Spatial-Temporal-Decoupled Masked Pre-training for Spatiotemporal Forecasting

IJCAI JEJU 2024

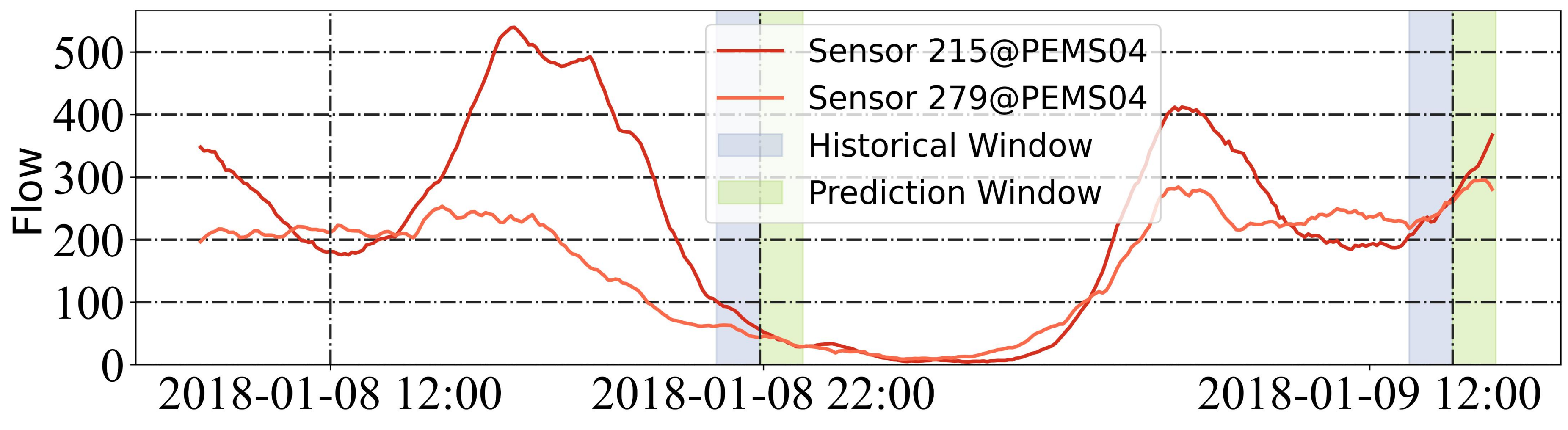
Haotian Gao[1,2], Renhe Jiang[1*], Zheng Dong[2], Jinliang Deng[3], Yuxin Ma[2], Xuan Song[2]

[1]The University of Tokyo, [2]Southern University of Science and Technology, [3]University of Technology Sydney

TL;DR

We propose a novel self-supervised pre-training framework, STD-MAE, which captures clear and comprehensive spatiotemporal heterogeneity through masked pre-training that decouples spatial and temporal dimensions, achieving state-of-the-art performance on multiple spatiotemporal forecasting benchmarks.

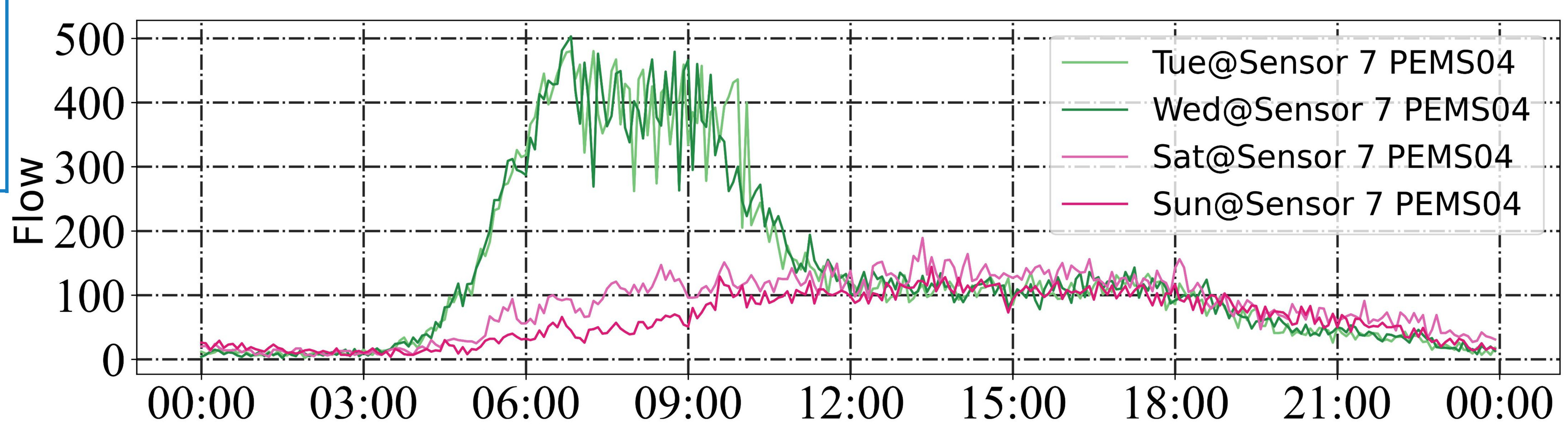
2. Spatiotemporal Mirages



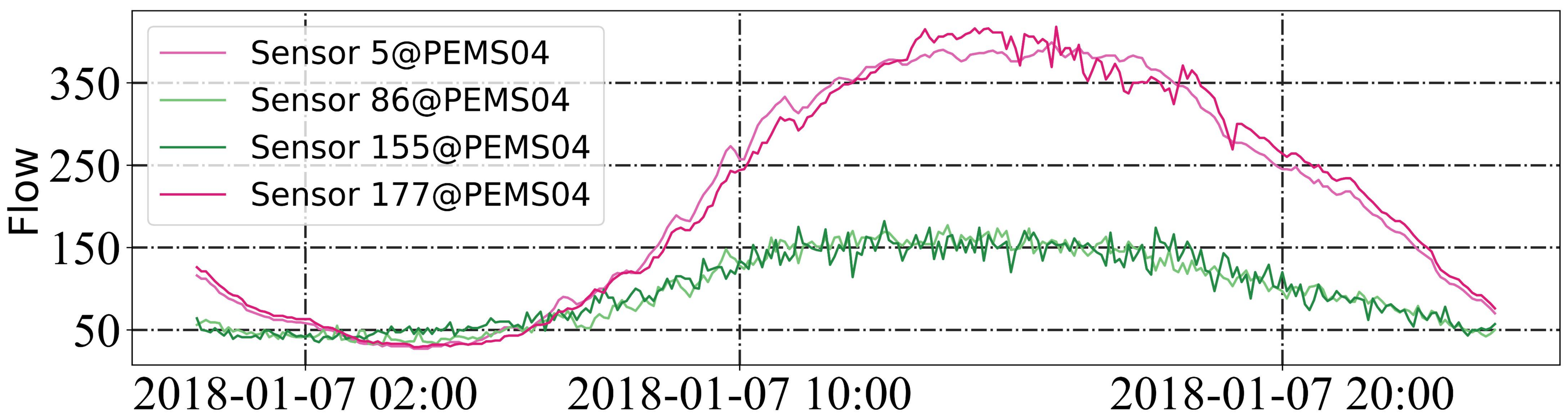
Two kinds of spatiotemporal mirages that may mislead prediction results:
i) dissimilar input time series followed by similar future values;
ii) similar input time series followed by dissimilar future values.

Introduction

1. Spatial-temporal Heterogeneity

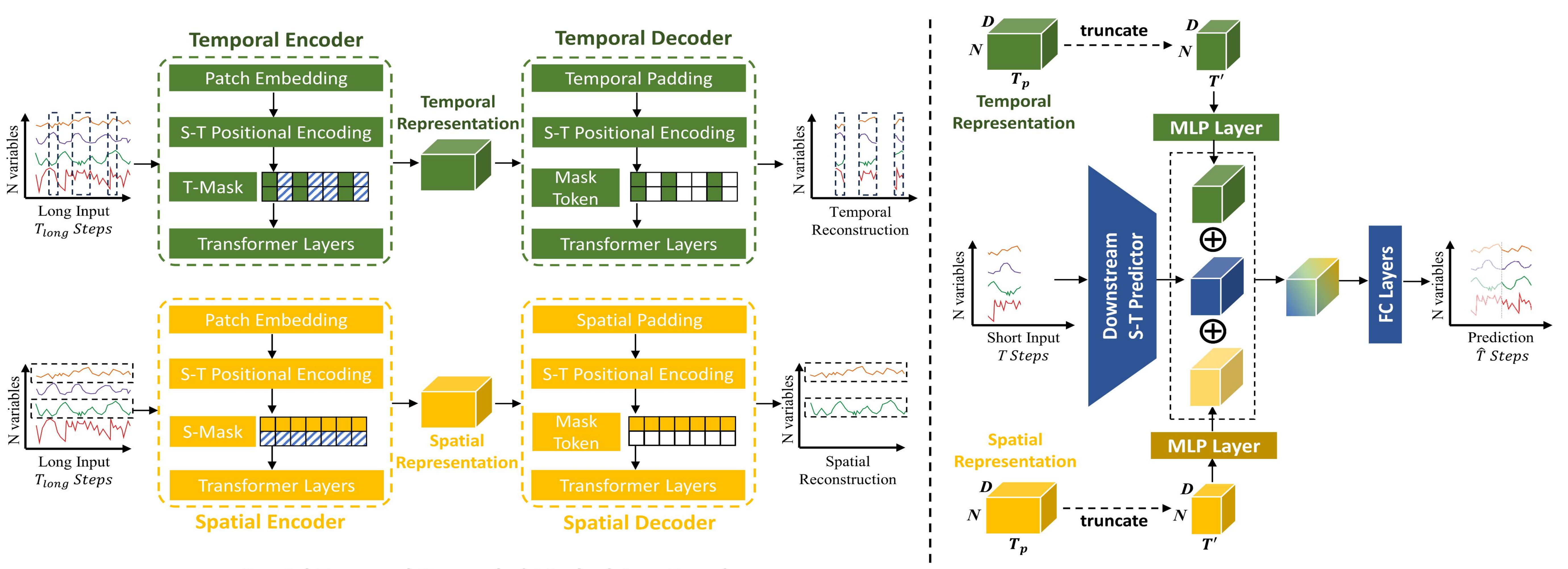


(a) Temporal Heterogeneity



(b) Spatial Heterogeneity

Methodology



Spatial-Temporal-Decoupled Masked AutoEncoder

Downstream Spatiotemporal Predictor

1. Pre-training Phase

Two decoupled masked autoencoders are trained to reconstruct the masked part via temporal masking strategy and spatial masking strategy, respectively.

2. Forecasting Phase

The learned spatial and temporal representations through pre-training are fed to downstream predictors without modifying original architecture.

Experiments

1. Benchmarks Summary

Datasets	#Sensors	#Time Steps	Time Interval
PEMS03	358	26208	5min
PEMS04	307	16992	5min
PEMS07	883	28224	5min
PEMS08	170	17856	5min

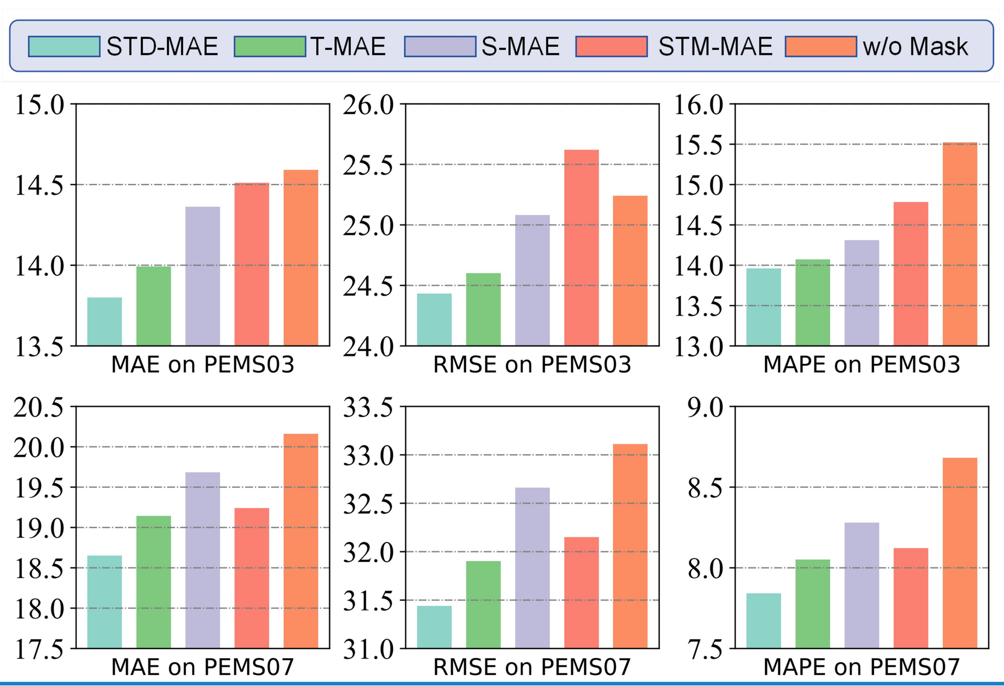
- The benchmarks record traffic flow data collected from the California Transportation Performance Management System (PEMS).
- The flow data is aggregated to 5 minutes, which means there are 12 points for each hour.
- The datasets involve hundreds of sensors and cover several months of data in total.
- We use data from the past hour to forecast the next hour.

2. Forecasting Performance

Model	PEMS03			PEMS04			PEMS07			PEMS08		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
ARIMA [Fang *et al.*, 2021]	35.31	47.59	33.78	33.73	48.80	24.18	38.17	59.27	19.46	31.09	44.32	22.73
VAR [Song *et al.*, 2020]	23.65	38.26	24.51	23.75	36.66	18.09	75.63	115.24	32.22	23.46	36.33	15.42
SVR [Song *et al.*, 2020]	21.97	35.29	21.51	28.70	44.56	19.20	32.49	50.22	14.26	23.25	36.16	14.64
LSTM [Song *et al.*, 2020]	21.33	35.11	23.33	27.14	41.59	18.20	29.98	45.84	13.20	22.20	34.06	14.20
TCN [Lan *et al.*, 2022]	19.31	33.24	19.86	31.11	37.25	15.48	32.68	42.23	14.22	22.69	35.79	14.04
Transformer [Vaswani *et al.*, 2017]	17.50	30.24	16.80	23.83	37.19	15.57	26.80	42.95	12.11	18.52	28.68	13.66
DCRNN [Li *et al.*, 2018]	18.18	30.31	18.91	24.70	38.12	17.12	25.30	38.58	11.66	17.86	27.83	11.45
STGCN [Yu *et al.*, 2018]	17.49	30.12	17.15	22.70	35.55	14.59	25.38	38.78	11.08	18.02	27.83	11.40
GWNet [Wu *et al.*, 2019]	19.85	32.94	19.31	25.45	39.70	17.29	26.85	42.78	12.12	19.13	31.05	12.68
STGODE [Fang *et al.*, 2021]	16.50	27.84	16.69	20.84	32.82	13.77	22.99	37.54	10.14	16.81	25.97	10.62
AGCRN [Bai *et al.*, 2020]	16.06	28.49	15.85	19.83	32.26	12.97	21.29	35.12	8.97	15.95	25.22	10.09
STNorm [Deng *et al.*, 2021]	15.32	25.93	14.37	19.21	32.30	13.05	20.59	34.86	8.61	15.39	24.80	9.91
STEP [Shao *et al.*, 2022]	14.22	24.55	14.42	18.20	29.71	12.48	19.32	32.19	8.12	14.00	23.41	9.50
PDFormer [Jiang *et al.*, 2023]	14.94	25.39	15.82	18.32	29.97	12.10	19.83	32.87	8.53	13.58	23.51	9.05
STAEformer [Liu *et al.*, 2023]	15.35	27.55	15.18	18.22	30.18	11.98	19.14	32.60	8.01	13.46	23.25	8.88
STD-MAE (Ours)	**13.80**	**24.43**	**13.96**	**17.80**	**29.25**	**11.97**	**18.65**	**31.44**	**7.84**	**13.44**	**22.47**	**8.76**

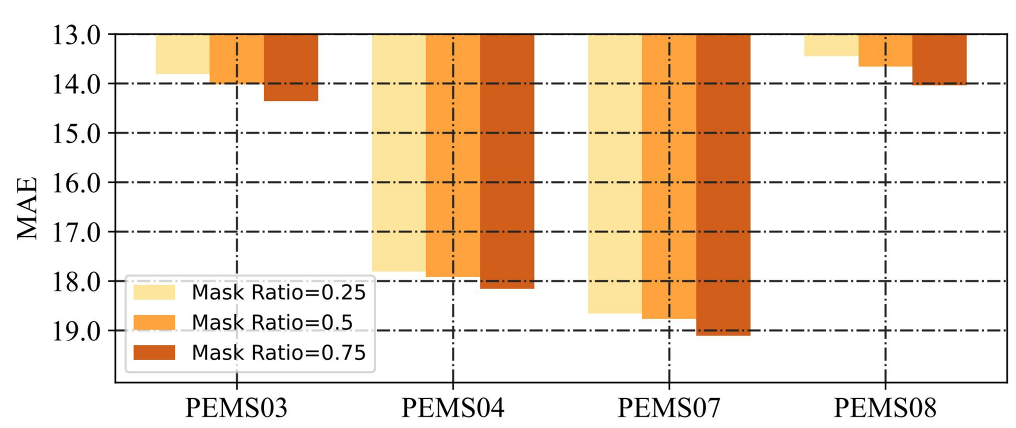
STD-MAE has sustained state-of-the-art performance on three out of these four spatiotemporal forecasting benchmarks for over six months.

3. Masking Mechanism Ablation



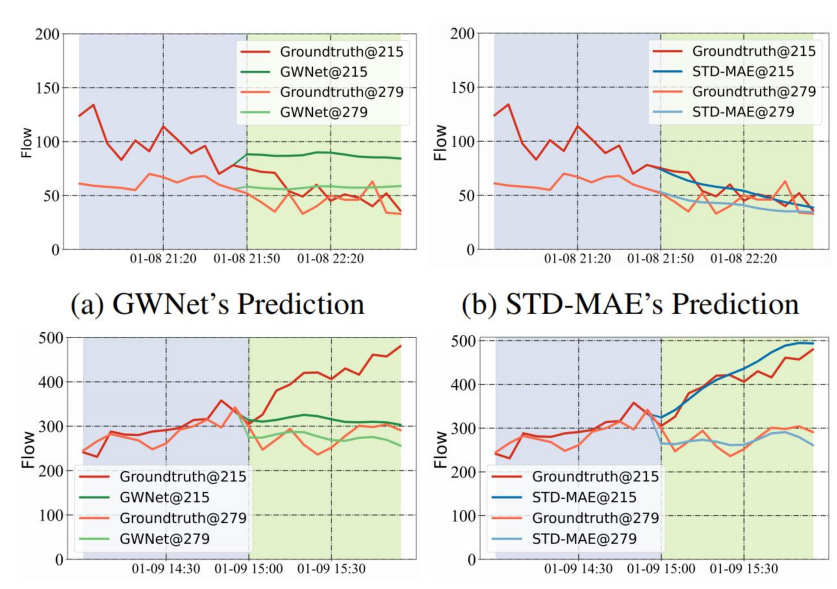
- **STD-MAE**: Using spatial-temporal-decoupled masking.
- **T-MAE**: Only masking on the temporal dimension.
- **S-MAE**: Only masking on the spatial dimension.
- **STM-MAE**: Using spatial-temporal-mixed masking.
- **w/o Mask**: Without applying any masked pre-training.

6. Masking Ratio Exploration



While an exact optimal is dataset-dependent, our results nonetheless show that relatively lower masking ratio (i.e. 25%) is preferable for spatiotemporal time series.

7. Case Study



(a) GWNet's Prediction (b) STD-MAE's Prediction

(c) GWNet's Prediction (d) STD-MAE's Prediction

Prediction under Spatiotemporal Mirages:
- **GWNet**: Without pre-training
- **STD-MAE**: With spatial-temporal-decoupled masked pre-training

GWNet exhibits a limitation under spatiotemporal mirages, which can make erroneous predictions about future trends. In contrast, STDMAE performs a significant accuracy in these situations.

QR Code for GitHub: QR Code for arXiv: QR Code for WeChat