Heterogeneity-Informed Meta-Parameter Learning for Spatiotemporal Time Series Forecasting

Zheng Dong[1,*], Renhe Jiang[2,*], Haotian Gao[2], Hangchen Liu[1], Jinliang Deng[3], Qingsong Wen[4], Xuan Song[5,1,†]

[1]Southern University of Science and Technology, [2]The University of Tokyo

[3]Hong Kong University of Science and Technology, [4]Squirrel AI, [5]Jilin University

[*]Equal Contribution, [†]Corresponding Author

KDD 2024 Research Track, Barcelona, Spain

GitHub

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

東京大学
THE UNIVERSITY OF TOKYO

KDD2024
BARCELONA, SPAIN

Challenge: Spatiotemporal Heterogeneity

- **Spatiotemporal heterogeneity is a key challenge in spatiotemporal forecasting.**

  - **Spatial heterogeneity**

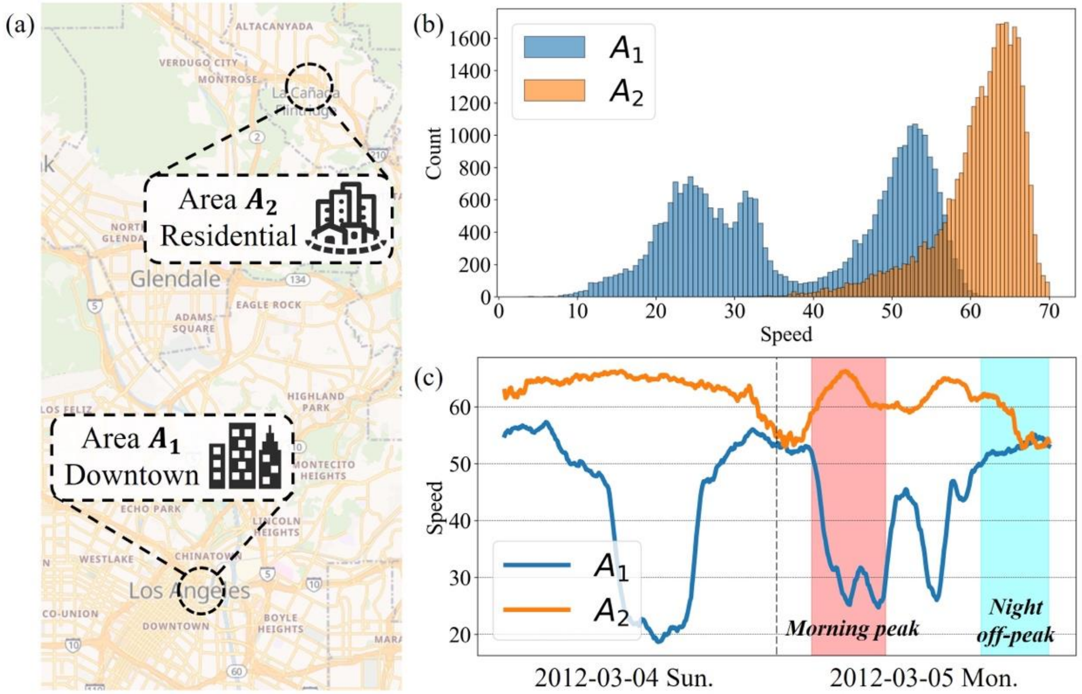
    Different locations + same time = different patterns

    e.g. different traffic speed distribution of A_1 and A_2

  - **Temporal heterogeneity**

    Same location + **different times** = different patterns

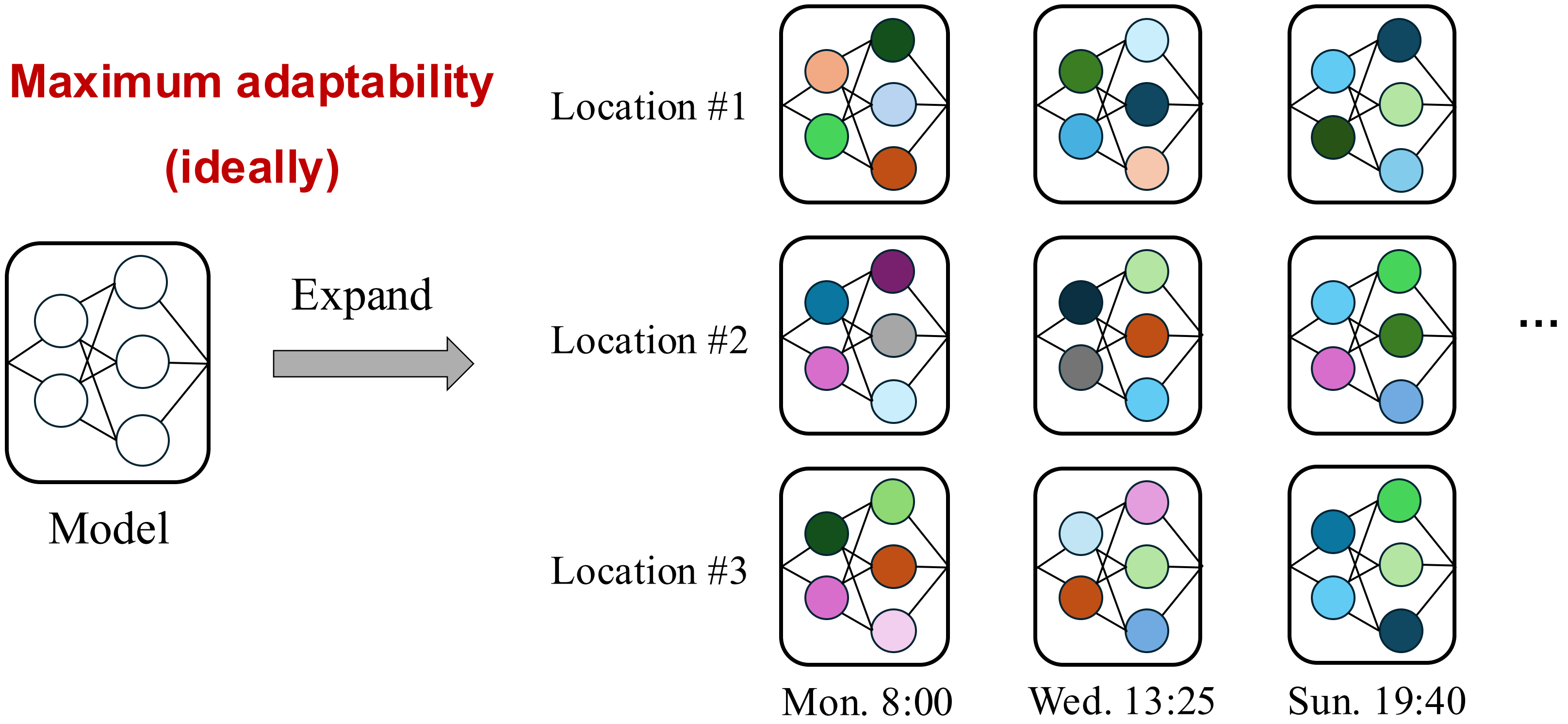
    e.g. different speed patterns between Sun. and Mon.



- Instead of designing new models, can we make the entire model **adaptive** to such different cases?

Motivation: Spatial/Temporal-Specific Parameters

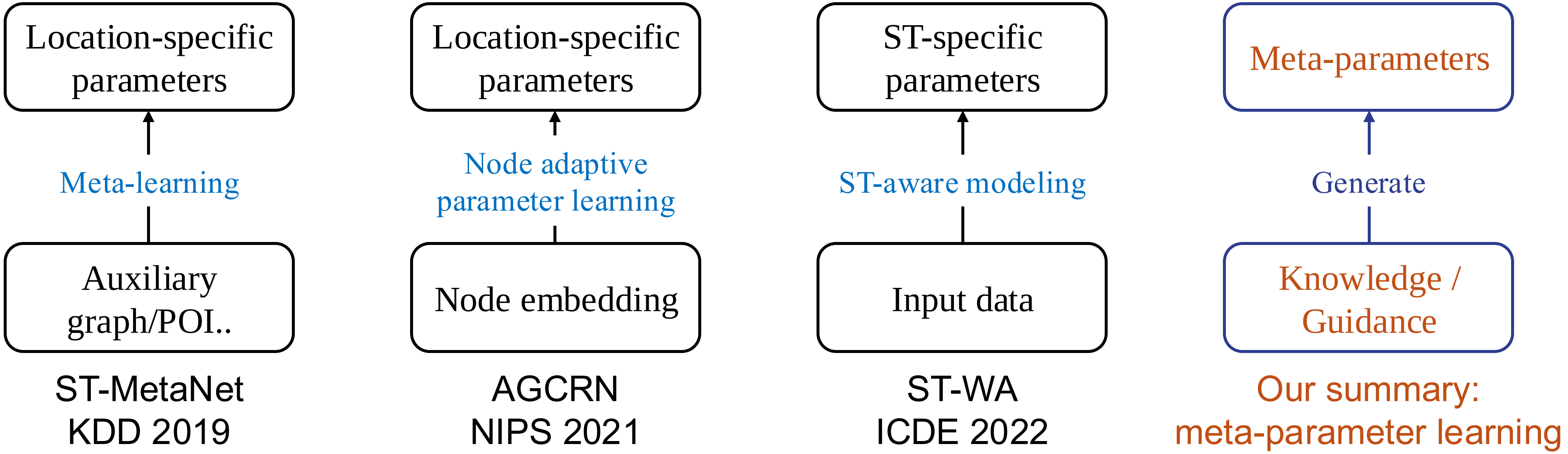
- A straightforward idea: maintain **different parameters** for each spatial location and time step



- **But not practical, because it is impossible to optimize such exploding parameters.**

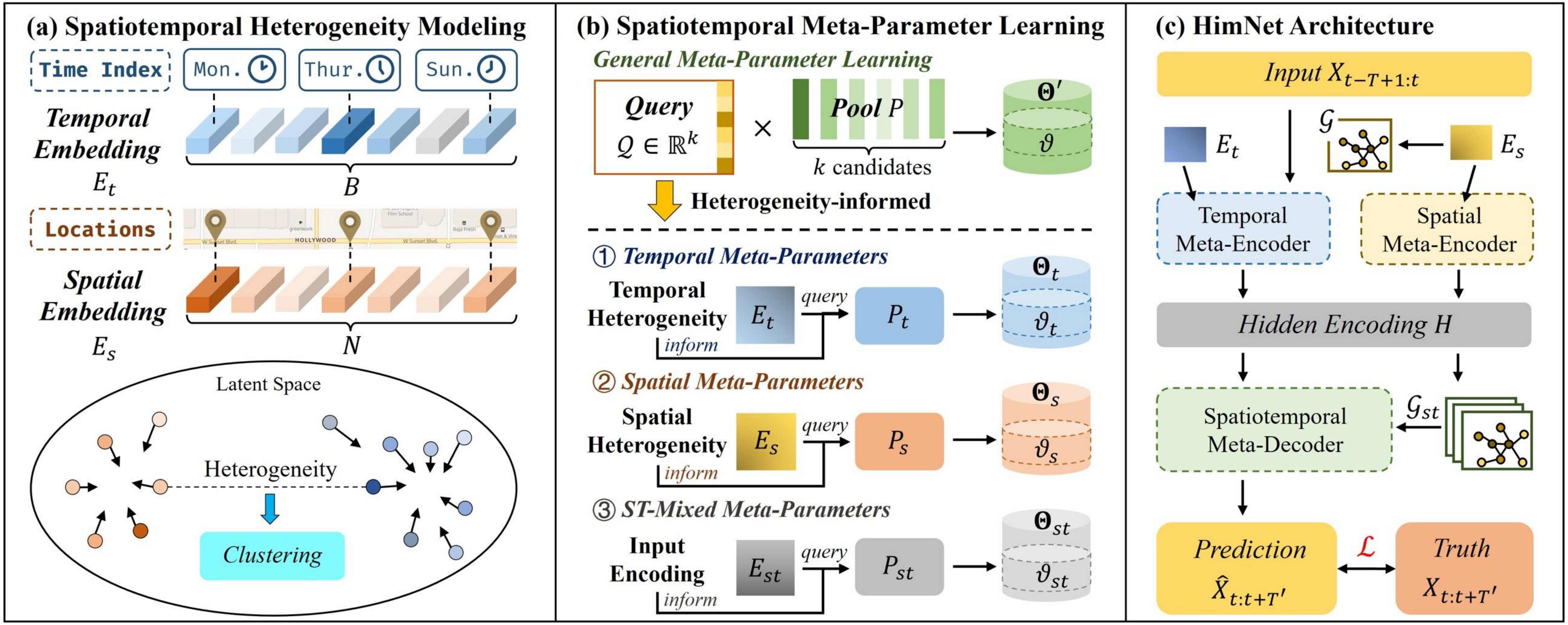
Motivation: Meta-Parameter Learning

- Meta-parameter learning: learn parameters dynamically on each input



- **Can heterogeneity be a guidance? → Heterogeneity-Informed Meta-Parameter Learning**

Methodology: Overview

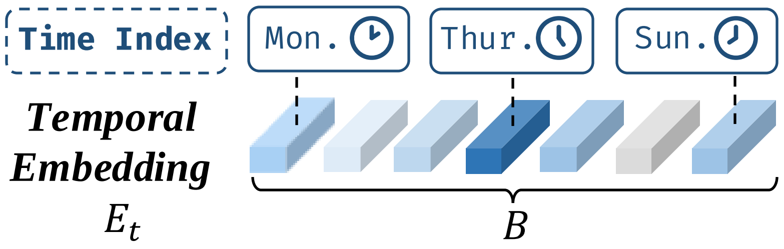


① Modeling Heterogeneity $\xrightarrow{\text{inform}}$ **② Learning Meta-Parameters** $\xrightarrow{\text{for}}$ **③ Predictor: HimNet**

Methodology: Heterogeneity Modeling

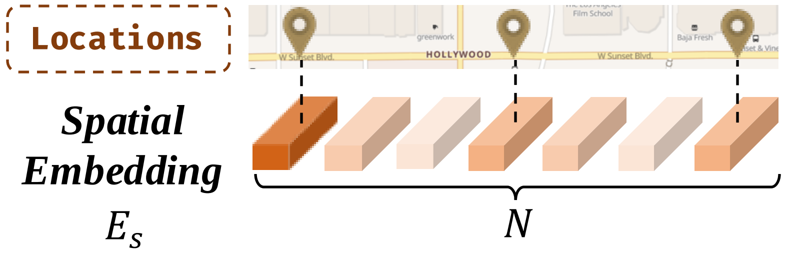
- E_t: embeddings for each **time step** in a week

  Capturing temporal heterogeneity



- E_s: embeddings for each **spatial location** in the dataset

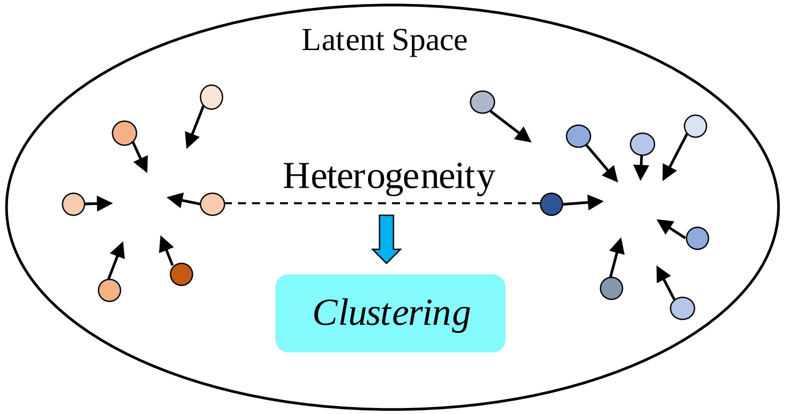
  Capturing spatial heterogeneity



- From the clustering view

  definition of heterogeneity

  - Times/locations having (dis)similar patterns will have (dis)similar embeddings after training
  - **Heterogeneity of ST data → heterogeneity of embeddings**, analogous to word embeddings in NLP models



Methodology: Meta-Parameter Learning

General Meta-Parameter Learning

- For a weight $W \in \mathbb{R}^{C \times D}$, initialize a meta-parameter **pool** $P = [W_1, W_2, \dots, W_k] \in \mathbb{R}^{k \times C \times D}$; for n query vectors $Q \in \mathbb{R}^{n \times k}$, calculate the **meta-parameter** $W_m \in \mathbb{R}^{n \times C \times D}$ by $W_m = QP$

Heterogeneity-Informed Meta-Parameter Learning

- ***Temporal Meta-Parameters***

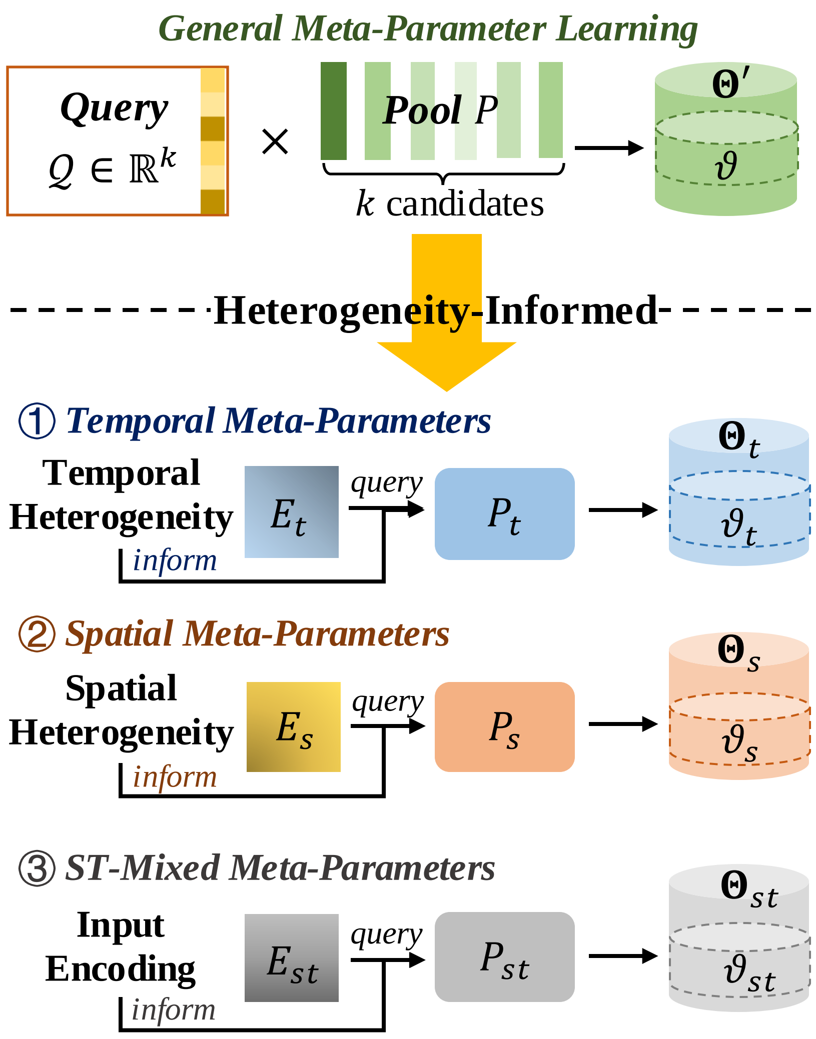
  Use E_t as queries $\rightarrow W_t = E_t P_t = [W_1, W_2, \dots, W_B] \in \mathbb{R}^{B \times C \times D}$

- ***Spatial Meta-Parameters***

  Use E_s as queries $\rightarrow W_s = E_s P_s = [W_1, W_2, \dots, W_N] \in \mathbb{R}^{N \times C \times D}$

- ***ST-Mixed Meta-Parameters***

  Queries $E_{st} = F_{enc}(X) \in \mathbb{R}^{B \times N \times d_{st}} \rightarrow W_{st} = E_{st} P_{st} \in \mathbb{R}^{B \times N \times C \times D}$



Methodology: HimNet

Heterogeneity-Informed Spatiotemporal Meta-Network

- Simple encoder-decoder architecture, based on Graph Convolutional Recurrent Unit (GCRU) [1] with adaptive graph learning mechanism [2,3].

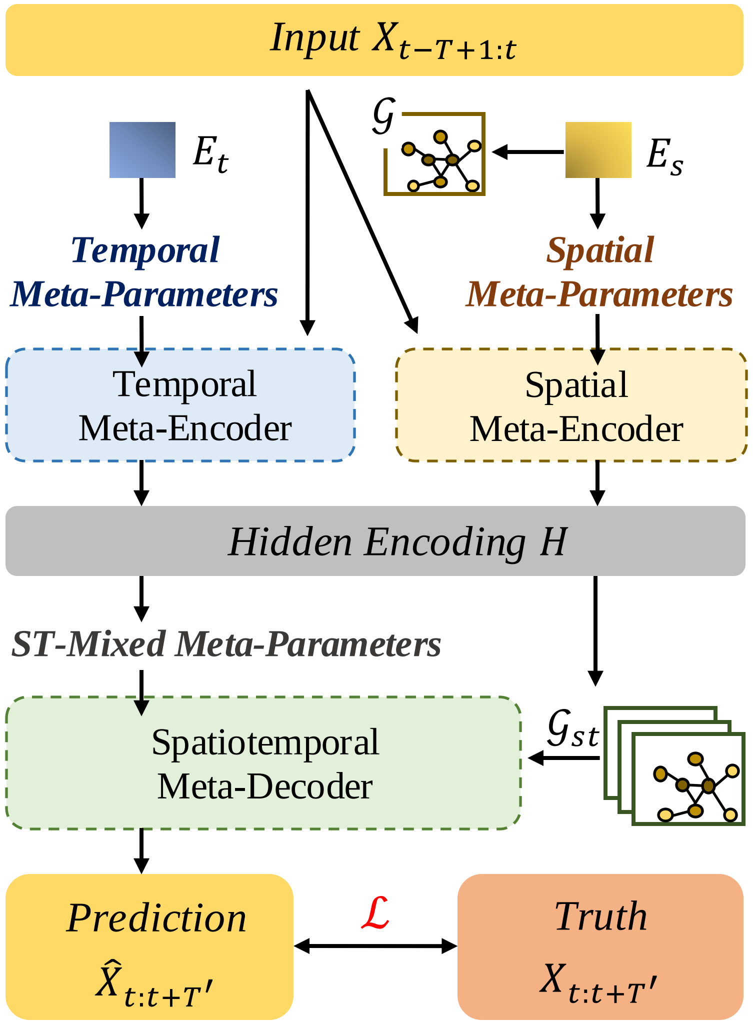
$$r_t = \sigma(\boldsymbol{\Theta}_r \star_{\mathcal{G}} [X_t, H_{t-1}] + b_r) \qquad u_t = \sigma(\boldsymbol{\Theta}_u \star_{\mathcal{G}} [X_t, H_{t-1}] + b_u)$$

$$c_t = \tanh(\boldsymbol{\Theta}_c \star_{\mathcal{G}} [X_t, (r_t \odot H_{t-1})] + b_c)$$

$$H_t = u_t \odot H_{t-1} + (1 - u_t) \odot c_t$$

$$X \star_{\mathcal{G}} = \mathrm{sum}_i(\tilde{A} X W_i) \qquad \tilde{A} = \mathrm{Softmax}(\mathrm{ReLU}(E \cdot E^{\top}))$$

- No parameter initialization: $\boldsymbol{\Theta}_{r/u/c}$ are dynamically generated by Meta-Parameter Learning with meta-parameter pools.



[1] Li, Y. et al. 2018. DCRNN. [2] Wu, Z. et al. 2019. Graph WaveNet. [3] Jiang, R. et al. 2023. MegaCRN.

Experiment: Overall Performance

- **HimNet** is SOTA on **5 benchmarks (METRLA, PEMSBAY, PEMS04/07/08)** against **12 baselines**.

Dataset		Metric	HI	GRU	STGCN	DCRNN	GWNet	AGCRN	GTS	STNorm	STID	ST-WA	PDFormer	MegaCRN	**HimNet**
METRLA	Step 3 15 min	MAE	6.80	3.07	2.75	2.67	2.69	2.85	2.75	2.81	2.82	2.89	2.83	2.65	**2.60**
		RMSE	14.21	6.09	5.29	5.16	5.15	5.53	5.27	5.57	5.53	5.62	5.45	5.08	**5.02**
		MAPE	16.72%	8.14%	7.10%	6.86%	6.99%	7.63%	7.12%	7.40%	7.75%	7.66%	7.77%	6.73%	**6.70%**
	Step 6 30 min	MAE	6.80	3.77	3.15	3.12	3.08	3.20	3.14	3.18	3.19	3.25	3.20	3.04	**2.95**
		RMSE	14.21	7.69	6.35	6.27	6.20	6.52	6.33	6.59	6.57	6.61	6.46	6.18	**6.06**
		MAPE	16.72%	10.71%	8.62%	8.42%	8.47%	9.00%	8.62%	8.47%	9.39%	9.22%	9.19%	8.22%	**8.11%**
	Step 12 60 min	MAE	6.80	4.88	3.60	3.54	3.51	3.59	3.59	3.57	3.55	3.68	3.62	3.51	**3.37**
		RMSE	14.21	9.75	7.43	7.47	7.28	7.45	7.44	7.51	7.55	7.59	7.47	7.39	**7.22**
		MAPE	16.71%	14.91%	10.35%	10.32%	9.96%	10.47%	10.25%	10.24%	10.95%	10.78%	10.91%	10.01%	**9.79%**
PEMSBAY	Step 3 15 min	MAE	3.05	1.44	1.36	1.31	1.30	1.35	1.37	1.33	1.31	1.37	1.32	1.28	**1.27**
		RMSE	7.03	3.15	2.88	2.76	2.73	2.88	2.92	2.82	2.79	2.88	2.83	2.71	**2.68**
		MAPE	6.85%	3.01%	2.86%	2.73%	2.71%	2.91%	2.85%	2.76%	2.78%	2.86%	2.78%	2.67%	**2.64%**
	Step 6 30 min	MAE	3.05	1.97	1.70	1.65	1.63	1.67	1.72	1.65	1.64	1.70	1.64	1.60	**1.57**
		RMSE	7.03	4.60	3.84	3.75	3.73	3.82	3.86	3.77	3.73	3.81	3.79	3.69	**3.60**
		MAPE	6.84%	4.45%	3.79%	3.71%	3.73%	3.81%	3.88%	3.66%	3.73%	3.81%	3.71%	3.60%	**3.52%**
	Step 12 60 min	MAE	3.05	2.70	2.02	1.97	1.99	1.94	2.06	1.92	1.91	2.00	1.91	1.90	**1.84**
		RMSE	7.01	6.28	4.63	4.60	4.60	4.50	4.60	4.45	4.42	4.52	4.43	4.49	**4.32**
		MAPE	6.83%	6.72%	4.72%	4.68%	4.71%	4.55%	4.88%	4.46%	4.55%	4.63%	4.51%	4.53%	**4.33%**
PEMS04	Average	MAE	42.35	25.55	19.57	19.63	18.53	19.38	20.96	18.96	18.38	19.06	18.36	18.72	**18.14**
		RMSE	61.66	39.71	31.38	31.26	29.92	31.25	32.95	30.98	29.95	31.02	30.03	30.53	**29.88**
		MAPE	29.92%	17.35%	13.44%	13.59%	12.89%	13.40%	14.66%	12.69%	12.04%	12.52%	12.00%	12.77%	**12.00%**
PEMS07	Average	MAE	49.29	26.74	21.74	21.16	20.47	20.57	22.15	20.50	19.61	20.74	19.97	19.83	**19.21**
		RMSE	71.34	42.78	35.27	34.14	33.47	34.40	35.10	34.66	32.79	34.05	32.95	32.91	**32.75**
		MAPE	22.75%	11.58%	9.24%	9.02%	8.61%	8.74%	9.38%	8.75%	8.30%	8.77%	8.55%	8.36%	**8.03%**
PEMS08	Average	MAE	34.66	19.36	16.08	15.22	14.40	15.32	16.49	15.41	14.21	15.41	13.58	14.75	**13.57**
		RMSE	50.45	31.20	25.39	24.17	23.39	24.41	26.08	24.77	23.28	24.62	23.41	23.73	**23.22**
		MAPE	21.63%	12.43%	10.60%	10.21%	9.21%	10.03%	10.54%	9.76%	9.27%	9.94%	9.05%	9.48%	**8.98%**

- **w/o $\boldsymbol{E}$**: replaces embedding with a constant matrix of all ones, removing heterogeneity modeling
- **w/o MP**: removes meta-parameters by downgrading to randomly initialized parameters

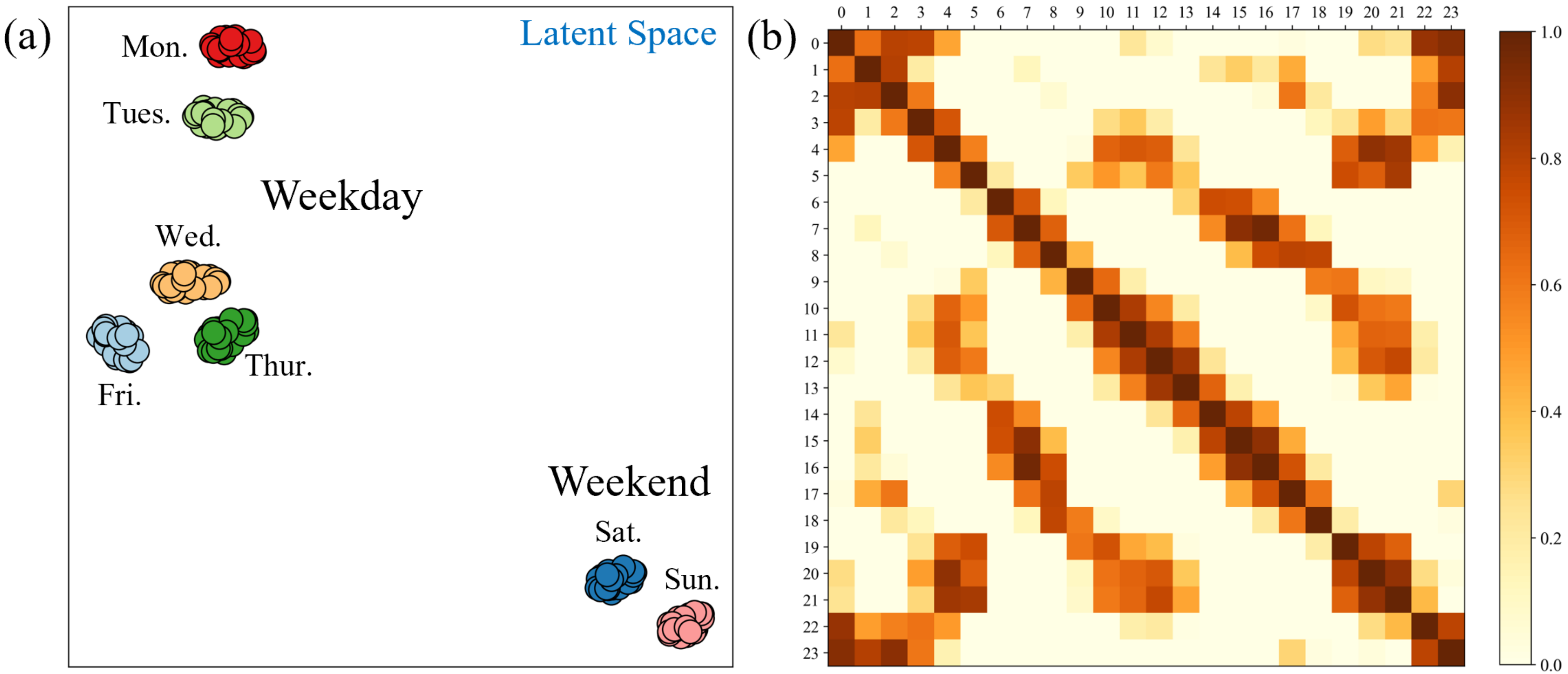
Average prediction MAE of the ablated variants.

Model	METRLA	PEMSBAY	PEMS04	PEMS07	PEMS08
w/o E_t	2.94	1.53	18.35	22.00	14.14
w/o E_s	3.49	1.74	21.30	19.26	14.79
w/o E_{st}	3.07	1.55	18.55	19.77	13.61
w/o TMP	2.94	1.54	18.65	19.58	14.44
w/o SMP	3.53	1.75	21.41	22.26	14.07
w/o STMP	3.01	1.57	18.65	19.86	13.72
HimNet	**2.92**	**1.51**	**18.14**	**19.21**	**13.57**

- The complete HimNet consistently outperforms the ablated versions. Every part is necessary.

Temporal Meta-Parameters

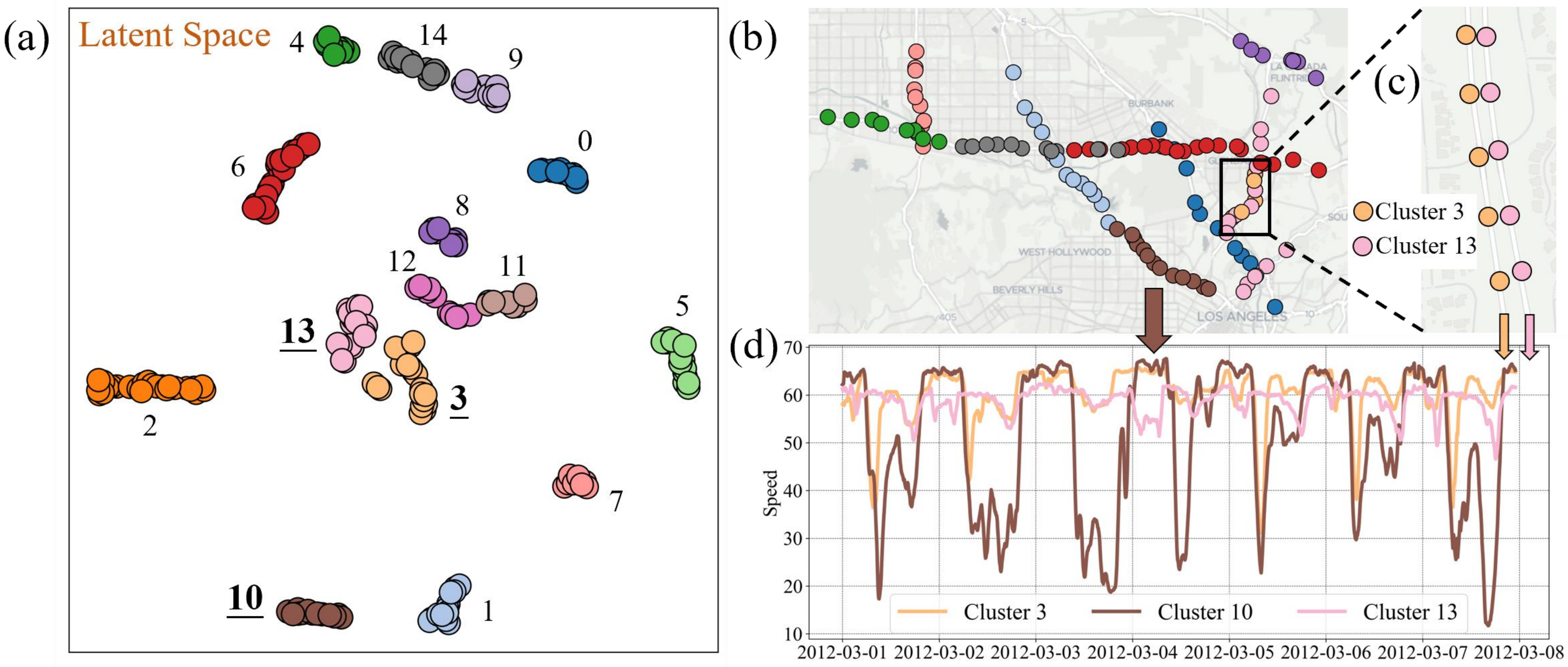
- Can **cluster** and clearly distinguish weekdays and weekends, modeling temporal heterogeneity
- Can capture relations of adjacent hours and peak hours, in consistent with our daily travel patterns



Experiment: Interpretability Analysis

Spatial Meta-Parameters

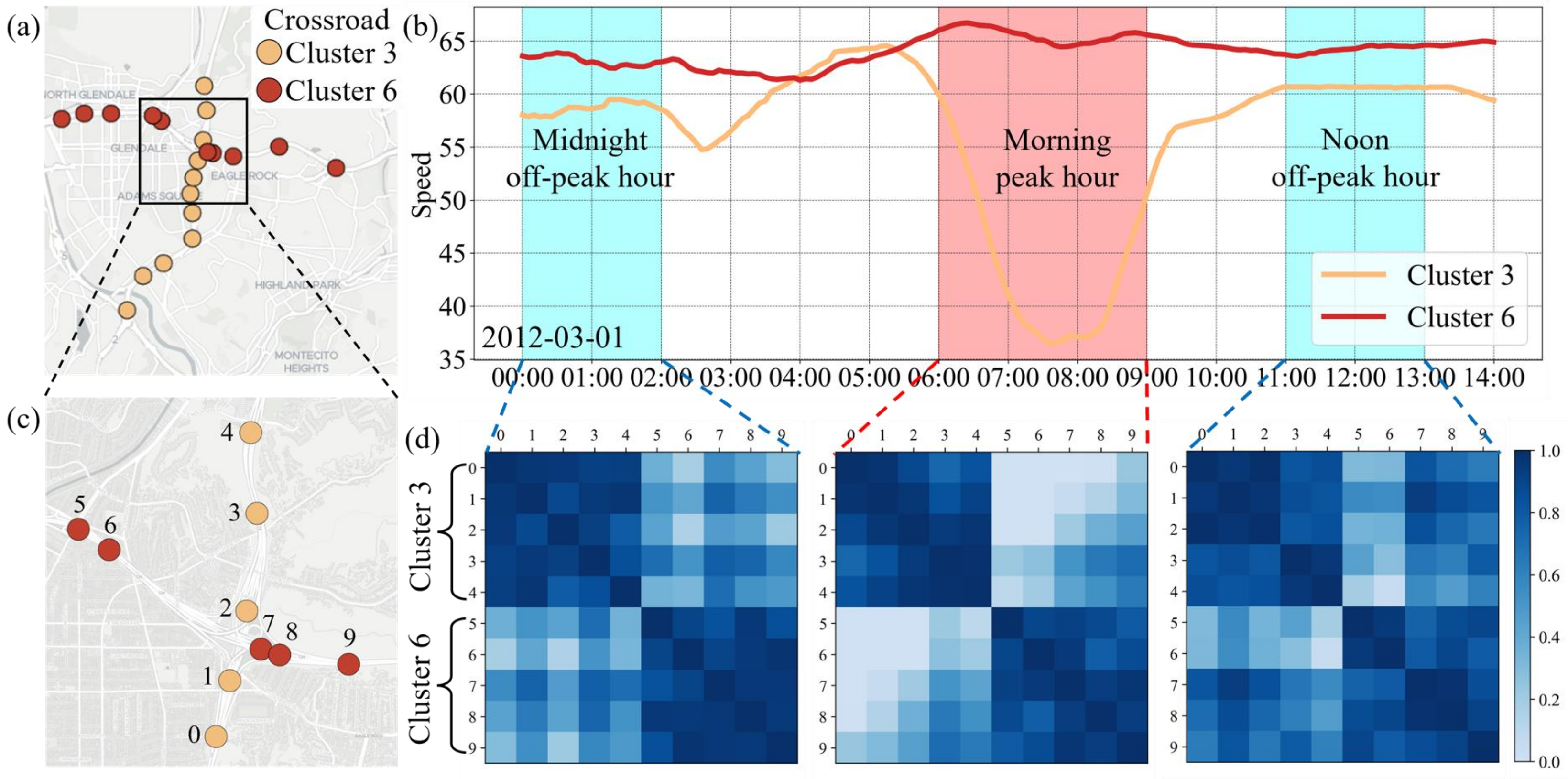
- Can **cluster** spatial locations that matches real road segments on map, modeling spatial heterogeneity
- Can reflect the similarity of raw time series in the dataset, *e.g.* clusters 3, 13 and 10 in the latent space



Experiment: Interpretability Analysis

Spatiotemporal-Mixed Meta-Parameters

- Within-cluster similarity: always **high and stable** during the three selected peak and off-peak hours
- Cross-cluster similarity: **evolves** according to time series patterns, *e.g.* the sharp drop in peak hours



Conclusion

KDD2024
BARCELONA, SPAIN

- Our **Heterogeneity-Informed Meta-Parameter Learning** is the **first** method to model and **fully leverage** spatiotemporal heterogeneity.
- **HimNet** achieves **SOTA** performance on **five** popular spatiotemporal benchmarks (METRLA, PEMSBAY, PEMS04/07/08) against **12 baselines.**
- The proposed meta-parameters show **superior interpretability** in real-world case studies.

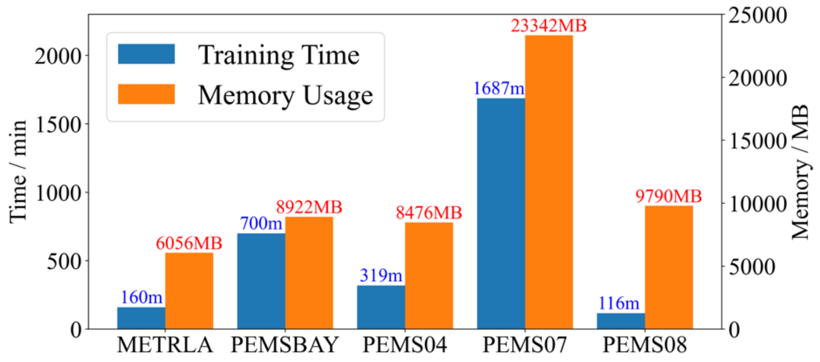
Code on GitHub | **Paper on arXiv** | **WeChat**

Visit our poster at board #131, Aug. 27th 6:30 - 9:30 PM!

■ Training efficiency of **HimNet**

- Faster than Transformer-based models: ST-WA and PDFormer
- Comparable with RNN-based models: DCRNN, AGCRN, GTS, and MegaCRN
- Slower than TCN-based (convolutional) models: STGCN, Graph WaveNet, and STNorm
- HimNet-Θ' confirms that optimizing the expanded spatial/temporal-specific parameters is impossible (refer to page 3).

Training cost of HimNet on each dataset.

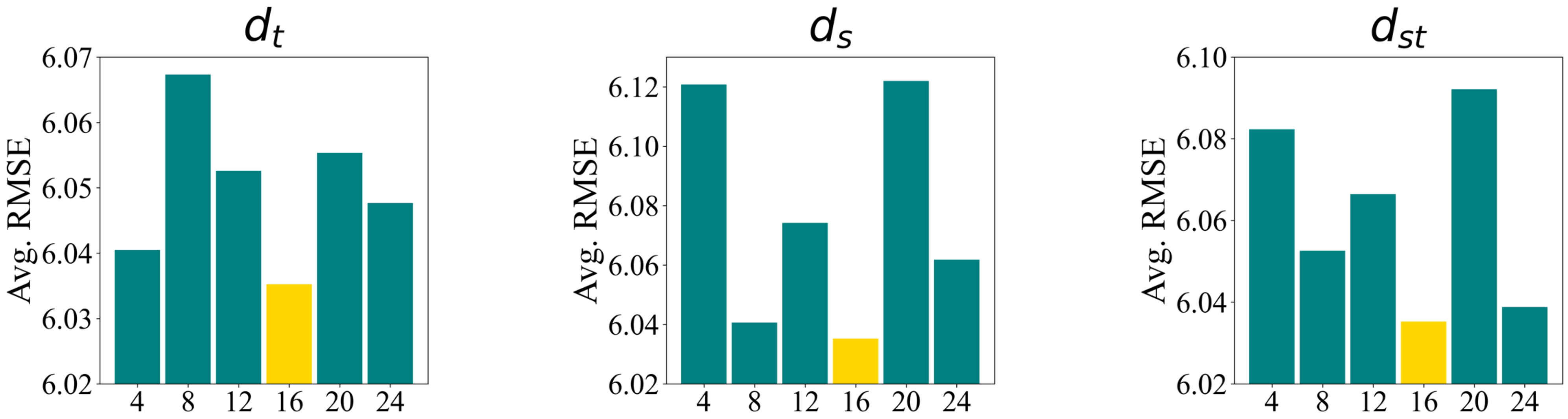


Efficiency comparison of the baselines on METRLA dataset.

	Model	#Params	Time / Batch	Time / Epoch	Mem Usage
Linear	STID [51]	118K	8ms	12s	1420MB
TCN-based	STGCN [64]	246K	23ms	34s	1650MB
	GWNet [60]	309K	40ms	60s	1994MB
	STNorm [11]	224K	39ms	59s	1818MB
RNN-based	DCRNN [36]	372K	189ms	284s	2134MB
	AGCRN [2]	752K	54ms	82s	2492MB
	GTS [49]	38.5M	114ms	171s	4096MB
	MegaCRN [28]	389K	89ms	134s	1962MB
Transformer-based	ST-WA [8]	375K	135ms	203s	2668MB
	PDFormer [26]	531K	173ms	260s	6938MB
HimNet (Also RNN-based)	HimNet-Θ'	10.9B	N/A	N/A	N/A
	HimNet	1251K	97ms	144s	6056MB

Appendix: Hyper-Parameter Study

- An embedding size of 16 generally provides good performance with controllable #parameters.



Average prediction RMSE w.r.t. S/T/ST embedding dimensions on METRLA.