2.47M	Memory:	0.51MB

$$L_s^{shot} = l_{ce}\left(\mathbf{F}\left(\mathrm{X}_s^{shot};\theta_f\right),\mathrm{Y}_s^{shot}\right) \tag{5}$$

where θ_f is the parameter of **F**. X_s^{shot} is the few-shot of the SD. Then, f_1 and f_2 are concatenated to obtain $f_{1,2}$. Finally, $f_{1,2}$ is inputted into the **D** to predict the group label. Based on the predicted group label, the loss L_{pa} for positive adversarial and the loss L_{na} for negative adversarial are calculated. L_{pa} can be calculated as follows:

$$\begin{aligned} L_{pa} = \mathrm{E}\Bigg[&\sum_{g=C+1}^{2C} l_{ce}\left(\mathbf{D}\left(\mathcal{C}\left(f_1^{\mathcal{G}_g},f_2^{\mathcal{G}_g}\right)\right),\mathrm{Y}_{\mathcal{G}_{g-C}}\right) \\ &+\sum_{g=3C+1}^{4C} l_{ce}\left(\mathbf{D}\left(\mathcal{C}\left(f_1^{\mathcal{G}_g},f_2^{\mathcal{G}_g}\right)\right),\mathrm{Y}_{\mathcal{G}_{g-C}}\right)\Bigg] \end{aligned} \tag{6}$$

where $f_1^{\mathcal{G}_g}$ is the feature extracted by **F** from sample $\mathrm{X}_1^{\mathcal{G}_g}$. $\mathrm{X}_1^{\mathcal{G}_g}$ represents the first sample of the g-th group. L_{na} can be calculated as follows:

$$L_{na} = -\mathrm{E}\left[\sum l_{ce}\left(\mathbf{D}\left(\mathcal{C}\left(\mathbf{F}(\mathrm{X}_1^{\mathcal{G}_q}),\mathbf{F}(\mathrm{X}_2^{\mathcal{G}_q})\right)\right),\mathrm{Y}_{\mathcal{G}_v}\right)\right] \tag{7}$$

where $\mathrm{X}_1^{\mathcal{G}_q} \in \mathcal{G}_q$ and $\mathrm{X}_2^{\mathcal{G}_q} \in \mathcal{G}_q$. Therefore, the loss for FMAA can be computed using (8):

$$L_{\mathrm{FMAA}} = \lambda_1 L_{pa} + \lambda_2 L_{na} \tag{8}$$

where λ_1 is the adaptive attenuation weight [12] and λ_2 is a constant.

D. The Method of LSMMD

Firstly, the SD dataset $\mathcal{D}_s$ and the TD dataset $\mathcal{D}_t$ are inputted into the **F** to obtain features f_s and f_t. Then, CMMD is used to reduce the feature distance between the same category in f_s and f_t. Although CMMD can reduce the conditional distribution difference between SD and TD, it significantly relies on the credibility of pseudo-labels. There is poor confidence in the pseudo-labels predicted by the model because of the huge distribution discrepancy between the SD and TD caused by the complicated working

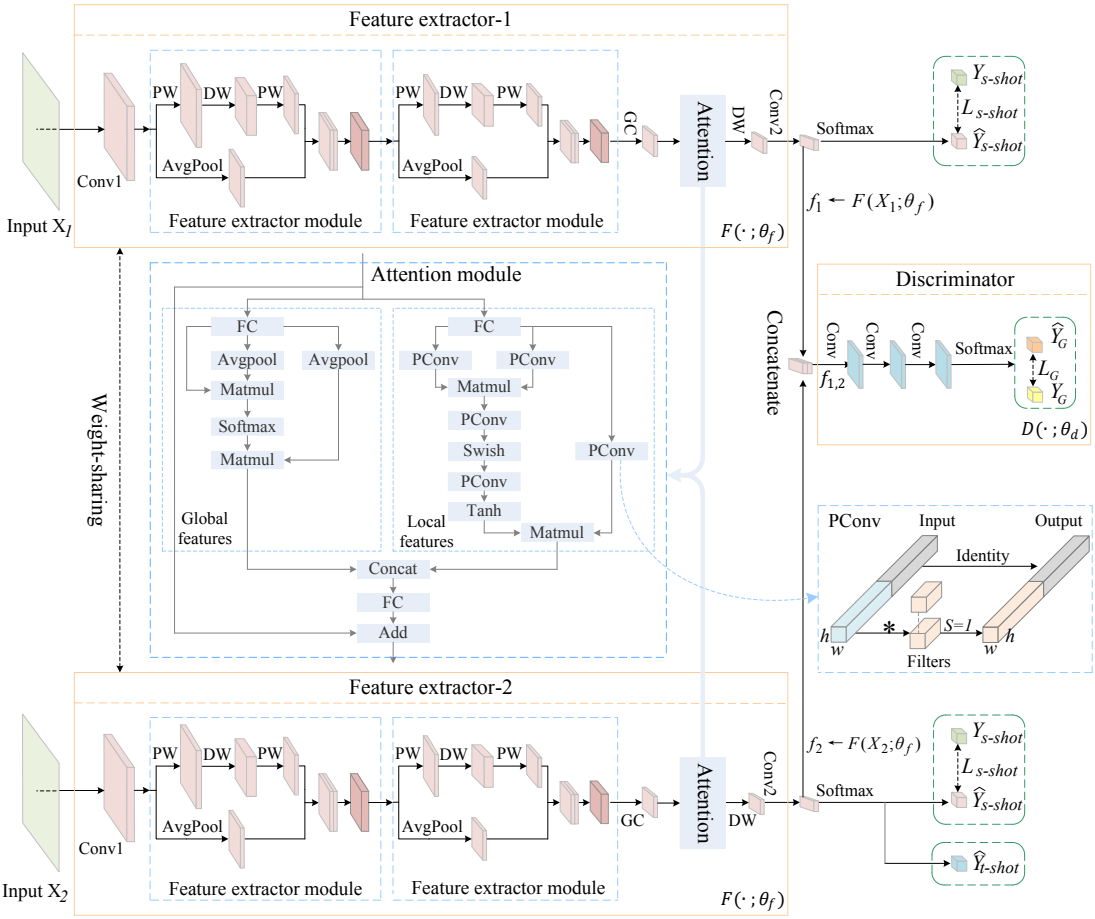


Fig. 2. The overall structure of the cross-machine diagnostic model.

conditions of rotating machines. Therefore, we enhance the credibility of pseudo-labels by reducing the feature distance between the samall samples and complete dataset in the TD, and combine it with CMMD to correct the model's misclassifications. The loss function of LSMMD is calculated as follows:

$$L_{\mathrm{LSMMD}} = \left\| \frac{1}{C}\sum_{c=1}^{C}\Phi(x_c^{t_{shot}}) - \frac{1}{N}\sum_{i=1}^{N}\Phi(x_j^t)) \right\|_{\mathcal{H}} + \lambda_3 \sum_{c=1}^{C} \left\| \frac{1}{M_c}\sum_{y_i^s=c}\Phi(x_i^s) - \frac{1}{\widehat{N}_c}\sum_{\hat{y}_j^t=c}\Phi(x_j^t) \right\|_H \tag{9}$$

where $x_c^{t_{shot}}$ comes from $\mathcal{D}_t^{shot}$, x_j^t comes from $\mathcal{D}_t$, and λ_3 is the adaptive amplification weight [12]. The first norm indicates the reduction of the data distribution difference between the small samples from the TD and the complete dataset of the TD through MMD. The second norm represents the reduction of the conditional distribution difference between each category in the SD and the TD through MMD.

E. The Process of Model Optimization

The optimization process includes pretraining the feature extractor **F**, pretraining the discriminator **D**, and alternating training between **F** and **D**. The specific optimization steps are as follows:

Step1: We use the SD dataset $\mathcal{D}_s$ to pretrain **F** for strong feature extraction capability. During this process, the loss function L_s is used to optimize **F**. L_s can be calculated as follows:

$$L_s = l_{ce}\left(\mathbf{F}\left(x_s;\theta_f\right), y_s\right) \tag{10}$$

Step2: We use X_s^{shot} and X_t^{shot} to pretrain the discriminator **D** for strong discriminative ability. During this process, the loss function L_G is used to optimize **D**. L_G can be calculated as follows:

$$L_G = \mathrm{E}\left[\sum_{g=1}^{4\mathrm{C}} l_{ce}\left(\mathbf{D}\left(f_{1,2}^{\mathcal{G}_g};\theta_d\right), \mathrm{Y}_{\mathcal{G}_g}\right)\right] \tag{11}$$

where θ_d is the parameter of **D**, and $\mathrm{Y}_{\mathcal{G}_g}$ is the true labels of the g-th data pairs.

Step3: After the pretraining phase, **F** and **D** enter the process of alternating adversarial training. In this process, the loss function L_F used to optimize **F** can be calculated as follows:

$$L_F = L_s^{shot} + L_{\mathrm{FMAA}} + L_{\mathrm{LSMMD}} \tag{12}$$

The optimization of **D** can still be accomplished using (11).

The model can extract domain-invariant features from both the SD and TD when **F** and **D** achieve a Nash equilibrium [15] in the adversarial process.

IV. EXPERIMENTS AND ANALYSIS

A. Case1: Experimental Platform of Ball Screw

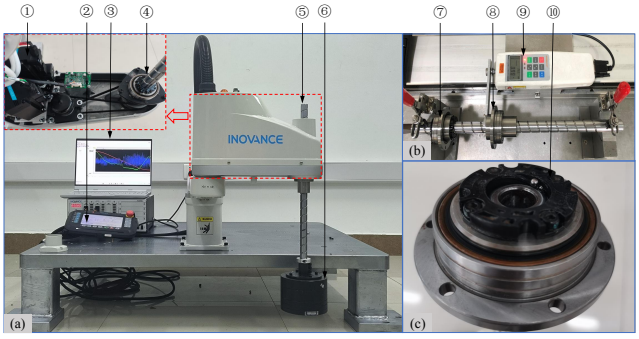
This section of the experiment focuses on the diagnostic analysis of ball screws in industrial robots. The ball screws exhibit four different health states: normal (Norm), ball shedding (Ball), helical nut stuck (Stuck1) and spline nut stuck (Stuck2), as shown in Fig. 3. We collected the current signals of the ball screw's drive motor from two robots. The signal's sampling frequency and sampling length are 8kHz and 1024, respectively.

1) Parameter Settings: Model training experiments were conducted on a workstation equipped with an i7-9700 CPU and RTX2080 GPU. The hyperparameter settings in the experiment are shown in Table II, where i represents the current epoch and w denotes warm steps. In the experiment, warm steps are set to 40. The model's learning rate l_r is set to 0.001, the number of training iterations is 100, and the Batch size is 64. λ_1, λ_2, and λ_3 are the trade-off parameters for the loss. In the fault diagnosis task, the degree of data distribution difference between the SD and the TD has the greatest impact on diagnostic performance. Among these hyperparameters, warm steps exhibit the strongest sensitivity to the diagnostic task. For transfer tasks with significant data distribution differences, a larger warm steps setting is necessary to allow the model to learn more comprehensively. However, to effectively reduce conditional distribution differences, the value of warm steps should not exceed half of the Epochs.

2) The Results of Cross-machine Fault Diagnosis: The diagnostic results for the robot's ball screws are shown in Table III. T0-9 indicates the SD of 0 kg load from the first machine and the TD of 9 kg load from another machine. Note that there is only one label available in the TD. Ten repetitions are made of each transfer task, and the average is taken to mitigate the randomness of the results. The unsupervised methods include ERM (Experience Risk Minimization), Deep Adaptation Network (DAN [16]), Adversarial Domain Adaptation (ADA [12]), DANN [4], and DDC [3]. In particular, the models of unsupervised methods all utilized attention mechanisms, with the learning rate set to 0.001. Unsupervised techniques don't need human annotation, but because cross-machine diagnosis tasks usually include large disparities in data distribution, they frequently fail to produce satisfactory results.

FADA can provide superior diagnostic performance with just one labeled sample in the TD, compared to unsupervised approaches. FMAA adds a metric learning strategy based on FADA, which can increase the distance among different categories. LSMMD introduces a label self-correction strategy based on FMAA, resulting in a 12.1% improvement in diagnostic accuracy. The lightweight attention module can extract both high-frequency and low-frequency information from the current signals, further enhancing the diagnostic performance of the model. Ultimately, the proposed method achieves a diagnostic result of 94.5% in the cross-machine task for ball screws. The computational time required for each method to train for 100 epochs is presented in Table III. The proposed method achieves a training completion time of 90.7 seconds.

3) Convergence and Robustness Analysis of the Model: The training process of the T0-0 transfer task is illustrated in Fig. 4(a), where the primary axis represents accuracy and the secondary axis represents loss. The testing accuracy in the TD closely follows the training accuracy in the SD, indicating that FMAA embeds the features of few-shots from the TD into the feature space of the same category in the SD, thereby reducing the data distribution difference between the two domains. The model's testing loss begins to converge around 20 epochs and gradually approaches zero. Fig. 4(b) demonstrates the interference-resistant process of FMAA, where the secondary axis represents the MMD distance between each category in the SD and TD. For example, S-T-Stuck1-MMD represents the MMD distance between the Stuck1 class in the SD and TD. Adversarial



① J3 drive motor	② Control panel	③ Computer	④ J3 axis	⑤ Screw
⑥ Load	⑦ Helical nut	⑧ Spline nut	⑨ Push-pull gauge	⑩ Ball shedding

Fig. 3. Experimental platform for industrial robot with ball screw. (a) Robot platform. (b) Stuck test. (c) Ball shedding.

TABLE II
HYPERPARAMETER SETTINGS FOR THE MODEL

l_r	Epochs	Batch size	λ_1	λ_2	λ_3	p
0.001	100	64	$\frac{1}{(1+10p)^{0.75}}$	0.2	$\frac{2}{1+exp(-10p)}-1$	$\frac{i-w}{Epochs}$

TABLE III
RESULTS OF CROSS-MACHINE DIAGNOSTICS FOR BALL SCREWS.

Methods	T0-0	T0-3	T0-6	T0-9	T3-3	T6-6	T9-9	Average	Time/s
ERM	25.2%	29.1%	32.2%	32.0%	27.4%	27.7%	40.2%	30.5%	13.9
DAN	25.0%	26.9%	55.5%	43.6%	34.1%	25.9%	78.4%	**41.3%**	41.1
ADA	25.7%	32.3%	33.5%	42.9%	25.8%	39.2%	40.9%	34.3%	26.1
DANN	26.9%	26.6%	48.0%	28.5%	25.5%	36.5%	54.6%	35.2%	43.8
DDC	36.7%	29.9%	26.0%	35.2%	28.9%	42.6%	26.4%	32.2%	33.2
FADA	90.4%	79.0%	57.1%	56.0%	61.8%	56.4%	68.3%	67.0%	35.3
FMAA	94.3%	85.1%	69.4%	63.1%	75.4%	65.6%	79.0%	76.0%	35.4
LSMMD	99.6%	96.6%	93.8%	73.7%	81.0%	86.4%	85.8%	88.1%	68.7
Attention	97.9%	93.9%	93.8%	94.8%	97.8%	90.3%	92.9%	**94.5%**	90.7

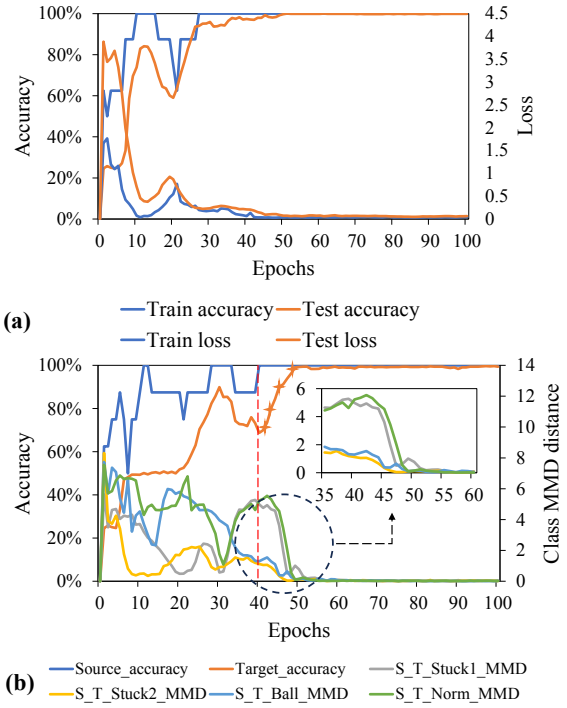


Fig. 4. Training process of Task T0-0. (a) Convergence process of the model. (b) Anti-interference process of the model.

training may interfere with the model's convergence and even lead to training collapse [12]. The model's accuracy starts to decline gradually due to adversarial interference at the 30th epoch, but after incorporating the LSMMD and Attention strategies at the 40th epoch, the model's accuracy begins to improve gradually. Therefore, LSMMD and Attention strategies effectively enhance the model's robustness. Additionally, after adding the LSMMD and Attention strategies, the MMD distance between the same categories in the SD and TD starts to decrease rapidly, significantly reducing the conditional distribution difference between the SD and TD.

4) Feature Visualization: The diagnostic model's efficacy is illustrated using the t-SNE [12] approach. The results of feature visualization are shown in Fig. 5. S-Stuck-1 and T-Stuck-1 denote the stuck faults in the SD and TD, respectively. ADA struggles to effectively differentiate fault categories due to significant data distribution differences between the SD and TD. Although FADA enlarges the feature distance among different categories by leveraging a labeled small sample, it still results in confusion between some categories (Stuck and Norm). FMAA alleviates the confusion of model classification by increasing the distribution distance among different categories through negative adversarial strategies. While CMMD reduces the feature distance between the same categories in the source and target domains, it still cannot differentiate between the Norm and Stuck categories. Through a self-correction method, LSMMD reduces the feature distance between the same category in both domains and fixes the model's misclassifications. The attention module enhances the diagnostic performance of the model while also perfectly accomplishing the clustering task.

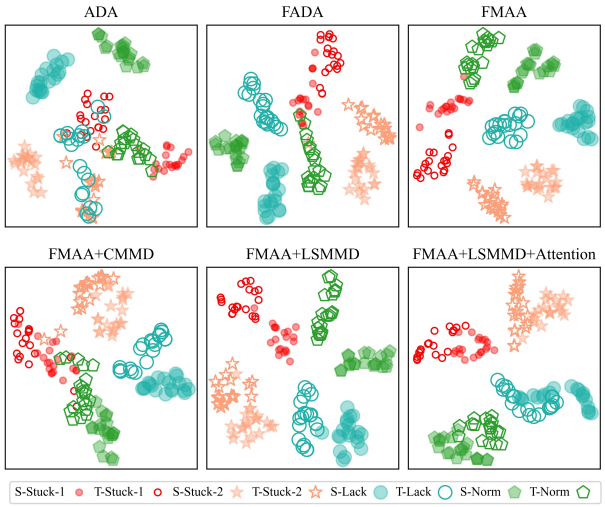


Fig. 5. Feature visualization results of different methods.

B. Case2: Experimental Platform of Bearing

The experiments for cross-machine diagnosis of bearings were conducted using the CWRU (Case Western Reserve University) dataset [5] and the SCU (Soochow University) dataset [12], as shown in Fig. 6. Both datasets consist of four different health conditions: normal (Norm), inner race fault (Inner), ball fault (Ball), and outer race fault (Outer). The sampling frequency of the vibration signals in the CWRU and SCU datasets is 12 kHz and 10 kHz. The CWRU dataset includes four operating conditions: 0hp, 1hp, 2hp, and 3hp. The SCU dataset includes four operating conditions: 0kN, 1kN, 2kN, and 3kN.

1) The Results of Cross-machine Fault Diagnosis: The hyperparameter settings were the same as in Case 1, and the diagnostic results are shown in Table IV. Among them, the T0-1 task represents training the model using the 0hp data from the CWRU dataset and testing the model using the 1kN data from the SCU dataset. The diagnostic results of ERM indicate that there is a huge distribution difference between the CWRU and SCU datasets. DDC achieved the best diagnostic results in the unsupervised method, but it does not meet the requirements of industrial applications. Compared to FADA, the proposed FMAA method shows improved diagnostic performance after incorporating metric learning strategies. LSMMD can improve the accuracy by

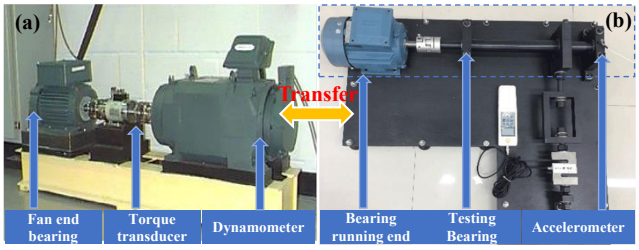


Fig. 6. Bearing experimental platform for CWRU and SCU. (a) CWRU dataset. (b) SCU dataset.

TABLE IV

RESULTS OF CROSS-MACHINE DIAGNOSTICS FOR BEARINGS.

Methods	T0-0	T0-1	T0-2	T0-3	T1-1	T2-2	T3-3	Average
ERM	43.7%	32.8%	34.5%	34.5%	25.0%	25.0%	25.0%	31.5%
DAN	25.0%	25.0%	55.7%	25.0%	25.0%	25.0%	25.0%	29.4%
ADA	25.0%	25.0%	25.0%	25.0%	25.0%	25.0%	25.0%	25.0%
DANN	25.0%	25.0%	27.5%	25.0%	25.0%	25.0%	47.5%	28.6%
DDC	25.6%	38.2%	43.7%	27.7%	44.8%	34.9%	33.6%	**35.5%**
FADA	81.0%	96.0%	92.6%	88.8%	94.7%	71.6%	79.5%	86.3%
FMAA	88.2%	91.7%	89.0%	93.8%	85.9%	85.2%	95.0%	89.8%
LSMMD	99.0%	94.3%	93.5%	97.5%	91.9%	97.4%	98.5%	96.0%
Attention	98.5%	97.4%	98.7%	98.2%	99.9%	97.3%	99.8%	**98.6%**

6.2% through its label self-correction strategy. Furthermore, the diagnostic performance can be further enhanced by incorporating the attention module into the model.

2) The Analysis of the Model's Classification Performance: The confusion matrix of the few-shot methods for the T3-3 task is shown in Fig. 7. FADA only correctly identifies the Outer category. There is confusion between some Norm and Inner samples due to the small feature distance. Some samples from Norm and Ball are predicted as Outer, indicating that it is necessary to increase the feature distance among the three categories to avoid confusion. FMAA eliminates the confusion between Norm and Inner after incorporating metric learning strategies, and the classification performance of Ball improves by 34.5%. The LSMMD strategy further enhances the classification performance of Ball. Finally, the attention module improves the classification performance of Ball to 99.5% by extracting global and local features of the model.

V. CONCLUSIONS

The objective of this study is to develop a robust and versatile method for cross-machine diagnosis of rotating machinery. Firstly, FMAA uses metric learning method to increase the feature distance among different categories, while employing adversarial learning methods to reduce the data distribution difference between the SD and TD. Then, LSMMD further enhances the diagnostic performance of the model by correcting its misclassification. Additionally, we incorporate a lightweight attention module into the model to extract local and global features of signals, thereby enhancing the diagnostic performance of the FMAA. Finally, the effectiveness of the cross-machine diagnostic method is validated using ball screw and bearing datasets.

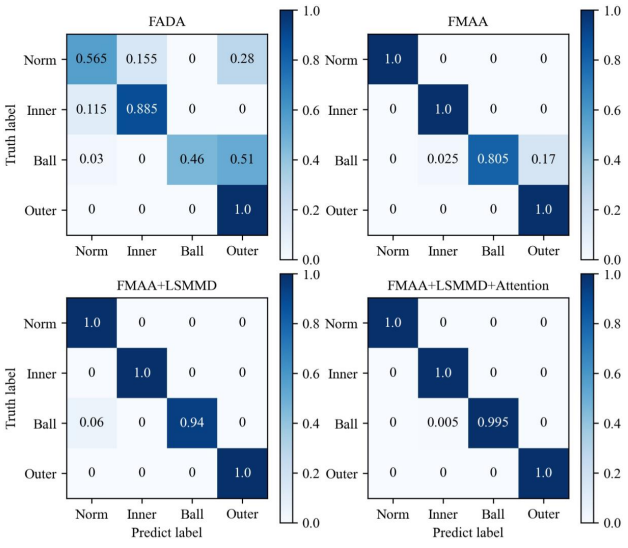


Fig. 7. The confusion matrix of few-shot methods in the T3-3 task.

REFERENCES

[1] J. Zhu, N. Chen, and C. Shen, "A new multiple source domain adaptation fault diagnosis method between different rotating machines," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 4788–4797, 2020.

[2] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 9, pp. 7316–7325, 2018.

[3] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," *arXiv preprint arXiv:1412.3474*, 2014.

[4] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *International conference on machine learning*. PMLR, 2015, pp. 1180–1189.

[5] Q. Qian, Y. Qin, J. Luo, Y. Wang, and F. Wu, "Deep discriminative transfer learning network for cross-machine fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 186, p. 109884, 2023.

[6] J. Zhang and L. Zhang, "Cross-machine few-shot fault diagnosis for rotating machinery with asymmetric distribution measure network," in *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2023, pp. 3404–3409.

[7] K. Yue, J. Li, J. Chen, R. Huang, and W. Li, "Multiscale wavelet prototypical network for cross-component few-shot intelligent fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–11, 2022.

[8] J. Wu, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Few-shot transfer learning for intelligent fault diagnosis of machine," *Measurement*, vol. 166, p. 108202, 2020.

[9] M. Azamfar, X. Li, and J. Lee, "Intelligent ball screw fault diagnosis using a deep domain adaptation methodology," *Mechanism and Machine Theory*, vol. 151, p. 103932, 2020.

[10] H. Yan, Z. Li, Q. Wang, P. Li, Y. Xu, and W. Zuo, "Weighted and class-specific maximum mean discrepancy for unsupervised domain adaptation," *IEEE Transactions on Multimedia*, vol. 22, no. 9, pp. 2420–2433, 2019.

[11] S. Motiian, Q. Jones, S. Iranmanesh, and G. Doretto, "Few-shot adversarial domain adaptation," *Advances in neural information processing systems*, vol. 30, 2017.

[12] Q. Chen, L. Chen, Q. Li, J. Shi, Z. Zhu, and C. Shen, "A lightweight and robust model for engineering cross-domain fault diagnosis via feature fusion-based unsupervised adversarial learning," *Measurement*, vol. 205, p. 112139, 2022.

[13] J. Chen, S.-h. Kao, H. He, W. Zhuo, S. Wen, C.-H. Lee, and S.-H. G. Chan, "Run, don't walk: Chasing higher flops for faster neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 12 021–12 031.

[14] Q. Fan, H. Huang, J. Guan, and R. He, "Rethinking local perception in lightweight vision transformer," *arXiv preprint arXiv:2303.17803*, 2023.

[15] D. M. Kreps, "Nash equilibrium," in *Game theory*. Springer, 1989, pp. 167–177.

[16] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," *arXiv: Learning,arXiv: Learning*, Feb 2015.