

# Filling the Missings: Spatiotemporal Data Imputation by Conditional Diffusion

Wenying He<sup>1</sup>, Jieliang Huang<sup>1</sup>, Junhua Gu<sup>1,\*</sup>, Ji Zhang<sup>3</sup> and Yude Bai<sup>2,\*</sup>

<sup>1</sup>School of Artificial Intelligence, Hebei University of Technology, Tianjin, China

<sup>2</sup>School of Software, Tiangong University, Tianjin, China

<sup>3</sup>University of Southern Queensland, Queensland, Australia

{hwyang1234, jhgu}@hebut.edu.cn, 202332805027@stu.hebut.edu.cn, Ji.Zhang@unisq.edu.au, baiyude@tiangong.edu.cn

## Abstract

Missing data in spatiotemporal systems presents a significant challenge for modern applications, ranging from environmental monitoring to urban traffic management. The integrity of spatiotemporal data often deteriorates due to hardware malfunctions and software failures in real-world deployments. Current approaches based on machine learning and deep learning struggle to model the intricate interdependencies between spatial and temporal dimensions effectively and, more importantly, suffer from cumulative errors during the data imputation process, which propagate and amplify through iterations. To address these limitations, we propose CoFILL, a novel Conditional Diffusion Model for spatiotemporal data imputation. CoFILL builds on the inherent advantages of diffusion models to generate high-quality imputations without relying on potentially error-prone prior estimates. It incorporates an innovative dual-stream architecture that processes temporal and frequency domain features in parallel. By fusing these complementary features, CoFILL captures both rapid fluctuations and underlying patterns in the data, which enables more robust imputation. The extensive experiments reveal that CoFILL's noise prediction network successfully transforms random noise into meaningful values that align with the true data distribution. The results also show that CoFILL outperforms state-of-the-art methods in imputation accuracy. The source code is publicly available at <https://github.com/joyHJL/CoFILL>.

## 1 Introduction

Spatiotemporal data integrates temporal dynamics with spatial relationships to represent complex real-world phenomena. This data type has proven valuable across diverse domains, from environmental science to urban systems [Yi *et al.*, 2016; Li *et al.*, 2017; Wu *et al.*, 2019]. Despite its importance, the collection of spatiotemporal data faces significant challenges in real-world deployments. Sensor malfunc-

tion, system instability, and environmental interference often lead to missing or corrupted measurements. These data quality issues compromise the reliability of information extracted from spatiotemporal systems and impair the performance of downstream applications [Lian *et al.*, 2023; Tan *et al.*, 2018; He *et al.*, 2025a].

Various methods have been proposed to handle missing data in spatiotemporal datasets. Traditional approaches rely on statistical and data analytic techniques, and make assumptions about the underlying probability distributions and use techniques such as linear interpolation [Challu *et al.*, 2023], expectation-maximization (EM) algorithm [Cong and Li, 2024; Bao *et al.*, 2025], or k-nearest neighbors (KNN) [Xu *et al.*, 2024] to estimate missing data. Their effectiveness is constrained by their inherent mathematical foundations, which rely on simplified models of relationships between variables. When applied to real-world data, where relationships between spatial and temporal components often follow complex, nonlinear patterns, these methods fail to reconstruct the true data distribution accurately.

Deep learning methods have also been applied for spatiotemporal data imputation to better process the inherent complexity in spatiotemporal relationships. Recurrent neural networks (RNN) [Shi *et al.*, 2024] model temporal relationships through hidden state propagation, while graph neural networks (GNN) [Agafonov, 2020; Jin *et al.*, 2023] learn spatial dependencies through node-level representations. These methods, however, face fundamental limitations. RNN [Chen *et al.*, 2022a] operate through an autoregressive mechanism, which causes early prediction errors to propagate through subsequent time steps. GNN [Cini *et al.*, 2021] depend on features from neighboring nodes, allowing errors to disseminate across the graph structure. The recursive nature of their processing leads to error accumulation over time. Their reliance on historical data also constrains their ability to capture the full range of possible values during imputation.

Alternatively, data imputation can be treated as a generative task [Lin *et al.*, 2024]. Recent work has shown that diffusion probabilistic models (DDPM) [Kingma, 2013] and generative adversarial networks (GAN) [Goodfellow *et al.*, 2020]. However, current DDPM-based methods face two key challenges. First, they fail to model both temporal and spatial aspects simultaneously [Wu *et al.*, 2019] - for example, Tashiro *et al.* [Tashiro *et al.*, 2021] consider only tem-

\*Corresponding author: Junhua Gu, Yude Bai.

poral features, ignoring crucial spatial relationships. Second, they rely on problematic pre-imputation approaches - Liu *et al.* [Liu *et al.*, 2023] use linear interpolation that creates artificial fluctuations, causing generated data to deviate from actual patterns. Our work addresses these limitations by developing a model that integrates both spatial and temporal information while avoiding error-prone pre-imputation steps.

In this paper, we introduce CoFILL (Conditional Diffusion Model based on Temporal-Frequency Spatiotemporal Imputation), a novel approach that addresses the limitations of current methods. Our model enhances imputation by utilizing temporal measurements from sensors and geographical connections between monitoring locations. The temporal data captures changing patterns over time, while spatial relationships reveal location interdependencies. The architecture features modules for temporal and frequency domain feature extraction to capture rapid changes and persistent patterns. A cross-attention mechanism combines these features into rich conditional information. This information guides the noise prediction network to transform random noise into meaningful data values through iterative denoising, which enables robust imputation by learning the underlying data distribution rather than direct estimation.

Our contributions are summarized as follows:

- We propose a spatiotemporal imputation framework, CoFILL, based on diffusion models. It applies the non-recursive property of diffusion models to effectively reduce error accumulation during the imputation process.
- We develop a dual-stream feature processing architecture to model complete spatiotemporal dynamics. We extract temporal and frequency information through separate pathways, then integrate them by cross-attention to capture both rapid fluctuations and underlying patterns.
- We conduct extensive experiments on three real-world datasets to validate our approach. The results demonstrate that CoFILL outperforms current state-of-the-art methods in spatiotemporal data imputation tasks.

## 2 Related Works

### 2.1 Spatiotemporal Data Imputation

Methods for spatiotemporal data imputation mainly fall into three main categories: statistical approaches, machine learning techniques, and deep learning models [He *et al.*, 2024]. Statistical methods, such as mean and median imputation [Amiri and Jensen, 2016], assume the data follows specific probability distributions, which is too restrictive for real-world data. Machine learning approaches provide more flexibility like random forest (RF) [Li *et al.*, 2014] and matrix factorization [Wang *et al.*, 2018]. Other machine learning methods such as Bayesian Gaussian process tensor completion (BGCP) [Chen *et al.*, 2019b] and time-regularized tensor factorization (TRTF) [Baggag *et al.*, 2019] show promise in processing complex data. But they still struggle with nonlinear relationships and scalability to large datasets. Deep learning methods, such as recurrent neural networks (RNN) [Vaswani *et al.*, 2017] and their variants like GRU-D [Shahidi *et al.*, 2024], excel at learning temporal patterns

but accumulate errors over prediction sequences. Convolutional neural networks, such as TimesNet [Wu *et al.*, 2022], only analyze time series in Euclidean space using Fourier transforms. Graph neural networks [Marisca *et al.*, 2022; Chen *et al.*, 2022b] improve imputation accuracy by modeling spatial relationships, but their reliance on fixed graph structures reduces their adaptability to new scenarios. These existing approaches cannot fully capture the nonlinear dependencies and complex patterns inherent in large-scale spatiotemporal data.

### 2.2 Diffusion Models for Imputation

Compared with traditional methods and deep learning methods, the generative data imputation method by diffusion models generates high-quality data. The conditional score-based diffusion model (CSDI) [Tashiro *et al.*, 2021] uses a subset of observed data as conditions to generate residual values. Meanwhile, PriSTI [Liu *et al.*, 2023] introduces spatiotemporal relationships and attention mechanisms to achieve better imputation results. GP-VAE [Fortuin *et al.*, 2020] models temporal patterns using Gaussian processes in latent space, while E2GAN [Li *et al.*, 2025] offers a simplified approach to generate plausible values for time series data. However, current diffusion models struggle to effectively extract and use conditional information - the known patterns and relationships in the existing data that guide the generation of missing data. This limitation impacts imputation accuracy because of the generative values that fail to maintain the underlying temporal and spatial dependencies of the data.

### 2.3 Temporal and Frequency Domain Feature

Feature extraction plays a vital role in spatiotemporal data imputation. Temporal feature extraction through recurrent structures like GRU-D [Shahidi *et al.*, 2024] captures immediate state changes but cannot handle variations at different time scales. Meanwhile, frequency domain approaches via Fourier transforms, such as TimesNet [Wu *et al.*, 2022], identify periodic and cyclical patterns but miss important transient changes. GNN-based methods [Marisca *et al.*, 2022] focus on modeling spatial relationships but overlook these temporal dynamics. We hope to address these issues by integrating temporal and frequency domain features through a cross-attention mechanism. The temporal features capture immediate state transitions, while frequency features reveal long-term trends.

## 3 Problem Formulation

Consider a spatiotemporal system with  $N$  observation nodes (such as air monitoring stations or traffic sensors) measured over  $L$  time steps. The measurements form a sequence  $X_{1:L} = X_1, X_2, \dots, X_L \in \mathbb{R}^{N \times L}$ , where  $X_l \in \mathbb{R}^N$  represents the values observed at time step  $l$ . The spatial relationships between nodes are represented by an adjacency matrix  $A \in \mathbb{R}^{N \times N}$ , where  $A_{i,j} = 1$  indicates a connection between nodes  $i$  and  $j$ , and  $A_{i,j} = 0$  indicates no connection. In real-world deployments, some observations may be missing due to sensor failures or system errors. We use a binary mask  $M \in \{0, 1\}^{N \times L}$  to indicate the presence or absence of

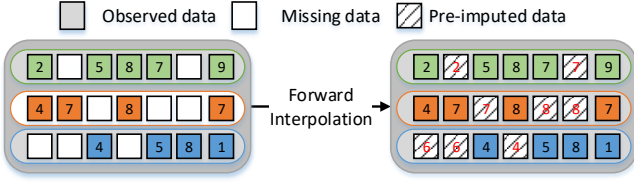


Figure 1. Pre-imputing by forward interpolation.

measurements, where  $m_{i,l} = 1$  denotes a valid observation at node  $i$  and time  $l$ , while  $m_{i,l} = 0$  indicates missing data. The spatiotemporal imputation problem aims to reconstruct missing data in  $X$  given the observed measurements and the adjacency matrix  $A$ . The goal is to develop an imputation method that preserves both the temporal evolution of measurements and their spatial relationships across nodes.

## 4 Methodology

### 4.1 Preprocessing of CoFILL

CoFILL performs a two-step preprocessing on the input series to ensure high-quality inputs for training. We first mask some of the observed values in the original data based on a random masking strategy. This not only simulates missing patterns but also enhances the model's generalization to diverse scenarios [Tashiro *et al.*, 2021]. Two sets of initial imputed data are then generated from the masked data. The first set uses forward interpolation, as shown in Figure 1. It fills in missing data by using the known values from the previous time step, thus preserving the continuity and periodicity of the spatiotemporal data. This also reduces information loss due to over-smoothing. The second set fills the missing data with Gaussian noise. We introduce noise consistent with the original data distribution to enhance data diversity and introduce appropriate randomness for the model [Fortuin *et al.*, 2020; He *et al.*, 2025b]. These two imputed data sets are adopted as inputs for subsequent modules, which further supports CoFILL in learning spatiotemporal dependencies and missing data distributions.

### 4.2 Conditional Information Module

We extract features from the imputed data sets of preprocessing to provide effective conditional information for the noise prediction network. These features, by identifying the spatiotemporal dependencies from both the observed and pre-imputed data, guide the network to predict noise more accurately. The process works as follows. Initially, we map the imputed data  $X_1$  to a latent space  $H_{in} = \text{Conv}(X_1) \in \mathbb{R}^{d \times N \times L}$ , where  $\text{Conv}(\cdot)$  refers to a  $1 \times 1$  convolution operation, and  $d$  denotes the channel size. Next, we extract features from both the temporal and frequency domains, and the combined feature is obtained through cross-attention.

In the temporal domain, data reflects instantaneous changes, which is crucial for short-term prediction. The temporal feature extraction module separately obtains both temporal and spatial dependencies. We map  $X_1$  to the latent space  $H_{in}$  and pass it through a Temporal Convolutional Network (TCN) to learn temporal dependencies. The TCN em-

ployes padded 1-dimensional gated causal convolutions, ensuring that the output and input maintain the same shape. The convolution kernel  $\Gamma_\tau \in \mathbb{R}^{K \times C_{in} \times C_{out}}$  generates outputs  $P, Q \in \mathbb{R}^{C_{out} \times N \times L}$  of the same shape. The definition and specific formulas of the TCN structure can be, respectively, expressed as follows:

$$\bar{H}_{in} = \Gamma_\tau(H_{in}) = P \odot \sigma(Q) \in \mathbb{R}^{C_{out} \times N \times L} \quad (1)$$

$$\begin{aligned} \tilde{H}_{in} = & \text{Dropout} \left( \text{Chomp} \left( \text{Conv2D}_{(k \times k)}^{(dilation)}(\bar{H}_{in}) \right) \right) \\ & + H_{in} \end{aligned} \quad (2)$$

where  $\odot$  represents the Hadamard product, and  $\sigma$  is the sigmoid activation function, introducing nonlinearity and acting as a gating mechanism. We denote the output of the TCN as  $\bar{H}_{in} \in \mathbb{R}^{d \times N \times L}$ , the convolution kernel size is  $k$ , the dilation size is 1, which indicates short-term feature extraction, and the convolution is expanded along the time axis. We use a Graph Convolutional Network (GCN) to learn spatial dependencies. The definition and specific formulas of the GCN structure can be, respectively, expressed as follows:

$$\tilde{H}_{in} = \Gamma_G(\bar{H}_{in}) = \sigma(\varphi(A_{gcn}, \bar{H}_{in})W_i) \quad (3)$$

$$\begin{aligned} \tilde{H}_{in} = & \text{ReLU} \left( \text{Conv2D} \left( \text{Concat} \left( \bar{H}_{in}, \right. \right. \right. \\ & \left. \left. \left. \{(\bar{H}_{in} A_{gcn})^{(o)}\} \right) \right) \right) \end{aligned} \quad (4)$$

where  $W_i \in \mathbb{R}^{C_{in}^g \times C_{out}^g}$  is a learnable parameter,  $\sigma$  is the sigmoid activation function, and  $\varphi(\cdot)$  represents the aggregation function that determines how neighboring node features are aggregated into the target node. The graph-normalized adjacency matrix is  $A_{gcn} = D^{-1/2}(A + I)D^{-1/2} \in \mathbb{R}^{N \times N}$ , where  $I$  is the identity matrix, and  $D$  is the diagonal matrix with  $D_{ii} = \sum_j \partial(A + I)_{ij}$ .  $A_{gcn}$  is the static adjacency matrix,  $o$  is the convolution order of the graph, and  $\text{Conv2D}$  indicates that convolution is used for channel fusion. We denote the output of both TCN and GCN as  $\tilde{H}_{in} \in \mathbb{R}^{d \times N \times L}$ . By processing dynamic characteristics in the temporal dimension with the TCN, we extract the temporal changes in the data. At the same time, we adopt the GCN to learn the relationships between nodes in the spatial dimension. This combination effectively identifies the complex changes in spatiotemporal data over time.

In the frequency domain, data reveals periodicity and trends, which are essential for long-term forecasting and recognizing periodic patterns. Many time series models fail to account for the frequency impact, which leads to an inability to extract the intrinsic features of the series. To this end, we apply the Discrete Cosine Transform (DCT) to extract frequency domain features  $\hat{H}_{in}$ . Unlike Fourier transforms, DCT avoids the Gibbs phenomenon. It also provides superior energy compaction compared to Wavelet transforms [Jiang *et al.*, 2023]. Its specific formula is as follows:

$$\hat{H}_{in} = \sum_{t=0}^{T-1} H_{in} \cdot \cos \left[ \frac{\pi}{T} \left( t + \frac{1}{2} \right) m \right], m = 0, 1, \dots, T-1 \quad (5)$$

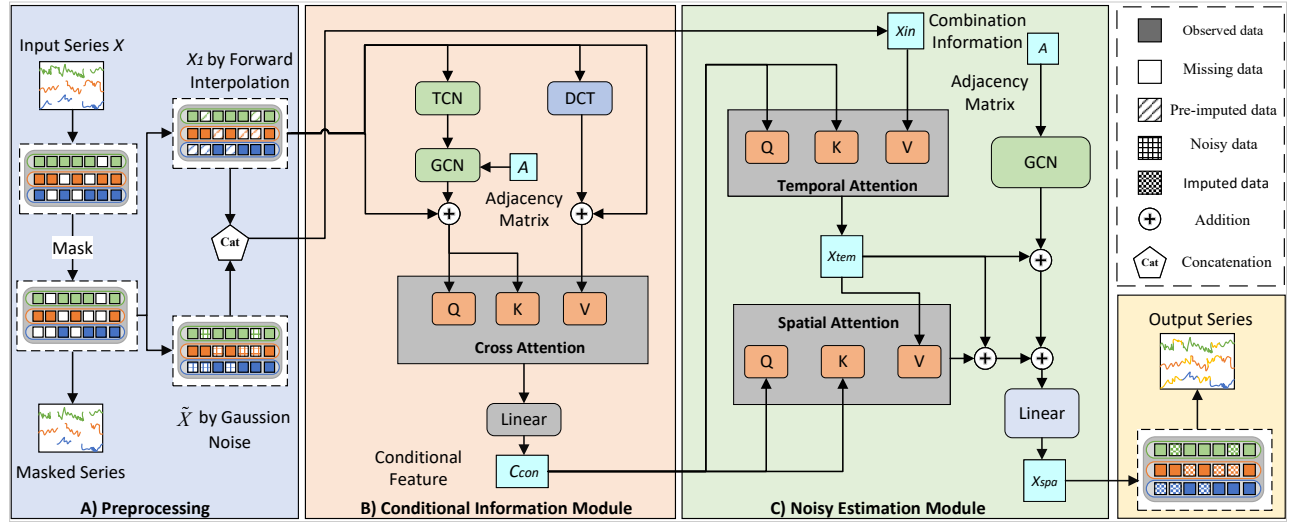


Figure 2. The framework of CoFILL.

where  $H_{in} \in \mathbb{R}^{d \times N \times L}$ ,  $\hat{H}_{in} \in \mathbb{R}^{d \times N \times L}$ ,  $T$  represents the time step. As  $m \rightarrow 0$ , the corresponding DCT component captures low-frequency signals, reflecting long-term stable trends. Conversely, as  $m \rightarrow T$ , it captures high-frequency components, indicating periodic structures. Together, the different frequency components in DCT represent both the long-term trends and periodic patterns within the sequence.

As shown in Figure 2 B), we apply cross-attention to combine the temporal-domain feature  $\hat{H}_{in}$  and the frequency-domain feature  $\hat{H}_{in}$  into a mixed feature representation. Temporal and frequency-domain features offer different perspectives. Temporal features reveal short-term behavior, while frequency-domain features identify long-term trends. The fusion of both perspectives enables the system to process multi-scale data within a unified framework. The temporal feature  $Q$  retains time sequence, local changes, and spatial features. It encodes immediate information along the time axis. The frequency-domain feature  $KV$  focuses on global periodicity and trends, which reduces noise interference while preserving key signal characteristics. The combination of both captures both local temporal information and global periodic trends. The time-frequency domain attention weights and the final conditioning information  $C_{con}$  are defined as:

$$\begin{aligned} Q_{STF} &= \hat{H}_{in} \cdot W_{STF}^Q, \\ K_{STF} &= \hat{H}_{in} \cdot W_{STF}^K, \\ V_{STF} &= \hat{H}_{in} \cdot W_{STF}^V \end{aligned} \quad (6)$$

$$\begin{aligned} C_{con} &= \text{Attn}_{stf}(Q_{STF}, K_{STF}, V_{STF}) \\ &= \text{softmax}\left(\frac{Q_{STF} K_{STF}^T}{\sqrt{d}}\right) \cdot V_{STF} \end{aligned} \quad (7)$$

### 4.3 Noisy Estimation Module

The noise prediction network  $\varepsilon$  in this module is employed during the reverse denoising process to progressively sam-

ple noise, as illustrated in Figure 2 C). It uses conditional information  $C_{con}$  to guide the extraction of spatiotemporal dependencies in the noise data  $X_{in}$ . Additionally, it leverages an attention mechanism to mitigate the randomness introduced by Gaussian noise injected into the data distribution. The model accepts three inputs, including noise information  $X_{in} = (X_1 \parallel \tilde{X}^t)$ , conditional information  $C_{con}$ , and the adjacency matrix  $A$ . It first generates temporal features  $X_{tem}$  through a temporal attention module  $\Upsilon_T(\cdot)$ . The data is then passed into a spatial attention module  $\Upsilon_S(\cdot)$ , equipped with a GNN, to produce spatial features  $X_{spa}$ . This process is formalized as follows:

$$X_{tem} = \Upsilon_T(X_{in}) = \text{Attn}_{tem}(X_{in}) \quad (8)$$

$$\begin{aligned} X_{spa} &= \Upsilon_S(X_{tem}, A) \\ &= \text{MLP}(\text{Norm}(\text{Attn}_{spa}(X_{tem})) + X_{tem}) \\ &\quad + \text{Norm}(G(X_{tem}, A) + X_{tem}) \end{aligned} \quad (9)$$

$\text{Attn}_{tem}(\cdot)$  and  $\text{Attn}_{spa}(\cdot)$  represent temporal and spatial attention, respectively. The mechanisms are the dot-product multi-head self-attention from the transformer architecture. We also put conditional information into the attention calculations to reduce the noise accumulated during the diffusion steps. Specifically, for  $\text{Attn}_{tem}(\cdot)$ , the dot-product attention  $\text{Attn}_{tem}(Q_T, K_T, V_T) = \text{softmax}((Q_T K_T^T)/\sqrt{d}) \cdot V_T$  is rewritten as  $\text{Attn}_{tem}(A_T, V_T) = A_T \cdot V_T$ , where  $A_T = \text{softmax}((Q_T K_T^T)/\sqrt{d})$  denotes the attention weights. The attention weight  $A_T$  is calculated by the conditional information  $C_{con}$ . We set the inputs  $Q_T$ ,  $K_T$ , and  $V_T$  as follows:

$$Q_T = C_{con} \cdot W_T^Q, \quad K_T = C_{con} \cdot W_T^K, \quad V_T = X_{in} \cdot W_T^V \quad (10)$$

where  $W_T^Q, W_T^K, W_T^V \in \mathbb{R}^{d \times d}$  are learnable projection parameters. Similarly, the spatial attention  $\text{Attn}_{spa}(A_S, V_S)$  computes the attention weight in the same manner:

---

**Algorithm 1** Training of CoFILL
 

---

**Input:** Incomplete observed data  $X$ , the adjacency matrix  $A$ , the number of iterations  $N$ , the number of diffusion steps  $T$ , noise levels sequence  $\bar{\alpha}_t$ .

**Output:** Optimized noise prediction model  $\epsilon_\theta$

```

for  $i = 1$  to  $N$  do
     $X_{\text{mask}} \leftarrow \text{Mask}(X)$ ;
     $X_1 \leftarrow \text{Interpolate}(X_{\text{mask}})$ ;
    Sample  $t \sim \text{Uniform}(\{1, \dots, T\})$ ,  $\epsilon \sim \mathcal{N}(0, I)$ ;
     $\tilde{X}^t = \sqrt{\bar{\alpha}_t} X^0 + \sqrt{1 - \bar{\alpha}_t} \epsilon$ ;
    Updating the gradient  $\nabla_\theta \left\| \epsilon - \epsilon_\theta(\tilde{X}^t, X_1, A, t) \right\|^2$ 
end for
    
```

---

$$Q_S = C_{\text{con}} \cdot W_S^Q, \quad K_S = C_{\text{con}} \cdot W_S^K, \quad V_S = X_{\text{tem}} \cdot W_S^V \quad (11)$$

The noise prediction network consists of multiple stacked attention layers that progressively refine spatial representations. Each layer processes features through a gated activation unit, producing spatial features  $X_{\text{spa}}$  that split into two paths. The primary path forms residual connections feeding into subsequent layers, while a secondary path creates skip connections that aggregate features across network depths. The network combines these accumulated features and processes them through two 1-dimensional convolution layers, which contains only the values of the interpolation targets.

#### 4.4 Training and Prediction

We apply a random masking strategy, during the training, to the input observations  $X$  to derive imputation targets  $\tilde{X}$ , while the remaining observations are used to extract pre-imputation features  $X_1$  as conditional information. Similar to CSDI, we implement masking strategies including point strategy, block strategy, and hybrid strategy. As shown in Algorithm 1, the noise prediction model of CoFILL is trained by iterating over the observed data. Specifically, once obtaining the training imputation targets  $\tilde{X}$  and conditional information  $X_1$ , the training objective for spatiotemporal imputation is:

$$L(\theta) = \mathbb{E}_{\tilde{X}^0 \sim q(\tilde{X}^0), \epsilon \sim \mathcal{N}(0, I)} \left\| \epsilon - \epsilon_\theta(\tilde{X}^t, X_1, A, t) \right\|^2 \quad (12)$$

In the imputation, we deploy the trained noise prediction model  $\epsilon_\theta$  and the available observation masks  $M$  to perform imputation. The imputation targets consist of all missing data in the spatiotemporal data. As shown in Algorithm 2, the trained model will fill in the missing data in the spatiotemporal data by performing reverse diffusion.

## 5 Experiments

### 5.1 Datasets

We use three spatiotemporal datasets for imputation algorithms. The AQI-36 dataset [Zheng *et al.*, 2014] serves as a good benchmark due to its inherent missing data patterns and complex spatiotemporal dependencies. The hourly PM2.5

---

**Algorithm 2** Imputation of CoFILL
 

---

**Input:** A sample of incomplete observed data  $X$ , the adjacency matrix  $A$ , the number of diffusion steps  $T$ , the optimized noise prediction model  $\epsilon_\theta$ .

**Output:** Missing data of the imputation target  $X_{\text{out}}$

```

 $X_1 \leftarrow \text{Interpolate}(X)$ 
Set  $\tilde{X}^t \sim \mathcal{N}(0, I)$ 
for  $t = T$  to 1 do
     $\mu_\theta(\tilde{X}^t, X_1, A, t) \leftarrow \frac{1}{\sqrt{\bar{\alpha}_t}} (\tilde{X}^t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(\tilde{X}^t, X_1, A, t))$ 
     $\tilde{X}^{t-1} \leftarrow \mathcal{N}(\mu_\theta(\tilde{X}^t, X_1, A, t), \sigma_t^2 I)$ 
end for
    
```

---

measurements across Beijing’s 36 monitoring stations over a full year handle both short-term fluctuations and long-term season trends. The traffic datasets, METR-LA and PEMS-BAY [Li *et al.*, 2017], present different scales to test our model’s capabilities. METR-LA’s 207 sensors and PEMS-BAY’s 325 sensors represent large-scale networks with dense spatial correlations. Their five-minute sampling frequency verifies model ability to face rapid temporal variations, while their different geographic contexts - Los Angeles and San Francisco Bay Area - validate results across urban traffic.

### 5.2 Baselines for Comparison

**Traditional Statistical Baselines.** The fundamental statistical approaches include Mean, which imputes values using historical node averages, and Daily Average (DA), which uses time-specific averages. We also include K-Nearest Neighbors (KNN) [Bai *et al.*, 2023; Xu *et al.*, 2024], which leverages geographical proximity for imputation, and Linear interpolation for time series reconstruction. For robust sequential data processing, we incorporate Kalman Filter (KF), which estimates missing data through state-space modeling.

**Matrix and Tensor-Based Baselines.** These intermediate approaches bridge classical statistics and modern machine learning by decomposing complex data structures. MICE [White *et al.*, 2011] uses conditional specifications with chained equations for multivariate imputation. Vector Autoregression (VAR) predicts single-step values through temporal dependencies. Time-Regularized Matrix Factorization (TRMF) [Yu *et al.*, 2016] specializes in temporal data patterns, and Bayesian Augmented Tensor Factorization (BATF) [Chen *et al.*, 2019a] integrates spatiotemporal knowledge through tensor decomposition.

**Deep Learning Baselines.** For probabilistic modeling with uncertainty quantification, we employ V-RIN [Mulyadi *et al.*, 2021] and GP-VAE [Fortuin *et al.*, 2020]. To leverage adversarial learning, we use rGAIN [Yoon *et al.*, 2018], which enhances GAIN through bidirectional processing. We further employ BRITS [Cao *et al.*, 2018] for RNN-based temporal modeling and GRIN for graph neural network processing with bidirectional GRU. Our evaluation also encompasses recent diffusion-based methods: CSDI [Tashiro *et al.*, 2021], which captures feature interactions through transformers, and PriSTI [Liu *et al.*, 2023], which implements spatiotemporal attention mechanisms.

### 5.3 Evaluation Metrics

We investigate the performance by three complementary metrics - Mean Absolute Error (MAE), Mean Squared Error (MSE), and Continuous Ranked Probability Score (CRPS) [Berrisch and Ziel, 2023]. MAE and MSE provide measures of imputation accuracy by calculating the distance between predicted values and ground truth. CRPS evaluates the quality of probabilistic predictions by measuring how well the predicted probability distributions align with given values.

### 5.4 Experimental Design

To ensure reproducibility and meaningful evaluation, we establish a methodical data partitioning strategy that respects the temporal characteristics of our datasets. The AQI-36 dataset requires careful temporal consideration due to its seasonal nature. We distribute the data across seasons by selecting March, June, September, and December for testing, which captures seasonal variations throughout the year. The validation set draws from the final 10% of data in February, May, August, and November to maintain seasonal representation. The remaining months form the training set.

To evaluate model performance under different missing data scenarios, we design two types of data corruption schemes. For AQI-36, we implement a simulated failure (SF) pattern that replicates real-world sensor malfunction distributions. For PEMS-BAY and METR-LA traffic datasets, we create controlled missing data scenarios through mask matrices. These scenarios include random point missing (Point), where we mask 25% of observations uniformly at random, and structured block missing (Block), which combines 5% random masking with continuous missing segments. These segments span 1 to 4 hours per sensor and occur with 0.15% probability, simulating extended sensor outages.

Our implementation runs on an NVIDIA RTX 4090 GPU with 24GB VRAM. We optimize the model using Adam with cosine annealing learning rate decay. The learning rate starts at  $10^{-3}$  and decays to  $10^{-5}$ . Table 1 details the complete hyperparameter configuration for each dataset. To establish statistical significance, we conduct five independent runs with different random seeds for each experiment.

	AQI-36	METR-LA	PEMS-BAY
Batch size	16	16	16
Time length $L$	36	24	24
Epochs	200	300	300
Learning rate	0.001	0.001	0.001
Channel size $d$	64	64	64
Noise estimation layer	4	4	4
Min noise level $\beta_1$	0.0001	0.0001	0.0001
Max noise level $\beta_T$	0.2	0.2	0.2
Diffusion steps $T$	100	50	50
Virtual nodes $k$	16	64	64

Table 1. The hyperparameters of CoFILL in each dataset

### 5.5 Comparative Performance Study

Our experimental evaluation demonstrates CoFILL’s consistent superiority across multiple datasets and missing data scenarios. The quantitative results in Table 2 reveal CoFILL’s

strong performance in terms of MAE and MSE metrics. For probabilistic imputation assessment, Table 3 presents the CRPS comparisons. CoFILL achieves the best performance in 12 out of 15 experimental configurations spanning different metrics and missing patterns. The improvement becomes particularly apparent when compared with PriSTI, the state-of-the-art method. On the METR-LA dataset with Block, CoFILL reduces both MAE and MSE by 10.22%. On the PEMS-BAY point-missing pattern, CoFILL performs on par with PriSTI (only  $\leq 1\%$  difference with identical MSE).

The comparative analysis reveals clear performance patterns across different methodological categories. Traditional statistical and machine learning approaches, such as Mean and KNN, show limited effectiveness due to their simplified assumptions about spatiotemporal relationships. The deep learning approaches, represented by GRIN and BRITS, face performance degradation from accumulated errors in their autoregressive processing. While recent diffusion-based models like CSDI and PriSTI demonstrate strong baseline performance, CoFILL still outperforms them with advanced conditional information construction and feature fusion methods.

### 5.6 Ablation Study

We evaluate the impact of key components of CoFILL by comparing the following settings.

- w/o Forward (Forward Interpolation): Pre-imputation information  $X_1$  is removed, and conditional feature extraction is performed directly on the missing data.
- w/o Temporal (Temporal Domain): The temporal module in the conditional feature extraction block is removed, leaving only the frequency domain module.
- w/o Frequency (Frequency Domain): The frequency module in the conditional feature extraction block is removed, leaving only the temporal module.
- w/o Cross (Cross Attention): The cross-attention mechanism between temporal and frequency domains is removed and replaced by direct addition.

Table 4 presents the impact of each setting on the performance. We only analyze AQI-36 and METR-LA due to the computational cost of re-running diffusion-based CoFILL, especially on the larger PEMS-BAY. This aligns with the experimental setup of the baseline PriSTI [Liu *et al.*, 2023]. The removal of forward interpolation produces the most significant performance degradation across datasets. For AQI-36, this modification increases the MAE from 8.7 to 9.15, representing a substantial 0.45 unit deterioration. The temporal domain component emerges as the second most critical element, with its removal causing more severe performance decline than eliminating the frequency domain component from the conditional information module described in Section 4.2. The cross-attention mechanism shows the least impact on model performance, particularly evident in the METR-LA dataset where its removal results in minimal MAE changes.

### 5.7 Hyperparameter Sensitivity Analysis

This experiment analyzes two important parameters in the diffusion model: max noise level ( $\beta_T$ ) and channel size of the



	AQI-36		METR-LA				PEMS-BAY			
	SF		Block		Point		Block		Point	
	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
Mean	53.48±0.00	4578.08±0.00	7.48±0.00	139.54±0.00	7.56±0.00	142.22±0.00	5.46±0.00	87.56±0.00	5.42±0.00	86.59±0.00
DA	50.51±0.00	4416.10±0.00	14.53±0.00	445.08±0.00	14.57±0.00	448.66±0.00	3.30±0.00	43.76±0.00	3.35±0.00	44.50±0.00
KNN	30.21±0.00	2892.31±0.00	7.79±0.00	124.61±0.00	7.88±0.00	129.29±0.00	4.30±0.00	49.90±0.00	4.30±0.00	49.80±0.00
Linear	14.46±0.00	673.92±0.00	3.26±0.00	33.76±0.00	2.43±0.00	14.75±0.00	1.54±0.00	14.14±0.00	0.76±0.00	1.74±0.00
KF	54.09±0.00	4942.26±0.00	16.75±0.00	534.69±0.00	16.66±0.00	529.96±0.00	5.64±0.00	93.19±0.00	5.68±0.00	93.32±0.00
MICE	30.37±0.00	2594.06±7.17	4.22±0.05	51.07±1.25	4.42±0.07	55.07±1.46	2.94±0.02	28.28±0.37	3.09±0.02	31.43±0.41
VAR	15.64±0.09	833.46±13.85	3.11±0.08	28.00±0.76	2.69±0.00	21.10±0.02	2.09±0.10	16.06±0.73	1.30±0.00	6.52±0.01
TRMF	15.46±0.08	1379.05±34.83	2.96±0.00	22.65±0.13	2.86±0.00	20.39±0.02	1.95±0.01	11.21±0.06	1.85±0.00	10.03±0.00
BATF	15.21±0.27	662.87±29.55	3.56±0.01	35.39±0.03	3.58±0.01	36.05±0.02	2.05±0.00	14.48±0.01	2.05±0.00	14.90±0.06
V-RIN	10.00±0.10	838.05±24.74	6.84±0.17	150.08±6.13	3.96±0.08	49.98±1.30	2.49±0.04	36.12±0.66	1.21±0.03	6.08±0.29
GP-VAE	25.71±0.30	2589.53±59.14	6.55±0.09	122.33±2.05	6.57±0.10	127.26±3.97	2.86±0.15	26.80±2.10	3.41±0.23	38.95±4.16
rGAIN	15.37±0.26	641.92±33.89	2.90±0.00	21.67±0.15	2.83±0.01	20.03±0.09	2.18±0.01	13.96±0.20	1.88±0.02	10.37±0.20
BRITS	14.50±0.35	622.36±65.16	2.34±0.01	17.00±0.14	2.34±0.00	16.46±0.05	1.70±0.01	10.50±0.07	1.47±0.00	7.94±0.03
GRIN	12.08±0.47	523.14±57.17	2.03±0.00	13.26±0.05	1.91±0.00	10.41±0.03	1.14±0.01	6.60±0.10	0.67±0.00	1.55±0.01
CSDI	9.51±0.10	352.46±7.50	1.98±0.00	12.62±0.60	1.79±0.00	8.96±0.08	0.86±0.00	4.39±0.02	0.57±0.00	1.12±0.03
PriSTI	9.03±0.07	310.39±7.03	1.86±0.00	10.70±0.02	1.72±0.00	8.24±0.05	0.78±0.00	3.31±0.01	<b>0.55±0.00</b>	<b>1.03±0.00</b>
CoFILL	<b>8.70±0.06</b>	<b>296.52±7.29</b>	<b>1.67±0.00</b>	<b>9.42±0.00</b>	<b>1.62±0.00</b>	<b>8.01±0.05</b>	<b>0.74±0.00</b>	<b>2.99±0.02</b>	0.57±0.00	<b>1.03±0.00</b>
Improvement	3.65%	4.47%	10.22%	11.96%	5.81%	2.79%	5.13%	9.67%	—	—

Table 2. The results of MAE and MSE for spatiotemporal imputation

	AQI-36	METR-LA		PEMS-BAY	
	SF	Block	Point	Block	Point
V-RIN	0.3154	0.1283	0.0781	0.0394	0.0191
GP-VAE	0.3377	0.1118	0.0977	0.0436	0.0568
CSDI	0.1056	0.0260	0.0235	0.0127	0.0067
PriSTI	0.0997	0.0244	0.0227	0.0093	<b>0.0064</b>
CoFILL	<b>0.09598</b>	<b>0.0220</b>	<b>0.0213</b>	<b>0.0091</b>	0.0069
Improvement	3.73%	9.80%	6.17%	2.15%	—

Table 3. The results of CRPS for spatiotemporal imputation

	AQI-36	METR-LA	
	SF	Block	Point
w/o Forward	9.15	1.95	1.83
w/o Temporal	9.00	1.89	1.78
w/o Frequency	8.81	1.71	1.72
w/o Cross	8.82	1.68	1.63
CoFILL	<b>8.70</b>	<b>1.67</b>	<b>1.62</b>

Table 4. The result of MAE for ablation studies

hidden state ( $d$ ) that represent fundamental trade-offs. The former controls the intensity of the noise added during the diffusion process. A high noise level ( $\beta_T$ ) may lead to distorted generated data, while a low noise level ( $\beta_T$ ) may make it difficult to learn the various relationships within the data. The latter determines the model’s ability to represent and handle information at each time step. A larger  $d$  allows the model to learn more features, but it may also lead to higher overhead.

The experimental results are illustrated in Figure 3 (only for AQI-36 and METR-LA due to the same reasons discussed in Section 5.6). For AQI-36, the model maintains stable MAE and MSE values with maximum noise level  $\beta_T$  between 0.2 and 0.4. The METR-LA dataset shows higher sensitivity to  $\beta_T$  variations within this range. The channel size  $d$  produces opposing effects, including degrading performance on AQI-36 (increasing  $d$ ) but improving results on METR-LA. This behavior reflects the different complexities in air quality and traffic data. Furthermore, we set  $d$  to 16 for AQI-36 and 64 for METR-LA. We apply METR-LA’s parameter to PEMS-BAY due to their similar data characteristics, as detailed in

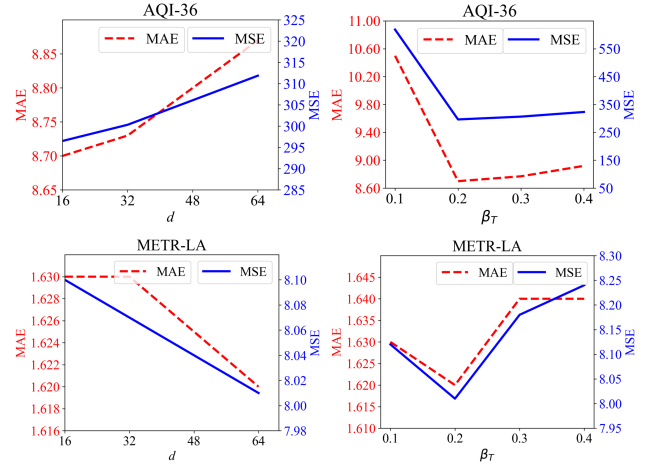

 Figure 3. Analysis of hyperparameters  $\beta_T$  and  $d$ .

Table 1. This enables CoFILL to outperform PriSTI across most scenarios, which obtains better performance even under PEMS-BAY’s Point where the metrics remain close to PriSTI.

## 6 Conclusion

In this paper, we present CoFILL, a novel conditional diffusion framework for spatiotemporal data imputation. CoFILL first overcomes the error accumulation problem by implementing a non-recursive processing structure through our diffusion framework. Second, it resolves the challenge of capturing complex temporal dependencies by introducing a dual-stream architecture that simultaneously processes temporal and frequency domain features. Extensive experiments on real-world datasets demonstrate CoFILL’s superior performance. Our model achieves better imputation accuracy than SOTAs across various missing data scenarios. We will develop adaptive fusion and extend CoFILL to handle multi-resolution spatiotemporal data in the future.

## Acknowledgments

This work is supported by the National Natural Science Foundation of China (62302148), Hebei Natural Science Foundation (F2024202076), and Hebei Yanzhao HuangJintai Talents Program (Postdoctoral Platform) (B2024003002).

## References

- [Agafonov, 2020] Anton Agafonov. Traffic flow prediction using graph convolution neural networks. In *Proceedings of the 10th international conference on information science and technology (ICIST)*, pages 91–95. IEEE, 2020.
- [Amiri and Jensen, 2016] Mehran Amiri and Richard Jensen. Missing data imputation using fuzzy-rough methods. *Neurocomputing*, 205:152–164, 2016.
- [Baggag et al., 2019] Abdelkader Baggag, Sofiane Abbar, Ankit Sharma, Tahar Zanouda, Abdulaziz Al-Homaid, Abhiraj Mohan, and Jaideep Srivastava. Learning spatiotemporal latent factors of traffic via regularized tensor factorization: Imputing missing values and forecasting. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 33(6):2573–2587, 2019.
- [Bai et al., 2023] Yude Bai, Sen Chen, Zhenchang Xing, and Xiaohong Li. Argusdroid: detecting android malware variants by mining permission-api knowledge graph. *Science China Information Sciences (SCIS)*, 66(9):1–19, 2023.
- [Bao et al., 2025] Peijun Bao, Zihao Shao, Wenhan Yang, Boon Poh Ng, and Alex C Kot. E3m: zero-shot spatiotemporal video grounding with expectation-maximization multimodal modulation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 227–243. Springer, 2025.
- [Berrisch and Ziel, 2023] Jonathan Berrisch and Florian Ziel. Crps learning. *Journal of Econometrics*, 237(2):105221, 2023.
- [Cao et al., 2018] Wei Cao, Dong Wang, Jian Li, Hao Zhou, Lei Li, and Yitan Li. Brits: Bidirectional recurrent imputation for time series. *Advances in neural information processing systems (NeurIPS)*, 31, 2018.
- [Challu et al., 2023] Cristian Challu, Kin G Olivares, Boris N Oreshkin, Federico Garza Ramirez, Max Mergenthaler Canseco, and Artur Dubrawski. Nhits: Neural hierarchical interpolation for time series forecasting. In *Proceedings of the AAAI conference on artificial intelligence (AAAI)*, volume 37, pages 6989–6997, 2023.
- [Chen et al., 2019a] Xinyu Chen, Zhaocheng He, Yixian Chen, Yuhuan Lu, and Jiawei Wang. Missing traffic data imputation and pattern discovery with a bayesian augmented tensor factorization model. *Transportation Research Part C: Emerging Technologies*, 104:66–77, 2019.
- [Chen et al., 2019b] Xinyu Chen, Zhaocheng He, and Lijun Sun. A bayesian tensor decomposition approach for spatiotemporal traffic data imputation. *Transportation research part C: emerging technologies*, 98:73–84, 2019.
- [Chen et al., 2022a] Jiawen Chen, Yajun Yang, Chenyang Wang, Heng Zhang, Chao Qiu, and Xiaofei Wang. Multitask offloading strategy optimization based on directed acyclic graphs for edge computing. *IEEE Internet of Things Journal*, 9(12):9367–9378, 2022.
- [Chen et al., 2022b] Jiawen Chen, Yajun Yang, Chenyang Wang, Heng Zhang, Chao Qiu, and Xiaofei Wang. Multitask offloading strategy optimization based on directed acyclic graphs for edge computing. *IEEE Internet of Things Journal*, 9(12):9367–9378, 2022.
- [Cini et al., 2021] Andrea Cini, Ivan Marisca, and Cesare Alippi. Filling the gaps: Multivariate time series imputation by graph neural networks. *arXiv preprint arXiv:2108.00298*, 2021.
- [Cong and Li, 2024] Yulai Cong and Sijia Li. Big learning expectation maximization. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, pages 11669–11677, 2024.
- [Fortuin et al., 2020] Vincent Fortuin, Dmitry Baranchuk, Gunnar Rätsch, and Stephan Mandt. Gp-vae: Deep probabilistic time series imputation. In *Proceedings of the International conference on artificial intelligence and statistics (AISTATS)*, pages 1651–1661. PMLR, 2020.
- [Goodfellow et al., 2020] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.
- [He et al., 2024] Wenying He, Haolu Zhou, Yun Zuo, Yude Bai, and Fei Guo. Muse: A deep learning model based on multi-feature fusion for super-enhancer prediction. *Computational Biology and Chemistry*, 113:108282, 2024.
- [He et al., 2025a] Wenying He, Feiyu Wang, Yude Bai, Neal N. Xiong, Guangquan Xu, and Fei Guo. Pefn: A patches enhancement and hierarchical fusion network for robust vehicle re-identification. *IEEE Internet of Things Journal*, pages 1–1, 2025.
- [He et al., 2025b] Wenying He, Jialu Xu, Yun Zuo, Yude Bai, and Fei Guo. Ensemble: identification of super-enhancers based on ensemble learning. *Briefings in Functional Genomics*, 24:elaf003, 04 2025.
- [Jiang et al., 2023] Maowei Jiang, Pengyu Zeng, Kai Wang, Huan Liu, Wenbo Chen, and Haoran Liu. Fecan: Frequency enhanced channel attention mechanism for time series forecasting. *Advanced Engineering Informatics*, 58:102158, 2023.
- [Jin et al., 2023] Di Jin, Bingdao Feng, Siqi Guo, Xiaobao Wang, Jianguo Wei, and Zhen Wang. Local-global defense against unsupervised adversarial attacks on graphs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(7):8105–8113, Jun. 2023.
- [Kingma, 2013] Diederik P Kingma. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.



- [Li *et al.*, 2014] Yuebiao Li, Zhiheng Li, and Li Li. Missing traffic data: comparison of imputation methods. *IET Intelligent Transport Systems*, 8(1):51–57, 2014.
- [Li *et al.*, 2017] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*, 2017.
- [Li *et al.*, 2025] Lujun Li, Haosen Sun, Shiwen Li, Peijie Dong, Wenhan Luo, Wei Xue, Qifeng Liu, and Yike Guo. Auto-gas: automated proxy discovery for training-free generative architecture search. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 38–55. Springer, 2025.
- [Lian *et al.*, 2023] Zheng Lian, Lan Chen, Licai Sun, Bin Liu, and Jianhua Tao. Gcnet: Graph completion network for incomplete multimodal learning in conversation. *IEEE Transactions on pattern analysis and machine intelligence (TPAMI)*, 45(7):8419–8432, 2023.
- [Lin *et al.*, 2024] Bill Yuchen Lin, Yicheng Fu, Karina Yang, Faeze Brahman, Shiyu Huang, Chandra Bhagavatula, Prithviraj Ammanabrolu, Yejin Choi, and Xiang Ren. Swiftsage: A generative agent with fast and slow thinking for complex interactive tasks. *Advances in Neural Information Processing Systems (NeurIPS)*, 36, 2024.
- [Liu *et al.*, 2023] Mingzhe Liu, Han Huang, Hao Feng, Leilei Sun, Bowen Du, and Yanjie Fu. Pristi: A conditional diffusion framework for spatiotemporal imputation. In *Proceedings of the 39th International Conference on Data Engineering (ICDE)*, pages 1927–1939. IEEE, 2023.
- [Marisca *et al.*, 2022] Ivan Marisca, Andrea Cini, and Cesare Alippi. Learning to reconstruct missing data from spatiotemporal graphs with sparse observations. *Advances in Neural Information Processing Systems (NeurIPS)*, 35:32069–32082, 2022.
- [Mulyadi *et al.*, 2021] Ahmad Wisnu Mulyadi, Eunji Jun, and Heung-Il Suk. Uncertainty-aware variational-recurrent imputation network for clinical time series. *IEEE Transactions on Cybernetics*, 52(9):9684–9694, 2021.
- [Shahidi *et al.*, 2024] Faezehsadat Shahidi, M Ethan MacDonald, Dallas Seitz, Rebecca Barry, and Geoffrey Messier. Enhancing risk prediction in mental health using ensemble hybrid models and administrative healthcare data with irregular intervals. In *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 3673–3678. IEEE, 2024.
- [Shi *et al.*, 2024] Yang Shi, Chenling Ding, Shuai Li, Bin Li, and Xiaobing Sun. New rnn algorithms for different time-variant matrix inequalities solving under discrete-time framework. *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, 2024.
- [Tan *et al.*, 2018] Qiaoyu Tan, Guoxian Yu, Carlotta Domeniconi, Jun Wang, and Zili Zhang. Incomplete multi-view weak-label learning. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2703–2709, 2018.
- [Tashiro *et al.*, 2021] Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. Csd: Conditional score-based diffusion models for probabilistic time series imputation. *Advances in Neural Information Processing Systems (NeurIPS)*, 34:24804–24816, 2021.
- [Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *advances in neural information processing systems (NeurIPS)*, 30(2017), 2017.
- [Wang *et al.*, 2018] Yang Wang, Yong Zhang, Xinglin Piao, Hao Liu, and Ke Zhang. Traffic data reconstruction via adaptive spatial-temporal correlations. *IEEE Transactions on Intelligent Transportation Systems (TITS)*, 20(4):1531–1543, 2018.
- [White *et al.*, 2011] Ian R White, Patrick Royston, and Angela M Wood. Multiple imputation using chained equations: issues and guidance for practice. *Statistics in medicine*, 30(4):377–399, 2011.
- [Wu *et al.*, 2019] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. Graph wavenet for deep spatial-temporal graph modeling. *arXiv preprint arXiv:1906.00121*, 2019.
- [Wu *et al.*, 2022] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*, 2022.
- [Xu *et al.*, 2024] Lihui Xu, Jinfeng Lin, Yuxuan Yang, Zhihao Zhao, Xiaoming Shi, Guanglong Ge, Jin Qian, Cheng Shi, Guohui Li, Simin Wang, et al. Ultrahigh thermal stability and piezoelectricity of lead-free knn-based texture piezoceramics. *Nature Communications*, 15(1):9018, 2024.
- [Yi *et al.*, 2016] Xiuwen Yi, Yu Zheng, Junbo Zhang, and Tianrui Li. St-mvl: Filling missing values in geo-sensory time series data. In *Proceedings of the 25th international joint conference on artificial intelligence (IJCAI)*, 2016.
- [Yoon *et al.*, 2018] Jinsung Yoon, James Jordon, and Michaela Schaar. Gain: Missing data imputation using generative adversarial nets. In *Proceedings of the International conference on machine learning (ICML)*, pages 5689–5698. PMLR, 2018.
- [Yu *et al.*, 2016] Hsiang-Fu Yu, Nikhil Rao, and Inderjit S Dhillon. Temporal regularized matrix factorization for high-dimensional time series prediction. *Advances in neural information processing systems (NeurIPS)*, 29, 2016.
- [Zheng *et al.*, 2014] Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3):1–55, 2014.