MNP-GDA: MULTIMODAL NORMALITY PROTOTYPICAL GRAPH DOMAIN ADAPTATION FOR MULTIVARIATE TIME SERIES ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Traditional multivariate time series anomaly detection (MTS-AD) methods are primarily developed under the single-domain assumption that both training and test data originate from the same domain, while this rarely holds in real-world applications. The newly emerging cross-domain MTS-AD problem aims to transfer knowledge from a labeled source domain to assist MTS-AD in an unlabeled target domain. However, how to explicitly model spatio-temporal dependencies and transfer multimodal normalities across heterogeneous MTS domains remains largely under-explored. To address this, we propose a Multimodal Normality Prototypical Graph Domain Adaptation (MNP-GDA) framework that captures latent spatio-temporal dependencies by adaptive graph modeling. Furthermore, MNP-GDA introduces a prototypical domain adaptation module with theoretical justification, which models and transfers diverse normal modes across domains upon multiple normal prototypes, by compacting normal samples around their nearest prototypes while separating anomalies from all prototypes. Finally, a label-conditioned adaptive reliability weighting mechanism supported by theoretical analysis is devised to weight pseudo-normal and -anomalous target samples oppositely, mitigating negative influence of unreliable pseudo-labels during domain adaptation. Extensive experiments on real-world server-monitoring and industrial-control benchmark datasets demonstrate state-of-the-art performance of MNP-GDA on cross-domain MTS-AD tasks.

1 INTRODUCTION

Multivariate time series (MTS) data are ubiquitous in real-world systems, ranging from online services (Sun et al., 2023), spacecraft (Herrmann et al., 2024), healthcare (Niu et al., 2023), finance (Cao & Guo, 2024), to Internet of Things (IoT) systems (Zheng et al., 2023). Multivariate time series anomaly detection (MTS-AD) (Wang et al., 2025) aims to identify samples that deviate significantly from general temporal dynamics or normal inter-variable dependencies, facilitating reliable IoT system operation, timely fault diagnosis, spacecraft monitoring, and financial risk mitigation.

Pioneering MTS-AD methods typically adopt an unsupervised learning paradigm to identify anomalies in the absence of labeled data (Shen et al., 2020; Su et al., 2019), however, without explicit semantic guidance on anomalous patterns, they may suffer from limited precision (Wang et al., 2024; Lai et al., 2024). To address this, recent semi-supervised methods (Zheng et al., 2023; Xie et al., 2025) leverage a limited number of labeled anomalies to guide the learning process, yielding more discriminative representations and enhancing anomaly detection performance.

Although unsupervised and semi-supervised approaches have substantially advanced MTS-AD, they are primarily developed under the **single-domain** assumption that both training and test data originate from the same domain and share the same variable space (Zhang et al., 2024; Xie et al., 2025). However, this assumption rarely holds in real-world applications that require frequent system upgrades or cross-platform deployment (He et al., 2024). In such cases, the MTS-AD model trained on a source domain cannot be directly applied to a new target domain, due to different data distributions and unshared variable sets across domains. This highlights the need to incorporate cross-domain data to assist MTS-AD. The newly emerging **domain adaptation for MTS-AD** problem (Darban

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

et al., 2025) aims to detect anomalies in an unlabeled target domain by transferring knowledge from a labeled source domain. However, three fundamental challenges (CHs) remain less explored.

CH1: **Spatio-temporal dependencies**. The MTS data contain essential spatial and temporal dependencies between variables and time steps (Wang et al., 2023). Graph-based MTS-AD methods (Zhao et al., 2020; Cui et al., 2025) can effectively model spatio-temporal dependencies, however, they are limited to single-domain scenario, unable to transfer knowledge across heterogeneous MTS domains with unshared variable sets and distinct dependency structures. In addition, existing domain adaptation methods for MTS-AD explicitly model either temporal (Lai et al., 2023; Darban et al., 2025) or spatial (He et al., 2024) dependency solely, where the latter merely constructs a K-nearest-neighbor (K-NN) graph with fixed neighborhood size for each variable, limiting its flexibility in modeling spatial dependency structure. *How to develop adaptive graph modeling to jointly capture latent spatio-temporal dependencies in cross-domain MTS-AD remains largely under-explored*.

CH2: **Transfer multimodal normalities across domains**. In MTS data, each variable has its own distinct normal mode, and normal samples often comprise multiple categories or subgroups, arising from different operating conditions, system states, and temporal patterns (Wang et al., 2023). Recent MTS-AD methods leverage multiple prototypes to capture diverse normal patterns (Li et al., 2023; Shen, 2025), however, they are limited to the single-domain setting, unable to transfer knowledge across domains. In addition, existing domain adaptation for MTS-AD methods (Lai et al., 2023; He et al., 2024; Darban et al., 2025; Zhang & Xu, 2025) generally overlook how to model and transfer multimodal normalities across domains. Besides, recent prototypical domain adaptation methods (Shao et al., 2024; Lin et al., 2025; Shen et al., 2026; Ma et al., 2025) typically represent each class with a single prototype. However, in the context of MTS-AD, using a single normal prototype may lead to mis-classification of obscure normal samples as anomalies when their representations lie far away from that single prototype (Dong et al., 2024; Li et al., 2023). This presents the key motivation of this work: *How to develop prototypical domain adaptation explicitly tailored for cross-domain MTS-AD, to model and transfer multimodal normalities across heterogeneous MTS domains*?

CH3: **Reliability weighting for pseudo-labels**. In the studied cross-domain MTS-AD problem, labels are unavailable in the target domain. However, prototypical domain adaptation requires the target label information to conduct domain alignment with the source prototypes. Target pseudo-labels therefore provide the required supervision, whereas their reliabilities vary across target samples and would directly affect domain alignment. Treating all pseudo-labeled target samples equally would make the low-confidence predictions exert the same influence as the reliable ones, resulting in negative transfer. Moreover, the same anomaly evidence carries the opposite reliability implications for pseudo-normal and -anomalous samples, where lower evidence supports the former, whereas higher evidence supports the latter. This raises a key question: *How to design adaptive reliability weighting mechanism to dynamically modulate the contribution of different pseudo-normal and -anomalous target samples during prototypical domain adaptation*?

To address the aforementioned challenges, we propose a Multimodal Normality Prototypical Graph Domain Adaptation (MNP-GDA) framework for cross-domain MTS-AD. First, to model spatio-temporal dependencies between variables and time steps within each domain, MNP-GDA learns adaptive inter-variable graph for each domain and constructs domain-specific spatio-temporal encoder to learn dependency-aware representations. To accommodate adaptation across heterogeneous MTS domains with unshared variable sets, MNP-GDA aligns the representations of both domains into a common latent space through a shared projection head. Second, to model and transfer diverse normal modes across domains, a multimodal normality prototypical domain adaptation module with theoretical justification is proposed, where multiple source-supervised normal prototypes serve as the shared normality anchors across domains, and the target normal samples would be attracted toward their nearest normal prototypes, while the target anomalies would be repelled from all normal prototypes. Finally, to alleviate negative effect of noisy pseudo-labels during domain adaptation, a theoretically grounded label-conditioned adaptive reliability weighting mechanism is devised to weight pseudo-normal and -anomalous target samples oppositely and modulate their influence upon the reliability of their pseudo-labels. The contributions of this work are summarized as follows:

- **Novel Perspective**. We identify core issues remain unresolved in the newly emerging cross-domain MTS-AD studies, i.e., modeling spatio-temporal dependencies and transferring multimodal normalities across heterogeneous MTS domains. We provide a novel perspective for cross-domain MTS-AD by reformulating it as a time-series-based graph domain adaptation problem.

- **Theoretically Grounded Method**. We propose a MNP-GDA method to capture latent spatio-temporal dependencies by adaptive graph modeling, transfer multimodal normalities across domains by normal attraction towards and anomaly repulsion from multiple normal prototypes, and mitigate negative influence of unreliable target pseudo-labels by label-conditioned adaptive reliability weighting. Theoretical analyses support the benefits of multiple prototypes and reliability weighting under explicit conditions.
- **High Performance**. Extensive experiments on five real-world server-monitoring and industrial-control benchmark datasets demonstrate state-of-the-art (SOTA) performance of MNP-GDA on cross-domain MTS-AD tasks. Compared with the strongest baseline, MNP-GDA significantly improves the average F1, AUPR, and AUROC by 6.60%, 9.48%, and 5.93%, respectively.

2 Related Work

Multivariate Time Series Anomaly Detection. Traditional MTS-AD methods primarily adopt an unsupervised paradigm to learn normal patterns without anomaly labels, by reconstruction and forecasting (Zhang et al., 2024), prototype-based normality learning (Shen, 2025), or one-class (Shen et al., 2020) objectives. Due to the lack of explicit semantic guidance on anomalous patterns, unsupervised methods struggle to distinguish subtle anomalies from complex normal variations (Wang et al., 2024). Semi-supervised methods incorporate available label information to improve normal–anomaly discrimination (Zheng et al., 2023; Xie et al., 2025). Recently, a line of graph-based methods focuses on explicitly modeling spatio-temporal dependencies (Liu et al., 2024b; Cui et al., 2025). For example, MTAD-GAT (Zhao et al., 2020) leverages feature- and time-oriented graph attention to capture inter-variable and temporal dependencies. MTGFlow (Zhou et al., 2024) integrates dynamically learned graphs with normalizing flows to identify low-likelihood observations. HSTVAE (Zhang et al., 2024) couples temporal stochasticity with latent graph relationships within a hierarchical variational framework. Despite notable progress, existing MTS-AD methods are predominantly developed for the single-domain setting, incapable of transferring knowledge across heterogeneous MTS domains with inherent temporal distribution shifts and different variable sets.

Domain Adaptation for Multivariate Time Series Anomaly Detection. Current research on domain adaptation for MTS-AD remains at an early stage. VANDA (He et al., 2024) assumes both domains are unlabeled and transfers inter-domain variate associations through optimal transport. DACAD (Darban et al., 2025) transfers knowledge from a labeled source domain to an unlabeled target domain through anomaly injection, contrastive adaptation, and adversarial alignment. ContexTDA (Lai et al., 2023) proposes a deep reinforcement learning method to transfer knowledge for cross-domain MTS-AD. PRACTISED (Zhang & Xu, 2025) addresses a different problem setting where the source domain only contains normal samples, and synthesizes pseudo-anomalies to learn transferable anomaly semantics. However, these methods *explicitly model either temporal or spatial dependency solely*. Moreover, they *overlook the transfer of multimodal normalities across heterogeneous MTS domains*. To fill such gaps, we propose MNP-GDA to jointly capture spatio-temporal dependencies by adaptive graph modeling, transfer multimodal normalities by multiple normal prototypes, and mitigate negative influence of unreliable pseudo-labels through adaptive reliability weighting.

Graph Domain Adaptation. Recent graph domain adaptation methods aim to transfer knowledge from a labeled source graph to an unlabeled target graph under domain discrepancy, addressing cross-graph prediction tasks on node classification (Shen et al., 2021; Dai et al., 2023; Shen et al., 2025; Wen et al., 2026; Wu et al., 2026; Fang et al., 2026; Chen et al., 2026; Zhang & He, 2025) and anomaly detection (Li et al., 2024; Huang et al., 2025; Zheng et al., 2026). However, they are mainly designed for static graphs, without modeling inherent spatio-temporal dependencies in MTS data. Moreover, they typically represent each class with a single prototype, which is insufficient to capture multimodal normalities in MTS-AD. Thus, it is impractical to directly apply off-the-shelf graph domain adaptation methods to address cross-domain MTS-AD.

3 Proposed MNP-GDA

Problem Definition. We consider cross-domain MTS-AD with a labeled source domain $\mathcal{D}_s = \{(\boldsymbol{X}_i^s, y_i^s)\}_{i=1}^{n_s}$ and an unlabeled target domain $\mathcal{D}_t = \{\boldsymbol{X}_i^t\}_{i=1}^{n_t}$. Each window $\boldsymbol{X}_i^r \in \mathbb{R}^{T\times d_r}$ in domain $r \in \{s,t\}$ contains T time steps of d_r variables, and $y_i^s \in \{0,1\}$ is the observed label of the

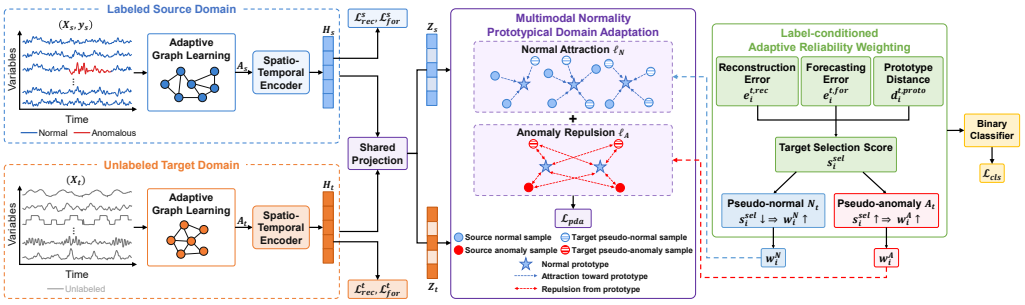

Figure 1: Overall framework of MNP-GDA, with adaptive graph learning, multimodal normality prototypical domain adaptation, and adaptive reliability weighting.

i-th source window with 0 and 1 representing normal and anomalous, respectively. The two domains possess inherent temporal distribution shifts and different sets of variables. The goal is to learn an anomaly scoring function for the target domain $S_\theta : \mathbb{R}^{T\times d_t} \to \mathbb{R}$ upon $\mathcal{D}_s$ and $\mathcal{D}_t$.

Overall Framework. We propose a novel MNP-GDA method to address cross-domain MTS-AD, its overall framework is illustrated in Figure 1. First, adaptive graph modeling captures latent spatio-temporal dependencies within each domain. Second, multimodal normality prototypical domain adaptation transfers knowledge across domains by normal attraction and anomaly repulsion upon multiple normal prototypes. Finally, the contributions of pseudo-labeled target samples are dynamically modulated during adaptation through label-conditioned adaptive reliability weighting.

3.1 Adaptive Graph Modeling on Spatio-Temporal Dependencies

Adaptive Graph Learning. To construct inter-variable graph, previous cross-domain MTS-AD method (He et al., 2024) selects a fixed number of K-NN for each variable, limiting its flexibility in modeling inter-variable dependency structure. To address this, we opt for adaptive graph modeling (Wu et al., 2019) to learn an adaptive inter-variable dependency matrix $\boldsymbol{A}_r \in \mathbb{R}^{d_r\times d_r}$ for each domain $r \in \{s, t\}$ to capture the hidden spatial dependencies, as:

$$\boldsymbol{A}_r = \mathrm{GumbelSoftmax}\big(\boldsymbol{E}_{1,r}\boldsymbol{E}_{2,r}^\top + \beta\boldsymbol{I}_{d_r}; \tau_g\big)\,, \tag{1}$$

where $\boldsymbol{E}_{1,r}$ and $\boldsymbol{E}_{2,r}$ are learnable variable embeddings that parameterize pairwise relation logits, $\beta\boldsymbol{I}_{d_r}$ introduces a self-loop bias with $\boldsymbol{I}_{d_r}$ denoting an identity matrix, and τ_g controls the Gumbel–Softmax temperature. As the source and target domains inherently exhibit different variable sets and distinct dependency structures, $\boldsymbol{A}_s \in \mathbb{R}^{d_s\times d_s}$ and $\boldsymbol{A}_t \in \mathbb{R}^{d_t\times d_t}$ have different dimensions.

Spatio-Temporal Encoder. Given a window $\boldsymbol{X}_i^r$ and adaptive inter-variable dependency matrix $\boldsymbol{A}_r$, a domain-specific spatio-temporal encoder is devised to capture the complementary local and global dependencies by dual branches. The local branch applies graph convolution (Cai et al., 2025) over $\boldsymbol{A}_r$ to capture inter-variable dependencies at each time step, thereby learning the local representation $\boldsymbol{H}_{i,\mathrm{loc}}^r$. Additionally, the global branch employs block-causal multi-head self-attention (Dai et al., 2024) over hidden states indexed by both time and variable to capture long-range dependencies across variables and time steps, yielding the global representation $\boldsymbol{H}_{i,\mathrm{glob}}^r$. The local and global representations are fused and applied attention-based pooling to obtain the window representation, as:

$$\boldsymbol{h}_i^r = \mathrm{Pool}_r\big(\mathrm{Fuse}_r\big(\boldsymbol{H}_{i,\mathrm{loc}}^r, \boldsymbol{H}_{i,\mathrm{glob}}^r\big)\big)\,, \tag{2}$$

where $\boldsymbol{h}_i^r \in \mathbb{R}^{d_h}$ and d_h denotes the dimension of the window representation. Detailed computations of the local and global representations are provided in Appendix A.1.

3.2 Multimodal Normality Prototypical Domain Adaptation

Multiple Normal Prototypes. In MTS data, normal samples often comprise multiple modes, arising from different operating conditions, system states, and temporal patterns (Wang et al., 2023). To transfer multimodal normalities across domains, MNP-GDA maintains a shared bank of ℓ_2-normalized normal prototypes $\mathcal{P} = \{\boldsymbol{p}_k\}_{k=1}^K$, where $\boldsymbol{p}_k \in \mathbb{R}^{d_p}$ denotes the k-th normal prototype, d_p is the

dimension of prototype representation, and K is the number of normal prototypes. To facilitate adaptation across heterogeneous MTS domains with unshared variable spaces, we further align the window representations of both domains into a common latent space through a shared projection head $g : \mathbb{R}^{d_h} \to \mathbb{R}^{d_p}$ and apply ℓ_2-normalization as $\boldsymbol{z}_i^r = g(\boldsymbol{h}_i^r)/\left\|g(\boldsymbol{h}_i^r)\right\|_2$. For each projected window representation $\boldsymbol{z}_i^r$, we denote its nearest normal prototype as $\boldsymbol{p}_{\boldsymbol{z}_i^r} := \arg\max_{\boldsymbol{p}_k \in \mathcal{P}} \langle \boldsymbol{z}_i^r, \boldsymbol{p}_k \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the inner product similarity. In the studied cross-domain MTS-AD problem, the fully labeled source domain provides reliable supervision on normal and anomalous patterns. Accordingly, the normal prototypes are initialized by applying K-means on the projected source normal representations. Then, during training, each k-th normal prototype is updated by a momentum moving average strategy (Wei et al., 2023) for every source mini-batch as: $\boldsymbol{p}_k \leftarrow \mu \boldsymbol{p}_k + (1 - \mu) \sum_{i \in \mathcal{B}_{s,k}} \boldsymbol{z}_i^s / |\mathcal{B}_{s,k}|$, where $\mathcal{B}_{s,k} = \{i \in \mathcal{B}_s \mid y_i^s = 0, \boldsymbol{p}_{\boldsymbol{z}_i^s} = \boldsymbol{p}_k\}$ denotes the normal samples assigned to prototype k in the source mini-batch $\mathcal{B}_s$, and $\mu \in [0, 1)$ is a momentum coefficient.

Next, all labeled source samples and selected pseudo-labeled target samples would be engaged in prototypical domain adaptation. Let $\mathcal{N}_s = \{i \mid y_i^s = 0\}$ and $\mathcal{A}_s = \{i \mid y_i^s = 1\}$ denote the source normal and anomalous sets, while $\mathcal{N}_t$ and $\mathcal{A}_t$ denote the target sets of pseudo-normal and -anomalous samples, their selection will be described in Section 3.3.

Normal Attraction Loss. Each normal representation $\boldsymbol{z}_i^r$ from both domains is pushed close to its nearest normal prototype $\boldsymbol{p}_{\boldsymbol{z}_i^r}$, while far away from other normal prototypes, as:

$$\ell_N(\boldsymbol{z}_i^r) = -\log \frac{\exp\big(\langle \boldsymbol{z}_i^r, \boldsymbol{p}_{\boldsymbol{z}_i^r} \rangle / \tau\big)}{\sum_{k=1}^{K} \exp(\langle \boldsymbol{z}_i^r, \boldsymbol{p}_k \rangle / \tau)}, \qquad i \in \mathcal{N}_r, \quad r \in \{s, t\}, \tag{3}$$

where τ is the temperature parameter. Minimizing ℓ_N promotes compact representations within each normal mode shared by two domains, while distinguishable representations between different normal modes, thereby effectively transferring multimodal normalities across domains.

Anomaly Repulsion Loss. Each anomalous representation $\boldsymbol{z}_i^r$ across domains is pulled far away from all normal prototypes, as:

$$\ell_A(\boldsymbol{z}_i^r) = \log \sum_{k=1}^{K} \exp(\langle \boldsymbol{z}_i^r, \boldsymbol{p}_k \rangle / \tau)\,, \qquad i \in \mathcal{A}_r, \quad r \in \{s, t\}. \tag{4}$$

Minimizing ℓ_A separates both source anomalous and target pseudo-anomalous samples from all the normal prototypes, thereby facilitating MTS-AD across domains.

Theoretical Justification for Multiple Prototypes. We theoretically quantify the benefit of representing multimodal normalities with multiple prototypes. For a normal representation $\boldsymbol{z} \sim P_N$, given K normal prototypes, the optimal expected cosine distance to its nearest prototype is given by:

$$D_K^* = \inf_{\substack{|\mathcal{P}|=K \\ \|\boldsymbol{p}\|_2=1,\ \forall\, \boldsymbol{p} \in \mathcal{P}}} \mathbb{E}_{\boldsymbol{z} \sim P_N} \left[\min_{\boldsymbol{p} \in \mathcal{P}} \big(1 - \langle \boldsymbol{z}, \boldsymbol{p} \rangle\big) \right]. \tag{5}$$

Suppose $P_N = \sum_{m=1}^{M} \pi_m P_m$ contains $M \geq 2$ normal modes, where $\pi_m > 0$ and $\sum_{m=1}^{M} \pi_m = 1$. Let $\boldsymbol{\mu}_m = \mathbb{E}_{\boldsymbol{z} \sim P_m}[\boldsymbol{z}]$, $r_m = \|\boldsymbol{\mu}_m\|_2 > 0$, and $\boldsymbol{v}_m = \boldsymbol{\mu}_m / r_m$ denote the mean vector, magnitude, and normalized mode direction of the normal mode m, respectively. Assume $\langle \boldsymbol{v}_a, \boldsymbol{v}_b \rangle \leq 1 - 2\delta$ for $a \neq b$, where $\delta > 0$ characterizes the angular separation between normal modes. With $a_m = \pi_m r_m$ representing the weight of normal mode m, $C = \sum_{m=1}^{M} a_m$, and $\sigma^2 = \sum_{m=1}^{M} \pi_m (1 - r_m)$, the following Theorem characterizes the benefit of multiple prototypes under distinct normal modes.

Theorem 1. For every $K \geq 1$, adding a prototype cannot increase the expected distance. Moreover, for $K \geq M$,

$$D_{K+1}^* \leq D_K^*, \qquad D_K^* \leq \sigma^2, \qquad D_1^* - D_K^* \geq \frac{2\delta}{C} \sum_{1 \leq a < b \leq M} a_a a_b > 0. \tag{6}$$

Remark. According to Theorem 1, adding prototypes would not worsen the nearest-prototype fit, and when $K \geq M$, distinct normal modes would admit a strictly smaller expected distance than using a single normal prototype, where such improvement is related to the separation between normal modes through δ. The proof is provided in Appendix B. This theoretical analysis supports our multimodal normality prototypical domain adaptation design, which explicitly preserves diverse normal modes and provides more reliable normality anchors for MTS-AD knowledge transfer across domains.

3.3 LABEL-CONDITIONED ADAPTIVE RELIABILITY WEIGHTING

Target Selection Score. The reconstruction and forecasting errors have been widely adopted as the indicators for MTS anomaly detection (Sun et al., 2025; Chen et al., 2023). To further exploit multimodal normalities in MTS, we define a comprehensive target selection score to jointly consider reconstruction consistency, temporal predictability, and prototype proximity, as:

$$s_i^{\text{sel}} = \sigma(e_i^{t,\text{rec}}) + \sigma(e_i^{t,\text{for}}) + \sigma(d_i^{t,\text{proto}}), \tag{7}$$

where $\sigma(\cdot)$ denotes min–max normalization, $e_i^{t,\text{rec}}$ and $e_i^{t,\text{for}}$ denote the reconstruction and forecasting errors of the target window i, respectively, detailed computations are provided in Appendix A.2, and $d_i^{t,\text{proto}} = 1 - \max_{\boldsymbol{p}_k \in \mathcal{P}} \langle \boldsymbol{z}_i^t, \boldsymbol{p}_k \rangle$ is the distance between $\boldsymbol{z}_i^t$ and its nearest normal prototype. Higher score s_i^{sel} indicates that the target window i is more likely to be anomalous, and vice versa.

For each target mini-batch, we retain top-$\rho_a\%$ of samples with the highest selection score as the target pseudo-anomalous set $\mathcal{A}_t$. However, normal samples in MTS generally comprise multiple normal modes (Li et al., 2023), thus directly selecting target samples with the lowest score might bias toward dominant normal modes, leaving less frequent normal modes underrepresented. We therefore perform **prototype-wise normal selection** to preserve diverse normal modes. Specifically, for each normal prototype $\boldsymbol{p}_k$, we retain top-$\rho_n\%$ of its assigned target samples with the lowest selection score as the corresponding pseudo-normal candidates $\mathcal{N}_t^k$. Then, the target pseudo-normal set $\mathcal{N}_t$ is obtained by integrating pseudo-normal candidates from all the K normal prototypes, and removing the samples overlapping with $\mathcal{A}_t$ to avoid contradictory pseudo-labels, i.e., $\mathcal{N}_t = \left(\bigcup_{k=1}^{K} \mathcal{N}_t^k\right) \setminus \mathcal{A}_t$.

Label-conditioned Adaptive Reliability Weights. The selected pseudo-labeled target samples are not equally reliable, and treating them uniformly would amplify noisy target supervision. Notably, the same selection score has the opposite reliability implications for the two pseudo-labeled groups, i.e., lower scores provide stronger evidence of normal, whereas higher scores provide stronger evidence of anomaly. Accordingly, MNP-GDA assigns label-conditioned adaptive reliability weights w_i^N and w_i^A for pseudo-normal and -anomalous target samples, respectively, by min-max normalizing the target selection score s_i^{sel} in the opposite directions, as:

$$w_i^A = \lambda + (1-\lambda)\frac{s_i^{\text{sel}} - \min_{j \in \mathcal{A}_t} s_j^{\text{sel}}}{\max_{j \in \mathcal{A}_t} s_j^{\text{sel}} - \min_{j \in \mathcal{A}_t} s_j^{\text{sel}} + \epsilon}, \qquad i \in \mathcal{A}_t, \tag{8}$$

$$w_i^N = \lambda + (1-\lambda)\frac{\max_{j \in \mathcal{N}_t^k} s_j^{\text{sel}} - s_i^{\text{sel}}}{\max_{j \in \mathcal{N}_t^k} s_j^{\text{sel}} - \min_{j \in \mathcal{N}_t^k} s_j^{\text{sel}} + \epsilon}, \qquad i \in \mathcal{N}_t^k, \tag{9}$$

where $w_i^A, w_i^N \in [\lambda, 1]$ and λ sets the minimum contribution of the selected pseudo-labeled target samples, and $\epsilon > 0$ is a small constant to avoid division by zero. In Eq. 8, a higher selection score s_i^{sel} yields a larger w_i^A, suggesting higher reliability of pseudo-anomalous label. Conversely, in Eq. 9, a lower selection score s_i^{sel} yields a larger w_i^N, indicating higher reliability of pseudo-normal label. Moreover, we apply prototype-wise normalization within each normal mode $\mathcal{N}_t^k$ to avoid interference from score variations across different normal modes.

Weighted Prototypical Domain Adaptation Loss. By assigning unit weights to all labeled source samples $\omega_i^s = 1$, whereas adaptive reliability weights to pseudo-labeled target samples, i.e., $\omega_i^t = w_i^N$ for $i \in \mathcal{N}_t$ and $\omega_i^t = w_i^A$ for $i \in \mathcal{A}_t$, the weighted prototypical domain adaptation loss is given by:

$$\mathcal{L}_{\text{pda}} = \frac{\sum_{r \in \{s,t\},\, i \in \mathcal{N}_r} \omega_i^r \ell_N(\boldsymbol{z}_i^r)}{\sum_{r \in \{s,t\},\, i \in \mathcal{N}_r} \omega_i^r} + \alpha_{\text{push}} \frac{\sum_{r \in \{s,t\},\, i \in \mathcal{A}_r} \omega_i^r \ell_A(\boldsymbol{z}_i^r)}{\sum_{r \in \{s,t\},\, i \in \mathcal{A}_r} \omega_i^r}, \tag{10}$$

where $\ell_N(\boldsymbol{z}_i^r)$ and $\ell_A(\boldsymbol{z}_i^r)$ are the normal attraction and anomaly repulsion losses in Eq. 3 and Eq. 4, and α_{push} controls the strength of anomaly repulsion. Minimizing $\mathcal{L}_{\text{pda}}$ aligns cross-domain normal samples with their corresponding normal prototypes, while pulling cross-domain anomalous samples far away from all normal prototypes, and adaptively modulates the influence of target pseudo-labeled samples according to their reliability weights.

Theoretical Justification for Adaptive Reliability Weighting. We theoretically analyze how adaptive reliability weighting mitigates the effect of noisy target pseudo-labels. For a selected

group $g \in \{N, A\}$, let $\tilde{y}_g$ denote its pseudo-label and y^* denote the unobserved ground-truth label. We define $\eta_g(\boldsymbol{X}) = \Pr(\tilde{y}_g \neq y^* \mid \boldsymbol{X})$ as the conditional pseudo-label error. Let $\ell_\theta(\boldsymbol{X}, y)$ denote the sample-wise prototypical domain adaptation loss, corresponding to the normal-attraction loss ℓ_N in Eq. 3 for $y = 0$ and the anomaly-repulsion loss ℓ_A in Eq. 4 for $y = 1$, and assume $|\ell_\theta(\boldsymbol{X}, 1) - \ell_\theta(\boldsymbol{X}, 0)| \leq B_g$, where B_g bounds the loss discrepancy induced by label corruption within group g. The condition $\mathrm{Cov}_g(w, \eta_g) \leq 0$ indicates that larger reliability weights are, on average, associated with lower pseudo-label error. We further define $\bar{\eta}_{w,g} = \mathbb{E}_g[w\eta_g]/\mathbb{E}_g[w]$ and $\bar{\eta}_g = \mathbb{E}_g[\eta_g]$ as the weighted and unweighted average pseudo-label error rates, respectively.

Theorem 2. For the reliability-weighted pseudo-label surrogate and its true-label counterpart,

$$R_{w,g}(\theta) = \frac{\mathbb{E}_g[w(\boldsymbol{X})\ell_\theta(\boldsymbol{X}, \tilde{y}_g)]}{\mathbb{E}_g[w(\boldsymbol{X})]}, \qquad R^*_{w,g}(\theta) = \frac{\mathbb{E}_g[w(\boldsymbol{X})\ell_\theta(\boldsymbol{X}, y^*)]}{\mathbb{E}_g[w(\boldsymbol{X})]}, \\ \left|R_{w,g}(\theta) - R^*_{w,g}(\theta)\right| \leq B_g\bar{\eta}_{w,g} \leq B_g\bar{\eta}_g. \tag{11}$$

Remark. According to Theorem 2, under the covariance condition, reliability weighting bounds the discrepancy between the pseudo-label and true-label objectives no larger than its unweighted counterpart. The proof is provided in Appendix C. This theoretical analysis supports our label-conditioned adaptive reliability weighting design, which mitigates negative influence of unreliable target supervision and promotes robust prototypical domain adaptation under noisy pseudo-labels.

For anomaly detection, MNP-GDA constructs a linear binary classifier $f_{\mathrm{cls}}(\cdot)$ on the top of each window representation, together with the three anomaly signals in the target selection score, and outputs the anomaly score, as:

$$S_i^r = f_{\mathrm{cls}}\left(\left[\boldsymbol{h}_i^r \| \sigma(e_i^{r,\mathrm{rec}}) \| \sigma(e_i^{r,\mathrm{for}}) \| \sigma(d_i^{r,\mathrm{proto}})\right]\right), \tag{12}$$

where $[\cdot\|\cdot]$ denotes the concatenation operation, and $S_i^r \in [0, 1], r \in \{s, t\}$ is the predicted anomaly score of window i with a larger value indicating higher likelihood of anomaly, and vice versa.

Overall Loss. Training the binary classifier upon observed source labels and selected target pseudo-labels, we compute the classification loss for each domain $\mathcal{L}^r_{\mathrm{cls}}$. Besides, following MTS-AD studies (Su et al., 2019; Zhao et al., 2020), for each domain, we adopt the reconstruction loss $\mathcal{L}^r_{\mathrm{rec}}$ to capture normal patterns of observed windows, and the forecasting loss $\mathcal{L}^r_{\mathrm{for}}$ to model temporal dependencies through next-step prediction. Finally, integrating these losses with the weighted prototypical domain adaptation loss $\mathcal{L}_{\mathrm{pda}}$, the overall loss of MNP-GDA is given by:

$$\mathcal{L} = \sum_{r\in\{s,t\}} (\alpha_{\mathrm{rec}}\mathcal{L}^r_{\mathrm{rec}} + \alpha_{\mathrm{for}}\mathcal{L}^r_{\mathrm{for}} + \alpha_{\mathrm{cls}}\mathcal{L}^r_{\mathrm{cls}}) + \alpha_{\mathrm{pda}}(e)\mathcal{L}_{\mathrm{pda}}. \tag{13}$$

where α_{rec}, α_{for}, α_{cls}, and $\alpha_{\mathrm{pda}}(e)$ are the weights of different loss terms. To mitigate negative influence of unstable prototypes and pseudo-labels during early training, $\alpha_{\mathrm{pda}}(e)$ increases linearly during the first five epochs and remains fixed thereafter. Detailed computations of the classification, reconstruction and forecasting losses are provided in Appendix A. The Algorithm of MNP-GDA is given in Appendix A.3 and complexity analysis is reported in Appendix G.

4 Experiments

4.1 Experimental Setup

Datasets and Tasks. We conducted experiments on five MTS-AD benchmark datasets: SMD (Su et al., 2019), PSM (Abdulaal et al., 2021), JD1 (Ma et al., 2021), SWaT (Mathur & Tippenhauer, 2016), and WADI (Ahmed et al., 2017), where the first three were collected from server-monitoring IT systems, whereas the other two were gathered from industrial water-treatment platforms. We constructed eight cross-domain MTS-AD tasks, with six among SMD, PSM, and JD1, as well as two between SWaT and WADI. Details of datasets and preprocessing are reported in Appendix D.1.

Baselines and Metrics. We compared MNP-GDA against nine SOTA baselines in three categories: 1) *Single-domain MTS-AD*: **OmniAnomaly** (Su et al., 2019) and **THOC** (Shen et al., 2020); 2) *Single-domain Graph-based MTS-AD*: **MTAD-GAT** (Zhao et al., 2020), **HSTVAE** (Zhang et al., 2024), and **MTGFlow** (Zhou et al., 2024); 3) *Domain Adaptation on MTS*, including for *classification*,

Table 1: F1, AUPR, and AUROC (%) on eight cross-domain MTS-AD tasks. The best and second-best results are shown in bold and underlined, respectively. S, P, J, Sw, and W are short for SMD, PSM, JD1, SWaT, and WADI.

Method	Metric	S→P	S→J	P→S	P→J	J→S	J→P	Sw→W	W→Sw	Avg.
OmniAnomaly	F1	22.96±0.19	11.06±0.34	9.99±0.38	11.41±0.70	9.96±0.26	23.32±0.09	23.54±4.25	22.37±0.00	16.83
	AUPR	12.92±0.40	5.98±0.16	4.96±0.09	6.00±0.17	4.95±0.10	13.47±0.10	15.60±4.78	9.31±1.41	9.15
	AUROC	56.07±0.90	54.82±1.73	58.94±1.10	55.25±2.58	58.78±0.70	57.41±0.21	61.73±3.13	80.29±0.74	60.41
THOC	F1	25.19±4.56	12.15±0.78	8.84±0.91	12.63±0.94	9.97±1.41	23.88±4.99	13.16±2.15	56.82±18.88	20.33
	AUPR	16.78±8.87	6.39±1.21	4.53±1.23	7.38±1.18	5.17±0.70	13.39±5.68	6.19±1.93	47.09±20.34	13.37
	AUROC	60.57±3.66	54.33±3.21	55.71±1.39	56.01±3.67	52.51±2.49	57.40±3.45	60.88±7.25	77.31±3.69	59.34
MTAD-GAT	F1	23.00±0.19	11.44±0.66	10.06±0.41	11.03±0.24	10.25±0.32	22.89±0.22	24.52±3.84	22.37±0.00	16.95
	AUPR	12.93±0.31	5.91±0.13	4.96±0.11	5.73±0.22	5.00±0.07	12.90±0.34	18.45±4.20	9.34±1.65	9.40
	AUROC	56.21±0.77	54.68±1.98	58.55±1.00	56.74±2.49	59.22±0.57	56.15±0.64	63.13±1.41	80.44±0.71	60.64
HSTVAE	F1	23.17±0.18	11.34±0.90	10.31±0.49	11.85±0.84	10.27±0.28	23.07±0.25	20.34±6.58	22.37±0.00	16.59
	AUPR	13.74±0.27	6.09±0.48	5.46±0.17	6.16±0.16	5.16±0.23	13.43±0.25	15.12±6.81	9.84±1.56	9.38
	AUROC	57.70±0.77	56.95±2.04	59.27±0.91	53.56±3.02	59.01±0.85	57.16±0.60	59.65±5.12	80.44±1.04	60.47
MTGFlow	F1	21.01±0.19	12.89±2.34	26.40±2.17	16.71±1.62	9.33±2.13	20.94±0.11	14.22±3.91	64.53±3.80	23.25
	AUPR	9.21±0.81	6.86±1.63	25.24±2.01	10.93±0.62	5.54±2.31	10.22±1.90	5.97±2.34	49.03±9.09	15.38
	AUROC	59.22±4.90	57.07±4.42	75.08±2.63	62.75±1.48	55.60±5.59	58.78±4.63	63.35±6.52	82.68±1.54	64.32
CLUDA	F1	25.11±5.77	10.89±0.00	17.28±9.72	12.68±2.15	11.29±3.03	24.75±3.51	20.05±11.27	56.76±20.84	22.35
	AUPR	16.44±7.58	4.58±0.43	9.38±7.51	6.66±1.28	5.54±1.21	14.91±4.18	11.74±8.32	46.89±27.98	14.52
	AUROC	57.88±6.32	58.97±4.50	56.70±3.47	55.99±5.34	55.07±3.60	57.31±3.42	60.66±3.80	68.99±11.43	58.95
ACON	F1	25.73±4.13	12.83±1.81	9.06±1.80	13.54±2.54	23.99±3.30	24.85±4.32	29.58±5.21	29.23±13.69	21.10
	AUPR	21.63±8.54	7.30±1.68	5.03±2.57	6.78±1.51	17.99±3.28	16.03±3.99	16.11±3.62	14.01±7.99	13.11
	AUROC	58.46±5.85	54.24±4.95	55.26±3.88	54.11±3.34	64.86±6.01	61.82±5.01	64.68±9.84	77.15±1.39	61.32
DACAD	F1	23.06±3.02	11.15±0.29	15.02±2.29	11.52±0.47	13.36±1.80	23.84±2.05	26.66±5.04	**74.90±1.76**	24.94
	AUPR	15.09±2.52	6.56±0.99	12.19±1.31	6.69±0.33	10.73±1.13	15.10±2.08	18.33±5.04	74.79±1.53	19.94
	AUROC	56.20±6.15	54.32±2.76	56.71±1.69	52.18±0.82	53.62±1.74	57.38±4.31	73.19±3.56	85.58±0.69	61.15
VANDA	F1	35.60±4.75	20.10±0.64	22.04±6.51	19.85±0.47	23.56±2.40	33.06±2.35	30.29±3.20	63.89±7.29	31.05
	AUPR	23.31±3.83	14.53±0.30	16.94±4.61	14.50±0.34	18.71±2.16	22.70±1.63	18.41±5.48	43.82±13.06	21.62
	AUROC	72.39±5.01	71.35±0.29	72.44±1.92	71.01±0.51	73.54±3.71	71.94±4.03	77.79±3.27	83.36±2.64	74.23
MNP-GDA (Ours)	F1	**38.41±2.47**	**25.05±0.92**	**35.08±3.21**	**25.31 ±0.58**	**35.30±1.35**	**35.27±1.62**	**32.90 ±3.28**	73.90±1.06	**37.65**
	AUPR	**29.37±3.09**	**19.87±2.36**	**27.97±4.93**	**17.53±2.87**	**27.26±4.92**	**26.35±1.32**	**24.52±4.13**	**75.96±3.96**	**31.10**
	AUROC	**80.10±2.30**	**75.54±0.90**	**84.11±1.09**	**75.20±1.80**	**82.35±0.93**	**76.14±1.68**	**79.81±2.80**	**88.00±0.85**	**80.16**

i.e., **CLUDA** (Ozyurt et al., 2023) and **ACON** (Liu et al., 2024a), and for *anomaly detection*, i.e., **DACAD** (Darban et al., 2025) and **VANDA** (He et al., 2024). ContexTDA (Lai et al., 2023) and PRACTISED (Zhang & Xu, 2025) were not included for comparison, due to unavailable source code. Following MTS-AD literature (Darban et al., 2025; Hermary et al., 2026), we adopted F1, AUPR, and AUROC as evaluation metrics. Each method was evaluated over five random seeds on every task, with mean and standard deviation reported. Detailed implementations of MNP-GDA and adaptation of baselines to cross-domain MTS-AD setting are provided in Appendix D.2 and Appendix D.3.

4.2 Results and Analysis

Table 1 presents the results of eight cross-domain MTS-AD tasks, we have following key observations: The proposed MNP-GDA achieves the best overall performance. Compared with the second-best method VANDA, MNP-GDA still improves the average F1, AUPR, and AUROC by 6.60%, 9.48%, and 5.93%, respectively. The outperformance of MNP-GDA might benefit from three novel designs: 1) adaptive graph modeling to capture latent spatio-temporal dependencies, 2) multimodal normality prototypical domain adaptation to transfer diverse normal modes across domains, and 3) label-conditioned adaptive reliability weighting to mitigate negative influence of unreliable pseudo-labels.

Moreover, the single-domain graph-based methods generally achieve better performance than non-graph ones, demonstrating the necessity of explicitly modeling spatio-temporal dependencies for MTS-AD. On average, the domain adaptation baselines outperform the single-domain baselines, highlighting the effectiveness of transferring cross-domain knowledge to assist MTS-AD.

4.3 Ablation Study

We conducted extensive ablation studies to investigate the contribution of key designs in MNP-GDA. As shown in Table 2, 1) Without adaptive graph modeling, i.e., constructing a fixed K-NN inter-variable graph instead would lead to significantly worse performance, verifying the benefit of adaptive graph learning in capturing latent spatial dependency. 2) Without modeling either spatial or

Table 2: Ablation study on three representative tasks.

Variant	PSM→SMD			JD1→SMD			WADI→SWaT		
	F1	AUPR	AUROC	F1	AUPR	AUROC	F1	AUPR	AUROC
MNP-GDA (full)	**35.08**	**27.97**	**84.11**	**35.30**	**27.26**	**82.35**	**73.90**	**75.96**	**88.00**
w/o adaptive graph modeling	31.17	25.64	82.82	28.49	20.74	78.75	72.79	71.88	82.65
w/o spatial dependency	33.81	24.79	81.49	30.50	21.90	78.79	72.36	72.36	83.96
w/o temporal dependency	26.28	15.10	78.14	31.19	22.88	78.02	70.33	62.61	86.19
w/o multiple normal prototypes	28.86	21.71	80.48	32.18	24.34	80.41	71.63	69.72	86.75
w/o normal attraction loss	30.38	21.08	79.46	33.87	26.60	76.89	72.90	67.11	87.34
w/o anomaly repulsion loss	28.62	20.56	75.06	33.06	26.48	76.74	71.90	67.45	86.38
w/o adaptive reliability weighting	33.71	24.96	82.34	33.41	22.53	79.12	73.02	72.24	80.94

temporal dependency also degrades the performance, demonstrating the necessity of capturing both spatio-temporal dependencies in MTS-AD. 3) Without multiple normal prototypes, i.e., using a single prototype instead would significantly degrade performance. This perfectly aligns with our theoretical analysis, highlighting the necessity of modeling multimodal normalities in MTS-AD. 4) Removing either normal attraction loss ℓ_N or anomaly repulsion loss ℓ_A in prototypical domain adaptation also significantly degrades performance, highlighting that compacting representations within each normal mode and repelling anomalies from all normal prototypes are both indispensable for cross-domain MTS-AD. 5) Removing adaptive reliability weighting, performance would drop considerably. This is consistent with our theoretical analysis that adaptive reliability weighting can promote robust prototypical domain adaptation under noisy pseudo-labels. Additional ablation studies on training objectives, anomaly score components, and target selection are provided in Appendix F.

4.4 Visualization of Representations

As shown in Figure 2, compared with SOTA cross-domain MTS-AD baselines, i.e., DACAD and VANDA, the proposed MNP-GDA effectively captures multiple normal modes, produces more compact representations within each normal mode, and clearer separation between normal and anomalous. This again validates the effectiveness of the proposed multimodal normality prototypical domain adaptation design with normal attraction and anomaly repulsion.

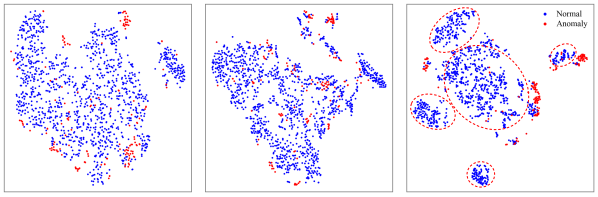

Figure 2: t-SNE visualization of target representations learned by DACAD, VANDA, and MNP-GDA on the JD1→PSM task. Blue and red denote normal and anomalous samples, respectively.

5 Conclusion

This work provides a novel perspective for cross-domain MTS-AD by reformulating it as a time-series-based graph domain adaptation problem. A multimodal normality prototypical graph domain adaptation framework MNP-GDA is proposed, with adaptive graph modeling to capture latent spatio-temporal dependencies, multimodal normality prototypical domain adaptation to transfer diverse normal modes across domains by normal attraction and anomaly repulsion upon multiple normal prototypes, and label-conditioned adaptive reliability weighting to mitigate negative influence of unreliable target pseudo-labels. Theoretical analyses characterize the benefits of multiple prototypes and reliability weighting under explicit conditions. Extensive experiments on real-world benchmark datasets demonstrate SOTA performance of MNP-GDA on cross-domain MTS-AD tasks, and the ablation and visualization analyses validate the proposed model designs.

AI USE STATEMENT

In this work, we used generative AI tools to assist with language editing, literature search, and improve the clarity and presentation of technical content. All AI-assisted outputs were reviewed and verified by the authors. The authors take full responsibility for the final content of this work, including text and claims produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

To support the reproducibility of our work, detailed method formulations and the overall training procedure are provided in Appendix A, and complete proofs of Theorem 1 and Theorem 2 are provided in Appendices B and C, respectively. Detailed descriptions of the datasets and preprocessing procedures are provided in Appendix D.1, while implementation details and task-specific hyperparameter settings are reported in Appendix D.2. Details of the baseline implementations are provided in Appendix D.3.

REFERENCES

Ahmed Abdulaal, Zhuanghua Liu, and Tomer Lancewicki. Practical approach to asynchronous multivariate time series anomaly detection and localization. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2485–2494, 2021.

Chuadhry Mujeeb Ahmed, Venkata Reddy Palleti, and Aditya P. Mathur. WADI: A water distribution testbed for research in the design of secure cyber physical systems. In *Proceedings of the 3rd international workshop on cyber-physical systems for smart water networks*, pp. 25–28, 2017.

Jinyu Cai, Yuan Xie, Glynnis Lim, Yifang Yin, Roger Zimmermann, and See-Kiong Ng. Self-perturbed anomaly-aware graph dynamics for multivariate time-series anomaly detection. In *Advances in Neural Information Processing Systems*, volume 38, pp. 84785–84805, 2025.

Feifei Cao and Xitong Guo. Automated financial time series anomaly detection via curiosity-guided exploration and self-imitation learning. *Engineering Applications of Artificial Intelligence*, 135: 108663, 2024.

Katrina Chen, Mingbin Feng, and Tony S. Wirjanto. Multivariate time series anomaly detection via dynamic graph forecasting. *arXiv preprint arXiv:2302.02051*, 2023.

Wei Chen, Xingyu Guo, Shuang Li, Yan Zhong, Zhao Zhang, Fuzhen Zhuang, Hongrui Liu, Libang Zhang, Guo Ye, and Huimei He. Learning structure-semantic evolution trajectories for graph domain adaptation. In *International Conference on Learning Representations*, pp. 109348–109368, 2026.

Zhe Cui, Di Zang, Junqi Zhang, and Keshuang Tang. MGCL: Multiorder graph neural network with cross-learning for multivariate time-series anomaly detection. *IEEE Transactions on Instrumentation and Measurement*, 74:1–14, 2025.

Quanyu Dai, Xiao-Ming Wu, Jiaren Xiao, Xiao Shen, and Dan Wang. Graph transfer learning via adversarial domain adaptation with graph convolution. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):4908–4922, 2023.

Zhihao Dai, Ligang He, Shuang-Hua Yang, and Matthew Leeke. SARAD: Spatial association-aware anomaly detection and diagnosis for multivariate time series. In *Advances in Neural Information Processing Systems*, volume 37, pp. 48371–48410, 2024.

Zahra Zamanzadeh Darban, Yiyuan Yang, Geoffrey I. Webb, Charu C. Aggarwal, Qingsong Wen, Shirui Pan, and Mahsa Salehi. DACAD: Domain adaptation contrastive learning for anomaly detection in multivariate time series. *IEEE Transactions on Knowledge and Data Engineering*, 37 (8):4485–4496, 2025.

Zhijin Dong, Hongzhi Liu, Boyuan Ren, Weimin Xiong, and Zhonghai Wu. Multi-normal prototypes learning for weakly supervised anomaly detection. *arXiv preprint arXiv:2408.14498*, 2024.

Ruiyi Fang, Jingyu Zhao, Shuo Wang, Ruizhi Pu, Bingheng Li, Jiale Cai, Zhihao Li, Zihao Jing, Jian Zhu, Song Tang, Charles Ling, and Boyu Wang. SAGA: Structural aggregation guided alignment with dynamic view and neighborhood order selection for multiview graph domain adaptation. In *International Conference on Learning Representations*, pp. 95585–95607, 2026.

Yifan He, Yatao Bian, Xi Ding, Bingzhe Wu, Jihong Guan, Ji Zhang, and Shuigeng Zhou. Variate associated domain adaptation for unsupervised multivariate time series anomaly detection. *ACM Transactions on Knowledge Discovery from Data*, 18(8):1–24, 2024.

Romain Hermary, Samet Hicsonmez, Dan Pineau, Abd El Rahman Shabayek, and Djamila Aouada. ASTER: Latent pseudo-anomaly generation for unsupervised time-series anomaly detection. In *International Conference on Pattern Recognition*, pp. 318–333, 2026.

Lars Herrmann, Marie Bieber, Wim J.C. Verhagen, Fabrice Cosson, and Bruno F. Santos. Unmasking overestimation: a re-evaluation of deep anomaly detection in spacecraft telemetry. *CEAS Space Journal*, 16(2):225–237, 2024.

Changqin Huang, Xinxing Shi, Chengling Gao, Qintai Hu, Xiaodi Huang, Qionghao Huang, and Ali Anaissi. Cross-domain graph anomaly detection via graph transfer and graph decouple. *CAAI Transactions on Intelligence Technology*, 10(4):1089–1103, 2025.

Kwei-Herng Lai, Lan Wang, Huiyuan Chen, Kaixiong Zhou, Fei Wang, Hao Yang, and Xia Hu. Context-aware domain adaptation for time series anomaly detection. In *Proceedings of the SIAM International Conference on Data Mining (SDM)*, pp. 676–684, 2023.

Thomas Lai, Thi Kieu Khanh Ho, and Narges Armanfard. Open-set multivariate time-series anomaly detection. In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*, pp. 2003–2010, 2024.

Yuxin Li, Wenchao Chen, Bo Chen, Dongsheng Wang, Long Tian, and Mingyuan Zhou. Prototype-oriented unsupervised anomaly detection for multivariate time series. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 19407–19424, 2023.

Zhong Li, Sheng Liang, Jiayang Shi, and Matthijs van Leeuwen. Cross-domain graph level anomaly detection. *IEEE Transactions on Knowledge and Data Engineering*, 36(12):7839–7850, 2024.

Chuanyun Lin, Xi Zhou, and Xiao Shen. Hierarchical graph contrastive domain adaptation for multi-source cross-network node classification. *Expert Systems with Applications*, 284:127900, 2025.

Mingyang Liu, Xinyang Chen, Yang Shu, Xiucheng Li, Weili Guan, and Liqiang Nie. Boosting transferability and discriminability for time series domain adaptation. In *Advances in Neural Information Processing Systems*, volume 37, pp. 100402–100427, 2024a.

Qinghua Liu and John Paparrizos. The elephant in the room: Towards a reliable time-series anomaly detection benchmark. In *Advances in Neural Information Processing Systems*, volume 37, pp. 108231–108261, 2024.

Zhe Liu, Xiang Huang, Jingyun Zhang, Zhifeng Hao, Li Sun, and Hao Peng. Multivariate time-series anomaly detection based on enhancing graph attention networks with topological analysis. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pp. 1555–1564, 2024b.

Minghua Ma, Shenglin Zhang, Junjie Chen, Jim Xu, Haozhe Li, Yongliang Lin, Xiaohui Nie, Bo Zhou, Yong Wang, and Dan Pei. Jump-starting multivariate time series anomaly detection for online service systems. In *2021 USENIX Annual Technical Conference (USENIX ATC 21)*, pp. 413–426, 2021.

Xin Ma, Yifan Wang, Siyu Yi, Wei Ju, Junyu Luo, Yusheng Zhao, Xiao Luo, and Jiancheng Lv. Dual prototype-enhanced contrastive framework for class-imbalanced graph domain adaptation. In *Advances in Neural Information Processing Systems*, volume 38, pp. 48456–48482, 2025.

Aditya P. Mathur and Nils Ole Tippenhauer. SWaT: A water treatment testbed for research and training on ICS security. In *2016 international workshop on cyber-physical systems for smart water networks (CySWater)*, pp. 31–36, 2016.

Mengjia Niu, Yuchen Zhao, and Hamed Haddadi. Effective abnormal activity detection on multivariate time series healthcare data. In *Proceedings of the 29th annual international conference on mobile computing and networking*, pp. 1–3, 2023.

Yilmazcan Ozyurt, Stefan Feuerriegel, and Ce Zhang. Contrastive learning for unsupervised domain adaptation of time series. In *International Conference on Learning Representations*, 2023.

Mengqiu Shao, Peng Xue, Xi Zhou, and Xiao Shen. Contrastive domain-adaptive graph selective self-training network for cross-network edge classification. *Pattern Recognition*, 152:110448, 2024.

Ke-Yuan Shen. Learn hybrid prototypes for multivariate time series anomaly detection. In *International Conference on Learning Representations*, pp. 53724–53746, 2025.

Lifeng Shen, Zhuocong Li, and James Kwok. Timeseries anomaly detection using temporal hierarchical one-class network. In *Advances in Neural Information Processing Systems*, volume 33, pp. 13016–13026, 2020.

Xiao Shen, Quanyu Dai, Sitong Mao, Fu-lai Chung, and Kup-Sze Choi. Network together: Node classification via cross-network deep network embedding. *IEEE Transactions on Neural Networks and Learning Systems*, 32(5):1935–1948, 2021.

Xiao Shen, Zhihao Chen, Shirui Pan, Shuang Zhou, Laurence T. Yang, and Xi Zhou. Open-set cross-network node classification via unknown-excluded adversarial graph domain alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 20398–20408, 2025.

Xiao Shen, Chuanyun Lin, Zhihao Chen, Junwei Gong, and Xi Zhou. Prototypical multi-source graph contrastive domain adaptation. *Pattern Recognition*, 180:114312, 2026.

Ya Su, Youjian Zhao, Chenhao Niu, Rong Liu, Wei Sun, and Dan Pei. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2828–2837, 2019.

Xin Sun, Heng Zhou, and Chao Li. Multivariate time series anomaly detection with idempotent reconstruction. In *Advances in Neural Information Processing Systems*, volume 38, pp. 160900–160958, 2025.

Yongqian Sun, Minghan Liang, Zeyu Che, Dongwen Li, Tinghua Zheng, Shenglin Zhang, Pengtian Zhu, Yuzhi Zhang, and Dan Pei. Efficient multivariate time series anomaly detection through transfer learning for large-scale web services. In *2023 IEEE International Conference on Web Services (ICWS)*, pp. 145–151, 2023.

Fengling Wang, Yiyue Jiang, Rongjie Zhang, Aimin Wei, Jingming Xie, and Xiongwen Pang. A survey of deep anomaly detection in multivariate time series: Taxonomy, applications, and directions. *Sensors*, 25(1):190, 2025.

Rui Wang, Xudong Mou, Renyu Yang, Kai Gao, Pin Liu, Chongwei Liu, Tianyu Wo, and Xudong Liu. CutAddPaste: Time series anomaly detection by exploiting abnormal knowledge. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3176–3187, 2024.

Yucheng Wang, Yuecong Xu, Jianfei Yang, Zhenghua Chen, Min Wu, Xiaoli Li, and Lihua Xie. Sensor alignment for multivariate time-series unsupervised domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 10253–10261, 2023.

Yujie Wei, Jiaxin Ye, Zhizhong Huang, Junping Zhang, and Hongming Shan. Online prototype learning for online continual learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 18764–18774, 2023.

Hongwei Wen, Can Zhang, Haoyu He, Hanyuan Hang, and Minglong Lei. Progressive graph structure adjustment for homophily shift adaptation. In *International Conference on Machine Learning*, 2026.

Hanrui Wu, Yanxin Wu, Lei Tian, Zhao-Rong Lai, Jinyi Long, Michael K. Ng, and C. L. Philip Chen. Local and high-order consistency coding and adaptation for cross-hypergraph node classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48:9997–10011, 2026.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. Graph wavenet for deep spatial-temporal graph modeling. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pp. 1907–1913, 2019.

Yongzheng Xie, Hongyu Zhang, and Muhammad Ali Babar. Multivariate time series anomaly detection by capturing coarse-grained intra-and inter-variate dependencies. In *Proceedings of the ACM on Web Conference 2025*, pp. 697–705, 2025.

Anyi Zhang and Peng Xu. PRACTISED: Pattern-based representation alignment for cross-domain time series anomaly detection. In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*, pp. 2291–2298, 2025.

Xiao Zhang, Shuqing Xu, Huashan Chen, Zekai Chen, Fuzhen Zhuang, Hui Xiong, and Dongxiao Yu. Rethinking robust multivariate time series anomaly detection: A hierarchical spatio-temporal variational perspective. *IEEE Transactions on Knowledge and Data Engineering*, 36(12):9136–9149, 2024.

Zhen Zhang and Bingsheng He. Towards unsupervised open-set graph domain adaptation via dual reprogramming. In *Advances in Neural Information Processing Systems*, volume 38, pp. 53320–53343, 2025.

Hang Zhao, Yujing Wang, Juanyong Duan, Congrui Huang, Defu Cao, Yunhai Tong, Bixiong Xu, Jing Bai, Jie Tong, and Qi Zhang. Multivariate time-series anomaly detection via graph attention network. In *2020 IEEE international conference on data mining (ICDM)*, pp. 841–850, 2020.

Lecheng Zheng, Dongqi Fu, Zihao Li, and Jingrui He. OwlEye: Zero-shot learner for cross-domain graph data anomaly detection. In *International Conference on Learning Representations*, pp. 33497–33522, 2026.

Minglei Zheng, Junfeng Man, Dian Wang, Yanan Chen, Qianqian Li, and Yong Liu. Semi-supervised multivariate time series anomaly detection for wind turbines using generator SCADA data. *Reliability Engineering & System Safety*, 235:109235, 2023.

Qihang Zhou, Shibo He, Haoyu Liu, Jiming Chen, and Wenchao Meng. Label-free multivariate time series anomaly detection. *IEEE Transactions on Knowledge and Data Engineering*, 36(7): 3166–3179, 2024.

A METHOD DETAILS

A.1 SPATIO-TEMPORAL ENCODER

For a window $\boldsymbol{X}_i^r$ from domain $r \in \{s, t\}$, each scalar observation is first combined with temporal and variable embeddings as:

$$\boldsymbol{H}_{i,t,j}^{r,(0)} = \boldsymbol{W}_{\text{in}} X_{i,t,j}^r + \boldsymbol{e}_t^{\text{time}} + \boldsymbol{e}_{r,j}^{\text{var}}. \tag{14}$$

The source and target spatio-temporal encoders are independently parameterized to capture the domain-specific dependency structures within each domain. Each domain-specific spatio-temporal encoder follows the local message passing and global attention structure described in Section 3.1. Specifically, at layer ℓ, the local and global branches operate in parallel on $\boldsymbol{H}_i^{r,(\ell)}$ for each domain $r \in \{s, t\}$, as:

$$\boldsymbol{H}_{i,\text{loc}}^{r,(\ell)} = \text{BN}_r\Big(\boldsymbol{H}_i^{r,(\ell)} + \text{GConv}_r(\boldsymbol{H}_i^{r,(\ell)}, \boldsymbol{A}_r)\Big), \tag{15}$$

$$\boldsymbol{H}_{i,\text{glob}}^{r,(\ell)} = \text{BN}_r\Big(\boldsymbol{H}_i^{r,(\ell)} + \text{vec}^{-1}\Big(\text{MHA}_r\big(\text{vec}(\boldsymbol{H}_i^{r,(\ell)}); \boldsymbol{M}\big)\Big)\Big), \tag{16}$$

$$\boldsymbol{H}_i^{r,(\ell+1)} = \text{BN}_r\Big(\boldsymbol{H}_{i,\text{loc}}^{r,(\ell)} + \boldsymbol{H}_{i,\text{glob}}^{r,(\ell)} + \text{FFN}_r\Big(\boldsymbol{H}_{i,\text{loc}}^{r,(\ell)} + \boldsymbol{H}_{i,\text{glob}}^{r,(\ell)}\Big)\Big), \tag{17}$$

where $\text{GConv}_r(\cdot), \text{MHA}_r(\cdot), \text{BN}_r(\cdot)$, and $\text{FFN}_r(\cdot)$ denote the graph convolution, multi-head self attention, batch normalization, and feedforward network operations, respectively, applied on domain r, $\text{vec}(\cdot)$ flattens the T time steps and d_r variables into a sequence of $T \times d_r$ hidden states for global self-attention, and $\text{vec}^{-1}(\cdot)$ reshapes the output sequence back to the $T \times d_r$ layout. To prevent the global attention branch from incorporating the future time step information during one-step forecasting, a block-causal mask $\boldsymbol{M}$ is added to the attention logits, where $M_{(t,j),(t',j')} = 0$ for $t' \le t$ and $-\infty$ otherwise. After the final block, the fused local and global representations are summarized by attention-based pooling to obtain $\boldsymbol{h}_i^r$.

A.2 RECONSTRUCTION, FORECASTING AND CLASSIFICATION LOSSES

For $r \in \{s, t\}$, let $\hat{\boldsymbol{X}}_i^{r,\text{rec}} = \boldsymbol{W}_{\text{rec}}^r \text{GRU}_{\text{rec}}^r(\boldsymbol{h}_i^r) + \boldsymbol{b}_{\text{rec}}^r$ and $\hat{X}_{i,t,j}^{r,\text{for}} = \boldsymbol{W}_{\text{for}}^r \boldsymbol{H}_{i,t,j}^{r,(L)} + b_{\text{for}}^r$ denote the outputs of the reconstruction and one-step forecasting heads, respectively, where $\text{GRU}_{\text{rec}}^r(\cdot)$ denotes the domain-specific Gated Recurrent Unit (GRU) decoder in the reconstruction head. The corresponding per-window reconstruction and forecasting errors are computed as:

$$e_i^{r,\text{rec}} = \frac{\|\hat{\boldsymbol{X}}_i^{r,\text{rec}} - \boldsymbol{X}_i^r\|_F^2}{T d_r}, \tag{18}$$

$$e_i^{r,\text{for}} = \frac{\|\hat{\boldsymbol{X}}_{i,1:T-1}^{r,\text{for}} - \boldsymbol{X}_{i,2:T}^r\|_F^2}{(T-1) d_r}. \tag{19}$$

The reconstruction loss and forecasting loss are averaged over all windows within each mini-batch of domain r, as:

$$\mathcal{L}_{\text{rec}}^r = \frac{1}{|\mathcal{B}_r|} \sum_{i \in \mathcal{B}_r} e_i^{r,\text{rec}}, \tag{20}$$

$$\mathcal{L}_{\text{for}}^r = \frac{1}{|\mathcal{B}_r|} \sum_{i \in \mathcal{B}_r} e_i^{r,\text{for}}. \tag{21}$$

Given the anomaly scores S_i^r predicted by the binary classifier in Eq.12, the observed source labels and pseudo target labels, the source and target classification losses are defined by the binary cross-entropy (BCE) losses, as:

$$\mathcal{L}_{\text{cls}}^s = \frac{1}{|\mathcal{B}_s|} \sum_{i \in \mathcal{B}_s} \text{BCE}(S_i^s, y_i^s), \tag{22}$$

$$\mathcal{L}_{\text{cls}}^t = \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} \text{BCE}(S_i^t, \tilde{y}_i^t), \qquad \mathcal{I}_t = \mathcal{N}_t \cup \mathcal{A}_t. \tag{23}$$

Algorithm 1 MNP-GDA.

Input: Labeled source domain $\mathcal{D}_s$ and unlabeled target domain $\mathcal{D}_t$.
Output: Target-domain anomaly scores $\{S_i^t\}$.

1: Initialize learnable model parameters Θ.
2: Obtain the initial source-normal projections using initialized source encoder and shared projection head.
3: Initialize the normal prototype bank $\mathcal{P} = \{\boldsymbol{p}_k\}_{k=1}^K$ by K-means on these projections.
4: **for** $e = 1, \ldots, E$ **do**
5: Set the warm-up coefficient $\alpha_{\mathrm{pda}}(e)$ for prototypical domain adaptation.
6: **for** each paired source and target mini-batch $\mathcal{B}_s$ and $\mathcal{B}_t$ **do**
7: Learn adaptive inter-variable dependency matrices $\boldsymbol{A}_s$ and $\boldsymbol{A}_t$ by Eq. 1, and obtain the corresponding spatio-temporal representations by Eq. 2.
8: Compute the reconstruction and forecasting losses $\mathcal{L}_{\mathrm{rec}}^r$ and $\mathcal{L}_{\mathrm{for}}^r$, for each domain $r \in \{s, t\}$ by Eqs. 18–21.
9: Align window representations of both domains into a common latent space through a shared projection head to obtain the projected representation $\boldsymbol{z}_i^r$, and compute the nearest-prototype distance $d_i^{r,\mathrm{proto}}$.
10: Min–max normalize the three anomaly signals within each mini-batch and compute the target selection score s_i^{sel} by Eq. 7.
11: Select the top-$\rho_a\%$ highest-scoring target samples as the pseudo-anomalous set $\mathcal{A}_t$.
12: **for** $k = 1, \ldots, K$ **do**
13: Select the top-$\rho_n\%$ lowest-scoring target samples assigned to $\boldsymbol{p}_k$ as the pseudo-normal subset $\mathcal{N}_t^k$.
14: **end for**
15: Set the pseudo-normal set $\mathcal{N}_t = \left(\bigcup_{k=1}^K \mathcal{N}_t^k\right) \setminus \mathcal{A}_t$.
16: Compute the anomalous and normal adaptive reliability weights w_i^N and w_i^A by Eqs. 8 and 9.
17: Compute the source classification loss $\mathcal{L}_{\mathrm{cls}}^s$ and target classification loss $\mathcal{L}_{\mathrm{cls}}^t$ by Eqs. 22 and 23.
18: **if** $\alpha_{\mathrm{pda}}(e) > 0$ **then**
19: Compute the weighted prototypical domain adaptation loss $\mathcal{L}_{\mathrm{pda}}$ by Eq. 10.
20: Update each normal prototype upon its assigned source-normal projections by momentum moving averaging strategy.
21: **else**
22: Set the weighted prototypical domain adaptation loss $\mathcal{L}_{\mathrm{pda}} = 0$.
23: **end if**
24: Compute the overall training loss $\mathcal{L}$ by Eq. 13 and update Θ.
25: **end for**
26: **end for**
27: Using the optimized model parameters Θ^* to obtain representations, three anomaly signals, and normal prototype bank.
28: Predict target anomaly scores $\{S_i^t\}$ by the binary classifier by Eq. 12.

A.3 Algorithm of MNP-GDA

Algorithm 1 summarizes the overall training procedure of MNP-GDA. The learnable model parameters are first initialized, after which the initial source-normal projections are obtained using the initialized source encoder and shared projection head and clustered by K-means to initialize the normal prototype bank (lines 1–3). During each training iteration, the source and target mini-batches are encoded through their domain-specific adaptive graphs and spatio-temporal encoders, while reconstruction error, forecasting error, and nearest-prototype distance are integrated to compute the target selection score (lines 7–10). The pseudo-normal and pseudo-anomalous target samples are then selected, whose contributions to prototypical domain adaptation are modulated by the label-conditioned adaptive reliability weights (lines 11–16). The source and target classification losses are computed based on the predicted anomaly score by the binary classifier, and the observed source labels and target pseudo-labels (line 17). The weighted prototypical domain adaptation loss $\mathcal{L}_{\mathrm{pda}}$ is computed upon the selected pseudo-labeled target samples and the source samples with observed labels. Meanwhile, the normal prototypes are updated upon the source normal projections using a momentum moving average strategy, preserving the source-supervised normality anchors throughout training (lines 18–23). The model is optimized by minimizing the overall loss (line 24). After training, the optimized model and prototype bank are used to compute the target-domain anomaly scores (lines 27–28).

B Proof of Theorem 1

Adding a prototype cannot increase the pointwise nearest-prototype distance, and therefore $D^*_{K+1} \leq D^*_K$. For $K \geq M$, choose one prototype at each mode direction $\boldsymbol{v}_m$. Then

$$D^*_K \leq \sum_{m=1}^{M} \pi_m \, \mathbb{E}_{\boldsymbol{z}\sim P_m} \left[1 - \langle \boldsymbol{z}, \boldsymbol{v}_m \rangle\right] = \sigma^2. \tag{24}$$

For a single unit prototype $\boldsymbol{p}$, $D_1(\boldsymbol{p}) = 1 - \langle \sum_{m=1}^{M} a_m \boldsymbol{v}_m, \boldsymbol{p} \rangle$, which is minimized when $\boldsymbol{p}$ is aligned with $\sum_{m=1}^{M} a_m \boldsymbol{v}_m$. Hence $D^*_1 = 1 - \| \sum_{m=1}^{M} a_m \boldsymbol{v}_m \|_2$. Since $\sigma^2 = 1 - C$,

$$D^*_1 - D^*_K \geq C - \left\| \sum_{m=1}^{M} a_m \boldsymbol{v}_m \right\|_2. \tag{25}$$

Moreover,

$$\begin{aligned} C^2 - \left\| \sum_{m=1}^{M} a_m \boldsymbol{v}_m \right\|_2^2 &= 2 \sum_{1 \leq a < b \leq M} a_a a_b \left(1 - \langle \boldsymbol{v}_a, \boldsymbol{v}_b \rangle\right) \\ &\geq 4\delta \sum_{1 \leq a < b \leq M} a_a a_b. \end{aligned} \tag{26}$$

Dividing by $C + \| \sum_{m=1}^{M} a_m \boldsymbol{v}_m \|_2 \leq 2C$ gives

$$C - \left\| \sum_{m=1}^{M} a_m \boldsymbol{v}_m \right\|_2 \geq \frac{2\delta}{C} \sum_{1 \leq a < b \leq M} a_a a_b > 0, \tag{27}$$

which proves Theorem 1.

C Proof of Theorem 2

By the definition of covariance,

$$\mathbb{E}_g[w\eta_g] = \mathbb{E}_g[w] \, \mathbb{E}_g[\eta_g] + \mathrm{Cov}_g(w, \eta_g). \tag{28}$$

Dividing by $\mathbb{E}_g[w] > 0$ gives

$$\bar{\eta}_{w,g} = \bar{\eta}_g + \frac{\mathrm{Cov}_g(w, \eta_g)}{\mathbb{E}_g[w]} \leq \bar{\eta}_g. \tag{29}$$

Conditional on $\boldsymbol{X}$, the pseudo-label loss differs from the true-label loss only when $\tilde{y}_g \neq y^*$. The bounded label-change assumption therefore implies

$$|\mathbb{E}[\ell_\theta(\boldsymbol{X}, \tilde{y}_g) \mid \boldsymbol{X}] - \ell_\theta(\boldsymbol{X}, y^*)| \leq B_g \eta_g(\boldsymbol{X}). \tag{30}$$

Multiplying by $w(\boldsymbol{X})$, taking $\mathbb{E}_g$, and normalizing by $\mathbb{E}_g[w]$ yields

$$\left| R_{w,g}(\theta) - R^*_{w,g}(\theta) \right| \leq B_g \bar{\eta}_{w,g} \leq B_g \bar{\eta}_g. \tag{31}$$

The reduction relative to the unweighted upper bound is

$$B_g(\bar{\eta}_g - \bar{\eta}_{w,g}) = -\frac{B_g \, \mathrm{Cov}_g(w, \eta_g)}{\mathbb{E}_g[w]}, \tag{32}$$

which is nonnegative under the covariance condition and completes the proof.

D Details of Experimental Setup

D.1 Datasets and Preprocessing

We conducted experiments on five real-world MTS-AD benchmark datasets, covering the server-monitoring systems and industrial control systems:

- **Server Machine Dataset (SMD)** (Su et al., 2019) is a server-monitoring dataset collected from a large Internet company over five weeks. The original dataset contains monitoring data from 28 machines, each characterized by 38 system metrics.
- **Pooled Server Metrics (PSM)** (Abdulaal et al., 2021) consists of application-server monitoring data collected from eBay and contains 25 system metrics.
- **JD1** (Ma et al., 2021) was collected from a large-scale content platform and contains monitoring data from 30 online services, each characterized by 19 system metrics. The dataset comprises 20 training days and 25 test days.
- **Secure Water Treatment (SWaT)** (Mathur & Tippenhauer, 2016) was collected from a water-treatment testbed and contains measurements from 51 sensors and actuators over 11 days, including seven days of normal operation and four days containing cyber-physical attacks.
- **Water Distribution (WADI)** (Ahmed et al., 2017) was collected from a water-distribution testbed and contains measurements from 123 sensors and actuators over 16 days, including 14 days of normal operation and two days containing attack scenarios.

For SMD and PSM, we adopted the multivariate versions released in TSB-AD-M (Liu & Paparrizos, 2024), which were constructed through systematic flaw identification and label-quality assessment. For data preprocessing, each domain was standardized independently, and the raw multivariate time series were partitioned into fixed-length windows using a sliding window. Following cross-domain MTS-AD literature (He et al., 2024), the window length was set to 15 for all datasets, with a stride of 5 for training to reduce redundancy among highly overlapping windows, and a stride of 1 for test to provide dense coverage of the test sequence. For SWaT and WADI, we further applied down-sampling factors of 4 and 2, respectively, and removed invariant variables, resulting in 44 and 96 variables. A window is labeled as anomalous if it contains at least one anomalous timestamp. Table 3 summarizes the statistics of the five datasets after preprocessing. Their variable dimensions range from 19 to 96, reflecting different variable sets across heterogeneous MTS domains.

Table 3: Statistics of five benchmark datasets.

Dataset	# Variables	# Samples	# Anomalous samples	Anomalous ratio (%)
SMD	38	207137	7517	3.63
PSM	25	217624	24381	11.20
JD1	19	216000	11348	5.25
SWaT	44	112480	13681	12.16
WADI	96	86401	4996	5.78

D.2 Implementation Details

We implemented MNP-GDA[1] in PyTorch 2.0.0 and PyTorch Geometric 2.3.1, and optimized it with the Adam optimizer. All experiments were conducted on a single NVIDIA GeForce RTX 3090 GPU with 24 GB of memory and an Intel Xeon Gold 6330 CPU. For each task, the proposed MNP-GDA was trained with 20 epochs, a batch size of 128, two layers of spatio-temporal encoder, two attention heads, dropout of 0.2, and a prototype momentum of 0.999. In addition, Table 4 specifies the settings of other hyper-parameters on eight tasks, including the learning rate (LR), the window representation dimension d_h, the projection dimension d_p, the number of normal prototypes K, the temperature parameter τ, the minimum reliability weight λ, the pseudo-normal and -anomaly selection ratios ρ_n and ρ_a, the weights of anomaly repulsion, reconstruction, forecasting, and prototypical domain adaptation losses α_{push}, α_{rec}, α_{for}, α_{pda}. The min–max normalization $\sigma(\cdot)$ of the three anomaly signals is computed within each mini-batch during training. During inference, the three anomaly signals are first computed for all target windows, and $\sigma(\cdot)$ is then applied over the whole target set.

D.3 Details of Baselines

We compared MNP-GDA against nine SOTA baselines in three categories, as follows:

1) *Single-domain MTS-AD*:

[1]The code of MNP-GDA will be made publicly available upon acceptance of the paper.

Table 4: Hyperparameter settings of MNP-GDA on eight tasks.

Task	LR	d_h	d_p	K	τ	λ	$\rho_n(\%)$	$\rho_a(\%)$	α_{push}	α_{rec}	α_{for}	α_{pda}
SMD→PSM	1e-3	64	32	5	0.5	0.3	50	1	0.1	10	1	0.1
SMD→JD1	1e-3	64	32	10	0.5	0.5	50	1	10	0.1	5	0.1
PSM→SMD	1e-3	64	64	5	0.1	0.3	50	5	5	0.1	5	0.1
PSM→JD1	1e-3	64	64	10	0.1	0.5	50	5	10	0.1	10	0.1
JD1→SMD	1e-3	64	32	5	0.1	0.5	50	5	0.1	0.1	10	1
JD1→PSM	1e-3	64	32	10	0.1	0.5	90	1	10	10	1	1
SWaT→WADI	1e-4	32	32	5	0.5	0.7	50	1	0.1	10	10	1
WADI→SWaT	1e-4	32	32	5	0.5	0.3	50	10	1.0	0.01	10	1

- **OmniAnomaly** (Su et al., 2019) learns a stochastic recurrent reconstruction model and scores deviations from reconstructed normal behavior.
- **THOC** (Shen et al., 2020) uses a temporal hierarchical one-class objective to represent normal sequences.

2) *Single-domain graph-based MTS-AD*:

- **MTAD-GAT** (Zhao et al., 2020) combines temporal and feature-oriented graph attention with reconstruction and forecasting.
- **HSTVAE** (Zhang et al., 2024) models hierarchical spatio-temporal variation with a variational autoencoder.
- **MTGFlow** (Zhou et al., 2024) estimates entity-aware densities with a dynamic graph and normalizing flows.

To tailor the single-domain baselines to cross-domain MTS-AD setting, they were trained on the source-domain windows and tested on the target-domain windows. To accommodate unequal variable dimensions across domains, linear projections were applied to map the original data from the source and target domains into a common dimension before inputting to each baseline.

3) *Domain adaptation on MTS*:

- **CLUDA** (Ozyurt et al., 2023) is an unsupervised domain adaptation method for time series through contrastive contextual representations and domain alignment.
- **ACON** (Liu et al., 2024a) conducts time series classification adaptation through temporal–frequency co-learning and domain alignment.
- **DACAD** (Darban et al., 2025) combines supervised source contrast, self-supervised target triplets, adversarial alignment, anomaly injection, and a DeepSVDD-style center detector for MTS-AD across domains. It adopts the target representation's distance from the learned center as the anomaly score.
- **VANDA** (He et al., 2024) models intra-domain variate associations and transfers inter-domain associations through similarity and optimal transport for MTS-AD across domains.

It is worth noting that CLUDA and ACON were originally developed for MTS classification across domains. To tailor CLUDA to cross-domain MTS-AD, we coupled its representation and adaptation objectives with a binary classification head trained on the source labels. Besides, for ACON, we added binary classification heads for the temporal and frequency branches while retaining its adaptation objectives.

DACAD and VANDA are SOTA cross-domain MTS-AD methods. While DACAD assumes that the source and target domains share the same variable set, to tailor it to different variable dimensions across heterogeneous MTS domains, we added linear projections to map the source and target inputs into a common dimension. In addition, VANDA originally computes the anomaly score for each timestamp, to tailor it to window-level MTS-AD task, we aggregated the timestamp-level anomaly scores within each window by mean-pooling to obtain the window-level anomaly scores.

E PARAMETER SENSITIVITY

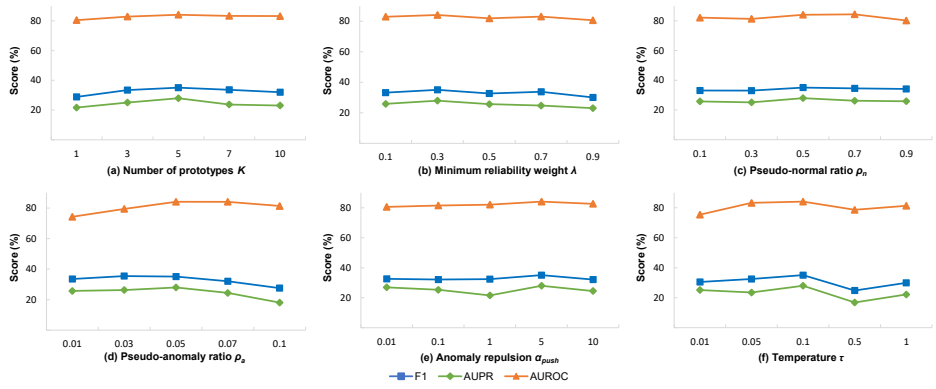

Figure 3: The sensitivity of MNP-GDA to the hyper-parameters K, λ, ρ_n, ρ_a, α_{push}, and τ on the representative task PSM→SMD.

We investigate the sensitivity of MNP-GDA to the hyper-parameters, and the results are presented in Figure 3. **1) Impact of the number of normal prototypes K.** As shown in Figure 3(a), MNP-GDA achieves the best performance when the number of normal prototypes K is set to 5, while a smaller or larger K would degrade the performance. This might be because too few prototypes can hardly capture multimodal normalities, while too many prototypes may introduce redundant normal prototypes. **2) Impact of the minimum reliability weight λ.** As shown in Figure 3(b), MNP-GDA achieves the best performance when λ is set to 0.3, while a larger λ would gradually degrade the performance. This might be because too large λ would reduce the difference between the reliability weights of target samples, thereby weakening the effect of adaptive reliability weighting. **3) Impact of the pseudo-normal and -anomaly selection ratios ρ_n and ρ_a.** As shown in Figure 3(c), MNP-GDA can achieve good performance when the pseudo-normal selection ratio ρ_n is not too small, while too small ρ_n would degrade the performance. This might be because too few pseudo-normal samples can hardly provide sufficient target normality information for domain adaptation. Additionally, as shown in Figure 3(d), MNP-GDA achieves good overall performance when the pseudo-anomaly selection ratio ρ_a is relatively small, while too large ρ_a would degrade the performance. This might be because increasing ρ_a would introduce more unreliable pseudo-anomalous samples into model training. Thus, a larger selection ratio is adopted for pseudo-normal samples, while a relatively smaller ratio is adopted for pseudo-anomalous samples. **4) Impact of the anomaly repulsion α_{push}.** As shown in Figure 3(e), MNP-GDA achieves the best performance when α_{push} is set to 5, while a larger or smaller α_{push} would degrade the performance. This might be because too small α_{push} can hardly separate anomalous samples from normal prototypes, while too large α_{push} would impose an excessive penalty on anomalous samples. **5) Impact of the temperature τ in prototypical domain adaptation.** As shown in Figure 3(f), MNP-GDA yields the best AUROC and AUPR when τ is set to 0.1, while maintaining competitive F1. Although a smaller τ can yield a higher F1, its AUROC is substantially degraded, while larger τ generally leads to worse performance. This might be because too small τ would impose overly sharp prototype discrimination, while too large τ would weaken the discrimination between different normal prototypes.

F ADDITIONAL ABLATION STUDY

1) Effect of different training objectives. As shown in Table 5, removing either $\mathcal{L}_{\text{cls}}$ or $\mathcal{L}_{\text{pda}}$ produces the two largest average performance drops, verifying the essential role of prototypical domain adaptation and classification losses in cross-domain MTS-AD. Removing either $\mathcal{L}_{\text{rec}}$ or $\mathcal{L}_{\text{for}}$ also degrades performance, showing that reconstruction and forecasting losses provide complementary supervision for learning normal dynamics.

2) Effect of different anomaly scoring components. We analyze the effectiveness of different anomaly scoring components, i.e., reconstruction error, forecasting error, and prototype distance. As

shown in Figure 4, among the three individual components, prototype distance achieves the strongest discriminative performance, further validating the effectiveness of the proposed multimodal normality prototypical domain adaptation module, as deviations from the learned normal prototypes provide informative evidence for distinguishing anomalies from diverse normal patterns. Meanwhile, the final anomaly score achieves the highest AUROC and AUPR, demonstrating the effectiveness of jointly considering the complementary signals for MTS-AD.

3) Effect of target selection and adaptive reliability weighting. For selected pseudo-normal and -anomaly target sets, we compare the target class ratio with the unweighted and reliability-weighted selection precision. The target class ratio represents the expected proportion under random selection, thus, a higher selection precision indicates that the corresponding class is effectively concentrated in the selected set. The reliability-weighted selection precision is computed on the same set using adaptive reliability weights, reflecting whether more reliable samples receive larger weights during target adaptation. As shown in Figure 5, the pseudo-anomaly unweighted selection precision substantially exceeds the target anomaly class ratio, demonstrating effective enrichment of rare anomalies during selection. Moreover, the reliability-weighted selection precision for pseudo-anomalies consistently exceeds the unweighted counterpart, indicating that **true anomalies indeed receive larger reliability weights, while incorrectly selected samples are actually downweighted to mitigate their negative effect**. While for pseudo-normal set, as the unweighted selection precision is already high enough, there is a smaller gap between reliability-weighted and unweighted selection precision. Together with the performance degradation under uniform target weights in Table 2, these results again validate the effectiveness of adaptive reliability weighting on mitigating negative influence of unreliable target pseudo-labels.

Table 5: Ablation of training objectives on three representative tasks.

Variant	PSM→SMD			JD1→SMD			WADI→SWaT		
	F1	AUPR	AUROC	F1	AUPR	AUROC	F1	AUPR	AUROC
MNP-GDA (full)	**35.08**	**27.97**	**84.11**	**35.30**	**27.26**	**82.35**	**73.90**	**75.96**	**88.00**
w/o $\mathcal{L}_{\mathrm{pda}}$	18.98	13.34	71.08	23.35	16.18	67.41	61.38	46.43	83.05
w/o $\mathcal{L}_{\mathrm{rec}}$	30.04	20.81	74.32	29.61	21.31	74.41	73.43	73.13	87.87
w/o $\mathcal{L}_{\mathrm{for}}$	27.62	22.81	77.47	25.99	21.02	76.66	71.64	70.50	86.88
w/o $\mathcal{L}_{\mathrm{cls}}$	23.39	19.45	76.76	25.51	17.77	75.19	70.81	57.71	86.60

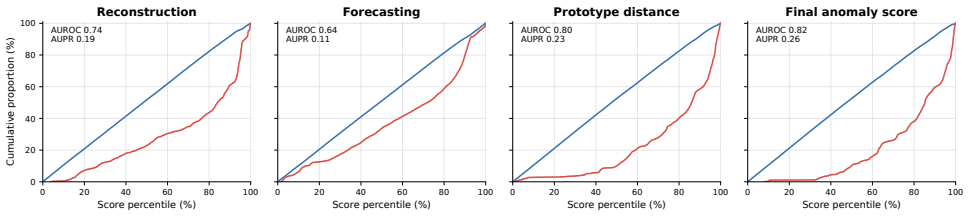

Figure 4: Empirical cumulative distributions of reconstruction error, forecasting error, prototype distance, and the final anomaly score for normal (blue) and anomalous (red) target windows on the SWaT→WADI task.

G Complexity Analysis

Let B denote the batch size, T be the window length, d_r be the number of variables in domain r, d_h be the representation dimension, K be the number of normal prototypes, and d_p be the projection dimension. In spatio-temporal encoder, the time complexity of adaptive graph learning is $\mathcal{O}(\sum_{r\in\{s,t\}} d_r^2 d_h)$, of local graph convolution is $\mathcal{O}(BT\sum_{r\in\{s,t\}}(d_r^2 d_h + d_r d_h^2))$, and of global self-attention is $\mathcal{O}(B\sum_{r\in\{s,t\}}(Td_r)^2 d_h)$. The time complexity of prototypical domain adaptation is $\mathcal{O}(BKd_p)$. The overall time complexity of MNP-GDA is $\mathcal{O}(B\sum_{r\in\{s,t\}}(Td_r)^2 d_h)$.

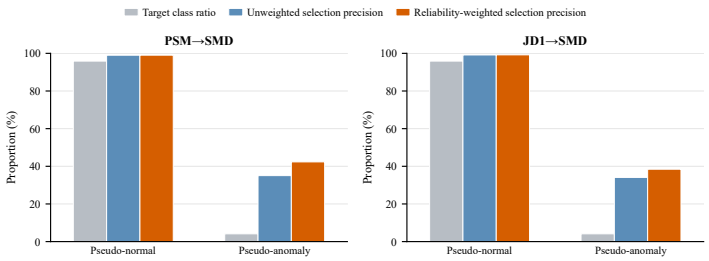

Figure 5: Analysis of target selection and reliability weighting on two representative tasks. Target class ratio is compared with the unweighted and reliability-weighted selection precision of the selected sets of pseudo-normal and -anomaly.

H Efficiency Analysis

Figure 6 compares the training time and AUPR of the domain adaptation methods on the representative SMD→PSM task. ACON has the shortest training time but lower AUPR than VANDA and MNP-GDA, whereas VANDA substantially outperforms CLUDA at a similar training time. DACAD requires the longest training time and yields the lowest AUPR among the comparing methods. In contrast, MNP-GDA achieves the highest AUPR with a moderate training time. Although MNP-GDA requires more training time than some baselines, it remains faster than DACAD while achieving substantially better detection performance. The additional training overhead of MNP-GDA may stem primarily from the dual-branch spatio-temporal encoder, which jointly models local inter-variable and global time-variable dependencies. Overall, MNP-GDA achieves a favorable trade-off between efficiency and cross-domain MTS-AD performance.

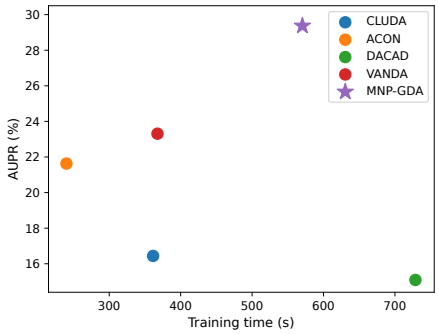

Figure 6: Training-time and AUPR comparison of domain adaptation methods on SMD→PSM task.

I Broader Impact

Advancements in cross-domain MTS-AD could lead to more reliable AI systems in real-world applications, such as online service monitoring, industrial process control, fault diagnosis, spacecraft monitoring, and financial risk mitigation. Our work provides a novel perspective for cross-domain MTS-AD by reformulating it as a time-series-based graph domain adaptation problem. The proposed MNP-GDA offers a principled solution to transfer multimodal normalities across heterogeneous MTS domains and mitigate negative influence of unreliable target pseudo-labels. Moreover, the theoretically grounded ideas of multimodal normality and adaptive reliability weighting are general and have the potential applicability to other cross-domain anomaly detection tasks beyond time series or graph data. Overall, this work not only contributes to cross-domain MTS-AD, but also provides methodological and theoretical insights that may facilitate the development of more accurate and label-efficient anomaly detection on newly deployed systems.