# Conditional Implicit Neural Representations via Cross-Attention for Multivariate Time-Series Imputation

Eunho Shin
*Department of Artificial Intelligence*
*Hanbat National University*
Daejeon, Republic of Korea
eunho@edu.hanbat.ac.kr

Janghun Hyeon
*Department of Artificial Intelligence*
*Hanbat National University*
Daejeon, Republic of Korea
jhhyeon@hanbat.ac.kr

Yunho Jeon†
*Department of Artificial Intelligence*
*Hanbat National University*
Daejeon, Republic of Korea
yhjeon@hanbat.ac.kr

***Abstract*—Existing Transformer-based approaches for time-series imputation typically rely on full-sequence reconstruction, which leads to computational inefficiency. To address this, we propose Conditional Implicit Neural Representation via Cross-Attention (CINR-CA), a novel framework designed for high-efficiency imputation. We formulate this task as a Conditional Implicit Neural Representation (INR) problem. By adopting a coordinate-based query mechanism, our model selectively computes outputs only for missing timestamps. This strategy significantly reduces inference costs by avoiding unnecessary computations on observed data. Experiments on the ETTh1 dataset demonstrate that CINR-CA reduces computational complexity (MFLOPs) by approximately 43% (Large settings) compared to the state-of-the-art SAITS model, offering a practical solution for resource-constrained environments where efficiency is important.**

***Index Terms*—Time series imputation, Cross-attention, Implicit Neural Representations, Computational efficiency, Multivariate time series**

## I. Introduction

Multivariate time series data is widely utilized across diverse fields such as transportation, economics, healthcare, and meteorology. While modern deep learning models demonstrate excellent performance when complete observations exist, real-world data often contains significant missing values due to sensor failures, transmission errors, and system instability. Representative examples include medical monitoring devices, environmental sensing systems, and IoT-based infrastructure networks. Missing values degrade data reliability and negatively impact subsequent tasks like prediction, classification, and anomaly detection. Therefore, effective missing value imputation remains a core element for robust time series analysis.

Traditional approaches to handling missing values are categorized into deletion strategies and imputation strategies. The deletion approach removes partially observed records, potentially causing information loss or statistical bias. The imputation approach preserves data by estimating missing values based on observed information using methods such as mean or median replacement, regression analysis, and k-nearest neighbors. However, these methods rely on strong distributional assumptions and fail to capture the complex nonlinear temporal dependencies commonly found in real-world data.

To address these limitations, this study proposes a novel framework, Conditional Implicit Neural Representation via Cross-Attention (CINR-CA), inspired by the coordinate-based paradigm of Implicit Neural Representations (INR). While INR frameworks typically reconstruct signals by mapping continuous coordinates (e.g., spatial locations in images) to signal values, we extend this concept to temporal imputation. The proposed CINR-CA model treats the time index of missing data as a query coordinate. By performing attention mechanisms exclusively on these target positions, the model avoids the redundancy of full-sequence reconstruction inherent in standard methods, thereby significantly reducing inference costs.

The main contributions of this paper are summarized as follows:

- **Novel Architecture for Efficiency:** We propose CINR-CA, a coordinate-based imputation framework that leverages cross-attention to selectively target missing timestamps. Unlike existing state-of-the-art models that require full-sequence reconstruction, our approach explicitly queries only the missing positions.
- **Inference-Focused Optimization:** We demonstrate that our selective querying mechanism significantly reduces computational redundancy specifically during the inference phase, making the model highly suitable for real-time applications where latency is critical.
- **Superior Efficiency-Accuracy Trade-off:** Experiments on the ETTh1 dataset show that CINR-CA reduces computational complexity (MFLOPs) by approximately 43% compared to SAITS. While maintaining highly competitive accuracy (with a marginal gap), it offers a practical solution for resource-constrained environments.

*This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00240379).

†Corresponding author

## II. Related Work

### A. Time Series Imputation

Traditional methods for restoring missing values have been widely used due to their simplicity of implementation and computational efficiency [1]. Mean imputation is the most basic approach, filling missing values with the average of observed values [2]. Linear regression predicts missing values by leveraging their linear relationship with other variables, while k-NN-based methods find samples with similar patterns to estimate missing values [3], [4]. The moving average filter considers temporal proximity, averaging values from surrounding time points for use.

These traditional methods offer the advantages of high computational efficiency and interpretability. However, they struggle to adequately capture the complex nonlinear patterns and interactions between variables in multivariate time series data [5]. Furthermore, most methods heavily rely on the missing-at-random (MAR) assumption, which assumes missing values occur randomly [6]. This makes them vulnerable to bias when dealing with systematic missing patterns frequently encountered in real-world data [7].

### B. Attention-Based Models: SAITS

Recent transformer architectures have attracted significant attention in the field of time series data imputation due to their ability to model long-range temporal dependencies [8]. Notably, advanced architectures have been proposed to better capture multivariate correlations. iTransformer [9] introduced an inverted structure that embeds the entire time series of each variate independently to learn multivariate representations. Similarly, Crossformer [10] utilized a two-stage attention mechanism to explicitly model cross-dimension dependencies between different variables. Building upon these architectural advancements, SAITS (Self-Attention-based Imputation for Time Series) [11] has achieved major progress in the specific task of imputation by applying a self-attention mechanism to distinguish observed and missing values. SAITS integrates the input sequence with a missing indicator mask, enabling the model to explicitly learn structural patterns associated with missingness.

SAITS utilizes a Transformer-based architecture, employing self-attention mechanisms to explicitly model temporal dependencies between observed and missing values. Despite its effectiveness, it remains constrained by the fundamental limitations of standard self-attention. First, its reliance on linear projections limits the capacity to capture complex nonlinear dynamics inherent in multivariate time series. More critically, SAITS reconstructs the entire sequence regardless of missing locations, resulting in significant computational overhead and a lack of flexibility for selective imputation.

These considerations necessitate an architecture that can efficiently perform selective completion tailored to missing timestamps while more flexibly modeling nonlinear temporal relationships.

### C. Implicit Neural Representations

Implicit Neural Representations (INR) [12] have emerged as a powerful framework for modeling continuous signals by parameterizing a signal as a mapping from a coordinate space to its corresponding values via a neural network. Unlike traditional discrete representations, INR enable resolution-independent reconstruction by optimizing the network to fit observed samples, proving particularly effective for restoring sparse or partially observed signals in domains such as images and 3D shapes.

The key advantage of INR lies in their ability to represent complex signals as continuous functions, allowing for flexible reconstruction where the model is queried only at desired coordinates. Motivated by this, recent works have extended INR to the temporal domain, treating time as a coordinate to efficiently predict values at specific timestamps. This coordinate-centric approach provides a compelling foundation for time-series imputation, enabling the selective reconstruction of missing values while capturing complex nonlinear dynamics through MLP-based architectures.

### D. Cross-Attention for Selective Reconstruction

Drawing inspiration from the coordinate-based nature of INR, we propose a novel imputation framework that reformulates time-series restoration as a query-based task.

Analogous to how image INR map spatial coordinates $(x, y)$ to pixel values, our approach treats missing timestamps as query coordinates within a Cross-Attention mechanism [13], [14]. This formulation enables selective reconstruction, significantly reducing inference computational costs by targeting only the specific time points requiring imputation, rather than reconstructing the entire sequence.

By adopting this coordinate-driven strategy, the proposed framework efficiently integrates the query mechanism with standard attention operations. This architecture effectively eliminates the redundancy of full-sequence generation, ensuring that computational resources are allocated exclusively to the restoration of missing values, thereby achieving a highly efficient imputation process tailored for resource-constrained environments.

## III. Method

In this section, we describe the proposed Cross-Attention–based multivariate time-series imputation model as illustrated in Fig. 1. The model effectively leverages the correlations among variables within a time series through an attention mechanism, and integrates an additional MLP layer to enhance nonlinearity in the Multi-Head Attention (MHA) structure. Through this design, the model is constructed to achieve lower reconstruction error compared to conventional Cross-Attention architectures.

### A. Encoder: Context Extraction with Mask-Aware Self-Attention

In the input time-series data $\mathbf{X} \in \mathbb{R}^{T \times D}$, missing values are zero-imputed, and a binary mask $\mathbf{M} \in \{0, 1\}^{T \times D}$ is provided,

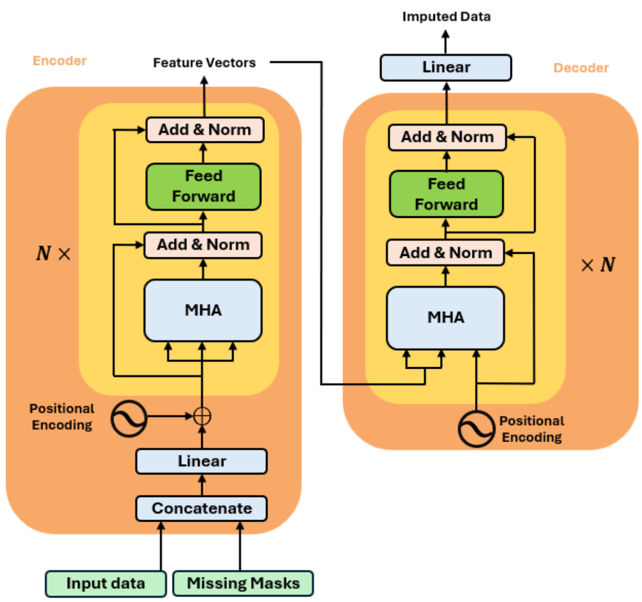


Fig. 1: Overall Framework of the Proposed Model

where a value of 1 explicitly denotes that the timestamp $t$ belongs to the set of missing indices $I$ (i.e., $M_{t,d} = 1$ if $t \in I$). The encoder combines these two pieces of information to extract global contextual features. First, the input data and the mask are concatenated along the channel dimension and linearly projected to map them into a $d_{model}$-dimensional space. Then, sinusoidal positional encoding ($\mathbf{PE}$) is added to incorporate ordering information.

$$\mathbf{H}_{in} = \text{Linear}(concat(\mathbf{X}, \mathbf{M})) + \mathbf{PE} \tag{1}$$

The latent representation $\mathbf{H}_{in}$ is processed by a Transformer encoder with $L$ layers, producing a contextual vector $\mathbf{Z}$ that reflects the temporal relationships of the observed data.

$$\mathbf{Z} = \text{Encoder}(\mathbf{H}_{in}) \in \mathbb{R}^{T \times d_{model}} \tag{2}$$

The resulting $\mathbf{Z}$ serves as the Key and Value in the decoder's cross-attention.

### B. Decoder Input: Coordinate-Based Query Formulation

Adopting the paradigm of INR, the decoder formulates the imputation task as a coordinate-based prediction problem. It utilizes the discrete position indices of the target timestamps as query coordinates to reconstruct the corresponding values. To perform selective reconstruction, the input query $\mathbf{Q}_{in}$ is generated by extracting the sinusoidal positional encodings corresponding exclusively to the missing timestamps ($I$).

$$\mathbf{Q}_{in} = \mathbf{PE}(I) \tag{3}$$

The resulting representation $\mathbf{Q}_{in}$ encapsulates the discrete positional information of the missing points, serving as the Query input for the CINR-CA module.

### C. Conditional INR via Cross-Attention

The proposed CINR-CA module, the core component of our architecture, is integrated into the decoder to selectively aggregate relevant contextual information from the Encoder representation $\mathbf{Z}$ for imputation.

Unlike conventional self-attention mechanisms that compute interactions across the entire sequence, CINR-CA restricts the attention operation to the specific target timestamps. The Query ($\mathbf{Q}$), Key ($\mathbf{K}$), and Value ($\mathbf{V}$) are derived via standard linear projections to maintain model efficiency:

$$\mathbf{Q} = \mathbf{Q}_{in}\mathbf{W}_Q \tag{4}$$
$$\mathbf{K} = \mathbf{Z}\mathbf{W}_K \tag{5}$$
$$\mathbf{V} = \mathbf{Z}\mathbf{W}_V \tag{6}$$

Here, $\mathbf{W}_Q$, $\mathbf{W}_K$, and $\mathbf{W}_V$ denote the learnable weight matrices. By utilizing the coordinate-based query $\mathbf{Q}_{in}$, which corresponds exclusively to the missing positions, the module avoids redundant computations for observed data. This design significantly reduces the computational overhead while effectively capturing temporal dependencies from the global context $\mathbf{Z}$.

Finally, the reconstructed values are computed through the Multi-Head Attention operation.

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V} \tag{7}$$

### D. Loss Function

To optimize the parameters of the proposed CINR-CA, we utilize the Mean Squared Error (MSE) as the training objective function. Consistent with our selective reconstruction strategy, the loss is computed over the entire sequence, Given the binary mask $\mathbf{M} \in \{0,1\}^{T \times D}$ (where 1 denotes observed and 0 denotes missing), the ground truth $\mathbf{X}$, and the model prediction $\hat{\mathbf{X}}$, the objective consists of two components: the reconstruction loss ($\mathcal{L}_{\text{obs}}$) and the imputation loss ($\mathcal{L}_{\text{miss}}$). We utilize the Mean Squared Error (MSE) for both terms, calculated only on the valid indices for each respective region:

$$\mathcal{L}_{\text{obs}} = \frac{\sum_{t,d} M_{t,d} \left\| X_{t,d} - \hat{X}_{t,d} \right\|_2^2}{\sum_{t,d} M_{t,d}}, \tag{8}$$

$$\mathcal{L}_{\text{miss}} = \frac{\sum_{t,d} (1 - M_{t,d}) \left\| X_{t,d} - \hat{X}_{t,d} \right\|_2^2}{\sum_{t,d} (1 - M_{t,d})}. \tag{9}$$

The final objective function is defined as a weighted sum of these two terms, controlled by the balancing hyperparameters $\alpha$ and $\beta$:

$$\mathcal{L} = \alpha \mathcal{L}_{\text{obs}} + \beta \mathcal{L}_{\text{miss}} \tag{10}$$

By adjusting $\alpha$ and $\beta$, the model can balance the trade-off between learning global temporal consistency from observed data and minimizing prediction errors in the missing intervals

## IV. Experiments

### A. Dataset and Missing Simulation

To ensure rigorous evaluation and reproducibility, we utilized the ETTh1 [15] (Electricity Transformer Temperature — Hourly Level 1) dataset, processed via the PyPOTS [16] ecosystem, a comprehensive Python toolbox for data mining on partially-observed time series. The dataset preparation and missing value injection were executed using the *preprocess_ett* function from the TSI-Bench [17] suite within PyPOTS [16]. Through this standardized pipeline, the data was restructured into fixed-length sequences ($T = 100$) with $D = 7$ features and partitioned into training (102 samples), validation (36 samples), and test (35 samples) sets. For the missingness simulation, we applied a point-wise missing mechanism consistent with standard benchmarking protocols. Specifically, 10% of the observed values were randomly masked (missing rate = 0.1) across the dataset to serve as the ground truth for evaluating the imputation performance of our proposed CINR-CA.

### B. Experimental Setup

In this experiment, we compared the proposed CINR-CA with SAITS [11], a state-of-the-art self-attention-based imputation model, to validate its effectiveness.

For a fair comparison, we adopted the training protocol proposed in the SAITS framework [11]. Accordingly, both models were trained using the MSE loss computed over the entire input sequence, Specifically, for the proposed CINR-CA, the balancing hyperparameters $\alpha$ and $\beta$ in the joint objective function were both set to 1 Crucially, to quantify the practical efficiency gains, we distinguished the operational strategy between training and inference. During training, the model utilized the full sequence length ($T = 100$) to learn global contexts. However, during the inference phase for measuring MFLOPs and accuracy, we applied the proposed selective reconstruction mechanism. Specifically, for a given batch, a timestamp $t$ was targeted for reconstruction only if a missing value existed in at least one of the $D = 7$ channels. Consequently, in our test set evaluation, the model reconstructed an average of 64 timestamps out of the full 100 time steps per sequence. The reported MFLOPs and accuracy metrics were calculated based on this actual reduced workload.

Finally, the comparative analysis focuses on two key metrics representing efficiency and accuracy:

- Inference MFLOPs (Mega Floating Point Operations): Used to quantify the computational complexity. This reflects the reduced cost derived from the average of 64 reconstructed timestamps.
- Test MAE: Used to evaluate the reconstruction accuracy at the missing positions.

### C. Comparative Analysis

Table I presents the quantitative comparison between the proposed CINR-CA and the baseline model, SAITS [11], on the ETTh1 [15] dataset. The experiment was conducted under two different model capacity settings (small and large) to evaluate scalability.

TABLE I: Comparison of Inference Efficiency (MFLOPs) and Accuracy between SAITS and CINR-CA on ETTh1 (miss_rate = 0.1, point missing)

| **Model** | **MFLOPs** | **MAE** |
|---|---|---|
| SAITS | 2.0 | 0.2003(0.012) |
| | 5.1 | 0.1549(0.009) |
| ours (Proposed) | 0.94 | 0.2067(0.008) |
| | 2.81 | 0.1646(0.004) |

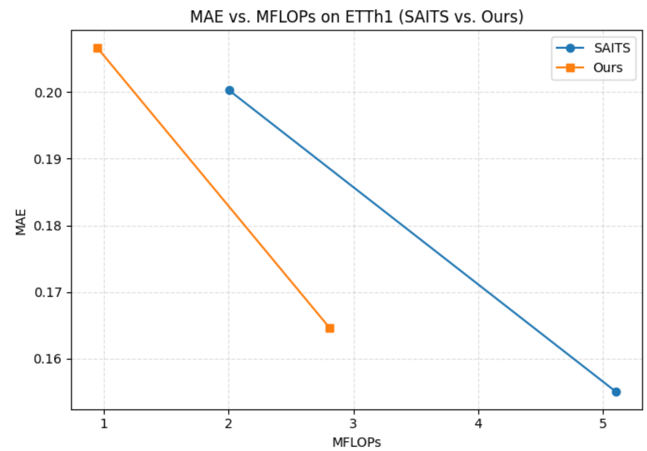


Fig. 2: Performance–Efficiency Trade-off on ETTh1 (SAITS vs. Ours)

**Computational Efficiency** The most significant advantage of CINR-CA is its computational efficiency during inference phase. As shown in the table, our model achieves a substantial reduction in inference MFLOPs compared to SAITS [11], as visualized in the efficiency-performance trade-off plot in Fig. 2. In the smaller setting, CINR-CA reduces computational cost by approximately 53% (2.0 $\rightarrow$ 0.94 MFLOPs), and in the larger setting, by roughly 45% (5.1 $\rightarrow$ 2.81 MFLOPs). This efficiency stems from our selective reconstruction strategy, which queries only the missing timestamps, whereas SAITS [11] reconstructs the entire sequence including observed values.

**Imputation Accuracy** In terms of accuracy, CINR-CA demonstrates performance highly competitive with the state-of-the-art. While SAITS [11] achieves marginally lower MAE (0.1549 vs. 0.1646 in the large setting), the performance gap is minimal considering the nearly two-fold reduction in computational resources. This result indicates that CINR-CA effectively optimizes the trade-off between accuracy and efficiency, making it a more practical solution for resource-constrained environments.

## V. Discussion

The proposed CINR-CA framework demonstrates that high computational efficiency and competitive imputation accuracy can coexist in time-series reconstruction. Despite operating with significantly fewer inference MFLOPs than SAITS [11], our method attains accuracy comparable to the state-of-the-art on the ETTh1 [15] dataset. This result indicates that querying exclusively the missing coordinates enables the model to strategically focus its computational effort on the regions most critical for reconstruction, thereby optimizing the resource allocation.

A primary factor of these efficiency gains is the selective reconstruction strategy, inspired by INR. Unlike existing approaches that reconstruct the full sequence, CINR-CA generates queries solely for missing timestamps and processes encoder representations accordingly. This design minimizes redundant computation on observed data, allowing the decoder to allocate greater attention to the missing segments. Consequently, the model effectively optimizes the conventional trade-off between efficiency and accuracy, a trend clearly observable in the Pareto frontier shown in Fig. 2.

Nevertheless, the proposed framework has several limitations that could be addressed in future research. First, while our focus was on demonstrating efficiency, we observed that merely scaling up model parameters to maximize absolute performance yielded diminishing returns, with accuracy plateauing at certain levels (e.g., around 0.16 MAE). This suggests that a systematic hyperparameter optimization is required to identify the optimal configuration that fully exploits the model's capacity beyond simple scaling. Second, our current experiments were confined to fixed-length windows within regular grid settings on a single dataset. Given that our coordinate-based mechanism intrinsically handles continuous time, it holds significant potential for irregularly sampled time-series. Future work will aim to extend this framework to irregular domains and validate its generalization capability across a broader range of multivariate datasets and varying missing rates, including block-missing patterns.

## VI. Conclusion

In this paper, we proposed CINR-CA, a novel framework designed for highly efficient time-series imputation. By adopting a coordinate-based query mechanism inspired by INR [12], our model shifts the paradigm from full-sequence reconstruction to selective reconstruction, targeting exclusively the missing timestamps.

Experimental results on the ETTh1 [15] dataset demonstrate that CINR-CA significantly outperforms the state-of-the-art model, SAITS [11], in terms of computational efficiency—reducing inference MFLOPs by up to 53% while maintaining competitive reconstruction accuracy. This study successfully optimizes the trade-off between model capacity and efficiency, offering a practical solution for resource-constrained environments where low latency is critical.

Future work will focus on two key directions to further advance this framework. First, we aim to extend the coordinate-based mechanism to handle irregularly sampled time-series, capitalizing on its inherent flexibility. Second, we plan to generalize our findings by evaluating the model on diverse multivariate time-series datasets and extending the coordinate-based mechanism to handle complex scenarios, such as irregularly sampled data and variable missing patterns.

## References


[1] S. Aimin *et al.*, "Missing data imputation: A comprehensive review," *Open Journal of Statistics*, 2024.

[2] D. B. Rubin, *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons, 1987.

[3] S. Faisal *et al.*, "Multiple imputation using nearest neighbor methods," *Information Sciences*, 2021.

[4] Y. Khan *et al.*, "A novel ranked k-nearest neighbors algorithm for missing data imputation," *BMC Medical Research Methodology*, 2024.

[5] X. Zheng, B. Dumitrescu, J. Liu, and C. D. Giurcăneanu, "Multivariate time series imputation: An approach based on dictionary learning," *Entropy*, vol. 24, no. 8, p. 1057, 2022.

[6] J. C. Jakobsen *et al.*, "Missing data in clinical research: A practical guide," *BMC Medical Research Methodology*, 2017.

[7] S. van Buuren, *Flexible Imputation of Missing Data*. CRC Press, 2nd ed., 2018.

[8] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, and L. Sun, "Transformers in time series: A survey," *arXiv preprint arXiv:2202.07125*, 2022.

[9] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long, "itransformer: Inverted transformers are effective for time series forecasting," *arXiv preprint arXiv:2310.06625*, 2023.

[10] Y. Zhang and J. Yan, "Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting," in *The eleventh international conference on learning representations*, 2023.

[11] W. Du, D. Côté, and Y. Liu, "Saits: Self-attention-based imputation for time series," *Expert Systems with Applications*, vol. 219, p. 119619, 2023.

[12] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," *Advances in neural information processing systems*, vol. 33, pp. 7462–7473, 2020.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[14] A. Jaegle, S. Borgeaud, J.-B. Alayrac, C. Doersch, C. Ionescu, D. Ding, S. Koppula, D. Zoran, A. Brock, E. Shelhamer, *et al.*, "Perceiver io: A general architecture for structured inputs & outputs," *arXiv preprint arXiv:2107.14795*, 2021.

[15] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Conference*, vol. 35, pp. 11106–11115, AAAI Press, 2021.

[16] W. Du, "PyPOTS: a Python toolbox for data mining on Partially-Observed Time Series," *arXiv preprint arXiv:2305.18811*, 2023.

[17] W. Du, J. Wang, L. Qian, Y. Yang, Z. Ibrahim, F. Liu, Z. Wang, H. Liu, Z. Zhao, Y. Zhou, *et al.*, "Tsi-bench: Benchmarking time series imputation," *arXiv preprint arXiv:2406.12747*, 2024.