CROSS-DOMAIN FAULT DIAGNOSIS FOR BEARINGS VIA MULTI-BRANCH CONVOLUTIONAL NEURAL NETWORK AND MIXED DISCREPANCY MEASUREMENT

Meng Luo[1], Quanfeng Li[2*], Wei Cui[3]

[1]*Shanghai DianJi University, Shanghai, China*
[2]*Shanghai DianJi University, Shanghai, China*
[3]*Shanghai Yingbeide Intelligent Technology Co., Ltd, Shanghai, China*
**liqf@sdju.edu.cn*

Keywords: FAULT DIAGNOSIS, ROLLING BEARING, MULTI-BRANCH CNN, DISCREPANCY MEASUREMENT, DOMAIN ADAPTATION

Abstract

To address the challenges of cross-domain fault diagnosis for industrial rotating machinery under strong noise interference and time-varying operating conditions, this paper proposes a novel method based on Multi-Branch Convolutional Neural Network (MBCNN) and a mixed discrepancy measurement strategy. The proposed MBCNN model is designed as a feature extractor, integrating parallel branches for moving average smoothing, Gaussian filtering, and original vibration signals, which effectively enhances time-frequency feature representation while suppressing noise interference. Furthermore, to improve feature distribution alignment between source and target domains, a Mean-Variance Mixed Discrepancy (MVMD) based domain adaptation module is introduced. This module jointly leverages Maximum Mean Square Discrepancy (MMSD) and Variance Discrepancy Representation (VDR) to perform multi-granularity statistical alignment. Extensive experiments on the CWRU bearing dataset, including noise robustness and cross-condition transfer tasks, Showcase the robust noise immunity and multi-domain diagnostic prowess of the proposed approach, offering a practical and dependable approach for intelligent fault diagnosis in complex industrial environments.

1 Introduction

Rolling bearings are essential in rotating equipment and are extensively employed across different industries, including power generation systems, new energy vehicles, and aerospace. The safety, stability, and reliability of the entire system largely depend on their operating conditions [1]. The swift rise of artificial intelligence in recent times has led to the widespread adoption of deep learning techniques for diagnosing machinery faults, thanks to its prowess in extracting key features and identifying issues effectively [2]. However, in practical industrial environments, vibration signals are often affected by random noise, variable loads, and time-varying operating conditions, resulting in significant distribution discrepancies between training and testing datasets. This distribution shift greatly hinders the effectiveness of traditional deep learning techniques in performing cross-domain fault diagnosis.

Transfer learning overcomes the scarcity of labeled data in the target domain by leveraging knowledge from a related source domain, and has become a popular research direction in intelligent fault diagnosis[3]. Among these approaches, domain adaptation (DA) enhances model accuracy in the target domain by minimizing feature distribution gaps between source and target data. Current DA techniques primarily fall into two categories: statistical discrepancy-based distribution alignment and adversarial-based domain-invariant feature learning. For instance, Rathore et al.[4] introduced a multiple kernel MMD to perform multi-scale domain mean embedding alignment; Song et al.[5] partitioned sub-domains according to fault types and employed Local Maximum Mean Discrepancy (LMMD) to reduce sub-domain distribution differences; Qin et al.[6] To optimize the feature extractor and improve domain confusion, an imbalanced adversarial training strategy is proposed, along with the incorporation of a covariance alignment loss (CORAL). Although these methods have achieved promising results in cross-domain fault diagnosis of rotating machinery, their robustness and feature representation capabilities remain limited under complex noise interference and significant condition variations. Moreover, most approaches rely on single-statistic-based distribution alignment, which fails to comprehensively model the complex distribution shift characteristics of vibration signals.

To address these challenges, a new cross-domain diagnosis method for rolling bearings is developed by integrating a Multi-Branch Convolutional Neural Network (MBCNN) with a Mean-Variance Mixed Discrepancy (MVMD) metric. The MBCNN employs a multi-branch denoising structure that integrates moving average filtering, Gaussian filtering, and raw vibration signals to improve noise-robust feature extraction and time-frequency representation. Meanwhile, MVMD jointly combines Maximum Mean Square Discrepancy (MMSD) and Variance Discrepancy Representation (VDR) to enable multi-level statistical alignment between domains. Extensive experiments on the CWRU bearing dataset, involving variable noise and cross-condition transfer tasks, confirm that the designed approach delivers superior diagnosis accuracy, noise resistance, and cross-domain adaptability in complex industrial scenarios.

2. Theoretical Background

2.1 Problem description of domain adaptation

As a crucial branch of transfer learning, DA focuses on mitigating distribution discrepancies between domains and improving model performance in the target domain by transferring knowledge from the source domain. In practical applications, the annotations for the source domain dataset $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$ are available, where n_s indicates the quantity of labeled samples in the source domain, and each sample x_i^s has a corresponding fault label y_i^s. Similarly, the unlabeled target domain dataset is denoted as $D_t = \{(x_j^t)\}_{j=1}^{n_t}$, where n_t represents target domain sample count, and x_j^t denotes the j-th instance in the target domain. Although the feature space is identical for both domains, their probability distributions differ, due to distribution shifts, the generalization ability of source-trained models significantly degrades in the target domain, which severely restricts the performance of cross-domain fault diagnosis.

2.2 Maximum mean square discrepancy

Maximum Mean Discrepancy (MMD) is a statistical measure for quantifying the difference between two probability distributions. When two datasets are nonlinearly separable in the original space, they can be projected into a Reproducing Kernel Hilbert Space (RKHS) using a mapping function. In this new space, MMD measures the distributional difference by computing the squared distance between the mean embeddings of the two datasets.

To further enhance the ability to characterize distribution differences, Qian et al.[7] proposed the Maximum Mean Square Discrepancy (MMSD) method based on MMD. This approach employs a tensor-product kernel to map data into a higher-dimensional feature space and measures the discrepancy between the second-order moment embeddings of two data domains with distinct distributions. Specifically, MMSD is defined as the squared norm of the difference between the mean square embeddings of the two domains in the tensor-product space:

$$\begin{aligned}&\mathrm{MMSD}^2[H \otimes H, D_s, D_t] = \\ &\left\| E_p[k(x_s,\cdot) \otimes k(x_s,\cdot)] - E_q[k(\cdot,x_t) \otimes k(\cdot,x_t)] \right\|_{H\otimes H}^2\end{aligned} \quad (1)$$

Where $K(\cdot)$ is the feature mapping function in the RKHS, typically implemented as a Gaussian kernel function. In practice, for finite-sample scenarios, the expectations are typically replaced by empirical averages.

2.3 Variance difference representation

Similar to the MMD principle, Variance Difference Representation (VDR) is based on kernel methods and constructs an explicit measure of variance differences in the feature space, effectively representing distribution discrepancies through the variance information of data points.

Specifically, VDR constructs the following basis function by computing the difference between the kernel function value and its expectation, in order to reflect the variance information of data points in the RKHS:

$$\tau(x,\cdot) = \left(K(x,\cdot) - E\left[K(x,\cdot)\right]\right)^{\otimes 2} \quad (2)$$

Where $K(x,\cdot)$ denotes the kernel function, $E\left[\kappa(x,\cdot)\right]$ represents the expectation over the data distribution, and $\otimes$ is the tensor product operator. Based on this basis function, an new RKHS tensor-product space could describe the variance discrepancies:

$$\mathrm{VDR}^2\left[H_1 \otimes H_2, D_s, D_t\right] = \left\| E_p\tau(X_s,\cdot) - E_q\tau(X_t,\cdot) \right\|_{H_1\otimes H_2}^2 \quad (3)$$

3. Framework for Fault Diagnosis Modeling

This study introduces a cross-domain fault detection framework for rolling bearings based on a MBCNN and a MVMD metric, as illustrated in Figure 1. The model includes a feature extraction component and a classification module. The feature extractor employs a multi-branch denoising network to project labeled source and unlabeled target domain features into a shared subspace, enhancing the extraction of common fault features under different conditions. The classifier comprises two fully connected layers (FC1 and FC2), with the DA module integrated into the FC2 layer to minimize the MVMD distance between features. the model achieves effective distribution alignment and robust fault diagnosis across domains.

3.1 Shared feature extraction network

To effectively extract domain-invariant shared features, designs an MBCNN-based shared feature extraction framework. A multi-modal input module integrates identity-mapped signals, moving average smoothed signals, and Gaussian-filtered signals as parallel branches to enhance feature diversity. Each branch adopts a convolutional structure with shared parameters, where convolutional layers with different kernel sizes are stacked to extract features at multiple scales, ranging from global to local levels. Multi-scale information is thus effectively captured within each branch. The features from all branches are then fused to improve the overall feature representation capability.

The MBCNN network comprises five convolutional-pooling layers, and ReLU activation functions to introduce nonlinearity. The detailed network parameter configuration is shown in Table 1. The kernel sizes of the convolutional layers decrease progressively from 6 to 3. By designing multi-scale convolution kernels, the network can better adapt to signal variations across different frequencies and time scales, thereby improving its robustness to noise.

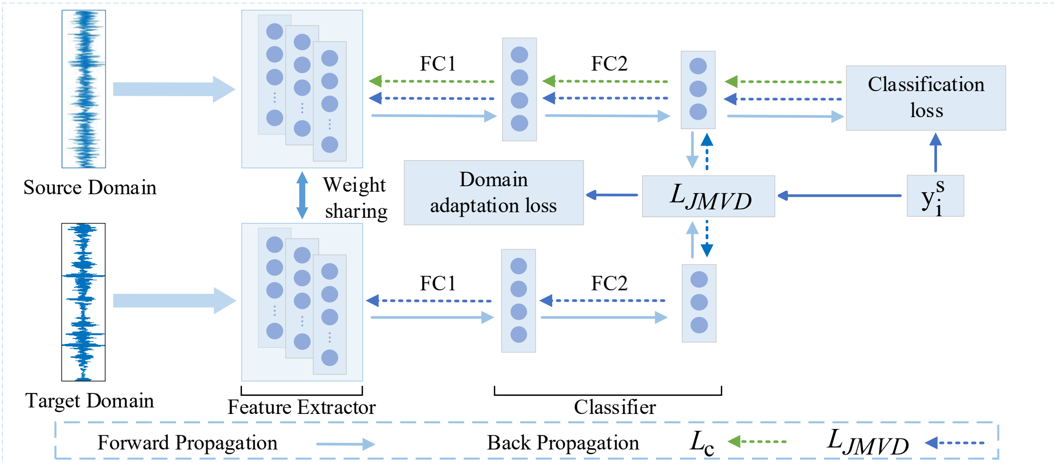


Fig. 1 Fault diagnosis model architecture

Table 1. Network model parameter table

Layer type	Table column subhead
Moving average layer	Window size=5, stride=1
Gaussian filtering layer	Window size=5, stride=1
Conv1d-1	Conv(16×(1×6)); BN; ReLU; MP(2×2)
Conv1d-2	Conv(32×(16×5)); BN; ReLU; MP(2×2)
Conv1d-3	Conv(64×(32×4)); BN; ReLU; MP(2×2)
Conv1d-4	Conv(128×(64×3)); BN; ReLU;MP(2×2)
Conv1d-5	Conv(256×(128×3)); BN; ReLU; MP(2×2)
Flatten	Global Avg Pool, Feature fusion
FC1	FC(768×128);ReLU
FC2	FC(128×class_number)

3.2 Mean-Variance mixed discrepancy

Given the limitations of a single metric in characterizing distribution discrepancies this paper proposes an improved distance measurement method called Mean-Variance Mixed Discrepancy (MVMD). This approach integrates the advantages of MMSD in optimizing mean square discrepancies and VDR in capturing variance differences, thereby improving the ability to capture complex distribution characteristics.

MVMD is jointly formulated based on Equations (1) and (3) as follows:

$$MVMD(D_s,D_t)=\alpha\cdot MMSD^2[H\otimes H,D_s,D_t]+ (1-\alpha)\cdot VDR^2[H_1\otimes H_2,D_s,D_t] \tag{4}$$

Where α is a trade-off parameter.

3.3 Loss function

During training, the model optimizes a composite loss function to enhance feature extraction capability while learning the distribution alignment relationship between domains, leading to enhanced accuracy in diagnosing faults across different domains.

Among them, the classification loss is used to evaluate the classification performance of the classifier on labeled source domain samples, which is computed by the cross-entropy loss function as follows:

$$L_c=\frac{1}{n_s}\sum_{i=1}^{n_s}J\left(\hat{y}_i^s,y_i^s\right) \tag{5}$$

The DA loss L_d is based on the MVMD metric defined in Equation (4), and is designed to achieve distribution alignment by minimizing the loss function.

$$L_d=L_{MVMD(D_s,D_t)} \tag{6}$$

The total loss function comprises two components, with the final objective defined as their weighted sum of classification and DA losses:

$$\min_{\theta\to(w,b)}(L_c+\lambda L_d) \tag{7}$$

Where λ is a trade-off parameter controlling the relative importance of the two losses. By optimizing this loss via backpropagation, the model parameters are iteratively updated to simultaneously improve classification accuracy and cross-domain adaptability.

4. Experimental Verification

4.1 Dataset description

In this study, we utilized the well-known CWRU bearing dataset sourced from Case Western Reserve University for our validation experiments. Illustrated in Figure 2, the experimental platform consists of a motor, bearings, a torque sensor, and a dynamometer. The fault data were gathered at the drive-end bearing with a sampling rate of 12 kHz, covering four operating states: normal (NO), outer race fault (OF), inner race fault (IF), and ball fault (BF). Each fault type includes three defect sizes (0.007 inch, 0.014 inch, and 0.021 inch), resulting in ten labeled fault categories in total.

To evaluate the model's adaptation capability under variable working conditions, four operating loads (0 HP, 1 HP, 2 HP, and 3 HP) are configured to construct transfer tasks. A sliding window technique is applied to segment 200 samples per health condition, with each sample containing 1024 consecutive data points.

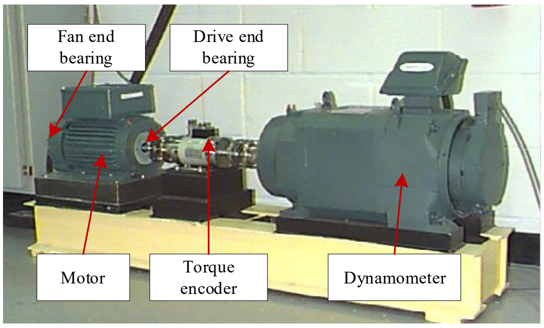


Fig. 2 CWRU bearing fault test rig

4.2 Noise robustness test

In industrial applications, rolling bearings often operate in environments with severe noise interference and complex working conditions, where vibration signals are highly susceptible to noise contamination. To evaluate the noise robustness of the proposed MBCNN feature extraction model, Experimental data were subjected to the addition of Gaussian white noise at varying signal-to-noise ratio levels, resulting in five distinct noise scenarios. Comparative experiments were conducted under the same conditions with existing models, including CNN, WDCNN, and MSCNN, to assess fault diagnosis performance in noisy environments. The experiments were performed using the 2 HP load dataset, and the results are presented in Figure 3.

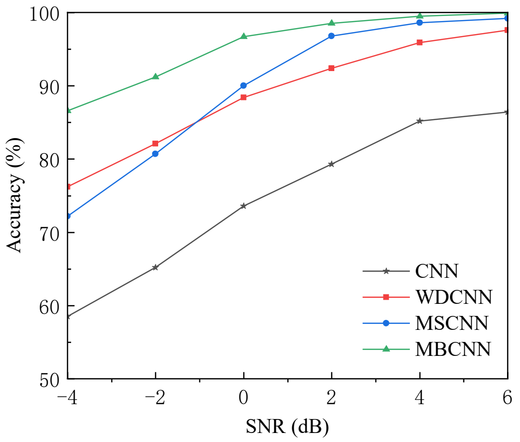


Fig. 3 Accuracy under noise variation

The LeNet-based 1D CNN, limited by its shallow architecture and restricted receptive field, shows weak feature extraction ability, achieving only 86.4% accuracy even at 6dB. WDCNN improves time-domain feature capture with wide convolution kernels but struggles under strong noise due to its fixed-scale design. Although MSCNN enhances feature diversity through a multi-scale structure, the absence of adaptive feature fusion causes critical fault features to be overwhelmed when SNR < -2dB. Conversely, the MBCNN that we introduce demonstrates consistently high accuracy across various noise levels, outperforming the compared models.

4.3 Cross-condition fault diagnosis experiment

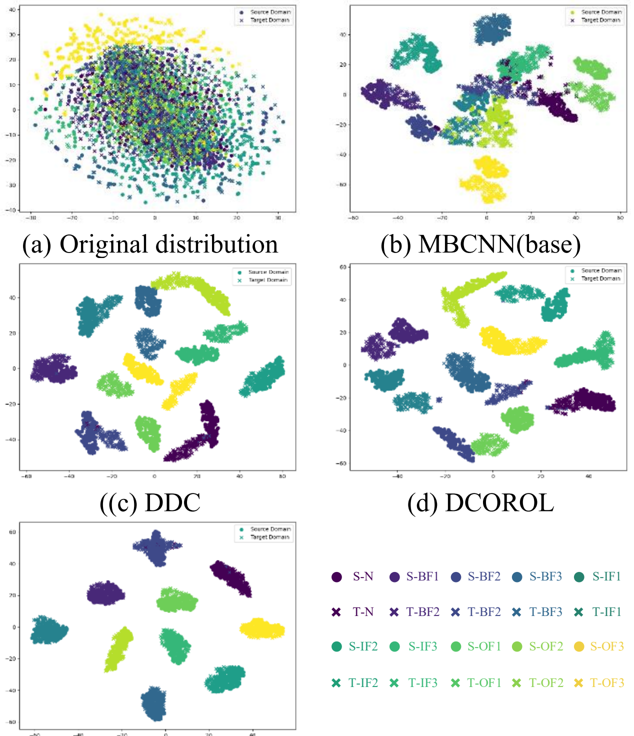
Assessing the performance and advantages of our newly developed DA model for cross-condition fault detection, we performed comparative tests with traditional approaches that focus on aligning distributions. with emphasis on the advantages of the MVMD metric. The models under comparison encompass DDC[8], which utilizes MMD for alignment at the DA layer, the CORAL-based DCORAL [9], and the baseline MBCNN without transfer learning.

All models share the same network architecture and hyperparameter settings. The results are averaged over five separate runs to ensure statistical reliability. Table 2 presents the detailed findings. The proposed MVMD model achieved an average diagnostic accuracy of 98.90% across 12 cross-condition transfer tasks, significantly outperforming other methods. The baseline MBCNN model without transfer learning reached only 80.25%, confirming the necessity of DA in cross-condition fault diagnosis. DDC and DCORAL, which rely on single-statistic alignment (MMD or CORAL), obtained average accuracies of 93.06% and 94.38%, respectively, indicating limitations in fully capturing distribution discrepancies. In contrast, MVMD improves feature distribution alignment by jointly optimizing both mean and variance differences, demonstrating the superiority of multi-statistic alignment.

Table 2. Diagnostic accuracy (%) of cross-condition (0-3 hp)

Transfer task	MBCNN (base)	DDC	DCOROL	MBCNN-MVMD
0→1	82.75	94.67	95.37	98.35
0→2	79.62	90.65	92.63	99.60
0→3	78.43	94.65	87.81	98.25
1→0	81.87	95.21	93.82	99.10
1→2	80.41	97.92	96.15	99.90
1→3	78.28	90.70	93.59	98.85
2→0	77.42	95.50	98.92	99.70
2→1	81.98	97.65	98.20	99.15
2→3	84.80	96.28	96.50	99.57
3→0	78.66	84.60	96.79	98.30
3→1	80.32	88.50	91.35	97.50
3→2	82.65	93.70	94.30	99.48
Average	80.25	93.06	94.38	98.90

To evaluate the feature transferability and model DA performance, T-SNE was employed to visualize the input data and the features extracted from the FC2 layer in the 2HP→1HP scenario. As shown in Figure 4. The baseline MBCNN exhibited severe overlap with poor class separability, while DDC and DCORAL presented limited capability in distinguishing inter-class distributions, resulting in insufficient feature fusion across domains. In comparison, the MVMD method produced highly compact clusters with significant overlap between source and target domain samples for all 10 fault categories, confirming its superior domain alignment and adaptation capability at the feature level.



(a) Original distribution (b) MBCNN(base)

((c) DDC (d) DCOROL

(e)MBCNN-JMVD

Fig. 4 Visualization of different transfer tasks.

5 Conclusion

To address the challenges of cross-domain fault diagnosis caused by strong noise interference and time-varying operating conditions in industrial equipment, the study develops a a cross-condition fault diagnosis framework based on a Multi-Branch CNN (MBCNN) and a Mean-Variance Mixed Discrepancy (MVMD) metric. Systematic experiments on the CWRU bearing dataset under variable working conditions validated the effectiveness of the approach introduced. The pivotal conclusions are outlined as follows:

(1) The multi-branch feature enhancement mechanism significantly improves noise robustness. The designed MBCNN integrates moving average filtering, Gaussian filtering, and raw signals through parallel branches, effectively enhancing feature representation and suppressing noise interference. Noise robustness was confirmed through variable-noise experiments.

(2) The hybrid discrepancy measurement MVMD enables accurate cross-domain alignment. Through the simultaneous adjustment of the mean and variance of the inter-domain distributions, MVMD effectively mitigates feature distribution shifts caused by condition variations. The proposed model achieved an average diagnostic accuracy of 98.90% across 12 cross-condition transfer tasks, outperforming existing methods and demonstrating superior cross-domain adaptability.

5 References

[1] J. Cai and Y. Wang, Application Technology of Motor Bearings. Beijing, China: China Machine Press, 2020.

[2] X. Chen, R. Yang, Y. Xue, M. Huang, R. Ferrero, and Z. Wang, “Deep Transfer Learning for Bearing Fault Diagnosis: A Systematic Review Since 2016,” IEEE Trans. Instrum. Meas., vol. 72, pp. 1–21, 2023.

[3] H. Shao, Y. Xiao, and S. Yan, “Simulation data-driven improved unsupervised domain adaptation for bearing fault diagnosis,” Journal of Mechanical Engineering, vol. 59, no. 3, pp. 76–85, 2023.

[4] M. S. Rathore and S. P. Harsha, “Rolling bearing prognostic analysis for domain adaptation under different operating conditions,” Eng. Fail. Anal., vol. 139, p. 106414, 2022.

[5] X. Song, W. Sun, G. Liu, et al., “Deep subdomain adaptation network for cross-condition fault diagnosis of motor rolling bearings,” Transactions of China Electrotechnical Society, vol. 39, no. 1, pp. 182–193, 2024.

[6] Y. Qin, Q. Yao, Y. Wang, and Y. Mao, “Parameter sharing adversarial domain adaptation networks for fault transfer diagnosis of planetary gearboxes,” Mech. Syst. Signal Process., vol. 160, p. 107936, 2021.

[7] Q. Qian, Y. Wang, T. Zhang, and Y. Qin, “Maximum mean square discrepancy: a new discrepancy representation metric for mechanical fault transfer diagnosis,” Knowl.-Based Syst., vol. 276, p. 110748, 2023.

[8] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, “Deep Domain Confusion: Maximizing for Domain Invariance,” Dec. 10, 2014, arXiv: arXiv:1412.3474. doi: 10.48550/arXiv.1412.3474.

[9] B. Sun and K. Saenko, “Deep CORAL: Correlation Alignment for Deep Domain Adaptation,” in Computer Vision – ECCV 2016 Workshops, vol. 9915, G. Hua and H. Jégou, Eds., in Lecture Notes in Computer Science, vol. 9915, Cham: Springer International Publishing, 2016, pp. 443–450. doi: 10.1007/978-3-319-49409-8_35.