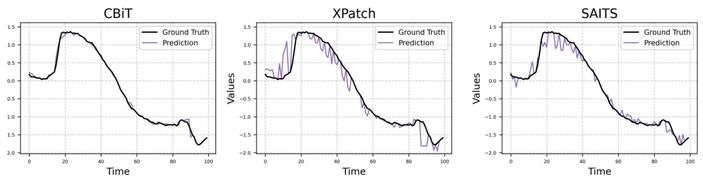
Rethinking Time-Series Imputation as Conditional Inference along Temporal Evolution

Yu Fan[1], Yang Yang[1], Yufan Guo[1], Huazhong Yang[1], Pengjun Wang[1]
[1] Department of Electronic Engineering, Tsinghua University, China

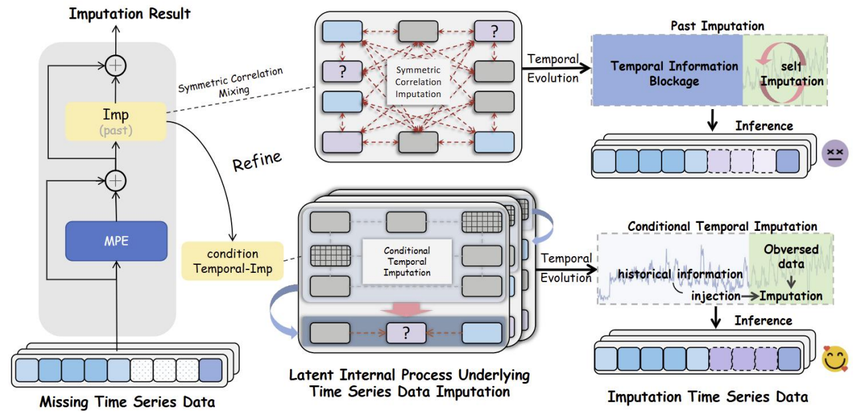
Abstract



Visualization of imputation results on the Power dataset under a 50% missing rate. The black line denotes the ground truth, while the colored lines correspond to the imputed values produced by SAITS, XPatch, and the proposed CBiT model.

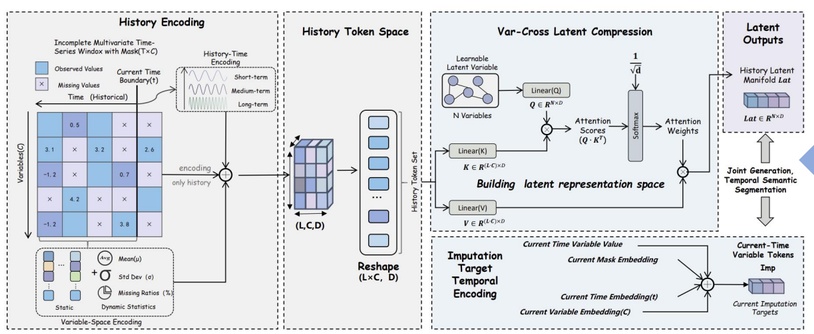
Real-world time-series data often suffer from missing observations, hindering long-range temporal modeling. However, most existing imputation methods formulate imputation as conditional reconstruction over limited context, which restricts temporal information propagation and fails to explicitly model temporal evolution. To overcome this limitation, we propose the Conditional Temporal Inference Paradigm (CTIP), which formulates time-series imputation as conditional inference along temporal evolution. Under this paradigm, we introduce CBiT, which leverages a history compression mechanism to encode long-range history into a compact latent space for history-conditioned temporal imputation.

In addition, we adopt a partitioned modeling strategy that distinguishes historical context and temporal imputation targets with only linear-time complexity. Extensive experiments on multiple public benchmarks show that CBiT improves imputation accuracy by reducing Masked MAE and Masked RMSE by 27.3% and 18.6%, respectively, across different missing rates.

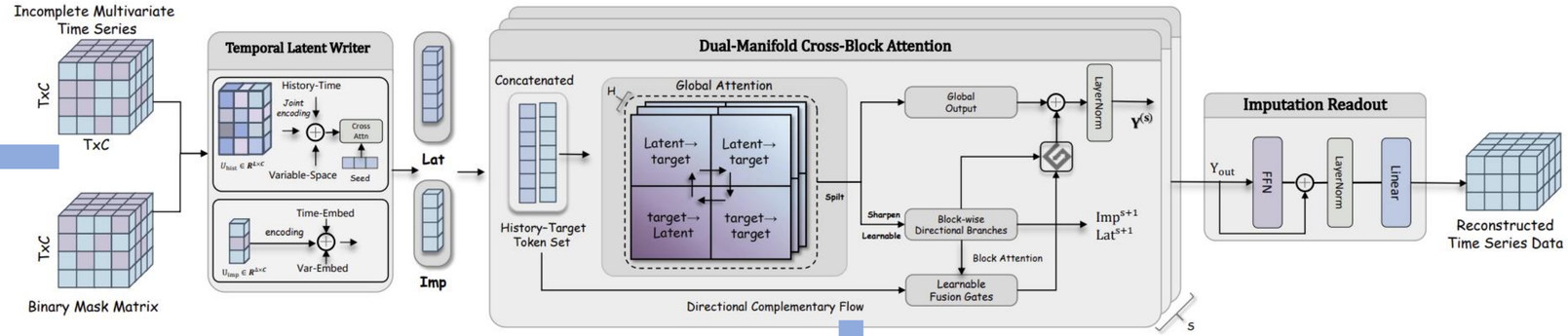


Conceptual illustration of the proposed CTIP paradigm in contrast to conventional local reconstruction-based imputation approaches.

Temporal Latent Writer architecture



The CBiT Architecture



Theoretical Motivation

1. Conditional risk over missing entries

$$R(\theta \mid I) = \mathbb{E}_{X_{\text{imp}}^{(m=1)} \sim p(\cdot \mid I)} \left[\ell_{\Omega(M)} \left(X_{\text{imp}}^{(m=1)}, f_\theta(I) \right) \right].$$

2. Non-degenerate historical information

$$p_\theta \left(X_{\text{imp}}^{(m=1)} \mid X_{\text{hist}}, X_{\text{imp}}^{(m=0)}, M, \Delta \right) \neq p_\theta \left(X_{\text{imp}}^{(m=1)} \mid X_{\text{imp}}^{(m=0)}, M, \Delta \right)$$

3. Local imputation marginalizes historical context

$$p \left(X_{\text{imp}}^{(m=1)} \mid X_{\text{imp}}^{(m=0)}, M, \Delta \right) = \int p \left(X_{\text{imp}}^{(m=1)} \mid X_{\text{hist}}, X_{\text{imp}}^{(m=0)}, M, \Delta \right) p \left(X_{\text{hist}} \mid X_{\text{imp}}^{(m=0)}, M, \Delta \right) dX_{\text{hist}}$$

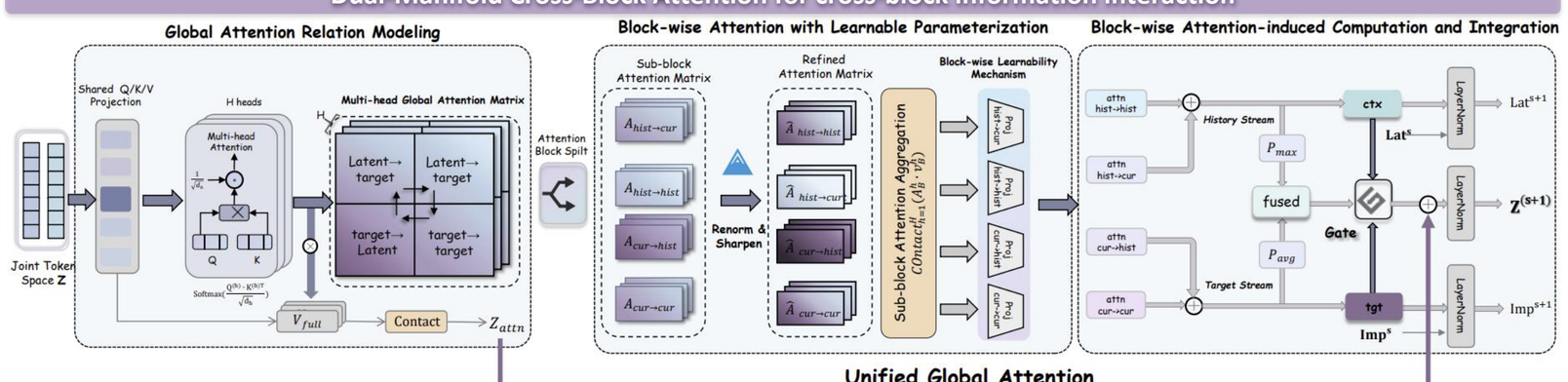
4. Full CTIP conditioning

$$I_{\text{full}} = \left\{ X_{\text{hist}}, X_{\text{imp}}^{(m=0)}, M, \Delta \right\}, \qquad I_{\text{loc}} = \left\{ X_{\text{imp}}^{(m=0)}, M, \Delta \right\}$$

5. Risk comparison between full and local conditioning

$$\inf_{f_\theta} R(\theta \mid I_{\text{full}}) \leq \inf_{f_\theta} R(\theta \mid I_{\text{loc}})$$

Dual-Manifold Cross-Block Attention for cross-block information interaction



Experimental Setup & Results

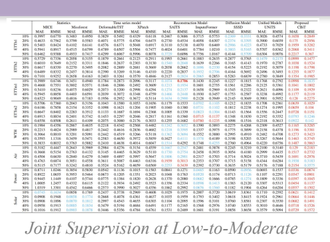
Evaluation Metrics and Training Objectives

$$\text{Masked-MAE} = \frac{\sum_{t,c} M_{t,c} \left| \hat{X}_{t,c} - X_{t,c} \right|}{\sum_{t,c} M_{t,c}},$$

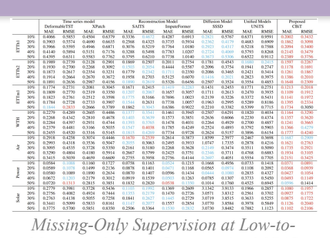
$$\text{Masked-RMSE} = \sqrt{\frac{\sum_{t,c} M_{t,c} \left(\hat{X}_{t,c} - X_{t,c} \right)^2}{\sum_{t,c} M_{t,c}}}$$

Training Objectives:

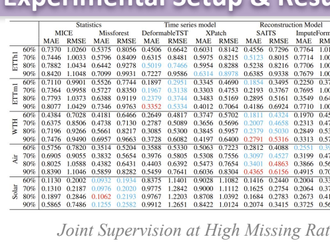
- **Missing-only Supervision:** $\mathcal{L}_{\text{miss}} = \mathcal{L}_{\text{masked}}$
- **Joint Supervision:** $\mathcal{L}_{\text{joint}} = \mathcal{L}_{\text{masked}} + \lambda \mathcal{L}_{\text{obs}}$



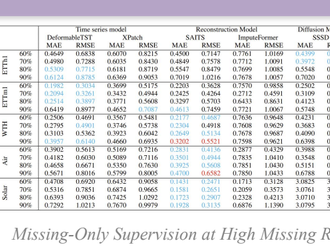
Joint Supervision at Low-to-Moderate Missing Rates under MCAR



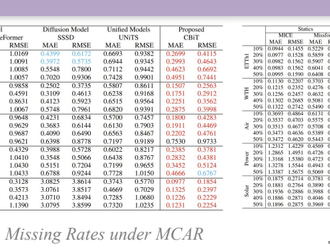
Missing-Only Supervision at Low-to-Moderate Missing Rates under MCAR



Joint Supervision at High Missing Rates under MCAR



Missing-Only Supervision at High Missing Rates under MCAR



Joint Supervision under MAR