RESEARCH ARTICLE

M-SAITS: A Dual-Stage Time Series Imputation Network via Decoupled Large-Kernel Convolution and Diagonally-Masked Attention

Tingli Su[1], **Gongxin Wang**[1], **Yuting Bai**[1,*] **and Rui Wan**[1]

[1] School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China

Abstract

Missing value imputation in multivariate time series is a critical challenge in the field of data mining. Although Transformer-based methods excel in modeling long-range dependencies, their inherent point-wise attention mechanisms often lack explicit modeling of local inductive biases in time series, making it difficult to effectively capture local smoothness and evolutionary trends. Furthermore, existing feature embedding strategies struggle to fully decouple the internal temporal evolution of variables from complex cross-variable dependencies. To address these limitations, this paper proposes a novel dual-stage imputation framework named M-SAITS. This framework innovatively introduces a decoupled feature encoder based on large-kernel depthwise convolutions. By utilizing an extended effective receptive field, it explicitly enhances the model's perception of local trends. Additionally, it employs a grouped convolution structure to achieve decoupled modeling of intra-variable temporal patterns and inter-variable interaction features. On this basis, combined with a Diagonally-Masked Self-Attention mechanism, the framework physically blocks information leakage paths while achieving lossless global context aggregation. Relying on a "Preliminary Inference–Iterative Refinement" cascade strategy and a masked weighted joint optimization objective, the model achieves high-fidelity data reconstruction. Extensive experiments on multiple benchmark datasets, such as Electricity and Air Quality, demonstrate that this method significantly outperforms existing state-of-the-art models across multiple evaluation metrics. Notably, in high-dimensional electricity data imputation tasks, M-SAITS achieves substantial performance improvements over baseline models such as CSDI and Transformer, with the Mean Absolute Error significantly reduced (up to approximately 60% under low missing rates).

Keywords: time series imputation, large-kernel convolution, decoupled feature representation, diagonally-masked attention, self-supervised learning.

1 Introduction

With the rapid proliferation of the Internet of Things (IoT), smart healthcare, and intelligent transportation systems, the scale of collected multivariate time series

Submitted: 09 January 2026
Accepted: 20 February 2026
Published: 08 March 2026

Vol. 2, **No.** 2, 2026.
10.62762/TMI.2026.671182

***Corresponding author:**
✉ Yuting Bai
baiyuting@btbu.edu.cn

Citation
Su, T., Wang, G., Bai, Y., & Wan, R. (2026). M-SAITS: A Dual-Stage Time Series Imputation Network via Decoupled Large-Kernel Convolution and Diagonally-Masked Attention. *ICCK Transactions on Machine Intelligence*, 2(2), 106–115.

data has grown exponentially. These data contain critical status information regarding system operations and are vital for anomaly detection, trend prediction, and decision-making [1–3]. However, in practical applications, collected data often suffers from severe missing values due to sensor malfunctions, network transmission packet loss, or external environmental interference. As noted by Wang et al. [4] in a recent survey, high-quality data imputation is not only a core step in data preprocessing but also the cornerstone for guaranteeing the performance of downstream tasks.

To address the missing value problem, the academic community has witnessed an evolution from statistical methods to deep learning approaches. Early deep imputation models were primarily based on Recurrent Neural Networks (RNNs). For instance, Che et al. [5] and the BRITS model proposed by Cao et al. [6] utilized bidirectional LSTMs to capture time lags, inferring missing values through historical and future contexts. To further enhance generation realism, Generative Adversarial Networks (GANs) were introduced to the field, such as GAIN [7] and SSGAN [8], which attempt to simulate real data distributions through adversarial training. However, the serial computation nature of RNNs limits their training efficiency, while GANs face challenges regarding training instability and mode collapse.

In recent years, the Transformer architecture has gradually become the mainstream paradigm in this field due to its powerful global modeling capabilities. The self-attention mechanism proposed by Vaswani et al. [9] enables models to capture long-range dependencies in parallel. Building on this, models such as SAITS [10] and Autoformer [11] have achieved significant breakthroughs in imputation accuracy by improving attention mechanisms. Nevertheless, existing Transformer-based methods still possess significant limitations: their point-wise attention calculation often neglects the inherent local inductive bias of time series, leading to suboptimal performance in waveform continuity and local smoothness of imputation results. Moreover, existing feature embedding methods struggle to effectively decouple the internal temporal evolution of variables from complex cross-variable dependencies [12, 13].

Addressing the aforementioned challenges, this paper proposes a novel dual-stage imputation framework named M-SAITS. The main contributions of this paper are summarized as follows:

First, we propose a dual-stage imputation architecture by fusing large-kernel convolution and attention mechanisms. Specifically, the large-kernel design philosophy was introduced into the imputation task, compensating for the deficiency of pure Transformer architectures in local inductive bias. Combined with the Diagonally-Masked Self-Attention (DMSA) mechanism, this approach physically blocks information leakage while achieving lossless global context aggregation.

Second, a feature encoder was designed based on decoupled large-kernel convolutions. Through a hierarchical encoding mechanism, we utilize an extended effective receptive field to explicitly enhance the model's perception of local trends. Furthermore, the grouped convolution structure effectively decouples the internal temporal evolution of variables from complex spatial dependencies across variables.

Finally, a "Preliminary Inference–Iterative Refinement" cascade optimization strategy was constructed in this work. Relying on a dual-stage progressive generation strategy and a masked weighted joint loss function, the model recovers missing data from coarse to fine. Extensive experiments on multiple benchmark datasets demonstrate that M-SAITS significantly surpasses state-of-the-art models, exhibiting superior accuracy and robustness across both random point and continuous block missing scenarios.

The structure of this paper is organized as follows: Section 2 reviews related work in the field of time series imputation, including Recurrent Neural Networks, Transformers, and recently emerging probabilistic and convolutional models; Section 3 details the core architecture, including the decoupled feature encoder design, and dual-stage training strategy of M-SAITS; Section 4 presents comparative experimental results on three real-world datasets, along with ablation studies and visualization analysis; Section 5 concludes the paper and outlines future research directions.

2 Related Work

This section reviews related research in the field of time series imputation from three dimensions: recurrent and generative models, Transformer-based attention models, as well as probabilistic and modern convolutional models. On this basis, we deeply analyze the limitations of existing methods in capturing local evolutionary trends and decoupling complex variable dependencies, and clarify how the M-SAITS framework proposed in this paper

compensates for these deficiencies by fusing large-kernel convolution with attention mechanisms, thereby establishing the motivation for improvement and the unique contributions of this study.

2.1 Imputation Based on Recurrent Neural Networks and Generative Adversarial Networks

Early deep imputation research primarily focused on improving RNNs to adapt to irregular time intervals. Che et al. [5] introduced a time-decay mechanism into GRUs to handle missing patterns by simulating information decay over time. The BRITS model proposed by Cao et al. [6] constructed a bidirectional recurrent dynamical system that takes the missing mask as a conditional input, while enforcing consistency constraints using forward and backward data flows. To address the lack of diversity in deterministic imputation, Yoon et al. [7] proposed GAIN, which introduces a Hint Vector to assist the discriminator, enabling the generator to synthesize data conforming to the original distribution. Miao et al. [8] further introduced semi-supervised learning into the GAN framework, utilizing label information to guide the imputation process. Although these methods laid the foundation for deep imputation, RNNs face gradient vanishing problems when processing long sequences, and GANs face challenges regarding training instability and mode collapse.

2.2 Transformer-Based Self-Attention Imputation

To overcome the limitations of RNNs, the Transformer [9] has been widely adopted due to its parallel computing advantages and global receptive field. Autoformer, proposed by Wu et al. [11], revolutionized long-term time series modeling through sequence decomposition and auto-correlation mechanisms, which have subsequently been introduced to imputation tasks. As a representative work in this field, SAITS, proposed by Du et al. [10], introduced DMSA and a weighted joint optimization objective, significantly improving imputation accuracy by eliminating information leakage. Additionally, Qiu et al. [12] explored the combination of tensor ring decomposition and graph regularization, attempting to capture latent structures in high-dimensional spaces. However, pure attention mechanisms often lack inductive bias for local neighborhood features, leading to suboptimal performance on high-frequency fluctuating data.

2.3 Probabilistic Models and Modern Convolutional Networks

Recent research presents two new trends. One is the resurgence of convolutional networks. TimesNet, proposed by Wu et al. [13], captured complex temporal features through multi-period 2D variation modeling. More critically, ModernTCN [14] , proposed by Luo and Wang, demonstrated that by introducing large convolution kernels and decoupled structures, pure convolutional architectures can achieve effective receptive fields comparable to Transformers. Inspired by this, this paper has tried to introduce large-kernel convolution into imputation tasks to compensate for the deficiencies of pure attention mechanisms in local feature extraction. The other trend is the rise of probabilistic generative models. CSDI, proposed by Tashiro et al. [15], applies diffusion models to time series imputation, generating high-quality samples through progressive denoising, becoming one of the current SOTA baselines. GP-VAE, proposed by Fortuin et al. [16], combines variational autoencoders with Gaussian processes to provide uncertainty estimates for missing values.

3 Methodology

This section details the mathematical definition and core architecture of M-SAITS. As illustrated in Figure 1, the framework comprises three key components: a Decoupled Feature Encoder based on large-kernel convolution, a DMSA module, and a dual-stage progressive imputation strategy.

Given a multivariate time series $X \in \mathbb{R}^{T \times D}$, where T denotes the time step and D represents the variable dimension, we define a binary mask matrix $M \in \{0,1\}^{T \times D}$. In this matrix, $M_{t,d} = 1$ indicates that the observation is present, while $M_{t,d} = 0$ denotes a missing value. The model takes the observed components $\tilde{X}_{\text{in}} = X \odot M$ and the mask M as inputs, with the objective of predicting the missing values to recover the complete sequence.

To construct the self-supervised training objective, an indicating mask M_{ind} is randomly generated in each iteration to artificially mask a subset of the observed values:

$$M_{\text{ind}}^{(t,d)} = \begin{cases} 1, & \text{if } x_t^d \text{ is masked artificially} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

At this point, the augmented data $\tilde{X}$ and the mask $\tilde{M}$ input into the model are redefined as:

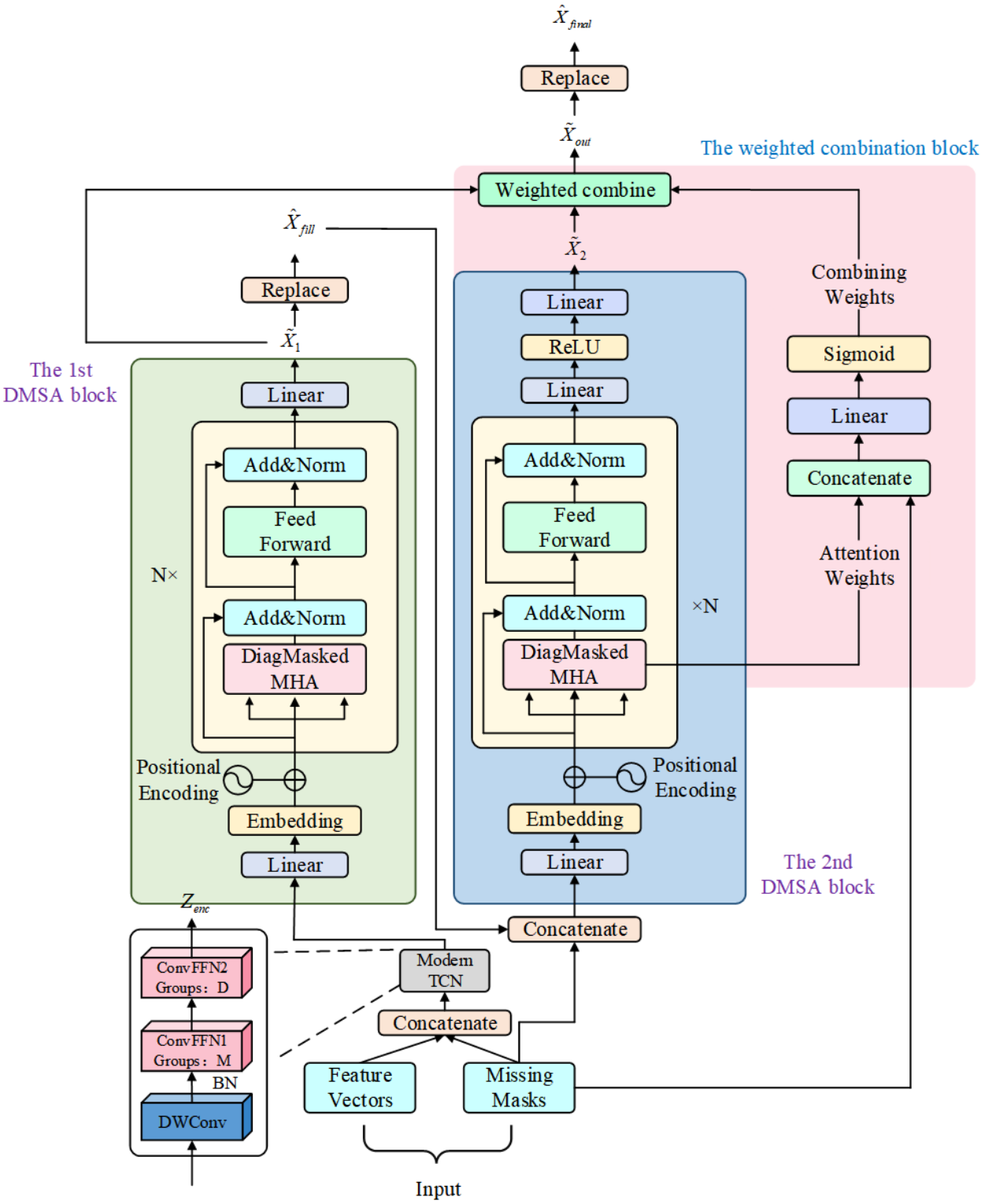

Figure 1. The overall architecture of the proposed M-SAITS framework.

$$\tilde{X} = X \odot (M - M_{\text{ind}}), \quad \tilde{M} = M - M_{\text{ind}} \tag{2}$$

where $\tilde{M}$ is set to 1 exclusively for positions corresponding to the original observations that have not been artificially masked, thereby preventing the input from containing any original or artificial missing data.

To fully decouple the complex dependencies within multivariate time series, we introduce a hierarchical encoding mechanism inspired by ModernTCN. This module strictly partitions the feature extraction process into three independent phases: temporal relationship

modeling, channel relationship modeling, and variable relationship modeling.

First, to establish the foundation for feature extraction, the input Concat$(\tilde{X}, \tilde{M})$ is mapped into a high-dimensional latent space H_0. Subsequently, the model proceeds to decoupled modeling across three dimensions. Temporal relationship modeling utilizes large-kernel depthwise convolution (DWConv) to capture long-range temporal dependencies. The DWConv slides along the time axis within each variable's independent channel, ensuring the extraction of pure temporal trends exclusively:

$$Z_{\text{time}} = \text{BN}(\text{DWConv}_K(H_0)) \tag{3}$$

where K denotes the size of the large convolution kernel (e.g., $K = 51$). Channel relationship modeling utilizes the first convolutional feed-forward network (ConvFFN$_1$) to capture interactions within the feature dimensions. This module typically consists of pointwise convolutions that operate on the hidden dimensions to enhance the representation capacity of the features:

$$Z_{\text{chn}} = \text{GELU}(\text{ConvFFN}_1(Z_{\text{time}})) + Z_{\text{time}} \tag{4}$$

Variable relationship modeling utilizes a second convolutional feed-forward network (ConvFFN$_2$) to specifically capture cross-variable spatial dependencies. This module is designed to fuse the cross-correlation information among different sensors or variables, thereby achieving the decoupling of the spatial dimension:

$$Z_{\text{enc}} = \text{GELU}(\text{ConvFFN}_2(Z_{\text{chn}})) + Z_{\text{chn}} \tag{5}$$

To prevent information leakage during the aggregation of global context (specifically, the "self-viewing" problem), we employ a DMSA mechanism.

A diagonal mask matrix $M_{\text{diag}} \in \mathbb{R}^{T\times T}$ is defined, where the diagonal elements are set to $-\infty$ and the remaining elements are 0. The attention is then calculated as follows:

$$\text{DMSA}(Q,K,V) = \underbrace{\text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}} + M_{\text{diag}}\right)}_{\text{attention weights } A} V = AV \tag{6}$$

where $Q, K, V \in \mathbb{R}^{T\times d_k}$ respectively represent the Query, Key, and Value matrices derived through linear transformations, and d_k is the dimension of the projected feature subspace (e.g., $d_k = 64$). A denotes the attention weight matrix after diagonal masking, which captures the global inter-dependencies across time steps; and AV is the global feature representation obtained by the weighted aggregation of the information vectors V via the weight matrix A. Due to $M_{\text{diag},ii} = -\infty$, the diagonal elements of the normalized attention weight matrix A are forced to zero. This implies that the prediction at time step t relies exclusively on its context set $\{x_1, \ldots, x_{t-1}, x_{t+1}, \ldots, x_T\}$, thereby physically blocking the label leakage path.

To achieve a coarse-to-fine reconstruction, the model adopts a cascaded strategy of "initial inference followed by iterative refinement." In the first stage, an initial prediction $\tilde{X}_1$ is generated based on Z_{enc}. Subsequently, a hybrid filling strategy is employed to construct a complete sequence $\tilde{X}_{\text{fill}}$, filling all missing cavities:

$$\tilde{X}_{\text{fill}} = M \odot \tilde{X}_{in} + (1 - M) \odot \tilde{X}_1 \tag{7}$$

In the second stage, to effectively aggregate the preliminary global estimation $\tilde{X}_1$ and the refined residual features $\tilde{X}_2$, we introduce a learnable gating coefficient η. This generates the fused reconstruction sequence $\tilde{X}_{\text{out}}$, which represents the model's best estimation for the entire timeline:

$$\tilde{X}_{\text{out}} = (1 - \eta) \odot \tilde{X}_1 + \eta \odot \tilde{X}_2 \tag{8}$$

As illustrated in the 'Replace' module in Figure 1, to ensure the integrity of the observed information, the final imputed data $\hat{X}_{\text{final}}$ is obtained by strictly replacing the predictions at observed positions with the original input data:

$$\hat{X}_{\text{final}} = M \odot \tilde{X}_{in} + (1 - M) \odot \tilde{X}_{\text{out}} \tag{9}$$

To balance the inferential capability for missing values with the reconstructive capability for observed data, a weighted joint loss function is employed. It consists of two components: the Masked Imputation Task (MIT) and the Observation Reconstruction Task (ORT). Specifically, the loss corresponding to MIT, denoted as L_{MIT}, calculates only the absolute error of the

artificially missing points ($M_{\text{ind}} = 1$) to train the model's inferential power as follows:

$$L_{\text{MIT}} = \frac{\sum_{t,d}\left(|\tilde{X}_{\text{out}}^{(t,d)} - X^{(t,d)}| \cdot M_{\text{ind}}^{(t,d)}\right)}{\sum_{t,d} M_{\text{ind}}^{(t,d)} + \delta} \tag{10}$$

L_{ORT}, corresponding to ORT calculates the error of the retained observations ($M_{\text{rec}} = M - M_{\text{ind}}$) to constrain the manifold consistency of the feature representations as follows:

$$L_{\text{ORT}} = \sum_{k\in\{1,2,\text{out}\}} \frac{1}{3}\left(\frac{\sum_{t,d}\left(|\tilde{X}_{k}^{(t,d)} - X^{(t,d)}| \cdot M_{\text{rec}}^{(t,d)}\right)}{\sum_{t,d} M_{\text{rec}}^{(t,d)} + \delta}\right) \tag{11}$$

The total loss function is defined as:

$$L_{\text{total}} = L_{\text{ORT}} + \lambda L_{\text{MIT}} \tag{12}$$

where δ is a small constant to prevent division by zero, and λ is a hyper-parameter to balance the weights of the two tasks.

4 Experiments

To comprehensively evaluate the model's generalization ability across different dimensions and temporal dependency patterns, we conducted experiments on three widely used real-world multivariate time series benchmark datasets.

1. Air Quality [17]: With a total length of 35,065 time steps and 132 feature variables, this dataset focuses on the meteorological domain and has an original missing rate of 1.6%.
2. Electricity [18]: Containing 26,304 time steps and 370 feature variables, this dataset focuses on the power domain and has an original missing rate of 0%.
3. ETT [19]: With a total length of 17,420 time steps and 7 feature variables, this dataset also focuses on the power domain and has an original missing rate of 0%.

Although two datasets possess an original missing rate of 0%, data missingness is inevitable during the monitoring process, necessitating the use of imputation methods for restoration. To simulate realistic missing scenarios, we employ two masking strategies to generate test data:

1. Random Point Missing: Observed points are randomly discarded at ratios of 10% and 50%.
2. Continuous Block Missing: A continuous segment of time steps is randomly masked to simulate sensor failure or communication interruption, ensuring a 50% missing rate.

To quantitatively evaluate the model's capability in missing data imputation, this study selects Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Relative Error (MRE) as the core evaluation metrics. During the testing phase, to ensure that the errors are calculated exclusively for the artificially masked ground-truth observations, we need to define a test mask matrix M_{test}. In this matrix, $m_d^t = 1$ represents the "masked ground truth" used for testing at that position, while $m_d^t = 0$ represents all other positions. All metrics are derived by performing an Hadamard product between the prediction error matrix and M_{test}. The calculation formulas are as follows:

$$\text{MAE} = \frac{\sum_t \left(|x_d^t - \hat{x}_d^t| \cdot m_d^t\right)}{\sum_t m_d^t} \tag{13}$$

where x_d^t denotes the ground truth, $\hat{x}_d^t$ the imputed values, $m_d^t \in M_{\text{test}}$ the test mask indicator, and $\sum_t m_d^t$ the total number of test samples.

Table 1. Experimental results of 10% point missing imputation.

	Electricity			ETT			Air Quality		
	MAE	MSE	MRE	MAE	MSE	MRE	MAE	MSE	MRE
Transformer	1.316	3.240	0.704	0.178	0.066	0.210	0.142	0.109	0.190
BRITS	0.971	2.194	0.520	0.145	0.064	0.171	0.127	0.101	0.169
CSDI	1.483	117.914	0.793	0.151	0.057	0.178	0.102	0.507	0.135
GP-VAE	1.152	2.778	0.616	0.329	0.202	0.388	0.240	0.187	0.320
M-SAITS	**0.523**	**1.105**	**0.281**	**0.122**	**0.039**	**0.145**	**0.098**	**0.092**	**0.132**

As shown in Table 1, under the low missing rate setting of 10%, M-SAITS achieved optimal performance across all datasets and metrics. Notably, on the Electricity dataset, which possesses the highest dimensionality, the MAE of M-SAITS was only 0.523, significantly outperforming Transformer (1.316) and CSDI (1.483). This demonstrates that the decoupled convolutional structure introduced in the model effectively leverages the complex dependencies among the 370 variables, exhibiting distinct advantages on high-dimensional data. On the Air Quality dataset, the MAE of M-SAITS was 0.098, superior to the runner-up Transformer (0.142), indicating the model's strong capability in capturing local variations in meteorological data. Figure 2 visualizes the imputation results under the

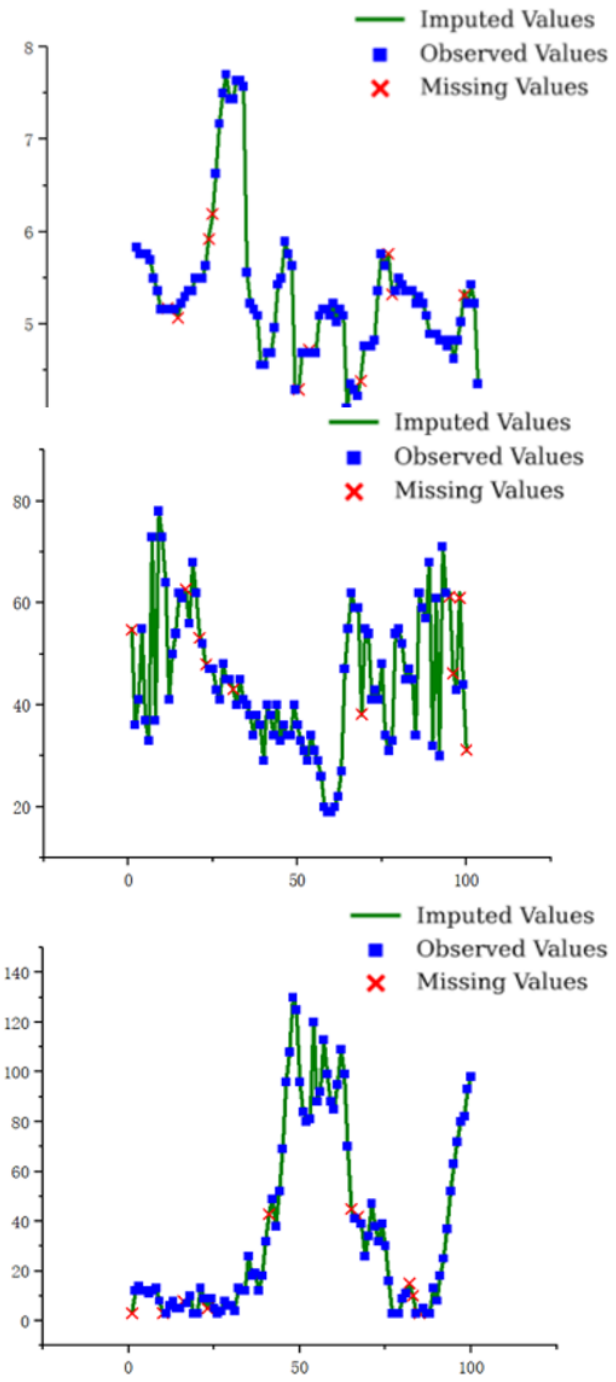

Figure 2. Experimental results of 50% point missing imputation. The sub-figures from left to right correspond to the ETT, Electricity, and Air Quality datasets, respectively.

50% point missing scenario, where the proposed M-SAITS most closely fits the ground truth across all three datasets, particularly in capturing peak-valley fluctuations.

As the missing rate increases to 50%, as shown in Table 2, the errors of all models inevitably rise; however, M-SAITS demonstrates exceptional robustness. On the ETT dataset, the MAE of M-SAITS is maintained at 0.201, whereas under the same conditions, CSDI records 0.318 and GP-VAE reaches 0.414. This performance is attributed to the extended receptive field provided by the large-kernel convolution, which enables the capture of long-range temporal trends even within sparse data. The visualization results in Figure 3 further corroborate this observation; even with sparse data points, the model remains capable of fitting reasonable trajectories.

Table 2. Experimental results of 50% point missing imputation.

	Electricity			ETT			Air Quality		
	MAE	MSE	MRE	MAE	MSE	MRE	MAE	MSE	MRE
Transformer	1.365	3.554	0.731	0.274	0.162	0.325	0.185	0.192	0.245
BRITS	1.124	2.828	0.602	0.238	0.127	0.281	0.169	0.194	0.224
CSDI	0.798	21.850	0.427	0.318	0.207	0.376	0.144	0.472	0.192
GP-VAE	1.099	2.973	0.588	0.414	0.301	0.490	0.258	0.234	0.343
M-SAITS	**0.756**	**1.769**	**0.405**	**0.201**	**0.095**	**0.238**	**0.135**	**0.181**	**0.179**

Table 3. Imputation results under 50% block missing.

	Electricity			ETT			Air Quality		
	MAE	MSE	MRE	MAE	MSE	MRE	MAE	MSE	MRE
Transformer	1.431	3.964	0.759	0.629	0.850	0.708	0.221	0.267	0.293
BRITS	1.230	3.354	0.653	0.730	1.065	0.822	0.193	0.265	0.256
CSDI	0.922	6.614	0.489	0.558	0.783	0.628	0.211	0.512	0.280
GP-VAE	1.242	3.624	0.659	0.806	1.265	0.908	0.321	0.343	0.426
M-SAITS	**0.876**	**2.135**	**0.465**	**0.517**	**0.674**	**0.598**	**0.171**	**0.218**	**0.223**

Table 3 presents the results for the most challenging 50% continuous block missing experiment, a scenario designed to simulate prolonged sensor malfunctions. On the Electricity dataset, M-SAITS achieved an MAE of 0.876, representing a reduction of approximately 30%–40% compared to Transformer (1.431) and GP-VAE (1.242). Furthermore, on the ETT dataset, the MSE of M-SAITS was 0.674, significantly lower than the 1.065 achieved by BRITS.

Combined with the visualization results in Figure 4, it can be observed that even in scenarios where large segments of data are continuously missing, the imputation curves generated by M-SAITS closely fit the true observed values. The model accurately restores the peak-valley trends of the waveforms without exhibiting significant oscillation or mode collapse. This validates the effectiveness of the "Preliminary Inference–Iterative Refinement" dual-stage strategy in handling large-area missingness.

5 Conclusion

Addressing the issues of insufficient local feature perception and difficulties in decoupling cross-variable dependencies in multivariate time series imputation, this study proposes a novel dual-stage imputation framework named M-SAITS. The primary contribution of this research lies in the innovative fusion of

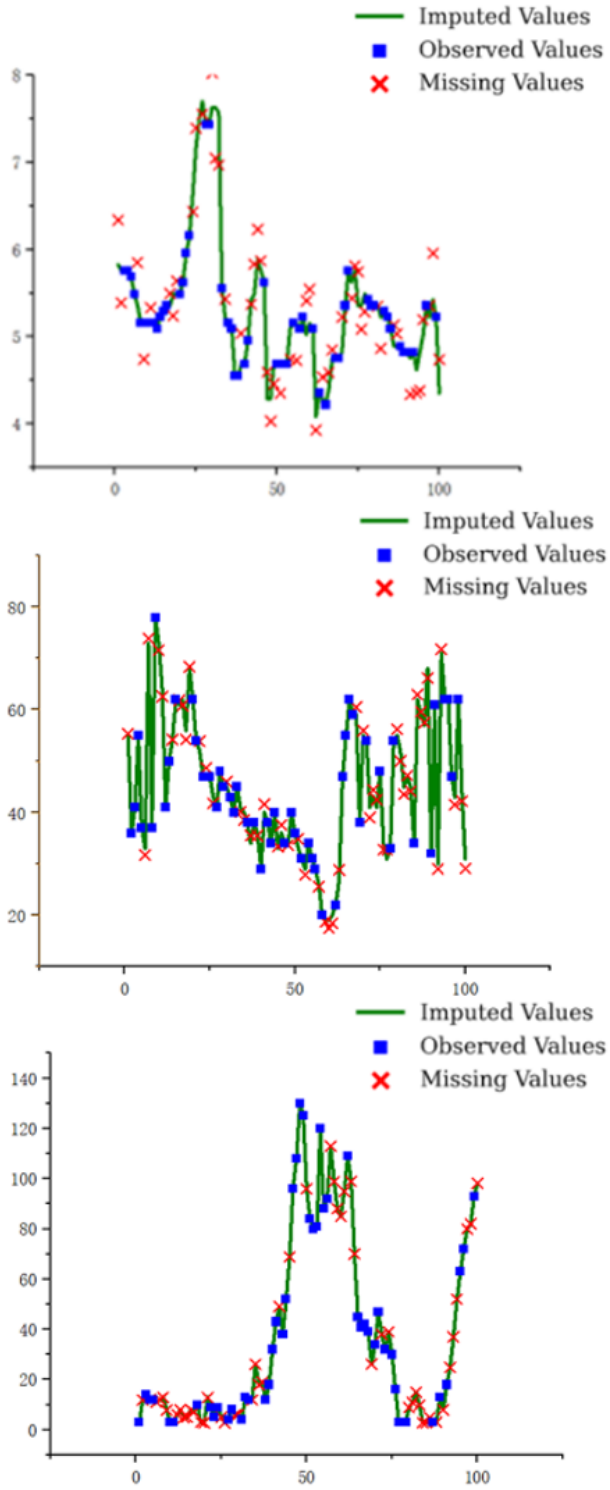

Figure 3. Experimental results of 50% point missing imputation. The sub-figures from left to right correspond to the ETT, Electricity, and Air Quality datasets, respectively.

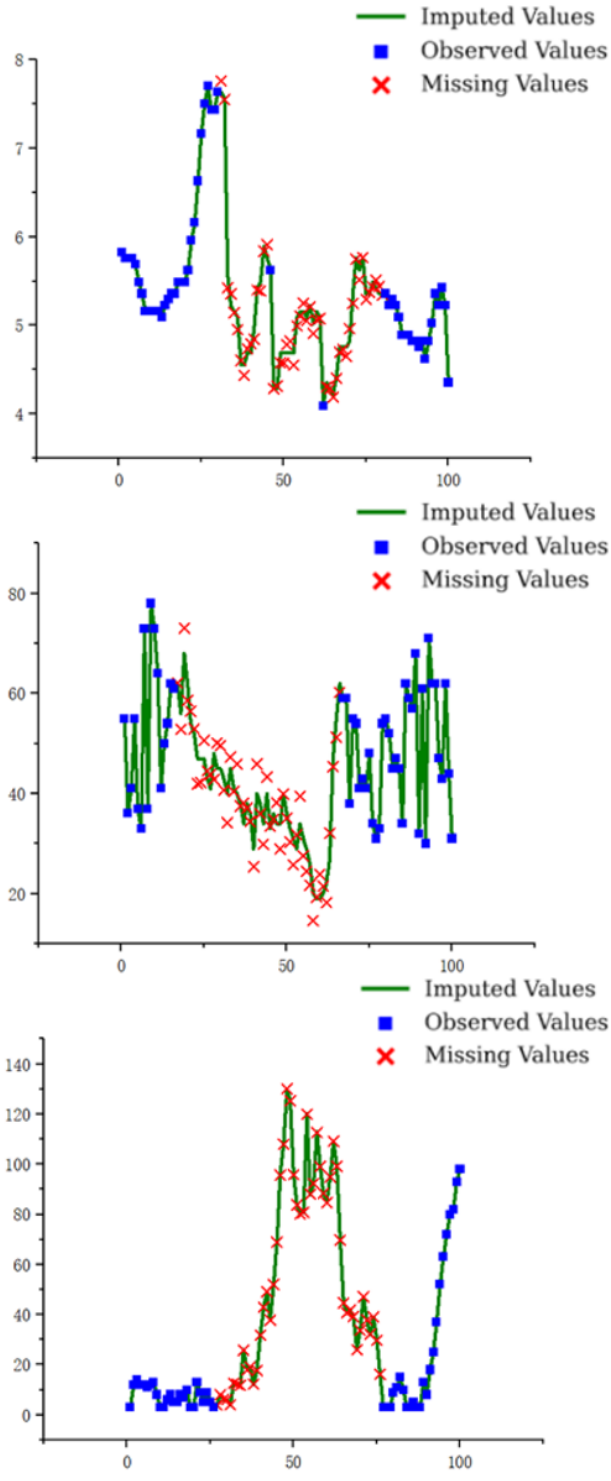

Figure 4. Experimental results of 50% block missing imputation. The sub-figures from left to right correspond to the ETT, Electricity, and Air Quality datasets, respectively.

modern large-kernel convolutional networks and Diagonally-Masked Self-Attention mechanisms to construct a high-precision time series reconstruction model. Inspired by ModernTCN, M-SAITS employs a hierarchical encoding mechanism that strictly divides feature extraction into three independent stages: temporal relations, channel relations, and variable relations. By leveraging a large receptive field, it enhances the inductive bias for local trends while effectively decoupling the internal temporal evolution of variables from cross-variable spatial

dependencies. On this basis, the Diagonally-Masked Self-Attention mechanism physically blocks information leakage paths, ensuring the rigor of global context aggregation. Furthermore, relying on the "Preliminary Inference–Iterative Refinement" cascade strategy and a masked weighted joint optimization objective, the model achieves progressive restoration from coarse to fine. Extensive experiments on three benchmark datasets—Electricity, Air Quality, and ETT—demonstrate that M-SAITS significantly outperforms existing state-of-the-art models, such as Transformer, BRITS, CSDI, and GP-VAE, in terms of MAE, MSE, and MRE metrics under both random point missing and continuous block missing patterns. Notably, in extreme scenarios involving high-dimensional data and continuous large-block missingness, the model exhibits exceptional robustness and generalization capability, offering an efficient and reliable solution for missing data recovery in fields such as IoT and industrial monitoring.

Data Availability Statement

Data will be made available on request.

Funding

This work was conducted as part of a collaborative research project of Beijing Technology and Business University under Grant 2024254.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Li, L., Zhang, J., Wang, Y., & Ran, B. (2018). Missing value imputation for traffic-related time series data based on a multi-view learning method. *IEEE Transactions on Intelligent Transportation Systems, 20*(8), 2933-2943. [CrossRef]

[2] Ren, H., Wang, Y., & Ma, H. (2024). Deep prediction network based on covariance intersection fusion for sensor data. *ICCK Transactions on Intelligent Systematics, 1*(1), 10-18. [CrossRef]

[3] Wang, J., Du, W., Yang, Y., Qian, L., Cao, W., Zhang, K., ... & Wen, Q. (2024). Deep learning for multivariate time series imputation: A survey. *arXiv preprint arXiv:2402.04059*. [CrossRef]

[4] Wang, J., Du, W., Yang, Y., Qian, L., Cao, W., Zhang, K., ... & Wen, Q. (2025, August). Deep learning for multivariate time series imputation: a survey. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence* (pp. 10696-10704). [CrossRef]

[5] Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. *Scientific Reports, 8*(1), 6085. [CrossRef]

[6] Cao, W., Wang, D., Li, J., Zhou, H., Li, L., & Li, Y. (2018). BRITS: Bidirectional recurrent imputation for time series. *Advances in Neural Information Processing Systems, 31*.

[7] Yoon, J., Jordon, J., & Schaar, M. (2018, July). Gain: Missing data imputation using generative adversarial nets. *In International conference on machine learning* (pp. 5689-5698). PMLR.

[8] Miao, X., Wu, Y., Wang, J., Gao, Y., Mao, X., & Yin, J. (2021). Generative semi-supervised learning for multivariate time series imputation. In *AAAI Conference on Artificial Intelligence* (Vol. 35, No. 10, pp. 8983-8991). [CrossRef]

[9] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*.

[10] Du, W., Côté, D., & Liu, Y. (2023). Saits: Self-attention-based imputation for time series. *Expert Systems with Applications, 219*, 119619. [CrossRef]

[11] Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems, 34*, 22419-22430.

[12] Qiu, Y., Zhou, G., Zhao, Q., & Xie, S. (2022). Noisy tensor completion via low-rank tensor ring. *IEEE Transactions on Neural Networks and Learning Systems, 35*(1), 1127-1141. [CrossRef]

[13] Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., & Long, M. (2023). TimesNet: Temporal 2D-variation modeling for general time series analysis. In *International Conference on Learning Representations*.

[14] Luo, D., & Wang, C. (2024). ModernTCN: A modern pure convolution structure for general time series analysis. In *International Conference on Learning Representations*.

[15] Tashiro, Y., Song, J., Song, Y., & Ermon, S. (2021). CSDI: Conditional score-based diffusion models for probabilistic time series imputation. *Advances in Neural Information Processing Systems, 34*, 24804-24816.

[16] Fortuin, V., Baranchuk, D., Rätsch, G., & Mandt, S. (2020, June). Gp-vae: Deep probabilistic time series imputation. In *International conference on artificial intelligence and statistics* (pp. 1651-1661). PMLR.

[17] Zhang, S., Guo, B., Dong, A., He, J., Xu, Z., & Chen, S. (2017). Cautionary tales on air-quality improvement in Beijing. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 473*(2205), 20170457. [CrossRef]

[18] Dua, D., & Graff, C. (2017). UCI machine learning repository, 2017. *URL http://archive. ics. uci. edu/ml, 7*(1), 62.

[19] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021, May). Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 35, No. 12, pp. 11106-11115). [CrossRef]

Tingli Su received her B.E. degree in Mechatronic Engineering and her Ph.D. degree in Control Science and Engineering from the Beijing Institute of Technology, China. From 2009 to 2012, she was a visiting student at the University of Bristol, where she conducted research on networked control systems. She is currently an Associate Professor at the Beijing Technology and Business University. Her research interests include multi-sensor fusion, data analytics, and time series-based state estimation. (Email: sutingli@btbu.edu.cn)

Gongxin Wang received the B.Eng. degree in Intelligent Manufacturing Engineering from Hefei University, China, in 2023. He is currently pursuing the M.Eng. degree in Control Engineering at Beijing Technology and Business University, Beijing, China. His research interests include missing data imputation, time series forecasting, deep learning, and related areas. (Email: 2330602075@st.btbu.edu.cn)

Yuting Bai received the Ph.D. degree in control science and engineering from Beijing Institute of Technology, the M.S. degree in management science and engineering from Beijing Technology and Business University, and the B.S. degree in automation from Beijing Technology and Business University. He is now an associate professor in Beijing Technology and Business University. His research mainly covers information fusion, machine learning and decision-making method. (Email: baiyuting@btbu.edu.cn)

Rui Wan received the B.Eng. degree in Industrial Automation from East China Jiaotong University, China, in 2023. He is currently pursuing the M.Eng. degree in Control Engineering at Beijing Technology and Business University, Beijing, China. His research interests include medical image segmentation, medical image classification, deep learning, and related areas. (Email: 2330602073@st.btbu.edu.cn)