Deep Autoencoder for Missing Data Imputation Based on Coherent Denoising and Spatio-Temporal Neighborhood-Preserving Embedding

XXX, XXX, *Senior Member, IEEE*, XXX,

***Abstract*—Missing data imputation remains a challenging task in multivariate time series analysis, particularly when handling complex spatio-temporal dependencies under arbitrary missing patterns. While deep autoencoders have demonstrated promising capabilities in capturing nonlinear relationships, existing approaches often overlook the intrinsic manifold structures and coherent patterns across multiple masking scenarios. To address these limitations, this paper proposes a deep AutoEncoder based on Coherent denoising and Spatio-temporal neighborhood-preserving embedding (AE-CS). The proposed method incorporates three fundamental constraints: multi-mask consistency regularization enhances robustness against diverse missing mechanisms, spatial locality preservation maintains geometric relationships among variables, and temporal coherence embedding captures dynamic patterns across time steps. These constraints are integrated through an adaptive fusion mechanism that dynamically balances their contributions based on data characteristics. Extensive evaluations on real-world datasets from industrial monitoring and healthcare domains demonstrate that CSAE achieves superior imputation accuracy and better preserves underlying data structures compared to state-of-the-art methods.**

***Index Terms*—Unsupervised neural networks, explainable residual generators, martingale difference inequality, process monitoring, fault detection.**

I. Introduction

INDUSTRIAL process monitoring and control systems rely heavily on complete and accurate sensor measurements to ensure operational safety, product quality, and energy efficiency. Missing data represents a pervasive challenge in such environments, arising from sensor malfunctions, communication failures, or maintenance activities, which can severely compromise system performance and decision-making reliability. The imputation of missing values thus constitutes a critical preprocessing step for subsequent data analysis tasks, including fault detection, soft sensor development, and process optimization. Traditional approaches to handling missing data often employ simple statistical methods or deletion techniques, which may introduce biases or discard valuable information. With the increasing complexity of modern industrial systems and the proliferation of high-dimensional sensory data, there emerges an urgent need for advanced imputation techniques capable of capturing intricate temporal dependencies and spatial correlations among process variables.

Existing methodologies for missing data imputation can be broadly categorized into three classes: statistical approaches, machine learning methods, and deep learning frameworks. Statistical methods typically rely on parametric assumptions about data distribution and employ expectation-maximization algorithms or multiple imputation techniques to estimate missing values. Machine learning approaches leverage neighborhood-based algorithms or matrix factorization to capture local data structures and global correlations. Deep learning paradigms, as hybrid frameworks combining representation learning with sequence modeling, utilize autoencoders, recurrent neural networks, and generative adversarial networks to learn complex data manifolds and temporal patterns. This taxonomy reflects the methodological evolution from model-driven to data-driven imputation strategies, with each category offering distinct advantages and limitations in handling different missing data mechanisms and patterns.

Statistical imputation methods have established solid theoretical foundations through decades of development, with techniques such as multiple imputation by chained equations and expectation-maximization algorithms providing principled frameworks for handling missing data. These methods demonstrate particular strengths in scenarios with well-understood data distributions and moderate missing rates, offering interpretable results with theoretical guarantees. Rubin [1] formalized the multiple imputation framework, establishing rigorous standards for statistical inference with missing data. However, such approaches frequently struggle with high-dimensional industrial datasets where the underlying distributional assumptions may not hold, and they often fail to capture complex nonlinear relationships among process variables, leading to suboptimal performance in practical applications involving dynamic process behaviors.

The advent of machine learning introduced more flexible imputation techniques capable of handling complex data structures without strong distributional assumptions. Neighborhood-based methods such as k-nearest neighbors imputation leverage local similarity patterns in the data space, while matrix factorization approaches model global correlations through low-rank approximations. Stekhoven [2] developed missForest, a random forest-based imputation method

This work was supported in part by the National Natural Science Foundation of China (U1911401, 62003373), and in part by the National Key Research and Development Program of China (2020YFB1713800, 2018AAA0101603).
(Zhuofu Pan and Hongtian Chen contributed equally to this work.) (Corresponding author: Biao Huang.)

Zhuofu Pan, Yalin Wang, and Weihua Gui are with the School of Automation, Central South University, Changsha, 410083, China (e-mail: panfuzz@csu.edu.cn; ylwang@csu.edu.cn; gwh@csu.edu.cn).

Hongtian Chen and Biao Huang are with the Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB T6G 1H9, Canada. (e-mail: chtbaylor@163.com; biao.huang@ualberta.ca).

that handles mixed-type data and complex interactions. These data-driven methods demonstrate superior adaptability to diverse data types and missing patterns compared to traditional statistical approaches. Nevertheless, they often exhibit limitations in capturing long-range temporal dependencies and may suffer from computational inefficiency when applied to large-scale industrial time series data with intricate dynamic characteristics.

Deep learning architectures have recently emerged as powerful frameworks for missing data imputation, combining the representational capacity of neural networks with specialized mechanisms for handling incomplete observations. Autoencoder variants, including denoising autoencensors and variational autoencoders, learn robust feature representations that facilitate accurate reconstruction of missing values. Recurrent neural networks with gating mechanisms, such as long short-term memory and gated recurrent units, effectively model temporal dependencies in sequential data. Yoon [3] proposed Generative Adversarial Imputation Nets, leveraging adversarial training to generate realistic imputations. Despite their impressive performance, these methods often overlook the inherent geometric structure of industrial data manifolds and may not adequately preserve local neighborhood relationships in both spatial and temporal domains, potentially leading to distorted data topology in the imputed results.

To address these limitations, a novel deep AutoEncoder based on Coherent denoising and Spatio-temporal neighborhood-preserving embedding (AE-CS) is proposed for industrial missing data imputation. The methodology integrates three complementary constraints: multi-mask consistency regularization enhances robustness to varying missing patterns; spatial locality preservation maintains geometric relationships among process variables; temporal coherence embedding captures local dynamic structures within time series. Primary contributions of this work include:

1) the development of an integrated imputation framework that simultaneously addresses data corruption and manifold preservation;
2) the formulation of weighted embedding losses that adaptively maintain spatio-temporal neighborhood structures;
3) extensive experimental validation demonstrating superior performance compared to state-of-the-art benchmarks across multiple industrial datasets.

The remainder of this paper is organized as follows. Section II introduces fundamental concepts and formally defines the missing data imputation problem. Section III elaborates on the proposed AE-CS framework, detailing its architectural components and optimization objectives. Section **??** presents comprehensive experimental results and comparative analysis with baseline methods. Finally, Section **??** summarizes key findings and discusses potential research directions.

II. Preliminaries and Problem Formulation

This section presents the fundamental concepts and theoretical foundations underlying the proposed methodology. The discussion encompasses neighborhood-preserving embedding techniques, deep autoencoder architectures, and a formal description of the missing data imputation problem addressed in this work.

A. Neighborhood-Preserving Embedding

Neighborhood-preserving embedding techniques aim to construct low-dimensional representations that maintain the intrinsic geometric structure of high-dimensional data. These methods operate under the manifold hypothesis, which posits that high-dimensional observations reside on or near a lower-dimensional manifold embedded within the ambient space. The fundamental objective involves learning a mapping function that transforms data points from the original high-dimensional space to a lower-dimensional latent space while preserving local neighborhood relationships.

The embedding process can be formulated as a transformation from the original space to the latent representation

$$\mathbf{z}_i = f(\mathbf{x}_i; \theta), \tag{1}$$

where $\mathbf{x}_i \in \mathbb{R}^D$ denotes the original high-dimensional data point, $\mathbf{z}_i \in \mathbb{R}^d$ represents the corresponding low-dimensional embedding with $d \ll D$, and $f(\cdot; \theta)$ signifies the embedding function parameterized by θ. The optimization objective focuses on maintaining local geometric relationships through the minimization of a neighborhood preservation loss

$$\mathcal{L}_{\text{embed}} = \sum_{i=1}^{N} \sum_{j \in \mathcal{N}(i)} w_{ij} \|\mathbf{z}_i - \mathbf{z}_j\|^2, \tag{2}$$

where $\mathcal{N}(i)$ indicates the neighborhood of point i and w_{ij} represents the affinity weight quantifying the similarity between points $\mathbf{x}_i$ and $\mathbf{x}_j$ in the original space. This formulation ensures that data points sharing high similarity in the original space remain proximate in the embedded representation, thereby preserving the intrinsic data manifold structure.

B. Deep Autoencoder

Deep autoencoders constitute a class of neural networks designed to learn compressed representations through unsupervised reconstruction. The architecture comprises two fundamental components: an encoder network $f_\theta : \mathcal{X} \to \mathcal{Z}$ that maps input data to a latent space, and a decoder network $g_\phi : \mathcal{Z} \to \mathcal{X}$ that reconstructs the original input from the latent representation. The forward propagation through the encoder involves a series of nonlinear transformations

$$\mathbf{z} = f_\theta(\mathbf{x}) = \sigma(\mathbf{W}^{(L)}\sigma(\cdots\sigma(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)})\cdots) + \mathbf{b}^{(L)}), \tag{3}$$

where $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ denote weight matrices and bias vectors at layer l, with $\sigma(\cdot)$ representing an element-wise activation function. The reconstruction is generated through a similar hierarchical process in the decoder. The network parameters are optimized by minimizing the discrepancy between inputs and reconstructions

$$\mathcal{L}_{\text{recon}}(\theta, \phi) = \frac{1}{N} \sum_{i=1}^{N} \ell(\mathbf{x}_i, g_\phi(f_\theta(\mathbf{x}_i))), \tag{4}$$

where $\ell(\cdot,\cdot)$ typically represents the mean squared error or cross-entropy loss. For incomplete datasets with missing values, the reconstruction loss is computed exclusively over observed entries, enabling the model to learn patterns from available data while developing imputation capabilities.

C. Problem Formulation

The problem under consideration involves multivariate time series data subject to arbitrary missing patterns. Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T]^\top \in \mathbb{R}^{T\times N}$ represent a time series with T observations across N variables. A binary mask matrix $\mathbf{M} \in \{0,1\}^{T\times N}$ indicates data availability, where $m_{t,n} = 1$ if element $x_{t,n}$ is observed and $m_{t,n} = 0$ otherwise. The observed data corresponds to the Hadamard product $\mathbf{X} \odot \mathbf{M}$.

The objective of missing data imputation involves estimating plausible values for unobserved entries, thereby recovering the complete matrix $\hat{\mathbf{X}}$. This task presents significant challenges due to the potential complexity of missing mechanisms, which may include missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR) scenarios. The proposed methodology aims to address these challenges through a unified framework that leverages both spatial correlations among variables and temporal dependencies across observations.

Successful imputation requires preserving the underlying data structure while accommodating irregular sampling patterns. The approach developed in this work simultaneously exploits local consistency in both spatial and temporal domains, employing a deep autoencoder architecture regularized by neighborhood-preserving constraints. This integration facilitates the learning of representations that remain robust to missing data patterns while maintaining fidelity to the intrinsic geometry of the complete data manifold.

III. Proposed Methodology

Contemporary approaches for multivariate time series imputation frequently encounter limitations in simultaneously addressing complex missing patterns while preserving both spatial correlations and temporal dependencies. The proposed AE-CS framework introduces a unified architecture that integrates coherent denoising with dual-domain neighborhood preservation, enabling robust imputation under diverse missing data mechanisms while maintaining structural integrity across variable and time dimensions.

A. Deep Autoencoder Architecture with Coherent Denoising

The AE-CS architecture employs a sophisticated deep autoencoder framework that systematically integrates coherent denoising mechanisms with spatio-temporal neighborhood preservation. The encoder network transforms input data into latent representations through hierarchical feature extraction, processing both observed values and missing indicators concurrently to distinguish between actual zeros and missing entries

$$\mathbf{Z}^{\text{orig}} = f_\theta(\mathbf{X}\odot\mathbf{M}, \mathbf{M}) = \sigma(\mathbf{W}^{(2)}\sigma(\mathbf{W}^{(1)}[\mathbf{X}\odot\mathbf{M}\|\mathbf{M}]+\mathbf{b}^{(1)})+\mathbf{b}^{(2)}), \tag{5}$$

where $\mathbf{W}^{(1)} \in \mathbb{R}^{2N\times d_h}$ and $\mathbf{W}^{(2)} \in \mathbb{R}^{d_h\times d_z}$ represent learnable weight matrices, $\mathbf{b}^{(1)} \in \mathbb{R}^{d_h}$ and $\mathbf{b}^{(2)} \in \mathbb{R}^{d_z}$ denote bias vectors, and $\sigma(\cdot)$ implements the ReLU activation function for introducing nonlinear transformations. The explicit concatenation of missing indicators enables the network to adaptively handle diverse missing patterns while preventing confounding during feature learning.

The coherent denoising component introduces stochastic robustness through multiple randomly corrupted versions of input data, functioning as an advanced regularization technique specifically designed for missing data scenarios. During each training iteration, K augmented masks undergo generation through systematic corruption processes

$$\mathbf{M}^{(k)} = \mathbf{M} \odot \mathbf{B}^{(k)}, \quad \mathbf{B}^{(k)}_{t,n} \sim \text{Bernoulli}(1 - p_{\text{drop}}), \tag{6}$$

where p_{drop} controls augmentation intensity and $\odot$ denotes element-wise multiplication. The consistency constraint ensures embeddings from corrupted inputs maintain proximity to original representations through minimization of adaptively weighted discrepancies

$$\mathcal{L}_{\text{consist}} = \frac{1}{K}\sum_{k=1}^{K} w^{(k)}\|f_\theta(\mathbf{X}\odot\mathbf{M}^{(k)}, \mathbf{M}^{(k)}) - f_\theta(\mathbf{X}\odot\mathbf{M}, \mathbf{M})\|_F^2, \tag{7}$$

where $w^{(k)} = \exp\left(-\frac{\|\mathbf{M}^{(k)}-\mathbf{M}\|_1}{\sigma_c^2}\right)$ provides adaptive weighting based on corruption severity. This formulation encourages learning of representations that remain stable under minor perturbations while adapting meaningfully to significant pattern variations, effectively enhancing model generalization across diverse missing mechanisms.

Spatial neighborhood preservation operates on individual time instances by identifying similar patterns across the variable dimension, leveraging correlated behaviors within temporal snapshots. For each time step t, the effective observation set comprises variables where $m_{t,n} = 1$, forming reduced-dimensional vectors used for affinity computation

$$w^{\text{space}}_{t,s} = \exp\left(-\frac{\|\mathbf{x}^{\text{obs}}_t - \mathbf{x}^{\text{obs}}_s\|_2^2}{\sigma_s^2}\right), \tag{8}$$

where σ_s controls spatial kernel bandwidth. The spatial manifold module subsequently computes refined embeddings through neighborhood aggregation

$$\mathbf{z}^{\text{space}}_t = \frac{\sum_{s\in\mathcal{N}^{\text{space}}_t} w^{\text{space}}_{t,s}\mathbf{z}^{\text{orig}}_s}{\sum_{s\in\mathcal{N}^{\text{space}}_t} w^{\text{space}}_{t,s}}, \tag{9}$$

with $\mathcal{N}^{\text{space}}_t$ containing indices of spatial neighbors identified through k-NN search in observed variable space. This spatial aggregation enables information propagation across time steps with similar variable expressions, effectively leveraging cross-sectional patterns for imputation.

Temporal neighborhood preservation focuses on individual variables across time, capturing dynamic patterns and evolutionary trends through similarity assessment in temporal domains. For variable n, observed time points form subsequences used for temporal affinity computation

$$w^{\text{time}}_{n,m} = \exp\left(-\frac{\|\mathbf{x}^{\text{obs}}_{:,n} - \mathbf{x}^{\text{obs}}_{:,m}\|_2^2}{\sigma_t^2}\right), \tag{10}$$

with temporal manifold processing generating refined embeddings through weighted aggregation

$$\mathbf{z}_n^{\text{time}} = \frac{\sum_{m\in\mathcal{N}_n^{\text{time}}} w_{n,m}^{\text{time}} \mathbf{z}_m^{\text{orig}}}{\sum_{m\in\mathcal{N}_n^{\text{time}}} w_{n,m}^{\text{time}}}, \tag{11}$$

where $\mathcal{N}_n^{\text{time}}$ represents temporal neighbors identified through k-NN search in observed temporal pattern space. This formulation captures inherent dynamics of time series data, ensuring variables exhibiting similar temporal behaviors maintain similar representations.

The adaptive fusion mechanism combines three complementary feature streams using learned attention weights that dynamically adjust based on input characteristics and missing patterns. Fusion coefficients undergo generation through gating network processing

$$\boldsymbol{\alpha} = \text{softmax}(\mathbf{W}_a[\bar{\mathbf{z}}^{\text{orig}} \| \bar{\mathbf{z}}^{\text{space}} \| \bar{\mathbf{z}}^{\text{time}} \| \rho] + \mathbf{b}_a), \tag{12}$$

where $\bar{\mathbf{z}}^{\cdot}$ denotes global average pooling, $\rho = 1 - \frac{1}{TN}\sum_{t,n} m_{t,n}$ represents overall missing rate, and $\mathbf{W}_a \in \mathbb{R}^{(3d_z+1)\times 3}$, $\mathbf{b}_a \in \mathbb{R}^3$ constitute learnable parameters. Fused embeddings subsequently undergo computation through weighted combination

$$\mathbf{Z}^{\text{fused}} = \alpha_1 \mathbf{Z}^{\text{orig}} + \alpha_2 \mathbf{Z}^{\text{space}} + \alpha_3 \mathbf{Z}^{\text{time}}, \tag{13}$$

enabling dynamic prioritization of different feature aspects based on data characteristics, with spatial features receiving higher weights for strong cross-sectional correlations and temporal features dominating for pronounced dynamic patterns.

The decoder network reconstructs complete data from fused representations through nonlinear transformations mapping latent embeddings to original data space

$$\hat{\mathbf{X}} = g_\phi(\mathbf{Z}^{\text{fused}}) = \mathbf{W}^{(4)}\sigma(\mathbf{W}^{(3)}\mathbf{Z}^{\text{fused}} + \mathbf{b}^{(3)}) + \mathbf{b}^{(4)}, \tag{14}$$

with $\mathbf{W}^{(3)} \in \mathbb{R}^{d_z\times d_h}$, $\mathbf{W}^{(4)} \in \mathbb{R}^{d_h\times N}$, and corresponding bias terms parameterizing decoding transformations. The complete optimization objective integrates all constraints through multi-task learning formulation

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + \lambda_1\mathcal{L}_{\text{consist}} + \lambda_2\mathcal{L}_{\text{space}} + \lambda_3\mathcal{L}_{\text{time}}, \tag{15}$$

where reconstruction loss $\mathcal{L}_{\text{recon}} = \frac{1}{\sum_{t,n} m_{t,n}}\|(\mathbf{X}-\hat{\mathbf{X}})\odot\mathbf{M}\|_F^2$ focuses exclusively on observed entries, spatial preservation loss $\mathcal{L}_{\text{space}} = \sum_{t=1}^{T}\sum_{s\in\mathcal{N}_t^{\text{space}}} w_{t,s}^{\text{space}}\|\mathbf{z}_t - \mathbf{z}_s\|_2^2$ enforces neighborhood consistency in spatial domains, temporal preservation loss $\mathcal{L}_{\text{time}} = \sum_{n=1}^{N}\sum_{m\in\mathcal{N}_n^{\text{time}}} w_{n,m}^{\text{time}}\|\mathbf{z}_n - \mathbf{z}_m\|_2^2$ maintains temporal pattern consistency, and $\lambda_1, \lambda_2, \lambda_3$ control relative importance of regularization terms. This comprehensive objective ensures simultaneous achievement of accurate reconstruction while preserving both spatial and temporal structures in learned representations.

Remark 1. *The dual neighborhood preservation strategy fundamentally diverges from conventional approaches by explicitly maintaining both cross-sectional and longitudinal relationships through separate but complementary mechanisms. This dual consideration proves particularly advantageous for multivariate time series with complex missing patterns, where information propagation must occur across both variable and time dimensions simultaneously, providing comprehensive characterization of underlying data structure.*

Remark 2. *The coherent denoising mechanism introduces specialized regularization that enhances model robustness against diverse missing patterns, functioning as data augmentation specifically tailored for missing data scenarios. By encouraging consistency between embeddings from original and corrupted inputs, the model develops increased resilience to variations in missingness patterns, ultimately improving generalization performance on test data with potentially different missing mechanisms.*

B. Imputation Framework Implementation

The complete AE-CS framework implements the architectural components through end-to-end trainable pipeline that jointly optimizes all elements rather than employing separate training stages. The training procedure iteratively refines model parameters through gradient-based optimization while maintaining delicate balance between reconstruction accuracy and structural preservation.

Algorithm 1: AE-CS Training Procedure

Input : Incomplete time series $\{\mathbf{X}_i\}_{i=1}^{B}$, mask matrices $\{\mathbf{M}_i\}_{i=1}^{B}$
Output: Optimized model parameters θ, ϕ

Initialize model parameters θ, ϕ;
while *training not converged* **do**
  foreach *mini-batch* $i = 1$ *to* B **do**
    $\mathbf{Z}_i^{\text{orig}} \leftarrow f_\theta(\mathbf{X}_i \odot \mathbf{M}_i, \mathbf{M}_i)$;
    for $k = 1$ *to* K **do**
      $\mathbf{M}_i^{(k)} \leftarrow \mathbf{M}_i \odot \text{Bernoulli}(1 - p_{\text{drop}})$;
      $\mathbf{Z}_i^{(k)} \leftarrow f_\theta(\mathbf{X}_i \odot \mathbf{M}_i^{(k)}, \mathbf{M}_i^{(k)})$;
    end
    Identify $\mathcal{N}_t^{\text{space}}$ via k-NN search on observed variables;
    Compute $\mathbf{Z}_i^{\text{space}}$ via weighted aggregation;
    Identify $\mathcal{N}_n^{\text{time}}$ via k-NN search on observed time points;
    Compute $\mathbf{Z}_i^{\text{time}}$ via weighted aggregation;
    Compute $\boldsymbol{\alpha}_i$ via gating network;
    $\mathbf{Z}_i^{\text{fused}} \leftarrow \alpha_{i,1}\mathbf{Z}_i^{\text{orig}} + \alpha_{i,2}\mathbf{Z}_i^{\text{space}} + \alpha_{i,3}\mathbf{Z}_i^{\text{time}}$;
    $\hat{\mathbf{X}}_i \leftarrow g_\phi(\mathbf{Z}_i^{\text{fused}})$;
    Compute $\mathcal{L}_{\text{recon}} \leftarrow \frac{1}{\sum \mathbf{M}_i}\|(\mathbf{X}_i - \hat{\mathbf{X}}_i)\odot\mathbf{M}_i\|_F^2$;
    Compute $\mathcal{L}_{\text{consist}} \leftarrow \frac{1}{K}\sum_{k=1}^{K} w^{(k)}\|\mathbf{Z}_i^{(k)} - \mathbf{Z}_i^{\text{orig}}\|_F^2$;
    Compute $\mathcal{L}_{\text{space}} \leftarrow \sum_{t=1}^{T}\sum_{s\in\mathcal{N}_t^{\text{space}}} w_{t,s}^{\text{space}}\|\mathbf{z}_{i,t} - \mathbf{z}_{i,s}\|_2^2$;
    Compute $\mathcal{L}_{\text{time}} \leftarrow \sum_{n=1}^{N}\sum_{m\in\mathcal{N}_n^{\text{time}}} w_{n,m}^{\text{time}}\|\mathbf{z}_{i,n} - \mathbf{z}_{i,m}\|_2^2$;
    $\mathcal{L}_{\text{total}} \leftarrow \mathcal{L}_{\text{recon}} + \lambda_1\mathcal{L}_{\text{consist}} + \lambda_2\mathcal{L}_{\text{space}} + \lambda_3\mathcal{L}_{\text{time}}$;
    Update θ, ϕ via gradient descent on $\mathcal{L}_{\text{total}}$;
  end
end

During inference, the computational graph simplifies considerably as coherent denoising branches become inactive, requiring only single forward pass through encoder, manifold modules, fusion mechanism, and decoder. This efficiency ensures practical deployment in real-world applications where computational resources may be constrained and rapid imputation is essential.

Algorithm 2: AE-CS Inference Procedure

Input : Incomplete time series $\mathbf{X}$, mask matrix $\mathbf{M}$
Output: Imputed time series $\hat{\mathbf{X}}$

$\mathbf{Z}^{\text{orig}} \leftarrow f_\theta(\mathbf{X} \odot \mathbf{M}, \mathbf{M})$;
Identify $\mathcal{N}_t^{\text{space}}$ via k-NN search on observed variables;
Compute $\mathbf{Z}^{\text{space}}$ via weighted aggregation;
Identify $\mathcal{N}_n^{\text{time}}$ via k-NN search on observed time points;
Compute $\mathbf{Z}^{\text{time}}$ via weighted aggregation;
Compute $\boldsymbol{\alpha}$ via gating network;
$\mathbf{Z}^{\text{fused}} \leftarrow \alpha_1 \mathbf{Z}^{\text{orig}} + \alpha_2 \mathbf{Z}^{\text{space}} + \alpha_3 \mathbf{Z}^{\text{time}}$;
$\hat{\mathbf{X}} \leftarrow g_\phi(\mathbf{Z}^{\text{fused}})$;

The neighborhood identification steps employ efficient k-nearest neighbor search with optimized data structures, achieving computational complexity of $O(TN \log k)$ for spatial neighborhoods and $O(NT \log k)$ for temporal neighborhoods through utilization of ball trees and spatial partitioning techniques. This efficient implementation ensures scalability to large multivariate time series datasets commonly encountered in practical applications, with overall computational complexity remaining linear with respect to data size.

The training process incorporates several strategic considerations to enhance stability and performance, including gradient clipping to prevent explosion, learning rate scheduling to facilitate convergence, and early stopping based on validation performance to prevent overfitting. The coherent denoising component operates exclusively during training, serving as regularizer that improves generalization without affecting inference efficiency. The neighborhood size parameters typically undergo cross-validation to determine optimal values that balance locality preservation against over-smoothing, with larger neighborhoods generally preferred for datasets exhibiting global structures and smaller neighborhoods suitable for capturing fine-grained local patterns.

The proposed methodology demonstrates particular effectiveness for datasets exhibiting both spatial correlations among variables and temporal dependencies across observations, with dual neighborhood preservation mechanism enabling comprehensive pattern capture. The adaptive fusion weights provide interpretable insights into data characteristics, with spatial dominance indicating strong cross-sectional relationships and temporal dominance reflecting pronounced dynamic patterns. The coherent denoising component enhances robustness against diverse missing mechanisms, making the approach suitable for real-world scenarios where missing patterns may differ between training and deployment environments.

IV. Experimental Verification

In this section, the effectiveness of the proposed explainable FD framework is validated on the widely recognized continuous stirred tank reactor (CSTR) simulation benchmark.

V. Conclusions

References

[1] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*. John Wiley & Sons, 2019.
[2] D. J. Stekhoven and P. Bühlmann, "Missforest—non-parametric missing value imputation for mixed-type data," *Bioinformatics*, vol. 28, no. 1, pp. 112–118, 2012.
[3] J. Yoon, J. Jordon, and M. Schaar, "Gain: Missing data imputation using generative adversarial nets," *arXiv preprint arXiv:1806.02920*, 2018.

Zhuofu Pan received the B.Eng. and M.Eng. degrees from the School of Civil Engineering, Changsha University of Science & Technology and Central South University, Changsha, China, in 2014 and 2017, respectively. He is currently working toward the Ph.D. degree in the School of Automation, Central South University, Changsha, China.