Glocal Information Bottleneck for Time Series Imputation

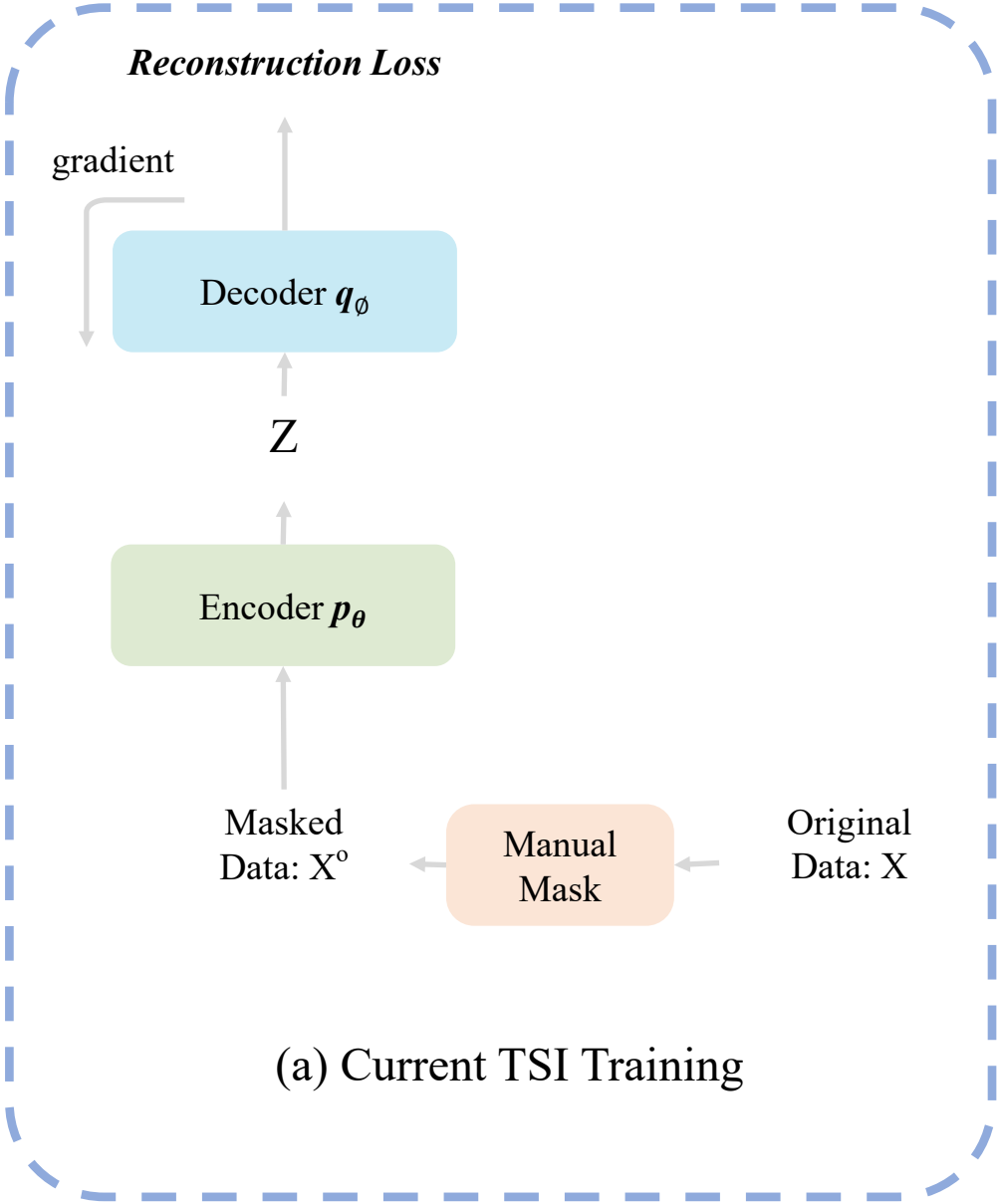
NeurIPS 2025

Jie Yang, Kexin Zhang, Guibin Zhang, Philip S. Yu, Kaize Ding

Current Time Series Imputation

Goal: Reconstruct the Missing Values

Encoder-Decoder Architecture



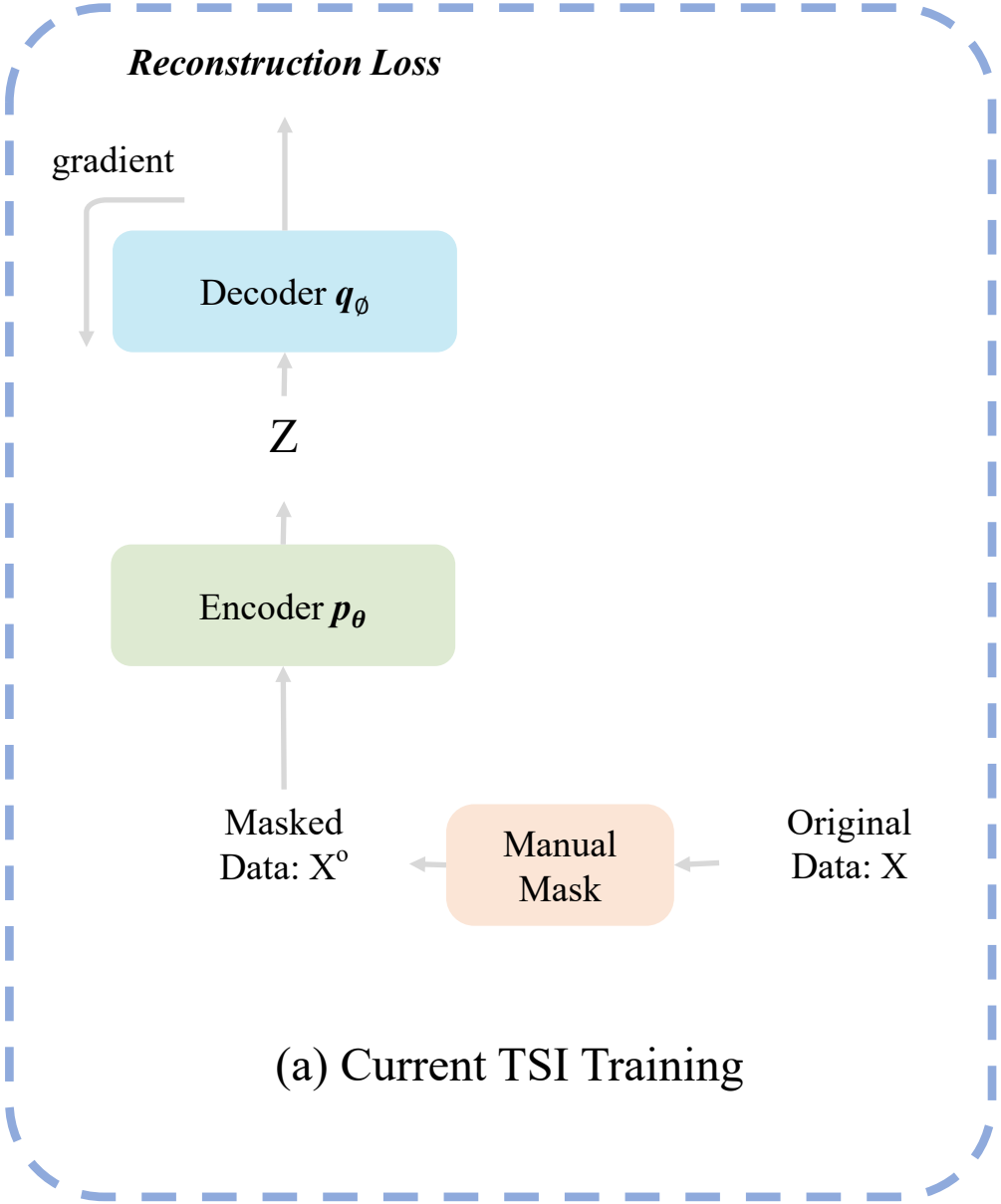
Current Time Series Imputation

Training Process:

- **Randomly masking** observed values to simulate missingness. The goal is to **learn the global data distribution from corrupted observations**, enabling the model to reconstruct masked values during training

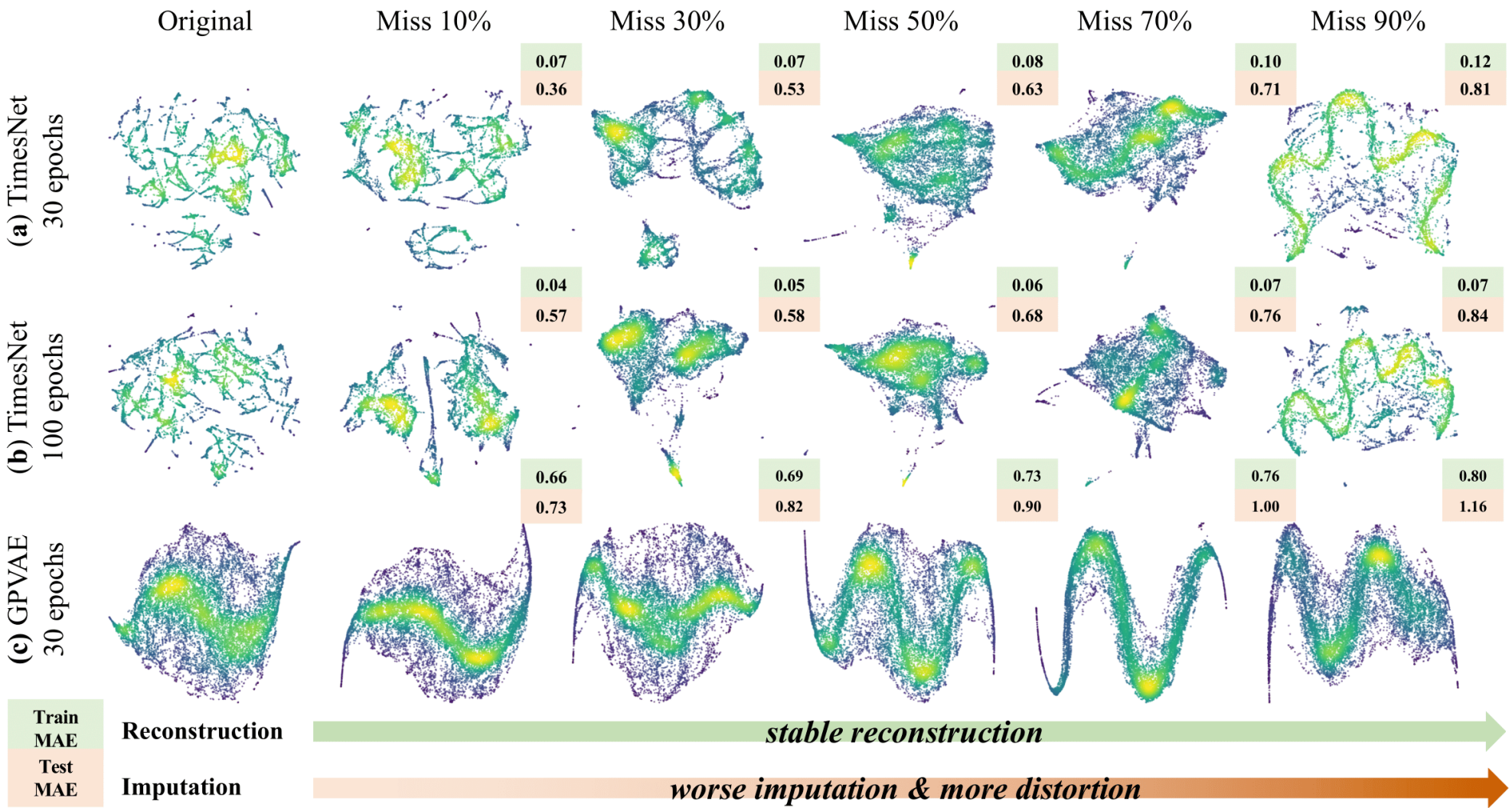
Inference Process:

- Serve as a **conditional generative model** with partially observed data to impute missing values



(a) Current TSI Training

Optimization Dilemma

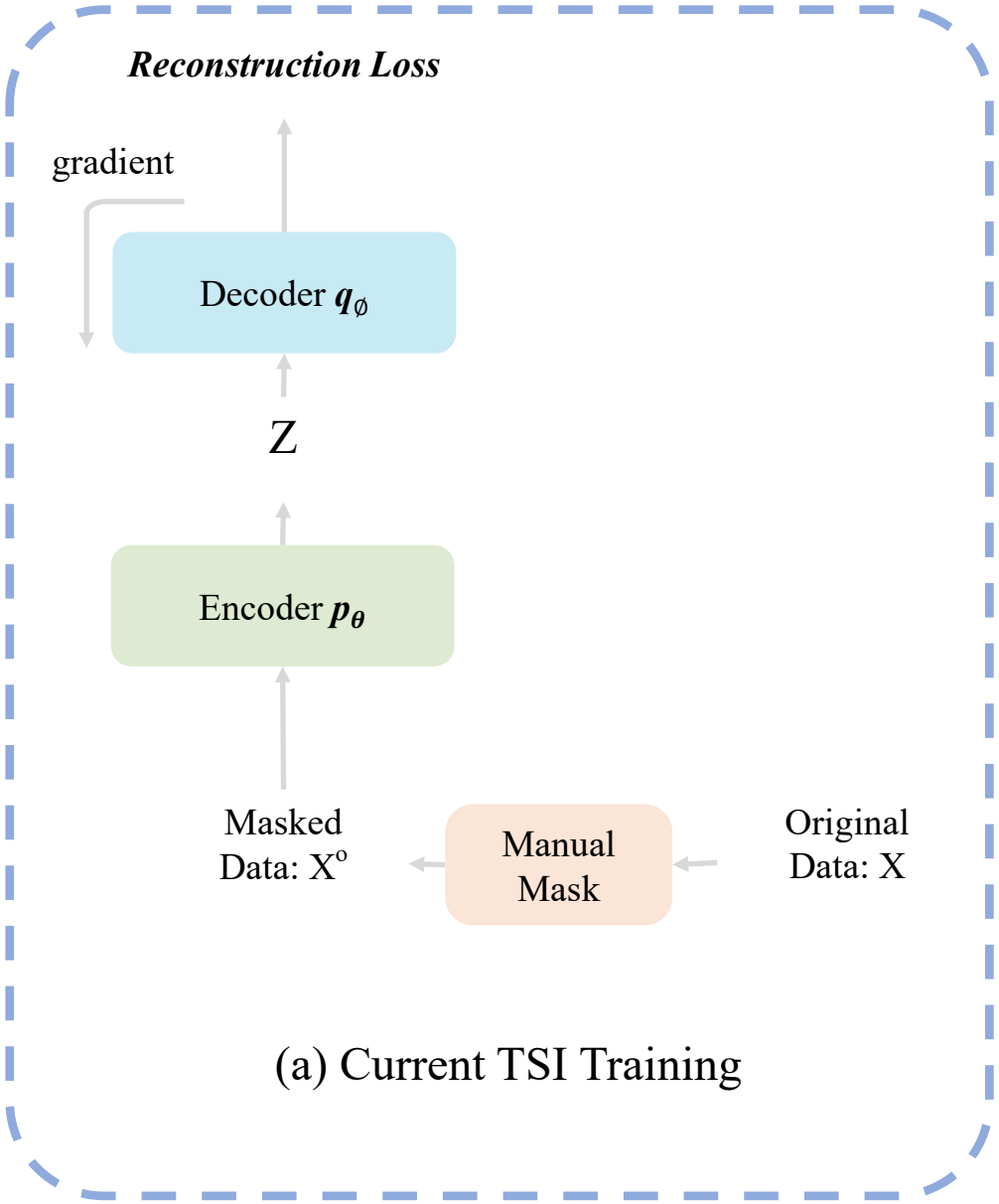


Optimization Dilemma

- Low training loss does not necessarily imply good imputation.

- Well-aligned representations are strongly related to good imputation.

Optimization Dilemma



(a) Current TSI Training

$$\mathbb{E}_{p(x)}\left[||x - \hat{x}||^2\right]$$

Focus on **local numerical accuracy** at each timestamp makes these models **sensitive** to temporal noise

Core Motivation

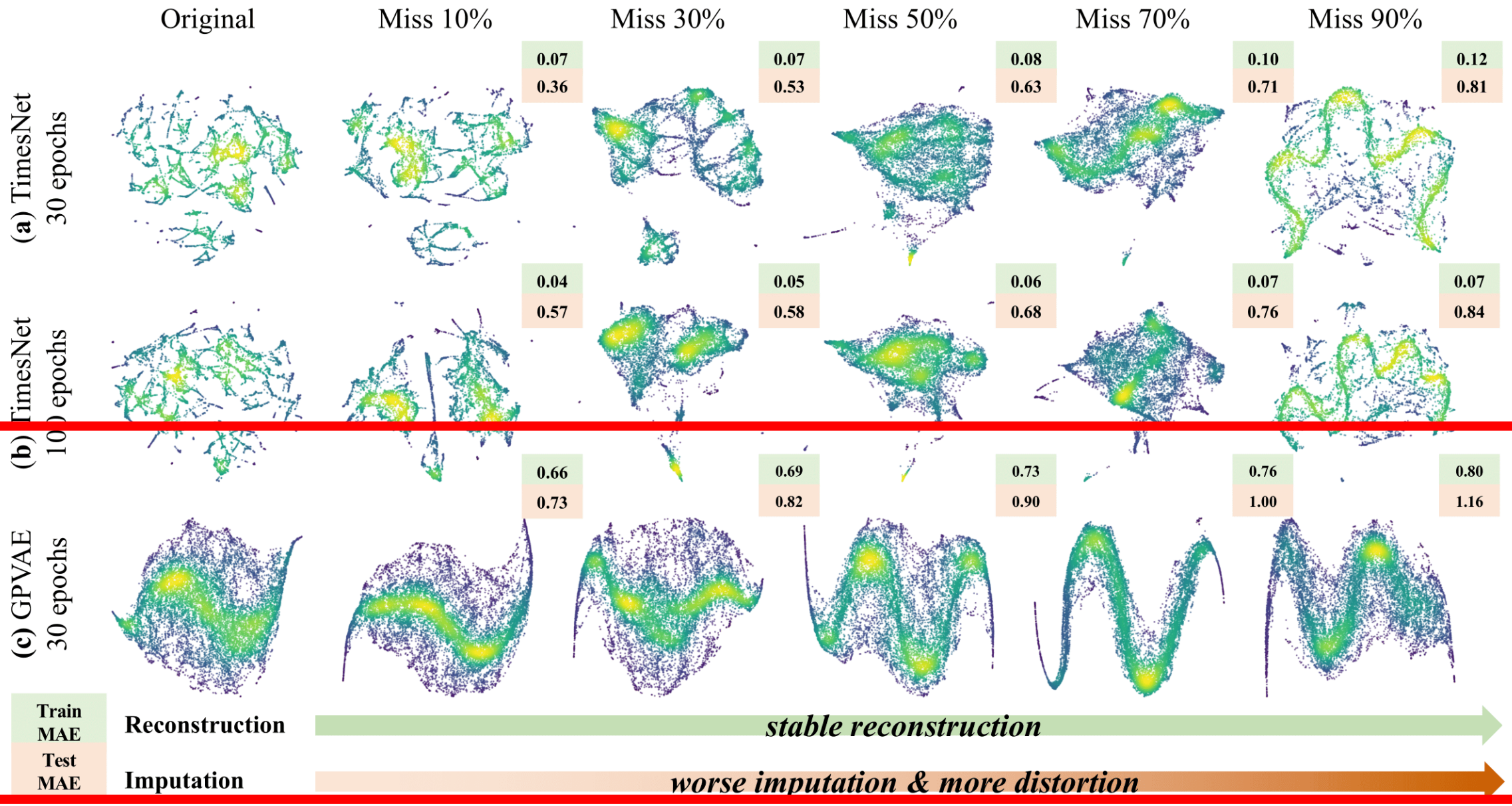
Can we design a training paradigm that encourages TSI models to capture both global and local information from incomplete data, without overfitting to noise?

Information Bottleneck

$$\min \; [I(Z^{\mathrm{IB}}; X^{\mathrm{IB}}) - \beta \cdot I(Y^{\mathrm{IB}}; Z^{\mathrm{IB}})]$$

$$\min_{\theta,\, \phi} [I_\theta(Z; X^{\mathrm{o}}) - \beta \cdot I_\phi(X; Z)]$$

Information Bottleneck



Information Bottleneck

- **Local:**

$$I(X;Z) \geq \mathbb{E}_{p(x,z)}[\log q_\phi(x|z)] \stackrel{\text{def}}{=} -\mathcal{L}^{\phi}_{\text{Loc}}$$

- **Global:**

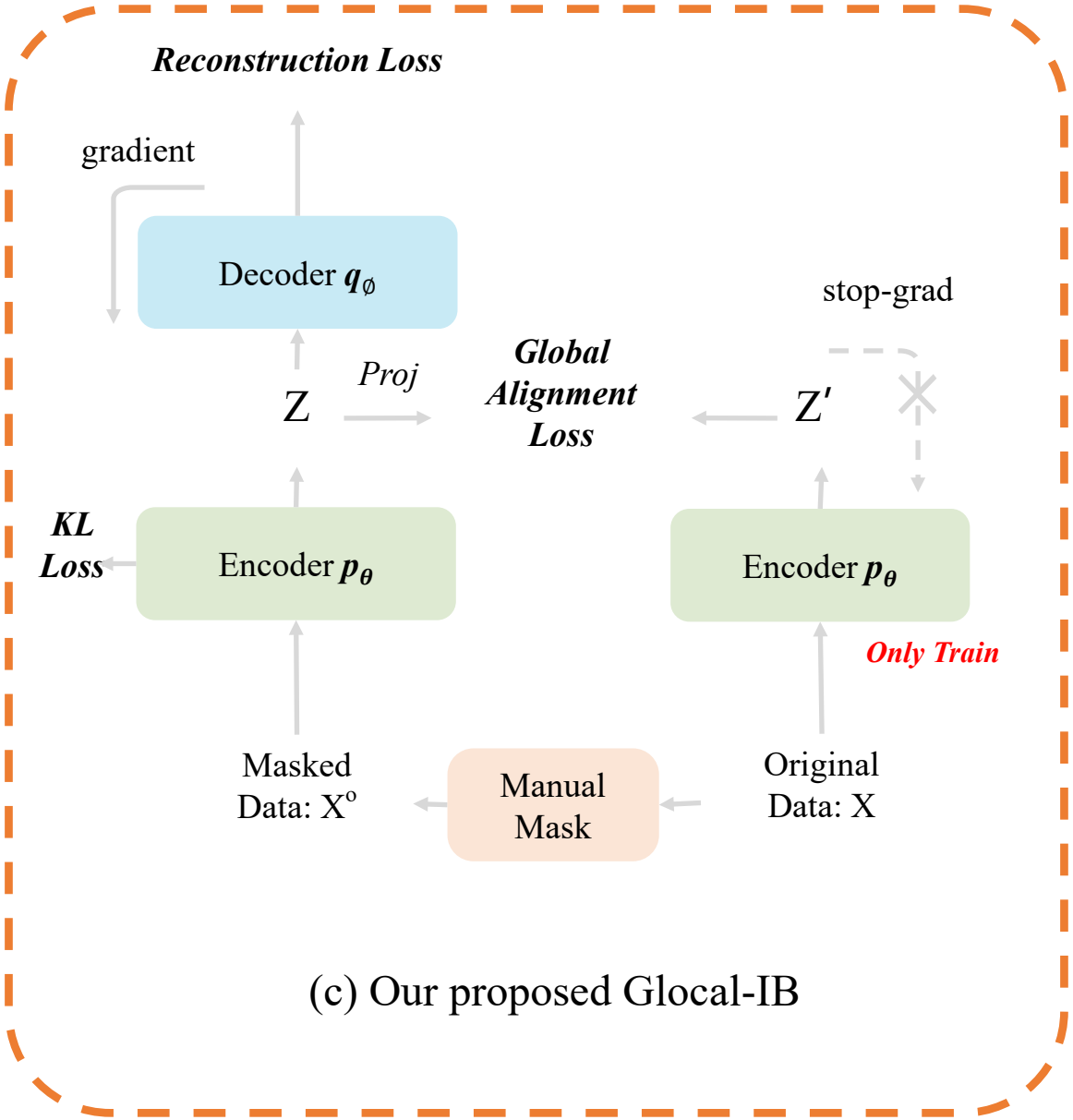
$$I(X;Z) \geq \mathbb{E}_{p(x,z)}\left[\log\left(\frac{f(x,z)}{f(x,z)+\sum\limits_{x_j \in X^{\text{neg}}} f(x_j,z)}\right)\right] \stackrel{\text{def}}{=} -\mathcal{L}^{\phi}_{\text{Glo\_1}}$$

$$I(X;Z) \geq \mathbb{E}_{p(x,z)}\left[\exp\left(\text{proj}(z)^\top \cdot \text{enc}(x)\right)\right] \stackrel{\text{def}}{=} \mathcal{L}^{\phi}_{\text{Glo\_2}}$$

Overall Objective

$$\min\left[\alpha \cdot \mathcal{L}^{\theta}_{\text{Reg}} + \beta_1 \cdot \mathcal{L}^{\phi}_{\text{Loc}} + \beta_2 \cdot \mathcal{L}^{\phi}_{\text{Glo}}\right]$$

Methodology



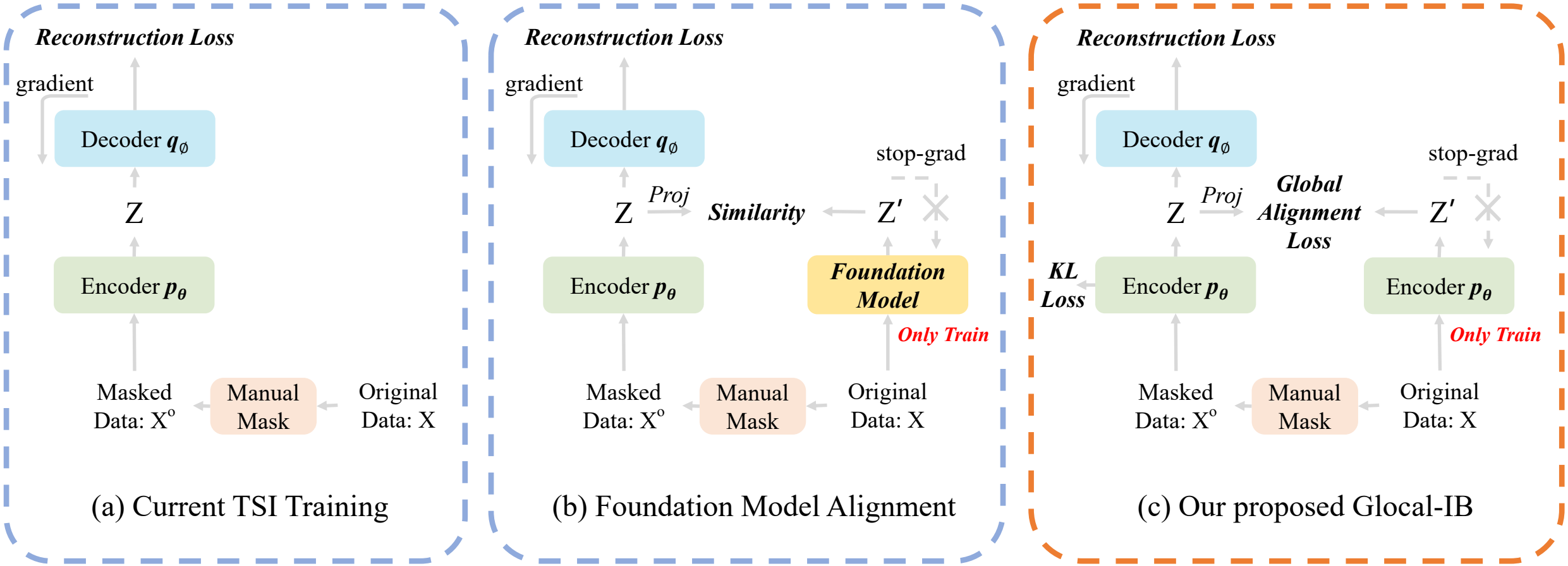
(c) Our proposed Glocal-IB

Experiments

Table 1: Imputation performance on 9 datasets (average MAE and MSE across 10% to 90% missing rates). Best is **bold** and second-best is underlined. We use OOM to denote out of memory.

Models	IMP	Ours		SAITS		Transformer		DLinear		TimesNet		FreTS		PatchTST		iTransformer		GPVAE		TimeMixer	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
ETTh1	38%	**0.283**	**0.197**	0.402	0.376	0.399	0.373	0.390	0.316	0.602	0.702	0.446	0.394	0.624	0.780	0.441	0.406	0.731	0.928	0.678	0.886
ETTh2	40%	**0.249**	**0.132**	0.340	0.256	0.307	0.218	0.352	0.243	0.800	1.140	0.434	0.370	0.525	0.575	0.413	0.321	0.686	0.769	0.529	0.535
ETTm1	28%	**0.157**	**0.069**	0.206	0.099	0.202	0.096	0.284	0.172	0.789	1.087	0.310	0.195	0.294	0.188	0.315	0.208	0.588	0.627	0.359	0.242
ETTm2	25%	**0.157**	**0.069**	0.206	0.099	0.202	0.096	0.284	0.172	0.789	1.087	0.310	0.195	0.294	0.188	0.315	0.208	0.588	0.627	0.359	0.242
Beijing Air	7%	**0.223**	**0.320**	0.256	0.353	0.268	0.375	0.279	0.338	0.264	0.370	0.289	0.349	0.365	0.472	0.365	0.473	0.380	0.483	0.448	0.628
PEMS-Traffic	6%	**0.318**	**0.630**	0.336	0.674	0.355	0.695	0.401	0.696	0.336	0.683	0.441	0.745	0.472	0.870	OOM	OOM	0.383	0.680	0.528	1.018
Electricity	8%	**0.372**	**0.296**	0.397	0.343	0.410	0.358	0.483	0.433	0.390	0.322	0.544	0.534	0.676	0.783	0.440	0.363	0.443	0.394	0.648	0.724
Weather	34%	**0.096**	**0.056**	0.136	0.093	0.139	0.091	0.161	0.089	0.262	0.211	0.167	0.085	0.188	0.111	0.182	0.102	0.278	0.195	0.239	0.180
Metr-LA	17%	**0.267**	**0.293**	0.301	0.392	0.306	0.387	0.387	0.412	0.289	0.354	0.414	0.453	0.423	0.544	0.427	0.477	0.420	0.463	0.607	1.000

Experiments



Experiments

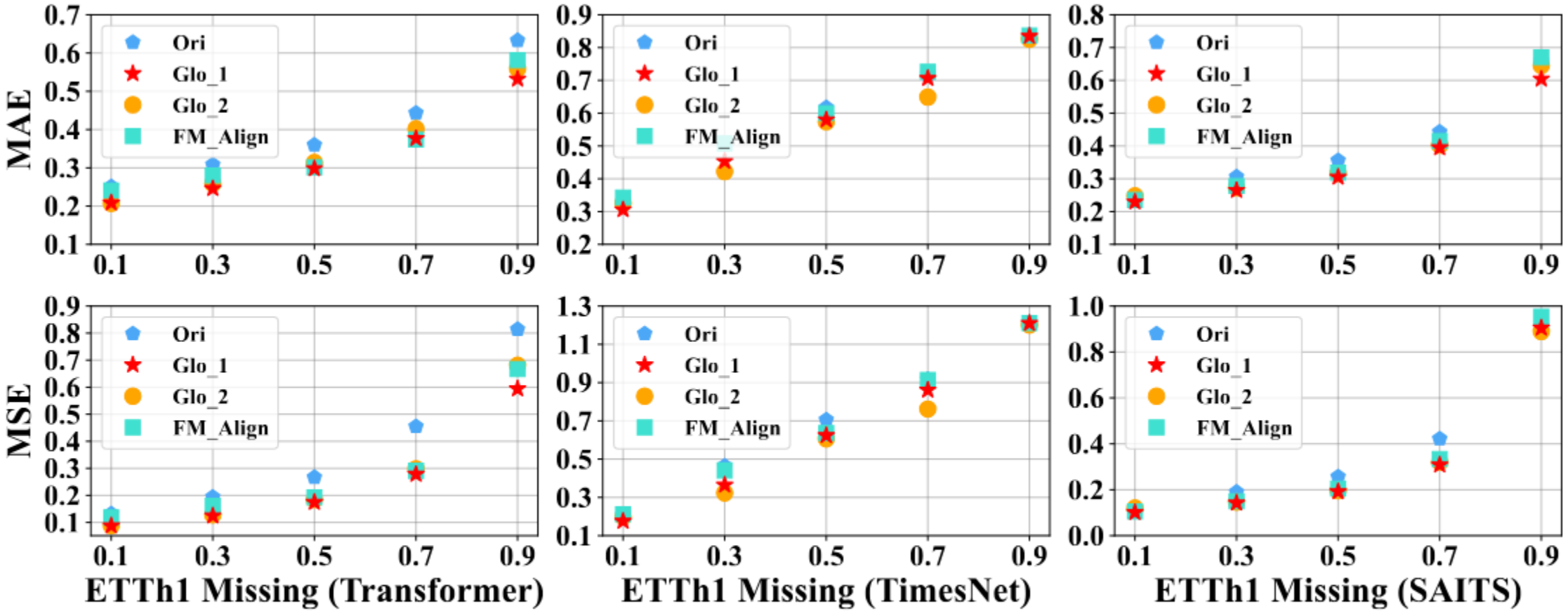


Figure 3: Imputation performance on the ETTh1 dataset of Transformer, TimesNet, and SAITS with four different training methods.

Experiments

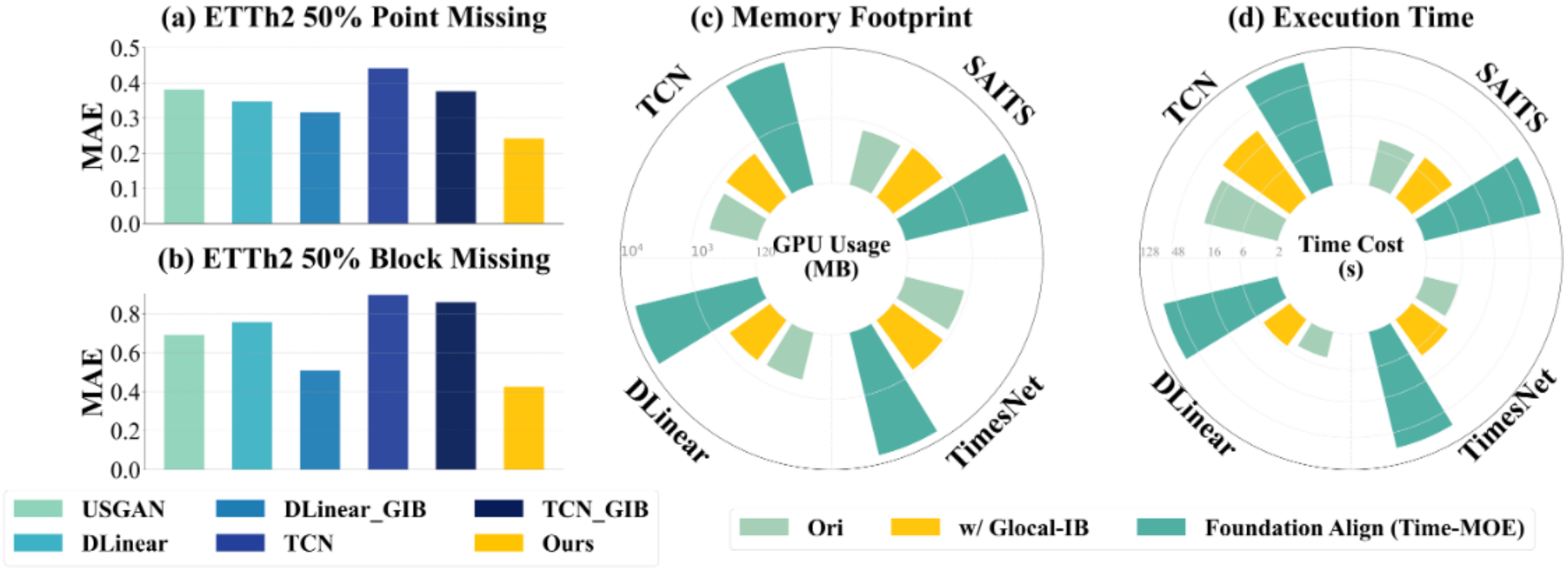


Figure 5: Different Missing Pattern Imputation and Efficiency results. **(a, b)**: Comparison of imputation performance on the ETTh2 dataset with 50% Point and Block missing rates. Additional results for various missing rates are presented in Appendix D.3. **(c, d)**: Efficiency comparison of four representative models on the ETTh1 dataset, evaluating the original models against their variants with Glocal-IB and foundation model alignment (Time-MoE). The radial axes are on a **logarithmic** scale.

Experiments

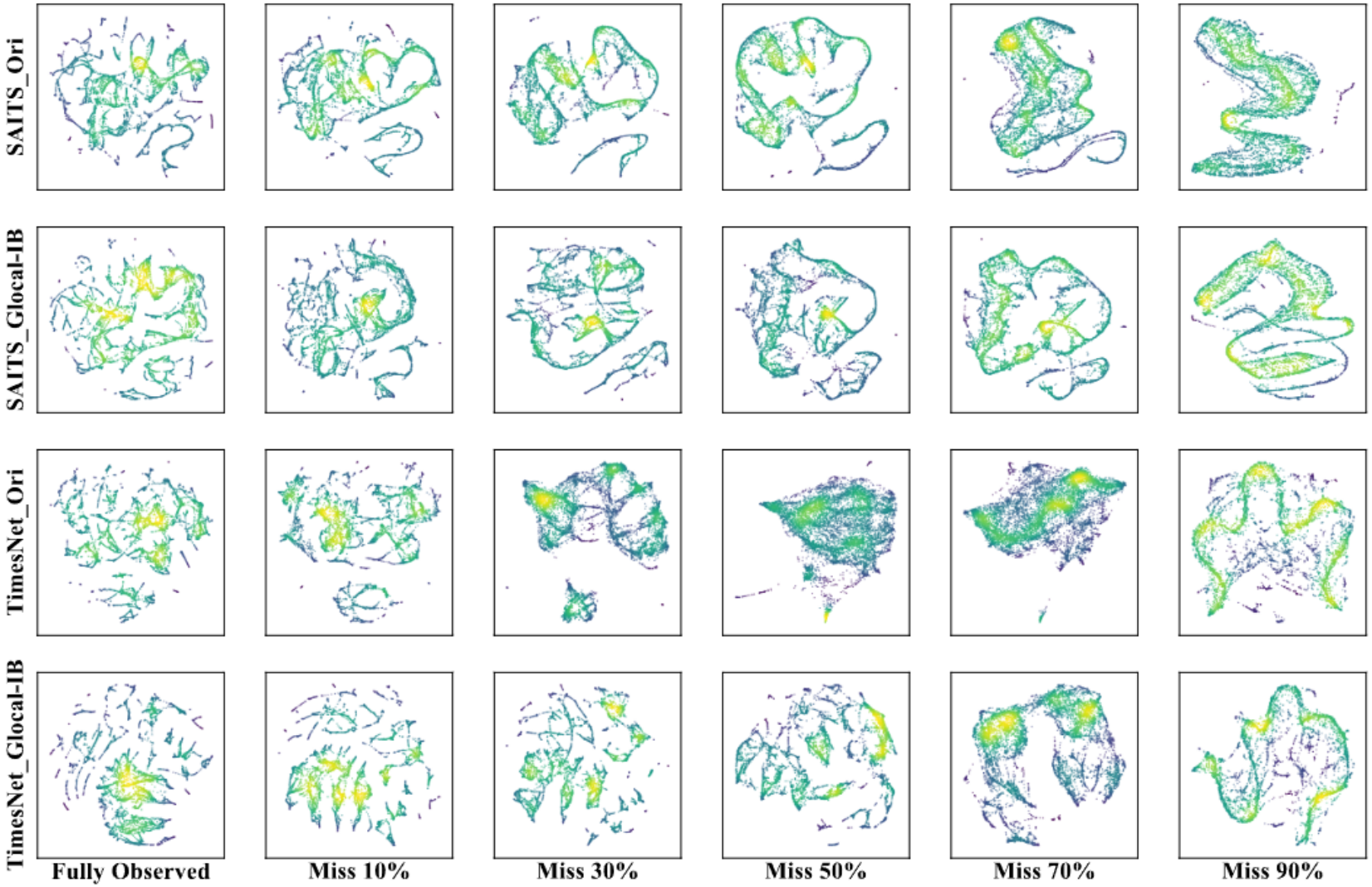


Figure 9: Latent space comparison of TimesNet and SAITS with Glocal-IB against their original implementations.

Experiments

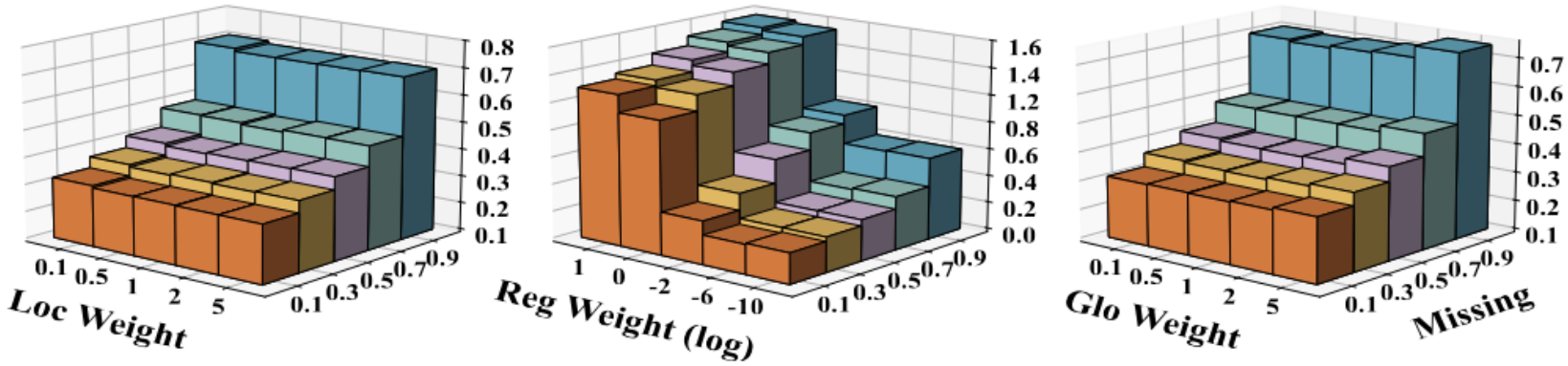
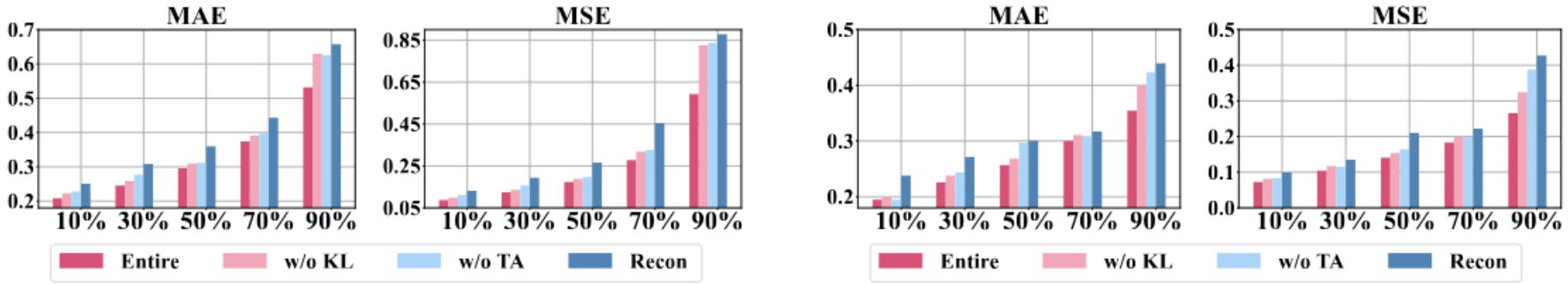


Figure 6: Hyperparameter sensitivity experiment results. More results are in Appendix D.4.



(a) Ablation study results on ETTh1 dataset

(b) Ablation study results on ETTh2 dataset

Figure 7: Visualization of ablation study results on ETTh1 dataset. More results are in Appendix D.4.

Glocal-IB

Thanks For Your Attention

Jie Yang (杨杰)

Code: https://github.com/Muyiiiii/NeurIPS-25-Glocal-IB

Paper: https://arxiv.org/abs/2510.04910