数据智能实验室
Data Intelligence Lab

NEURAL INFORMATION PROCESSING SYSTEMS

# Get Rid of Task Isolation: A Continuous Multi-task Spatio-Temporal Learning Framework

Zhongchao Yi   Zhengyang Zhou*   Qihe Huang

Yanjiang Chen   Liheng Yu   Xu Wang   Yang Wang*

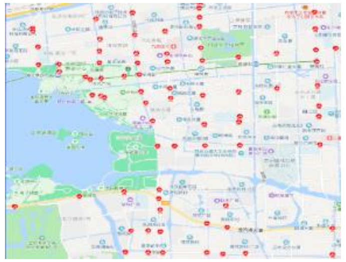
# Motivations

## ➢ Exploring ST Intelligence across Tasks

**OOD Generalization for ST Prediction**
- Temporal ➔ Distribution shift over time series (distribution shift)
- Spatial ➔ Model transfer across cities (inter-city transfer, …)


Remote sense


Trajectory


Road sensor

Traffic accident


**Multi-task intelligence**

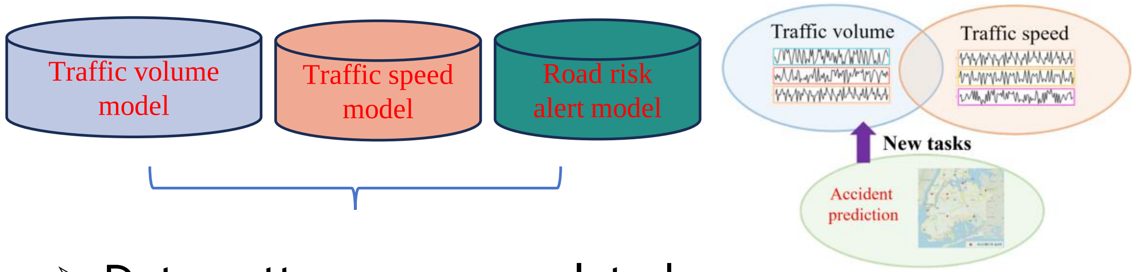
**A unified model for fast transfer, fine-tuning on different tasks of the same ST domain is highly required !**

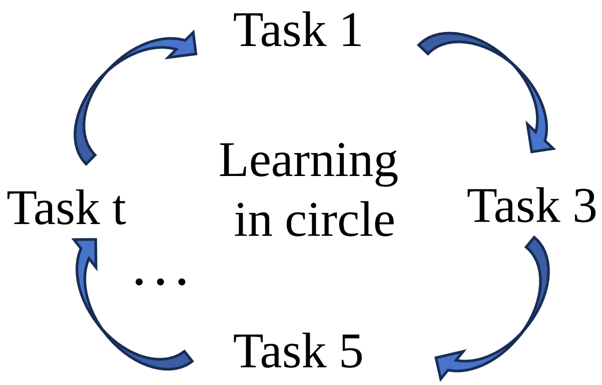
- ✓ **Common ST associations for cold-start and challenging tasks**
- ✓ **Without re-training model for a different task**

➔ **Task-guided Model Evolution Green ST computing**

# Challenges & Solutions

## ➢ Challenges for ST Task Learning




- ➢ Data patterns are correlated
- ➢ Models are isolated, not shared and interactive
- ➢ Cold-start challenges for new tasks

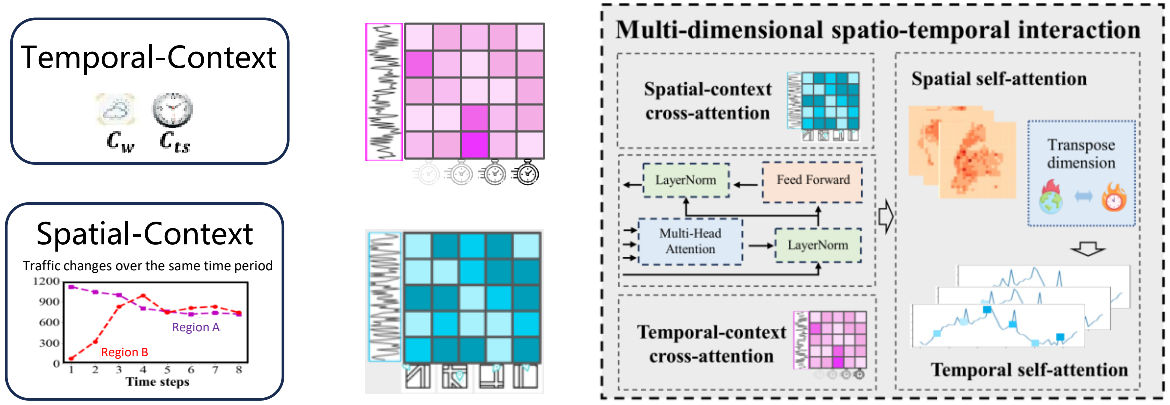
- ❖ **How to capture commonalities among tasks ?**
- ❖ **How to utilize commonalities and personalities to enhance each individual tasks and new tasks ?**

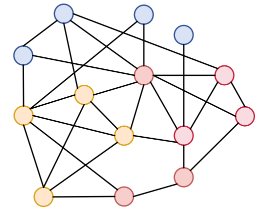
- ✓ **Cross-interaction of different dimensions of data**
- ✓ **Capture the model behavior through rolling training, and decouple the common/individual patterns**

# Continuous Multi-task SpatioTemporal (CMuST)

## ➢ ST Intelligence for Task Continuous Learning

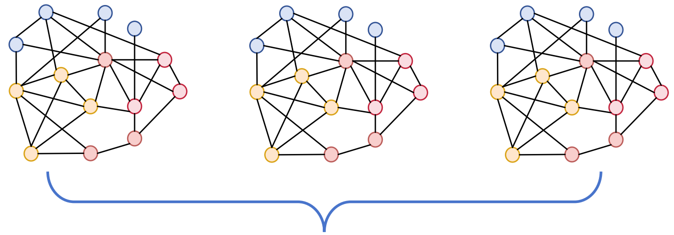
### ➢ Multi-dimensional Spatio-Temporal Interaction (MSTI)



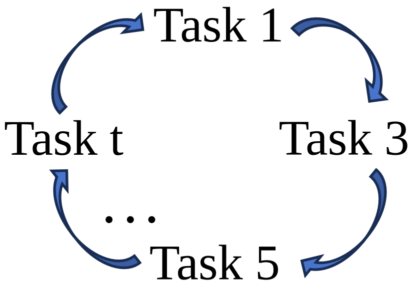
More informative and dimension-level relations encapsulated

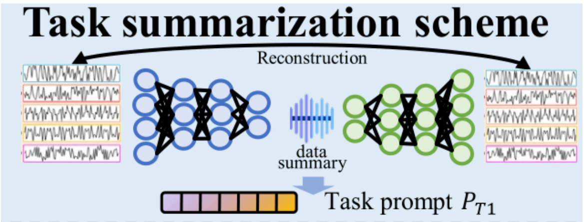
Multi-dimensional, multi-perspective spatial and temporal interaction



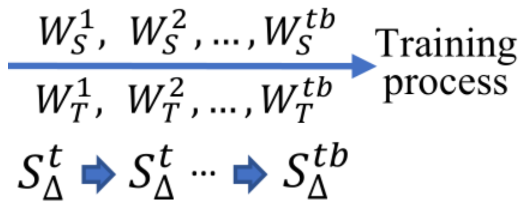
Ability to capture common dependencies on data dimensions across tasks

### ➢ Rolling Adaptation (RoAda)



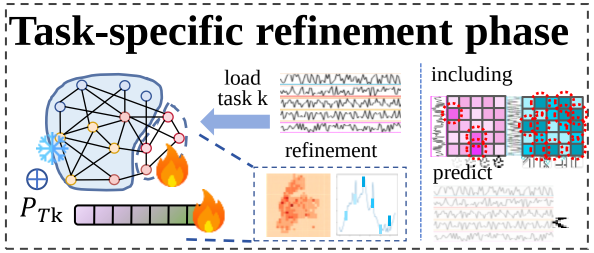




Task personality profiling based on data reconstruction




$$\mathbf{W}^{(\mathcal{T}_2)}_{\text{stable}}, \mathbf{W}^{(\mathcal{T}_2)}_{\text{dynamic}} = \{w_{ij} \in \mathcal{W}^{(\mathcal{T}_2)} : \mathrm{Var}(w_{ij}) < \delta\}, \{w_{ij} \in \mathcal{W}^{(\mathcal{T}_2)} : \mathrm{Var}(w_{ij}) \geq \delta\}$$

Model behavior modeling, decoupling task-level common patterns

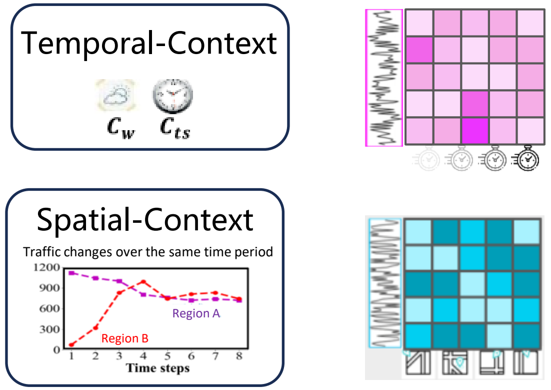


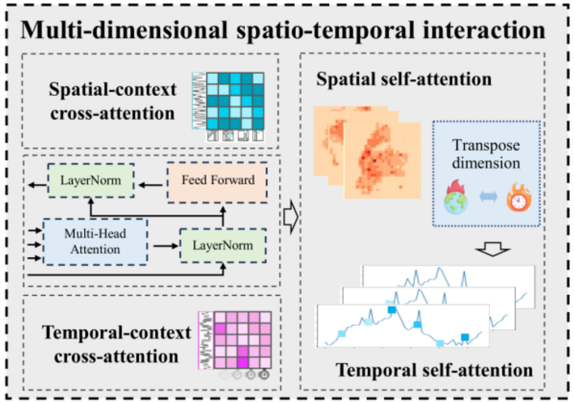

Task-specific fine-tuning

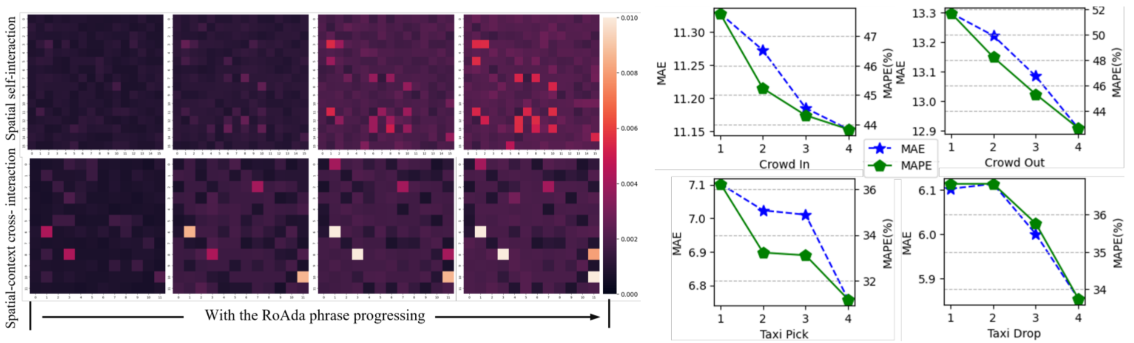
# Continuous Multi-task SpatioTemporal (CMuST)

## ➢ ST Intelligence for Task Continuous Learning

### ➢ Multi-dimensional Spatio-Temporal Interaction (MSTI)

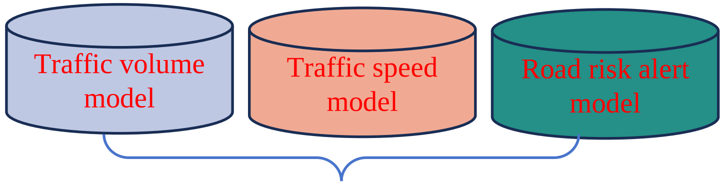


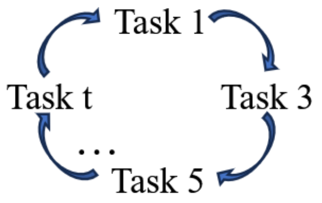

(a) Visualizing attention across training phases.

(b) Performance variation along with task increasing.

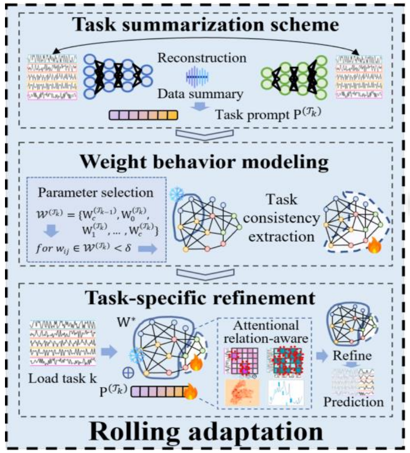


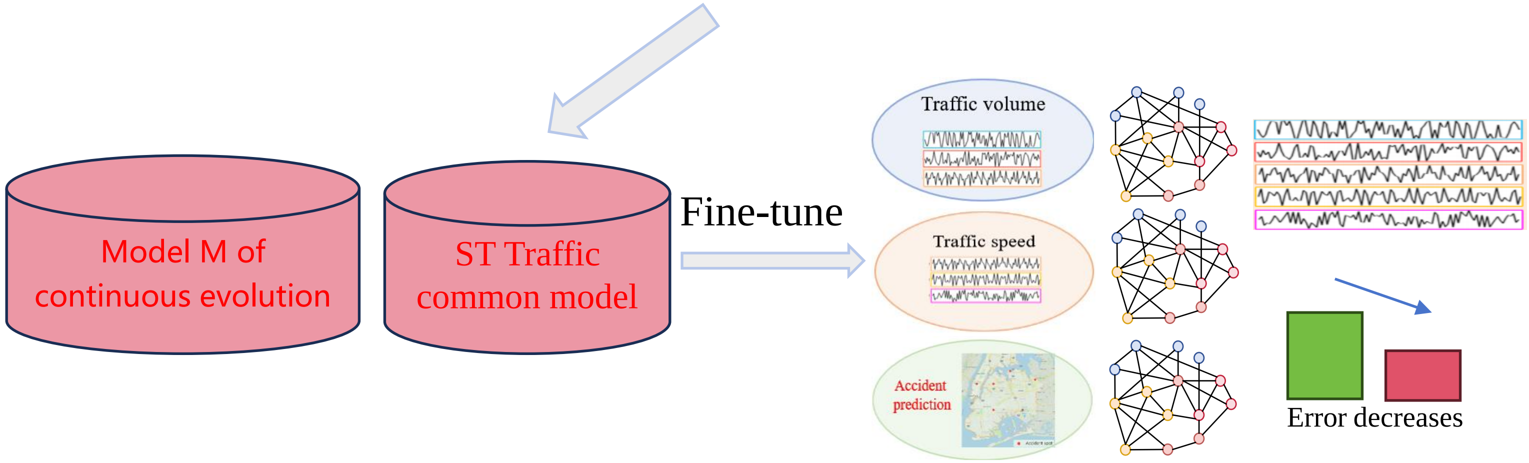

*Collective intelligence*

### ➢ Rolling Adaptation (RoAda)




| City | Task | #Records | Time Span | #Regions |
|---|---|---|---|---|
| NYC | Taxi Drop<br>Taxi Pick<br>Crowd In<br>Crowd Out | 15,245k<br>12,748k<br>17,231k<br>16,493k | 01/01/2016-<br>03/31/2016 | 206 |
| SIP | Traffic Flow<br>Traffic Speed | 1,237k<br>307k | 01/01/2017-<br>03/31/2017 | 108 |
| Chicago | Taxi Drop<br>Taxi Pick<br>Risk | 1,326k<br>1,289k<br>61k | 06/01/2023-<br>12/31/2023 | 220 |



# Experiments

## ➢ Datasets

- **NYC**: Includes three months of crowd flow and taxi hailing from Manhattan and its surrounding areas in New York City. Tasks: *Crowd In*, *Crowd Out*, *Taxi Pick*, and *Taxi Drop*.
- **SIP**: Contains records of *Traffic Flow* and *Traffic Speed* within Suzhou Industrial Park over a period of three months.
- **Chicago**: Comprises of traffic data collected in the second half of 2023 from Chicago, including three tasks: *Taxi Pick*, *Taxi Drop*, and *Risk*.

| City | Task | #Records | Time Span | #Regions | #Time Steps | Time Interval |
|---|---|---|---|---|---|---|
| NYC | Taxi Drop<br>Taxi Pick<br>Crowd In<br>Crowd Out | 30,245k | 01/01/2016-<br>03/31/2016 | 206 | 4368 | 30mins |
| SIP | Traffic Flow<br>Traffic Speed | 1,237k<br>307k | 01/01/2017-<br>03/31/2017 | 108 | 25920 | 5mins |
| Chicago | Taxi Drop<br>Taxi Pick<br>Risk | 3,291k<br>61k | 06/01/2023-<br>12/31/2023 | 220 | 10272 | 30mins |

## ➢ Comparison with existing ST models

| Datesets | | NYC | | | | SIP | | Chicago | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Methods | Metrics | Crowd In | Crowd Out | Taxi Pick | Taxi Drop | Traffic Flow | Traffic Speed | Taxi Pick | Taxi Drop | Risk |
| DCRNN | MAE | 17.5289 | 19.5667 | 10.8188 | 9.6142 | 12.5326 | 0.7044 | 3.0624 | 2.5793 | 1.1174 |
| | MAPE | 0.5939 | 0.5695 | 0.4330 | 0.4818 | 0.2455 | 0.2686 | 0.4237 | 0.4816 | 0.2504 |
| AGCRN | MAE | 11.5135 | 13.1569 | 7.0675 | 6.0066 | 15.8319 | 0.6924 | 2.3542 | 2.0884 | 1.1183 |
| | MAPE | 0.5094 | 0.4773 | 0.3753 | 0.3665 | 0.2926 | 0.2744 | 0.4092 | 0.4046 | 0.2505 |
| GWNET | MAE | 11.4420 | 13.2992 | 7.0701 | 6.1171 | 13.0529 | 0.6900 | 2.3671 | 2.0434 | 1.1197 |
| | MAPE | 0.4778 | 0.6171 | 0.3713 | 0.3514 | 0.2483 | 0.2655 | 0.3912 | 0.4044 | 0.2514 |
| STGCN | MAE | 11.3766 | 13.3522 | 7.1259 | 5.9268 | 15.3501 | 0.7111 | 2.3781 | 2.1427 | 1.1184 |
| | MAPE | 0.5018 | 0.4318 | 0.3234 | **0.3339** | 0.3041 | 0.2660 | 0.4074 | 0.4331 | 0.2507 |
| GMAN | MAE | 11.3414 | 13.1923 | 7.0662 | 6.0912 | 13.0368 | 0.6952 | 2.3663 | 2.0316 | 1.1182 |
| | MAPE | 0.4782 | 0.6065 | 0.3652 | 0.3468 | 0.2464 | 0.2678 | 0.3953 | 0.4036 | 0.2516 |
| ASTGCN | MAE | 14.2847 | 17.1582 | 9.1430 | 7.7063 | 16.4896 | 0.6980 | 2.5091 | 2.1520 | 1.1175 |
| | MAPE | 0.6396 | 0.5922 | 0.4607 | 0.4524 | 0.3104 | 0.2682 | 0.4593 | 0.4413 | **0.2502** |
| STTN | MAE | 12.1994 | 14.1966 | 7.6716 | 6.3816 | 15.1751 | 0.6939 | **2.2996** | 2.0355 | 1.1214 |
| | MAPE | 0.4757 | 0.4744 | 0.3600 | 0.3763 | 0.2881 | 0.2625 | 0.3893 | 0.4133 | 0.2518 |
| MTGNN | MAE | 11.4350 | 13.3072 | 7.0736 | 6.1162 | 13.0486 | 0.6989 | 2.3692 | 2.0361 | 1.1201 |
| | MAPE | 0.4785 | 0.6185 | 0.3782 | 0.3502 | 0.2475 | 0.2687 | 0.3979 | 0.4073 | 0.2578 |
| STEP | MAE | 11.2328 | 13.1043 | 6.9619 | 5.9101 | 12.0032 | 0.6970 | 2.3592 | 2.0168 | 1.1190 |
| | MAPE | 0.4537 | 0.4361 | 0.3248 | 0.3379 | 0.2391 | 0.2638 | 0.3914 | 0.4019 | 0.2507 |
| PromptST | MAE | **11.0036** | 13.0237 | 6.8711 | 5.8797 | 11.8620 | 0.6921 | 2.3576 | 2.0065 | 1.1186 |
| | MAPE | 0.4465 | 0.4358 | 0.3265 | 0.3382 | 0.2375 | 0.2632 | 0.3913 | 0.4012 | 0.2511 |
| CMuST | MAE | 11.1533 | **12.9088** | **6.7581** | **5.8546** | **11.5811** | **0.6843** | 2.3264 | **2.0034** | **1.1172** |
| | MAPE | **0.4384** | **0.4265** | **0.3118** | 0.3375 | **0.2279** | **0.2585** | **0.3872** | **0.4009** | 0.2503 |

## ➢ Experiment Design

- **Single-task learning:** Treat different task sets as separate datasets to train and test the models respectively, train 5 times and take the average results.
- **Multi-task learning:** Align the features of different types of data in the same city into the same graph, and concatenates the data features for training the model.

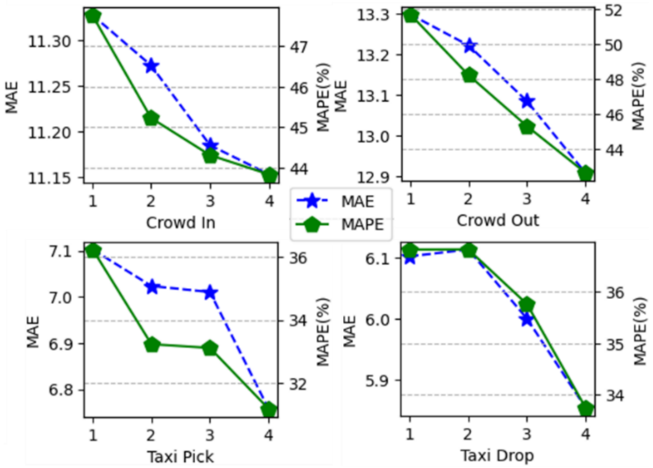
# Experiments & Conclusion

## ➢ Robustness in data-scarce scenarios

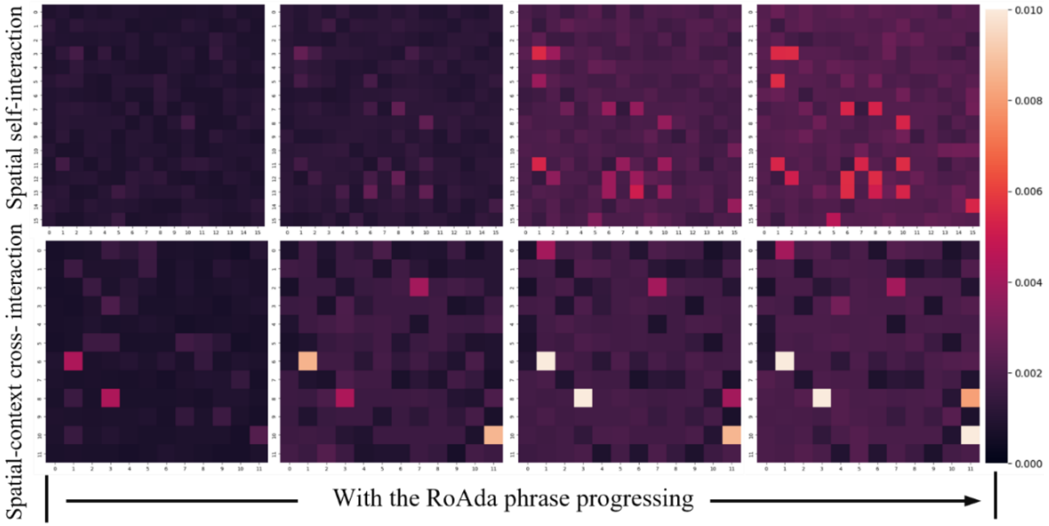
Reducing part of the spatial nodes, as well as extending the time interval to reduce the number of samples, to investigate the robustness in challenging with limited data.

| model | NYC for Crowd In | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 25% nodes | | 50% nodes | | 2 times interval | | 4 times interval | |
| | MAE | MAPE | MAE | MAPE | MAE | MAPE | MAE | MAPE |
| GWNET | 13.7648 | 0.4825 | 12.4637 | 0.4731 | 20.2547 | 0.4465 | 20.6487 | 0.4958 |
| STEP | 13.1827 | 0.4772 | 12.2393 | 0.4612 | 20.1936 | 0.4436 | 20.1465 | 0.4915 |
| PromptST | 12.8362 | 0.4719 | 12.0361 | 0.4607 | 19.8465 | 0.4384 | 19.5238 | 0.4872 |
| **CMuST** | **12.1611** | **0.4506** | **11.2864** | **0.4470** | **18.2925** | **0.4279** | **18.4084** | **0.4797** |

## ➢ Performance with task increasing




## ➢ Visualizing attention weights




## ➢ Contributions

- ✓ The first continuous multi-task spatiotemporal learning framework, reinforcing individual correlated learning task in collective perspective.
- ✓ Propose two innovative learning modules, MSTI and RoAda, enabling common and individual pattern extraction.
- ✓ Construct multi-task ST learning benchmarks for three cities.

# Thank You for Listening

## References


Zhengyang Zhou et al. (2023). "Maintaining the status quo: Capturing invariant relations for ood spatiotemporal learning." In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 3603-3614.

Qihe Huang et al. (2024). "Leret: Language-empowered retentive network for time series forecasting." In: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, pp. 4165-4173.

Zezhi Shao et al. (2022). "Spatial-temporal identity: A simple yet effective baseline for multivariate time series forecasting." In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, pp. 4454-4458.



**Contact**

Zhongchao Yi: zhongchaoyi@mail.ustc.edu.cn
Zhengyang Zhou: zzy0929@ustc.edu.cn
Yang Wang: angyan@ustc.edu.cn