

Full length article

Interpretable modulated differentiable STFT and physics-informed balanced spectrum metric for freight train wheelset bearing cross-machine transfer fault diagnosis under speed fluctuations

Chao He, Hongmei Shi^{*}, Ruixin Li, Jianbo Li, ZuJun Yu

State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing 100044, China

School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University, Beijing 100044, China

ARTICLE INFO

Keywords:

Modulation differentiable STFT
Balanced spectrum quality
Cross-machine diagnosis
Speed fluctuations
Heavy haul freight trains

ABSTRACT

The service conditions of wheelset bearings has a direct impact on the safe operation of railway heavy haul freight trains as the key components. However, speed fluctuation of the trains and few fault samples are the two main problems that restrict the accuracy of bearing fault diagnosis. Therefore, a cross-machine transfer diagnosis (pyDSN) network coupled with interpretable modulated differentiable short-time Fourier transform (STFT) and physics-informed balanced spectrum quality metric is proposed to learn domain-invariant and discriminative features under time-varying speeds. Firstly, due to insufficiency in extracting extract frequency components of time-varying speed signals using fixed windows, a modulated differentiable STFT (MDSTFT) that is interpretable with STFT-informed theoretical support, is proposed to extract the robust time-frequency spectrum (TFS). During training process, multiple windows with different lengths dynamically change. Also, in addition to the classification metric and domain discrepancy metric, we creatively introduce a third kind of metric, referred to as the physics-informed metric, to enhance transferable TFS. A physics-informed balanced spectrum quality (BSQ) regularization loss is devised to guide an optimization direction for MDSTFT and model. With it, not only can model acquire high-quality TFS, but also a physics-restricted domain adaptation network can be also acquired, making it learn real-world physics knowledge, ultimately diminish the domain discrepancy across different datasets. The experiment is conducted in the scenario of migrating from the laboratory datasets to the freight train dataset, indicating that the hybrid-driven pyDSN outperforms existing methods and has practical value. Partial code is available at <https://github.com/ligue/PyDSN>.

1. Introduction

Parallel to the ongoing increase in load weight and operational mileage, heavy haul freight trains are experiencing a decline in service condition. Bearing failures can compromise the dynamic performance of train operations, potentially leading to severe issues such as axle overheating and axle breakage, which may result in significant operational safety incidents. Consequently, monitoring the health status of bearings and diagnosing faults have followed with interest for railway departments and researchers.

Specially focused on wheelset bearings, schemes based on signal processing or traditional machine learning have been investigated. Mykhalkiv et al. [1] presented a signal processing algorithm based on minimum entropy deconvolution and square envelope spectra for axle-box bearing faults. Wang et al. [2] designed a feature engineering

algorithm for freight train air brakes. Si et al. [3] verified the effectiveness of utilizing temperature data for monitoring bearing faults. Wu et al. [4] applied a two-stage strategy that combined autocorrelation and an improved flow Gaussian mixture. These plans all rely on expert experience. Nowadays, artificial intelligence (AI)-based intelligent diagnosis, with end-to-end feature representation capabilities, has drawn considerable acclaim in the current era. Liu et al. [5] contributed to electronically controlled pneumatic utilizing feature fusion and ensemble learning techniques. Li et al. [6] proposed a transfer learning wheelset bearing fault diagnosis algorithm based on fast Fourier transform (FFT) and fine-tuned convolutional neural networks (CNN) for heavy haul freight train, but it required target domain annotated dataset that come from the same equipment at constant speed. In summary, owing to the little research, it is currently quite immature within the realm of AI-driven heavy haul freight train wheel bearing

^{*} Corresponding author at: State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing 100044, China.

E-mail addresses: chaohe@bjtu.edu.cn (C. He), hmsi@bjtu.edu.cn (H. Shi), 20221357@bjtu.edu.cn (R. Li), jbli@bjtu.edu.cn (J. Li), zjyu@bjtu.edu.cn (Z. Yu).

<https://doi.org/10.1016/j.aei.2024.102568>

Received 5 March 2024; Received in revised form 10 April 2024; Accepted 23 April 2024

Available online 6 May 2024

1474-0346/© 2024 Elsevier Ltd. All rights reserved.

monitoring. For another matter, the aforementioned techniques failed to concurrently address the principal pain points from two fronts in heavy haul freight train wheel diagnosis.

- † For the first challenge, in the process of actual operation, the heavy haul freight trains inevitably experience acceleration and deceleration process, generating copious non-stationary signals possessing fault information that cannot be detectable at constant speed. It impedes the efficacy of the algorithms predicated on constant speed [7].
- † Meanwhile, it is time-consuming and labor-intensive to obtain sufficient annotated fault data, which hampers the practicality of supervised learning-based intelligent fault diagnosis schemes.

In summary, the issue can be defined as the unsupervised cross-machine transfer learning diagnosis for heavy haul freight train wheel bearings under conditions of small samples and fluctuating speeds, which represents a novel issue.

In response to the first challenge, extensive advancements [7–9] have been made. The traditional signal processing approaches covers order tracking [10], cyclic spectrum correlation, generalized demodulation [11], and time–frequency analysis [12]. Although aforementioned methods exhibit an unambiguous interpretability ascribed to rigorous derivation, the dependence on expert experience hinders the execution of automated diagnosis under time-varying speeds. As the second dimension, the AI-assisted time-varying speed intelligent fault diagnosis can enable the end-to-end mode. Li et al. [13] and Shao et al. [14] introduced machine learning to identify small infrared thermal images of various faults at variable speeds. Xu et al. [15] encoded signals under linear varying speeds into images utilizing STFT, and subsequently input them to capsule network and Transformer to examine the diagnosis capacity under small samples. Rao et al. [16] argued that speed changes can engender imbalanced fault information, and thus suggested a speed normalization module to weaken this impact. Meanwhile, they further devised a plug and play branch named speed adaptive gate [17]. In addition, knowledge decoupling framework [18], PeriodNet [19], anti-symmetric Laplace-Stacked Autoencoder [20], alternative kernel convolutional networks [21], stack denoising autoencoders and gated recurrent unit [22], long short-term memory [23], multi-branch redundant adversarial net [24], restricted sparse networks [25], deep nonlinear order-cyclic convolutional network [26], gcontrast learning [27], attention mechanism [28–30], and generation network [31] have also achieved success in diverse variable speed research. While the methodologies outlined alleviate the first issue, they fails to deal with the cross-domain challenge. It is because that obtaining a multitude of heavy haul freight train data is impractical. Given the availability of abundant annotated data from the laboratory equipment, the cross-machine transfer diagnosis [32] under fluctuating speed scenarios is increasingly being recognized as a promising solution to the second challenge posed by the targeted device — the challenge of limited sample availability.

Despite few research dedicated to cross-machine transfer diagnosis within fluctuating speed scenarios, transfer-learning diagnosis within the same machine has yielded noteworthy outcomes [33–35]. For domain discrepancy metric, Liang et al. [36] leveraged continuous wavelet transform (CWT), deformable convolution, and local maximum mean discrepancy (LMMD) to propose a diagnosis method for the same machine from single source domain variable speed data to multiple target domains. Subsequently, their investigation [37] primarily concentrated on two tasks of constant speed to variable speeds, and variable speeds to variable speeds for the same machine. Cao et al. [38] primarily targeted the research of constant speed to variable speeds and suggested Cauchy kernel induced maximum mean dispersion (CK-MMD) to address the domain shift. Si et al. [39] designed a multi-order moment matching loss. Zhao et al. [40] put forward a graph convolutional neural network based on dynamic multi-kernel maximum mean

discrepancy (DMK-MMD) for aircraft engine diagnosis under different loads, which utilized the order spectrum analysis to obtain the order spectrum signal.

Concerning an alternate category of approaches that involve data preprocessing and model design, Shi et al. [41] proposed a reliable feature-assisted contrastive generalization net, which efficiently achieved the migration of variable speed data from multiple source domains to single target domain. Zhou et al. [42] modified the self-attention module to depthwise separable convolution and utilized comparative regularization to assist the improved Transformer to extract transferable features for low and high speeds both, enabling the detection of medium-speed faults. Subsequently, Gao et al. [43] established a dual correlation model based on graph neural networks to facilitate the transferring from low-speed regions to high-speed regions, verifying the reliability of the proposed algorithm. Xu et al. [44] came up with an algorithm formulated based on synchronous compressed wavelet transform-time feature order spectrum and multi-scale domain adaptation networks (DAN) to substantiate its capability in detecting faults across all speeds and between acceleration and deceleration of the same machine. Lu et al. [45] devised a spectrum alignment and DAN to cope with unbalanced issue at a wide range of speed variation conditions. Pang et al. [46] introduced a time–frequency supervised contrastive learning framework for cross-speed fault diagnosis of bearings, which achieves the transfer diagnosis from constant speed to acceleration or deceleration in various bearing health conditions. Luo et al. [47] introduced a meta-learning approach that employs an elastic prototypical network (EProtoNet) for few-shot fault transfer diagnosis in scenarios ranging from stable to unstable speeds on the same device. However, the issue of cross-machine transfer diagnosis in scenarios on speed fluctuations has been overlooked. Concurrently, the aforementioned reviewed cross-domain algorithms are purely data-driven, which critically depend upon the quality and quantity of data. They disregard the beneficial and significant role of interpretable signal driving and physical constraints on in fostering cross-domain transfer diagnosis.

Notably, time–frequency representation using STFT serves as an effective way to process non-stationary signals under time-varying speeds. However, when the speed varies, the periodicity and amplitude of the impact caused by the fault will alter, and the fault characteristic frequency will also evolve over time, producing more complex non-stationary signals. When STFT is adopted, if the window length is too long, it results in higher frequency resolution, while time resolution will decrease. Therefore, for data under speed fluctuations, while differentiable STFT can automatically search for the optimal window length [48–50], utilizing a fixed window length still remains ineffective and tricky.

In conclusion, the current study does possess certain limitations.

- † To the best of our knowledge, in the realm of heavy haul freight trains, there remains few research on cross-machine bearing diagnosis under variable speeds, and it has not been documented and studied in detail yet. Also, the reported cross-domain algorithms disregard the beneficial impact of interpretable signal driving and physical constraints on transfer diagnosis.
- † Current cross-domain investigations concentrate on different speeds and loads from the same device. Owing to the inability to collect sufficient labels from heavy haul freight train wheels, cross-domain algorithms for the same machine are ill-equipped to address this problem of cross-machine diagnosis under time-varying speeds.
- † The existing approaches to cross-domain variable speeds follow the two-stage strategy, namely involving preliminary preprocessing followed by domain adaptation. The effectiveness hinges extensively upon the quality of preprocessing, while preprocessing relies on manual experience. It not only results in cumbersome parameter selection, but also hinders automated end-to-end diagnosis.

- † While differentiable STFT has shown promise in fault diagnosis, all signal samples adopt a fixed windows length, which is not conducive to capturing distinctive fault information contained in variable speeds, and the cost-sensitive loss does not contemplate the inherently physical information expression of STFT.

In accordance with our prior work—Interpretable and differentiable STFT domain adaptation network (IDSN) [32], interpretable differentiable STFT is considered as “prior knowledge”, DAN counts as “data-driven technique”. In this paper, Balanced Spectrum Quality metric is “physics-informed expression”. A knowledge-data-physics-informed one-stage triple-driven model, termed as pyDSN, is established. pyDSN can alleviate both severe dependence on homemade parameters and quality and quantity of datasets. With the support of domain discrepancy metric, pyDSN can also achieve bearing monitoring from laboratory simulation data to real-world heavy haul freight train wheel bearings.

In comparison with IDSN (DSTFT) [32], pyDSN (MDSTFT) has been substantively improved. For data under speed fluctuations, while IDSN can automatically search for the optimal window length, utilizing a fixed window length still remains ineffective and tricky. Because variable speed data contains some fault information that does not exist absent from constant one, a differentiable variable window length STFT algorithm is proposed. The time-varying fixed windows are modified to time-varying and variable windows to facilitate the fault information expression.

The physical information in this paper, specifically refers to the physics-informed constraints, which serve as a regularization term to alleviate overfitting and enhance the robustness of model. The physics-informed loss integrated with physical prior knowledge can constrain the optimization path, designing task-specific loss functions. The weights of each loss item are balanced, ensuring a lucid fusion of physical prior and neural networks, thereby enhancing interpretability [51,52]. Liao et al. [53] introduced kurtosis and $G - l_1/l_2$ norm for blind deconvolution networks. Yan et al. [54] proposed a physics-guided loss function for modeling healthy degradation; Russell et al. [55] incorporated frequency content into Auto-encoder. Chen et al. [56] represented physical degradation characteristics as loss function constraints to guide network training. Freeman et al. [57] constructed a physics-informed loss to bridge the gap between fault features and the environment. Xu et al. [58] designed a physics-aware loss for building damage assessment; Zhang et al. [59] developed a physical loss function for structural health monitoring to evaluate the difference between model output and finite element output. Nevertheless, for the arduous task of cross-machine transfer diagnosis under speed fluctuations, there remains still a lack of a physics-informed loss capable of achieving high-performance transfer diagnosis.

To bridge this gap, rather than mere classification and measurement losses as in IDSN, the physics-informed constraint, referred to as BSQ, has been enforced in the training process of pyDSN, assessing the time-frequency resolution of TFS. The physical loss is incorporated into the training procedure to extract more generalized feature representations. We argue that the infliction of physical constraints is essential to acquire transferable and discriminative features. This study proposes a taxonomy with two dimensions to assess the TFS quality generated by STFT: One type is the evaluation indicator, which evaluates the quality of the generated STFT through a certain calculation method, such as Intelligent spectrogram [60], Rényi entropy [61], and so forth. Another type is the differentiable loss index, encompassing all attributes of the evaluation indicator while possessing certain unique features. On the one hand, due to its capacity to backpropagate and adaptively modify pivotal parameters in MDSTFT, it can steer MDSTFT towards generating the superior STFT spectrogram; On the other hand, the differentiable loss index unifies the standard of distinct evaluation indicators, circumventing the scenario where one evaluation indicator ranks high and another evaluation indicator rates poorly, achieving trade-off among

various evaluation indicators. In differentiable loss indexes, the trusted metrics based on kurtosis and entropy are two commonly employed methodologies, but these metrics are usually ineffective for pulse signal components of the rotating mechanical equipment [62]. Consequently, balanced spectrum quality loss is devised to evaluate the physical properties of time-frequency spectrograms and ultimately serve as a physical constraint to augment the extrapolation capability of domain adaptation networks.

Lastly, in the realm of intelligent fault diagnosis, the cross-machine diagnosis from time-varying speeds (source domain) to time-varying speeds (target domain) remains an innovative topic that has yet to be explored. PyDSN is the first work, as a starting point food for thought.

The key findings of this investigation encompass the following four aspects.

- † A physics-informed cross-machine transfer method, named as pyDSN, is proposed for heavy haul freight train wheels to address fault detection under time-varying speed conditions. It features two state-of-the-art modules: modulated differentiable STFT (MDSTFT) and balanced spectrum quality loss (BSQ).
- † Distinct from employing a fixed window length, MDSTFT with time-varying window is devised, which adaptively adopts different time-varying window lengths to facilitate the extraction of fault details contained at various rotation speeds. The windows are adaptively adjusted and the length of each is different.
- † In addition to the classification metric and domain discrepancy metric, BSQ, as a third kind of metric to capture transferable features, is devised that can comprehensively embody the physical information of time-frequency spectrum and serve as a regularization. It can steer the differentiable time-varying STFT algorithm towards efficient TFS and also regulate DAN within physical boundaries to learn real-world physics knowledge.
- † Comprehensive analysis illustrates pyDSN demonstrates commendable functionality in time-varying operating conditions, substantiating the reliability and superior performance with the proposed methodology. This is the first successful attempt for cross-machine transfer diagnosis under speed fluctuation.

Subsequent sections are organized into the following structure. We start with defining the main problem and reviewing the prior theories. The various components of pyDSN are elaborately outlined in Section 3. Section 4 incorporates multiple empirical trials and assessments to emphasize the inherent advancements. Finally, Section 5 conducts a thorough interpretability analysis. It concludes in Section 6.

2. Preliminaries

2.1. Problem formulation

The present issue may be characterized as the transfer diagnosis problem for bearings on heavy haul freight train wheels, in scenarios involving small samples and speed fluctuations.

- † Owing to the complexity of data and models, “small sample” is difficult to delineate consistently and precisely. Typically, in machine learning, it can lead to over-fitting. In scenarios where the training data is scarce, the learned feature representation is limited and only fits the training data well, resulting in superior training accuracy yet diminished test precision. According to Ref. [15], the size of training set is 100K (K is the fault category), and thus the task belongs to small sample issue.
- † The concept of “speed fluctuation” can be summarized as the collected signals not being at the constant speed condition, but at a linearly varying speed, because heavy haul freight trains inevitably encounter the reality of acceleration or deceleration.

† Cross-machine transfer diagnosis stems from the fact that the source domain has numerous labels, whereas the target domain lacks labels. This is because a large amount of laboratory data can be available, but there is less labeled data on real-world heavy haul freight train wheel bearings.

The source domain and target domain can be respectively rendered as $D_s: X_m^s = \{(x_i^s, y_i^s) | i = 1, 2, 3, \dots, n_s\}$ and $D_t: X_m^t = \{(x_j^t, y_j^t) | j = 1, 2, 3, \dots, n_t\}$. Among them, m is the number of samples, i, j denote the number of sample categories, $m \leq 20$. x is the sample, and y is label. s represents the source domain, and t represents the target domain. The variable speed dataset amassed from the laboratory in the source domain has a data distribution P_s . The target domain is the real-world heavy haul freight train wheel dataset, with a data distribution P_t .

The task of the paper objective lies in minimizing the $\mathfrak{F} = \langle P_s, P_t \rangle$ at first during the data input stage. Simultaneously, in the classification stage, further diminish the size of \mathfrak{F} in Reproducing kernel Hilbert Space (RKHS). In conclusion, the risk loss can be mitigated via two strategies, namely physical-informed data augmentation and minimized domain discrepancy. In order to rectify these concerns, a physical-informed modulated differentiable STFT-based cross-machine diagnosis for heavy wagons is proposed.

2.2. Related works about differentiable STFT

Studies into DSTFT have predominantly employed two distinct methodologies [32,49,63].

2.2.1. Scheme one

In the seminal work pertaining to DSTFT applied to cross-machine transfer diagnosis, the differentiable coefficient (θ) is introduced; subsequently, the window length and window function are adapted via θ , making the time–frequency resolution ($F_W[t, f]$) trade-off as shown in Eq. (1).

$$F_W[t, f] = \sum_{n=1}^{N-1} x[t+n] W_{m,\theta}[n] e^{-j \frac{2\pi}{N} f n} \quad (1)$$

where t is the sample position, n is window length, j is the unit imaginary number, N is the support size.

Considering Kaiser window as an example, the differentiable window can be expressed as:

$$W_{m,\theta}^{\text{Kaiser}}[n] = \frac{I_0 \left[\beta \sqrt{1 - \left(\frac{2n}{\gamma\theta - 1} \right)^2} \right]}{I_0(\beta)} \quad (2)$$

where the hop/window ratio is γ , I_0 is the first type of modified Bessel function of zero order, and β is any non negative real number used to adjust the shape of the Kaiser window.

2.2.2. Scheme two

The initial step involves dividing signals (x) into sample slices (X) based on hop size and slice length. The time dimension proxy (ϖ) of the window function ($W_N^{\text{Kaiser}}[\varpi]$) is differentiable. ϖ as stipulated in Eq. (2) is differentiable, but the first idea is that n is differentiable, which is the main difference between the two. The obtained time–frequency spectrum is shown in Eq. (3):

$$F_W[t, f] = \text{DFFT}(X \otimes W_N^{\text{Kaiser}}[\varpi]) \quad (3)$$

Among them, DFFT is the discrete Fourier transform, and \otimes is the element by element-wise product.

However, it is noteworthy that both the aforementioned methods utilize the same window length.

3. The proposed method: pyDSN

3.1. The overview of pyDSN

The core idea lies in facilitating extraction of domain-invariant and discriminative features. The prior research predominantly achieve these objectives by meticulously designing backbones, decreasing domain metric discrepancy, or implementing sophisticated training methodologies. Nevertheless, this paper contends that the aforementioned tasks can be accomplished through prior-empowered of Signal processing knowledge, obviating intricate computations and architectural design. Given that the hyper-parameters of signal processing algorithms are generally determined by human experts, so a modulated differentiable STFT algorithm is proposed.

This paper constitutes the first exploration of the cross-machine transfer diagnosis scenario on speed fluctuation. The pseudo-code is as shown in Algorithm 1. The systematic diagnostic procedure encompasses the following fivefold fundamental stages.

- (1) Six vibration signals subjected to speed fluctuations are collected, encompassing one from University of Ottawa dataset, three from independently gathered bearing datasets, and two from gearbox datasets, also including a real-world heavy haul freight train wheelset bearing dataset. With the exception of the latter, the preceding five datasets serve as the source domain with an extensive volume of labels; Real-world dataset form the target domain.
- (2) The Source domain and target domain data are separately input into the proposed Modulated Differentiable STFT (MDSTFT), and output the STFT spectrogram. As the source and target domains originate from diverse mechanical devices, in order to extract domain invariant and discriminative features, MDSTFT modules do not share parameters.
- (3) The spectrogram obtained from MDSTFT is input into the unsupervised balanced spectrum quality (BSQ) loss. The lower the value of BSQ, the more concentrated energy of the spectrogram and the amplified distinctiveness of features. BSQ serves to direct the generation of spectrograms and imposes constraints on DAN. The source domain and target domain have different regularization coefficients, as shown in Eq. (4), where \mathcal{L} is the total loss function, and $\lambda_0, \lambda_1, \lambda_2$ are regularization coefficients, where $\lambda_0 = \frac{-4}{\sqrt{\frac{\text{epoch}}{\max_epoch}} + 1} + 4$, $\lambda_1 = 1.0$, $\lambda_2 = 0.01$.
- (4) The captured spectrogram is fed into DAN and further reduce the distribution shifted between the two domains by combining label smoothing classification loss and domain metric discrepancy.
- (5) In the testing scenario, bearing datasets from heavy haul freight train wheelsets are fed into the trained model to test the performance.

$$\mathcal{L} = \underbrace{\mathcal{L}_{cl}}_{\text{Classification Loss}} + \lambda_0 \underbrace{\mathcal{L}_m}_{\text{Domain Metric Loss}} + \lambda_1 \underbrace{\mathcal{L}_{S-BSQ}}_{\text{Source domain Balanced Spectrum Quality Loss}} + \lambda_2 \underbrace{\mathcal{L}_{T-BSQ}}_{\text{Target domain Balanced Spectrum Quality Loss}} \quad (4)$$

$$\mathcal{L}_m = \left\| \frac{1}{n_s} \sum_{i,j=1}^{n_s} \phi_1(x_i^s, x_j^s) \phi_2(z_i^s, z_j^s) - \frac{1}{n_t} \sum_{i,j=1}^{n_t} \phi_1(x_i^t, x_j^t) \phi_2(z_i^t, z_j^t) \right\|_{\mathcal{H}}^2 \quad (5)$$

where n is the number of samples, ϕ is the kernel function, x is feature, z is the output of the fully connected layer.

A physics-informed pyDSN is composed of MDSTFT and BSQ loss. The overall diagnostic process is depicted in Fig. 1.

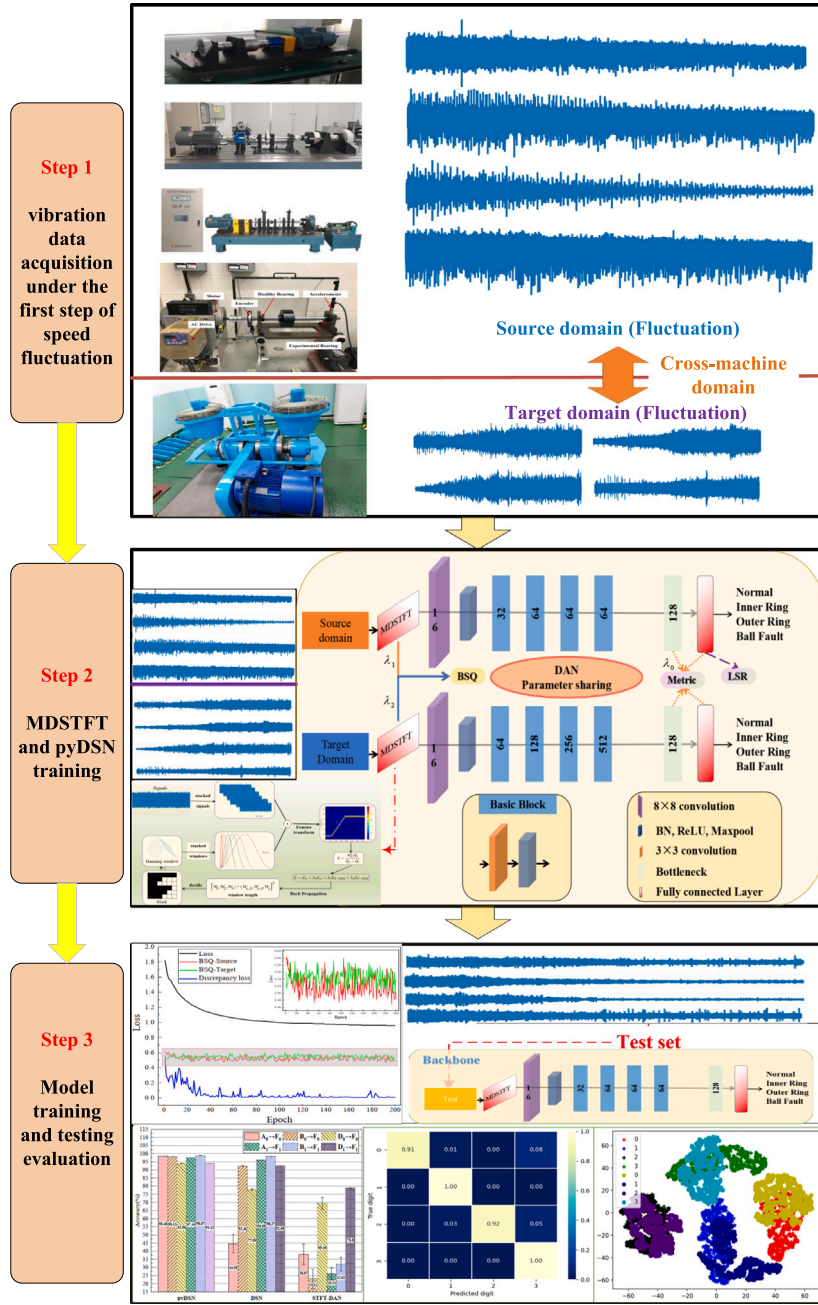


Fig. 1. The overview of pyDSN.

3.2. Modulated differentiable STFT (MDSTFT)

Given the inherent non-differentiability of vanilla STFT, an alternative approach known as DSTFT is proposed, which advocates for an adaptive adjustment of the window length according to data via gradient descent. Nevertheless, it should be noted that DSTFT is dependent on steady-state fault signals, falling short when dealing with variable speed signals.

In order to further amplify the capability of DSTFT to extract fault information, a signal modulation mechanism is proposed. With it, MDSTFT can not only perceive the most suitable window length, deriving from the probabilistic distribution of the data, but also modulate window lengths of different window functions to extract concealed variable speed fault details. In extreme cases, the window length can be

infinitely akin to or equivalent to zero, indicating that the signal slice in extracting fault information will be substantially diminished or have no benefit. Therefore, this modulation scheme endows MDSTFT with flexibility in selecting the appropriate windows for extracting unique fault information within speed fluctuations.

If MDSTFT is implemented as shown in Fig. 2, for a certain signal slice, the window function is not a function of window length (n), but rather a function of temporal dimension (m). Given frame (T), signal length (L), window length (N), hop size (r), the required number of frames is:

$$n_T = \left\lceil 1 + \frac{L' - N - 1}{S} \right\rceil \quad (6)$$

where n_T is the required number of frames and L' is L with padding zero.

The given sample is sliced according to the time dimension, as shown in Eq. (7).

$$X = \begin{bmatrix} X_1 & X_2 & \dots & X_{n_T} \end{bmatrix}^T$$

$$= \begin{bmatrix} x_1 & x_2 & \dots & x_N \\ x_{1+t} & x_{2+t} & \dots & x_{N+t} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1+(n_T-1)t} & x_{2+(n_T-1)t} & \dots & x_{N+(n_T-1)t} \end{bmatrix} \quad (7)$$

Scheme One utilizes the loop sliding window operation, which cannot implement back-propagation when adopts diverse length windows. So, Scheme Two is adopted. The time proxy $\bar{\omega}$ is seen as a one-dimensional vector, according to the number of frames (n_T), rather than a certain value in DSTFT, and B_t is a sequence of equal difference vectors with a tolerance of one on the basis of support size, named the resampled window time.

$$\bar{B}_t = \begin{bmatrix} 0 & 1 & 2 & \dots & N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 2 & \dots & N \end{bmatrix}_{n_T \times N} \quad (8)$$

$$\bar{\omega} \in [N_1, N_2, \dots, N_i]_{n_T \times 1}, (N_i \leq N) \quad (9)$$

STFT for x with the length of L is defined as:

$$F[t, f] = \sum_{n=0}^{N-1} x[t+n]W[n]e^{-j\frac{2\pi}{N}fn} \quad (10)$$

where f is FFT frequency bin.

Vanilla DSTFT regards the window function as function of time proxy, defined as:

$$\hat{F}_{\bar{\omega}}[t, f] = \sum_{n=0}^{N-1} x[\bar{\omega}]W[\bar{\omega}, B_t]e^{-j\frac{2\pi}{N}f\bar{\omega}} \quad (11)$$

In order to achieve differentiable and diverse window lengths, a modulation mechanism is introduced, which is ingeniously implemented through a mask matrix. The mask matrix is composed of continuous value one and ∞ . The resampled time and time proxy are compared element by element, as illustrated in Eq. (12). In this way, it is feasible to adaptively mask the useless resampled time upon the learnable time proxy, so that the window function is incapable of identifying this segment, where the value of the corresponding position of the window function is 0. Hence, a corresponding alteration and learnable ($\Delta_{m,i}$) can be achieved under varying and learnable window lengths ($\bar{\omega}_i$).

$$\Delta_{m,i} = \begin{cases} 1, & \bar{B}_{t,i} < \bar{\omega}_i \\ \infty, & \bar{B}_{t,i} \geq \bar{\omega}_i \end{cases} \quad (12)$$

For example, assuming $N = 5$, $n_T = 3$, then $\bar{B}_t = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix}$

and if the learned window length is $\bar{\omega} = [5, 3, 1]^T$, so $\Delta_M = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & \infty & \infty \\ 0 & \infty & \infty & \infty & \infty \end{bmatrix}$.

where $\lim_{i \rightarrow \infty} W[\Delta_M, \bar{\omega}, \bar{B}_t] = 0$.

The mask matrix is utilized to derive a modulated window function. Herein, the original form of the Kaiser window as an example:

$$W_{\bar{\omega}}^{\text{Kaiser}}[n] = \frac{I_0 \left[\beta \sqrt{1 - \left(1 - \frac{2n}{N-1}\right)^2} \right]}{I_0(\beta)} \quad (13)$$

The Kaiser window in DSTFT can be represented as,

$$W_{\bar{\omega}}^{\text{Kaiser}}[\bar{\omega}, B_t] = \frac{I_0 \left[\beta \sqrt{1 - \left(1 - \frac{2B_t}{\bar{\omega}-1}\right)^2} \right]}{I_0(\beta)} \quad (14)$$

In stark contrast to the aforementioned, the modulated Kaiser window can be obtained through the mask matrix:

$$W_{\bar{\omega}}^{\text{Kaiser}}[\Delta_M, \bar{\omega}, \bar{B}_t] = \frac{I_0 \left[\beta \sqrt{1 - \left(1 - \frac{2\Delta_M \bar{B}_t}{\bar{\omega}-1}\right)^2} \right]}{I_0(\beta)} \quad (15)$$

By employing Eqs. (7) and (15), MDSTFT is ultimately derived as displayed in Eq. (16).

$$\tilde{F}_{\bar{\omega}}[t, f] = \sum_{n=0}^{N-1} x[\bar{\omega}]W[\Delta_M, \bar{\omega}, \bar{B}_t]e^{-j\frac{2\pi}{N}f\bar{\omega}} \quad (16)$$

Comparing Eqs. (16) and (11), numerous differences can readily be discerned: Primarily, the window length is esteemed as a collection of parameters rather than a single variable, achieving adaptive adjustment of diverse window lengths. Secondly, with strictly following window function calculation in the STFT process, a modulation mechanism based on mask matrix is introduced, which not only upholds the differentiability of the window function, but also standardizes the calculation procedure, satisfying the different window functions corresponding to different lengths.

3.3. Physics-informed balanced spectrum quality

Intelligent spectrogram is emerging as an analytical tool for assessing the quality of the time-frequency spectrum. Nonetheless, as a measurement of the physical property of time-frequency resolution, this approach suffers from three significant drawbacks: (i): Intelligent spectrogram is only an evaluation indicator and cannot serve as a measure for the differentiable loss of the produced spectrogram within neural networks. (ii): There exist dual methodologies for quantifying the intelligence spectrogram, such as frequency-domain quality coefficient (\bar{Q}_f) and time-domain quality coefficient (\bar{Q}_t). For some generated spectrogram, it is inconsistent rankings in \bar{Q}_f and \bar{Q}_t , causing confusion for Engineers to select an appropriate parameter. As an outcome, there is remains a scarcity of a holistic evaluation index capable of amalgamating both methodologies. (iii): The larger \bar{Q}_f and \bar{Q}_t , the higher quality of the generated spectrogram, which contradicts the tenets of propelling the backpropagation algorithm towards minimal values.

Adhering to these issues leads us to propose the innovative solution—balanced spectrum quality loss, which helps to balance the time-frequency resolution of TFS and enhance energy aggregation, aiming at enhancing physical expression.

The physics-informed loss BSQ, serving as a regularization term, assists in mitigating over-fitting and diminishes the reliance on vast training datasets. This enhances the robustness of the trained model. It harnesses unlabeled data, facilitating the extraction of more generalized domain-invariant features. Furthermore, it ensures that the output of neural networks aligns more closely with physical principles, thus amplifying the interpretability of the output. BSQ not only steers MDSTFT in capturing high-quality TFS but also functions as a physical constraint, guiding learning domain-invariant features to augment the accuracy of cross-machine transfer diagnosis.

The quality coefficients serve as the proxies for two physical properties—temporal resolution and frequency resolution. As an initial step, compute the mean value of all rows and columns within the spectrogram:

$$\mu_f[t] = \frac{1}{F} \sum_{f=0}^{F-1} |\hat{F}_{\bar{\omega}}[t, f]| \quad (17)$$

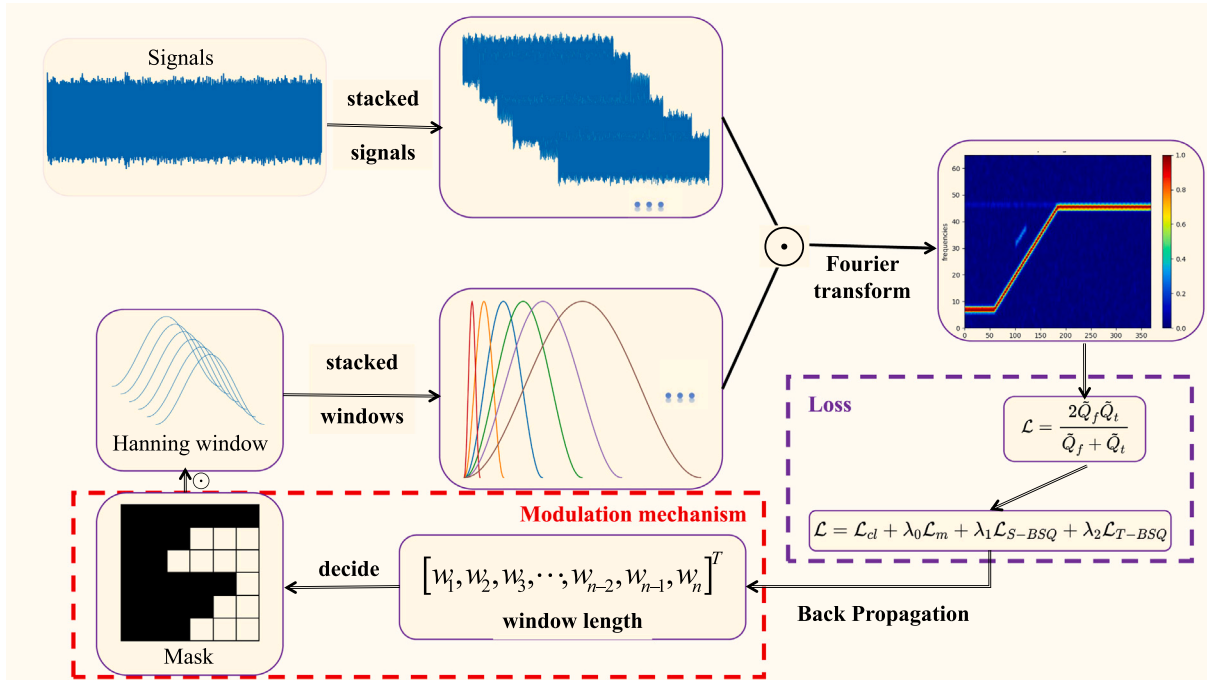


Fig. 2. \odot is element-wise product. Multiplying the signals with the window functions with variable window lengths, followed by DFFT yields MDSTFT. Subsequently, the balanced spectrum quality loss is calculated, the optimal window function is determined employing three evaluation metrics and backpropagation procedure.

$$\mu_t[f] = \frac{1}{T} \sum_{t=0}^{T-1} |\hat{F}_w[t, f]| \quad (18)$$

where $\mu_t[f]$ and $\mu_f[t]$ signify the average values of time and frequency variables, respectively.

Following, calculate their standard deviation ($\sigma_f[t]$) and ($\sigma_t[f]$), as shown in Eqs. (19) and (20).

$$\sigma_f[t] = \sqrt{\frac{1}{F-1} \sum_{f=0}^{F-1} \left| |\hat{F}_w[t, f]| - \mu_f[t] \right|^2} \quad (19)$$

$$\sigma_t[f] = \sqrt{\frac{1}{T-1} \sum_{t=0}^{T-1} \left| |\hat{F}_w[t, f]| - \mu_t[f] \right|^2} \quad (20)$$

The balanced coefficient of variation is designated as the quotient of the mean divided by the standard deviation, which is not necessarily better as it is larger, but better as it is smaller. This corresponds precisely to the direction of minimization pursued in the backpropagation algorithm and constitutes an appropriate loss function during the optimization process.

$$\tilde{c}_f[t] = \frac{\mu_f[t]}{\sigma_f[t]} \quad (21)$$

$$\tilde{c}_t[f] = \frac{\mu_t[f]}{\sigma_t[f]} \quad (22)$$

Finally, obtain the Rational Quality Coefficient of the given time-frequency spectrum:

$$\tilde{Q}_f = \frac{1}{T} \sum_{t=0}^{T-1} \frac{\tilde{c}_f[t]}{\tilde{c}_f[t]_{\max}} \in (0, 1) \quad (23)$$

$$\tilde{Q}_t = \frac{1}{F} \sum_{f=0}^{F-1} \frac{\tilde{c}_t[f]}{\tilde{c}_t[f]_{\max}} \in (0, 1) \quad (24)$$

As illustrated in Eqs. (23) and (24), to circumvent the occurrence of the phenomenon of gradient explosion or disappearance, with adjustment of \tilde{Q}_f and \tilde{Q}_t to a maximum value 1.0. In intelligent spectrogram, it is observed that for \tilde{Q}_f and \tilde{Q}_t of the specific spectrogram, there exists

an inconsistent ranking for a certain window length. For example, for $n = 8$, the ranking of \tilde{Q}_f is first, but the ranking of \tilde{Q}_t is at the end [60]. To remedy this predicament, a novel loss function based on harmonic mean is devised, which incorporates both time resolution and frequency resolution into consideration.

In adherence to the principles of harmonic series, \mathcal{L}_{BSQ} is dictated by \tilde{Q}_f and \tilde{Q}_t . The value of the harmonic mean (\mathcal{L}_{BSQ}) resides within the confines between \tilde{Q}_f and \tilde{Q}_t . If \mathcal{L}_{BSQ} diminish, it is imperative for both \tilde{Q}_f and \tilde{Q}_t to be contemporaneously decreased. The minimization of \mathcal{L}_{BSQ} implies that \tilde{Q}_f and \tilde{Q}_t are also optimized towards the minimal direction. Hence, the interplay between the two was holistically taken into account for achieving the optimal window function optimization result. According to the definition of harmonic mean, Eq. (25) can emerge:

$$\mathcal{L}_{BSQ} = \left(\frac{\tilde{Q}_f^{-1} + \tilde{Q}_t^{-1}}{2} \right)^{-1} = \left(\frac{\tilde{Q}_f + \tilde{Q}_t}{2} \right)^{-1} \quad (25)$$

$$= \frac{2\tilde{Q}_f\tilde{Q}_t}{\tilde{Q}_f + \tilde{Q}_t}$$

BSQ analysis can balance two physical properties: temporal resolution and frequency resolution.

3.4. Training pipeline of pyDSN for cross-machine transfer diagnosis

The diagnostic procedure exhibited by pyDSN is illustrated in Fig. 3. During the forward propagation phase, variable speed data are fed into MDSTFT and DAN to generate soft labels, and subsequently, AdamW algorithm is deployed to optimize the parameters within pyDSN.

$$\left\{ \begin{array}{l} \mathcal{L} = \mathcal{L}_{cl} + \lambda_0 \mathcal{L}_m + \lambda_1 \mathcal{L}_{S-BSQ} + \lambda_2 \mathcal{L}_{T-BSQ} \\ \theta_s \leftarrow \theta_s - \mu \left(\frac{\partial \mathcal{L}}{\partial \theta_s} \right) \\ \theta_t \leftarrow \theta_t - \mu \left(\frac{\partial \mathcal{L}}{\partial \theta_t} \right) \\ \theta_0 \leftarrow \theta_0 - \kappa \left(\frac{\partial \mathcal{L}}{\partial \theta_0} \right) \\ \theta_1 \leftarrow \theta_1 - \kappa \left(\frac{\partial \mathcal{L}}{\partial \theta_1} \right) \end{array} \right. \quad (26)$$

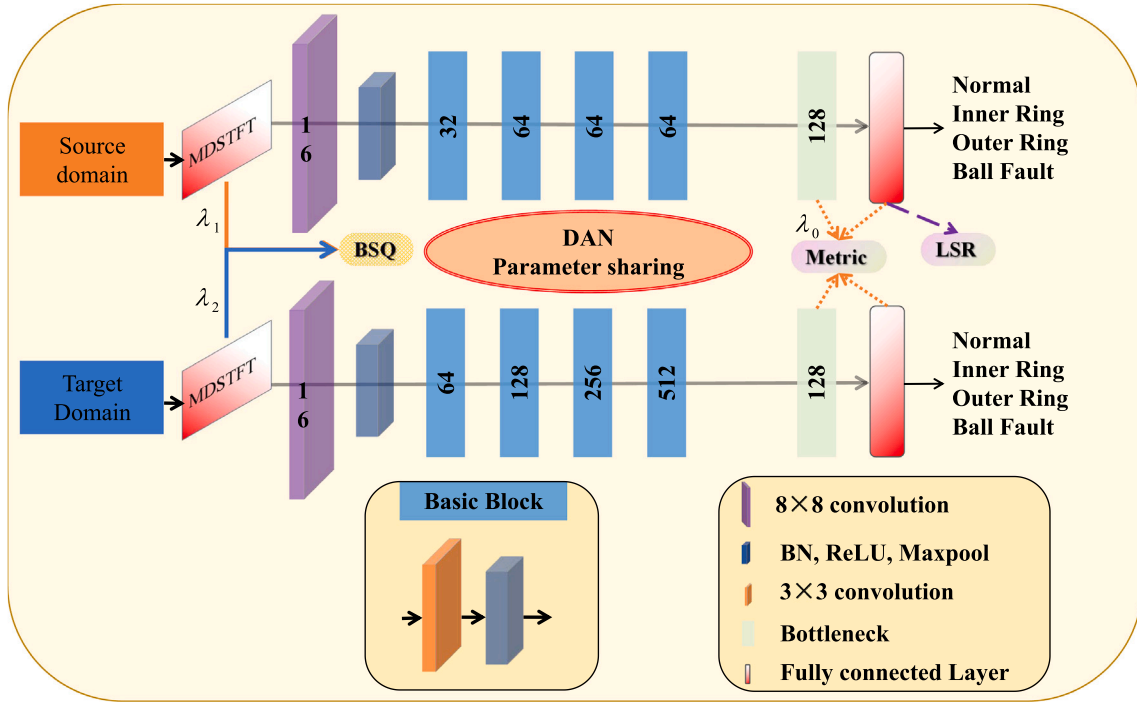


Fig. 3. Flowchart of a PyDSN model.

where $\mu = 0.001, \kappa = 100.0$ are the learning rate, and $\Theta_s, \Theta_t, \theta_0, \theta_1$ are the hyper-parameters of the source domain classifier, target domain classifier, source domain MDSTFT, and target domain MDSTFT, respectively.

The proposed BSQ loss serves not only to guide the adjustment of STFT window lengths, but also to bind DAN. The reverse gradient propagation of BSQ can be representationally encapsulated as Eq. (27):

$$\begin{aligned} \frac{\partial L_{BSQ}}{\partial \Theta} &= \frac{\partial \left(\frac{\hat{Q}_f^{-1} + \hat{Q}_t^{-1}}{2} \right)^{-1}}{\partial \Theta} \\ &= - \left(\frac{\hat{Q}_f^{-1} + \hat{Q}_t^{-1}}{2} \right)^{-2} \left[\frac{\partial (2\hat{Q}_f)^{-1}}{\partial \Theta} + \frac{\partial (2\hat{Q}_t)^{-1}}{\partial \Theta} \right] \end{aligned} \quad (27)$$

where Θ is the hyper-parameters of pyDSN.

The flowchart depicting the pseudo code for the total training of this model is shown as Algorithm 1.

4. Application to wheelset bearings of freight trains

4.1. The description of datasets and tasks

4.1.1. Data description

This investigation covers bearing and gearbox datasets: a dataset from Ottawa University; three datasets from the independently maintained experimental platform for bearings and gearboxes; and an additional dataset gathered from the real-world heavy haul freight train wheelset bearing platform designed and established by CRRC Qingdao Sifang. A total of five datasets are utilized to investigate the cross-machine transfer diagnosis capabilities of the proposed pyDSN algorithm. A concise overview of these datasets is delineated below (see Table 1).

- (1) BJTU₁: The bearing fault dataset originates from the double-span and double-rotor comprehensive fault experimentation platform of Houde Automation Measurement as illustrated in Fig. 4a. Utilized within the experiment is the NSK-6308 deep groove

ball bearing at a sampling rate of 20 kHz. The experimental bearings into four categories based on their respective locations: Inner race fault (IR), Outer race fault (OR), Ball fault (B) and the healthy bearing (H). Localized pitting failures are mimicked through electrospark machining methods across these various sections.

- (2) BJTU₂: The gearbox fault dataset comes from the planetary gearbox fault test bench of Houde Automation Measurement, as shown in Fig. 4b. The experimental setup, such as sensors and sampling frequency is identical with BJTU₁. It simulates three distinct gearbox failures at varying rotational speeds, encompassing cracks (C), tooth fractures (TF), tooth missing (TM), and health (H).
- (3) BJTU₃: The third fault database originates from the rotor gear comprehensive fault experimental platform, depicted in Fig. 4c, where the employed rolling element model is the NSK-6205 single row deep groove ball bearing. The remaining experimental parameters adhere to those of BJTU₁.
- (4) Ottawa [64]: The experimental setup of the variable speed rolling bearing dataset in the Mechanical Engineering Laboratory of the University of Ottawa, Canada is shown in Fig. 4d. The installation encompasses the ER16K ball bearing support shaft, replacing the selected experimental bearing with bearings presenting distinct health conditions. Furthermore, an accelerometer is positioned onto the external shell of the experimental bearing to acquire vibration data, with a sampling frequency 200 kHz.

The railway scenario dataset comes from the heavy haul freight train wheelset bearing platform developed by CRRC Qingdao Sifang, depicted in Fig. 5(a), which exhibits precision in simulating the payloads, axle bearing speeds, as well as the corresponding environment while under genuine operations, conforming to the real-world railway scenario. Specifically, vertical load is utilized to simulate axle load; Lateral load is used to simulate different line conditions, such as turning, passing through turnouts, and the lateral force on the track during uphill and downhill operations; The fan is employed to simulate

Algorithm 1 The training and testing process of pyDSN.

Input: $D_s = X_m^s$: source domain; $D_t = X_m^t$: target domain; $\lambda_0, \lambda_1, \lambda_2$: coefficients; M : training epoch;
Output: Testing accuracy of PyDSN.

- 1: Initialize PyDSN
- 2: **for** $i \leftarrow 0; i \leq M; i \leftarrow i + 1$ **do**
- 3: Get source-domain spectrogram $S(D_s)$ from by MDSTFT.
- 4: Calculate source-domain balanced spectrum quality loss \mathcal{L}_{S-BSQ} by Eq. (25).
- 5: Get target-domain spectrogram $S(D_t)$ from BJTU-QS dataset by MDSTFT.
- 6: Calculate target-domain balanced spectrum quality loss \mathcal{L}_{T-BSQ} by Eq. (25).
- 7: Get source-domain feature $F(D_s)$.
- 8: Get source-domain soft prediction $P(D_s)$.
- 9: Calculate smoothed cross-entropy loss \mathcal{L}_{cl} .
- 10: Get target-domain feature $F(D_t)$.
- 11: Get target-domain soft prediction $P(D_t)$.
- 12: Extract domain-invariant features using \mathfrak{F} .
- 13: Calculate domain metric loss \mathcal{L}_m .
- 14: Train the parameters θ_s of source-domain MDSTFT by \mathcal{L}_{cl} , $\lambda_0 \mathcal{L}_m$, $\lambda_1 \mathcal{L}_{S-BSQ}$ and $\lambda_2 \mathcal{L}_{T-BSQ}$.
- 15: Train the parameters θ_t of target-domain MDSTFT by \mathcal{L}_{cl} , $\lambda_0 \mathcal{L}_m$, $\lambda_1 \mathcal{L}_{S-BSQ}$ and $\lambda_2 \mathcal{L}_{T-BSQ}$.
- 16: Train the feature extractor parameters θ_1 by \mathcal{L} .
- 17: Train the classifier parameters θ_2 by \mathcal{L} .
- 18: **end for**
- 19: Save the trained model Θ .
- 20: Get test data and load trained model Θ .
- 21: Get test feature $F(D'_t)$ and prediction $P(D'_t)$.

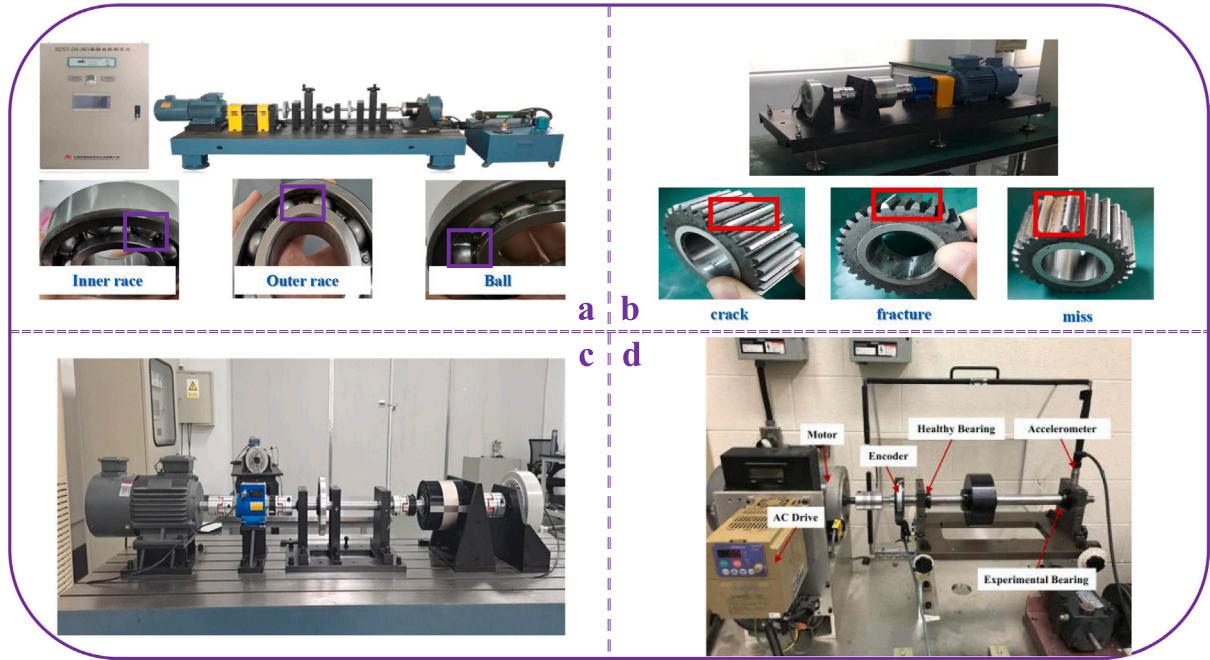


Fig. 4. Equipments: (a) is double-span and double-rotor comprehensive fault experimentation platform; (b) is gearbox fault experiment platform; (c) is rotor-gear comprehensive fault experiment platform; (d) is experimental setup from University of Ottawa.

crosswind during operation, setting between 8 m/s to 10 m/s. The standard bearings is equipped on one side of the wheelset and experimental bearing is mounted onto the opposite side. A vibration acceleration sensor is fixed on the outside of the axle box with a sampling frequency 16 kHz. The tested bearing is 352226X2-2RZ (Fig. 5(e)(f)), which is a sealed dual row tapered roller bearing with double inner rings and middle isolation. The two end surfaces of the bearing are equipped with skeleton rubber sealing rings, with an inner diameter of 130 mm. It is used for accelerating freight car bogies and consists of outer rings, inner rings, rollers, cages, middle separators, sealing devices and other components. Three types of faults consist of: inner ring, outer ring,

and balling, which are confirmed after dismantling and inspecting in the online monitoring system or manual check. In addition, it includes healthy bearings (H). The bearing rotation speed is set at 60 km/h ~ 120 km/h (see Table 1).

4.1.2. Task description

The heavy haul freight train inevitably encounters acceleration and deceleration during operation, which may harbor distinctive features compared to constant speed data. By utilizing self-collected and public variable speed data, the heavy haul freight train wheel bearing health monitoring is addressed through cross-machine transfer diagnosis methodology.

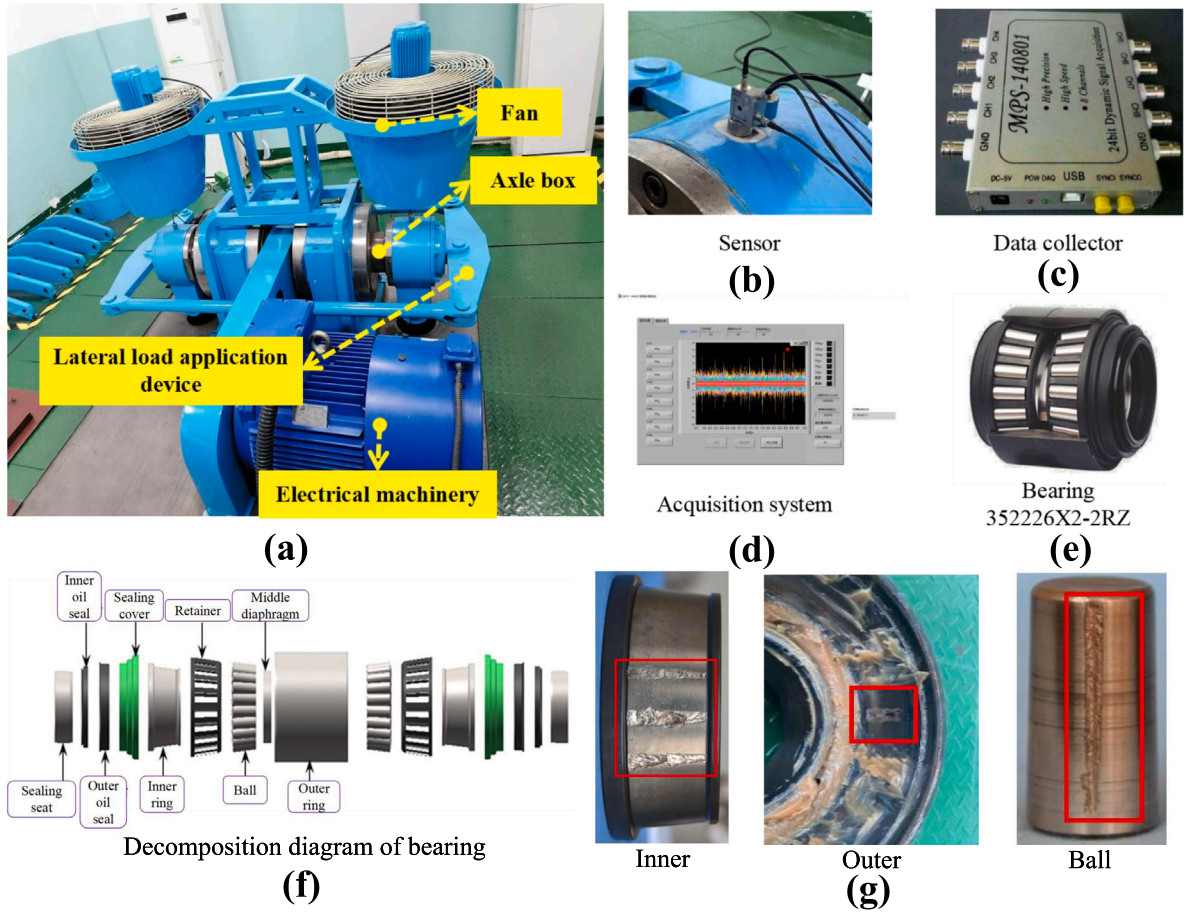


Fig. 5. Heavy haul freight train wheelset bearing related equipment: (a) is heavy haul freight train wheelset bearing experimental platform; (b) is the data collector; (c) is the sensor; (d) is the acquisition system; (e) is the bearing (352226X2-2RZ); (f) is the schematic diagram of bearing disassembly; (g) is the partial failure types of railway freight train bearing.

Table 1
The detailed description of datasets.

Name	Data source	Speed (r/min)	Sampling frequency (kHz)	Health status
A ₀	BJTU ₁	900~2400	20	H,IF,OF,BF
A ₁	BJTU ₁	2400~900		
B ₀	BJTU ₂ (bearing)	1000~2500	20	H,IF,OF,BF
B ₁	BJTU ₂ (bearing)	2500~1000		
C ₀	BJTU ₂ (gear)	1000~2500	200	H,C,TF,TM
C ₁	BJTU ₂ (gear)	2500~1000		
D ₀	Ottawa	Accelerate	20	H,IF,OF,BF
D ₁	Ottawa	Decelerate		
E ₀	BJTU ₃	500~2500	16	H,C,TF,TM
E ₁	BJTU ₃	2500~500		
F ₀	Heavy haul freight train	60 km/h~120 km/h	16	H,IF,OF,BF
F ₁	Heavy haul freight train	120 km/h~60 km/h		

The signal is segmented into samples using a sliding window, with a sample length 3072. In the scenario on time-varying speeds, there are 100 samples for each category in cross-machine transfer diagnosis. Therefore, each category encompasses 100K samples in source domain or target domain. The test set comes from the target domain, yielding a total collection of 900K samples (K represents the number of categories). The multiple cross-machine tasks are constructed, where A→F represents A as the source domain and F as the target domain. All relevant experiments are executed on RTX 4090 24 GB GPU equipped with Pytorch 2.0.

The experiment are repeated five times and record the average and standard deviation. The other parameter settings are shown in

Table 2. In addition to the bearing dataset, in order to improve the credibility and generalization, gearbox data is employed to validate the performance with pyDSN. Of these, A, B, D, and F are bearing datasets. C and E are gearbox datasets.

4.2. Case analysis

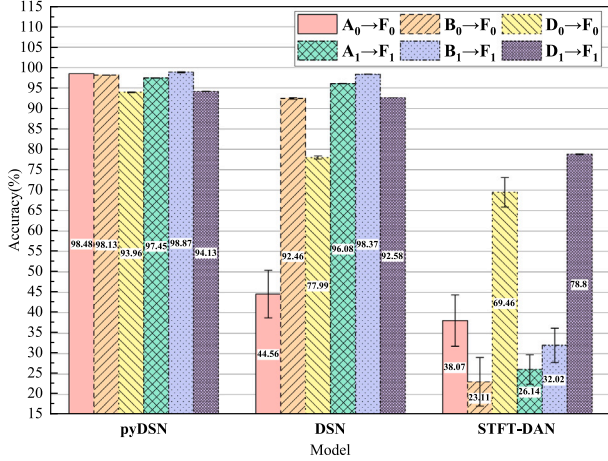
4.2.1. Comparison performance with DSN, STFT, and pyDSN

pyDSN is improved over the typical STFT and DSN (IDSN). To evaluate the health status of wheelset bearings, the wheelset bearing dataset is applied as the target domain, while other datasets are utilized as the source domain. As shown in Fig. 6, among the six acceleration

Table 2

Model parameter settings.

Parameter	Value
batch_size	20
Backbone	WDCNN
window function	Kaiser
Optimizer	AdamW
stride	16
max_epoch	200
seed	3407
training set	400
testing set	3600
lr	0.001
lr (MDSTFT)	100.0

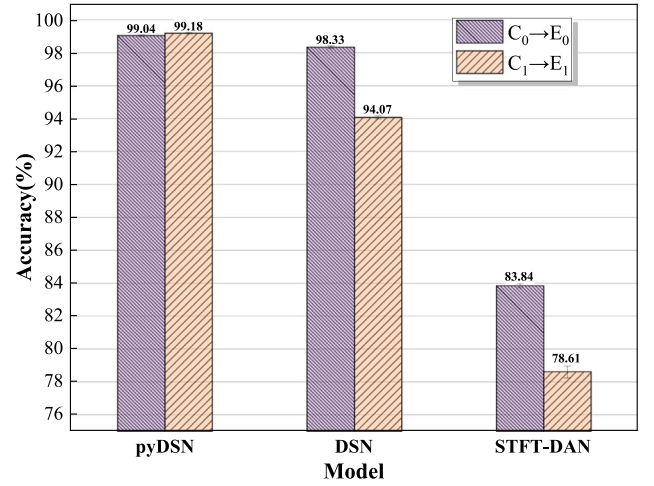
**Fig. 6.** Performance with three methods of different transfer tasks on bearings.

and deceleration tasks executed, pyDSN achieved an average accuracy of 96.85%, far surpassing DSN (83.67%). This is due to the fact that MDSTFT does not utilize a fixed window length, but instead uses multiple different windows, which can adaptively collect fault information on conditions at variable speeds. Moreover, pyDSN, under the influence of BSQ loss, enables MDSTFT to generate high-quality spectrograms, which helps to mitigate the distribution shifts between the source and target domains. The two-stage strategy, STFT-DAN, which uses labor experience to select window sizes, has limited performance. DSN with a fixed differentiable window faces a larger distribution shifts between domains, which makes it off-limits to extract discriminative fault feature information, resulting in an accuracy of less than 80%.

To verify comprehensive generalization capability of pyDSN, as shown in Fig. 7, experiments are conducted on the gearbox components. B and D are collected in-house. Compared pyDSN and DSN, self-collected datasets generally exhibit a high signal-to-noise ratio and obvious fault characteristics, making it relatively easy to obtain domain-invariant features. Despite pyDSN (99.11%) and DSN (96.20%) with high diagnostic accuracy ($>96.00\%$), pyDSN still improves by 2.91% in comparison to DSN. STFT-DAN (81.23%) still performs poorly, highlighting the significance of data enhancement in improving transferability under speed fluctuations and small sample conditions. In summary, although signal processing universally acknowledged as a preprocessing methodology, differentiable signal processing substantially enhances the data feature enhancement capacity of signal processing algorithms due to its robust ability to adaptively adjust pivotal parameters with the assistance of gradient-descent algorithms.

4.2.2. The plug-and-play MDSTFT and BSQ

As illustrated in Table 3, taking $A_0 \rightarrow F_0$ as an example, BSQ, MDSTFT, and DSTFT can all be seamlessly integrated into several

**Fig. 7.** Performance with three methods of different transfer tasks on gearboxes.**Table 3**

Different backbones being integrated BSQ and MDSTFT [65–69].

Model	BSQ(w/o)	DSTFT	MDSTFT
EfficientNet v2 [65]	85.56%	43.46%	91.61%(↑ 6.05%)
ResNet-18 [66]	88.69%	83.95%	95.83%(↑ 7.14%)
DenseNet-121 [67]	70.86%	59.55%	90.92%(↑ 20.06%)
WDCNN-2D [68]	94.56%	33.21%	97.31%(↑ 2.75%)
DRSN-CW [69]	76.67%	47.82%	96.11%(↑ 19.44%)
Average	83.27%	53.60%	94.36%(↑ 11.09%)

classic networks without conflicts. We have selected lightweight model (EfficientNet v2), deep models (ResNet-18 and DenseNet-121), as well as fault diagnosis models (WDCNN and DRSN-CW), as these models are extensively utilized and representative in fault diagnosis. BSQ (w/o) signifies the utilization of MDSTFT, without BSQ loss; MDSTFT represents the usage of BSQ loss; DSTFT denotes the usage of DSTFT and BSQ loss. Faced with demanding cross-machine and variable speed condition tasks, the performance with DSTFT depends on the integrated backbone, exhibiting significant difference in performance under different backbones. For example, ResNet-18 can achieve 83.95%, while it drops to 47.82% when utilizing lightweight DRSN-CW, indicating poor generalization when processing speed fluctuation datasets. MDSTFT can improve this by 40.76% compared with DSTFT, suggesting that employing different windows can effectively extract exclusive fault information. By comparing whether to use physics-informed BSQ, the hybrid-driven pyDSN improves by 11.09%, implying that the hybrid-driven model incorporating physical limitations can enhance the capability to obtain discriminative features. This proves that the physics-informed regularization processes a strong driving force for extracting domain-invariant features and improving the generalization capacity.

As shown in Fig. 8, the iterative process of the loss function is delineated during the learning phase. Loss represents the overall loss function (Eq. (4)), and joint maximum mean discrepancy (JMMD) loss (\mathcal{L}_m) is identified to measure the discrepancy between source and target domains. At the conclusion of iteration, both losses ultimately attain full convergence. In addition, the BSQ loss in the source domain decreased from 0.59438 to 0.46898, whilst the BSQ loss associated with the target domain decreased from 0.55964 to 0.50112. This observation highlights that the gradient-descent algorithm can effectively search for the optimal window lengths, enhance spectral energy clustering, and further facilitate the learning of domain-invariant features.

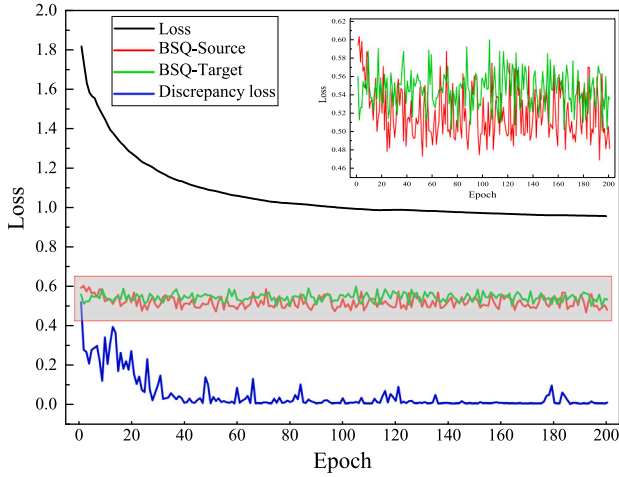


Fig. 8. The diagram of loss function iteration.

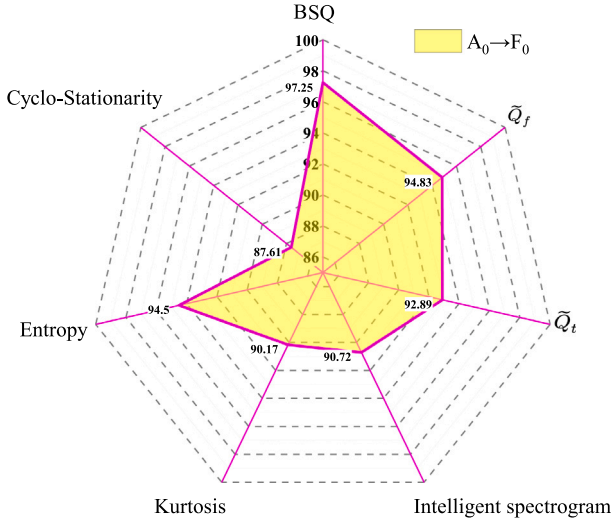


Fig. 9. The accuracy comparison with other loss functions (%).

4.2.3. The experiment of BSQ

To demonstrate the pioneering nature of the proposed BSQ. A comprehensive evaluation is conducted using prevalent physics-informed losses: Entropy and Kurtosis are the predominantly metrics for evaluating the spectrogram energy focusability; BSQ is a trailblazing iteration from Intelligent Spectrum; Cycle-Stationarity [62] is another method. \tilde{Q}_f only focuses on frequency resolution, while \tilde{Q}_t only considers time resolution, while BSQ aims to strike a balance between these two dimensions. The accuracy stands at 97.25%, when jointly considering time–frequency resolution, respectively yielding 2.42% and 4.36% over either consideration alone. This underscores that the spectrogram needs to balance time–frequency resolution to accurately represent fault information.

Moreover, compared to other physics-guided losses, BSQ also showcases superior performance, significantly improving the adaptability to cross-domain, with a 9.64% increase over Cycle-Stationarity. BSQ proves to a highly efficacious assessment tool to quantify the energy distribution and fault feature expression within spectrograms (see Fig. 9).

4.3. The performance analysis of pydsn on heavy haul freight train wheelset bearings

To assess the superior performance of pyDSN with other state of the arts (SOTA). The first type is the classical mapping-based models such as DDC (MMD), DCORAL, DJDA, and DSAN; The second type comprises the adversarial domain adaptation model DANN; The third type consists of emerging models the emerging models like CK-MMD [70], MMSD [71], and DDTLN [72] for cross-machine transfer diagnosis; The last type is the con-temporarily popular hybrid-driven deep models embedded signal knowledge, where signal processing is regarded as a form of data augmentation, and these models have achieved remarkable performance in their respective tasks, encompassing SincNet [73], MCN_WFK [74], NTScatNet [75], DFAWNet [76], GT-FENet [77], GaborNet [78], WIDAN [79], MWPN [80], and TFN [81], as illustrated in Table 4.

The first three algorithms mentioned are purely data-driven, which perform inadequately under cross-machine variable speed tasks, which indicates that data-driven models depend on the quantity and quality of available data. In the presence of small samples and larger disparity between the source and target domains, these algorithms are likely to fail. Conversely, the hybrid-driven methods demonstrate highly competitive performance in the face of such challenging tasks. The signal processing module can reduce the noises, highlight fault information, and mitigate the adverse effects of noise on the source and target domains. However, reasonable signal processing methods are required to achieve this effect. For instance, utilization of Gabor transform results in an average accuracy of only 40.23%, which is even worse than some purely data-driven methods.

The task encompasses two thorny challenges. One is across machines, as the data collected by different devices has greater distribution shifts; The other challenge is variable speed data, where the fault frequency is changing caused by changes in speed, rather than like constant speed. When pyDSN is tested against six tasks, the accuracy is 94.81%, 96.64%, 98.78%, 93.83%, 98.86%, and 99.00%, respectively, with an average accuracy is 96.97%. This represents a minimum improvement of 22.7%. The MDSTFT module in pyDSN aims to use differentiable windows and physical losses to extract fault information at variable speeds, while shortening the distribution differences between the source and target domains. However, other tasks perform poorly due to the lack of the aforementioned either.

The superior domain-invariant features can be explained from two perspectives. On one hand, domain-invariant features are extracted under the guidance of MDSTFT that furnishes an interpretable initialization input for acquiring domain-invariant features. Meanwhile, MDSTFT maintains interpretable with following the calculation process of STFT. It can be seamlessly integrated with domain adaptation networks without conflicts. On the other hand, domain-invariant features are extracted under the guidance of BSQ. BSQ measures energy focusability of TFS and balances two physical properties: time–frequency resolution. It directs the optimization trajectory of neural networks, allowing physical insights to influence the optimization process, and facilitates to extract more robust domain-invariant features, making the feature representations more closely with physical principles and enhancing interpretability.

According to Table 4, in the Task $A_0 \rightarrow F_0$, interpretable methods such as SincNet and NTScatNet demonstrate superior performance. For the specific task, These methods utilize the Sinc function and wavelet scattering structure, respectively, to extract generalized features from the source domain, thereby minimizing the discrepancy between the two domains. In the case of pyDSN, this subpar performance could be attributed to the influence of regularization weight parameters, which may result in extracted features that do not effectively represent domain-invariant features. Nevertheless, this performance decline is within acceptable limits, with the accuracy remaining above 94.00%.

Table 4
Performance comparison with other methods (%).

	$A_0 \rightarrow F_0$	$B_0 \rightarrow F_0$	$D_0 \rightarrow F_0$	$A_1 \rightarrow F_1$	$B_1 \rightarrow F_1$	$D_1 \rightarrow F_1$	Average
DDC	44.92 \pm 2.88	59.33 \pm 3.17	50.42 \pm 4.00	49.17 \pm 0.46	28.41 \pm 0.03	67.92 \pm 2.13	50.03
DCORAL	32.17 \pm 2.75	33.33 \pm 2.41	37.58 \pm 1.04	29.17 \pm 0.05	28.33 \pm 0.09	63.08 \pm 1.25	32.28
DJDA	29.18 \pm 0.93	50.96 \pm 4.57	50.32 \pm 8.38	45.57 \pm 5.25	33.11 \pm 0.09	61.75 \pm 1.88	45.15
DSAN	32.07 \pm 1.34	33.21 \pm 0.86	41.33 \pm 0.83	47.75 \pm 6.17	27.33 \pm 0.13	67.42 \pm 0.59	41.52
DANN	30.47 \pm 0.36	36.19 \pm 0.60	61.00 \pm 0.61	43.19 \pm 0.59	18.69 \pm 0.03	47.61 \pm 0.31	39.53
CK-MMD [70]	30.42 \pm 1.21	38.75 \pm 2.30	31.75 \pm 0.03	52.33 \pm 1.00	28.91 \pm 0.09	64.00 \pm 1.08	41.02
MMSD [71]	89.75 \pm 7.75	19.92 \pm 1.55	91.25 \pm 3.92	26.17 \pm 1.24	46.08 \pm 9.97	64.08 \pm 1.29	56.21
DDTLN [72]	30.14 \pm 0.78	33.21 \pm 1.95	64.14 \pm 3.32	29.54 \pm 6.45	52.33 \pm 2.41	55.12 \pm 1.37	44.08
SincNet [73]	98.67 \pm 0.34	33.08 \pm 1.24	67.25 \pm 1.46	65.08 \pm 5.45	27.17 \pm 0.51	76.00 \pm 0.77	61.21
MCN_WFK [74]	27.50 \pm 2.17	32.00 \pm 0.00	36.50 \pm 1.08	28.00 \pm 0.29	26.00 \pm 0.00	35.25 \pm 1.54	30.88
NTScatNet [75]	100.00 \pm 0.00	71.00 \pm 2.54	72.61 \pm 0.72	84.54 \pm 1.36	46.67 \pm 2.38	71.38 \pm 0.50	74.37
DFAWNet [76]	94.63 \pm 0.59	86.05 \pm 3.21	59.63 \pm 2.70	80.35 \pm 1.44	29.15 \pm 2.20	68.28 \pm 0.96	69.68
GTFENet [77]	57.92 \pm 1.32	40.96 \pm 1.03	58.58 \pm 3.76	60.33 \pm 3.19	60.50 \pm 0.57	89.88 \pm 2.42	61.36
GaborNet [78]	39.75 \pm 0.79	57.50 \pm 1.55	30.17 \pm 4.19	32.25 \pm 1.14	49.00 \pm 2.95	32.75 \pm 1.22	40.23
WIDAN [79]	96.88 \pm 5.81	22.13 \pm 4.21	70.63 \pm 4.86	71.88 \pm 1.30	72.92 \pm 1.85	59.38 \pm 0.87	65.64
MWPN [80]	50.17 \pm 1.48	28.08 \pm 0.84	75.58 \pm 0.71	58.58 \pm 2.13	44.83 \pm 0.60	65.18 \pm 4.14	53.74
TFN [81]	75.00 \pm 0.43	37.25 \pm 1.63	96.75 \pm 1.50	33.75 \pm 3.70	38.00 \pm 1.90	85.60 \pm 2.29	61.06
pyDSN	94.81 \pm 3.02	96.64 \pm 0.42	98.78 \pm 0.50	93.83 \pm 0.29	98.86 \pm 0.32	99.00 \pm 0.04	96.97

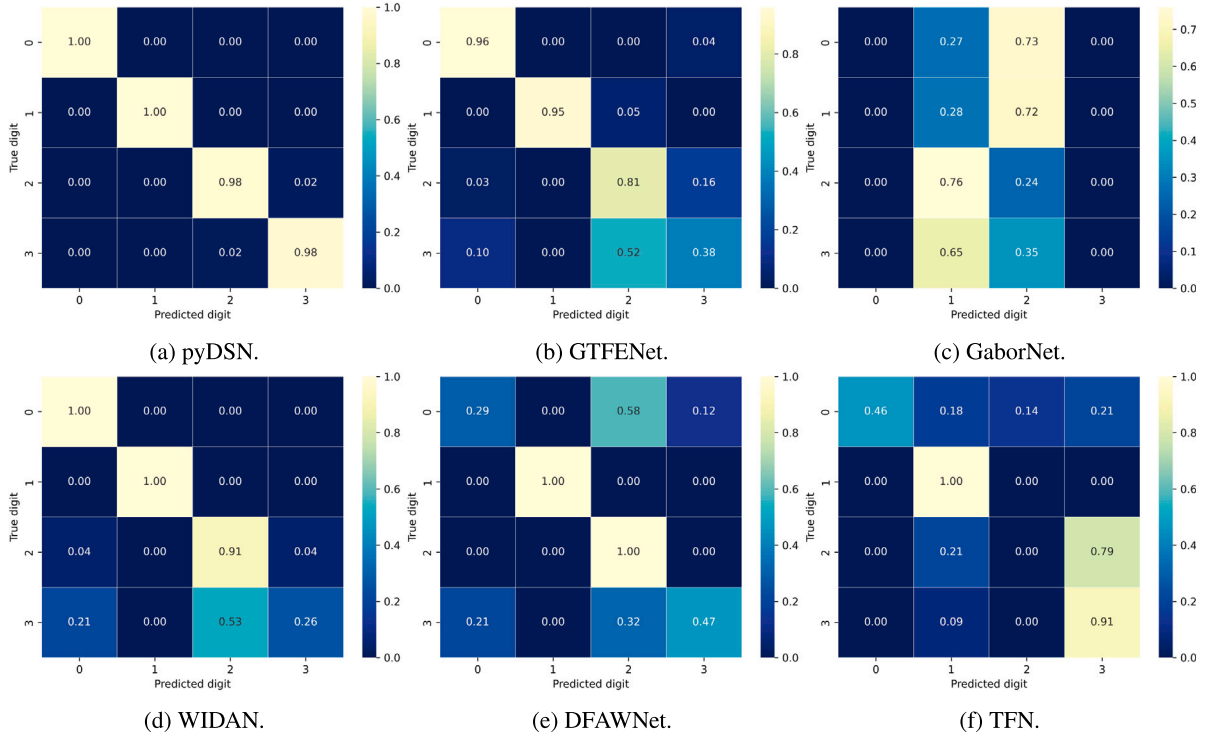


Fig. 10. The confusion matrix of six methods on $D_1 \rightarrow F_1$.

In contrast, in other tasks, SincNet and NTScatNet exhibit severe performance degradation, which underscores the stability, robustness, and best average precision of the proposed pyDSN.

As shown in Fig. 10, the performance of pyDSN and other methods across $D_1 \rightarrow F_1$ is evaluated through the confusion matrix. The confusion matrix can provide a comprehensive evaluation of different categories. Among these models, the most fundamental criterion is “Zero False Alarm” and “Zero Missed Alarm”, and it is clear that only pyDSN can satisfy this requirement. In tasks, identifying inner ring and ball faults proves to be challenging, while other categories can recognize over 99%. This finding serves as a valuable direction for future research endeavors.

5. The interpretability analysis of MDSTFT

The Spectra Quest Variable (SQV) speed dataset, with a sampling frequency of 25.6 kHz, is a publicly available dataset for variable speed analysis from Xi’an Jiaotong University. It is collected during a continuous speed change, encompassing a complete process of acceleration and deceleration. It begins from a stationary state, gradually increasing to 3000 rpm/min, where it stabilizes, and subsequently decelerates gradually back to zero. It serves as a good benchmark dataset and can be used as an example for research in this field. More details about the dataset can be found in [82].

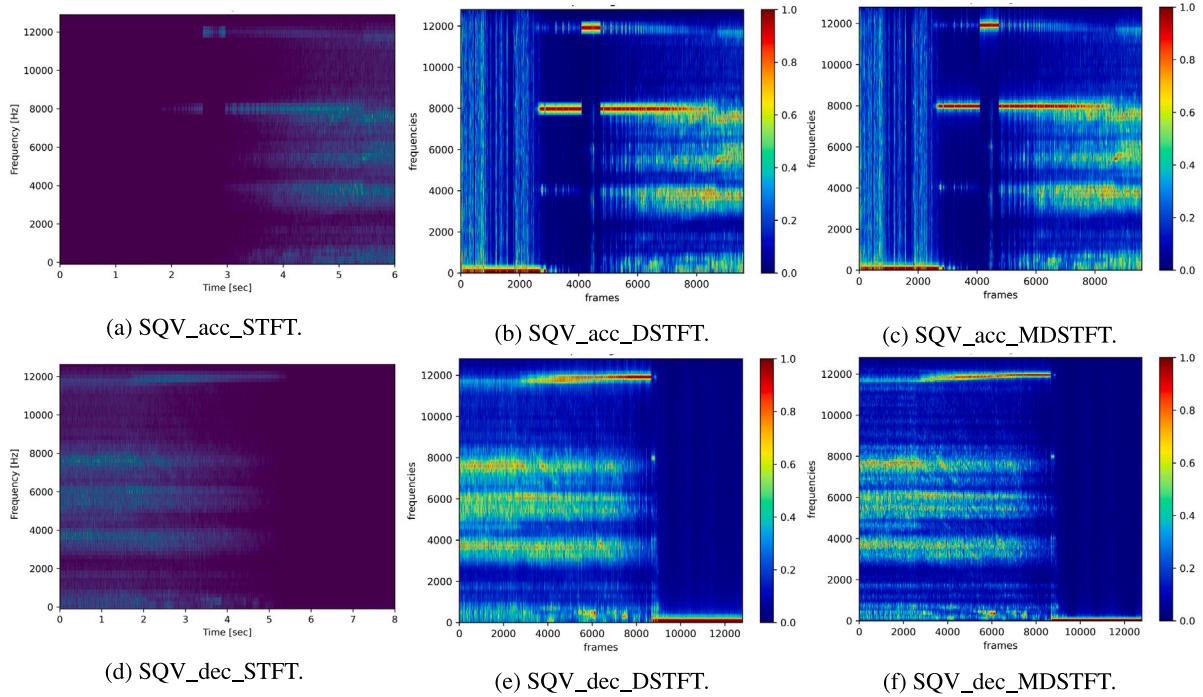


Fig. 11. The spectrograms on SQV dataset with STFT, DSTFT, and MDSTFT.

MDSTFT is considered inherently interpretable due to its unerring adherence to core principles of STFT and produces reliable results. Additionally, the learning rate of MDSTFT is 100.0, which is appreciably different from DAN. A larger learning rate serves to broaden the exploration spectrum of the time window and make it easier to find the optimal window lengths. In comparison to DSTFT, MDSTFT incorporates a signal modulation mechanism, which allows the raw signals to see a portion of the windows during FFT, thereby instigating different window lengths for capturing fault frequencies.

The rotational speed is continuously fluctuating. It is difficult to accurately measurement speed changes due to the high cost of tachometer. MDSTFT does not rely on the rotational speed to extract various fault features but achieves the same effect by using the time-varying window lengths under the guidance of the modulation mechanism. It does not require prior information such as rotational speed.

The calculation of bearing fault frequency necessitates prior knowledge of parameters including the rolling element diameter, ball pitch diameter, bearing contact angle, number of rolling elements, and rotational frequency. Not all devices can obtain these parameters, and the formulas are only applicable to constant rotational speed conditions. MDSTFT also do not need the information such as bearing dimensions.

Therefore, MDSTFT does not require prior knowledge of information such as speed and bearing size to obtain fault frequency.

The progressiveness of MDSTFT can be demonstrated from both qualitative and quantitative perspectives.

5.1. Qualitative perspective

As shown in Fig. 11, the spectrograms for the outer ring fault is depicted. The energy focusability achieved by vanilla STFT is deficient and poor. While upon employing differentiable STFT, the energy focusability has significantly improved, which demonstrates superior aggregation capabilities. It can be seen that the main fault frequencies in the speed fluctuation data are mainly localized around 0 Hz, 8000 Hz, and 12000 Hz during the acceleration process. The deceleration process tends to concentrate near 0 Hz, 4000 Hz, 6000 Hz, 8000 Hz, and 12000 Hz. In summary, the proposed MDSTFT method

Table 5
Rényi entropy and BSQ of SQV dataset.

		STFT	DSTFT	MDSTFT
SQV_acc	Rényi entropy	1.4880	1.4100	1.3980
	BSQ		0.5548	0.2515
SQV_dec	Rényi entropy	2.0830	1.9980	1.7410
	BSQ		0.5378	0.3220

exhibits the highest level of energy aggregation degree and more prominent fault characteristics.

5.2. Quantitative perspective

Furthermore, several methods are critically conducted quantitatively. Rényi entropy is extensively employed metric to evaluate energy aggregation, but it is incapable of back-propagate as it calculates the probability distribution density. Rényi entropy falls under the evaluation indicator, and BSQ pertains to the differentiable loss index. A decrease in both these indicators signifies higher spectrogram quality. From Table 5, the Rényi entropy of STFT is the highest, indicating that the poorer energy focusability, the more diminished the quality of spectrograms. The proposed MDSTFT has the lowest Rényi entropy (1.398 and 1.7411) and BSQ (0.2515 and 0.3220), whose quantitative values are lower than DSTFT, highlighting the necessity of introducing the mask modulation mechanism.

6. Conclusion

In this paper, we propose a physics-informed modulated differentiable STFT, which introduces modulation differentiable STFT to capture fault frequency information under time-varying speeds. Additionally, a loss function, termed as balanced spectrum, is devised to evaluate the quality of the generated spectrograms. By employing these two mechanisms, we have migrated the predicament of bearing transfer diagnosis in heavy haul freight trains under time-varying speed conditions, and have achieved the superior performance among analogous methods. For heavy haul freight equipment, and a future

exploration of signal processing informed neural networks integrated physical constraints is a promising avenue.

Nonetheless, several issues worth further consideration. Initially, the proposed pyDSN needs to train two distinct models, because acceleration and deceleration are the distinct nature. Potential future research directions can focus on the cross-machine transfer diagnosis at time-varying full speeds via one model. Secondly, differentiable signal processing has broad application potential, such as differentiable wavelets and differentiable mode decomposition. Essentially, the proposed paradigm is an end-to-end strategy that combines signal processing prior-empowered modules and classifiers without conflicts, yet the process from feature to decision in the classifier remains a black box. In the future, using explainable structures similar to capsule networks as classifiers is also an interesting topic. Lastly, this study is limited to closed set transfer diagnosis, and the performance with differentiable STFT in partial domain or open set domain adaptation warrants additional scrutiny.

CRedit authorship contribution statement

Chao He: Writing – original draft, Software, Validation, Visualization, Methodology, Investigation, Proof-reading. **Hongmei Shi:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Ruixin Li:** Writing – review & editing. **Jianbo Li:** Writing – review & editing, Supervision. **ZuJun Yu:** Supervision, Project administration.

Declaration of competing interest

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

Data availability

The data that has been used is confidential.

Acknowledgments

The authors are grateful for the supports of the National Natural Science Foundation of China (No. 52272429), and State Key Laboratory of Advanced Rail Autonomous Operation (Contract No. RAO2023ZZ003).

Su Li, thank you for bringing me luck and accompanying me all the way. Let us fall in love with each other forever, grow old together, and last forever.

References

- [1] S. Mykhalkiv, V. Ravlyuk, A. Khodakivskiy, V. Bereznyi, Identification of axle-box bearing faults of freight cars based on minimum entropy deconvolution and squared envelope spectra, *Int. J. Eng. Technol. (UAE)* 7 (2018) 167–173, <http://dx.doi.org/10.14419/ijet.v7i4.3.19729>.
- [2] Q. Wang, T. Gao, H. Tang, Y. Wang, Z. Chen, J. Wang, P. Wang, Q. He, A feature engineering framework for online fault diagnosis of freight train air brakes, *Measurement* 182 (2021) 109672, <http://dx.doi.org/10.1016/j.measurement.2021.109672>.
- [3] J. Si, H. Shi, J. Yang, Evaluation of Chinese freight train bearing condition based on spatiotemporal feature extraction, *Proc. Inst. Mech. Eng. F* 236 (9) (2022) 1047–1057, <http://dx.doi.org/10.1177/09544097211061944>.
- [4] J. Wu, Y. Li, L. Jia, G. An, Y.-F. Li, J. Antoni, G. Xin, Semi-supervised fault diagnosis of wheelset bearings in high-speed trains using autocorrelation and improved flow Gaussian mixture model, *Eng. Appl. Artif. Intell.* 132 (2024) 107861, <http://dx.doi.org/10.1016/j.engappai.2024.107861>.
- [5] Z. Liu, M. Zhang, F. Liu, B. Zhang, Multidimensional feature fusion and ensemble learning-based fault diagnosis for the braking system of heavy-haul train, *IEEE Trans. Ind. Inform.* 17 (1) (2021) 41–51, <http://dx.doi.org/10.1109/TII.2020.2979467>.
- [6] C. Li, K. Yang, H. Tang, P. Wang, J. Li, Q. He, Fault diagnosis for rolling bearings of a freight train under limited fault data: Few-shot learning method, *J. Transp. Eng. A* 147 (8) (2021) 04021041, <http://dx.doi.org/10.1061/JTEPBS.0000554>.
- [7] D. Liu, L. Cui, H. Wang, Rotating machinery fault diagnosis under time-varying speeds: A review, *IEEE Sens. J.* 23 (24) (2023) 29969–29990, <http://dx.doi.org/10.1109/JSEN.2023.3326112>.
- [8] Q. Qian, J. Luo, Y. Qin, Adaptive intermediate class-wise distribution alignment: A universal domain adaptation and generalization method for machine fault diagnosis, *IEEE Trans. Neural Netw. Learn. Syst.* (2024) 1–15, <http://dx.doi.org/10.1109/TNNLS.2024.3376449>.
- [9] T. Han, J. Tian, C. Chung, Y.M. Wei, Challenges and opportunities for battery health estimation: bridging laboratory research and real-world applications, *J. Electroanal. Chem.* 89 (2024) 434–436, <http://dx.doi.org/10.1016/j.jechem.2023.10.032>.
- [10] J. Yang, Z. Wang, Y. Guo, T. Gong, Z. Shan, A novel noise-aided fault feature extraction using stochastic resonance in a nonlinear system and its application, *IEEE Sens. J.* 24 (7) (2024) 11856–11866, <http://dx.doi.org/10.1109/JSEN.2024.3365105>.
- [11] K. Liang, M. Zhao, J. Lin, J. Jiao, C. Ding, Toothwise health monitoring of planetary gearbox under time-varying speed condition based on rotating encoder signal, *IEEE Trans. Ind. Electron.* 69 (6) (2022) 6267–6277, <http://dx.doi.org/10.1109/TIE.2021.3090713>.
- [12] X. Cheng, L. Yuan, Y. Lu, Y. Wang, N. Ding, Y. Gong, Fault identification of rolling bearings under linear varying speed based on the slope features of time-frequency ridges, *Mech. Syst. Signal Process.* 205 (2023) 110834, <http://dx.doi.org/10.1016/j.ymssp.2023.110834>.
- [13] X. Li, H. Shao, S. Lu, J. Xiang, B. Cai, Highly efficient fault diagnosis of rotating machinery under time-varying speeds using LSISMM and small infrared thermal images, *IEEE Trans. Syst. Man Cybern. Syst.* 52 (12) (2022) 7328–7340, <http://dx.doi.org/10.1109/TSMC.2022.3151185>.
- [14] H. Shao, W. Li, B. Cai, J. Wan, Y. Xiao, S. Yan, Dual-threshold attention-guided GAN and limited infrared thermal images for rotating machinery fault diagnosis under speed fluctuation, *IEEE Trans. Ind. Inform.* 19 (9) (2023) 9933–9942, <http://dx.doi.org/10.1109/TII.2022.3232766>.
- [15] Y. Xu, H. Tao, W. Li, Y. Zhong, CapsFormer: A novel bearing intelligent fault diagnosis framework with negligible speed change under small-sample conditions, *IEEE Trans. Instrum. Meas.* 72 (2023) 1–11, <http://dx.doi.org/10.1109/TIM.2023.3318693>.
- [16] M. Rao, M.J. Zuo, Z. Tian, A speed normalized autoencoder for rotating machinery fault detection under varying speed conditions, *Mech. Syst. Signal Process.* 189 (2023) 110109, <http://dx.doi.org/10.1016/j.ymssp.2023.110109>.
- [17] M. Rao, M.J. Zuo, Z. Tian, Speed adaptive gate: A novel auxiliary branch for enhancing deep learning-based rotating machinery fault classification under varying speed conditions, *Measurement* 217 (2023) 113016, <http://dx.doi.org/10.1016/j.measurement.2023.113016>.
- [18] Z. Shi, J. Chen, Y. Zi, Z. Chen, DecouplingNet: A stable knowledge distillation decoupling net for fault detection of rotating machines under varying speeds, *IEEE Trans. Neural Netw. Learn. Syst.* (2023) 1–15, <http://dx.doi.org/10.1109/TNNLS.2023.3258748>.
- [19] R. Li, J. Wu, Y. Li, Y. Cheng, PeriodNet: Noise-robust fault diagnosis method under varying speed conditions, *IEEE Trans. Neural Netw. Learn. Syst.* (2023) 1–15, <http://dx.doi.org/10.1109/TNNLS.2023.3274290>.
- [20] X. Yan, D. She, Y. Xu, Deep order-wavelet convolutional variational autoencoder for fault identification of rolling bearing under fluctuating speed conditions, *Expert Syst. Appl.* 216 (2023) 119479, <http://dx.doi.org/10.1016/j.eswa.2022.119479>.
- [21] Y. Chang, J. Chen, Q. Chen, S. Liu, Z. Zhou, CFs-focused intelligent diagnosis scheme via alternative kernels networks with soft squeeze-and-excitation attention for fast-precise fault detection under slow & sharp speed variations, *Knowl.-Based Syst.* 239 (2022) 108026, <http://dx.doi.org/10.1016/j.knosys.2021.108026>.
- [22] J. Yu, Y. Xu, K. Liu, Planetary gear fault diagnosis using stacked denoising autoencoder and gated recurrent unit neural network under noisy environment and time-varying rotational speed conditions, *Meas. Sci. Technol.* 30 (9) (2019) 095003, <http://dx.doi.org/10.1088/1361-6501/ab1da0>.
- [23] Z. An, S. Li, J. Wang, X. Jiang, A novel bearing intelligent fault diagnosis framework under time-varying working conditions using recurrent neural network, *ISA Trans.* 100 (2020) 155–170, <http://dx.doi.org/10.1016/j.isatra.2019.11.010>.
- [24] Z. Shi, X. Liu, J. Chen, Y. Zi, Z. Zhou, A multi-branch redundant adversarial net for intelligent fault diagnosis of multiple components under drastically variable speeds, *ISA Trans.* 129 (2022) 540–554, <http://dx.doi.org/10.1016/j.isatra.2022.01.011>.
- [25] H. Pu, K. Zhang, Y. An, Restricted sparse networks for rolling bearing fault diagnosis, *IEEE Trans. Ind. Inform.* 19 (11) (2023) 11139–11149, <http://dx.doi.org/10.1109/TII.2023.3243929>.
- [26] J. Guo, Q. He, F. Gu, DNOCNet: A novel end-to-end network for induction motor drive systems fault diagnosis under speed fluctuation condition, *IEEE Trans. Ind. Inform.* (2024) 1–10, <http://dx.doi.org/10.1109/TII.2024.3369239>.
- [27] Y. Dong, H. Jiang, R. Yao, M. Mu, Q. Yang, Rolling bearing intelligent fault diagnosis towards variable speed and imbalanced samples using multiscale dynamic supervised contrast learning, *Reliab. Eng. Syst. Saf.* 243 (2024) 109805, <http://dx.doi.org/10.1016/j.res.2023.109805>.

- [28] S. Yan, H. Shao, Y. Xiao, J. Zhou, Y. Xu, J. Wan, Semi-supervised fault diagnosis of machinery using LPS-DGAT under speed fluctuation and extremely low labeled rates, *Adv. Eng. Inform.* 53 (2022) 101648, <http://dx.doi.org/10.1016/j.aei.2022.101648>.
- [29] J.-X. Liao, H.-C. Dong, Z.-Q. Sun, J. Sun, S. Zhang, F.-L. Fan, Attention-embedded quadratic network (qttnet) for effective and interpretable bearing fault diagnosis, *IEEE Trans. Instrum. Meas.* 72 (2023) 1–13, <http://dx.doi.org/10.1109/TIM.2023.3259031>.
- [30] B. Liu, C. Yan, Y. Liu, Z. Wang, Y. Huang, L. Wu, Multiscale residual antinoise network via interpretable dynamic recalibration mechanism for rolling bearing fault diagnosis with few samples, *IEEE Sens. J.* 23 (24) (2023) 31425–31439, <http://dx.doi.org/10.1109/JSEN.2023.3328007>.
- [31] Z. Meng, H. He, W. Cao, J. Li, L. Cao, J. Fan, M. Zhu, F. Fan, A novel generation network using feature fusion and guided adversarial learning for fault diagnosis of rotating machinery, *Expert Syst. Appl.* 234 (2023) 121058, <http://dx.doi.org/10.1016/j.eswa.2023.121058>.
- [32] C. He, H. Shi, J. Li, IDSNet: A one-stage interpretable and differentiable STFT domain adaptation network for traction motor of high-speed trains cross-machine diagnosis, *Mech. Syst. Signal Process.* 205 (2023) 110846, <http://dx.doi.org/10.1016/j.ymssp.2023.110846>.
- [33] T. Han, W. Xie, Z. Pei, Semi-supervised adversarial discriminative learning approach for intelligent fault diagnosis of wind turbine, *Inform. Sci.* 648 (2023) 119496, <http://dx.doi.org/10.1016/j.ins.2023.119496>.
- [34] X. Li, Y. Wang, J. Yao, M. Li, Z. Gao, Multi-sensor fusion fault diagnosis method of wind turbine bearing based on adaptive convergent viewable neural networks, *Reliab. Eng. Syst. Saf.* 245 (2024) 109980, <http://dx.doi.org/10.1016/j.res.2024.109980>.
- [35] D. Sun, Z. Meng, Y. Guan, J. Liu, W. Cao, F. Fan, Intelligent fault diagnosis scheme for rolling bearing based on domain adaptation in one dimensional feature matching, *Appl. Soft Comput.* 146 (2023) 110669, <http://dx.doi.org/10.1016/j.asoc.2023.110669>.
- [36] P. Liang, B. Wang, G. Jiang, N. Li, L. Zhang, Unsupervised fault diagnosis of wind turbine bearing via a deep residual deformable convolution network based on subdomain adaptation under time-varying speeds, *Eng. Appl. Artif. Intell.* 118 (2023) 105656, <http://dx.doi.org/10.1016/j.engappai.2022.105656>.
- [37] P. Liang, L. Xu, H. Shuai, X. Yuan, B. Wang, L. Zhang, Semisupervised subdomain adaptation graph convolutional network for fault transfer diagnosis of rotating machinery under time-varying speeds, *IEEE/ASME Trans. Mechatronics* 29 (1) (2024) 730–741, <http://dx.doi.org/10.1109/TMECH.2023.3292969>.
- [38] H. Cao, H. Shao, X. Zhong, Q. Deng, X. Yang, J. Xuan, Unsupervised domain-share CNN for machine fault transfer diagnosis from steady speeds to time-varying speeds, *J. Manuf. Syst.* 62 (2022) 186–198, <http://dx.doi.org/10.1016/j.jmsy.2021.11.016>.
- [39] J. Si, H. Shi, T. Han, J. Chen, C. Zheng, Learn generalized features via multi-source domain adaptation: Intelligent diagnosis under variable/constant machine conditions, *IEEE Sens. J.* 22 (1) (2022) 510–519, <http://dx.doi.org/10.1109/JSEN.2021.3126864>.
- [40] X. Zhao, X. Zhu, J. Yao, W. Deng, Y. Cao, P. Ding, M. Jia, H. Shao, Intelligent health assessment of aviation bearing based on deep transfer graph convolutional networks under large speed fluctuations, *Sensors* 23 (9) (2023) 4379, <http://dx.doi.org/10.3390/s23094379>.
- [41] Z. Shi, J. Chen, X. Zhang, Y. Zi, C. Li, J. Chen, A reliable feature-assisted contrastive generalization net for intelligent fault diagnosis under unseen machines and working conditions, *Mech. Syst. Signal Process.* 188 (2023) 110011, <http://dx.doi.org/10.1016/j.ymssp.2022.110011>.
- [42] H. Zhou, X. Huang, G. Wen, S. Dong, Z. Lei, P. Zhang, X. Chen, Convolution enabled transformer via random contrastive regularization for rotating machinery diagnosis under time-varying working conditions, *Mech. Syst. Signal Process.* 173 (2022) 109050, <http://dx.doi.org/10.1016/j.ymssp.2022.109050>.
- [43] D. Gao, K. Huang, Y. Zhu, L. Zhu, K. Yan, Z. Ren, C.G. Soares, Semi-supervised small sample fault diagnosis under a wide range of speed variation conditions based on uncertainty analysis, *Reliab. Eng. Syst. Saf.* 242 (2023) 109746, <http://dx.doi.org/10.1016/j.res.2023.109746>.
- [44] Z. Xu, G. Tang, B. Pang, X. Qi, Rolling bearing fault diagnosis under time-varying speeds based on time-characteristic order spectrum and multi-scale domain adaptation network, *Meas. Sci. Technol.* 34 (12) (2023) 125118, <http://dx.doi.org/10.1088/1361-6501/acf332>.
- [45] F. Lu, Q. Tong, Z. Feng, Q. Wan, Unbalanced bearing fault diagnosis under various speeds based on spectrum alignment and deep transfer convolution neural network, *IEEE Trans. Ind. Inform.* 19 (7) (2023) 8295–8306, <http://dx.doi.org/10.1109/TII.2022.3217541>.
- [46] B. Pang, Q. Liu, Z. Sun, Z. Xu, Z. Hao, Time-frequency supervised contrastive learning via pseudo-labeling: An unsupervised domain adaptation network for rolling bearing fault diagnosis under time-varying speeds, *Adv. Eng. Inform.* 59 (2024) 102304, <http://dx.doi.org/10.1016/j.aei.2023.102304>.
- [47] J. Luo, H. Shao, J. Lin, B. Liu, Meta-learning with elastic prototypical network for fault transfer diagnosis of bearings under unstable speeds, *Reliab. Eng. Syst. Saf.* 245 (2024) 110001, <http://dx.doi.org/10.1016/j.res.2024.110001>.
- [48] C. Wang, Y. Sun, X. Wang, Image deep learning in fault diagnosis of mechanical equipment, *J. Intell. Manuf.* (2023) 1–41, <http://dx.doi.org/10.1007/s10845-023-02176-3>.
- [49] A. Zhao, K. Subramani, P. Smaragdakis, Optimizing short-time Fourier transform parameters via gradient descent, in: *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2021, Toronto, on, Canada, June 6–11, 2021*, 2021, pp. 736–740, <http://dx.doi.org/10.1109/ICASSP39728.2021.9413704>.
- [50] Y. Sun, W. Wang, Role of image feature enhancement in intelligent fault diagnosis for mechanical equipment: A review, *Eng. Fail. Anal.* 156 (2023) 107815, <http://dx.doi.org/10.1016/j.engfailanal.2023.107815>.
- [51] Y. Xu, S. Kohtz, J. Boakye, P. Gardoni, P. Wang, Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges, *Reliab. Eng. Syst. Saf.* 230 (2023) 108900, <http://dx.doi.org/10.1016/j.res.2022.108900>.
- [52] R. Yan, Z. Shang, Z. Wang, W. Xu, Z. Zhao, S. Wang, X. Chen, Challenges and opportunities of XAI in industrial intelligent diagnosis: Priori-empowered, *J. Mech. Eng.* 60 (2024) 1–20, <http://kns.cnki.net/kcms/detail/11.2187.TH.20240220.1637.010.html>.
- [53] J.-X. Liao, C. He, J. Li, J. Sun, S. Zhang, X. Zhang, Classifier-guided neural blind deconvolution: a physics-informed denoising module for bearing fault diagnosis under heavy noise, *arXiv preprint arXiv:2404.15341* (2024) <http://dx.doi.org/10.48550/arXiv.2404.15341>.
- [54] T. Yan, Y. Fu, M. Lu, Z. Li, C. Shen, D. Wang, Integration of a novel knowledge-guided loss function with an architecturally explainable network for machine degradation modeling, *IEEE Trans. Instrum. Meas.* 71 (2022) 1–12, <http://dx.doi.org/10.1109/TIM.2022.3193196>.
- [55] M. Russell, P. Wang, Physics-informed deep learning for signal compression and reconstruction of big data in industrial condition monitoring, *Mech. Syst. Signal Process.* 168 (2022) 108709, <http://dx.doi.org/10.1016/j.ymssp.2021.108709>.
- [56] X. Chen, M. Ma, Z. Zhao, Z. Zhai, Z. Mao, Physics-informed deep neural network for bearing prognosis with multisensory signals, *J. Dyn. Monit. Diagn.* 1 (4) (2022) 200–207, <http://dx.doi.org/10.37965/jdmd.2022.54>.
- [57] B. Freeman, Y. Tang, Y. Huang, J. VanZwieten, Physics-informed turbulence intensity infusion: A new hybrid approach for marine current turbine rotor blade fault detection, *Ocean Eng.* 254 (2022) 111299, <http://dx.doi.org/10.1016/j.oceaneng.2022.111299>.
- [58] S. Xu, H.Y. Noh, PhyMDAN: Physics-informed knowledge transfer between buildings for seismic damage diagnosis through adversarial learning, *Mech. Syst. Signal Process.* 151 (2021) 107374, <http://dx.doi.org/10.1016/j.ymssp.2020.107374>.
- [59] Z. Zhang, C. Sun, Structural damage identification via physics-guided machine learning: a methodology integrating pattern recognition with finite element model updating, *Struct. Health Monit.* 20 (4) (2021) 1675–1688, <http://dx.doi.org/10.1177/1475921720927488>.
- [60] A. Jablonski, K. Dziedzic, Intelligent spectrogram—A tool for analysis of complex non-stationary signals, *Mech. Syst. Signal Process.* 167 (2022) 108554, <http://dx.doi.org/10.1016/j.ymssp.2021.108554>.
- [61] A. Rényi, On measures of entropy and information, in: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics, Vol. 4, 1961*, pp. 547–562, https://static.renyi.hu/renyi_cikk/1961_on_measures_of_entropy_and_information.pdf.
- [62] D. Marx, K. Gryllias, Differentiable short-time Fourier transform window length selection driven by cyclo-stationarity, in: *Annual Conference of the PHM Society, Vol. 15, 2023*, <http://dx.doi.org/10.36001/phmconf.2023.v15i1.3566>.
- [63] M. Leiber, Y. Marnissi, A. Barrau, M.E. Badaoui, Differentiable adaptive short-time Fourier transform with respect to the window length, in: *IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2023, Rhodes Island, Greece, June 4–10, 2023*, 2023, pp. 1–5, <http://dx.doi.org/10.1109/ICASSP49357.2023.10095245>.
- [64] H. Huang, N. Baddour, Bearing vibration data collected under time-varying rotational speed conditions, *Data Brief* 21 (2018) 1745–1749, <http://dx.doi.org/10.1016/j.dib.2018.11.019>.
- [65] M. Tan, Q.V. Le, EfficientNetV2: Smaller models and faster training, in: *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18–24 July 2021, Virtual Event*, in: *Proceedings of Machine Learning Research, Vol. 139, PMLR, 2021*, pp. 10096–10106, <http://proceedings.mlr.press/v139/tan21a.html>.
- [66] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016*, pp. 770–778, <http://dx.doi.org/10.1109/CVPR.2016.90>.
- [67] G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21–26, 2017*, 2017, pp. 2261–2269, <http://dx.doi.org/10.1109/CVPR.2017.243>.
- [68] W. Zhang, G. Peng, C. Li, Y. Chen, Z. Zhang, A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals, *Sensors* 17 (2) (2017) 425, <http://dx.doi.org/10.3390/s17020425>.
- [69] M. Zhao, S. Zhong, X. Fu, B. Tang, M. Pecht, Deep residual shrinkage networks for fault diagnosis, *IEEE Trans. Ind. Inform.* 16 (7) (2020) 4681–4690, <http://dx.doi.org/10.1109/TII.2019.2943898>.

- [70] Y. Xiao, H. Shao, S. Han, Z. Huo, J. Wan, Novel joint transfer network for unsupervised bearing fault diagnosis from simulation domain to experimental domain, *IEEE/ASME Mech. Syst. Signal Process.* 27 (6) (2022) 5254–5263, <http://dx.doi.org/10.1109/TMECH.2022.3177174>.
- [71] Q. Qian, Y. Wang, T. Zhang, Y. Qin, Maximum mean square discrepancy: a new discrepancy representation metric for mechanical fault transfer diagnosis, *Knowl.-Based Syst.* 276 (2023) 110748, <http://dx.doi.org/10.1016/j.knosys.2023.110748>.
- [72] Q. Qian, Y. Qin, J. Luo, Y. Wang, F. Wu, Deep discriminative transfer learning network for cross-machine fault diagnosis, *Mech. Syst. Signal Process.* 186 (2023) 109884, <http://dx.doi.org/10.1016/j.ymssp.2022.109884>.
- [73] M. Ravanelli, Y. Bengio, Speaker recognition from raw waveform with SincNet, in: 2018 IEEE Spoken Language Technology Workshop, SLT, 2018, pp. 1021–1028, <http://dx.doi.org/10.1109/SLT.2018.8639585>.
- [74] R. Liu, X. Ding, Q. Wu, Q. He, Y. Shao, An interpretable multiplication-convolution network for equipment intelligent edge diagnosis, *IEEE Trans. Syst. Man Cybern. Syst.* (2024) 1–12, <http://dx.doi.org/10.1109/TSMC.2023.3346398>.
- [75] C. Liu, X. Ma, T. Han, X. Shi, C. Qin, S. Hu, NtScatNet: An interpretable convolutional neural network for domain generalization diagnosis across different transmission paths, *Measurement* 204 (2022) 112041, <http://dx.doi.org/10.1016/j.measurement.2022.112041>.
- [76] Z. Shang, Z. Zhao, R. Yan, Denoising fault-aware wavelet network: A signal processing informed neural network for fault diagnosis, *Chin. J. Mech. Eng.* 36 (1) (2023) 9, <http://dx.doi.org/10.1186/s10033-023-00838-0>.
- [77] L. Jia, T.W. Chow, Y. Yuan, GTFE-net: A gramian time frequency enhancement CNN for bearing fault diagnosis, *Eng. Appl. Artif. Intell.* 119 (2023) 105794, <http://dx.doi.org/10.1016/j.engappai.2022.105794>.
- [78] A. Alekseev, A. Bobe, GaborNet: Gabor filters with learnable parameters in deep convolutional neural network, in: 2019 International Conference on Engineering and Telecommunication, EnT, 2019, pp. 1–4, <http://dx.doi.org/10.1109/EnT47717.2019.9030571>.
- [79] C. He, H. Shi, X. Liu, J. Li, Interpretable physics-informed domain adaptation paradigm for cross-machine transfer diagnosis, *Knowl.-Based Syst.* 288 (2024) 111499, <http://dx.doi.org/10.1016/j.knosys.2024.111499>.
- [80] K. Yue, J. Li, J. Chen, R. Huang, W. Li, Multiscale wavelet prototypical network for cross-component few-shot intelligent fault diagnosis, *IEEE Trans. Instrum. Meas.* 72 (2023) 1–11, <http://dx.doi.org/10.1109/TIM.2022.3230480>.
- [81] Q. Chen, X. Dong, G. Tu, D. Wang, C. Cheng, B. Zhao, Z. Peng, TFN: An interpretable neural network with time-frequency transform embedded for intelligent fault diagnosis, *Mech. Syst. Signal Process.* 207 (2024) 110952, <http://dx.doi.org/10.1016/j.ymssp.2023.110952>.
- [82] Z. Shi, J. Chen, Y. Zi, Z. Zhou, A novel multitask adversarial network via redundant lifting for multicomponent intelligent fault detection under sharp speed variation, *IEEE Trans. Instrum. Meas.* 70 (2021) 1–10, <http://dx.doi.org/10.1109/TIM.2021.3055821>.