Noise Optimized Conditional Diffusion for Domain Adaptation

Lingkun Luo[1], Shiqiang Hu[1*], Liming Chen[2,3]
[1]School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai, China
[2]LIRIS, CNRS UMR 5205, Ecole Centrale de Lyon, France
[3]Institut Universitaire de France (IUF), France
lolinkun@gmail.com, sqhu@sjtu.edu.cn, liming.chen@ec-lyon.fr

Abstract

Pseudo-labeling is a cornerstone of Unsupervised Domain Adaptation (UDA), yet the scarcity of High-Confidence Pseudo-Labeled Target Domain Samples (**hcpl-tds**) often leads to inaccurate cross-domain statistical alignment, causing DA failures. To address this challenge, we propose **N**oise **O**ptimized **C**onditional **D**iffusion for **D**omain **A**daptation (**NOCDDA**), which seamlessly integrates the generative capabilities of conditional diffusion models with the decision-making requirements of DA to achieve task-coupled optimization for efficient adaptation. For robust cross-domain consistency, we modify the DA classifier to align with the conditional diffusion classifier within a unified optimization framework, enabling forward training on noise-varying cross-domain samples. Furthermore, we argue that the conventional $\mathcal{N}(\mathbf{0}, \mathbf{I})$ initialization in diffusion models often generates class-confused hcpl-tds, compromising discriminative DA. To resolve this, we introduce a class-aware noise optimization strategy that refines sampling regions for reverse class-specific hcpl-tds generation, effectively enhancing cross-domain alignment. Extensive experiments across 5 benchmark datasets and 29 DA tasks demonstrate significant performance gains of **NOCDDA** over 31 state-of-the-art methods, validating its robustness and effectiveness.

1 Introduction

The success of deep learning models relies heavily on abundant, well-labeled training data $(\mathcal{X}_\mathcal{S}, \mathcal{Y}_\mathcal{S})$, enabling impressive performance on test samples $(\mathcal{X}_\mathcal{T})$. These models typically assume that training and testing data are independent and identically distributed (*iid*), implying the same statistical distribution for both domains. However, this assumption often breaks in real-world scenarios [Pan and Yang, 2010; Lu *et al.*, 2020], where variations in environmental conditions (*e.g.*, lighting or temperature) or data capture methods cause significant discrepancies between training and testing data. Such *domain shifts* hinder the effectiveness of models trained on a specific labeled source domain $(\mathcal{D}_\mathcal{S})$ when applied to an unlabeled target domain $(\mathcal{D}_\mathcal{T})$. Unsupervised Domain Adaptation (UDA) addresses these challenges by leveraging models trained on $\mathcal{D}_\mathcal{S}$ for direct application to $\mathcal{D}_\mathcal{T}$, despite the inherent *domain shift*.

The theoretical foundation of UDA, as established in Eq.(1) [Ben-David *et al.*, 2010; Kifer *et al.*, 2004], provides an error bound for the hypothesis h on the target domain, offering a framework to optimize models for cross-domain tasks under non-*iid* conditions:

$$e_\mathcal{T}(h) \leq \underbrace{e_\mathcal{S}(h)}_{\textbf{Term.1}} + \underbrace{d_H(\mathcal{D}_\mathcal{S}, \mathcal{D}_\mathcal{T})}_{\textbf{Term.2}} + \underbrace{\min\left\{E_{\mathcal{D}_\mathcal{S}}\left[|f_\mathcal{S}(\mathbf{x}) - f_\mathcal{T}(\mathbf{x})|\right], E_{\mathcal{D}_\mathcal{T}}\left[|f_\mathcal{S}(\mathbf{x}) - f_\mathcal{T}(\mathbf{x})|\right]\right\}}_{\textbf{Term.3}} \tag{1}$$

where the performance $e_\mathcal{T}(h)$ of a hypothesis h on the $\mathcal{D}_\mathcal{T}$ is bounded by three critical terms on the right-hand side. Specifically, **Term.1** represents the classification error on the $\mathcal{D}_\mathcal{S}$. **Term.2** quantifies the $\mathcal{H}$*-divergence*[Kifer *et al.*, 2004] between the cross-domain distributions $(\mathcal{D}_\mathcal{S}, \mathcal{D}_\mathcal{T})$, and **Term.3** captures the discrepancy between the labeling functions of the two domains. In the context of UDA, optimizing Term.1 is typically achieved through supervised learning to ensure effective functional regularization. However, optimizing Term.2 and Term.3 depends on hcpl-tds from $\mathcal{D}_\mathcal{T}$, which poses a challenge in UDA setups where labels from $\mathcal{D}_\mathcal{T}$ are unavailable.

The scarcity of High-Confidence Pseudo-Labeled Target-Domain Samples (hcpl-tds). Both **Term 2** and **Term 3** rely heavily on hcpl-tds for optimization. However, real-world domain shifts often restrict the availability of these samples, limiting the reduction of the error bound in Eq. (1).

Impact on Term.2: The goal of Term.2 is to align cross-domain conditional distributions, *i.e.*, $Q(\mathcal{Y}_S \mid \mathcal{X}_S) \approx Q(\mathcal{Y}_\mathcal{T} \mid \mathcal{X}_\mathcal{T})$. In UDA, target domain pseudo-labels $\mathcal{Y}_\mathcal{T}$ are often inaccurate due to domain shifts. To mitigate this, we select hcpl-tds to construct $\mathcal{D}_\mathcal{T}^h = \{(\mathcal{X}_\mathcal{T}^i, \mathcal{Y}_\mathcal{T}^i) \mid \text{confidence}(\mathcal{Y}_\mathcal{T}^i) \geq \eta\}$, where η is a confidence threshold. However, $\mathcal{D}_\mathcal{T}^h$ is typically much smaller than $\mathcal{D}_\mathcal{T}$ (*i.e.*, $\left\|\mathcal{D}_\mathcal{T}^h\right\|_0 \ll \|\mathcal{D}_\mathcal{T}\|_0$), limiting its ability to fully capture the statistical characteristics of $\mathcal{D}_\mathcal{T}$. This sparsity impedes accurate distribution alignment, ultimately affecting DA performance.

Impact on Term 3. Let $f_\mathcal{S}(\mathbf{x})$ and $f_\mathcal{T}(\mathbf{x})$ denote the source and target labeling functions, respectively. We define $\mathcal{D}_\mathcal{T}^h = \{\mathbf{x} \in \mathcal{D}_\mathcal{T} \mid f_\mathcal{T}(\mathbf{x}) \geq \eta\}$, as the set of selected hcpl-tds, where $f_\mathcal{T}(\mathbf{x})$ exceeds a predefined confidence threshold η. Naturally, the size of $\mathcal{D}_\mathcal{T}^h$ is often smaller

than the total size of $\mathcal{D}_\mathcal{T}$, limiting the effective sample space available for estimating the expectation in Term 3 of Eq.(1). Since $\mathbb{E}_{\mathcal{D}_\mathcal{T}}$ depends on the target domain samples, this reduction compromises the quality of the estimated expectation $\mathbb{E}_{\mathcal{D}_\mathcal{T}}$. Mathematically, let $\hat{\mathbb{E}}_{\mathcal{D}^h_\mathcal{T}}[\|f_\mathcal{S}(\mathbf{x}) - f_\mathcal{T}(\mathbf{x})\|]$ denote the empirical expectation computed over $\mathcal{D}^h_\mathcal{T}$. When $\left\|\mathcal{D}^h_\mathcal{T}\right\|_0 \ll \|\mathcal{D}_\mathcal{T}\|_0$, the empirical estimation introduces a bias: $\left|\mathbb{E}_{\mathcal{D}_\mathcal{T}}[\|f_\mathcal{S}(\mathbf{x}) - f_\mathcal{T}(\mathbf{x})\|] - \hat{\mathbb{E}}_{\mathcal{D}^h_\mathcal{T}}[\|f_\mathcal{S}(\mathbf{x}) - f_\mathcal{T}(\mathbf{x})\|]\right| \to \varepsilon$, where ε represents the error due to the insufficient representation of $\mathcal{D}_\mathcal{T}$. This bias reduces the robustness of the alignment between $f_\mathcal{S}(\mathbf{x})$ and $f_\mathcal{T}(\mathbf{x})$, thereby hindering effective DA.

Recent studies have demonstrated the effectiveness of diffusion models [Song *et al.*, 2020b; Ho *et al.*, 2020; Lipman *et al.*, 2022] in sample generation and distribution inference. Building on this foundation, we propose a novel approach, **N**oise **O**ptimized **C**onditional **D**iffusion for **D**omain **A**daptation (**NOCDDA**), which seamlessly integrates diffusion-based generation with DA decision-making for synergistic optimization. Specifically, the forward diffusion process is coupled with DA classifier training, enabling robust cross-domain decision-making by injecting progressively increasing noise while refining cross-domain decision boundaries. Meanwhile, the reverse sampling phase employs noise-optimized strategies for class-aware **hcpl-tds** generation, enhancing both the quality and discriminative power of $\mathcal{D}_\mathcal{T}$ representations for effective DA. *In this framework, the superscript* (t) *denotes the diffusion model's time step, and the subscript* $\mathcal{T}$ *represents the target domain.*

Forward Diffusion Enabled Robust Cross-Domain Classifier Training: The forward diffusion process in NOCDDA incorporates two levels of constraints to achieve robust cross-domain classifier optimization. **Unified Optimization Framework:** By integrating the DA classifier with the conditional diffusion model's classifier, a shared optimization framework is established, promoting invariant feature extraction and synergistic decision-making. **Noise-Enhanced Training:** The shared classifier is trained on sequentially noised samples from $\mathcal{D}_\mathcal{S}$ and hcpl-tds, improving sensitivity to noise variations and adaptability in cross-domain scenarios. This dual-constraint strategy ensures that forward diffusion training not only reinforces cross-domain alignment but also enhances decision-making across diverse noise conditions, resulting in a robust and unified DA framework.

Class-Specific Noise Optimization in Reverse Diffusion:

In the reverse diffusion phase, traditional diffusion models initialize noise from $P(\mathbf{x}^{(T)}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$, assuming an infinite forward diffusion process ($T \to \infty$). However, real-world tasks with finite diffusion steps ($T < \infty$) yield terminal distributions $P(\mathbf{x}^{(T)}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ represent the mean and covariance of noisy samples. To address this mismatch and meet the class-specific needs of domain adaptation (DA), we model class-specific terminal distributions tailored to hcpl-tds subsets. For each class $c \in \mathcal{C}_\mathcal{T}$, the terminal distribution is defined as: $P_c(\mathbf{x}^{(T)}) = \mathcal{N}(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c), \boldsymbol{\mu}_c = \mathbb{E}[\mathbf{x}_c^{(T)}], \boldsymbol{\Sigma}_c = \mathrm{Cov}(\mathbf{x}_c^{(T)})$, where $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$ are computed from hcpl-tds for class c. This tailored initialization ensures reverse sampling trajectories align with class-aware hcpl-tds distributions, facilitating discriminative sample generation for robust cross-domain decision-making. Detailed visualizations in the supplementary material (Appendix.1) validate the effectiveness of this approach across various diffusion models, illustrating how noise optimization significantly enhances class-specific sample generation. These findings highlight the critical role of tailored noise sampling in improving the efficiency and effectiveness of DA.

The key contributions are summarized as follows:

- **Forward Diffusion Enhanced Robust Classifier Training:** During the forward diffusion training process in the NOCDDA model, the DA classifier is modified and integrated with the diffusion model's classifier alongside a unified optimization framework. This coupling promotes robust cross-domain decision-making by enabling the DA classifier to distinguish cross-domain samples with varying noise intensities. Such integration effectively extracts invariant domain features, reinforcing domain alignment and further improving the model's robustness.
- **Noise-Optimized Backward Sampling for Class-Specific hcpl-tds Generation:** Traditional diffusion models initialize reverse sampling with a uniform Gaussian prior $\mathcal{N}(\mathbf{0}, \mathbf{I})$, requiring an infinite forward diffusion process ($T \to \infty$), which is impractical for real-world tasks. This approach fails to meet the class-aware sample generation needs of DA. To address this, we propose a noise-optimized strategy, modeling class-specific terminal distributions $P_c(\mathbf{x}^{(T)}) = \mathcal{N}(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$, where $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$ are computed from hcpl-tds under finite diffusion steps. This tailored initialization ensures reverse sampling aligns with hcpl-tds distributions, enabling class-aware generation and improving cross-domain decision-making in DA.
- We conducted extensive experiments on 15 image classification and 14 time series domain adaptation tasks across 31 comparison methods, showcasing the superior performance of our proposed NOCDDA method over state-of-the-art DA methods.

2 Related Work

Existing domain adaptation techniques are categorized into shallow and deep approaches. Given recent trends, this chapter focuses on deep DA methods to comparative analyze the rationale behind our proposed NOCDDA approach.

Traditional deep-DA techniques are generally classified into two paradigms: **non-adversarial** and **adversarial** methods. Non-adversarial methods rely on predefined statistical metrics such as Wasserstein distance [Tang *et al.*, 2024], frequency spectrum [Yang and Soatto, 2020], and Maximum Mean Discrepancy (MMD) [Luo *et al.*, 2024b; Long *et al.*, 2018a] to quantify and minimize domain shift by reducing divergence in kernel feature spaces [Shu *et al.*, 2024]. These methods depend on the careful selection and design of these metrics. In contrast, adversarial learning-based DA methods use adversarial training [Goodfellow *et al.*, 2014] to align source and target domain features by making them indistinguishable through domain classifier training [Chen *et al.*, 2023;

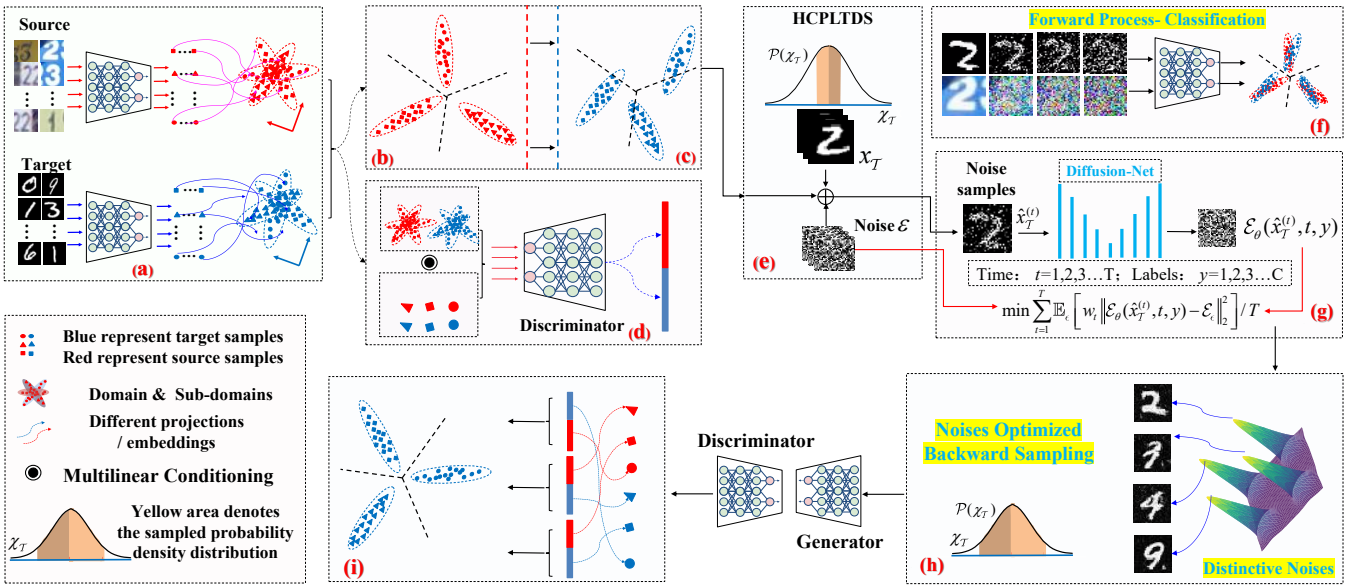

Figure 1: Illustration of the proposed **NOCDDA** method. Fig.1(a) shows source (red) and target (blue) data with distinct distributions and geometric structures, where shapes indicate class labels. Fig.1(b&c) depicts generator training on labeled source data and pseudo-label prediction for the target domain. Fig.1(d) aligns feature and label spaces via conditional adversarial learning. hcpl-tds selection is highlighted in Fig.1(e), followed by robust cross-domain classifier training under forward noise perturbation in Fig.1(f). Fig.1(g&h) illustrates class-specific backward sampling and enriched target sample generation. Fig.1(i) shows enhanced DA performance via improved target representation.

Jing *et al.*, 2024]. While this data-driven approach offers flexibility by removing the need for predefined metrics, it often results in unstable training due to the lack of explicit guidance for optimizing likelihood distributions [Bond-Taylor *et al.*, 2021; Yang *et al.*, 2022]. These DA techniques often focus on optimizing the mapping from samples to label space ($f_d(\mathbf{x}, \theta) \mapsto p(\mathbf{c})$), thus primarily operating as discriminative models.

In contrast to traditional discriminative models, which focus solely on inferring labels $\mathbf{c}$, generative models ($f_g(\theta) \mapsto p(\mathbf{x}, \mathbf{c})$) have recently gained prominence in the DA field as an alternative approach. The key advantage of generative models lies in their ability to jointly model both the label distribution $p(\mathbf{c})$ and the data distribution $p(\mathbf{x})$ during the training process, preserving critical structural information within feature representations. This regularized functional learning framework [Yang *et al.*, 2022; Jiang *et al.*, 2017] enhances the robustness and reliability of generative approaches for DA. For example, Zhang *et al.* [Zhang *et al.*, 2020] framed DA as a Bayesian inference problem in probabilistic graphical models, while VDI [Xu *et al.*, 2023] extended this framework by incorporating domain indices into a variational Bayesian approach for multi-domain data. Recently, diffusion generative models [Song *et al.*, 2020b; Ho *et al.*, 2020; Lipman *et al.*, 2022], leveraging Markov process-based sequential optimization, have emerged as a promising tool for DA, refining the joint probability distribution function $p_\theta(\mathbf{x}_\mathcal{T}, \mathbf{x}_\mathcal{S})$ through reverse diffusion, facilitating smooth domain transitions and robust alignment.

Building on diffusion models, DAD [Peng *et al.*, 2024] bridges source and target domains by simulating a sequence of distributions with minimal discrepancies between adjacent steps, achieving seamless cross-domain blending. TransDiff [Kang *et al.*, 2024] uses cross-domain prompts to generate semantically-aware pseudo-target domain images, significantly improving domain adaptation performance. DDA [Gao *et al.*, 2023] leverages a diffusion model to minimize domain divergence by aligning data structures in noisy latent spaces. Subsequently, Kamil *et al.* [Deja *et al.*, 2023] enhanced classifier guidance to improve conditional distribution-aware domain adaptation. Prof. Li *et al.* [Du and Li, 2024] further advanced cross-domain alignment through diffusion-based purification of target domain labels.

Existing diffusion models have advanced domain adaptation (DA) through sample augmentation and label optimization but often fail to integrate DA decision-making with generative objectives effectively. To address this limitation, we propose NOCDDA, a unified framework that seamlessly combines conditional diffusion generation with DA-specific objectives. In the forward diffusion phase, the joint classifier is trained on sequentially noised cross-domain data, enhancing its robustness and adaptability to domain discrepancies. In the reverse generation phase, NOCDDA incorporates noise-optimized sampling strategies, enabling class-specific hcpl-tds generation that aligns with cross-domain decision-making requirements. By unifying robust decision-making with class-aware sample augmentation, NOCDDA not only improves cross-domain alignment but also establishes a strong foundation for effective and discriminative DA performance.

3 Noise Optimized Conditional Diffusion Motivated Domain Adaptation

We begin by defining the notations used in this paper, followed by a clear explanation of the research motivation and a step-

by-step breakdown of the proposed **NOCDDA** approach.

3.1 Notations and Problem Statement

We use the cursive symbols $\mathcal{S}$ and $\mathcal{T}$ to represent the *source* or *target* domain, respectively. Matrices are denoted by boldface uppercase letters, *e.g.* $\mathbf{M} = (m_{ij})$, whose i-th row is $\mathbf{m}^i$ and j-th column is $\mathbf{m}_j$. Vectors are denoted by boldface lowercase letters, *e.g.* $\mathbf{x}$. A domain $\mathcal{D}$ is defined as an ℓ-dimensional feature space X together with a marginal distribution $P(\mathbf{x})$, *i.e.* $\mathcal{D} = \{X, P(\mathbf{x})\}$ with $\mathbf{x} \in X$. Given a domain $\mathcal{D}$, a task $\mathcal{TK}$ consists of a C-cardinality label set $\mathcal{Y}$ and a classifier $f(\mathbf{x})$, *i.e.* $\mathcal{TK} = \{\mathcal{Y}, f(\mathbf{x})\}$, where $f(\mathbf{x}) = \mathcal{Q}(y \mid \mathbf{x})$ interprets the class-conditional distribution over $\mathcal{Y}$. In unsupervised domain adaptation, we have a *source domain* $\mathcal{D}_\mathcal{S} = \{\mathbf{x}^i_\mathcal{S}, \mathbf{y}^i_\mathcal{S}\}_{i=1}^{n_s}$ with n_s labeled samples $\mathbf{X}_\mathcal{S} = [\mathbf{x}^1_\mathcal{S} \ldots \mathbf{x}^{n_s}_\mathcal{S}]$ and labels $\mathbf{Y}_\mathcal{S} = \{\mathbf{y}^1_\mathcal{S}, \ldots, \mathbf{y}^{n_s}_\mathcal{S}\}^\top \in \mathbb{R}^{n_s \times C}$, as well as an *unlabeled target domain* $\mathcal{D}_\mathcal{T} = \{\mathbf{x}^j_\mathcal{T}\}_{j=1}^{n_t}$ with n_t unlabeled samples $\mathbf{X}_\mathcal{T} = [\mathbf{x}^1_\mathcal{T} \ldots \mathbf{x}^{n_t}_\mathcal{T}]$ and unknown labels $\mathbf{Y}_\mathcal{T} = \{\mathbf{y}^1_\mathcal{T}, \ldots, \mathbf{y}^{n_t}_\mathcal{T}\}^\top \in \mathbb{R}^{n_t \times C}$. Although $X_\mathcal{S} = X_\mathcal{T}$ and $\mathcal{Y}_\mathcal{S} = \mathcal{Y}_\mathcal{T}$, the distributions and conditionals differ, *i.e.* $P(X_\mathcal{S}) \neq P(X_\mathcal{T})$, $\mathcal{Q}(\mathcal{Y}_\mathcal{S} \mid X_\mathcal{S}) \neq \mathcal{Q}(\mathcal{Y}_t \mid X_\mathcal{T})$. We define a *sub-domain* (class) in the source domain by $\mathcal{D}^{(c)}_\mathcal{S} = \{\mathbf{x}^i_\mathcal{S} \in \mathbf{X}_\mathcal{S} \mid \mathbf{y}^i_\mathcal{S} = c\}$. Sub-domains in the target domain, $\mathcal{D}^{(c)}_\mathcal{T}$, require a base classifier to assign pseudo labels to unlabeled samples in $\mathcal{D}_\mathcal{T}$. To distinguish the diffusion model's time step, we use the superscript (t) for time encoding and the subscript $\mathcal{T}$ to denote the target domain.

Time Embedding Aware Classifier Unification

As shown in Fig.1.b, to minimize **Term.1** in Eq.(1) for the source domain's Structural Risk Minimization, a well-trained classifier must be learned in a supervised manner. The hypothesis $f : \mathcal{X}_\mathcal{S} \to \mathcal{Y}_\mathcal{S}$ maps source domain samples $\mathcal{X}_\mathcal{S}$ to their labels $\mathcal{Y}_\mathcal{S}$. However, such a hypothesis trained solely on this mapping fails to align with the classification decisions required in conditional diffusion training, which demands incorporating a time embedding (t) to capture the temporal dynamics of the diffusion process. To address this, the predictive model is formulated as:

$$f = \arg\min_{f \in \mathcal{H}_\mathcal{K}} \mathbb{E}_{(x_\mathcal{S}, y_\mathcal{S}) \sim \mathcal{D}_\mathcal{S}} \Big[l(f(x_\mathcal{S}, (t)), y_\mathcal{S}) \Big] + \mathcal{R}(f). \quad (2)$$

Here, $f(x_\mathcal{S}, (t))$ integrates the time embedding (t), aligning the classifier's decisions with the conditional diffusion framework. For clean source domain images, (t) is set to 0, ensuring consistency with the DA framework and maintaining alignment for noise-free data. This incorporation of time embedding within the DA classifier represents a key innovation, enabling unified decision-making across both DA and conditional diffusion training. The supervised loss $l(f(x_\mathcal{S}, (t)), y_\mathcal{S})$ ensures accurate learning on the source domain, while the regularization term $\mathcal{R}(f)$ controls the complexity of the hypothesis space $\mathcal{H}_\mathcal{K}$.

Despite the implemented setup, cross-domain distribution divergence often leads to unreliable mappings for the target domain, $f : (\mathbf{x}_\mathcal{T}, t) \mapsto \mathbf{y}_\mathcal{T}$. This underscores the need for effective strategies to minimize conditional distribution discrepancies and ensure efficient domain adaptation (DA).

Adversarial Learning Enforced Domain Alignment

Despite Eq.(2) effectively reducing the classification error on the $\mathcal{D}_\mathcal{S}$ to minimize **Term.1** in Eq.(1), the learned hypothesis struggles to generalize to the $\mathcal{D}_\mathcal{T}$ due to domain shift. To address this, we focus on optimizing **Term.2** by minimizing the cross-domain conditional distribution divergence. This is achieved by incorporating conditional adversarial training to align the hybrid conditional distributions of feature representations f and labels y, as illustrated in Fig.1.(a-d). Specifically, Fig.1.(d) shows how cross-domain divergence is reduced through adversarial optimization of the **generator** (G) and **discriminator** (D). The updated objective is formulated as:

$$\begin{aligned} \min_G \max_D L_{adv} = -\mathbb{E}[\sum_{c=1}^{C} \mathbb{1}_{[y_s=c]} \log \sigma(G(x_\mathcal{S}, (t))] \\ +\mathbb{E}[\log D((f \otimes y)_\mathcal{S})] + \mathbb{E}[\log(1 - D((f \otimes y)_\mathcal{T}))] \end{aligned} \quad (3)$$

In the RHS of Eq.(3), the first term reformulates Eq.(2), while the remaining two terms optimize G and D within an adversarial framework to achieve model equilibrium. The symbol $\otimes$ represents the multilinear conditioning operation [Long *et al.*, 2018b], which leverages the f and y inferred by G to perform kernel embedding for joint distribution alignment. This minimizes the conditional divergence between the source $(\mathcal{X}_\mathcal{S}, \mathcal{Y}_\mathcal{S})$ and target $(\mathcal{X}_\mathcal{T}, \mathcal{Y}_\mathcal{T})$ domains, facilitating smoother cross-domain knowledge transfer. However, conditional adversarial methods depend on High-Confidence Pseudo-labeled Target Domain Samples (hcpl-tds). In real-world settings, as shown in Fig.1.(c), target domain pseudo-labels are often noisy, leading to functional training errors. To address this, we quantify classifier prediction uncertainty using the entropy criterion: $E(\mathbf{f}) = -\sum_{c=1}^{C} f_c \log f_c$, where C is the number of classes, and f_c is the probability of assigning a sample to class c. Samples with lower entropy are selected as hcpl-tds, thereby enhancing the reliability of training.

As discussed in Sec. (1), the limited availability of hcpl-tds fails to sufficiently represent the target domain's distribution or cover its discriminative decision space, hindering cross-domain alignment. To address this, we integrate conditional diffusion techniques to infer intrinsic distributions and generate diverse samples, thereby improving the target domain's representation and decision-making capacity, ultimately enabling robust and effective cross-domain adaptation.

Forward Diffusion Enabled Robust Cross-Domain Classifier Training

Given the limited number of hcpl-tds, directly estimating and sampling from their probability density function is impractical. To overcome this, we employ conditional diffusion techniques with label guidance to infer the target domain's distribution. Specifically, our approach uses a forward diffusion process based on the DDPM framework [Ho *et al.*, 2020], where noise is progressively added to the input data, training both the denoising model and the classifier. It is worth noting that the classifier $f_\phi(\mathbf{x}, (t))$ is trained with ***two levels of consistency constraints***, optimizing the model specifically for DA tasks.

1. Consistency Training With Noised Source and hcpl-tds Samples: To facilitate target-domain sample generation through conditional diffusion, our classifier f_ϕ learns from forward diffusion classification on hcpl-tds. To further enhance

training, we also incorporate noised source-domain data, imposing *consistency constraints* both *across domains* (source vs. target) and *across time* (multiple noise levels). Specifically, let $\mathbf{x}^{(0)}$ be a clean sample (at time step 0) drawn from the joint dataset $\mathcal{D}_{\text{joint}}$, which includes both source-domain data and hcpl-tds. We then minimize the following objective:

$$\min_{\phi} \quad \mathbb{E}_{\mathbf{x}^{(0)}\sim\mathcal{D}_{\text{joint}},\, t\sim\mathcal{U}(1,T)}\Big[\|f_\phi(\mathbf{x}_{\mathcal{S}}^{(t)},(t)) - y_{\mathcal{S}}\|^2 + \|f_\phi(\mathbf{x}_{\mathcal{T}}^{(t)},(t)) - y_{\mathcal{T}}\|^2\Big], \quad (4)$$

where $\mathbf{x}_{\mathcal{S}}^{(t)}$ and $\mathbf{x}_{\mathcal{T}}^{(t)}$ represent sequentially noised versions of the source-domain sample $\mathbf{x}_{\mathcal{S}}^{(0)}$ and hcpl-tds $\mathbf{x}_{\mathcal{T}}^{(0)}$, respectively. By exposing the classifier f_ϕ to varying noise intensities (*i.e.*, different time steps t) for both domains, this training process enforces *cross-time consistency* to improve robustness against noise, and *cross-domain consistency* to facilitate domain-invariant feature learning. Consequently, the trained model achieves improved cross-domain generalization, providing a solid foundation for robust classifier training and adaptation to distribution shifts.

2. Consistency Between the Diffusion-Based Classifier and the DA Classifier: As discussed in Eq.(2), the DA classifier is modified to share the same model parameters ϕ as the diffusion-based classifier. Specifically, let $f_\phi\colon (\mathbf{x}, t) \longmapsto \hat{y}$, where $\hat{y}$ represents the predicted label (or label distribution) and $t \in \{0, 1, \ldots, T\}$ denotes the diffusion time step. The DA classifier corresponds to $f_\phi(\mathbf{x}^{(0)}, 0)$, which operates on clean samples $\mathbf{x}^{(0)}$ (*i.e.*, $t = 0$), while the diffusion-based classifier corresponds to $f_\phi(\mathbf{x}^{(t)}, t)$ for $t > 0$, processing noised samples $\mathbf{x}^{(t)}$ obtained from $\mathbf{x}^{(0)}$ via the forward diffusion process. Since both classifiers use the same function f_ϕ, they naturally share label supervision. Specifically, let $\{(\mathbf{x}_i^{(0)}, y_i)\}$ represent labeled samples from the joint dataset $\mathcal{D}_{\text{joint}}$. Our training objective combines losses over both clean ($t = 0$) and noised ($t > 0$) inputs:

$$\begin{aligned}\min_{\phi} \quad & \underbrace{\mathbb{E}_{\mathbf{x}^{(0)}\sim\mathcal{D}_{\text{joint}}}\big[\|f_\phi(\mathbf{x}^{(0)}, 0) - y\|^2\big]}_{\text{DA on clean samples}} + \\ & \underbrace{\mathbb{E}_{\mathbf{x}^{(0)}\sim\mathcal{D}_{\text{joint}},\, t\sim\mathcal{U}(1,T)}\big[\|f_\phi(\mathbf{x}^{(t)}, t) - y\|^2\big]}_{\text{Diffusion on noised samples}}.\end{aligned} \quad (5)$$

Here, $\mathbf{x}^{(t)}$ is the noisy version of $\mathbf{x}^{(0)}$, and y is the same label supervision applied in both the DA (clean) and diffusion (noised) scenarios. While we do not explicitly impose a constraint like $\|f_\phi(\mathbf{x}^{(t)}, t) - f_\phi(\mathbf{x}^{(0)}, 0)\|^2$, the shared parameterization ϕ inherently promotes *temporal continuity* (across different t-values) and *mutual reinforcement* between the two tasks. Thus, improvements achieved on noised inputs ($t > 0$) transfer to the clean case ($t = 0$), and vice versa, enabling robust cross-domain alignment and enhancing overall adaptation performance.

By jointly optimizing Eq.(4) and Eq.(5), the proposed forward diffusion training ensures robust cross-domain decision alignment and temporal consistency between the DA and diffusion classifiers, thereby laying a solid foundation for effective reverse conditional sampling in hcpl-tds generation and enhanced adaptation.

Noise Optimized Backward Generation

Building on the trained denoising model and the jointly optimized classifier, we adopt a reverse conditional generation strategy inspired by BeatGAN [Dhariwal and Nichol, 2021] to enhance hcpl-tds quantity and leverage DDIM [Song *et al.*, 2020a] to reduce reverse diffusion steps. Traditional diffusion methods sample from a Gaussian prior $\mathcal{N}(\mathbf{0}, \mathbf{I})$, assuming an infinite forward diffusion process ($T \to \infty$), which is impractical for real-world tasks. Finite diffusion steps ($T < \infty$) lead to deviations in the terminal distribution, causing class overlap in generated hcpl-tds and hindering effective cross-domain decision-making (*see Appendix.1*). To address this, we model class-specific terminal distributions $\mathcal{N}(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$, where $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$ are derived from hcpl-tds under finite forward diffusion. This noise-optimized approach ensures class-aware, discriminative hcpl-tds generation, improving $\mathcal{D}_{\mathcal{T}}$ representation and cross-domain alignment for domain adaptation tasks.

1.Class-Specific Terminal Distributions: For C classes in $\mathcal{D}_{\mathcal{T}}$, we partition the hcpl-tds into subsets $\{\mathcal{D}_{\mathcal{T}}^{(1)}, \ldots, \mathcal{D}_{\mathcal{T}}^{(C)}\}$, where each subset corresponds to samples pseudo-labeled as class c. Instead of assuming a shared Gaussian distribution for all classes at $t = T$, we estimate class-specific terminal distributions as: $\mu_c = \mathbb{E}[\mathbf{x}_c^{(T)}]$, $\Sigma_c = \mathrm{Var}[\mathbf{x}_c^{(T)}]$, where $\mathbf{x}_c^{(T)} \in \mathcal{D}_{\mathcal{T}}^{(c)}$ represents the noised sample at the final forward step T. However, due to uniform noise injection during forward diffusion, Σ_c tends to be large and similar across classes, leading to overlapping distributions. To mitigate this, we apply a regularization strategy that uniformly scales down the variances as: $\Sigma_c = \frac{1}{C}\mathbf{I}$, where C is the number of classes. This adjustment minimizes inter-class overlap, enabling precise noise optimization and improving the quality of hcpl-tds generation.

2. Noise Optimized Reverse Sampling: For class-specific generation, we incorporate class-specific priors (μ_c, Σ_c) into the reverse sampling process, refining both the initialization and backward updates within the DDIM framework. Mathematically, let ϵ_θ denote the trained noise-prediction network, and $p_\phi(y \mid \mathbf{x}^{(t)})$ represent the classifier's label distribution function.

Case 1: Initialization at $t = T$ (Class-Specific Starting Point) The terminal state of the forward diffusion process at $t = T$ is modeled as a class-specific Gaussian distribution:

$$\mathbf{x}_c^{(T)} \sim \mathcal{N}(\mu_c, \Sigma_c), \quad s.t.\ \mu_c = \mathbb{E}[\mathbf{x}_c^{(T)}],\ \Sigma_c = \tfrac{1}{C}\,\mathbf{I}. \quad (6)$$

Here, μ_c represents the mean of noised samples for class c, while the scaled variance $\Sigma_c = \frac{1}{C}\,\mathbf{I}$ reduces inter-class overlap by decreasing variance as the number of classes C increases. This adjustment ensures distinct reverse sampling trajectories, avoiding the overlap that would occur with a uniform variance ($\Sigma_c = \mathbf{I}$). In practice, the initialization is performed as:

$$\mathbf{x}_c^{(T)} = \mu_c + \Sigma_c^{\frac{1}{2}}\mathbf{z}, \quad s.t.\ \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}),\ \Sigma_c = \tfrac{1}{C}\,\mathbf{I}. \quad (7)$$

This initialization anchors the reverse sampling process to the empirical distribution of hcpl-tds for each class c, ensuring the reverse trajectory begins within a class-specific region, thereby enabling class-aware hcpl-tds generation.

Case 2: Reverse Sampling for $t < T$ (Adapted from DDIM with Class-Specific Initialization). Unlike standard DDIM,

which initializes reverse diffusion with $\mathbf{x}^{(T)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, we begin from a class-specific distribution $\mathbf{x}_c^{(T)} \sim \mathcal{N}(\mu_c, \Sigma_c)$, as defined in Eq. (6)–(7). This initialization ensures that the entire reverse trajectory remains *class-aware*. At each step $t \to t-1$, the DDIM-style update is enhanced by incorporating the classifier gradient to align with target-domain decisions. Specifically, the *classifier-guided* noise term is defined as:

$$\hat{\epsilon}_c(\mathbf{x}^{(t)}) = \epsilon_\theta(\mathbf{x}^{(t)}, t) - \sqrt{1-\overline{\alpha}_t}\, \nabla_{\mathbf{x}^{(t)}} \log p_\phi(y \mid \mathbf{x}^{(t)}), \quad (8)$$

where $\epsilon_\theta(\mathbf{x}^{(t)}, t)$ predicts the noise component, α_t denotes the diffusion coefficient at time step t, and the classifier gradient term $-\nabla_{\mathbf{x}^{(t)}} \log p_\phi(y \mid \mathbf{x}^{(t)})$ steers the samples toward more discriminative regions.

Backward Update. Starting from $\mathbf{x}_c^{(T)}$ at $t = T$, the state at each subsequent step $t > 0$ is updated as:

$$\mathbf{x}_c^{(t-1)} = \sqrt{\overline{\alpha}_{t-1}}\Big(\frac{\mathbf{x}_c^{(t)} - \sqrt{1-\overline{\alpha}_t}\,\hat{\epsilon}_c(\mathbf{x}_c^{(t)})}{\sqrt{\overline{\alpha}_t}}\Big) + \sqrt{1-\overline{\alpha}_{t-1}}\,\hat{\epsilon}_c(\mathbf{x}_c^{(t)}), \quad (9)$$

where each $\mathbf{x}_c^{(t)}$ recursively depends on the previous state $\mathbf{x}_c^{(t+1)}$. Overall, this class-specific initialization and classifier-guided reverse sampling strategy ensure that the process aligns with the empirical terminal distribution of each class. This approach significantly enhances the generation of class-specific hcpl-tds, facilitating improved cross-domain separability and discriminative adaptation.

3. Merits and Discussion of Noise-Optimized Conditional Generation:

As detailed in the supplementary material (Appendix.1), we evaluate three diffusion-based approaches, namely DSM [Vincent, 2011], Consistency Distillation, and Consistency Training [Song *et al.*, 2023], to validate the effectiveness of noise optimization for class-aware sample generation. Visualized reverse sampling trajectories demonstrate that incorporating class-specific terminal distributions $\mathcal{N}(\mu_c, \frac{1}{C}\mathbf{I})$ under finite forward diffusion steps facilitates decision-boundary-aware hcpl-tds generation. Although constraining terminal noise may slightly limit generative diversity, it aligns with the fundamental objective of DA: optimizing cross-domain decision-making. To sum up, by resolving the inherent conflict between finite diffusion steps and unified noise distributions, the proposed noise-optimized strategy prioritizes classification-driven objectives, thereby enhancing cross-domain alignment and improving DA performance.

NOCDDA Motivated Domain Adaptation:
In summary, the proposed **NOCDDA** strategy is specifically tailored for DA, integrating robust cross-domain classifier optimization in the forward diffusion process with class-aware hcpl-tds generation in the reverse sampling process. By coupling conditional diffusion with DA models, NOCDDA achieves synergistic optimization, effectively enhancing cross-domain decision boundary alignment.

4 Experiments

The experimental section covers Dataset Description, Experimental Setup, 31 Baseline Methods, Experimental Results and Discussion, and includes an Ablation Study to further discuss the individual contributions of the proposed method's design.

4.1 Dateset Description

Image datasets: ***Digits***: We evaluate domain adaptation on three digit datasets: MNIST (M) with 60,000 training and 10,000 test samples, USPS (U) with 7,291 training and 2,007 test samples, and SVHN (S) with over 600,000 labeled street view digits. ***Office-31***: This benchmark dataset includes over 4,000 images across three domains: *Amazon* (A), *Webcam* (W), and *Dslr* (D), used to evaluate six transfer tasks such as *A* $\to$ *W* and *W* $\to$ *D*. ***ImageCLEF-DA***: This dataset contains 12 shared classes from *Caltech-256* (C), *ImageNet ILSVRC 2012* (I), and *Pascal VOC 2012* (P), covering six cross-domain tasks like *C* $\to$ *I* and *I* $\to$ *P*. **Time-series datasets**: ***CWRU***: The Case Western Reserve University (CWRU) dataset includes vibration data across four operating speeds (domains A, B, C, D) and 10 health states, enabling 12 cross-domain tasks. ***SEU***: The Southeast University (SEU) dataset includes bearing and gear data under two working conditions (tasks A and B) with imbalanced fault sampling, providing a challenging DA benchmark. *Detailed dataset descriptions are available in Appendix.2 of the supplementary material.*

4.2 Experimental Setup

Experiments were conducted on several datasets, including Digits, Office-31, ImageCLEF-DA, CWRU, and SEU, using the PyTorch framework and an Nvidia 4090 GPU. For the Digits dataset, the diffusion model was trained with 1000 diffusion steps using DDIM sampling (200-step schedule with 5-step jumps), a batch size of 36, a learning rate of 0.02, and momentum of 0.5 for 100 epochs. For the Office-31 and ImageCLEF-DA datasets, DDIM sampling was optimized with 50 steps (20-step jumps), generating 100 images per class, with learning rates of 0.001 (for Office-31) or 0.0003 (depending on the domain pair). For CWRU and SEU, DDIM with 10 steps (100-step jumps) was used, generating 50 samples per class, with a batch size of 64, a learning rate of 0.03, and 50 epochs for CWRU and 30 for SEU. The model employed a U-Net architecture (1D for CWRU and SEU, 2D for Office-31 and ImageCLEF-DA), incorporating encoder-decoder structures, skip connections, attention layers, and residual blocks. *For detailed settings, please refer to Appendix.3 in the supplementary material.* Model performance was evaluated on the target domain test set $\mathcal{D}_\mathcal{T}$ using accuracy, as defined in Eq. (10), a widely-used metric in prior research [Long *et al.*, 2018b; Li *et al.*, 2024].

$$Accuracy = \frac{|x \in \mathcal{D}_\mathcal{T} \wedge \hat{y}(x) = y(x)|}{|x \in \mathcal{D}_\mathcal{T}|}, \quad (10)$$

where $\hat{y}(x)$ is the predicted label, and $y(x)$ is the ground truth label for test data x.

4.3 Baseline Methods

The developed NOCDDA method was evaluated against **31** established DA techniques, including **DOLL** [Luo *et al.*, 2024b], **TLSR** [Luo *et al.*, 2024a], **MCADA** [Chen *et al.*, 2023], **CaCo** [Huang *et al.*, 2022], **SUDA** [Zhang *et al.*, 2022], **SATLN** [Qin *et al.*, 2023], **PfAReLU** [Chen *et al.*, 2024], **ATM** [Li *et al.*, 2020], **TPN** [Pan *et al.*, 2019], **BSP** [Chen *et al.*, 2019], **CAT** [Deng *et al.*, 2019], **MSTN** [Xie *et al.*,

2018], **MCD** [Saito *et al.*, 2018], **MADA** [Pei *et al.*, 2018], **GTA** [Sankaranarayanan *et al.*, 2018], **CDAN** [Long *et al.*, 2018b], **CyCADA** [Hoffman *et al.*, 2018], **UNIT**, **ADDA**, **JAN**, **DRCN**, **CoGAN**, **DANN**, **DAN**, **DDC**, **D-CORAL**, **RTN**, **MK-MMD**, **JMMD**, as well as baseline models such as **ResNet** and **CNN**. *For full references of the compared models, please refer to Appendix.4 of the supplementary material.* Performance metrics for these methods were obtained from their original publications or subsequent analyses [Luo *et al.*, 2024b; Chen *et al.*, 2024], reflecting their reported best results.

4.4 Experimental Results and Discussion

Digits: Table.1 shows the DA results on the digit datasets, with the best results highlighted in red. Among existing methods, ATM [Li *et al.*, 2020], which builds on CDAN, achieves a second-best accuracy of 97.1% by improving domain separation. In comparison, NOCDDA achieves the highest accuracy of 99.3% by coupling the forward diffusion process with DA decision-making, ensuring robust cross-domain consistency. The reverse generation process further enhances the target domain's statistical representation through optimized noise sampling. **Office-31**: Table.2 compares popular DA methods on the Office-31 dataset, which includes 31 office-related categories. Using ResNet as a baseline, methods like ADDA align features through adversarial learning, while MADA refines decision boundaries with multiple discriminators. In contrast, NOCDDA optimizes cross-domain representations through dual-classifier consistency and noise-optimized reverse generation, achieving the highest average accuracy of 89.4%. *Experimental results and detailed discussions for **ImageCLEF-DA** and cross-domain time-series datasets **SEU** and **CWRU** are provided in Appendix.5, 6 & 7, respectively.*

Method	DAN	DRCN	CDAN	ADDA	DOLL	MCD	CAT	TPN	BSP	GTA	CyCADA	UNIT	MSTN	ATM	NOCDDA
M→U	80.3	91.8	95.6	89.4	97.2	94.2	90.6	92.1	93.3	92.8	95.6	96.0	92.9	96.1	**99.2**
U→M	77.8	73.7	98.0	90.1	86.9	94.1	80.9	94.1	94.5	90.8	96.5	93.0	–	99.0	**99.7**
S→M	73.5	82.0	89.2	76.0	87.8	96.2	98.1	93.0	91.4	92.4	90.4	90.5	91.7	96.1	**98.9**
Average	77.2	82.5	94.3	85.2	90.6	94.8	89.9	93.1	93.1	92.0	94.2	93.2	–	97.1	**99.3**

Table 1: Classification Accuracy (%) on Digits Datasets.

4.5 Ablation Study

Although the NOCDDA method achieves state-of-the-art performance across 29 DA tasks in 5 datasets, outperforming 31 existing methods, its success hinges on three key components: *hcpl-tds selection*, *classifier unification-enhanced forward training*, and *noise-optimized reverse generation*. To evaluate the contributions of each component, an ablation study was conducted on the *SVHN-to-MNIST* DA task. As shown in Table. 3, the vertical axis, 'X% TDS', represents the proportion of high-confidence, low-entropy target domain samples, while the horizontal axis indicates the number of generated samples (*e.g.*, $G6000$ corresponds to 6000 samples generated using the diffusion model). The $G6000{+}CU$ model integrates classifier unification in the forward diffusion phase, while NOCDDA includes noise optimization during reverse sampling. Key observations include: **1. Full TDS Utilization** ($TDS = 100\%$): Utilizing all target domain samples achieves a baseline accuracy of 93.1%, demonstrating the importance of TDS for effective $\mathcal{D}_{\mathcal{T}}$ modeling. However, doubling the sample size from $TDS = 50\%$ to $TDS = 100\%$ yields only a marginal 0.4% improvement, revealing the limited contribution of low-confidence samples. This underscores the critical role of hcpl-tds in enhancing discriminative DA by prioritizing high-confidence target samples for cross-domain alignment. **2. Impact of Generated Samples**: At $TDS = 10\%$, generating 6000 additional samples (G_{6000}) boosts baseline accuracy by $+13.7\%$ (from 78.7% to 92.4%), demonstrating the importance of hcpl-tds generation in enhancing data representation for discriminative DA. **3. Joint Optimization** ($G_{6000} + CU$): Incorporating classifier unification (CU) into the forward diffusion training further improves performance. For example, at $TDS = 50\%$, accuracy increases from 97.2% (G_{6000}) to 98.4% ($G_{6000} + CU$), validating the efficacy of coupling DA classification with the generative model's conditional diffusion. **4. Noise Optimization**: **NOCDDA** achieves the highest accuracy of 99.1% at $TDS = 50\%$, demonstrating that noise-optimized sampling ($\mathcal{N}(\mu_c, \Sigma_c)$) effectively enhances class-specific hcpl-tds generation. This further refines cross-domain decision-making and underscores the importance of noise optimization in DA tasks.

Method	A→W	D→W	W→D	A→D	D→A	W→A	Average
ResNet-50	68.4	96.7	99.3	68.9	62.5	60.7	76.1
DAN	80.5	97.1	99.6	78.6	63.6	62.8	80.4
RTN	84.5	96.8	99.4	77.5	66.2	64.8	81.6
DANN	82.0	96.9	99.1	79.7	68.2	67.4	82.2
ADDA	86.2	96.2	98.4	77.8	69.5	68.9	82.9
JAN	85.4	97.4	99.8	84.7	68.6	70.0	84.3
GTA	89.5	97.9	99.8	87.7	72.8	71.4	86.5
CDAN	94.1	98.6	**100.0**	92.9	71.0	69.3	87.7
DDC	75.8	95.0	98.2	77.5	67.4	64.0	79.7
D-CORAL	77.7	97.6	99.7	81.1	64.6	64.0	80.8
MADA	90.0	97.4	99.6	87.8	70.3	66.4	85.3
SWD	90.4	98.7	**100.0**	**94.7**	70.3	70.5	87.4
CaCo	89.7	98.4	**100.0**	91.7	73.1	72.8	87.6
SUDA	90.8	98.7	**100.0**	91.2	72.2	71.4	87.4
NOCDDA	**95.1**	**99.3**	**100.0**	94.3	**74.5**	**72.9**	**89.4**

Table 2: Accuracy (%) on Office-31 Dataset

TDS Ratio	S→M	G 2000	G 6000	G 6000 + CU	NOCDDA
10% TDS	78.7	83.1	92.4	93.3	95.0
25% TDS	89.6	93.9	96.3	97.6	98.9
50% TDS	92.7	94.0	97.2	98.4	**99.1**
100% TDS	93.1	94.1	94.2	94.9	95.1

Table 3: Ablation Study on Digital Image Datasets

5 Conclusion

In this research, we propose Noise Optimized Conditional Diffusion for Domain Adaptation (NOCDDA), a novel approach to enhancing cross-domain functional learning. Unlike traditional diffusion-based DA methods, NOCDDA seamlessly integrates DA decision-making with the generative process of conditional diffusion, fostering mutual reinforcement between the two tasks. During the forward diffusion phase, the classifier is unified with the DA decision-maker, enabling robust cross-domain decision training. In the reverse sampling phase, noise optimization ensures the generation of class-specific hcpl-tds, enhancing discriminative performance. Experimental results on five benchmark datasets, along with comparisons against 31 state-of-the-art methods, validate the effectiveness and competitiveness of NOCDDA in DA tasks.

Acknowledgments

This work has been accepted by the International Joint Conference on Artificial Intelligence (IJCAI 2025). This work was supported by the National Natural Science Foundation of China(62006152), Natural Science Foundation of Shanghai (24ZR1434400), and China Aviation Science Foundation (2022Z071057002). Prof.Liming Chen in this work was in part supported by the French Research Agency, l'Agence Nationale de Recherche (ANR), through the projects Chiron (ANR-20-IADJ-0001-01), Aristotle (ANR-21-FAI1-0009-01), and Astérix (ANR-23-EDIA-0002) as well as the French national investment prioritary program PSPC FAIR WASTE project.

References

[Ben-David *et al.*, 2010] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79(1):151–175, 2010.

[Bond-Taylor *et al.*, 2021] Sam Bond-Taylor, Adam Leach, Yang Long, and Chris G Willcocks. Deep generative modelling: A comparative review of vaes, gans, normalizing flows, energy-based and autoregressive models. *IEEE transactions on pattern analysis and machine intelligence*, 2021.

[Chen *et al.*, 2019] Xinyang Chen, Sinan Wang, Mingsheng Long, and Jianmin Wang. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. In *International conference on machine learning*, pages 1081–1090. PMLR, 2019.

[Chen *et al.*, 2023] Haotian Chen, Yonghui Xu, Huanhuan Chen, Yong Liu, Haishu Tan, Yuguang Yan, Han Yu, et al. Multicomponent adversarial domain adaptation: A general framework. *IEEE Transactions on Neural Networks and Learning Systems*, 34(10):6824–6838, 2023.

[Chen *et al.*, 2024] Yongyi Chen, Dan Zhang, and Ruqiang Yan. Domain adaptation networks with parameter-free adaptively rectified linear units for fault diagnosis under variable operating conditions. *IEEE Trans. Neural Networks Learn. Syst.*, 35(11):16872–16885, 2024.

[Deja *et al.*, 2023] Kamil Deja, Tomasz Trzciński, and Jakub M Tomczak. Learning data representations with joint diffusion models. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 543–559. Springer, 2023.

[Deng *et al.*, 2019] Zhijie Deng, Yucen Luo, and Jun Zhu. Cluster alignment with a teacher for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9944–9953, 2019.

[Dhariwal and Nichol, 2021] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

[Du and Li, 2024] Zhekai Du and Jingjing Li. Diffusion-based probabilistic uncertainty estimation for active domain adaptation. *Advances in Neural Information Processing Systems*, 36, 2024.

[Gao *et al.*, 2023] Jin Gao, Jialing Zhang, Xihui Liu, Trevor Darrell, Evan Shelhamer, and Dequan 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*, pages 11786–11796, 2023.

[Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[Ho *et al.*, 2020] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[Hoffman *et al.*, 2018] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. CyCADA: Cycle-consistent adversarial domain adaptation. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1989–1998, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018. PMLR.

[Huang *et al.*, 2022] Jiaxing Huang, Dayan Guan, Aoran Xiao, Shijian Lu, and Ling Shao. Category contrast for unsupervised domain adaptation in visual tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1203–1214, 2022.

[Jiang *et al.*, 2017] Zhuxi Jiang, Yin Zheng, Huachun Tan, Bangsheng Tang, and Hanning Zhou. Variational deep embedding: An unsupervised and generative approach to clustering. In Carles Sierra, editor, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 1965–1972. ijcai.org, 2017.

[Jing *et al.*, 2024] Mengmeng Jing, Jingjing Li, Ke Lu, Lei Zhu, and Heng Tao Shen. Visually source-free domain adaptation via adversarial style matching. *IEEE Transactions on Image Processing*, 2024.

[Kang *et al.*, 2024] Yuhan Kang, Jie Wu, Qiang Liu, Jun Yue, and Leyuan Fang. Trans-diff: Heterogeneous domain adaptation for remote sensing segmentation with transfer diffusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024.

[Kifer *et al.*, 2004] Daniel Kifer, Shai Ben-David, and Johannes Gehrke. Detecting change in data streams. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*, pages 180–191. VLDB Endowment, 2004.

[Li *et al.*, 2020] Jingjing Li, Erpeng Chen, Zhengming Ding, Lei Zhu, Ke Lu, and Heng Tao Shen. Maximum density divergence for domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 43(11):3918–3930, 2020.

[Li *et al.*, 2024] Shugang Li, Renhu Bu, Shuang Li, Chi Harold Liu, and Keke Huang. Principal properties attention matching for partial domain adaptation in fault

diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 73:1–12, 2024.

[Lipman *et al.*, 2022] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

[Long *et al.*, 2018a] Mingsheng Long, Yue Cao, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Transferable representation learning with deep adaptation networks. *IEEE transactions on pattern analysis and machine intelligence*, 41(12):3071–3085, 2018.

[Long *et al.*, 2018b] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. *Advances in neural information processing systems*, 31, 2018.

[Lu *et al.*, 2020] Ying Lu, Lingkun Luo, Di Huang, Yunhong Wang, and Liming Chen. Knowledge transfer in vision recognition: A survey. *ACM Comput. Surv.*, 53(2):37:1–37:35, 2020.

[Luo *et al.*, 2024a] Hao Luo, Zhiqiang Tian, Kaibing Zhang, Guofa Wang, and Shaoyi Du. Semi-supervised domain adaptation via joint transductive and inductive subspace learning. *IEEE Transactions on Multimedia*, 2024.

[Luo *et al.*, 2024b] Lingkun Luo, Shiqiang Hu, and Liming Chen. Discriminative noise robust sparse orthogonal label regression-based domain adaptation. *International Journal of Computer Vision*, 132(1):161–184, 2024.

[Pan and Yang, 2010] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2010.

[Pan *et al.*, 2019] Yingwei Pan, Ting Yao, Yehao Li, Yu Wang, Chong-Wah Ngo, and Tao Mei. Transferrable prototypical networks for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2239–2247, 2019.

[Pei *et al.*, 2018] Zhongyi Pei, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Multi-adversarial domain adaptation. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[Peng *et al.*, 2024] Duo Peng, Qiuhong Ke, ArulMurugan Ambikapathi, Yasin Yazici, Yinjie Lei, and Jun Liu. Unsupervised domain adaptation via domain-adaptive diffusion. *IEEE Transactions on Image Processing*, 2024.

[Qin *et al.*, 2023] Yi Qin, Quan Qian, Jun Luo, and Huayan Pu. Deep joint distribution alignment: A novel enhanced-domain adaptation mechanism for fault transfer diagnosis. *IEEE Trans. Cybern.*, 53(5):3128–3138, 2023.

[Saito *et al.*, 2018] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3723–3732, 2018.

[Sankaranarayanan *et al.*, 2018] Swami Sankaranarayanan, Yogesh Balaji, Carlos D Castillo, and Rama Chellappa. Generate to adapt: Aligning domains using generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8503–8512, 2018.

[Shu *et al.*, 2024] Qin Shu, Fode Zhang, Lijuan Shen, and Hon Keung Tony Ng. Rul prediction with cross-domain adaptation based on reproducing kernel hilbert space. *IEEE Transactions on Reliability*, 2024.

[Song *et al.*, 2020a] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

[Song *et al.*, 2020b] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

[Song *et al.*, 2023] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. *arXiv preprint arXiv:2303.01469*, 2023.

[Tang *et al.*, 2024] Song Tang, An Chang, Fabian Zhang, Xiatian Zhu, Mao Ye, and Changshui Zhang. Source-free domain adaptation via target prediction distribution searching. *International journal of computer vision*, 132(3):654–672, 2024.

[Vincent, 2011] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011.

[Xie *et al.*, 2018] Shaoan Xie, Zibin Zheng, Liang Chen, and Chuan Chen. Learning semantic representations for unsupervised domain adaptation. In *International conference on machine learning*, pages 5423–5432. PMLR, 2018.

[Xu *et al.*, 2023] Zihao Xu, Guang-Yuan Hao, Hao He, and Hao Wang. Domain-indexing variational bayes: Interpretable domain index for domain adaptation. In *The Eleventh International Conference on Learning Representations*, 2023.

[Yang and Soatto, 2020] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4085–4095, 2020.

[Yang *et al.*, 2022] Xiangli Yang, Zixing Song, Irwin King, and Zenglin Xu. A survey on deep semi-supervised learning. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[Zhang *et al.*, 2020] Kun Zhang, Mingming Gong, Petar Stojanov, Biwei Huang, Qingsong Liu, and Clark Glymour. Domain adaptation as a problem of inference on graphical models. *Advances in neural information processing systems*, 33:4965–4976, 2020.

[Zhang *et al.*, 2022] Jingyi Zhang, Jiaxing Huang, Zichen Tian, and Shijian Lu. Spectral unsupervised domain adaptation for visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9829–9840, 2022.