ST-LoRA: Low-rank Adaptation for Spatio-Temporal Forecasting

Weilin Ruan, Wei Chen, Xilin Dang, Jianxiang Zhou, Weichuang Li, Xu Liu, Yuxuan Liang

香港科技大学（广州）
THE HONG KONG
UNIVERSITY OF SCIENCE AND
TECHNOLOGY (GUANGZHOU)

Presenter: Weilin Ruan
rwlinno@gmail.com +86 136 725 46637

Table of Content

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

香港中文大學
The Chinese University of Hong Kong

NUS National University of Singapore

CITY ND

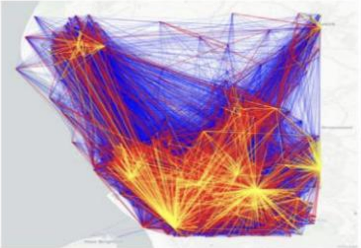
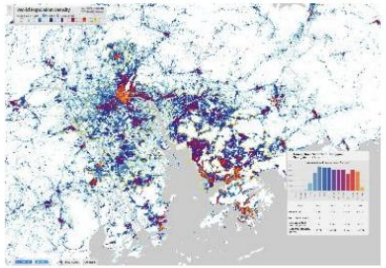
地理信息科学与技术全国重点实验室
State Key Laboratory of Resources and Environment Information System

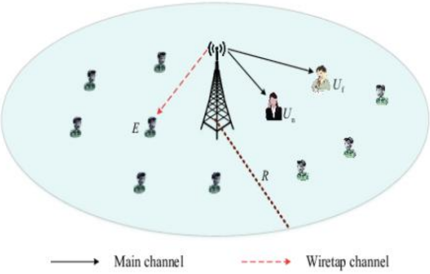
Spatio-temporal Forecasting

Given spatio-temporal graph $\mathcal{G}$ and input tensor X, we aim to learn function $\mathcal{F}$:

$$[X^{(t-s+1)},\ldots,X^{(t)};\mathcal{G}] \xrightarrow{\mathcal{F}(\cdot)} [X^{(t+1)},\ldots,X^{(t+h)}].$$

- **Vast Spatio-Temporal Scenarios**
- **Complexity of the Data**
- **Potential Application Value**

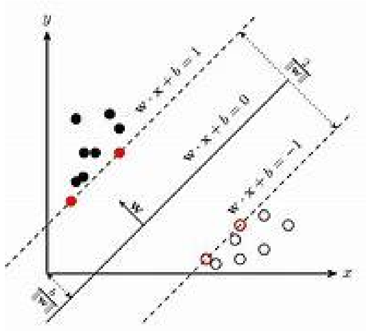


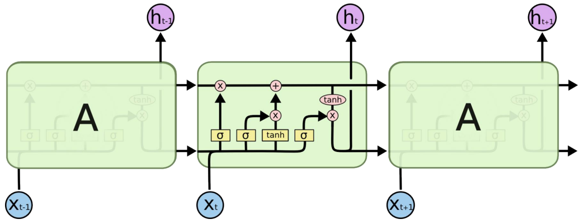


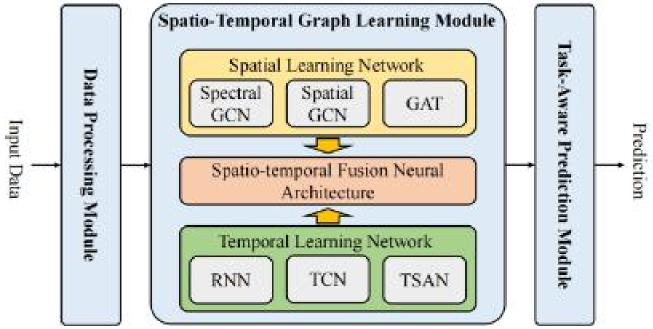
Introduction - Background

Development of Multivariate Time Series Forecasting Methods

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$







MileStones

1ST TERM

Prediction using traditional **statistical** methods such as VAR, HI, and ARIMA, but ignoring spatial dependence

2ND TERM

Using **machine learning** models such as SVMs to capture features, but it is difficult to deal with complex nonlinear relationships.

3RD TERM

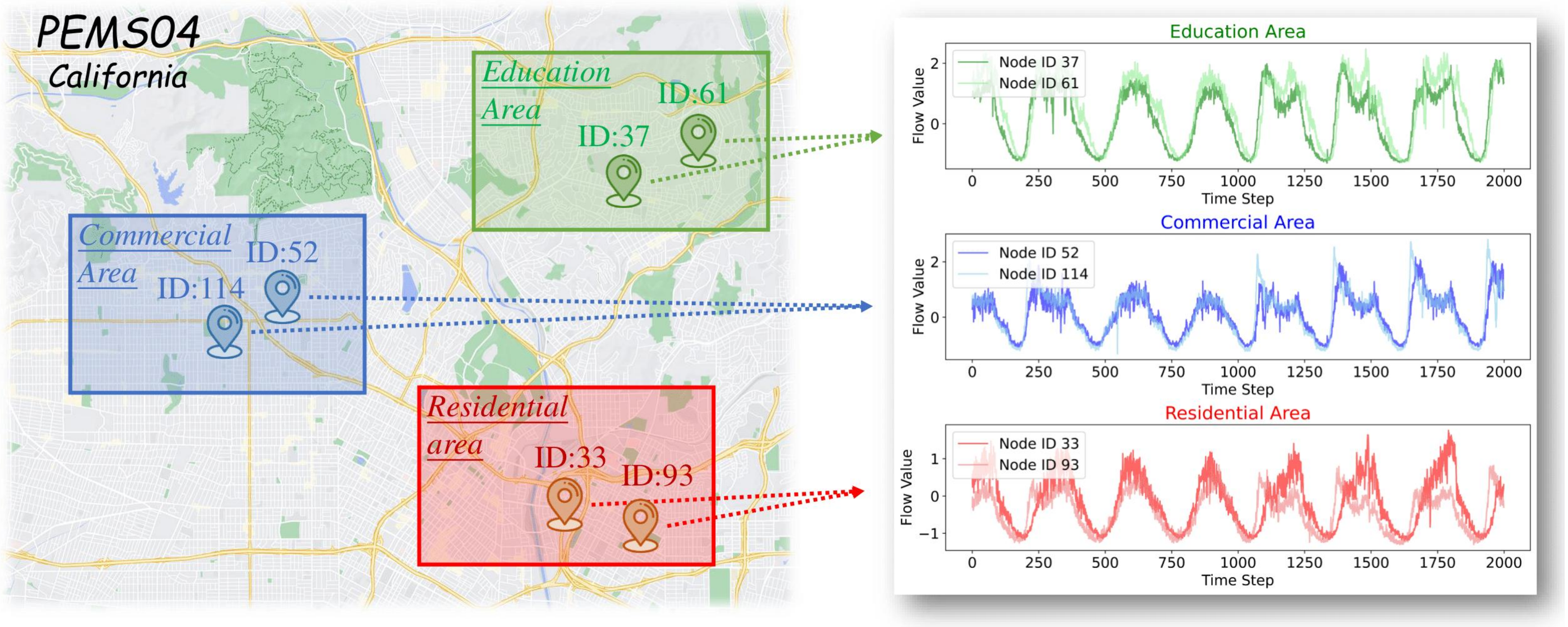
Deep learning models such as CNNs and LSTMs used to capture spatio-temporal dependencies a decade ago.

4TH TERM

After 2017, **STGNNs** model performance continues to break through and combine with cutting-edge technologies such as big language modeling.

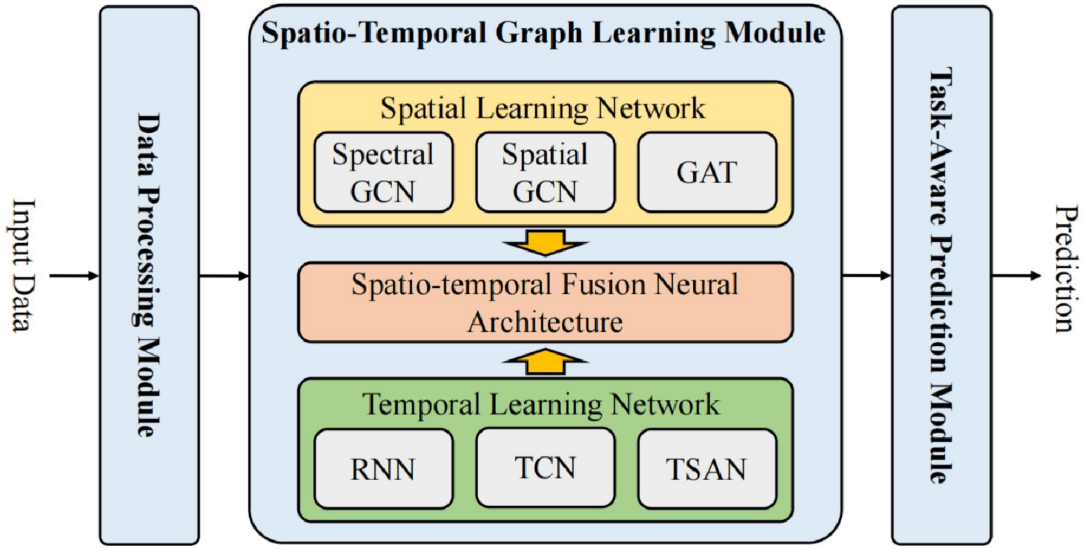
ST-LoRA - Challenges

Node Heterogeneity: The diverse differences and unique individual behavioral patterns that exist among different nodes stem from the complex diversity of the real world. Specifically, each eigenvalue of nodes has completely different spatial topologies and relationships in different scenarios, as well as completely different temporal patterns such as periodicity and trends.



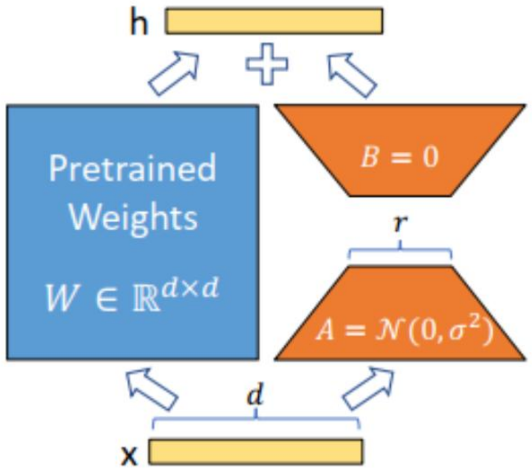
Spatio-temporal Forecasting

Spatio-Temporal Graph Neural Networks (STGNNs) have emerged as powerful tools for prediction tasks. However, these methods often introduce substantial computational overhead and model complexity.



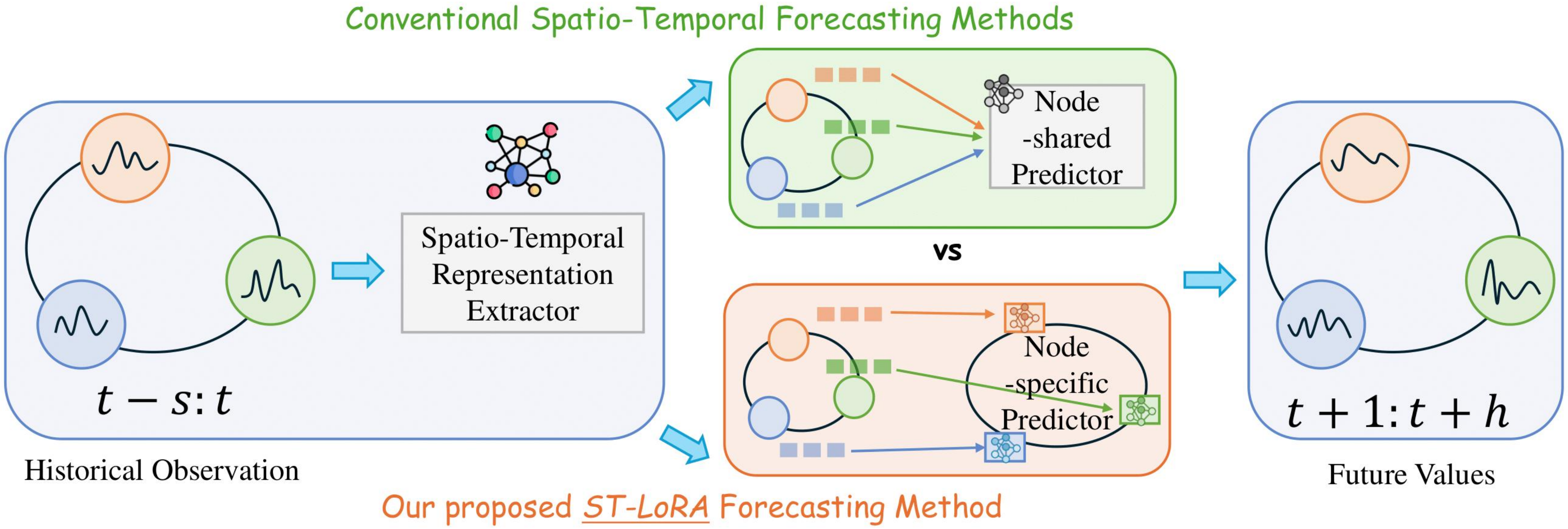
Low-rank Adapation

Low-rank Adaptation is the technique that decomposes high-dimensional parameter spaces into products of low-rank matrices, reducing computational complexity while preserving essential information. Our proposed framework adapts low-rank optimization for node-level heterogeneity in spatio-temporal forecasting.

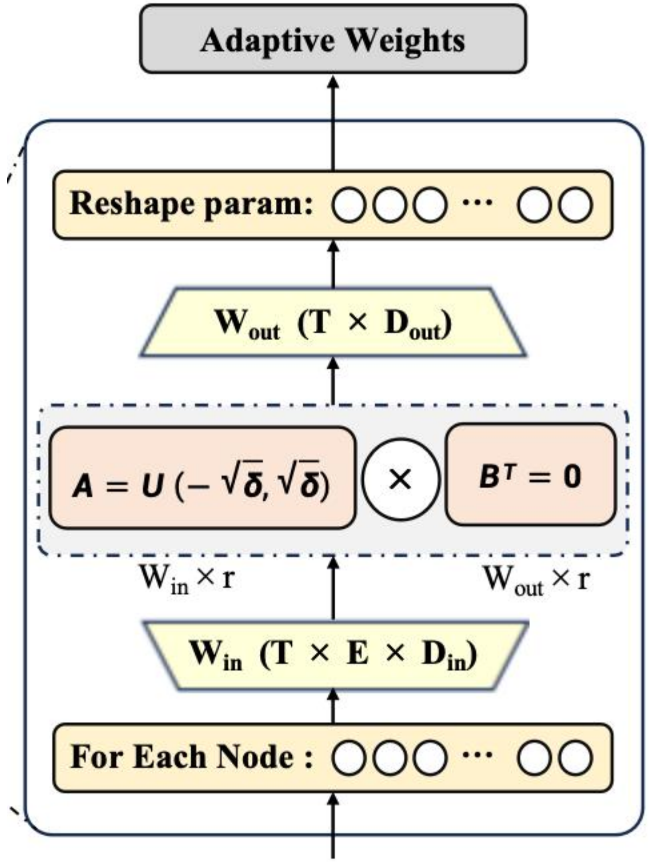


ST-LoRA - Comparison

Existing methods rely only on parameter-sharing node predictors, which are difficult to capture the different distributional changes among different nodes over time, resulting in limited prediction performance, and the improvement of model performance is always accompanied by the increase of model complexity. On this basis, this paper proposes a new method focusing on specific node predictors to improve the efficiency of existing STGNN methods.



Node-Adaptive Low-rank Layers



(c) Node Adaptive Low-rank Layer

Problem:

Overparameterization → shared predictors fail to capture diverse behaviors

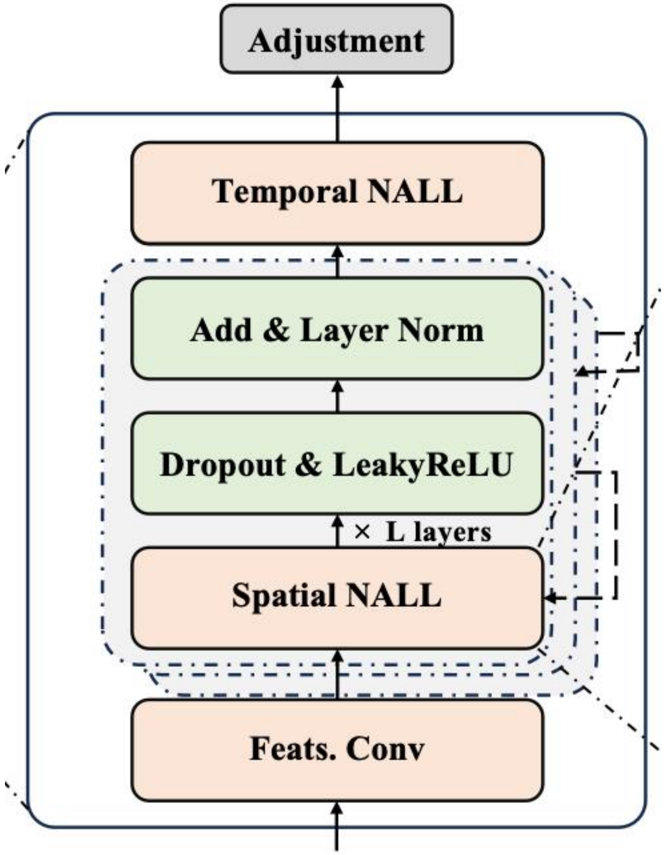
Key Idea:

Introduce low-rank decomposition for node-level parameter adaptation to capture node-level behaviors.

Formulation: $\Delta \mathbf{W}_{v_i} = \mathbf{B}\mathbf{A}_{v_i} \cdot \frac{\alpha}{r}, \quad \hat{\mathbf{y}}_{\mathbf{i}} = \sigma\left(\mathbf{W}\mathbf{x} + \Delta \mathbf{W}_{v_i}\mathbf{x} + \mathbf{b}\right),$

This formulation reduces the parameter complexity from $O(N \cdot d_{in} \cdot d_{out})$ to $O(r \cdot (d_{in} + d_{out}) + N \cdot r^2)$, representing a significant reduction when $r \ll \min(d_{in}, d_{out})$.

Node-specific Predictors



(b) Node-specific Predictor

Problem:
Node heterogeneity → One NALL is not enough
→ complex spatio-temporal dependencies need deeper modeling.

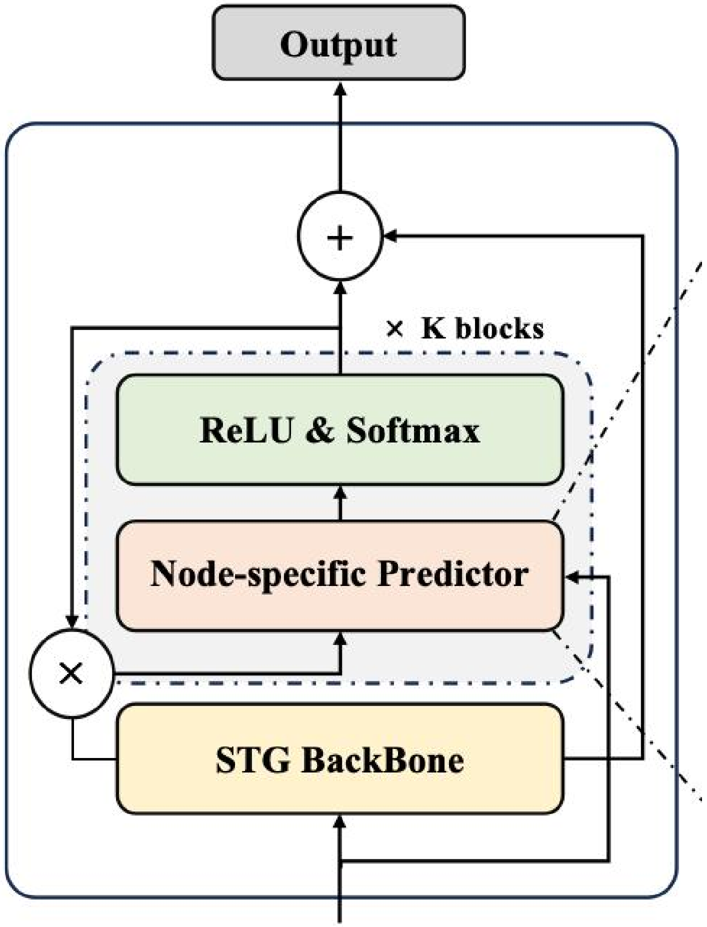
Key Idea: Stack multiple NALLs with residual connections.

Formulation:

$$\mathbf{H}^{(0)} = \mathrm{Conv2D}(\mathbf{X}_{t-T:t}),$$
$$\mathbf{H}^{(l)} = \mathbf{H}^{(l-1)} + \mathrm{NALL}^{(l)}(\sigma(\mathbf{H}^{(l-1)})), \quad l = 1, 2, ..., L$$
$$\hat{\mathbf{Y}}_t = \mathcal{G}_t(\mathbf{H}^{(L)}),$$

Conv2D → temporal feature extraction
Residual NALL layers → spatial heterogeneity modeling
Norm. + Dropout. → stable training

ST-LoRA Integration



(a) ST-LoRA

Goal:

Combine backbone model's generalization with node-specific adaptation.

Forward formulation:

$$
\begin{aligned}
\mathbf{Y}_{\text{base}} &= f_\theta(\mathbf{X}_{t-T:t}),\\
\mathbf{Z}^{(1)} &= \mathcal{H}^{(1)}(\sigma(\mathbf{Y}_{\text{base}})),\\
\mathbf{Z}^{(k)} &= \mathcal{H}^{(k)}(\sigma([\mathbf{X}_{t-T:t}, \mathbf{Z}^{(k-1)}])), \quad k = 2, \ldots, K
\end{aligned}
$$

$$
\mathbf{R} = \sigma(\mathcal{F}([\mathbf{X}_{t-T:t}, \frac{1}{K}\sum_{k=1}^{K}\mathbf{Z}^{(k)}])),
$$

$$
\mathbf{Y}_{\text{final}} = \mathbf{R} \odot \mathbf{Y}_{\text{base}} + (1-\mathbf{R}) \odot \frac{1}{K}\sum_{k=1}^{K}\mathbf{Z}^{(k)},
$$

Optimization:

$$
\mathcal{L} = \frac{1}{T'}\sum_{i=1}^{T'} \|\mathbf{X}_{t+i} - \mathbf{Y}_{\text{final}}^{(t+i)}\|_1 + \lambda\|\alpha\|_2,
$$

Experimental Setup

Datasets:

Table 1: Statistics and description of datasets we used.

Dataset	#Nodes	#Edges	#Frames	Time Range	Type
METR-LA	207	1515	34,272	03/01/2012 – 06/27/2012	Traffic speed
PEMS-BAY	325	2369	52,116	01/01/2017 – 06/30/2017	Traffic speed
PEMS03	358	547	26208	09/01/2018 – 11/30/2018	Traffic flow
PEMS04	307	340	16992	01/01/2018 – 02/28/2018	Traffic flow
PEMS07	883	866	28224	05/01/2017 – 08/06/2017	Traffic flow
PEMS08	170	295	17856	07/01/2016 – 08/31/2016	Traffic flow

Visualization of Traffic Speed/flow Sensor Map

Metrics: MAE / RMSE / MAPE

Baselines: LSTM / STGCN / GraphWaveNet / AGCRN / D2STGNN / STAEformer

Performance Comparisons

Table 2: The improvement of different models in the PEMS04 dataset. Here, lower values indicate better performance. All six baselines have achieved significant improvements, denoted by Δ. The subscripts indicate standard deviations.

Model	15min MAE ↓	15min RMSE ↓	15min MAPE% ↓	30min MAE ↓	30min RMSE ↓	30min MAPE% ↓	60min MAE ↓	60min RMSE ↓	60min MAPE% ↓	Average MAE ↓	Average RMSE ↓	Average MAPE% ↓
HA	$28.92_{\pm1.25}$	$42.69_{\pm1.82}$	$20.31_{\pm0.89}$	$33.73_{\pm1.28}$	$49.37_{\pm1.85}$	$24.01_{\pm0.91}$	$46.97_{\pm1.31}$	$67.43_{\pm1.89}$	$35.11_{\pm0.92}$	$38.03_{\pm1.28}$	$59.24_{\pm1.85}$	$27.88_{\pm0.91}$
VAR	$21.94_{\pm0.62}$	$34.30_{\pm1.02}$	$16.42_{\pm0.48}$	$23.72_{\pm0.71}$	$36.58_{\pm1.08}$	$18.02_{\pm0.52}$	$26.76_{\pm0.82}$	$40.28_{\pm1.23}$	$20.94_{\pm0.64}$	$23.51_{\pm0.72}$	$36.39_{\pm1.11}$	$17.85_{\pm0.55}$
SVR	$22.52_{\pm0.68}$	$35.30_{\pm1.12}$	$14.71_{\pm0.45}$	$27.63_{\pm0.78}$	$42.23_{\pm1.25}$	$18.29_{\pm0.49}$	$37.86_{\pm1.15}$	$56.01_{\pm1.70}$	$26.72_{\pm0.82}$	$28.66_{\pm0.87}$	$44.59_{\pm1.36}$	$19.15_{\pm0.59}$
LSTM	$21.94_{\pm0.59}$	$33.37_{\pm0.93}$	$15.32_{\pm0.40}$	$25.83_{\pm0.66}$	$39.10_{\pm1.04}$	$20.35_{\pm0.43}$	$36.41_{\pm0.82}$	$50.73_{\pm1.28}$	$29.92_{\pm0.56}$	$27.14_{\pm0.69}$	$41.59_{\pm1.08}$	$18.20_{\pm0.46}$
LSTM+	$18.89_{\pm0.58}$	$29.96_{\pm0.91}$	$13.02_{\pm0.40}$	$21.31_{\pm0.65}$	$34.22_{\pm1.04}$	$13.96_{\pm0.43}$	$26.34_{\pm0.80}$	$41.30_{\pm1.26}$	$18.26_{\pm0.56}$	$22.18_{\pm0.68}$	$35.16_{\pm1.07}$	$15.08_{\pm0.46}$
Δ	$-3.05_{\pm0.18}$	$-3.41_{\pm0.20}$	$-2.30_{\pm0.14}$	$-4.52_{\pm0.27}$	$-4.88_{\pm0.29}$	$-6.39_{\pm0.38}$	$-10.07_{\pm0.60}$	$-9.43_{\pm0.57}$	$-11.66_{\pm0.70}$	$-4.96_{\pm0.30}$	$-6.43_{\pm0.39}$	$-3.12_{\pm0.19}$
STGCN	$19.45_{\pm0.59}$	$30.12_{\pm0.92}$	$14.21_{\pm0.43}$	$21.85_{\pm0.62}$	$34.43_{\pm0.97}$	$14.13_{\pm0.44}$	$26.97_{\pm0.68}$	$41.11_{\pm1.06}$	$16.84_{\pm0.48}$	$22.70_{\pm0.63}$	$35.55_{\pm0.98}$	$14.59_{\pm0.45}$
STGCN+	$19.12_{\pm0.58}$	$29.72_{\pm0.91}$	$13.89_{\pm0.42}$	$19.92_{\pm0.61}$	$31.63_{\pm0.96}$	$13.77_{\pm0.42}$	$22.07_{\pm0.67}$	$34.47_{\pm1.05}$	$15.42_{\pm0.47}$	$20.37_{\pm0.62}$	$31.94_{\pm0.97}$	$14.36_{\pm0.44}$
Δ	$-0.33_{\pm0.02}$	$-0.40_{\pm0.02}$	$-0.32_{\pm0.02}$	$-1.93_{\pm0.12}$	$-2.80_{\pm0.17}$	$-0.36_{\pm0.02}$	$-4.90_{\pm0.29}$	$-6.64_{\pm0.40}$	$-1.42_{\pm0.09}$	$-2.33_{\pm0.14}$	$-3.61_{\pm0.22}$	$-0.23_{\pm0.01}$
GWNet	$18.65_{\pm0.57}$	$29.24_{\pm0.89}$	$13.82_{\pm0.42}$	$19.57_{\pm0.60}$	$30.62_{\pm0.92}$	$13.28_{\pm0.39}$	$23.07_{\pm0.70}$	$35.35_{\pm1.08}$	$17.34_{\pm0.53}$	$25.45_{\pm0.62}$	$39.70_{\pm0.97}$	$17.29_{\pm0.45}$
GWNet+	$17.89_{\pm0.55}$	$28.52_{\pm0.87}$	$12.64_{\pm0.39}$	$18.88_{\pm0.58}$	$29.38_{\pm0.89}$	$13.06_{\pm0.40}$	$20.89_{\pm0.64}$	$32.96_{\pm1.00}$	$14.92_{\pm0.46}$	$19.22_{\pm0.59}$	$30.62_{\pm0.93}$	$13.54_{\pm0.41}$
Δ	$-0.76_{\pm0.05}$	$-0.72_{\pm0.04}$	$-1.18_{\pm0.07}$	$-0.69_{\pm0.04}$	$-1.24_{\pm0.07}$	$-0.22_{\pm0.01}$	$-2.18_{\pm0.13}$	$-2.39_{\pm0.14}$	$-2.42_{\pm0.15}$	$-6.23_{\pm0.37}$	$-9.08_{\pm0.54}$	$-3.75_{\pm0.23}$
AGCRN	$18.12_{\pm0.55}$	$29.45_{\pm0.90}$	$12.85_{\pm0.39}$	$18.77_{\pm0.57}$	$30.08_{\pm0.92}$	$12.97_{\pm0.40}$	$20.41_{\pm0.62}$	$32.87_{\pm1.00}$	$14.38_{\pm0.44}$	$19.83_{\pm0.58}$	$32.26_{\pm0.94}$	$13.40_{\pm0.41}$
AGCRN+	$17.83_{\pm0.54}$	$29.16_{\pm0.89}$	$12.55_{\pm0.38}$	$18.63_{\pm0.57}$	$29.99_{\pm0.91}$	$12.82_{\pm0.39}$	$19.97_{\pm0.61}$	$32.37_{\pm0.99}$	$13.78_{\pm0.42}$	$18.81_{\pm0.57}$	$30.51_{\pm0.93}$	$13.05_{\pm0.40}$
Δ	$-0.29_{\pm0.02}$	$-0.29_{\pm0.02}$	$-0.30_{\pm0.02}$	$-0.14_{\pm0.01}$	$-0.09_{\pm0.01}$	$-0.15_{\pm0.01}$	$-0.44_{\pm0.03}$	$-0.50_{\pm0.03}$	$-0.60_{\pm0.04}$	$-1.02_{\pm0.06}$	$-1.75_{\pm0.11}$	$-0.35_{\pm0.02}$
STAE	$17.95_{\pm0.55}$	$29.12_{\pm0.89}$	$12.65_{\pm0.39}$	$18.92_{\pm0.58}$	$30.09_{\pm0.92}$	$13.35_{\pm0.41}$	$21.06_{\pm0.64}$	$33.37_{\pm1.02}$	$15.55_{\pm0.47}$	$19.31_{\pm0.59}$	$30.86_{\pm0.94}$	$13.85_{\pm0.42}$
STAE+	$17.65_{\pm0.54}$	$28.73_{\pm0.88}$	$12.45_{\pm0.38}$	$18.62_{\pm0.57}$	$29.55_{\pm0.90}$	$13.29_{\pm0.41}$	$20.40_{\pm0.62}$	$32.38_{\pm0.99}$	$15.00_{\pm0.46}$	$18.89_{\pm0.58}$	$30.22_{\pm0.92}$	$13.58_{\pm0.41}$
Δ	$-0.30_{\pm0.02}$	$-0.39_{\pm0.02}$	$-0.20_{\pm0.01}$	$-0.30_{\pm0.02}$	$-0.54_{\pm0.03}$	$-0.06_{\pm0.01}$	$-0.66_{\pm0.04}$	$-0.99_{\pm0.06}$	$-0.55_{\pm0.03}$	$-0.42_{\pm0.03}$	$-0.64_{\pm0.04}$	$-0.27_{\pm0.02}$
D2STGNN	$18.95_{\pm0.58}$	$29.85_{\pm0.91}$	$14.82_{\pm0.45}$	$19.96_{\pm0.61}$	$31.34_{\pm0.95}$	$15.52_{\pm0.47}$	$23.34_{\pm0.71}$	$35.89_{\pm1.09}$	$17.39_{\pm0.53}$	$20.75_{\pm0.63}$	$32.36_{\pm0.99}$	$15.91_{\pm0.49}$
D2STGNN+	$18.25_{\pm0.56}$	$28.92_{\pm0.88}$	$14.12_{\pm0.43}$	$19.21_{\pm0.59}$	$30.50_{\pm0.93}$	$13.46_{\pm0.41}$	$21.73_{\pm0.66}$	$33.73_{\pm1.03}$	$17.00_{\pm0.52}$	$19.73_{\pm0.60}$	$31.05_{\pm0.95}$	$14.86_{\pm0.45}$
Δ	$-0.70_{\pm0.04}$	$-0.93_{\pm0.06}$	$-0.70_{\pm0.04}$	$-0.75_{\pm0.05}$	$-0.84_{\pm0.05}$	$-2.06_{\pm0.12}$	$-1.61_{\pm0.10}$	$-2.16_{\pm0.13}$	$-0.39_{\pm0.02}$	$-1.02_{\pm0.06}$	$-1.31_{\pm0.08}$	$-1.05_{\pm0.06}$

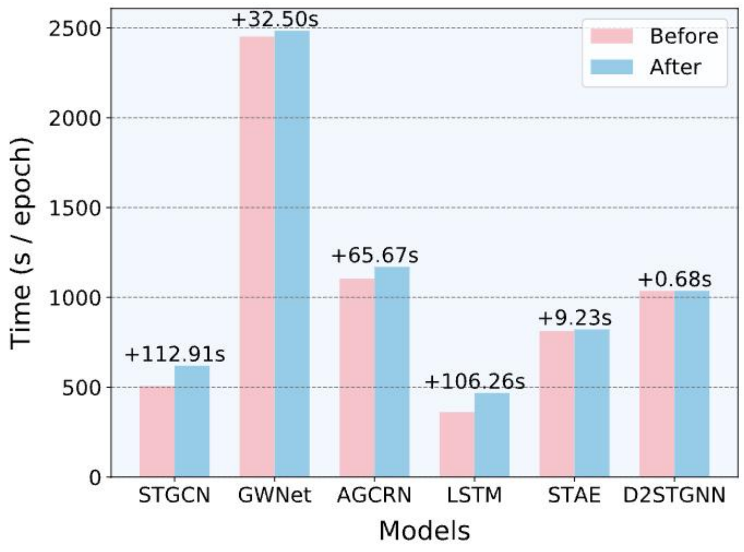
Table 3: Performance Improvements of one of the backbone STGNNs on Multiple Traffic Datasets. We use STGCN in the table as an example to illustrate the significant enhancement of our method from the perspective of the dataset.

Dataset	15min MAE	15min RMSE	15min MAPE%	30min MAE	30min RMSE	30min MAPE%	60min MAE	60min RMSE	60min MAPE%	Average MAE	Average RMSE	Average MAPE%
PEMS04	$-0.33_{\pm0.02}$	$-0.40_{\pm0.03}$	$-0.07_{\pm0.01}$	$-1.45_{\pm0.09}$	$-2.50_{\pm0.15}$	$-0.10_{\pm0.01}$	$-4.94_{\pm0.30}$	$-6.61_{\pm0.40}$	$-1.28_{\pm0.08}$	$-2.24_{\pm0.14}$	$-3.17_{\pm0.19}$	$-0.48_{\pm0.03}$
PEMS08	$-0.20_{\pm0.01}$	$-1.09_{\pm0.07}$	$-0.57_{\pm0.03}$	$-1.37_{\pm0.08}$	$-1.79_{\pm0.11}$	$-0.30_{\pm0.02}$	$-7.63_{\pm0.46}$	$-5.86_{\pm0.35}$	$-1.28_{\pm0.08}$	$-3.07_{\pm0.18}$	$-2.91_{\pm0.17}$	$-0.72_{\pm0.04}$
PEMS03	$-0.31_{\pm0.02}$	$-0.51_{\pm0.03}$	$-0.72_{\pm0.04}$	$-0.22_{\pm0.01}$	$-0.23_{\pm0.01}$	$-0.60_{\pm0.04}$	$-0.45_{\pm0.03}$	$-0.35_{\pm0.02}$	$-1.25_{\pm0.08}$	$-0.33_{\pm0.02}$	$-0.36_{\pm0.02}$	$-0.86_{\pm0.05}$
PEMS07	$-0.26_{\pm0.02}$	$-0.29_{\pm0.02}$	$-0.16_{\pm0.01}$	$-0.38_{\pm0.02}$	$-0.35_{\pm0.02}$	$-0.31_{\pm0.02}$	$-0.59_{\pm0.04}$	$-0.59_{\pm0.04}$	$-0.32_{\pm0.02}$	$-0.41_{\pm0.03}$	$-0.41_{\pm0.03}$	$-0.26_{\pm0.02}$
METR-LA	$-0.10_{\pm0.01}$	$-0.36_{\pm0.02}$	$-0.03_{\pm0.00}$	$-0.33_{\pm0.02}$	$-0.83_{\pm0.05}$	$-0.80_{\pm0.05}$	$-1.05_{\pm0.06}$	$-1.97_{\pm0.12}$	$-2.30_{\pm0.14}$	$-0.49_{\pm0.03}$	$-1.05_{\pm0.06}$	$-1.04_{\pm0.06}$
PEMSBAY	$-0.04_{\pm0.00}$	$-0.06_{\pm0.00}$	$-0.11_{\pm0.01}$	$-0.10_{\pm0.01}$	$-0.45_{\pm0.03}$	$-0.23_{\pm0.01}$	$-0.50_{\pm0.03}$	$-1.14_{\pm0.07}$	$-1.02_{\pm0.06}$	$-0.21_{\pm0.01}$	$-0.55_{\pm0.03}$	$-0.45_{\pm0.03}$

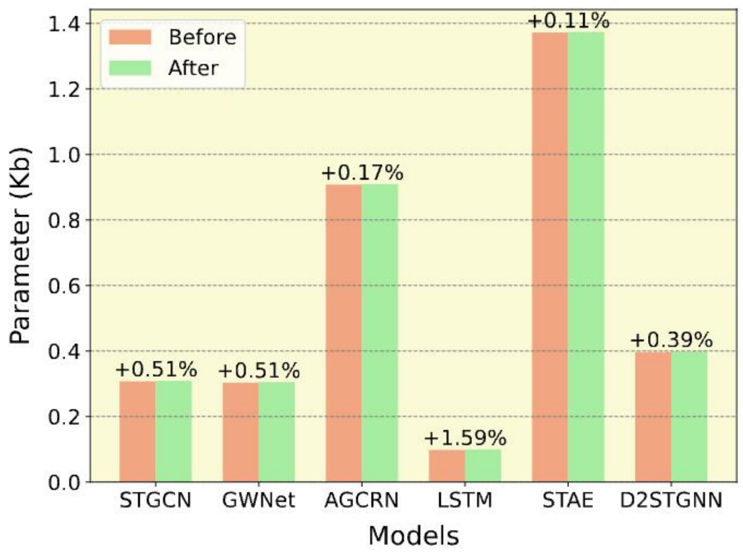
ST-LoRA consistently improves accuracy across diverse backbones and datasets, with the largest gains in long-horizon forecasts, all while adding less than 1% parameters.

Efficiency and scalability Studies (RQ3)

Time Efficiency. In Figure 4(a), we analyze the computational overhead when integrating our framework with existing models. The results demonstrate that ST-LoRA introduces minimal additional training time while delivering substantial performance gains.



(a) Processing time.



(b) Training parameter.

Parameter Sensitivity Analysis (RQ4)

1) Proportions of Dataset: our methods also works in the case of missing data;
2) The node embedding dimension represents the rank of the low-rank matrix required for the additional parameters of the nodes, which tends to increase as more feature information is included in the data;
3) After stacking multiple layers of NALL, the fine-tuning effect of these additional parameter spaces is amplified.

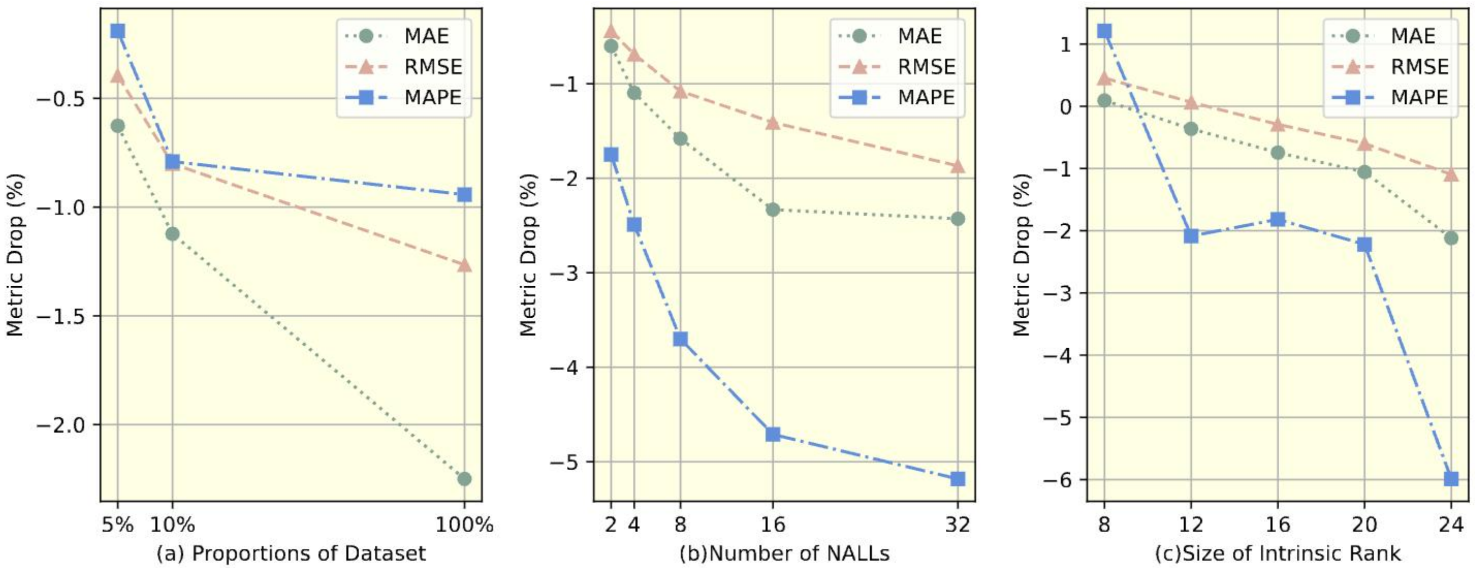


Fig. 5: Parameter sensitivity analysis examining the impact of varying dimensions, placements, and quantities of adaptive components on model performance.

Conclusion

In this paper, we focus on the challenges of node-level heterogeneity and overparameterization in spatio-temporal modeling. We introduce **ST-LoRA**, a novel framework featuring Node-Adaptive Low-rank Layers and Node-Specific Predictors that efficiently customize parameters while maintaining computational efficiency. Our proposed approach enhances existing models by effectively capturing heterogeneous features and distributional changes within independent nodes. The improvements demonstrate that efficient low-rank adaptation can significantly enhance forecasting in domains with heterogeneous node behaviors.

Contributions:

- A node-level heterogeneity perspective for STF;
- A general low-rank adaptation method for existing ST models;
- Extensive empirical studies.

Future work:

Develop a **unified, automated framework** for flexible LoRA variants, and benchmark across diverse spatio-temporal models and datasets.

GitHub

WeChat

GitHub

WeChat

Thanks for listening

Q&A

Now I am looking for RA and PhD opportunities.
If you are interested in my work, please feel free to talk to me!

rwlinno@gmail.com

FQA

Q1: “What’s the essential differences with parameter sharing/Lora at regional level?”
A1: Unlike sharing one predictor, ST-LoRA gives **each node** its own low-rank adaptation via node-specific matrices. Each node only learns a small low-rank matrix A_i, combined with a shared global matrix B.

Q2: “Can it be used for multimodal/multi-task/meteorological etc?”
A2: Yes, it’s **model-agnostic** and can extend to meteorology, mobility, or other spatio-temporal tasks.

Q3: “If the number of nodes * **is extremely large** * , is the parameter pressure still controllable?”
A3: Overhead is still controllable, scaling only with low-rank dimension r, not full parameters.

Q4: How does it avoid overfitting with node-specific parameters?
A4: Low-rank factorization + regularization keep adaptations lightweight and generalizable.