Diffusion Bridge-Based Gradual Domain Adaptation for Fault Diagnosis: Concept, Algorithms and Applications

Chuang Wang, Zidong Wang, *Fellow*, *IEEE*, and Hongli Dong, *Senior Member*, *IEEE*

***Abstract*—Gradual domain adaptation (GDA) has been recognized as an effective strategy for alleviating domain shift by decomposing large distribution discrepancies into a sequence of progressive adaptation steps, thereby enabling smoother knowledge transfer from source to target domains. However, existing GDA methods still face two major challenges: the construction of intermediate domains is often heuristic and poorly constrained, which may result in unstable or unreliable adaptation trajectories; meanwhile, semantic information can be progressively distorted during gradual adaptation, leading to distribution misalignment and performance degradation. To address these issues, a diffusion bridge-based data adaptation (DBDA) framework is proposed in this paper. Specifically, a forward-reverse diffusion bridge is constructed to progressively map target-domain features toward the source domain in a scalable and flexible manner. In addition, a knowledge retention mechanism is incorporated to preserve semantic consistency throughout the adaptation trajectory. Furthermore, a history-aware boundary extension loss is developed to enhance decision robustness under out-of-distribution conditions. Extensive experiments on cross-domain pipeline fault diagnosis datasets, as well as the public Sleep-EDF dataset, demonstrate that the proposed DBDA framework consistently outperforms existing GDA approaches in terms of accuracy, generalization capability, and stability.**

***Index Terms*—Gradual domain adaptation; Diffusion bridge; Cross-domain adaptation; Fault diagnosis; Knowledge retention; Diffusion models**

I. Introduction

Deep neural networks (DNNs) have demonstrated significant success in a wide range of applications, including computer vision, speech recognition, and industrial fault diagnosis, owing to their strong nonlinear modeling capability and powerful feature extraction ability [11], [22], [23], [32]. Nevertheless, the performance of DNNs has been shown to depend heavily on the availability of large-scale annotated datasets, the acquisition of which is often costly, labor-intensive, and, in some cases, infeasible due to privacy protection constraints [21], [42], [44], [45]. To alleviate these limitations, transfer learning has been widely adopted, where labeled data from a related source domain are exploited to facilitate learning in an unlabeled target domain [27], [35]. However, substantial discrepancies between source and target data distributions typically result in pronounced performance degradation on the target domain, a phenomenon commonly referred to as domain shift, as illustrated in Fig. 1(a). A representative example arises in natural gas pipeline fault diagnosis, where monitoring signals are generated under diverse operating conditions, causing models trained under one condition to generalize poorly to others due to distributional differences. Consequently, bridging the gap between the labeled source domain and the unlabeled target domain has become the central objective of cross-domain diagnostic model research.

Domain gaps between source and target domains are generally manifested in the form of covariate shift, label shift, or concept drift, all of which violate the assumption of independent and identically distributed data and thus impair model generalization. These challenges have motivated extensive research on domain adaptation (DA) methods, which typically employ transfer models equipped with distribution alignment regularizers or adversarial domain discriminators to reduce inter-domain discrepancies [9], [31], [34]. For instance, Su et al. [31] proposed a DA approach based on inter-domain decision discrepancy minimization to align source and target distributions. Similarly, Chen et al. [9] developed a dual adversarial-guided unsupervised multi-domain adaptation network to integrate multi-source knowledge for collaborative diagnosis. However, most existing DA methods perform domain alignment in a single step, as depicted in Fig. 1(b), where the decision boundary is adjusted only once within a shared high-dimensional feature space. Although such one-step adaptation can partially alleviate distribution discrepancies, it often fails to capture complex or gradual domain transitions, thereby leading to unstable or suboptimal decision boundary transformations.

Recently, increasing attention has been devoted to alleviating domain discrepancies through the introduction of intermediate domain embeddings. A representative paradigm, referred to as gradual domain adaptation (GDA), bridges large distribution gaps by constructing a sequence of proxy distributions, each exhibiting sufficiently small shifts relative

This work was supported in part by the National Natural Science Foundation of China under Grant 62403119, the Postdoctoral Fellowship Program of China Postdoctoral Science Foundation under Grant GZB20240136, the China Postdoctoral Foundation under Grant 2024MD753911, the Heilongjiang Provincial Postdoctoral Science Foundation of China under Grant LBH-TZ2405, the Engineering and Physical Sciences Research Council (EPSRC) of the UK, the Royal Society of the UK, and the Alexander von Humboldt Foundation of Germany. *(Corresponding author: Hongli Dong.)*

Chuang Wang and Hongli Dong are with the Artificial Intelligence Energy Research Institute, Northeast Petroleum University, Daqing 163318, China, also with the Heilongjiang Provincial Key Laboratory of Networking and Intelligent Control, Northeast Petroleum University, Daqing 163318, China, and also with Sanya Offshore Oil & Gas Research Institute, Northeast Petroleum University, Sanya 572025, China. (Emails: wangchuang64@126.com, shiningdhl@vip.126.com)

Zidong Wang is with the Department of Computer Science, Brunel University London, Uxbridge, Middlesex, UB8 3PH, United Kingdom. (Email: Zidong.Wang@brunel.ac.uk)

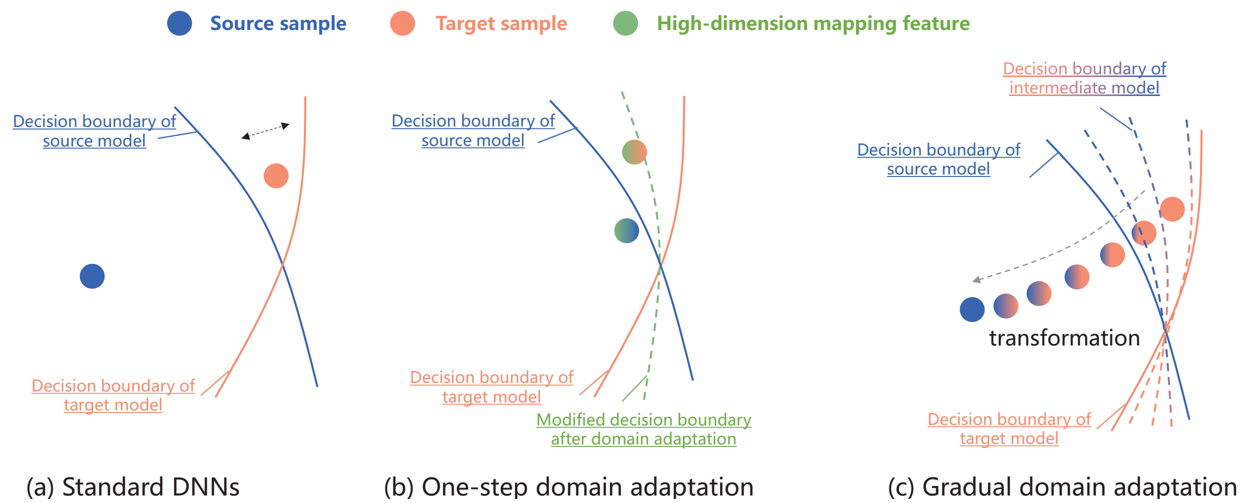


Fig. 1: Illustration of model decision boundaries under distribution discrepancies for standard DNNs, one-step DA, and GDA. (a) Standard DNNs: the source-trained model fails to generalize due to unaligned decision boundaries. (b) One-step DA: single-step alignment in a shared high-dimensional feature space reduces distribution discrepancies but may cause unstable decision boundary. (c) Gradual DA: progressive alignment through intermediate domains enables smooth boundary transformation and stable cross-domain generalization.

to its neighbors. For example, Chen et al. [5] proposed an intermediate domain labeler that discovers coarse intermediate domains using a progressive domain discriminator. Moreover, Yang et al. [43] introduced a federated semi-supervised transfer fault diagnosis method that generates intermediate distributions and minimizes discrepancies to achieve targeted transfer learning. Compared with traditional DA, the advantage of GDA lies in decomposing large-scale distribution shifts into a series of smaller and more tractable transitions, thereby avoiding excessive reliance on a single-step alignment strategy that is often difficult to realize in practice. However, collecting real-world datasets to serve as proxy distributions remains challenging, since target domains in many industrial scenarios often exhibit dynamic and discontinuous characteristics. In addition, the generation of reliable pseudo-samples is hindered by issues related to validity, interpretability, and the absence of fair evaluation criteria.

To address the aforementioned limitations, a **D**iffusion **B**ridge-based **D**ata **A**daptation (DBDA) model is introduced in this paper. The proposed DBDA progressively transforms the data distribution of the target domain into that of the source domain while enabling smooth evolution of the decision boundary across intermediate domains. The primary objective is to correct biased domain modeling along the distributional trajectory from the target domain to the source domain, while preserving valuable semantic information from the target domain. In this way, robust model evaluation can be conducted with flexible objective functions, without being constrained by specific cross-domain adaptation designs. Specifically, DBDA reformulates conventional DA as a data adaptation framework by integrating three key modules: (1) concatenated Target-to-Source Diffusion Bridges (TSDB), composed of a forward bridge trained on the target domain and a reverse bridge trained on the source domain, which encode target samples into a latent domain and decode distribution trajectories through an inverse diffusion process; (2) a knowledge retention mechanism that preserves dominant semantic information from the target domain along intermediate distribution trajectories to enhance generalization performance; and (3) a history-aware boundary extension loss that improves model robustness against the diversity of operating conditions introduced by the transformed samples.

The DBDA framework serves as a plug-and-play solution for cross-domain fault diagnosis and exhibits three salient properties: (1) *Scalable*: it is capable of accommodating various data modalities, including time series and images, and can be applied under both source-free and standard data adaptation settings; (2) *Flexible*: it establishes a transformation channel from the target domain to the source domain by decoupling and recombining two independently trained diffusion models, allowing domain-specific reverse bridges to be reused without retraining; and (3) *Simple*: it mitigates domain gaps directly at the data level, thereby avoiding excessive reliance on complex domain-specific matching modules that may overlook the requirements of downstream tasks.

The main contributions of this paper are summarized as follows.

1) A DBDA framework is proposed to address the challenges of GDA, which integrates effective domain transformation, refined knowledge retention, and robust decision boundary extension within a unified data adaptation paradigm.
2) A TSDB model is developed to connect source and target domains through a sequence of naturally evolving intermediate distributions, where diagnostic errors are progressively minimized by expanding the decision boundary during training to enable smooth and continuous knowledge transfer.
3) Extensive experiments conducted on pipeline datasets and other real-world datasets demonstrate the effectiveness and practicality of the proposed DBDA framework in comparison with existing GDA algorithms, highlight-

ing the potential of diffusion-based data adaptation for enhancing safety and operational maintenance performance.

The remainder of this paper is organized as follows. In Section II, related works on unsupervised domain adaptation, gradual domain adaptation, and diffusion models are reviewed. Section III presents the detailed implementation of each module in the DBDA framework. Experimental results and corresponding analyses are reported in Section IV. Finally, concluding remarks are provided in Section V.

II. Related Work

A. Unsupervised Domain Adaptation (UDA)

The fundamental assumption in classical deep learning that training and test data are drawn from the same distribution has often been violated in many real-world applications [8], [23]. To address this issue, unsupervised domain adaptation (UDA) methods have been developed to enable models to perform effectively in a target domain by leveraging knowledge from semantically related but distributionally different source domains. Existing UDA approaches can be broadly categorized into two classes: (1) metric learning-based distribution alignment methods and (2) adversarial learning-based distribution discrimination methods. For example, Wang et al. [36] proposed a prototype-assisted contrastive adversarial network for weak-shot learning, in which a contrastive adversarial discrepancy strategy is employed to simultaneously reduce domain gaps at both global and local levels. In addition, Liu et al. [18] developed a cross-machine deep subdomain adaptation network for wind turbine fault diagnosis under multiple operating conditions, where joint statistical and geometric metrics are constructed to enhance domain confusion.

Another research direction in UDA focuses on adversarial learning-based knowledge transfer. For instance, Su et al. [30] established a dual adversarial adaptation framework in which two domain discriminators participate in a two-player min-max game with the target model to learn reliable domain-invariant features. Moreover, Wang et al. [39] designed a dynamic collaborative adversarial domain adaptation network for unsupervised rotating machinery fault diagnosis, where a dual-system collaborative adversarial mechanism is introduced to dynamically adjust the network training architecture. Nevertheless, existing UDA methods, whether explicitly or implicitly designed, still encounter significant difficulties in constructing informative mapping strategies or effective regularization terms capable of simultaneously capturing domain-shared and domain-specific features across the entire domain. These challenges are particularly pronounced when balancing feature significance and fine-tuning hyperparameters.

Whether based on metric learning or adversarial learning, existing methods essentially reduce the domain gap by aligning the marginal or conditional distributions between the source and target domains in a single step, thereby learning transferable representations and achieving cross-domain generalization. However, when the distribution discrepancy is substantial, relying solely on one-step alignment in the feature space often leads to suboptimal performance. Consequently, recent studies have begun to explore transfer learning at the data level, namely, data adaptation. By means of domain translation, sample reweighting, and sample generation, data adaptation directly adjusts the distribution of the training data to better meet the model's generalization requirements. Existing data adaptation studies can generally be categorized into three types of paradigms: (1) **Domain translation.** Mapping the target domain to the style of source samples, or vice versa, to achieve cross-domain knowledge transfer by training a robust classifier without alignment regularization; (2) **Sample reweighting.** Assigning larger weights during training to samples that are more relevant to the target domain or harder to align, so as to emphasize the contribution to shaping the decision boundary; (3) **Sample generation.** Synthesizing samples that exhibit characteristics of the target distribution using generative models, thereby increasing the coverage of target-related training data and strengthening the model's ability to capture potential target patterns. In summary, the proposed DBDA essentially belongs to the domain translation branch of data adaptation.

B. Gradual Domain Adaptation

GDA exploits a sequence of continuous unlabeled samples as intermediate domains, allowing the source classifier to adapt to the target domain through self-training rather than explicit feature alignment [48]. Existing studies have investigated gradual adaptation mainly from two perspectives: (1) generating intermediate domains using generative models, and (2) utilizing real-world datasets with semantic relevance as intermediate domains.

For the first category, Chen et al. [4] proposed a generative appearance replay method for continual domain adaptation, where a segmentation model is sequentially adapted to new domains using unlabeled data. Similarly, Zhang et al. [47] introduced an approach for constructing evolving intermediate domains by progressively decreasing the proportion of source data while increasing the proportion of target data, with a sampling strategy designed to minimize domain distance between consecutive distributions.

For the second category, Wu et al. [41] proposed a gradual conditional domain adversarial network for bearing fault diagnosis, in which real-world intermediate datasets are employed to train a sequence of domain adaptation models, thereby bridging the source and target domains under varying fault categories and operating conditions. Nevertheless, the construction or selection of appropriate intermediate domains remains an open problem, since both generative and real-data-based approaches face challenges related to reliability, domain coverage, and the lack of unified evaluation criteria.

Different from the above two categories, the proposed DBDA aims to perform a controllable and progressive modeling of the cross-domain distribution evolution via the diffusion model. The generation trajectory is explicitly constrained to a feasible evolution path jointly defined by the source and target domains, rather than synthesizing pseudo intermediate domain that do not exist in real-world scenarios. On the other hand, DBDA ultimately aims to establish explicit target-to-source domain mapping relationships, ensuring that target samples retain semantically consistent and discriminative representations within the source domain. This enables the direct reuse of the discriminative capability of the source diagnostic model without the need for incremental alignment through intermediate domains.

C. Diffusion Model-Based Data Adaptation

Diffusion models represent a state-of-the-art class of generative modeling techniques in computer vision, characterized by the capability to generate diverse samples through iterative refinement along a Markov chain. In essence, diffusion models are trained to denoise samples and generate data from a simple Gaussian distribution via gradient-based optimization. For example, Chen et al. [6] proposed DiffusionDet, in which object detection is formulated as a denoising diffusion process that progressively transforms noisy bounding boxes into accurate object predictions.

More recently, diffusion models incorporating various bridge formulations, such as the Brownian bridge, Doob's h-transform, and Schrödinger bridge, have been explored to achieve image translation via optimal transport between arbitrary probability distributions. For instance, Li et al. [15] developed an image-to-image translation method based on the Brownian bridge diffusion model, where image translation is modeled as a stochastic Brownian bridge process. In parallel, diffusion models have also been introduced for domain adaptation and related transfer settings. Zhang et al. [46] proposed a domain-guided conditional diffusion model for UDA by generating domain-aligned samples, and Gao et al. [12] introduced a diffusion-driven test-time adaptation method that projects corrupted inputs back toward the source distribution using a source diffusion prior. In this paper, the primary objective of DBDA is conceptually related to such diffusion-based translation and adaptation paradigms; however, the focus is placed on GDA for time series fault diagnosis, where the goal is to transform temporal characteristics and conventional features without altering discriminative fault patterns. Motivated by this perspective, the proposed DBDA constructs a progressive target-source conversion path by concatenating two independent diffusion models via a bridge model, and further integrates trajectory-level semantic retention with the downstream boundary regularization mechanism to achieve robust fault diagnosis.

III. Diffusion Bridge-Based Data Adaptation

A. Notations and Overview

Following the notations adopted in the previous study [37], the source domain is denoted by $\mathcal{S}$ and the target domain by $\mathcal{T}$. The source domain and the target domain consist of instance datasets $\mathbf{x}^{(s)}$ and $\mathbf{x}^{(t)}$, which are drawn from d-dimensional marginal distributions $p^{(s)}(\mathbf{x})$ and $p^{(t)}(\mathbf{x})$, respectively, together with the corresponding label sets $\mathbf{y}^{(s)}$ and $\mathbf{y}^{(t)}$. A detailed description of the notations used throughout this paper is summarized in Table I, and an overview of the proposed DBDA framework is illustrated in Fig. 2.

TABLE I: Detailed description of the notations mentioned in this paper.

Notations	Definitions
$\mathbf{x}^{(s)}$	Source sample
$\hat{\mathbf{x}}^{(s)}$	Transformed sample
$\mathbf{x}^{(i)}$	Intermediate sample related to the TSDB
$\mathbf{x}^{(t)}$	Target sample
$p^{(s)}(\mathbf{x})$	Source data distribution
$p^{(t)}(\mathbf{x})$	Target data distribution
$\mathbf{y}^{(s)}$	Source label
$\mathbf{y}^{(t)}$	Target label
$\mathcal{S}$	Source domain
$\mathcal{T}$	Target domain
$\mathcal{I}$	Intermediate domain
$\mathcal{F}$	Fusion domain consisting of $\mathbf{x}^{(s)}$ and $\hat{\mathbf{x}}^{(s)}$
$t \in [0, T]$	Time steps associated with the diffusion process
p_0	Initial distribution (source or target domain)
$\mathbf{x}_0$	Sample from initial distribution
$p_{t\sim T}$	Intermediate distribution obtained by adding Gaussian noise

B. Diffusion Process

Given a time series sampled from the initial distribution $\mathbf{x}_0 \sim p_0$, the forward process progressively adds Gaussian noise to the sequence over T timesteps, thereby yielding a prior distribution p_T. This diffusion process can generally be described as the solution to the following Stochastic Differential Equations (SDEs) [14], [33] for any time step $t \in [0, T]$:

$$\mathrm{d}\mathbf{x}_t = \mathbf{f}(t)\mathbf{x}_t \mathrm{d}t + g(t)\mathrm{d}\boldsymbol{w}_t, \tag{1}$$

where $\boldsymbol{w}_t$ is a wiener process, $\mathbf{f}(t)$ is the vector-valued drift coefficient of $\mathbf{x}_t$, and $g(t)$ is a scalar diffusion coefficient of $\mathbf{x}_t$. By following the forward process in time, the final variable $\mathbf{x}_T$ is constrained to obey the prior distribution $p_{\text{prior}}(\mathbf{x})$. Correspondingly, a backward SDE from time T to 0 can be defined by starting from the marginal distribution:

$$\mathrm{d}\mathbf{x}_t = \left[\mathbf{f}(t)\mathbf{x}_t - g^2(t)\nabla_{\mathbf{x}} \log p_t(\mathbf{x})\right] \mathrm{d}t + g(t)\mathrm{d}\bar{\boldsymbol{w}}_t, \tag{2}$$

where $p_t(\mathbf{x}) := p(\mathbf{x}_t, t)$ is the marginal distribution of $\mathbf{x}$ at time t, $\bar{\boldsymbol{w}}_t$ is the standard wiener process in reversed process, and $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ represents the score function of the noise perturbed distribution at each time t. In practice, the noise prediction model $\epsilon_\theta(\mathbf{x}_t, t)$ that parameterized by θ is used to estimate the score function $-\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$.

During the diffusion process, the SDE simulates data evolution through the injection of random noise. However, due to its stochastic nature, undesired fluctuations may be introduced into the transformed samples, potentially leading to instability, particularly in domain adaptation tasks that require feature consistency. To address this issue, the probabilistic flow (PF) ordinary differential equation (ODE) is introduced in the TSDB following [28], which enables the transformation of samples into deterministic trajectories while preserving the same marginal distributions as the corresponding SDE. Accordingly, for the forward SDE defined in (1), an equivalent deterministic PF ODE can be formulated as:

$$\mathrm{d}\mathbf{x}_t = \left[\mathbf{f}(t)\mathbf{x}_t + \frac{g^2(t)}{2\sigma_t}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right] \mathrm{d}t. \tag{3}$$

In this paper, the DPM-Solver [20] is adopted to solve the aforementioned PF ODE, and its exact solution is provided

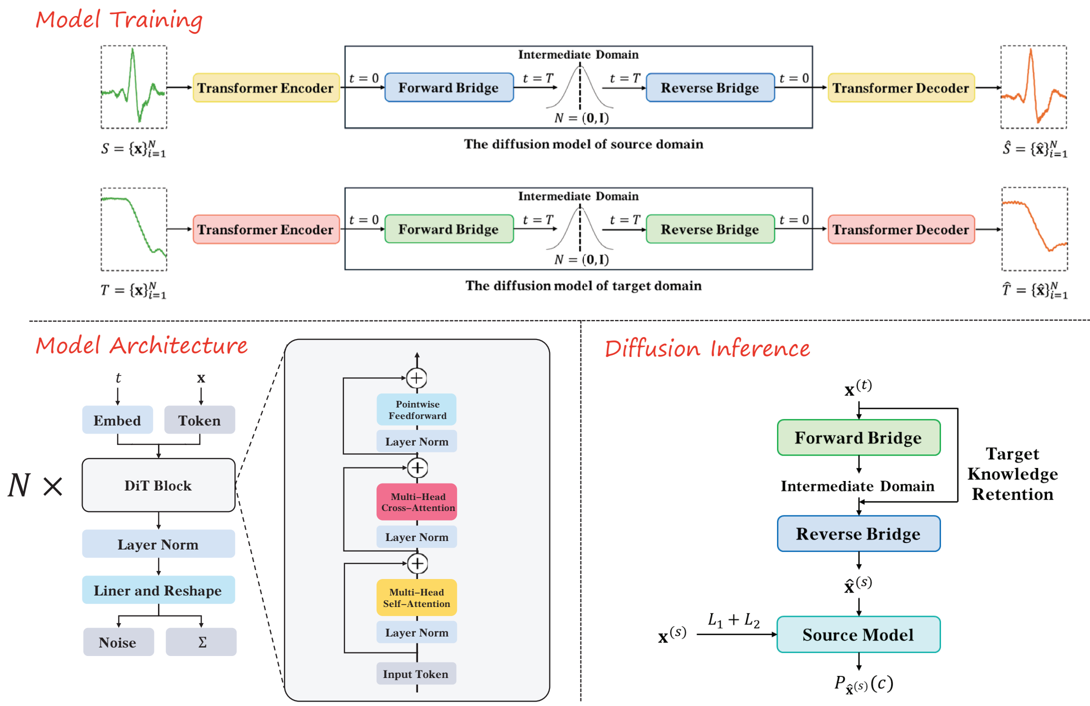


Fig. 2: The proposed DBDA framework including training strategy, inference process, and model architecture.

in Section III-E. For notational convenience, the operator DPMSolve is introduced (where $-\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ is replaced by $\boldsymbol{\epsilon}_\theta$) to denote the mapping from $\mathbf{x}_0$ to $\mathbf{x}_T$:

$$\mathrm{DPMSolve}(\mathbf{x}_0, \varphi_\theta, \lambda_0, \lambda_T) = \mathbf{x}_0 + \int_{\lambda_0}^{\lambda_T} \varphi_\theta\left(\mathbf{x}_t, \lambda\right) \mathrm{d}\lambda, \quad (4)$$

where $\lambda_0 = \log(\alpha_0/\sigma_0)$ is a strictly decreasing function of $t = 0$ corresponding to noise schedule, and λ_T is the same. $\varphi_\theta = \mathrm{d}\mathbf{x}/\mathrm{d}t$ denotes the θ-parameterized velocity field.

C. Target-to-Source Diffusion Bridge

Given a source sample $\mathbf{x}^{(s)}$, a target sample $\mathbf{x}^{(t)}$, a source diffusion model denoted by $\varphi_\theta^{(s)}$, and a target diffusion model denoted by $\varphi_\theta^{(t)}$, the TSDB first applies DPMSolve in the target domain to obtain the intermediate representation $\mathbf{x}^{(i)}$ at the terminal time T, which is regarded as the latent code associated with the prior distribution. Subsequently, the intermediate domain $\mathcal{I}$ is used as the initial condition for DPMSolve with the source diffusion model to generate the transformed target-domain sample. An illustration of the TSDB is provided in Fig. 3. In summary, the translation process from the target domain to the source domain is defined as

$$\mathbf{x}^{(i)} = \mathrm{DPMSolve}(\mathbf{x}^{(t)}, \varphi_\theta^{(t)}, \lambda_0, \lambda_T), \quad (5)$$

$$\hat{\mathbf{x}}^{(s)} = \mathrm{DPMSolve}(\mathbf{x}^{(i)}, \varphi_\theta^{(s)}, \lambda_T, \lambda_0), \quad (6)$$

where (5) solves the forward probability flow under the target diffusion model, mapping the target samples to the latent space. (6) then solves the reverse process under the source diffusion model, decoding the latent variables back to the source data manifold, which aims to generate samples with source-domain style transformation. Thus, TSDB couples the two diffusion trajectories through the combination of (5) and (6) to achieve the transformation from the target domain to the source domain.

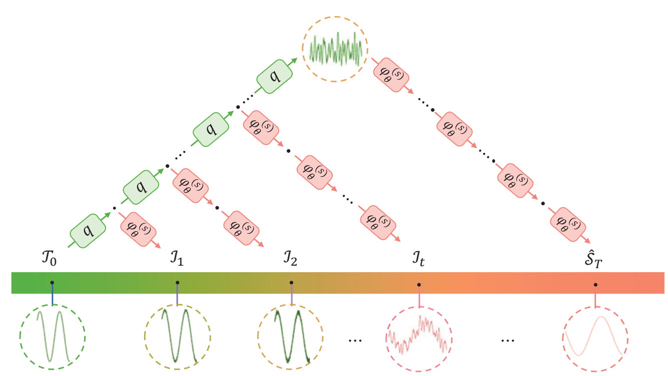


Fig. 3: The illustration of the TSDB. The green path represents the forward diffusion process in the target domain, which maps target samples into a latent Gaussian space. The red path denotes the reverse diffusion process in the source domain, which decodes latent samples back into diverse fault samples with a source-domain style. Each bridge concatenates two adjacent intermediate domains, with slight shifts in distribution. Consequently, the overall transformation is achieved through multiple progressive transitions rather than a single mapping. The color gradient bar at the bottom and the waveform example visually illustrate this gradual evolution process.

From a theoretical perspective, the PF-ODE represents the solution to a special Schrödinger Bridge Problem with linear or degenerate drift between the data and latent distributions

[7], which can be interpreted as an optimal transport problem. As a result, the TSDB is capable of achieving high translation efficiency compared with alternative approaches. As a result, the TSDB is capable of achieving high translation efficiency compared with alternative approaches.

Remark 1: The source diffusion model and the target diffusion model are trained independently within their respective domains. Accordingly, the translation process of the proposed TSDB can be interpreted as the concatenation of two diffusion bridges: a forward bridge that maps the target domain to the intermediate domain, and a reverse bridge that connects the intermediate domain to the source domain. By decoupling the pairwise training procedure inherent in traditional domain adaptation, TSDB allows domain-specific diffusion models to remain applicable across different domain pairs, regardless of whether a domain appears as the source or the target. Since pairwise training is not required, privacy preservation for domain data owners can be ensured, thereby rendering DBDA particularly suitable for source-free adaptation scenarios.

D. Target Knowledge Retention

The primary goal of TSDB is to transform target samples into the source domain while preserving as much target-domain semantic information as possible. This design helps avoid degradation in task-specific strategy learning and improves both transferability and discriminability of the decision boundary. However, the diffusion process does not provide explicit guidance for this transformation. As a result, balancing domain translation and discriminative semantic preservation in the fusion domain $\mathcal{F}$ which consists of real source samples $\mathbf{x}^{(s)}$ and transformed samples $\hat{\mathbf{x}}^{(s)}$, is difficult. To address this issue, an iterative latent refinement step conditioned on the target inputs is introduced during the reverse process, with the aim of guiding the entire trajectory modeling toward improved returns while avoiding excessive consistency.

The Fourier filter exploits frequency-domain statistics to disentangle time-variant and time-invariant components, thereby enabling effective characterization of transition dynamics along distribution trajectories. Specifically, by defining τ as a sliding window over the sequence, the Fourier filter performs the disentanglement of $\mathbf{x}$ as

$$\mathbf{x}_{\text{inv}} = \text{FT}^{-1}\left(\mathcal{G}_\delta\left(\text{FT}\left(\mathbf{x}\right)\right)\right), \tag{7}$$

$$\mathbf{x}_{\text{var}} = \text{FT}^{-1}\left(\bar{\mathcal{G}}_\delta\left(\text{FT}\left(\mathbf{x}\right)\right)\right), \tag{8}$$

$$\text{s.t.}\quad \text{FT}(\mathbf{x}) = \mathbf{x} - \mathbf{x}_{\text{inv}}, \tag{9}$$

where FT and FT^{-1} represent Fourier transform and its inverse. $\mathcal{G}_\delta$ denotes filter, and $\bar{\mathcal{G}}_\delta$ represents its complementary set, which is designed to satisfy $\text{FT}(\mathbf{x}) = \mathbf{x} - \mathbf{x}_{\text{inv}}$ to avoid information loss. At each step of the reverse process, an estimate $\hat{\mathbf{x}}_{t-1}$ at time $t-1$ can be obtained from $p(\hat{\mathbf{x}}_{t-1}|\mathbf{x}_t)$. Furthermore, since the semantic information contained in failure signals is primarily reflected in non-stationary components, $\hat{\mathbf{x}}_{t-1}$ is encouraged to move in a direction that reduces the discrepancy between $\mathbf{x}_{\text{var}}(T)$ and $\hat{\mathbf{x}}_{\text{var}}(t-1)$, as expressed by

$$\tilde{\mathbf{x}}^{(s)}_{t-1} = \hat{\mathbf{x}}^{(s)}_{t-1} + \boldsymbol{\gamma}\cdot\text{sign}\left(\nabla_{\mathbf{x}_t}\left\|\mathbf{x}^{(t)}_{\text{var},0} - \hat{\mathbf{x}}^{(s)}_{\text{var},(t-1)}\right\|_2\right), \tag{10}$$

where $\boldsymbol{\gamma}$ is a scaling hyperparameter that controls the refinement step size. Although TSDB facilitates domain translation, semantic drift can accumulate along the reverse denoising trajectory without explicit guidance. (10) introduces a lightweight per-step refinement that reduces the discrepancy between the target semantic component and its counterpart in the transformed sample. Repeatedly applying the update throughout the trajectory steers the transformation towards semantic consistency, with the step size controlling the strength of the guidance. The pseudocode for the target knowledge retention procedure is summarized in Algorithm 1.

Algorithm 1: Target Knowledge Retention

Input: target sample $\mathbf{x}^{(t)}$, source model $\varphi^{(s)}_\theta$, target model $\varphi^{(t)}_\theta$, diffusion step T, Fourier filter FT
Output: $\hat{\mathbf{x}}^{(s)}$
1: $\mathbf{x}^{(i)} = \text{DPMSolve}(\mathbf{x}^{(t)}, \varphi^{(t)}_\theta, \lambda_0, \lambda_T)$,
2: Sample $\mathbf{x}^{(i)}_T$ from $p^{(i)}(\mathbf{x})$.
3: **for** $t \leftarrow T \ldots 1$ **do**
4: $\quad\hat{\mathbf{x}}^{(s)}_{t-1} \sim p_\theta(\mathbf{x}^{(s)}_{t-1}|\mathbf{x}^{(s)}_t)$
5: $\quad\tilde{\mathbf{x}}^{(s)}_{t-1} = \hat{\mathbf{x}}^{(s)}_{t-1} + \boldsymbol{\gamma}\cdot\text{sign}\left(\nabla_{\mathbf{x}_t}\left\|\mathbf{x}^{(t)}_{\text{var},0} - \hat{\mathbf{x}}^{(s)}_{\text{var},(t-1)}\right\|_2\right)$
6: **end for**
7: **return** $\tilde{\mathbf{x}}^{(s)}$

E. Diffusion Objective Function and Inference Process

Since the TSDB employs two diffusion models that are trained separately on the source domain and the target domain, only the objective function and inference process of the source diffusion model are described, as the formulation for the target domain follows an identical procedure.

Given source samples drawn from the distribution $p^{(s)}(\mathbf{x}_0)$, the diffusion model is trained to learn a parameterized data distribution $p^{(s)}_\theta(\mathbf{x}_0)$ that approximates $p^{(s)}(\mathbf{x}_0)$. The denoising diffusion probabilistic model [10] is formulated as a latent variable model of the following form:

$$p^{(s)}_\theta(\mathbf{x}_0) = \int p^{(s)}_\theta(\mathbf{x}_{0:T})\mathrm{d}\mathbf{x}_{1:T}, \tag{11}$$

$$p^{(s)}_\theta(\mathbf{x}_{0:T}) := p^{(s)}_\theta(\mathbf{x}_T)\prod_{t=1}^{T} p^{(s)}_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \tag{12}$$

where $\mathbf{x}_1, \ldots, \mathbf{x}_T$ denote latent variables in the intermediate sample space. Training of the diffusion model is carried out by optimizing the variational lower bound of the data likelihood, thereby encouraging accurate reconstruction of the data from prior information, as expressed by

$$\max_\theta \mathbb{E}_{p^{(s)}(\mathbf{x}_0)}[\log p^{(s)}_\theta(\mathbf{x}_0)] \leq \tag{13}$$
$$\max_\theta \mathbb{E}_{p^{(s)}(\mathbf{x}_0,\ldots,\mathbf{x}_T)}\left[\log p^{(s)}_\theta(\mathbf{x}_{0:T}) - \log p^{(s)}(\mathbf{x}_{1:T}|\mathbf{x}_0)\right].$$

In practice, the variational lower bound in (13) is replaced by a simplified surrogate loss given by

$$L(\boldsymbol{\epsilon}_\theta) := \sum_{t=1}^{T}\mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}_t}\left[\left\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}^{(s)}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t, t)\right\|^2_2\right], \tag{14}$$

where $\boldsymbol{\epsilon}_\theta^{(s)}$ is a source noise prediction network intended to predict $\boldsymbol{\epsilon} \in N(\mathbf{0}, \mathbf{I})$ from $\mathbf{x}_T$. With a trained noise prediction network $\boldsymbol{\epsilon}_\theta^{(s)}$, the solution $\mathbf{x}_{t-1}^{(s)}$ at time $t-1$ for the diffusion PE ODE in (4) can be obtained using DPMSolve as

$$\hat{\mathbf{x}}_{t-1}^{(s)} = \frac{\alpha_{t-1}}{\alpha_t}\mathbf{x}_t^{(s)} - \alpha_{t-1}\int_{\lambda_t}^{\lambda_{t-1}} e^{-\lambda}\boldsymbol{\epsilon}_\theta(\hat{\mathbf{x}}_\lambda, \lambda)\mathrm{d}\lambda, \quad (15)$$

where the coefficients α_t and α_{t-1} represent the deterministic scaling terms in the diffusion parameterization $\mathbf{x}_t = \alpha_t\mathbf{x}_0 + \sigma_t\boldsymbol{\epsilon}$, while $\lambda_t = \log(\alpha_t/\sigma_t)$ and λ_{t-1} denote the corresponding log-signal-to-noise ratios (log-SNRs). The variable $\hat{\mathbf{x}}_\lambda$ is the intermediate latent state along the continuous reverse-time trajectory parameterized by λ, and $e^{-\lambda} = \sigma/\alpha$ acts as a weighting factor that scales the predicted noise according to the instantaneous SNR. The term $\epsilon_\theta(\hat{\mathbf{x}}_\lambda, \lambda)$ denotes the model-predicted noise at state $\hat{\mathbf{x}}_\lambda$, where θ represents the learnable parameters of the neural denoising network.

F. History-Aware Boundary Extension Strategy Learning

As an add-on procedure, a key advantage of the DBDA framework lies in its compatibility with both upstream data preprocessing and downstream strategy learning approaches. However, it should be noted that, in practice, the real source domain is naturally extended to the fusion domain through the injection of high-level semantic information from different and previously unseen target domains. This enlarged feature space inevitably results in increased diversity within the source domain.

Motivated by this observation, we design a boundary extension loss to cope with diverse and evolving fault patterns across domains. The loss includes a history-aware cross-entropy term and a confidence error regularization term for the source classifier. First, to enable progressive refinement of the decision boundary along the evolving fusion domains, a history-aware classification loss is introduced by integrating information from all preceding training steps. At each step t, rather than relying solely on samples from the current fusion domain $\mathcal{F}_t$, the source classifier exploits a history-aggregated soft label representation that accumulates the true labels associated with all previous steps from 1 to t through an exponentially decaying weighting scheme. This mechanism allows stable patterns learned from earlier domains to be retained while adapting to the most recent domain shift. The resulting history-aware classification loss is formulated as

$$L_1 = -\frac{1}{MT}\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{i=1}^{C}\tilde{\mathbf{y}}_{m,i}^{(t)}\log\left(\frac{\exp\big(z_{m,i}^{(t)}\big)}{\sum_{j=1}^{C}\exp\big(z_{m,j}^{(t)}\big)}\right), \quad (16)$$

where M denotes the number of elements in the fusion domain, C is the number of categories, and z represents the signal features extracted by the diagnostic model from the fusion domain $\mathcal{F}$ at each inference step. It is emphasized that a label-space consistent setting is adopted in this paper, where the source and target domains share an identical set of class labels. As a result, cross-domain knowledge transfer is focused exclusively on distributional discrepancies rather than categorical mismatches. The history-aggregated label $\tilde{\mathbf{y}}_{m,i}^{(t)}$ is computed as:

$$\tilde{\mathbf{y}}_m^{(t)} = \sum_{k=1}^{t}\omega_{t,k}\,\mathbf{y}_m^{(k)}, \quad (17)$$

$$\sum_{k=1}^{t}\omega_{t,k} = 1, \quad \omega_{t,k} \geq 0, \quad (18)$$

where $k = 1, \ldots, t$ is the index of the historical step, and the exponentially decaying weights are defined as:

$$\omega_{t,k} = \frac{(1-\gamma)\,\gamma^{t-k}}{1-\gamma^t}, \quad (19)$$

where $0 < \gamma < 1$ is the decay coefficient. In addition, a confidence error regularization term is incorporated as part of the boundary extension loss to ensure that hard samples, i.e., samples located near the decision boundary, are not assigned lower confidence rewards:

$$L_2 = \sum_{1\leq i,j\leq N}\Delta_{i,j}\mathbb{I}[e_i > e_j], \quad (20)$$

where e_i and e_j denote the errors associated with the i-th and j-th samples from the fusion domain, respectively. The indicator function $\mathbb{I}$ takes the value 1 if the error of the i-th sample exceeds that of the j-th sample, and 0 otherwise. The confidence discrepancy $\Delta_{i,j}$ is computed as

$$\Delta_{i,j} = \max\{0, \max_{c\in C}P_i(c) - \max_{c\in C}P_j(c)\}^2, \quad (21)$$

where $P_i(c)$ denotes the predicted probability of the diagnosis model for category c corresponding to the i-th sample. In this study, the source classifier is implemented using a three-layer Long Short-Term Memory (LSTM) network followed by a two-layer fully connected neural network. After constructing the fusion domain, the diagnostic model is trained on samples with increased diversity and varying degrees of translation. The history-aware cross-entropy term in (16)-(19) stabilizes boundary updates by leveraging historical decision information. The confidence regularization term in (20)-(21) further penalizes over-confident predictions on out-of-domain or uncertain features, preventing unreliable samples from dominating the training process. The aforementioned items collectively promote a smoother and more transferable evolution of the decision boundary during gradual domain migration. The training process of DBDA is formally presented in Algorithm 2.

Remark 2: If the classifier is trained at each step using only the current domain, unstable decision boundaries and catastrophic forgetting may arise due to the non-stationary nature of the feature space. To mitigate this issue, a history-aware learning mechanism is introduced to aggregate supervision information from all previous steps. Specifically, during training at step t, the source classifier does not rely solely on the current fusion domain $\mathcal{F}_t$; instead, information from all preceding domains $\{\mathcal{F}_1, \ldots, \mathcal{F}_t\}$ is jointly incorporated. By fusing historical labels through an exponentially decaying weighting scheme, discriminative cues from earlier domains are preserved while greater emphasis is placed on recent

Algorithm 2: The training process of the DBDA

Input: labeled source dataset $\mathcal{D}_S = \{(\mathbf{x}^{(s)}, \mathbf{y}^{(s)})\}$, unlabeled target dataset $\mathcal{D}_T = \{\mathbf{x}^{(t)}\}$, source diffusion model $\varphi_\theta^{(s)}$, target diffusion model $\varphi_\theta^{(t)}$, classifier f_ψ, diffusion step T, Fourier filter FT

Output: transformed target samples $\tilde{\mathbf{x}}^{(s)}$ and robust classifier f_ψ

Train $\varphi_\theta^{(t)}$ on $\mathcal{D}_T$ by minimizing the diffusion objective in (14).
Train $\varphi_\theta^{(s)}$ on $\mathcal{D}_S$ by minimizing the diffusion objective in (14).
Initialize transformed target set $\tilde{\mathcal{D}}_{T\to S} \leftarrow \emptyset$.
for each target sample $\mathbf{x}^{(t)} \in \mathcal{D}_T$ **do**
  $\mathbf{x}^{(i)} = \mathrm{DPMSolve}(\mathbf{x}^{(t)}, \varphi_\theta^{(t)}, \lambda_0, \lambda_T)$
  $\hat{\mathbf{x}}^{(s)} = \mathrm{DPMSolve}(\mathbf{x}^{(i)}, \varphi_\theta^{(s)}, \lambda_T, \lambda_0)$
  $\tilde{\mathbf{x}}^{(s)} =$ **TargetKnowledgeRetention**$(\hat{\mathbf{x}}^{(s)}, \mathbf{x}^{(t)}, \varphi_\theta^{(s)}, \varphi_\theta^{(t)}, T, \mathrm{FT})$
end for
Construct fusion domain $\mathcal{F} \leftarrow \{\tilde{\mathbf{x}}^{(s)}\} \cup \mathcal{D}_S$.
Train f_ψ on $\mathcal{F}$ with the history-aware boundary extension loss in (16)-(21).
return f_ψ and $\tilde{\mathcal{D}}_{T\to S}$

distribution-consistent information. This cumulative learning strategy enables the source classifier to capture the temporal evolution of domain alignment and to progressively refine its decision boundary as domain discrepancies are reduced. Consequently, more stable adaptation performance is achieved, together with improved generalization across evolving intermediate domains.

IV. Experiments

Across all evaluated domain shift scenarios, DBDA consistently achieves the best performance on both the pipeline dataset and the Sleep-EDF dataset. On the pipeline datasets, which include nine cross-domain tasks, DBDA achieves the highest average accuracy of 72.84% and ranks first in all nine tasks. The improvements of the DBDA are especially pronounced in the presence of larger domain gaps, thereby highlighting the effectiveness of the diffusion bridge-based DBDA in facilitating progressive alignment. On the Sleep-EDF dataset, DBDA achieves an average accuracy of 66.73%, surpassing the state-of-the-art TemSR method at 64.72%. The ablation results further validate the contributions of each component. Removing the Fourier-based retention module caused a 4.93% accuracy drop in the HP→LP task and an 8.25% drop in the HP→AWD task. Removing the L_2 confidence regularization term results in further accuracy declines for both tasks. The ablation results demonstrate that the target knowledge retention mechanism mitigates semantic drift, while the history-aware boundary extension strategy learning stabilizes the formation of decision boundaries within the fusion domain. Finally, the additional runtime introduced by components related to the diffusion transformation during training is evaluated. Experimental results demonstrate that the impact on overall computational cost is minimal, with the additional overhead being effectively manageable.

A. Dataset Descriptions

To comprehensively evaluate the performance of the proposed DBDA, two types of benchmarks are adopted, including custom natural gas pipeline datasets [37] and the public Sleep-EDF dataset [13].

The natural gas pipeline datasets are self-collected and consist of the negative pressure wave dataset (NPWD) and the acoustic wave dataset (AWD). The NPWD is simulated using a ZJ-CSGD type platform, with key parameters set to a pipeline length of 180.2 m, a sampling frequency of 1024 Hz, and a flow rate of $10\ \mathrm{m}^3/\mathrm{h}$. The NPWD includes three operating conditions, namely high pressure (HP), medium pressure (MP), and low pressure (LP), as illustrated in Fig. 4. The AWD is simulated using an HD-II type platform, with key parameters set to a pipeline length of 160 m, a sampling frequency of 5000 Hz, a flow rate of $60\ \mathrm{m}^3/\mathrm{h}$, and a pressure condition of 0.5 MPa, as shown in Fig. 5. For both the NPWD and AWD, pipeline data collected under each pressure condition are categorized into three fault types by switching valves, including large, medium, and small leakage (LL, ML, and SL).

(a) (b)

Fig. 4: Console and platform of the negative pressure wave dataset.

(a) (b)

Fig. 5: Console and platform of the acoustic wave dataset.

The Sleep-EDF dataset is employed as a benchmark for sleep stage classification (SSC) tasks. In this study, model performance is evaluated using the Epz-Cz EEG channel. The dataset contains recordings from 20 different individuals, with each individual treated as a distinct domain. For each subject, sleep stages are categorized into five classes: wake, light sleep stage 1 (N1), light sleep stage 2 (N2), deep sleep stage 3 (N3), and rapid eye movement (REM).

TABLE II: Pipeline fault diagnosis results (%) are listed in this table. Results are averaged from all random seeds.

Method	NPWD						NPWD→AWD			Ave
	HP→MP	HP→LP	MP→HP	MP→LP	LP→HP	LP→MP	HP→AWD	MP→AWD	LP→AWD	
SASA	57.84±2.59	59.61±3.43	60.89±1.10	66.73±1.98	64.37±2.75	67.07±4.06	60.56±1.91	57.67±1.82	56.86±2.27	61.29
FixBi	66.25±3.63	58.98±5.68	63.86±2.82	70.08±1.45	56.40±3.85	63.66±1.03	50.76±2.18	46.84±4.03	48.09±3.52	58.32
OUDA	66.95±1.41	64.69±3.29	62.93±2.24	67.33±2.79	71.13±1.74	69.02±2.46	62.00±1.78	62.67±1.37	60.69±2.32	65.27
CoTTA	68.33±1.75	70.38±2.25	67.97±1.42	70.23±1.02	69.78±1.28	71.37±0.87	64.33±2.23	61.80±2.05	63.92±1.98	67.56
DDA	70.55±1.85	71.53±1.51	74.82±2.60	70.84±0.80	70.24±1.54	71.10±1.76	66.32±1.57	64.50±2.01	65.56±1.69	69.50
DBDA(our)	74.82±1.20	74.07±1.26	71.15±1.69	74.97±2.85	73.49±1.72	75.54±1.50	70.75±1.01	71.25±1.37	69.55±0.85	72.84

B. Implementation Details

1) Neural Network Architecture: The Transformer architecture is adopted as the backbone of the DBDA framework following diffusion transformers (DiTs) [25], which have been shown to outperform traditional convolutional networks in visual recognition tasks. A sequence of 12 DiT blocks is employed, with each block configured with a hidden dimension size of 384 and 6 attention heads. During training, the model is optimized using the Adam optimizer, with the learning rate set to 1×10^{-4}. The total number of training epochs is 1000, and the batch size is fixed at 64.

2) Training Strategy: A key advantage of DBDA lies in its ability to overcome the limitations associated with pairwise training in existing UDA approaches, thereby exhibiting stronger generalization capability in data-scarce scenarios. Accordingly, the performance of the proposed model is evaluated through comparative analysis with other representative domain adaptation algorithms.

3) Baselines: To assess the performance of DBDA on pipeline datasets, five well-established models are selected as baselines, including DDA [12], CoTTA [38], FixBi [24], SASA [2], and OUDA [37]. For fairness, all hyperparameters are determined using a grid search strategy on the validation sets.

For the Sleep-EDF dataset, experimental results reported for five DA-based classification models in [26] and [40] are adopted for comparison, including NRC, AaD, BAIT, MAPU, and TemSR.

C. Main Results

1) Pipeline Datasets: The mean accuracy and standard error of all algorithms are reported in Table II, where the best results are highlighted in red and the second-best results are underlined. All results are obtained using five different random seeds. In comparison with other methods, DBDA achieves optimal performance on 9 tasks, attaining the highest average accuracy of 72.84%.

Notably, although OUDA explicitly models both domain-invariant and domain-specific features, its performance remains inferior to that of DBDA, suggesting that feature alignment-based approaches lack sufficient flexibility to effectively address data privacy-constrained scenarios. In addition, DDA, despite being regarded as a state-of-the-art algorithm in computer vision, exhibits limited effectiveness in several time-series classification tasks. On one hand, the linear low-pass filter employed in DDA is primarily designed to capture high-level features in image data. On the other hand, DDA relies on pairwise training, which restricts its adaptability to a wider range of transfer scenarios. By contrast, the proposed method preserves frequency characteristics during sequence transformation and is therefore better suited to handling such settings. These observations indicate that diffusion-based models are well suited for feature transformation and data adaptation, and that the proposed DBDA framework is capable of effectively addressing real-world transfer tasks, irrespective of the availability of source data.

2) Sleep-EDF datasets: Fig. 6 shows the performance of the proposed DBDA and its competitors on the public Sleep-EDF. First, NRC, AaD, and BAIT are source-free domain adaptation (SFDA) methods based on contrastive learning, which aim to achieve unsupervised adaptation by enhancing prediction consistency among local neighbors and promoting diversity in the global prediction distribution. However, the performance of NRC (56.84%<64.72%), AaD (59.11%<64.72%), and BAIT (61.53%<64.72%) does not reach the mean line in any of the five tasks. This phenomenon illustrates that these methods rely solely on static feature similarity between time series to construct pseudo-supervision, which is prone to failure in scenarios with ambiguous decision boundaries or high uncertainty. Moreover, in terms of average performance, the results of MAPU (64.04%) and TemSR (64.72%) are comparable to our DBDA method (66.73%), demonstrating that generative mechanism-based SFDA methods are able to efficiently achieve cross-domain transfer with adequate modeling of source domain structure information. Nevertheless, our algorithm still performs superior to TemSR on all tasks, benefiting from its structurally decoupled design.

D. Model Analysis

1) Representation Analysis: To directly assess the distribution transformation capability of the DBDA framework, we conduct a visualization study on a synthetic Concentric Rings (CR) dataset, which contains complex non-linear structures and therefore serves as a suitable testbed for evaluating whether a method can translate distributions without destroying geometry. As shown in the left figure of Fig. 7, each source point is assigned a unique color so that the same instance can be visually tracked after translation. We then translate these samples through the proposed cross-diffusion transformation to the target domain while preserving the color

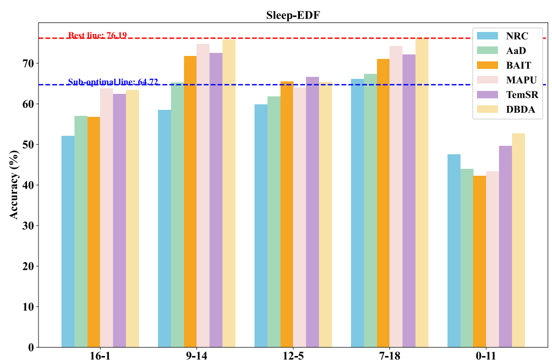


Fig. 6: Classification results for five cross-domain tasks on the Sleep-EDF dataset.

identities, which enables an explicit examination of point-wise correspondences along the mapping.

The key observation is that the DBDA produces a smooth and coherent translation that preserves structural relationships. In the right figure of Fig. 7, points sharing the same color remain locally consistent after translation and form continuous trajectories rather than discontinuous jumps. This indicates that the learned transformation is not merely matching global statistics but maintains neighborhood structure and avoids mode collapse. Moreover, the ring-wise geometry is largely retained, suggesting that DBDA captures the intrinsic non-linear structure of the target distribution and maps it to the source distribution in a stable manner. Overall, this visualization supports the effectiveness of DBDA in achieving distribution translation with structural consistency, which aligns with our design goal of progressive and semantically faithful cross-domain transformation.

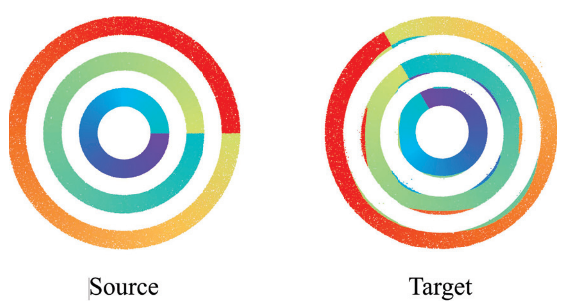


Fig. 7: Translation task on synthetic dataset. (Left) The source dataset CR. (Right) Result of translation to the target domains.

2) Ablation Studies: To verify the rational effectiveness of the DBDA components, detailed ablation studies are conducted by systematically adding and removing individual modules. TSDB, as the foundation of DBDA, enhances classification performance for subsequent FT and L_2 by implementing target-to-source transformation and progressive alignment based on diffusion bridge. As summarized in Table III, for the proposed DBDA framework, the removal of FT results in noticeable performance degradation, particularly on the HP→LP and 0→11 tasks, with decreases of 2.16% and 7.36%, respectively. Based on this simplified variant, further removal of the L_2 loss leads to additional performance deterioration, indicating that the L_2 loss plays an important role in regulating classification confidence errors associated with out-of-domain features. Moreover, when incorporating L_2 loss into strong baselines such as DDA and TemSR, consistent improvements are observed, including gains of 1.86% and 3.12% for DDA on the HP→LP and HP→AWD tasks, respectively, as well as improvements of approximately 2.5% on both Sleep-EDF tasks. These results demonstrate that DBDA consistently enhances cross-domain classification performance and that diffusion-based modeling provides an effective basis for robust adaptation.

TABLE III: Performance demotions (↓) or promotions (↑) are obtained by our components. The ablation of DBDA is conducted sequentially: we first remove **FT**, and then further remove L_2 on top of this reduced model.

Datasets		Pipeline		Sleep-EDF	
Tasks		HP→LP	HP→AWD	9→14	0→11
DBDA	Original	74.07%	70.75%	75.91%	52.74%
	- FT	71.91%	66.28%	72.08%	45.38%
	- L_2	70.42%	64.91%	71.69%	43.27%
	Demotion	4.93%↓	8.25%↓	5.56%↓	17.96%↓
DDA	Original	71.53%	66.32%	-	-
	+ L_2	72.86%	68.39%	-	-
	Promotion	1.86%↑	3.12%↑	-	-
TemSR	Original	-	-	72.60%	49.62%
	+ L_2	-	-	74.35%	50.84%
	Promotion	-	-	2.41%↑	2.46%↑

3) Hyperparameter Sensitivity: The sensitivity of the DBDA framework to several hyperparameters is evaluated, including the learning rate lr, the depth of the Transformer L, and the balancing parameter between the L_1 loss and the L_2 loss. The corresponding results are presented in Fig. 8. It is observed that the learning rate, as one of the most influential factors, requires careful selection. In addition, increasing the number of Transformer blocks in DBDA does not necessarily yield performance improvements, indicating that larger model configurations are not always advantageous. Furthermore, variations in the balancing parameters exhibit a limited impact on the overall performance of DBDA.

4) Computational Complexity and Runtime: The proposed DBDA consists of three steps: (a) the training of diffusion models in the source and target domains; (b) domain transformation via TSDB; (c) the training of a classifier on the fusion domain. Relative to standard DA that only trains a feature extractor and a classifier, the main additional overhead arises from (a) and (b). We set the reverse step to $T = 1000$, while adopting an accelerated solver with 50 steps at inference; therefore, a single solver run requires 50 decoder evaluations. For each target sample, TSDB executes one forward solve with the target diffusion model and one reverse solve with the source diffusion model, resulting in approximately 100 decoder evaluations per sample. The target-knowledge retention step is optional and introduces only a linear overhead with respect to the number of refinement steps. Classifier training on the fusion domain has complexity comparable to conventional

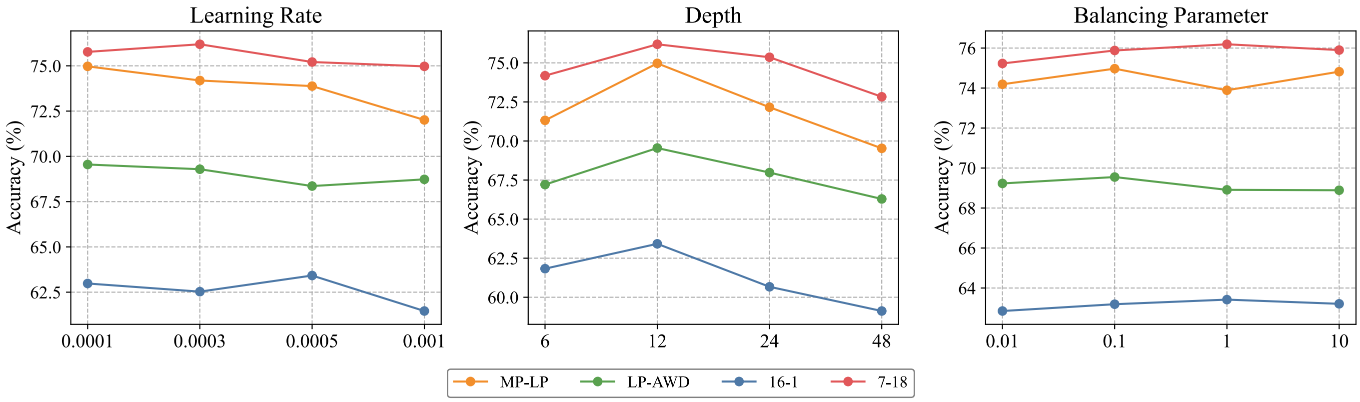


Fig. 8: Hyperparameter sensitivity with respect to the learning rate lr, the depth of Transformer L, and the balancing parameter for L_1 loss and L_2 loss.

domain adaptation, since optimization is performed only for a discriminative network on an augmented dataset. In practice, runtime can be effectively regulated by selecting the solver step budget and enabling retention only when necessary.

V. Conclusion

This paper has proposed a DBDA model for GDA, with a focus on fault diagnosis. Considering the non-stationary nature of operating conditions that gradually evolve due to factors such as transportation demands, the proposed DBDA has introduced a diffusion bridge-driven data adaptation framework, which aims to leverage the distribution transformation properties of diffusion models to achieve the mapping from the target domain to the source domain with semantic consistency, enabling the rapid reuse of the source diagnostic model. The framework is composed of three main components: (1) a bidirectional TSDB that facilitates latent-space mapping and reverse decoding to enable progressive target-to-source transformation; (2) a target knowledge retention mechanism that preserves critical semantic features from the target domain during adaptation, thereby mitigating semantic drift and enhancing generalization performance; and (3) a history-aware boundary extension loss that strengthens the robustness of the diagnostic model under diverse operating conditions by stabilizing decision boundaries on the fusion domain. Extensive experimental results have obtained on multiple datasets have demonstrated that the proposed DBDA method achieves superior diagnostic accuracy and efficiency, thereby confirming its effectiveness and practical applicability. In future work, we will focus on the following three areas: (a) Extending DBDA to more challenging scenarios, including multi-domain and continual adaptation by incorporating condition-aware diffusion modeling [3], [17]; (b) Improving efficiency and advancing the theoretical understanding of DBDA under diverse drift patterns [1], [16]; (c) Integrating DBDA with control theory to facilitate effective and rational implementation across a broader range of industrial applications [19], [29].

References

[1] G. Biroli, T. Bonnaire, V. De Bortoli, and M. Mézard, Dynamical regimes of diffusion models, *Nature Communications*, vol. 15, no. 1, art. no. 9957, 2024.

[2] R. Cai, J. Chen, Z. Li, W. Chen, K. Zhang, J. Ye, Z. Li, X. Yang, and Z. Zhang, Time series domain adaptation via sparse associative structure alignment, In: *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, vol. 35, no. 8, pp. 6859–6867.

[3] Z. Cai, Y. Huang, T. Zhang, Y. Zheng, and D. Yue, Multi-source domain adaptation by causal-guided adaptive multimodal diffusion networks, In: *International Journal of Computer Vision*, 2025, vol. 133, no. 7, pp. 4623–4645.

[4] B. Chen, K. Thandiackal, P. Pati, and O. Goksel, Generative appearance replay for continual unsupervised domain adaptation, *Medical Image Analysis*, vol. 89, art. no. 102924, 2023.

[5] H.-Y. Chen and W.-L. Chao, Gradual domain adaptation without indexed intermediate domains, In: *Advances in Neural Information Processing Systems*, 2021, vol. 34, pp. 8201–8214.

[6] S. Chen, P. Sun, Y. Song, and P. Luo, Diffusiondet: Diffusion model for object detection, In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 19830–19843.

[7] T. Chen, G.-H. Liu, and E. A. Theodorou, Likelihood training of schrödinger bridge using forward-backward sdes theory, *arXiv preprint arXiv:2110.11291*, 2021.

[8] W. Chen, J. Hu, Z. Wu, and S. Ma, A survey on fault detection for networked systems under communication constraints, *Systems Science & Control Engineering*, vol. 13, no. 1, art. no. 2460434, 2025.

[9] X. Chen, H. Shao, Y. Xiao, S. Yan, B. Cai, and B. Liu, Collaborative fault diagnosis of rotating machinery via dual adversarial guided unsupervised multi-domain adaptation network, *Mechanical Systems and Signal Processing*, vol. 198, art. no. 110427, 2023.

[10] J. S. Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli, Deep unsupervised learning using nonequilibrium thermodynamics, *arXiv preprint arXiv:1503.03585*, 2015.

[11] L. Enciso-Salas, G. Pérez-Zuñiga, J. Sotomayor-Moriano, E. Chanthery, E. H. Sepúlveda-Oviedo, A. Subias, L. Travé-Massuyès, R. Garcia, and J. Aguilar, A bibliometric literature review of integrated data and model based diagnosis approaches for the industry 4.0, *International Journal of Systems Science*, vol. 57, no. 8, pp. 2415–2442, 2025.

[12] J. Gao, J. Zhang, X. Liu, T. Darrell, E. Shelhamer, and D. Wang, Back to the source: Diffusion-driven adaptation to test-time corruption, In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 11786–11796.

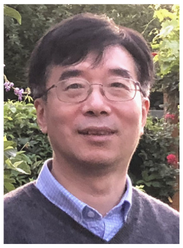
[13] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals, *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[14] D. P. Kingma, T. Salimans, B. Poole, and J. Ho, Variational diffusion models, In: *Advances in Neural Information Processing Systems*, 2021, vol. 34, pp. 21696–21707.

[15] B. Li, K. Xue, B. Liu, and Y.-K. Lai, BBDM: Image-to-image translation with Brownian bridge diffusion models, In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 1952–1961.
[16] R. Liang, B. Yang, P. Chen, X. Li, Y. Xue, and Z. Yu, Diffusion models as network optimizers: Explorations and analysis, *IEEE Internet of Things Journal*, vol. 12, no. 10, pp. 13183–13193, 2025.
[17] Y. C. Liao, J. J. Chen, C. P. Huang, C. S. Lin, M.-L. Wu, and Y.-C. F. Wang, Continual personalization for diffusion models, In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 15511–15520.
[18] J. Liu, L. Wan, F. Xie, Y. Sun, X. Wang, D. Li, and S. Wu, Cross-machine deep subdomain adaptation network for wind turbines fault diagnosis, *Mechanical Systems and Signal Processing*, vol. 210, art. no. 111151, 2024.
[19] S. Liu, L. Wang, Y. Zhang, Y.-A. Wang, and H. Dong, Recursive filtering of networked systems with communication protocol scheduling: a survey, *International Journal of Systems Science*, vol. 56, no. 11, pp. 2499–2516, 2025.
[20] C. Lu, Y. Zhou, F. Bao, J. Chen, C. Li, and J. Zhu, Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps, In: *Advances in Neural Information Processing Systems*, 2022, pp. 5775–5787.
[21] S. Lu, Z. Gao, Q. Xu, C. Jiang, A. Zhang, and X. Wang, Class-imbalance privacy-preserving federated learning for decentralized fault diagnosis with biometric authentication, *IEEE Transactions on Industrial Informatics*, vol. 18, no. 12, pp. 9101–9111, 2022.
[22] Q. Ma, D. Wu, and X. Luo, A review of deep learning-based power load forecasting methods, *International Journal of Network Dynamics and Intelligence*, vol. 4, no. 4, art. no. 100027, 2025.
[23] C. Mao and C. Wen, Fault diagnosis method for rolling bearings in high-noise environments based on COA-FMD-ITEO, *International Journal of Network Dynamics and Intelligence*, vol. 4, no. 3, art. no. 100016, 2025.
[24] J. Na, H. Jung, H. J. Chang, and W. Hwang, FixBi: Bridging domain spaces for unsupervised domain adaptation, In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 1094–1103.
[25] W. Peebles and S. Xie, Scalable diffusion models with transformers, In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 4195–4205.
[26] M. Ragab, E. Eldele, M. Wu, C. S. Foo, X. Li, and Z. Chen, Source-free domain adaptation with temporal imputation for time series data, In: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 1989–1998.
[27] B. Song, S. Zhao, L. Dang, H. Wang, and L. Xu, A survey on learning from data with label noise via deep neural networks, *Systems Science & Control Engineering*, vol. 13, no. 1, art. no. 2488120, 2025.
[28] J. Song, C. Meng, and S. Ermon, Denoising diffusion implicit models, *arXiv preprint arXiv:2010.02502*, 2020.
[29] Y. Song, B. Zhang, C. Wen, D. Wang, and G. Wei, Model predictive control for complicated dynamic systems: A survey, *International Journal of Systems Science*, vol. 56, no. 9, pp. 2168–2193, 2025.
[30] H. Su, J. Li, Z. Du, L. Zhu, K. Lu, and H.-T. Shen, Cross-domain recommendation via dual adversarial adaptation, *ACM Transactions on Information Systems*, vol. 42, no. 3, pp. 1–26, 2024.
[31] Z. Su, J. Zhang, J. Tang, Y. Wang, H. Xu, J. Zou, and S. Fan, A novel deep transfer learning method with inter-domain decision discrepancy minimization for intelligent fault diagnosis, *Knowledge-Based Systems*, vol. 259, art. no. 110065, 2023.
[32] B. Song, S. Zhao, L. Dang, H. Wang, and L. Xu, A survey on learning from data with label noise via deep neural networks, *Systems Science & Control Engineering*, vol. 13, no. 1, art. no. 2488120, 2025.
[33] Y. Song, J. Sohi-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, Score-based generative modeling through stochastic differential equations, *arXiv preprint arXiv:2011.13456*, 2020.
[34] J. Tian, J. Zhang, Y. Jiang, S. Wu, H. Luo, and S. Yin, A novel generalized source-free domain adaptation approach for cross-domain industrial fault diagnosis, *Reliability Engineering & System Safety*, vol. 243, art. no. 109891, 2024.
[35] G. Xu, X. Zhao, Y. He, and J. Hu, Transfer alignment models of inertial navigation system: A review and comparison, *International Journal of Systems Science*, vol. 57, no. 8, pp. 2474–2494, 2025.
[36] C. Wang, Z. Wang, and H. Dong, A novel prototype-assisted contrastive adversarial network for weak-shot learning with applications: Handling weakly labeled data, *IEEE/ASME Transactions on Mechatronics*, vol. 29, no. 1, pp. 533–543, 2024.
[37] C. Wang, Z. Wang, H. Liu, H. Dong, and G. Lu, An optimal unsupervised domain adaptation approach with applications to pipeline fault diagnosis: Balancing invariance and variance, *IEEE Transactions on Industrial Informatics*, vol. 20, no. 8, pp. 10019–10030, 2024.
[38] Q. Wang, O. Fink, V. G. Van, and D. Dai, Continual test-time domain adaptation, In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 7201–7211.
[39] X. Wang, H. Jiang, M. Mu, and Y. Dong, A dynamic collaborative adversarial domain adaptation network for unsupervised rotating machinery fault diagnosis, *Reliability Engineering & System Safety*, vol. 255, art. no. 110662, 2025.
[40] Y. Wang, P. Gong, M. Wu, F. Ott, X. Li, L. Xie, and Z. Chen, Temporal source recovery for time-series source-free unsupervised domain adaptation, *arXiv preprint arXiv:2409.19635*, 2024.
[41] C. Wu, D. Zhao, T. Han, and Y. Xia, Bearing fault diagnosis using gradual conditional domain adversarial network, *Applied Soft Computing*, vol. 158, art. no. 111580, 2024.
[42] X. Yan, C. Wang, and Y. Jin, Federated bimodal graph neural networks for text-image retrieval, *International Journal of Network Dynamics and Intelligence*, vol. 4, no. 2, art. no. 100009, 2025.
[43] B. Yang, Y. Lei, X. Li, and N. Li, Targeted transfer learning through distribution barycenter medium for intelligent fault diagnosis of machines with data decentralization, *Expert Systems with Applications*, vol. 244, art. no. 122997, 2024.
[44] X. Yin, X. Li, Y. Zhang, W. Zhou, and Z. Liu, Multi-task network based health status assessment of cutting tools, *International Journal of Network Dynamics and Intelligence*, vol. 4, no. 2, art. no. 100008, 2025.
[45] Y. Zhan, R. Yang, J. You, M. Huang, W. Liu, and X. Liu, A systematic literature review on incomplete multimodal learning: techniques and challenges, *Systems Science & Control Engineering*, vol. 13, no. 1, art. no. 2467083, 2025.
[46] Y. Zhang, S. Chen, W. Jiang, Y. Zhang, J. Lu, and J. T. Kwok, Domain-guided conditional diffusion model for unsupervised domain adaptation, *Neural Networks*, vol. 184, art. no. 107031, 2025.
[47] Y. Zhang, B. Deng, K. Jia, and L. Zhang, Gradual domain adaptation via self-training of auxiliary models, *arXiv preprint arXiv:2106.09890*, 2021.
[48] Z. Zhuang, Y. Zhang, and Y. Wei, Gradual domain adaptation via gradient flow, In: *The Twelfth International Conference on Learning Representations*, 2024.

Chuang Wang received the B.Sc. degree in automation and the Ph.D. degree in petroleum and natural gas engineering from the Northeast Petroleum University, Daqing, China, in 2017 and 2023, respectively.

From August 2022 to August 2023, he was a Visiting Ph.D. Student with the Department of Computer Science, Brunel University London, Uxbridge, U.K. He is currently a Professor with the Artificial Intelligence Energy Research Institute, Northeast Petroleum University, Daqing, China. His research interests include evolutionary computation and transfer learning techniques.



Zidong Wang (Fellow, IEEE) received the B.Sc. degree in mathematics in 1986 from Suzhou University, Suzhou, China, and the M.Sc. degree in applied mathematics in 1990 and the Ph.D. degree in electrical engineering in 1994, both from Nanjing University of Science and Technology, Nanjing, China.

He is currently Professor of Dynamical Systems and Computing in the Department of Computer Science, Brunel University London, U.K. From 1990 to 2002, he held teaching and research appointments in universities in China, Germany and the U.K. Prof. Wang's research interests include dynamical systems, signal processing, bioinformatics, control theory and applications. He has published a number of papers in international journals. He is a holder of the Alexander von Humboldt Research Fellowship of Germany, the JSPS Research Fellowship of Japan, William Mong Visiting Research Fellowship of Hong Kong.

Prof. Wang serves (or has served) as the Editor-in-Chief for *International Journal of Systems Science*, the Editor-in-Chief for *Neurocomputing*, the Editor-in-Chief for *Systems Science & Control Engineering*, and an Associate Editor for 12 international journals including IEEE TRANSACTIONS ON AUTOMATIC CONTROL, IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, IEEE TRANSACTIONS ON NEURAL NETWORKS, IEEE TRANSACTIONS ON SIGNAL PROCESSING, and IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS-PART C. He is a Member of the Academia Europaea, a Member of the European Academy of Sciences and Arts, an Academician of the International Academy for Systems and Cybernetic Sciences, a Fellow of the IEEE, a Fellow of the Royal Statistical Society and a member of program committee for many international conferences.

Hongli Dong (Senior Member, IEEE) received the Ph.D. degree in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2012.

From 2009 to 2010, she was a Research Assistant with the Department of Applied Mathematics, City University of Hong Kong, Hong Kong. From 2010 to 2011, she was a Research Assistant with the Department of Mechanical Engineering, The University of Hong Kong, Hong Kong. From 2011 to 2012, she was a Visiting Scholar with the Department of Information Systems and Computing, Brunel University London, London, U.K. From 2012 to 2014, she was an Alexander von Humboldt Research Fellow with the University of Duisburg-Essen, Duisburg, Germany. She is currently a Professor with the Artificial Intelligence Energy Research Institute, Northeast Petroleum University, Daqing, China. She is also the Director of the Heilongjiang Provincial Key Laboratory of Networking and Intelligent Control, Daqing, China. Her current research interests include robust control and networked control systems.

Dr. Dong is a very active reviewer for many international journals.