Cross-Domain Conditional Diffusion Models for Time Series Imputation

Kexin Zhang[1⋆], Baoyu Jing[2], K. Selçuk Candan[3], Dawei Zhou[4], Qingsong Wen[5], Han Liu[1], and Kaize Ding[1] (✉)

[1] Northwestern University, Evanston, IL 60208, USA kevin.kxzhang@gmail.com, {hanliu, kaize.ding}@northwestern.edu
[2] University of Illinois Urbana-Champaign, Champaign, IL 61820-5711, USA baoyuj2@illinois.edu
[3] Arizona State University, Tempe, AZ 85281, USA candan@asu.edu
[4] Virginia Tech, Blacksburg, VA 24061-0131, USA zhoud@vt.edu
[5] Squirrel Ai Learning, Bellevue, WA 98004, USA qingsongedu@gmail.com

Abstract. Cross-domain time series imputation is an underexplored data-centric research task that presents significant challenges, particularly when the target domain suffers from high missing rates and domain shifts in temporal dynamics. Existing time series imputation approaches primarily focus on the single-domain setting, which cannot effectively adapt to a new domain with domain shifts. Meanwhile, conventional domain adaptation techniques struggle with data incompleteness, as they typically assume the data from both source and target domains are fully observed to enable adaptation. For the problem of cross-domain time series imputation, missing values introduce high uncertainty that hinders distribution alignment, making existing adaptation strategies ineffective. Specifically, our proposed solution tackles this problem from three perspectives: **(i) Data:** We introduce a frequency-based time series interpolation strategy that integrates shared spectral components from both domains while retaining domain-specific temporal structures, constructing informative priors for imputation. **(ii) Model:** We design a diffusion-based imputation model that effectively learns domain-shared representations and captures domain-specific temporal dependencies with dedicated denoising networks. **(iii) Algorithm:** We further propose a cross-domain consistency alignment strategy that selectively regularizes output-level domain discrepancies, enabling effective knowledge transfer while preserving domain-specific characteristics. Extensive experiments on three real-world datasets demonstrate the superiority of our proposed approach. Our code implementation is available here[6].

Keywords: Time Series Imputation · Domain Adaptation · Conditional Diffusion Models.

⋆ Kexin Zhang is a research intern during the completion of this work.
[6] https://github.com/kexin-kxzhang/CD2-TSI

1 Introduction

Multivariate time series imputation is essential for various real-world applications, including environmental monitoring and energy management [44]. Missing values commonly arise due to sensor failures, transmission errors, or external disruptions, leading to incomplete data that could degrade the reliability of downstream tasks [18]. Effective imputation is thus critical for preserving the integrity of the data and ensuring reliable results in subsequent applications [16].

Many endeavors have been made to model the temporal patterns inherent in time series. Traditional statistical and machine learning methods for time series imputation often assume stationarity or linear relationships, which may not capture the full complexity of real-world time series data. Recurrent neural networks and attention-based models have improved the modeling of temporal dependencies by capturing nonlinear relationships [2]. More recently, deep generative models [38, 10], such as variational autoencoders and generative adversarial networks, have been explored for time series imputation. Diffusion-based models [31, 1, 19, 45] further advance imputation by learning a denoising process that iteratively refines missing values.

Despite their success, these methods struggle under high missing rates [31], as sparse observations hinder the effective modeling of the underlying temporal dependencies [7, 45, 11]. When observations are highly incomplete, it is natural to leverage related domains to improve imputation performance [5]. For instance, in air quality monitoring, neighboring cities' sensor networks may provide complementary temporal patterns when local sensors fail. Recent advancements in domain adaptation (DA) have shown promising results in transferring knowledge across domains in tasks such as time series forecasting and classification [14, 27]. In light of this, we propose to tackle the novel cross-domain time series imputation by adapting domain discrepancies between two related domains.

However, cross-domain time series imputation remains largely underexplored and directly applying DA techniques to time series imputation may easily fail due to the following challenges: **(1)** ***Data Challenge:*** Most of the DA problems usually assume the observed data from both source and target domains are complete [41], however, in time series imputation, it is hard for existing models to well characterize the real data distributions due to the high missingness in the observed data. **(2)** ***Model Challenge*****:** Existing approaches commonly rely on simply training a single shared model on mixed data, which cannot distinguish domain-shared and specific knowledge [42]. Therefore, it is necessary to develop a model that can both facilitate knowledge transfer and capture domain-specific patterns. **(3)** ***Algorithm Challenge*****:** Time series data from different domains often exhibit domain-specific temporal dependencies, leading to variations in seasonality, trend shifts, or periodic patterns [40], etc. Existing domain adaptation algorithms often enforce alignment without considering such differences across domains, thus failing to capture the cross-domain knowledge required for accurate imputation in the target domain.

To address these challenges, we propose a novel **C**ross-**D**omain **C**onditional **D**iffusion Model for **T**ime **S**eries **I**mputation **(CD^2-TSI)**, which improves im-

putation in the target domain by leveraging knowledge from a source domain while preserving domain-specific temporal patterns. Specifically, to counter the data challenge, we introduce a frequency-based time series interpolation strategy, which interpolates original missing values by integrating shared spectral components from both the source and target domains. The pre-interpolated values are used to construct the missing targets, providing more informative priors for training the imputation model. For the model challenge, we develop a diffusion-based imputation framework that learns domain-shared representations to capture common patterns across domains while maintaining dedicated denoising networks to model domain-specific temporal dependencies. To tackle the algorithm challenge, we propose a cross-domain consistency alignment algorithm that imposes alignment based on output-level discrepancy. The degree of alignment is adjusted according to the prediction difference between source and target networks for the same target samples. This approach facilitates cross-domain transfer while preserving target-specific temporal characteristics, preventing the model from overfitting to source domain patterns.

In summary, the main contributions of this work are summarized as follows: (1) We target the problem of cross-domain time series imputation, which is largely underexplored and requires research attention in the community. (2) We propose CD^2-TSI, a new diffusion model-based framework that solves the problem of cross-domain time series imputation from data, model, and algorithm perspectives. (3) We conduct extensive experiments on three real-world datasets, demonstrating that CD^2-TSI outperforms state-of-the-art models across various missing data patterns, highlighting its effectiveness in cross-domain settings.

2 Related Work

2.1 Time Series Imputation

Time series imputation (TSI) methods can be broadly categorized into predictive and generative approaches [7]: (1) Predictive methods [34, 6, 22] predict deterministic values but suffer from error accumulation and fail to capture the uncertainty of missing values. GRU-D [3] and BRITS [2] use deep autoregressive models with time decay, while GRIN [4] incorporates graph neural networks (GNN) for spatial relationships. (2) Generative methods, such as those based on Variational Autoencoders (VAE), Generative Adversarial Networks (GAN), and diffusion models, effectively circumvent the limitations faced by those predictive models. VAE-based methods [24, 15] optimize reconstruction error and regularize the latent space. GAN-based approaches [23] use adversarial training between the generator and discriminator but can be unstable and produce unrealistic results. Diffusion models show promise due to their ability to model complex data distributions and generate varied outputs for missing values. CSDI [31] and SSSD [1] use observed data as conditional information; PriSTI [19] extracts conditional information and considers spatiotemporal dependencies using geographic data; MTSCI [45] incorporates a complementary mask strategy and a mixup mechanism to realize intra-consistency and inter-consistency. However,

these methods focus primarily on modeling temporal dependencies within a single domain, overlooking the complexities posed by cross-domain scenarios where domain shifts in missing patterns or temporal dynamics exist. Our CD^2-TSI framework addresses this gap by combining diffusion models with domain adaptation to enhance imputation quality across domains.

2.2 Time Series Domain Adaptation

Domain Adaptation (DA) [9] seeks to transfer knowledge to a target domain by leveraging information from source domains. These methods can be categorized into three groups: (1) Adversarial-based methods train a domain discriminator to identify domains while learning transferable features. For example, CoDATS [33] employs a gradient reversal layer for adversarial training with weak supervision on multi-source data. SLARDA [26] aligns temporal dynamics across domains via autoregressive adversarial training. (2) Discrepancy-based methods use statistical distances to align features from source and target domains. AdvSKM [20] leverages maximum mean discrepancy (MMD) with a hybrid spectral kernel for temporal domain adaptation. RAINCOAT [12] tackles feature and label shifts via temporal and frequency feature alignment. (3) Self-supervision methods incorporate auxiliary tasks. DAF [14] uses a shared attention module for domain-invariant and specific features and reconstruction. While existing DA methods have proven effective in forecasting and classification tasks, their application to imputation remains underexplored, where temporal discrepancies as well as data deficiency introduced by missing values pose additional challenges. CD^2-TSI differs from these approaches by addressing the challenges introduced by incomplete observations in cross-domain time series imputation and adaptation.

3 Problem Definition

Cross-domain time series imputation aims to reconstruct missing values in a target domain by leveraging knowledge from a related source domain. Given multivariate time series data with potential missing values, we define the time series in both domains as $\mathbf{X}_d = (\mathbf{X}_{d,1}, \ldots, \mathbf{X}_{d,K}) \in \mathbb{R}^{K \times L}$, where K is the number of features, L is the length of the time series, and $d \in \{Src, Tgt\}$ denotes the source and target domains. We assume all time series in both domains have the same length. An observation mask $\mathbf{M}_d \in \{0,1\}^{K \times L}$ indicates missing values, where $m_{k,l} = 1$ if the value is observed for the k-th feature at the l-th timestamp, and $m_{k,l} = 0$ if the value is missing. Since real-world datasets often lack ground truth for missing data, we artificially mask a subset of observed values for training and evaluation. Following previous work [45, 31], the extended missing targets $\widetilde{\mathbf{X}}_d \in \mathbb{R}^{K \times L}$ include both the original missing values and artificially masked values, with a binary mask $\widetilde{\mathbf{M}}_d \in \{0,1\}^{K \times L}$.

4 Methodology

4.1 Model Overview (CD^2-TSI)

As shown in Fig. 1, CD^2-TSI incorporates a cross-domain diffusion-based framework, where source and target domains share representations while maintaining domain-specific denoising networks. A pre-interpolation strategy is proposed to integrate spectral components from both domains, providing priors for original missing values via cross-domain frequency mixup, while the artificial missing values are retained to construct missing targets. These targets are then corrupted by adding noise to obtain noisy inputs for training the denoising network. To ensure effective adaptation, we introduce cross-domain consistency alignment. This algorithm promotes adaptation based on output discrepancy while preventing excessive regularization that could force the target domain to overly conform to source domain patterns. Overall, CD^2-TSI effectively leverages cross-domain information to improve the imputation quality in the target domain.

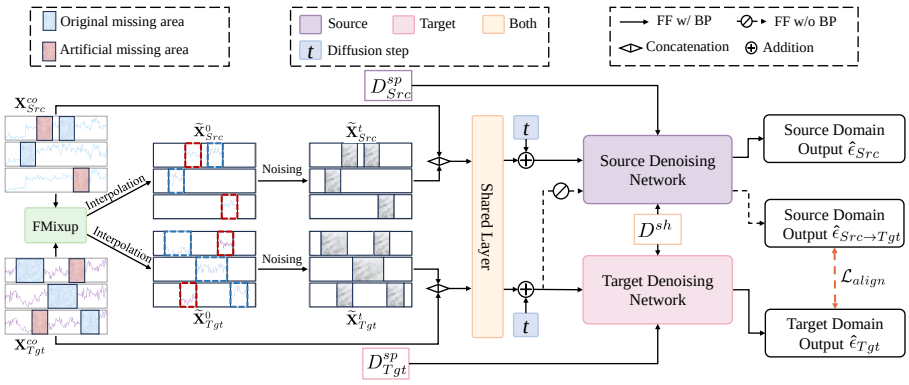

Fig. 1. Architecture of CD^2-TSI. FMixup is utilized to interpolate the original missing areas (blue), while artificial missing values (red) are retained to construct the missing targets $\widetilde{\mathbf{X}}^0$. These targets are then transformed into the noisy targets $\widetilde{\mathbf{X}}^t$ to train the denoising network, with the help of conditional information $\mathbf{X}^{co}$. The framework is optimized using a combination of denoising loss and consistency alignment loss.

4.2 Conditional Diffusion Model for Time Series Imputation

Imputing missing values in time series data requires capturing complex temporal dependencies while addressing challenges from data incompleteness. Our framework takes Denoising Diffusion Probabilistic Models (DDPM) [13] as the base model, where the imputation process is formulated as a conditional generative

task. In the forward process, Gaussian noise is step by step added to the missing targets $\widetilde{\mathbf{X}}^0$ across T diffusion steps, gradually transforming $\widetilde{\mathbf{X}}^0$ into a noisy version $\widetilde{\mathbf{X}}^T$. This process is formalized as follows:

$$q\left(\widetilde{\mathbf{X}}^{1:T} \mid \widetilde{\mathbf{X}}^0\right) = \prod_{t=1}^{T} q\left(\widetilde{\mathbf{X}}^t \mid \widetilde{\mathbf{X}}^{t-1}\right), q\left(\widetilde{\mathbf{X}}^t \mid \widetilde{\mathbf{X}}^{t-1}\right) = \mathcal{N}\left(\widetilde{\mathbf{X}}^t; \sqrt{1-\beta_t}\widetilde{\mathbf{X}}^{t-1}, \beta_t \boldsymbol{I}\right) \tag{1}$$

where β_t represents the noise level, and t indicates the diffusion step. According to DDPM, $\widetilde{\mathbf{X}}^t = \sqrt{\bar{\alpha}_t}\widetilde{\mathbf{X}}^0 + \sqrt{1-\bar{\alpha}_t}\epsilon$, where $\alpha_t = 1-\beta_t, \bar{\alpha}_t = \prod_{i=1}^{t}\alpha_i$, and $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ where $\mathcal{N}$ is Gaussian distribution. When T is large enough, $q\left(\widetilde{\mathbf{X}}^T \mid \widetilde{\mathbf{X}}^0\right)$ approximates a standard normal distribution.

The reverse process then reconstructs the missing targets by iteratively denoising imputed values, conditioned on the remaining observations $\mathbf{X}^{co}$:

$$p_\theta\left(\widetilde{\mathbf{X}}^{t-1} \mid \widetilde{\mathbf{X}}^t, \mathbf{X}^{co}\right) = \mathcal{N}\left(\mu_\theta\left(\widetilde{\mathbf{X}}^t, \mathbf{X}^{co}, t\right), \sigma_t^2 \boldsymbol{I}\right), \tag{2}$$

$$\mu_\theta\left(\widetilde{\mathbf{X}}^t, \mathbf{X}^{co}, t\right) = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\widetilde{\mathbf{X}}^t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta\left(\widetilde{\mathbf{X}}^t, \mathbf{X}^{co}, t\right)\right), \sigma_t^2 = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \tag{3}$$

where $\epsilon_\theta(\cdot)$ is the denoising network with learnable parameters θ. The model is trained to estimate the added noise ϵ given $\widetilde{\mathbf{X}}^t$, conditional observations $\mathbf{X}^{co}$ and current diffusion step t, and the training objective of time series imputation is:

$$\mathcal{L}(\theta) = \mathbb{E}_{\widetilde{\mathbf{X}}^0 \sim q\left(\widetilde{\mathbf{X}}^0\right), \epsilon \sim \mathcal{N}(0,I)} \left\| \epsilon - \epsilon_\theta\left(\widetilde{\mathbf{X}}^t, \mathbf{X}^{co}, t\right) \right\|^2 \tag{4}$$

4.3 Cross-Domain Time Series Frequency Interpolation

Modeling temporal dependencies from incomplete time series is challenging, especially under high missing rates. Severe missingness disrupts the real data distribution, making it difficult for the model to capture consistent temporal dependencies. Addressing this issue is crucial, as many existing methods [31, 1] simply replace original missing values with *zeros* when constructing the missing targets. However, *zeros* cannot reflect the real data distribution, and such a distribution shift makes it more challenging for the diffusion model to accurately recover missing values during the denoising process. Although linear interpolation in the time domain can partially address this issue, it often fails to capture complex non-linear temporal dynamics.

To solve this problem, our intuition is that time series data from related domains typically share low-frequency components, which represent long-term trends or periodic patterns (e.g., daily cycles in hydrology data), while high-frequency components reflect domain-specific details, such as sensor noise or transient fluctuations [35, 43, 36]. Formally, a signal can be decomposed into an amplitude spectrum, which captures the intensity of different frequency components, and a phase spectrum, which preserves local temporal structure. Hence, we propose a frequency-based time series interpolation strategy – FMixup. FMixup

is achieved through two key steps: (1) blending *low-frequency amplitude* spectra across domains and (2) retaining each domain's *high-frequency amplitude* and *phase* spectra. The augmented data can then be used to replace original missing values and refine the missing targets.

Domain-Shared Frequency Mixup. To exchange structural information across domains, we transform the conditional observations $\mathbf{X}^{co} \in \mathbb{R}^{K \times L}$ from both domains into the frequency domain using the Fast Fourier Transform (FFT) [25]:

$$\mathcal{F}(x)(u,v) = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} x(k,l) e^{-j2\pi\left(\frac{k}{K}u + \frac{l}{L}v\right)} \tag{5}$$

where u and v are frequency indices along the two dimensions, and j is the imaginary unit. This frequency space signal $\mathcal{F}(x)$ can be further decomposed into an amplitude spectrum $\mathcal{A} \in \mathbb{R}^{K \times L}$ and a phase spectrum $\mathcal{P} \in \mathbb{R}^{K \times L}$. To integrate common patterns, we introduce a binary mask $\mathcal{M} = \mathbb{1}_{(k,l) \in [-\alpha K:\alpha K, -\alpha L:\alpha L]}$ that selects low-frequency region of the amplitude spectrum, where $\alpha \in (0,1)$ determines the proportion of low-frequency information incorporated. The amplitude spectra of the source and target domains are then blended as follows:

$$\mathcal{A}_{Src \to Tgt} = \mathcal{A}_{Tgt} * (1 - \mathcal{M}) + (\lambda \mathcal{A}_{Tgt} + (1 - \lambda)\mathcal{A}_{Src}) * \mathcal{M} \tag{6}$$

where $\mathcal{A}_{Src}$ and $\mathcal{A}_{Tgt}$ represent the amplitude spectra of the source and target domains, respectively. $\mathcal{A}_{Src \to Tgt}$ is the newly mixed amplitude spectrum, and parameter λ adjusts the balance between the two spectra.

Domain-Specific Frequency Preserving. As mentioned above, high-frequency components often contain domain-specific fine-grained details. To preserve such information, we retain the high-frequency amplitude components of the target domain. Additionally, we do not modify the phase spectrum $\mathcal{P}_{Tgt}$, as it represents local structural information necessary for preserving the original sequence characteristics. The final augmented time series is obtained by combining the mixed amplitude spectrum $\mathcal{A}_{Src \to Tgt}$ with original phase spectrum $\mathcal{P}_{Tgt}$ of the target domain and applying the inverse Fourier transform:

$$X_{Src \to Tgt} = \mathcal{F}^{-1}(\mathcal{A}_{Src \to Tgt}, \mathcal{P}_{Tgt}) \tag{7}$$

This augmented series is used to fill in the original missing values in the target domain. Therefore, we obtain the refined missing targets $\widetilde{\mathbf{X}}^0_{Tgt}$. This frequency-based time series interpolation strategy ensures that the local temporal structure remains aligned with the target domain while benefiting from shared low-frequency trends. Similarly, we obtain the missing targets $\widetilde{\mathbf{X}}^0_{Src}$.

4.4 Cross-Domain Conditional Diffusion Model

Although cross-domain time series frequency mixup provides priors from the data perspective, it does not fully address domain shifts in temporal dynamics. Learning a single shared model for both source and target domains often fails

to capture domain-specific patterns, leading to suboptimal imputation performance in the target domain. To address this, we propose a novel cross-domain conditional diffusion model that enables domain-shared knowledge transfer while modeling domain-specific temporal dependencies.

Domain-Shared Temporal Knowledge Transfer. To facilitate knowledge transfer across domains, we try to learn domain-shared input representations using a shared convolution layer. The input representations for the source and target domains are formulated as: $H^{in}_{Src} = \text{Conv}\left(\mathbf{X}^{co}_{Src} \| \widetilde{\mathbf{X}}^{t}_{Src}\right)$ and $H^{in}_{Tgt} = \text{Conv}\left(\mathbf{X}^{co}_{Tgt} \| \widetilde{\mathbf{X}}^{t}_{Tgt}\right)$, respectively, where Conv is 1×1 convolution. To further integrate shared information and help the imputation, the model incorporates domain-shared side information D^{sh}, which includes: (1) a time embedding $s = \{s_1, \ldots, s_L\} \in \mathbb{R}^{L \times 128}$ for temporal dependencies, constructed using sine-cosine temporal encoding [32]; (2) a learnable feature embedding $f = \{f_1, \ldots, f_K\} \in \mathbb{R}^{K \times 16}$ to model shared feature relationships. We expand and concatenate s and f and obtain $D^{sh} \in \mathbb{R}^{K \times L \times C}$, where C is the channel size.

Domain-Specific Temporal Knowledge Modeling. After the common feature extraction, the model applies domain-specific attention mechanisms:

$$\begin{aligned} H^{tem} &= \text{Attn}_{tem}\left(H^{in} + \text{Linear}(t_{emb})\right) \\ H^{fea} &= \text{Attn}_{fea}\left(H^{tem}\right) \end{aligned} \tag{8}$$

where $\text{Attn}_{tem}(\cdot)$ captures temporal dependencies, and $\text{Attn}_{fea}(\cdot)$ models feature interactions. These attention layers are domain-specific, enabling the model to learn unique characteristics within each domain. The diffusion step embedding t_{emb} is constructed through sine-cosine temporal encoding as well and projected through a linear layer. Additionally, domain-specific side information D^{sp} includes the conditional mask $\widetilde{\mathbf{M}}$ of each domain, which explicitly indicates missing positions. The final output of each denoising network is computed as:

$$H^{out} = H^{fea} + \text{Conv}(D^{sh}) + \text{Conv}(D^{sp}). \tag{9}$$

The domain-specific modeling stacks multiple layers, where the output H^{out} of each layer is divided into a residual connection and a skip connection after a gated activation unit. The residual connection serves as the input to the next layer, while the skip connections from each layer are summed and passed through two layers of 1×1 convolution to obtain the final output.

4.5 Cross-Domain Consistency Alignment (CDCA)

To mitigate temporal discrepancies across domains while accounting for the uncertainty caused by missing values, we further propose cross-domain consistency alignment. Unlike conventional domain adaptation methods, which enforce rigid alignment regardless of the magnitude of domain discrepancies, CDCA selectively enforces prediction consistency based on the model output discrepancy for the same target domain samples.

Let $\hat{\epsilon}_{Tgt}$ denote the target network's prediction on a given target sample, and let $\hat{\epsilon}_{Src\to Tgt}$ denote the prediction from the source network (in evaluation mode) when the same target sample is used as input. The average absolute difference between these predictions, denoted as Δ, is then computed:

$$\Delta = \frac{1}{N}\sum_{i=1}^{N} \|\hat{\epsilon}_{Tgt,i} - \hat{\epsilon}_{Src\to Tgt,i}\| \tag{10}$$

where N is the number of target domain samples.

CDCA compares Δ against two thresholds: a lower threshold τ_l and an upper threshold τ_h. If $\Delta < \tau_l$, the discrepancy is within an acceptable range, where enforcing alignment could amplify the impact of missingness-induced noise rather than improving adaptation. If $\tau_l \leq \Delta \leq \tau_h$, the discrepancy is moderate. In this case, we impose a penalty proportional to the excess difference, specifically $\Delta - \tau_l$. If $\Delta > \tau_h$, the discrepancy is large, suggesting that strict alignment could cause overfitting to source domain patterns and distort intrinsic target structures. To prevent this, the penalty is capped at $\min(\Delta - \tau_l, \tau_h)$. Thus, the alignment loss is formulated as:

$$\mathcal{L}_{align} = \begin{cases} 0, & \Delta < \tau_l, \\ \min\left(\Delta - \tau_l, \tau_h\right), & \Delta \geq \tau_l. \end{cases} \tag{11}$$

By applying regularization only when discrepancies exceed a lower threshold and capping penalties for large deviations, CDCA achieves a trade-off between imposing cross-domain alignment and preserving target-specific characteristics.

4.6 Overall Loss Function

The overall loss function for our model integrates several components: time series imputation losses (Eq. 4) for both source and target domains, and an auxiliary loss that addresses cross-domain consistency alignment (Eq. 11), re-weighted by parameters μ_{align}. The overall loss function is defined as:

$$\mathcal{L} = \mathcal{L}_{Src} + \mathcal{L}_{Tgt} + \mu_{align}\mathcal{L}_{align} \tag{12}$$

This formulation ensures that the model not only learns to impute missing values within each domain but also mitigates domain discrepancies.

5 Experiments

5.1 Experimental Setting

We evaluate our model on three real-world datasets, with details described below and the statistics of datasets are presented in Table 1.

Air Quality [37] dataset contains PM2.5 measurements from Beijing (B) and Tianjin (T). Beijing data is collected from 36 stations, while Tianjin has

Table 1. Dataset Characteristics

Statistics \ Dataset	Air Quality		Hydrology		Electricity	
	Beijing	Tianjin	Discharge	Pooled	ETTh1	ETTh2
Samples	8759	8759	2726	2726	17420	17420
Length	36	36	16	16	48	48
Features	27	27	20	20	7	7
Original Missing Rate	12.36%	20.84%	0%	19.99%	0%	0%

data from 27 stations. For cross-domain setting, 27 stations with the fewest missing values were sampled from Beijing data.

Hydrology dataset records daily river flow and sediment concentration from 20 stations in the United States, collected from United States Geological Survey [17] and Water Quality Portal [28]. It consists of two domains: Discharge (D) and Pooled (P), spanning from March 1, 2017, to September 30, 2022.

Electricity [44] dataset consists of power load and oil temperature data. It includes two years of data (from July 2016 to July 2018) from two distinct regions in China, referred to as ETTh1 and ETTh2.

We follow the dataset splitting strategy used in prior work [45, 31, 19]. For Air Quality dataset, we select Mar., Jun., Sep., and Dec. as the test set, the last 10% of the data in Feb., May, Aug., and Nov. as the validation set, and the remaining data as the training set. For Hydrology and Electricity, we split the training/validation/test set by 70%/10%/20%.

Evaluation Metrics. We evaluate the performance using three metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Continuous Ranked Probability Score (CRPS). MAE and RMSE measure the error between imputed values and ground truth for deterministic methods. CRPS is used to measure how well the imputed probability distributions align with the observed values for methods that produce probability distributions.

Masking Strategy. Since original missing values within datasets lack ground truth, we consider two missing patterns to simulate the missing values for evaluation: (1) *Point missing*, where 10% of the observations is masked, and (2) *Block missing*, where we mask 5 % of the observed data and mask observations ranging from 1 to 4 data points for each feature with 0.15 % probability. For training strategies, we use two masking strategies for self-supervised learning: (1) *Point strategy* randomly selects $r(r \in [0\%, 100\%])$ of observed values; (2) *Block strategy* randomly select a sequence of length $[L/2, L]$ as missing targets with an additional 5 % of observed values randomly selected. Since Air Quality dataset has much original missing data in the training set, we adopt point missing pattern following previous work [31]. For Hydrology and Electricity, we apply both point and block missing patterns following [19].

Baselines and Implementation Details. Baselines for DA. The chosen baselines include various state-of-the-art methods that have been widely adopted in the time series classification and forecasting tasks: CORAL [30],

CDAN [21], DIRT-T [29], AdvSKM [20], CotMix [8]. **Baselines for TSI.** The baselines include RNN-based models M-RNN [39], BRITS [2], GNN-based models GRIN [4], SPIN and SPIN-H [22] and diffusion-based methods CSDI [31], SSSD [1], PriSTI [19], and MTSCI [45]. **Hyperparameters.** We set the batch size to 16 and the number of epochs to 200. The Adam optimizer is used with an initial learning rate of 1e-3, decaying to 1e-4 and 1e-5 at 75% and 90% of the total epochs, respectively. The frequency space mix ratio λ is sampled within [0.0, 1.0], and α in FMixup is empirically set as 0.003. As for the model, we use 4 residual layers, 64 residual channels (C), and 8 attention heads. For methods requiring an adjacency matrix, we use the identity matrix by default. We adopt the quadratic schedule for other noise levels following [31], with a minimum noise level $\beta_1 = 0.0001$ and a maximum noise level $\beta_T = 0.5$.

5.2 Overall Performance

The overall comparisons on three datasets are shown in Table 2. We summarize the observations as follows: (1) CD^2-TSI consistently outperforms baseline models across all datasets. Compared with existing time series imputation methods that rely solely on single-domain data, CD^2-TSI integrates cross-domain knowledge, leading to better imputation performance. Additionally, CD^2-TSI outperforms domain adaptation methods by specifically handling discrepancies in temporal dynamics and data deficiency caused by missing values, which are not adequately addressed by conventional DA approaches. (2) Incorporating domain adaptation techniques increases imputation accuracy compared to training solely on the target domain. However, the extent of this improvement varies across different missing patterns. For example, CDAN performs well on point missing but falls short on block missing scenarios due to its alignment strategy. In contrast, CD^2-TSI consistently improves upon the strongest DA baselines, with an average improvement of 1.92% (RMSE) and 1.34% (MAE), demonstrating its effectiveness in cross-domain alignment for imputation. (3) Our method achieves notable improvements over the best TSI baselines. Specifically, on the Air Quality dataset, CD^2-TSI provides a +4.41% improvement in RMSE and a +5.46% improvement in MAE. On the Hydrology dataset, it results in a +4.37% improvement in RMSE and a +4.61% improvement in MAE, while on the Electricity dataset, they are +4.13% in RMSE and +2.04% in MAE. Among TSI models, MTSCI's performance varies across datasets, with lower accuracy on Air Quality and Hydrology due to high missing rates. These results underscore CD^2-TSI's capacity to adapt effectively to different real-world datasets even under severe missing conditions.

5.3 Sensitivity and Ablation Study

Sensitivity Analysis. We investigate the sensitivity of our method to different missing rates and masking strategies. We conduct experiments on the Air Quality dataset to evaluate performance under various missing rates and on the Hydrology and Electricity datasets to assess different masking strategies.

Table 2. The overall performance comparison. Bold scores are the best performance, and underlined scores are the best time series imputation baseline performance.

Method	Air Quality $B \rightarrow T$		Hydrology $D \rightarrow P$				Electricity $h1 \rightarrow h2$			
	Point		Point		Block		Point		Block	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Coral	14.814	7.374	48.353	13.579	45.484	16.370	0.645	0.381	1.230	0.548
CDAN	14.594	7.203	47.968	13.370	44.725	16.403	0.644	0.381	1.270	0.563
Dirt-T	14.945	7.429	48.506	13.659	45.394	16.227	0.672	0.396	1.194	0.559
AdvSKM	14.786	7.311	48.776	13.289	43.996	16.012	0.643	0.380	1.283	0.567
CoTMix	14.632	7.232	48.470	13.685	44.810	15.954	0.651	0.383	1.235	0.568
M-RNN	46.226	29.497	58.975	19.465	47.000	21.524	7.338	5.386	11.309	4.428
BRITS	40.067	26.355	56.749	19.249	47.207	22.251	6.988	4.428	8.109	4.893
GRIN	26.274	15.773	60.845	23.690	55.254	27.094	3.744	1.587	3.273	1.854
SPIN	27.881	16.914	59.263	22.020	53.530	25.428	6.750	2.856	7.096	3.503
SPIN-H	30.895	18.617	58.915	19.501	49.234	19.893	6.947	2.941	8.001	6.064
CSDI	15.002	7.452	48.542	13.744	45.819	16.470	0.647	0.380	1.320	0.592
SSSD	15.536	8.086	49.132	14.668	47.079	18.282	0.787	0.501	1.307	0.677
PriSTI	15.546	7.686	48.760	14.064	47.868	18.032	0.683	0.407	1.329	0.615
MTSCI	16.252	8.793	48.941	14.477	48.094	18.523	0.715	0.483	1.240	0.562
CD^2-TSI	**14.339**	**7.045**	**46.852**	**13.182**	**43.407**	**15.626**	**0.635**	**0.378**	**1.161**	**0.542**

For the Air Quality dataset, we randomly select 10/20/30/40/50% of the observed values as ground truth in the test data. Fig. 2 (a) shows that our method consistently performs well across these rates. As missing data increases, imputation accuracy typically declines due to reduced availability of observed conditional information. However, our approach's use of frequency mixup interpolation and cross-domain alignment helps maintain high performance.

For the Hydrology and Electricity datasets, we use two settings: Point $\rightarrow$ Block (Point missing pattern in training set, Block missing pattern in testing set) and Block $\rightarrow$ Point (Block missing pattern in training set, Point missing pattern in testing set). Fig. 2 (b-c) shows that CD^2-TSI achieves relatively better performance with various missing patterns in the training and testing sets.

Ablation Study. We evaluate the impact of key components in CD^2-TSI on imputation performance. **(1) w/o FMixup**: Frequency mixup is excluded, and zero filling is used to construct the missing targets. **(2) w/ L.I.**: FMixup is replaced with linear interpolation to construct the missing targets. **(3) w/o CDCA**: Consistency alignment loss L_{align} is removed. Since Electricity dataset does not contain original missing values, no interpolation is required for original missing areas, and we only evaluate w/o CDCA on this dataset.

Table 3 presents the results of our ablation study. Removing frequency mixup interpolation (w/o FMixup) significantly degrades performance across all datasets and missing patterns, confirming that frequency mixup provides informative priors that enhance imputation accuracy. When FMixup is replaced with lin-

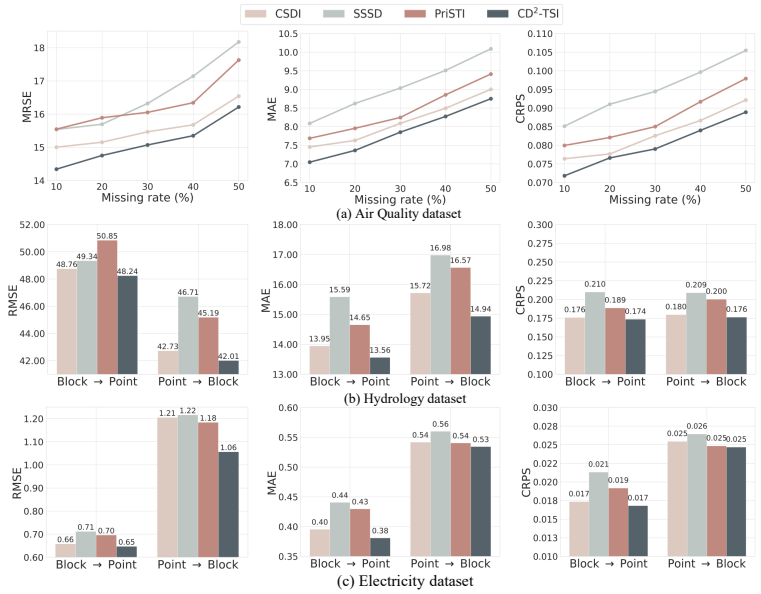

Fig. 2. Sensitivity Analysis for Air Quality, Hydrology, Electricity Datasets

Table 3. Ablation Study of CD^2-TSI.

Method	Air Quality $B \rightarrow T$		Hydrology $D \rightarrow P$				Electricity $h1 \rightarrow h2$			
	Point		Point		Block		Point		Block	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
w/o FMixup	14.817	7.292	47.662	13.365	44.204	15.956	-	-	-	-
w/ L.I.	14.921	7.357	48.524	13.557	44.261	16.163	-	-	-	-
w/o CDCA	14.782	7.319	47.288	13.390	43.679	15.876	0.641	0.380	1.255	0.554
CD^2-TSI	**14.339**	**7.045**	**46.852**	**13.182**	**43.407**	**15.626**	**0.635**	**0.378**	**1.161**	**0.542**

ear interpolation (w/ L.I.), performance further declines, demonstrating that frequency-domain interpolation captures temporal dependencies more effectively than simple interpolation in the time domain. Removing the cross-domain consistency alignment loss (w/o CDCA) results in performance degradation, particularly in block missing scenarios. For instance, in the Electricity dataset, RMSE increases from 1.161 to 1.255 in block missing pattern. This confirms that cross-domain consistency alignment helps in mitigating temporal discrepancies across domains. Overall, the findings of the ablation studies underscore the importance of each proposed component in improving cross-domain imputation.

5.4 Hyperparameter and Efficiency Study

Hyperparameter Study. We conduct a hyperparameter study on key parameters in CD^2-TSI to select the optimal settings across three datasets: the frequency space mix ratio λ, the lower and upper thresholds τ_l and τ_h. The results are shown in Fig. 3. The parameter λ controls the extent of frequency mixing between the source and target domains, while the thresholds τ_l and τ_h ensure cross-domain consistency alignment while preserving domain-specific variations. Our study finds that a moderate frequency mixing ratio and properly selected alignment thresholds ensure effective cross-domain time series imputation.

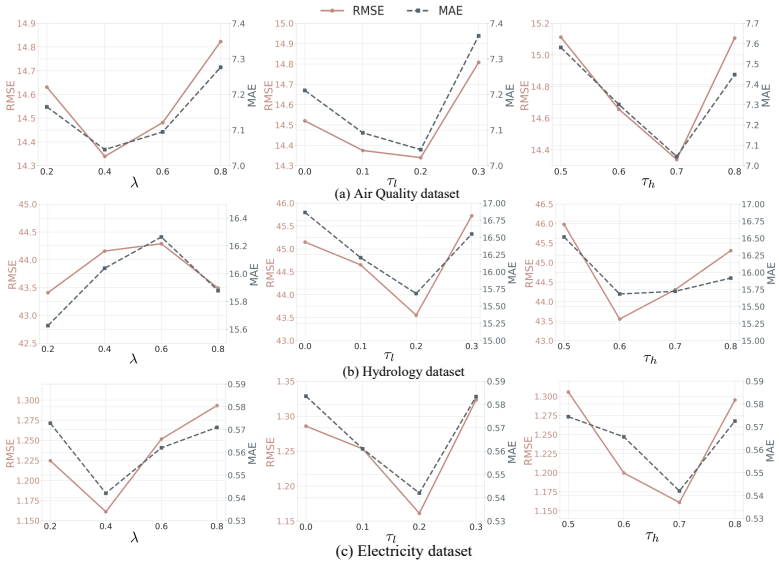

Fig. 3. Hyperparameter study on three key parameters of CD^2-TSI.

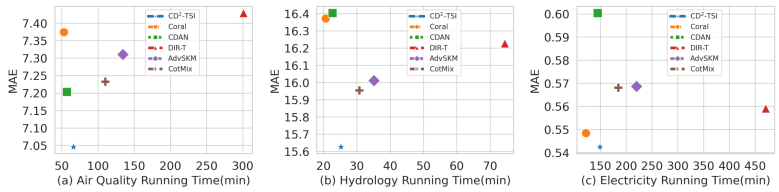

Fig. 4. Efficiency analysis on Air Quality, Hydrology and Electricity datasets.

Efficiency Study. We illustrate the total training time of DA models trained on all three datasets, and the experiments are conducted on an NVIDIA RTX

4090 GPU with 24G memory. Fig. 4 shows the Time-MAE curve, indicating the relationship between time complexity and model performance. Compared with models such as Coral and CDAN, which achieve the least running time, CD^2-TSI achieves better imputation results by leveraging frequency mixup and cross-domain adaptation at the cost of marginally increased training time.

6 Conclusion

In this paper, we introduce CD^2-TSI, a novel approach for cross-domain time series imputation, addressing the limitations of existing methods in handling high missing rates and domain shifts. Our approach effectively leverages cross-domain information through a diffusion-based framework while preserving domain-specific temporal dependencies. The proposed frequency mixup interpolation and selective consistency alignment strategies contribute to improved adaptation and imputation accuracy. CD^2-TSI has demonstrated superior performance on three real-world datasets through comprehensive experiments. Future work will explore more challenging real-world conditions with extreme missing rates and complex domain shifts.

Acknowledgments. This work was supported in part by USACE under Grant No. GR40695, "Designing nature to enhance resilience of built infrastructure in western US landscapes", and by the National Science Foundation under Grant No. 2311716, "CausalBench: A Cyberinfrastructure for Causal-Learning Benchmarking for Efficacy, Reproducibility, and Scientific Collaboration".

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alcaraz, J.M.L., Strodthoff, N.: Diffusion-based time series imputation and forecasting with structured state space models. arXiv preprint arXiv:2208.09399 (2022)
2. Cao, W., Wang, D., Li, J., Zhou, H., Li, L., Li, Y.: Brits: Bidirectional recurrent imputation for time series. Advances in neural information processing systems **31** (2018)
3. Che, Z., Purushotham, S., Cho, K., et al.: Recurrent neural networks for multivariate time series with missing values. Scientific reports (2018)
4. Cini, A., Marisca, I., Alippi, C.: Filling the g_ap_s: Multivariate time series imputation by graph neural networks. arXiv preprint arXiv:2108.00298 (2021)
5. Ding, K., Shu, K., Shan, X., Li, J., Liu, H.: Cross-domain graph anomaly detection. IEEE Transactions on Neural Networks and Learning Systems **33**(6), (2021)
6. Du, W., Côté, D., Liu, Y.: Saits: Self-attention-based imputation for time series. Expert Systems with Applications (2023)
7. Du, W., Wang, J., Qian, L., et al.: Tsi-bench: Benchmarking time series imputation. arXiv preprint arXiv:2406.12747 (2024)
8. Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C.K., Li, X.: Contrastive domain adaptation for time-series via temporal mixup. IEEE Transactions on Artificial Intelligence **5**(3), 1185–1194 (2023)

9. Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C.K., Li, X.: Label-efficient time series representation learning: A review. IEEE Transactions on Artificial Intelligence (2024)
10. Fortuin, V., Baranchuk, D., Rätsch, G., Mandt, S.: Gp-vae: Deep probabilistic time series imputation. In: International conference on artificial intelligence and statistics. pp. 1651–1661. PMLR (2020)
11. Gao, H., Shen, W., Qiu, X., et al.: Diffimp: Efficient diffusion model for probabilistic time series imputation with bidirectional mamba backbone. arXiv preprint arXiv:2410.13338 (2024)
12. He, H., Queen, O., Koker, T., Cuevas, C., Tsiligkaridis, T., Zitnik, M.: Domain adaptation for time series under feature and label shifts. In: International conference on machine learning. pp. 12746–12774. PMLR (2023)
13. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
14. Jin, X., Park, Y., Maddix, D., Wang, H., Wang, Y.: Domain adaptation for time series forecasting via attention sharing. In: International Conference on Machine Learning. pp. 10280–10297. PMLR (2022)
15. Kim, S., Kim, H., Yun, E., Lee, H., Lee, J., Lee, J.: Probabilistic imputation for time-series classification with missing data. In: International Conference on Machine Learning. pp. 16654–16667. PMLR (2023)
16. Kong, Y., Yang, Y., Hwang, Y., Du, W., Zohren, S., Wang, Z., Jin, M. and Wen, Q.: Time-MQA: Time Series Multi-Task Question Answering with Context Enhancement. arXiv preprint arXiv:2503.01875 (2025)
17. Konrad, C.P., Anderson, S.W., Restivo, D.E., et al.: Network analysis of usgs streamflow gages (2022)
18. Little, R.J., Rubin, D.B.: Statistical analysis with missing data. John Wiley & Sons (2019)
19. Liu, M., Huang, H., Feng, H., Sun, L., Du, B., Fu, Y.: Pristi: A conditional diffusion framework for spatiotemporal imputation. In: 2023 IEEE 39th International Conference on Data Engineering (ICDE). pp. 1927–1939. IEEE (2023)
20. Liu, Q., Xue, H.: Adversarial spectral kernel matching for unsupervised time series domain adaptation. In: IJCAI (2021)
21. Long, M., Cao, Z., Wang, J., Jordan, M.I.: Conditional adversarial domain adaptation. Advances in neural information processing systems **31** (2018)
22. Marisca, I., Cini, A., Alippi, C.: Learning to reconstruct missing data from spatiotemporal graphs with sparse observations. Advances in neural information processing systems **35**, 32069–32082 (2022)
23. Miao, X., Wu, Y., Wang, J., Gao, Y., Mao, X., Yin, J.: Generative semi-supervised learning for multivariate time series imputation. In: Proceedings of the AAAI conference on artificial intelligence. pp. 8983–8991 (2021)
24. Miyaguchi, K., Katsuki, T., Koseki, A., Iwamori, T.: Variational inference for discriminative learning with generative modeling of feature incompletion. In: International Conference on Learning Representations (2022)
25. Nussbaumer, H.J., Nussbaumer, H.J.: The fast Fourier transform. Springer (1982)
26. Ragab, M., Eldele, E., Chen, Z., Wu, M., Kwoh, C.K., Li, X.: Self-supervised autoregressive domain adaptation for time series data. IEEE Transactions on Neural Networks and Learning Systems **35**(1), 1341–1351 (2022)
27. Ragab, M., Eldele, E., Tan, W.L., Foo, C.S., Chen, Z., Wu, M., Kwoh, C.K., Li, X.: Adatime: A benchmarking suite for domain adaptation on time series data. ACM Transactions on Knowledge Discovery from Data **17**(8), 1–18 (2023)

28. Read, E.K., Carr, L., De Cicco, L., et al.: Water quality data for national-scale aquatic research: The water quality portal. Water Resources Research (2017)
29. Shu, R., Bui, H.H., Narui, H., Ermon, S.: A dirt-t approach to unsupervised domain adaptation. arXiv preprint arXiv:1802.08735 (2018)
30. Sun, B., Saenko, K.: Deep coral: Correlation alignment for deep domain adaptation. In: Computer vision–ECCV 2016 workshops (2016)
31. Tashiro, Y., Song, J., Song, Y., Ermon, S.: Csdi: Conditional score-based diffusion models for probabilistic time series imputation. Advances in neural information processing systems **34**, 24804–24816 (2021)
32. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
33. Wilson, G., Doppa, J.R., Cook, D.J.: Multi-source deep domain adaptation with weak supervision for time-series sensor data. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. (2020)
34. Wu, H., Hu, T., Liu, Y., et al.: Timesnet: Temporal 2d-variation modeling for general time series analysis. arXiv preprint arXiv:2210.02186 (2022)
35. Xu, Q., Zhang, R., Zhang, Y., Wang, Y., Tian, Q.: A fourier-based framework for domain generalization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14383–14392 (2021)
36. Yang, X., Sun, Y., Chen, X., et al.: Frequency-aware generative models for multivariate time series imputation. Advances in Neural Information Processing Systems **37**, 52595–52623 (2024)
37. Yi, X., Zheng, Y., Zhang, J., Li, T.: St-mvl: filling missing values in geo-sensory time series data. In: Proceedings of the 25th international joint conference on artificial intelligence (2016)
38. Yoon, J., Jordon, J., Schaar, M.: Gain: Missing data imputation using generative adversarial nets. In: International conference on machine learning. (2018)
39. Yoon, J., Zame, W.R., Van Der Schaar, M.: Estimating missing data in temporal data streams using multi-directional recurrent neural networks. IEEE Transactions on Biomedical Engineering **66**(5), 1477–1490 (2018)
40. Yue, Z., Wang, Y., Duan, J., Yang, T., Huang, C., Tong, Y., Xu, B.: Ts2vec: Towards universal representation of time series. In: Proceedings of the AAAI conference on artificial intelligence. pp. 8980–8987 (2022)
41. Zhang, K., Liu, S., Wang, S., Shi, W., Chen, C., Li, P., Li, S., Li, J., Ding, K.: A survey of deep graph learning under distribution shifts: from graph out-of-distribution generalization to adaptation. arXiv preprint arXiv:2410.19265 (2024)
42. Zhang, K., Wang, Y., Li, X., Tang, R., Zhang, R.: Incmsr: An incremental learning approach for multi-scenario recommendation. In: Proceedings of the 17th ACM International Conference on Web Search and Data Mining. pp. 939–948 (2024)
43. Zhang, X., Zhao, Z., Tsiligkaridis, T., Zitnik, M.: Self-supervised contrastive pre-training for time series via time-frequency consistency. Advances in neural information processing systems **35**, 3988–4003 (2022)
44. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond efficient transformer for long sequence time-series forecasting. In: Proceedings of the AAAI conference on artificial intelligence. pp. 11106–11115 (2021)
45. Zhou, J., Li, J., Zheng, G., Wang, X., Zhou, C.: Mtsci: A conditional diffusion model for multivariate time series consistent imputation. In: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. pp. 3474–3483 (2024)