The Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI-23)

ST-SSL

Spatio-**T**emporal **S**elf-**S**upervised **L**earning for Traffic Flow Prediction

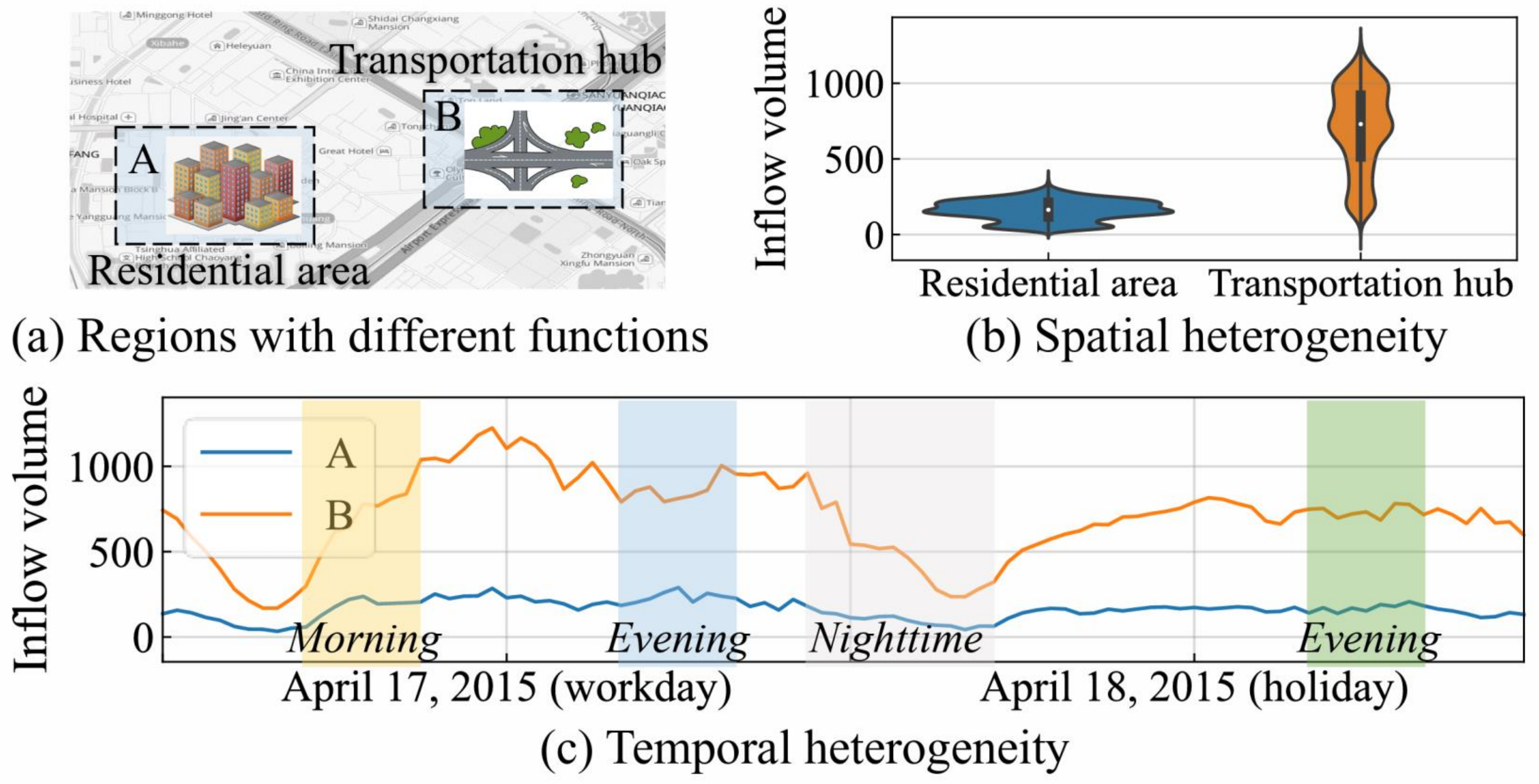
23.10.24

Presented by Yyyq

01

问题描述

- 预测的鲁棒性（Robust）不足 ⬅ 针对数据稀疏、数据尖峰
- 不考虑空间异质性（heterogeneity） ⬅ 容易偏向于交通流量较大的热门区域
- 无法捕捉时间异质性（heterogeneity） ⬅ 在所有时间段内共享参数空间



(a) Regions with different functions (b) Spatial heterogeneity

(c) Temporal heterogeneity

02 创新点

- **针对噪声干扰：** 自适应异质性感知**数据增广**方法
  - Traffic-level
  - Graph Topology-level

- **识别时间和空间异质性：** 两个自监督学习任务
  - 空间异质性建模
  - 时间异质性建模

03 算法实现：问题定义

➢ 空间区域：城市划分为 $N = I \times J$个网格

$$r_n(1 \leq n \leq N) \qquad \mathcal{V} = \{r_1, \ldots, r_N\}$$

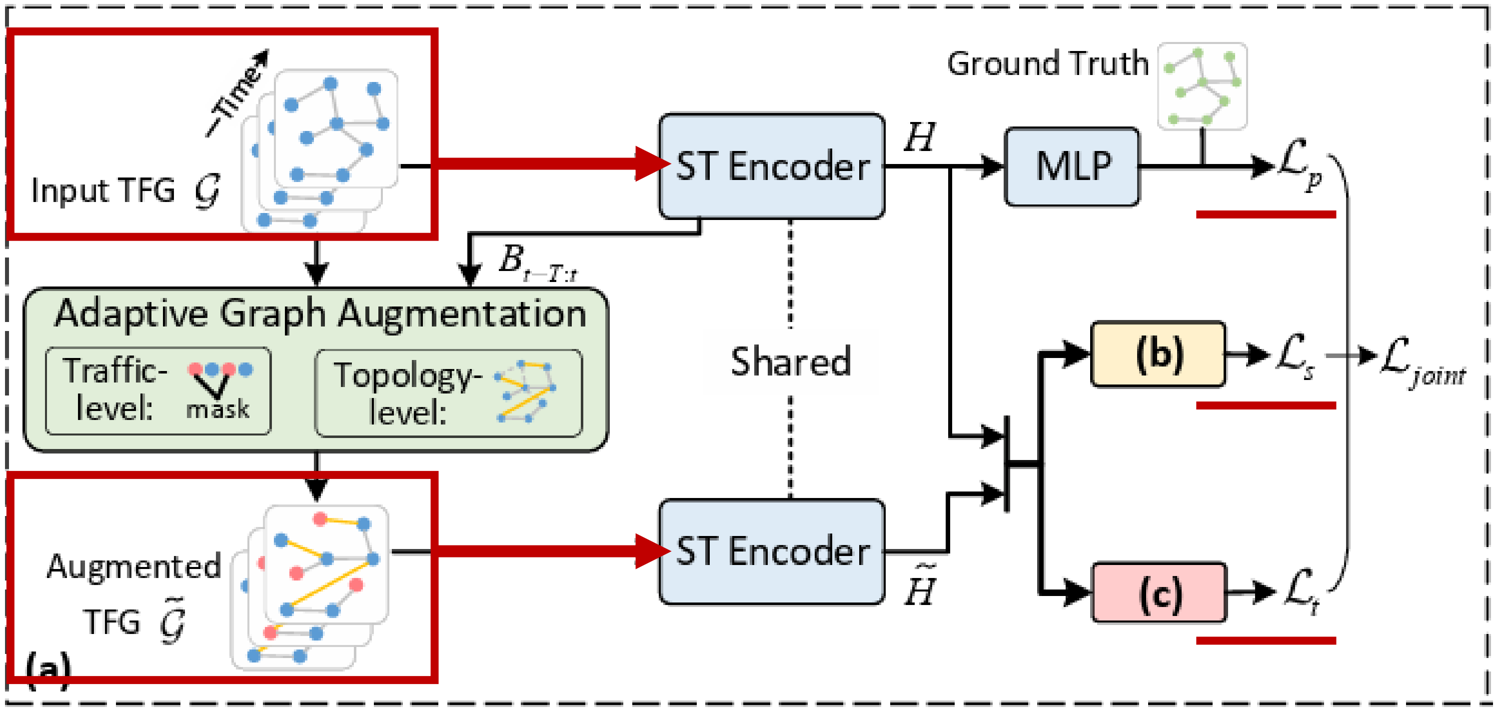
➢ 交通流量图 TFG： $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{A}, \mathcal{X}_{t-T:t})$

（流入和流出流量） $\mathcal{X}_{t-T:t} \in \mathbb{R}^{T \times N \times 2} = (\boldsymbol{X}_{t-T}, \ldots, \boldsymbol{X}_t).$

➢ 问题描述：用过去 t 个预测未来 t + 1 个时间步

03

算法实现：整体结构



Time
Input TFG $\mathcal{G}$
ST Encoder
H
MLP
Ground Truth
$\mathcal{L}_p$
$B_{t-T:t}$
Adaptive Graph Augmentation
Traffic-level:
mask
Topology-level:
Shared
(b)
$\mathcal{L}_s$
$\mathcal{L}_{joint}$
Augmented TFG $\tilde{\mathcal{G}}$
ST Encoder
$\tilde{H}$
(c)
$\mathcal{L}_t$
(a)

03 算法实现：ST-Encoder

- **时间相关性（TC）：** 具有门控机制的沿时间维度的一维因果卷积

$$\left(\boldsymbol{B}_{t-T_{out}}, \ldots, \boldsymbol{B}_t\right) = \mathrm{TC}\left(\boldsymbol{X}_{t-T}, \ldots, \boldsymbol{X}_t\right),$$

- **空间（地理）相关性（SC）：** 空间图卷积（ChebNet）

$$\boldsymbol{E}_t = \mathrm{SC}\left(\boldsymbol{B}_t, \boldsymbol{A}\right).$$

- 三明治结构：**TC → SC → TC**

- **最终输出：时间维度降为1，** $\boldsymbol{H} \in \mathbb{R}^{N \times D}$

算法实现：由异质性指导的自适应图增广

➢ 异质性度量：

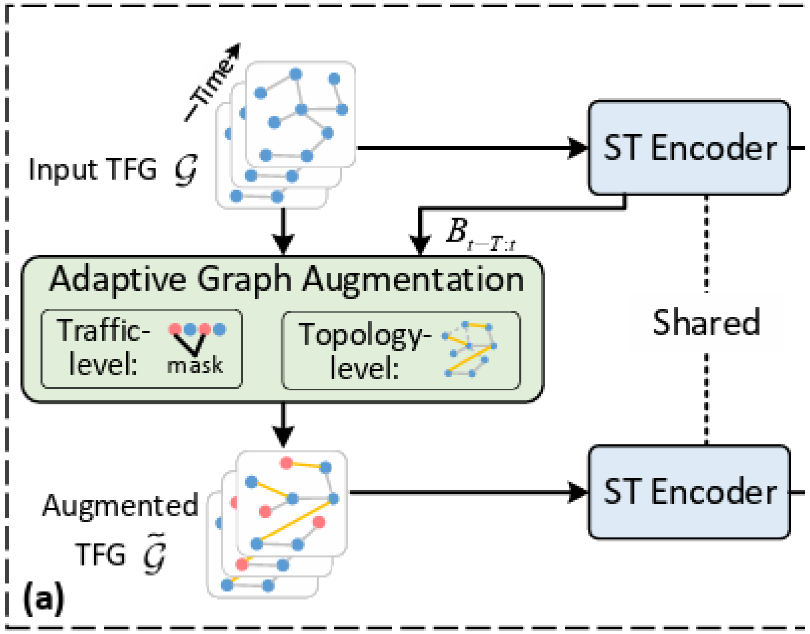
- 聚合权重

$$p_{\tau,n} = \boldsymbol{b}_{\tau,n}^{\top} \cdot \boldsymbol{w}_0.$$

$$\boldsymbol{u}_n = \sum_{\tau=t-T}^{t} p_{\tau,n} \cdot \boldsymbol{b}_{\tau,n},$$

- 两个区域之间的异质性

$$q_{m,n} = \frac{\boldsymbol{u}_m^{\top} \boldsymbol{u}_n}{\|\boldsymbol{u}_m\| \|\boldsymbol{u}_n\|}.$$



03 算法实现：数据增广

➢ **Traffic-level Augmentation：** 用伯努利概率判断是否mask掉数据扰动

$$\rho_{\tau,n} \sim Bern(1 - p_{\tau,n})$$

➢ **Graph Topology-level Augmentation：**

- 地理相邻 $\rho_{m,n} \sim Bern(1 - q_{m,n})$

- 地理不相邻 $\rho_{m,n} \sim Bern(q_{m,n})$

$$\tilde{\mathcal{G}} = (\mathcal{V}, \tilde{\mathcal{E}}, \tilde{A}, \tilde{\mathcal{X}}_{t-T:t})$$

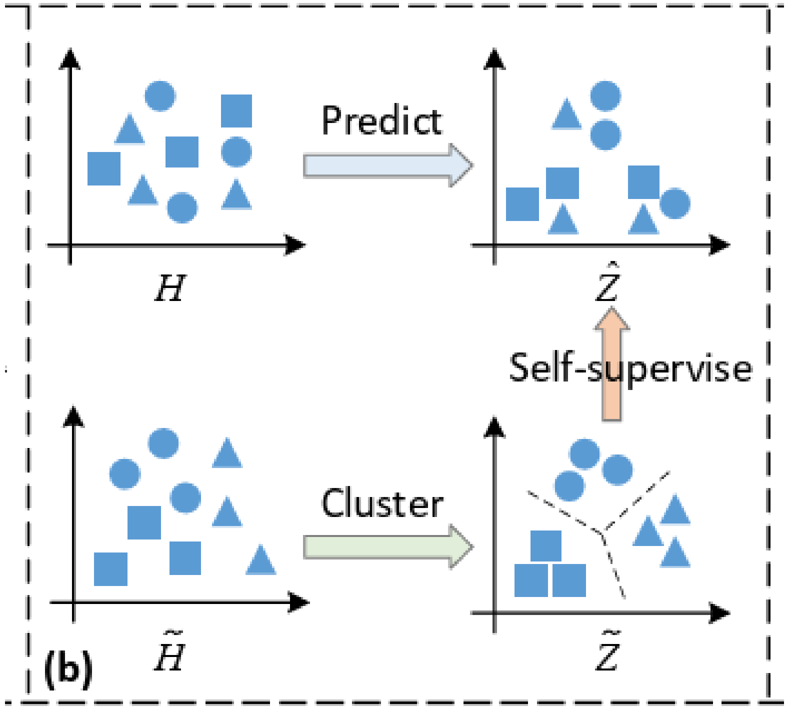
算法实现：空间异质性建模

➢ 基于软聚类的区域自监督学习(SSL)任务

- **K个聚类**

$$\{c_1, \ldots, c_K\}, \quad \tilde{z}_{n,k} = c_k^\top \tilde{h}_n$$

$$\tilde{z}_n = (\tilde{z}_{n,1}, \ldots, \tilde{z}_{n,K})$$



- 自监督学习任务

从原始TFG编码的区域嵌入预测不同城市区域功能的聚类分配。

$$\hat{z}_{n,k} = c_k^\top h_n \qquad \mathcal{L}_s = \sum_{n=1}^{N} l(h_n, \tilde{z}_n) = \sum_{n=1}^{N} \sum_k \tilde{z}_{n,k} \log \frac{\exp(\hat{z}_{n,k}/\gamma)}{\sum_j \exp(\hat{z}_{n,j}/\gamma)}$$

03 算法实现：空间异质性建模

➢ 基于软聚类的区域自监督学习(SSL)任务

- **自监督学习任务**

  从原始TFG编码的区域嵌入预测不同城市区域功能的聚类分配。

$$\hat{z}_{n,k} = c_k^\top h_n \qquad \mathcal{L}_s = \sum_{n=1}^{N} l(h_n, \tilde{z}_n) = \sum_{n=1}^{N}\sum_{k} \tilde{z}_{n,k} \log \frac{\exp\left(\hat{z}_{n,k}/\gamma\right)}{\sum_j \exp\left(\hat{z}_{n,j}/\gamma\right)}$$

- ①聚类分配矩阵不能保证每个区域的簇分配总和为1
- ②避免每个区域具有相同分配的平凡解

$$\tilde{\mathcal{Z}} = \{\tilde{Z} \in \mathbb{R}_+^{N\times K} | \tilde{Z}1_K = 1_N, \tilde{Z}1_N = \frac{N}{K}1_K\}$$

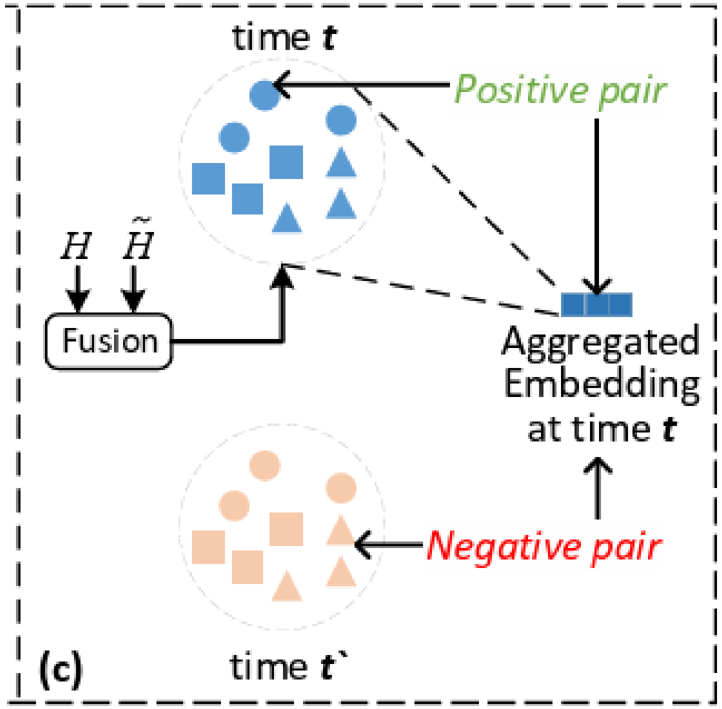
03 算法实现：时间异质性建模

$$v_{t,n} = w_i \odot h_{t,n} + w_2 \odot \tilde{h}_{t,n}$$

$$s_t = \sigma(\frac{1}{N}\sum_{n=1}^{N} v_{t,n})$$

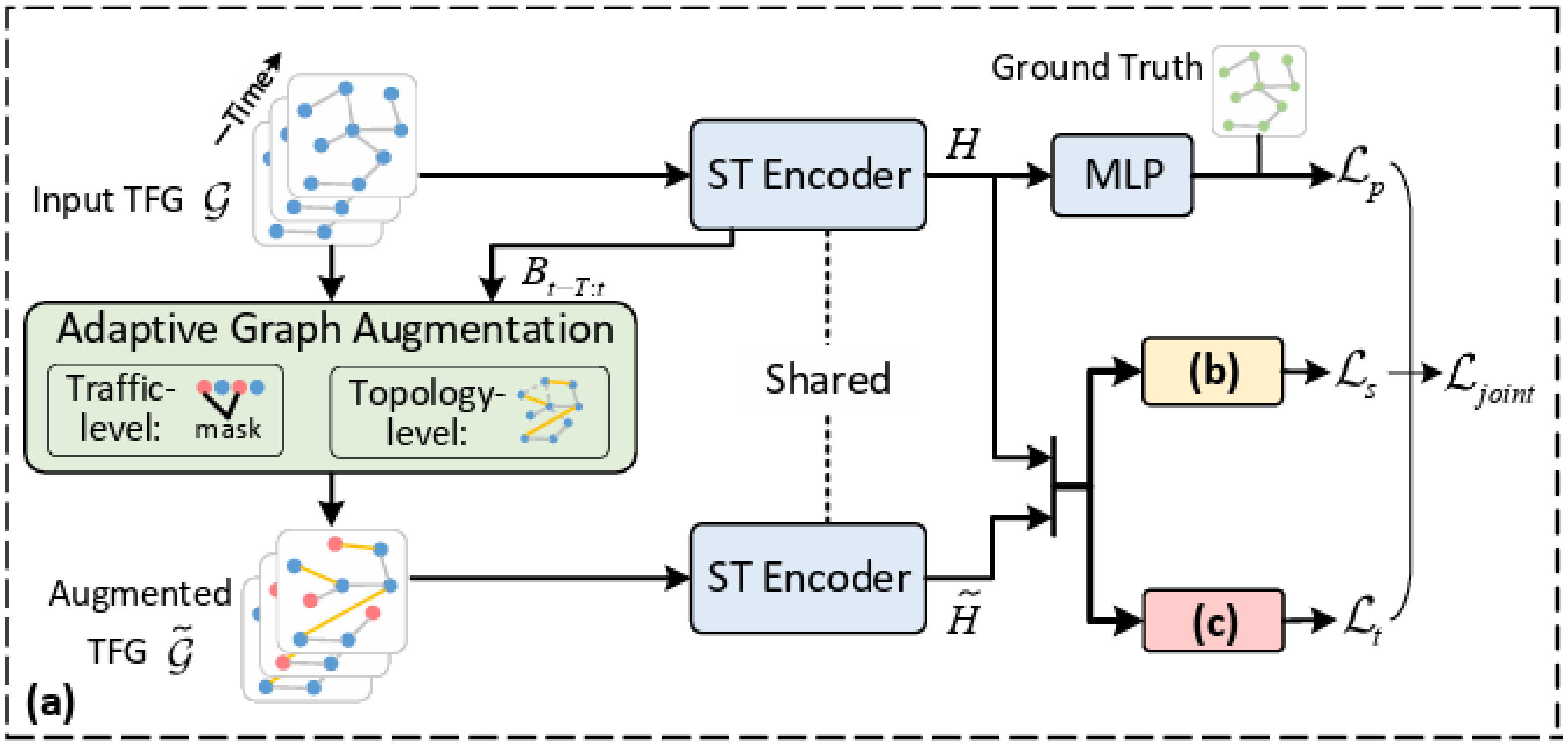
- positive pairs：同一时间步的区域级和城市级嵌入
- negative pairs：不同时间步的区域级和城市级嵌入

$$\mathcal{L}_t = -(\sum_{n=1}^{N} \log g(v_{t,n}, s_t) + \sum_{n=1}^{N} \log(1 - g(v_{t',n}, s_t)))$$



03

算法实现：训练过程



Input TFG $\mathcal{G}$
Time
Ground Truth
ST Encoder
H
MLP
$\mathcal{L}_p$
$B_{t-T:t}$
Adaptive Graph Augmentation
Traffic-level:
mask
Topology-level:
Shared
(b)
$\mathcal{L}_s$
$\mathcal{L}_{joint}$
Augmented TFG $\tilde{\mathcal{G}}$
ST Encoder
$\tilde{H}$
(c)
$\mathcal{L}_t$
(a)

04 实验部分

RQ1：与各种基线相比，ST-SSL的整体交通预测性能如何？

RQ2：设计的不同子模块如何有助于模型性能？

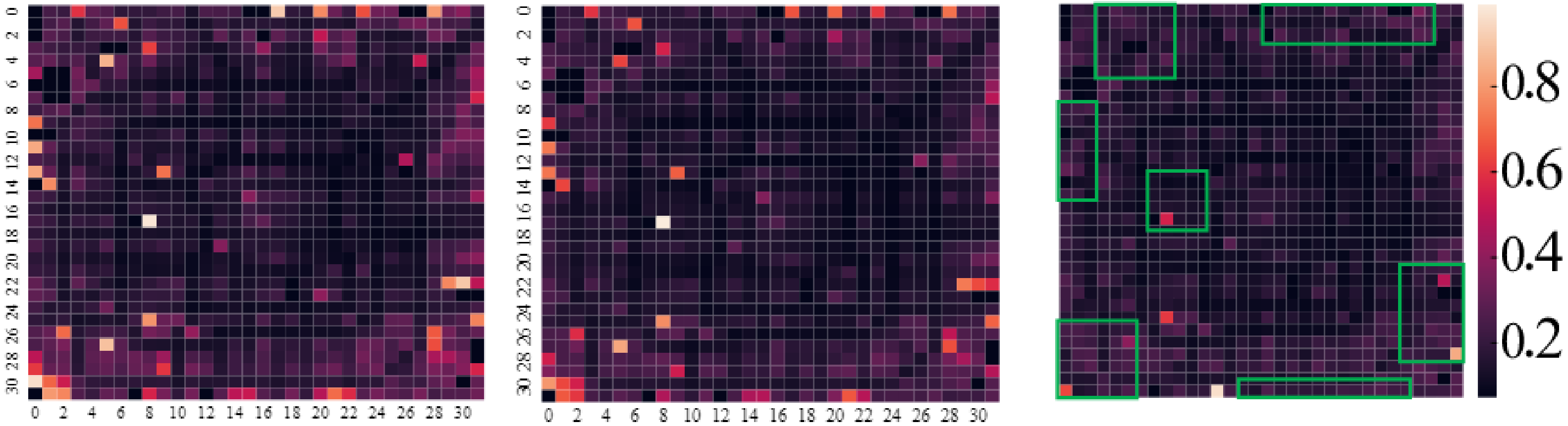
RQ3：ST-SSL在异质空间区域和不同时间段上的表现如何？

RQ4：增强图和学习表示如何使模型受益？

Data type	Bike rental		Taxi GPS	
Dataset	NYCBike1	NYCBike2	NYCTaxi	BJTaxi
Time interval	1 hour	30 min	30 min	30 min
# regions	16×8	10×20	10×20	32×32
# taxis/bikes	6.8k+	2.6m+	22m+	34k+

实验部分

Table 1: Statistics of Datasets.



(a) ST-ResNet (b) AGCRN (c) ST-SSL

04

实验部分：对比实验

Dataset	Metric	Type	ARIMA	SVR	ST-ResNet	STGCN	GMAN	AGCRN	STSGCN	STFGNN	ST-SSL
NYCBike1	MAE	In	10.66	7.27	5.53±0.06	5.33±0.02	6.77±3.42	5.17±0.03	5.81±0.04	6.53±0.10	**4.94±0.02**
		Out	11.33	7.98	5.74±0.07	5.59±0.03	7.17±3.61	5.47±0.03	6.10±0.04	6.79±0.08	**5.26±0.02**
	MAPE	In	33.05	25.39	25.46±0.20	26.92±0.08	31.72±12.29	25.59±0.22	26.51±0.32	32.14±0.23	**23.69±0.11**
		Out	35.03	27.42	26.36±0.50	27.69±0.14	34.74±17.04	26.63±0.30	27.56±0.39	32.88±0.19	**24.60±0.27**
NYCBike2	MAE	In	8.91	12.82	5.63±0.14	5.21±0.02	5.24±0.13	5.18±0.03	5.25±0.03	5.80±0.10	**5.04±0.03**
		Out	8.70	11.48	5.26±0.08	4.92±0.02	4.97±0.14	4.79±0.04	4.94±0.05	5.51±0.11	**4.71±0.02**
	MAPE	In	28.86	46.52	32.17±0.85	27.73±0.16	27.38±1.13	27.14±0.14	29.26±0.13	30.73±0.49	**22.54±0.10**
		Out	28.22	41.91	30.48±0.86	26.83±0.21	26.75±1.14	26.17±0.22	28.02±0.23	29.98±0.46	**21.17±0.13**
NYCTaxi	MAE	In	20.86	52.16	13.48±0.14	13.12±0.04	15.09±0.61	12.13±0.11	13.69±0.11	16.25±0.38	**11.99±0.12**
		Out	16.80	41.71	10.78±0.25	10.35±0.03	12.06±0.39	9.87±0.04	10.75±0.17	12.47±0.25	**9.78±0.09**
	MAPE	In	21.49	65.10	24.83±0.55	21.01±0.18	22.73±1.20	18.78±0.04	22.91±0.44	24.01±0.30	**16.38±0.10**
		Out	21.23	64.06	24.42±0.52	20.78±0.16	21.97±0.86	18.41±0.21	22.37±0.16	23.28±0.47	**16.86±0.23**
BJTaxi	MAE	In	21.48	52.77	12.12±0.11	12.34±0.09	13.13±0.43	12.30±0.06	12.72±0.03	13.83±0.04	**11.31±0.03**
		Out	21.60	52.74	12.16±0.12	12.41±0.08	13.20±0.43	12.38±0.06	12.79±0.03	13.89±0.04	**11.40±0.02**
	MAPE	In	23.12	65.51	15.50±0.26	16.66±0.21	18.67±0.99	15.61±0.15	17.22±0.17	19.29±0.07	**15.03±0.13**
		Out	20.67	65.51	15.57±0.26	16.76±0.22	18.84±1.04	15.75±0.15	17.35±0.17	19.41±0.07	**15.19±0.15**

04

实验部分：消融实验

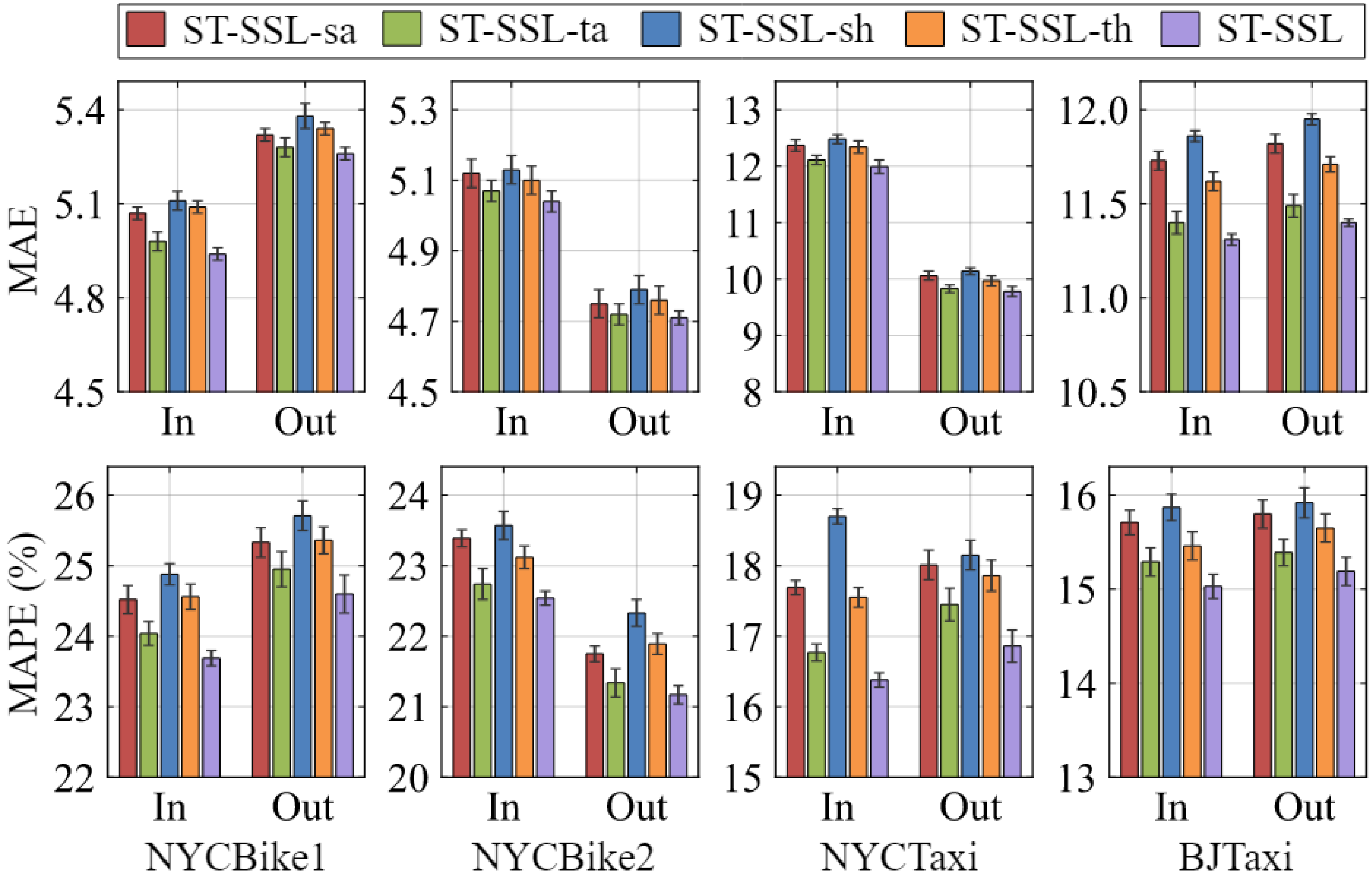
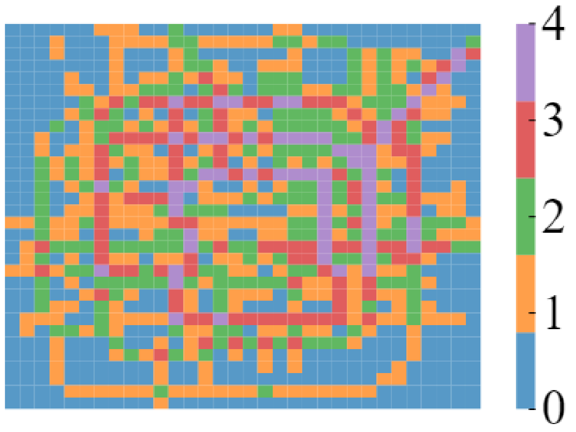


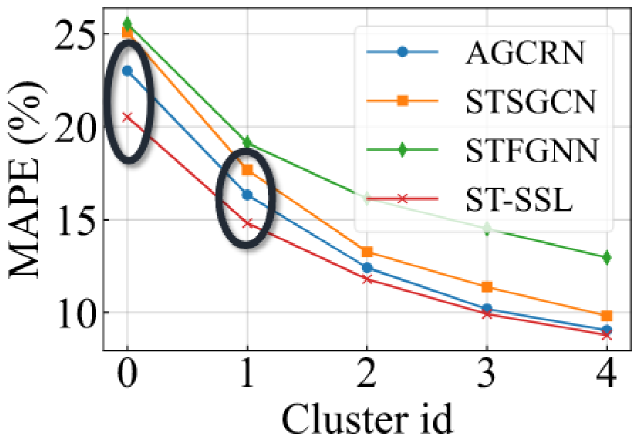
Figure 4: Ablation study of our proposed ST-SSL.

04

实验部分：对于异质性的预测性能



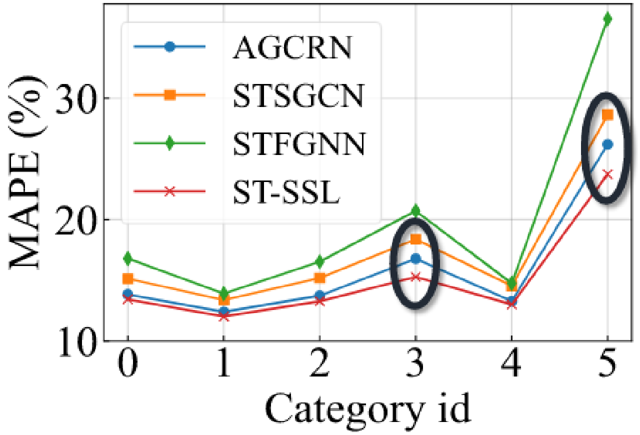
(a) Spatial clusters



(b) Spatial performance

Type	Time period	Category (id)
Workday	7:00-10:00	Morning (0)
Workday	10:00-17:00	Regular (1)
Workday	17:00-20:00	Evening (2)
Workday	20:00-7:00	Night (3)
Holiday	9:00-22:00	Day (4)
Holiday	22:00-9:00	Night (5)

(c) Temporal categories



(d) Temporal performance

04

实验部分：鲁棒性

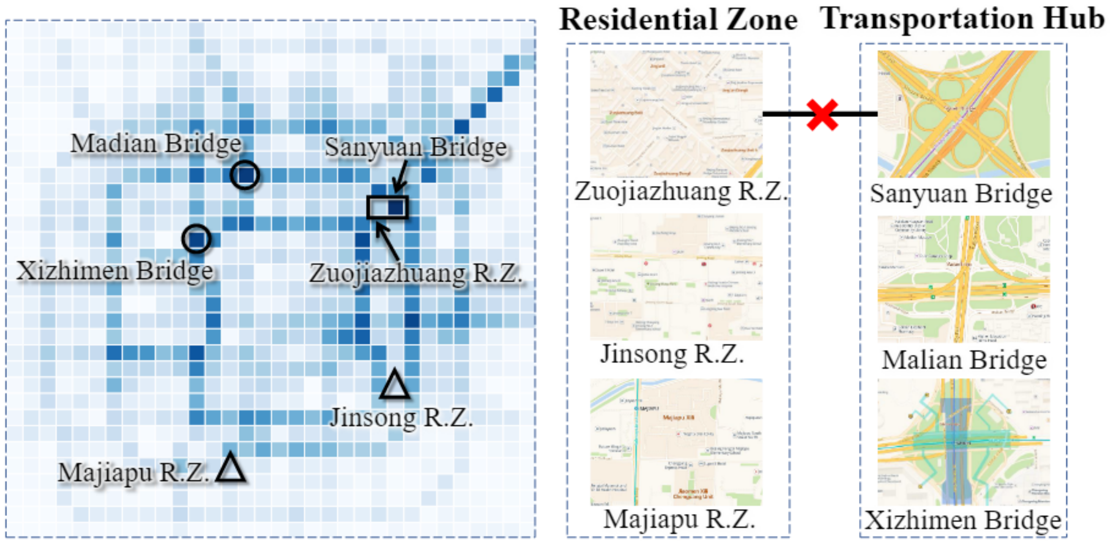


Figure 6: Case study on the adaptive graph augmentation.

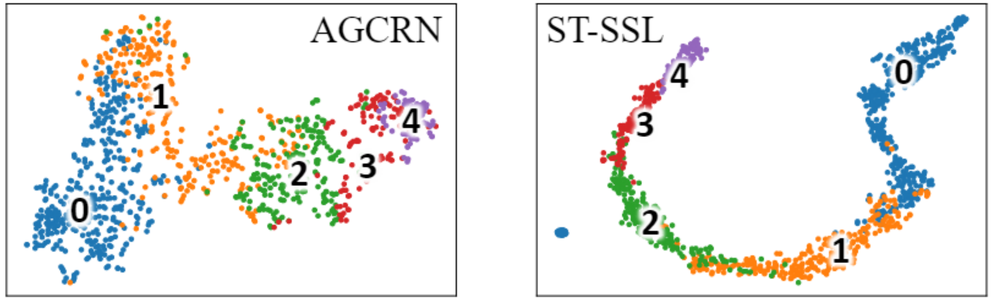


Figure 7: t-SNE visualization of embeddings on BJTaxi.

谢谢观看

MANY THANKS !

23.10.24