

# Non-Aligned Multi-Scale Data Completion for Sparse Mobile CrowdSensing



Wenbin Liu<sup>1,2</sup>, Hao Du<sup>1,2</sup>, En Wang<sup>1,2</sup>, Dongming Luan<sup>1,2</sup>,  
Bo Yang<sup>1,2</sup>, Yongjian Yang<sup>1,2</sup> and Jie Wu<sup>3</sup>

1 College of Computer Science and Technology, Jilin University

2 Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education

3 Telecom Cloud Computing Research Institute



# Outline



- I. Background and Motivation
- II. Problem and Workflow
- III. Method
- IV. Performance Evaluation
- V. Conclusion

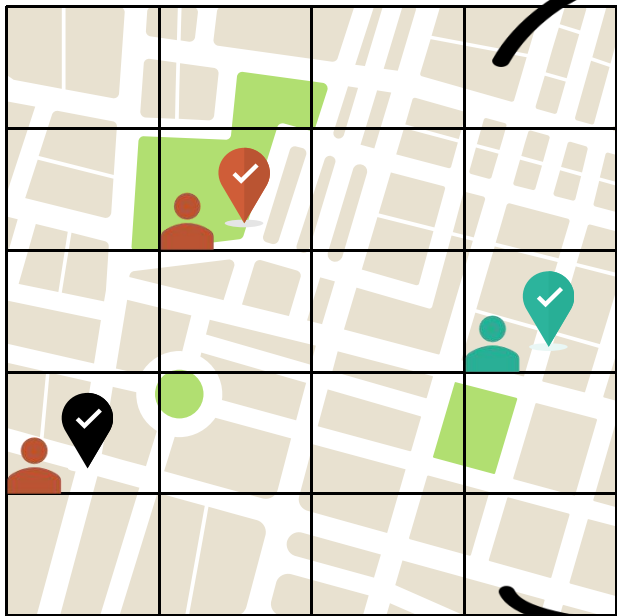


# Outline



- I. **Background and Motivation**
- II. Problem and Workflow
- III. Method
- IV. Performance Evaluation
- V. Conclusion

- Recruit users with mobile devices to perform various sensing tasks



Traffic Monitoring



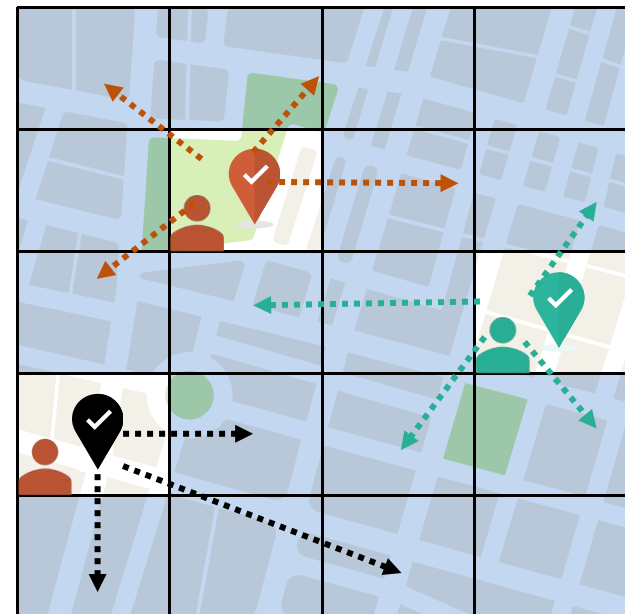
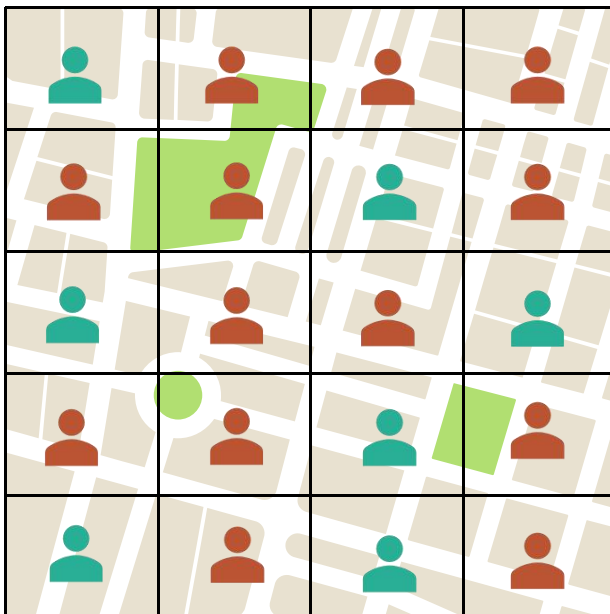
Air Quality Sensing

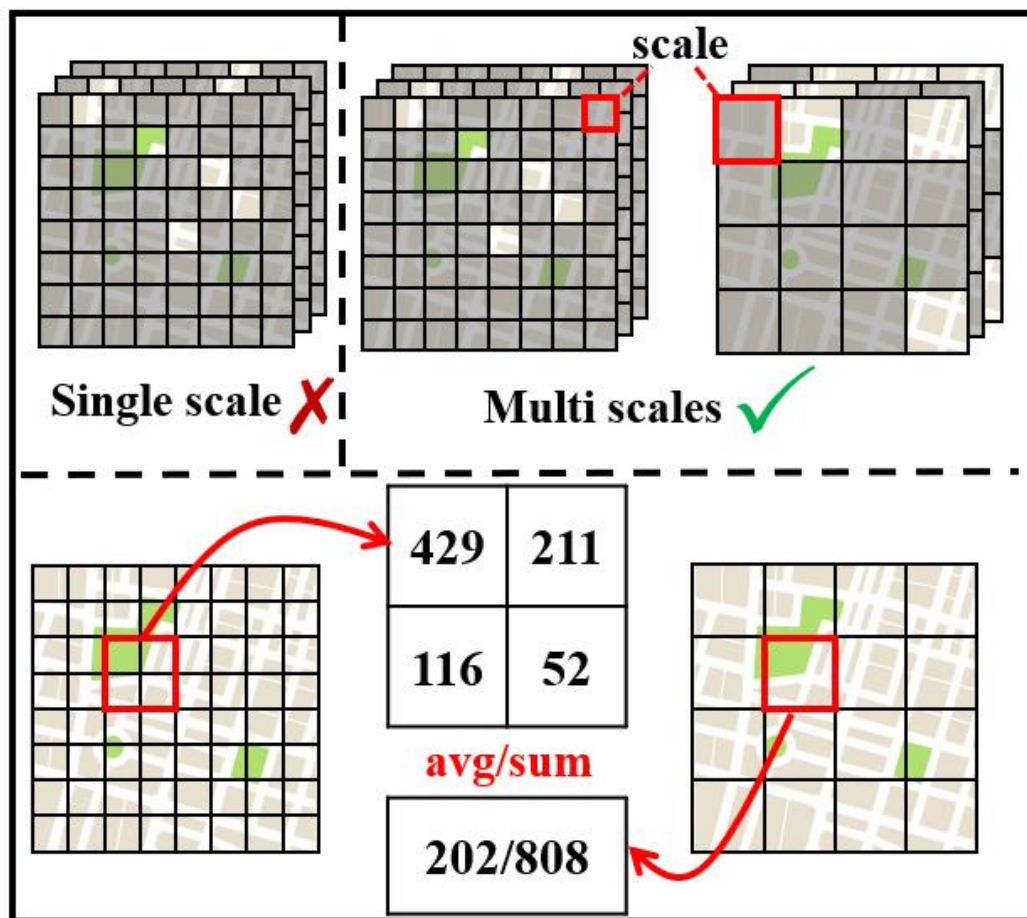
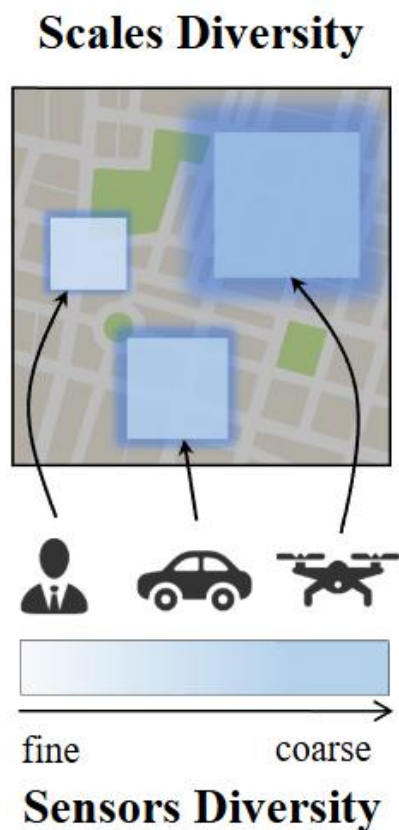


Crowdsourced Parking

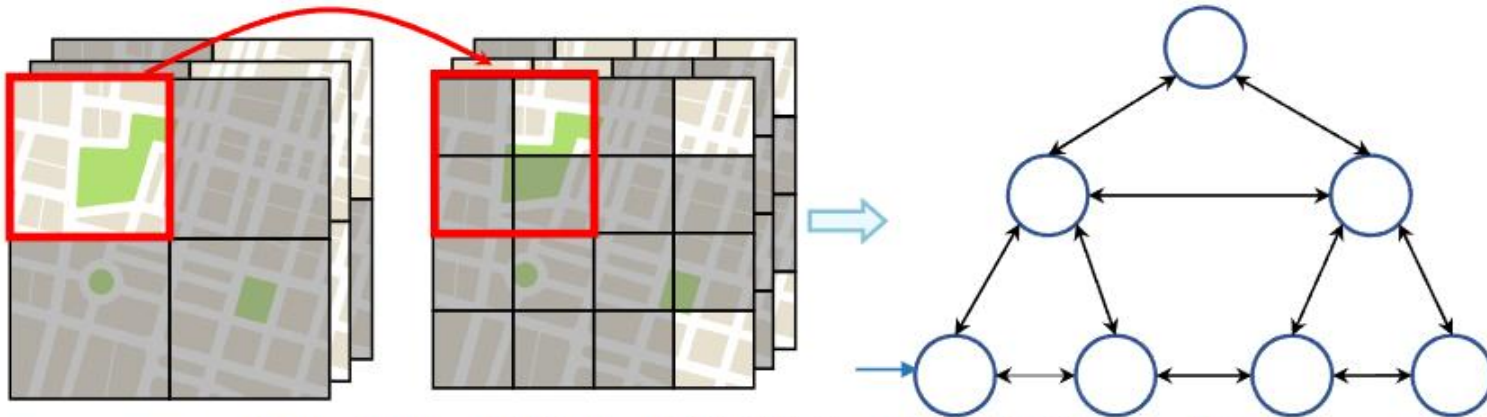


- MCS: a large number of workers
- Sparse MCS: sense a few and infer the rest



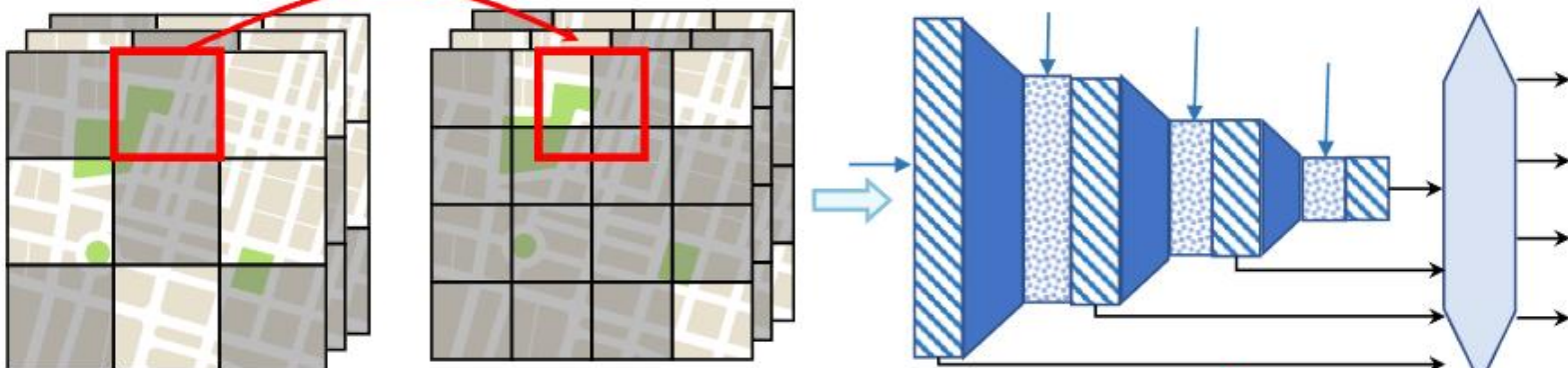


# Aligned vs. Non-aligned



Previous (Aligned Multi-Scale Data)

X



Our (Non-Aligned Multi-Scale Data)

✓



# Challenge



- How to handle multi-scale data with non-aligned compositional relationships?
- How to effectively capture intra-scale and inter-scale correlations?
- How to construct a lightweight and effective model for handling multi-scale data?



# Outline



- I. Background and Motivation
- II. Problem and Workflow**
- III. Method
- IV. Performance Evaluation
- V. Conclusion

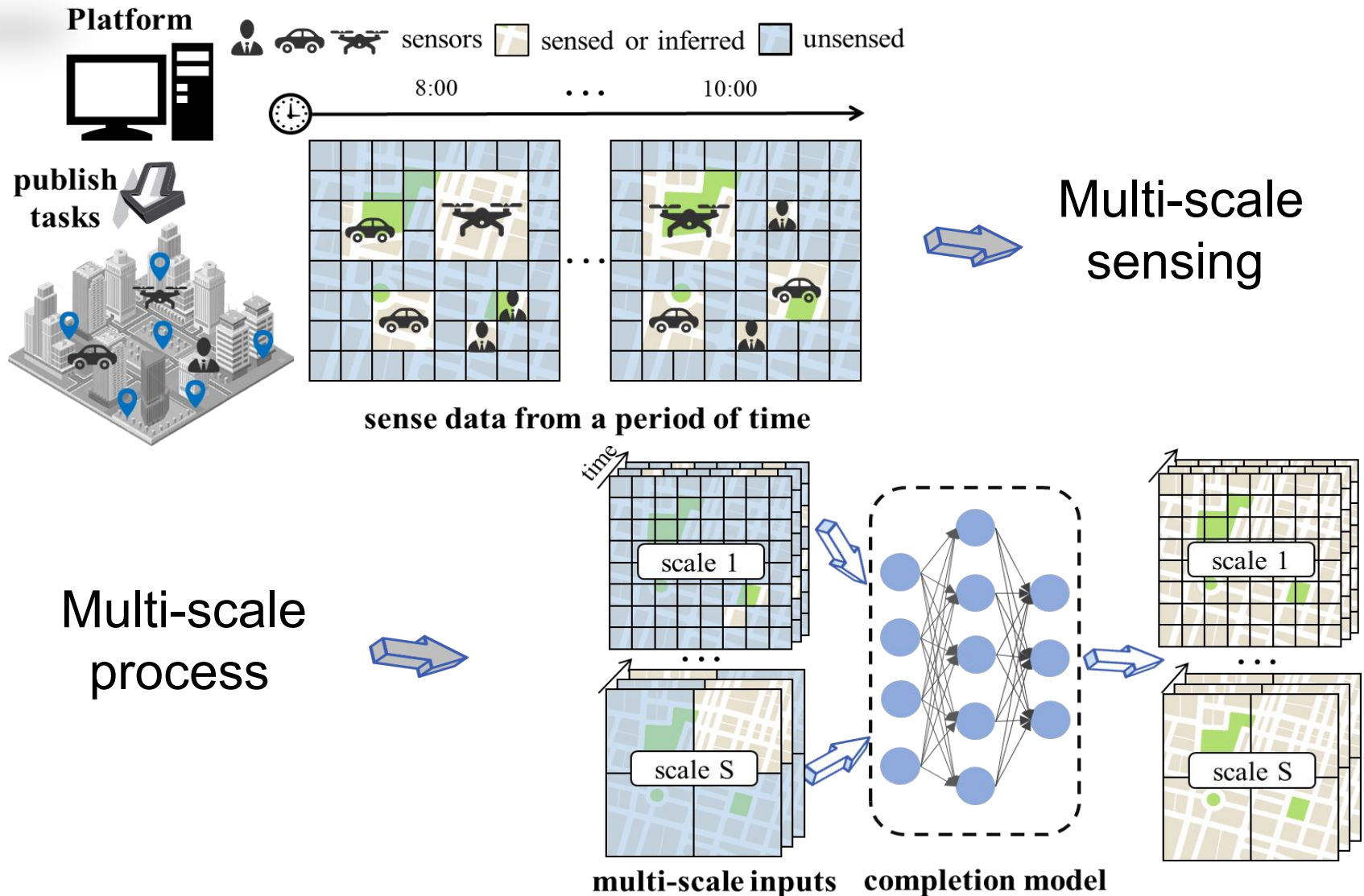


# Problem Formulation



- **Problem** [Non-Aligned Multi-Scale Data Completion]:
- Given  $T$  time slices,  $S$  different scales of data, and  $H*W$  subareas of the finest scale size, we aim to sense data from a limited number of subareas across different scales and utilize this data to reconstruct complete data

$$\min \delta(\hat{\mathbf{Y}}^{(1)}, \mathbf{Y}^{(1)}) = \sum_{i=1}^T \sum_{j=1}^H \sum_{k=1}^W \left| y_{i,j,k}^{(1)} - \hat{y}_{i,j,k}^{(1)} \right|$$

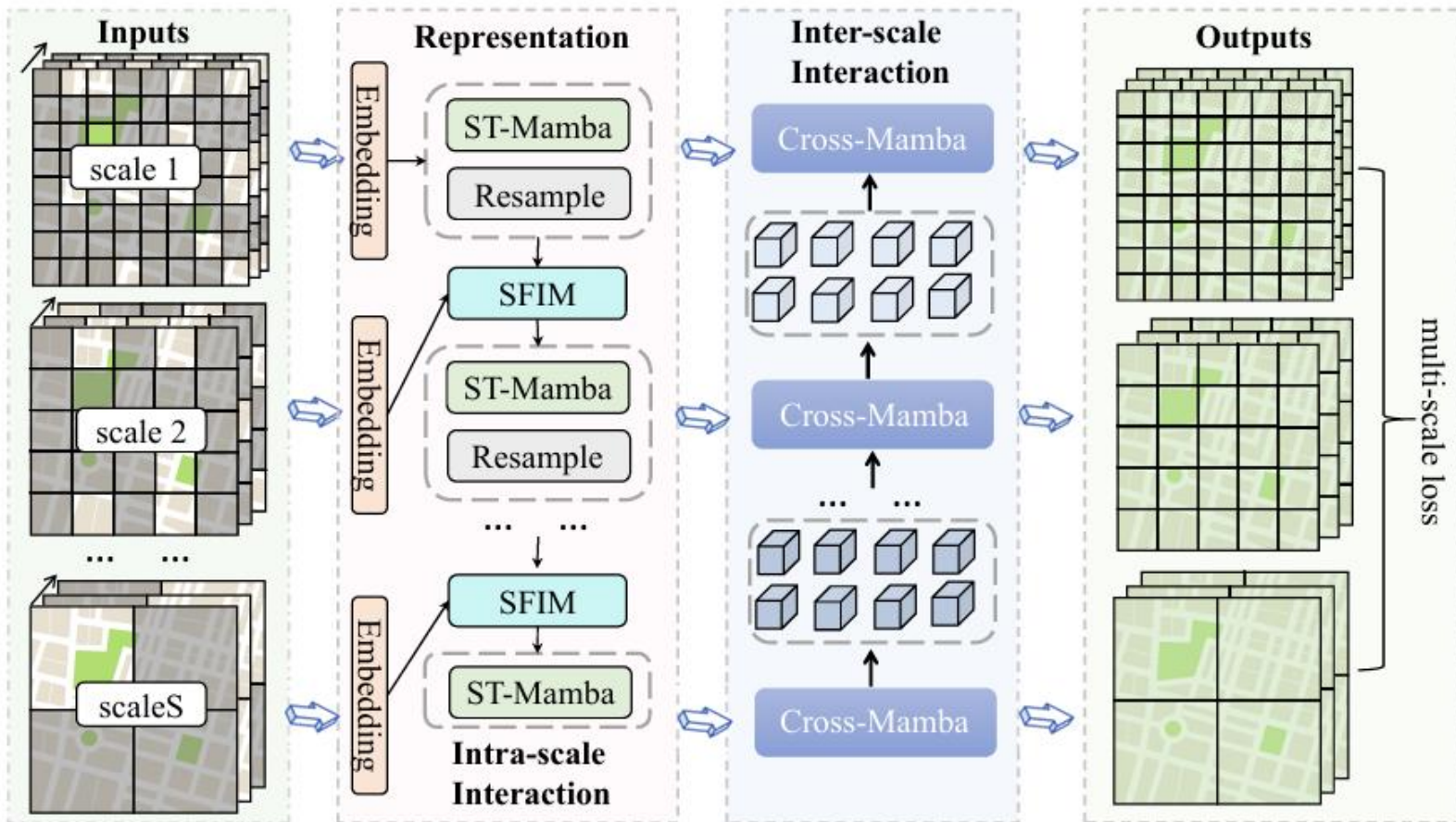


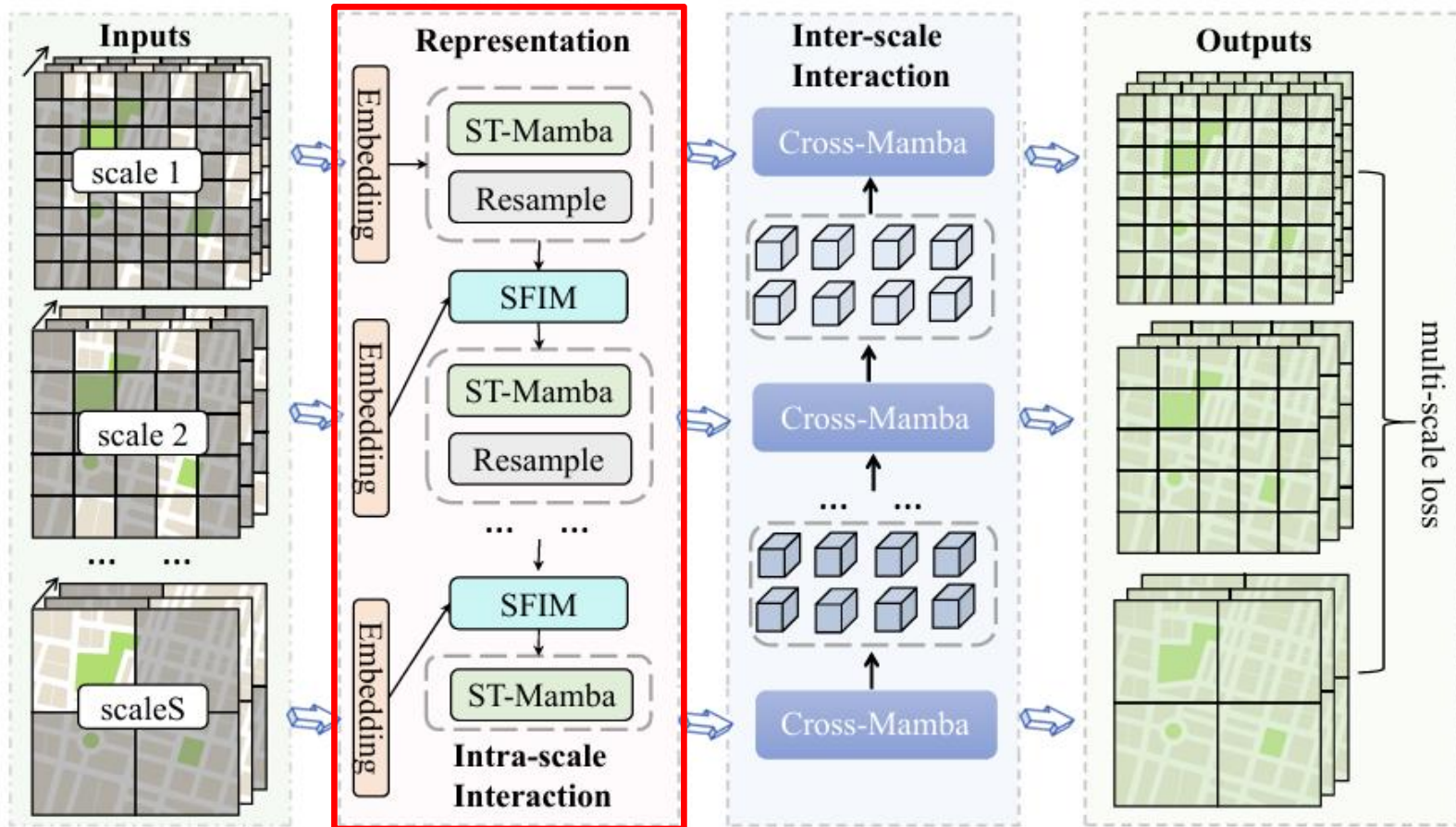


# Outline

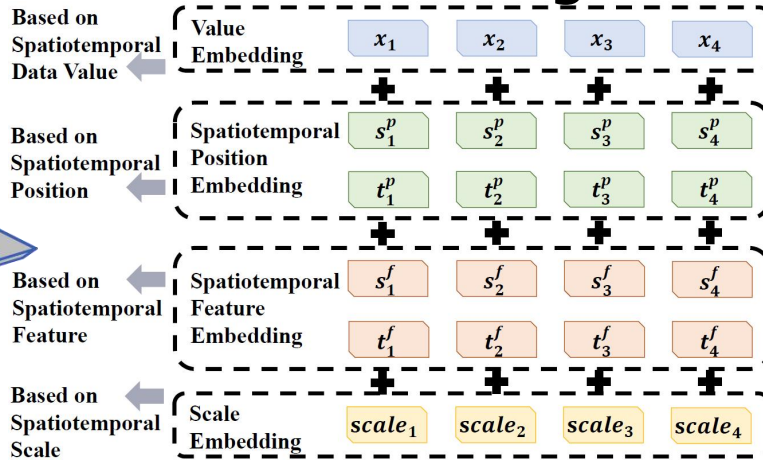


- I. Background and Motivation
- II. Problem and Workflow
- III. Method**
- IV. Performance Evaluation
- V. Conclusion

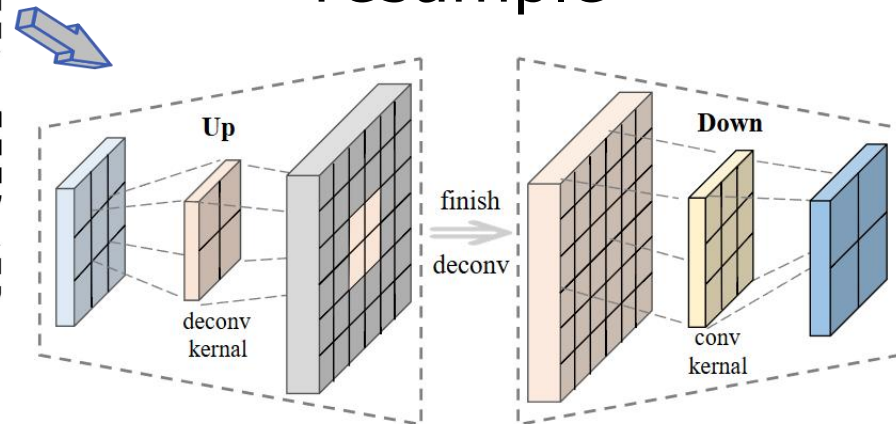




## Embedding

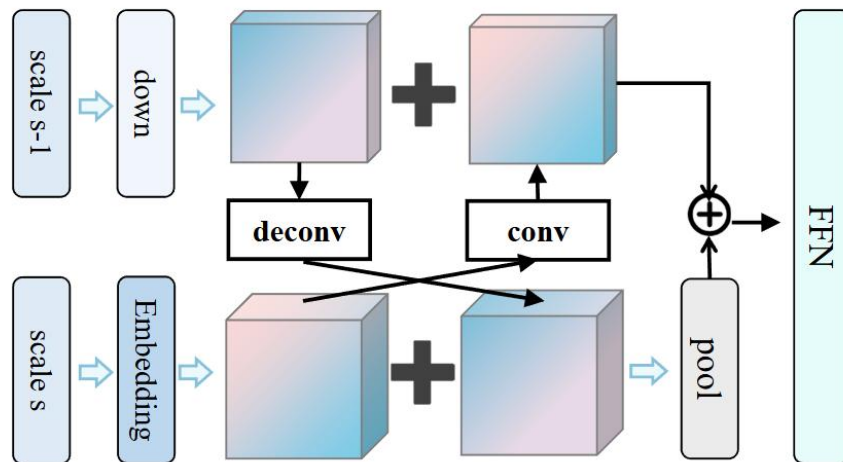


resample

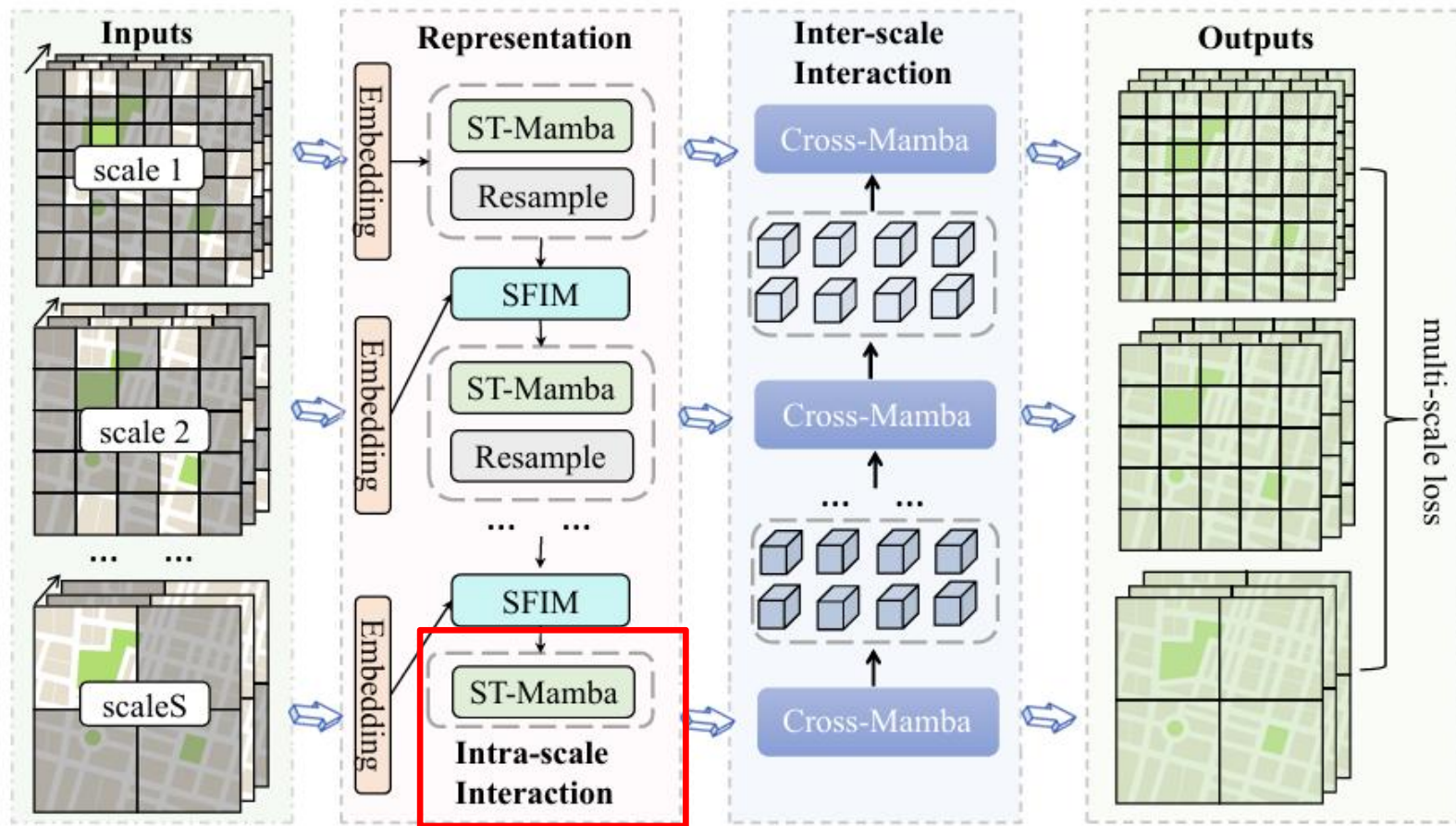


Scale  $s$

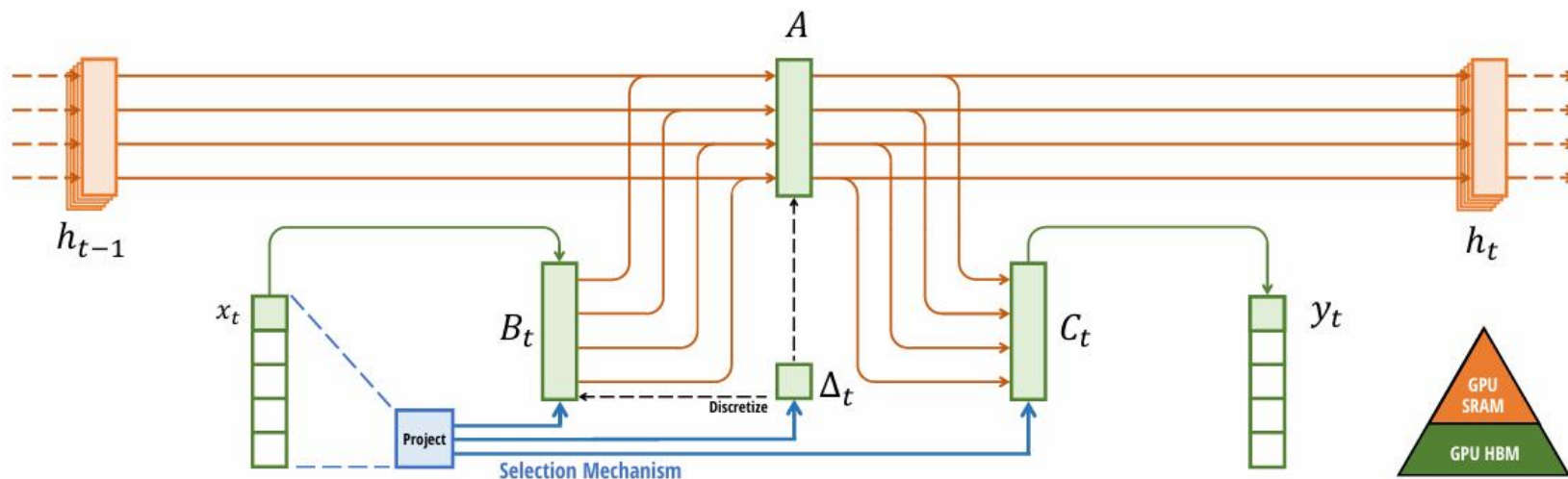
Scale  $s+1$



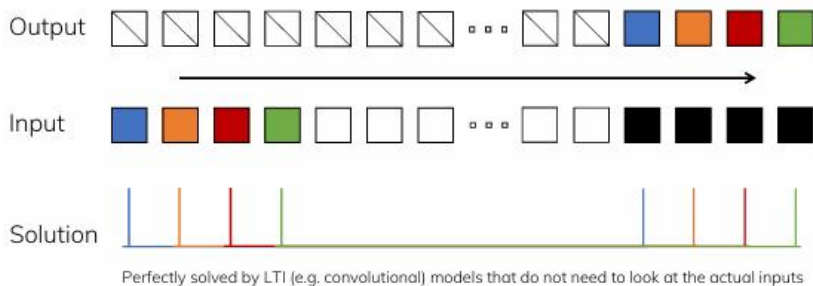
SFIM



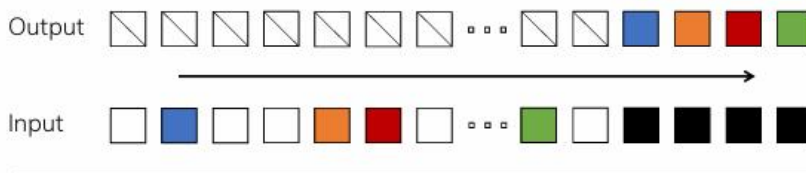
## Selective State Space Model with Hardware-aware State Expansion



### Copying



### Selective Copying



### Induction Heads



- Mamba is designed for sequence, insensitivity to spatial

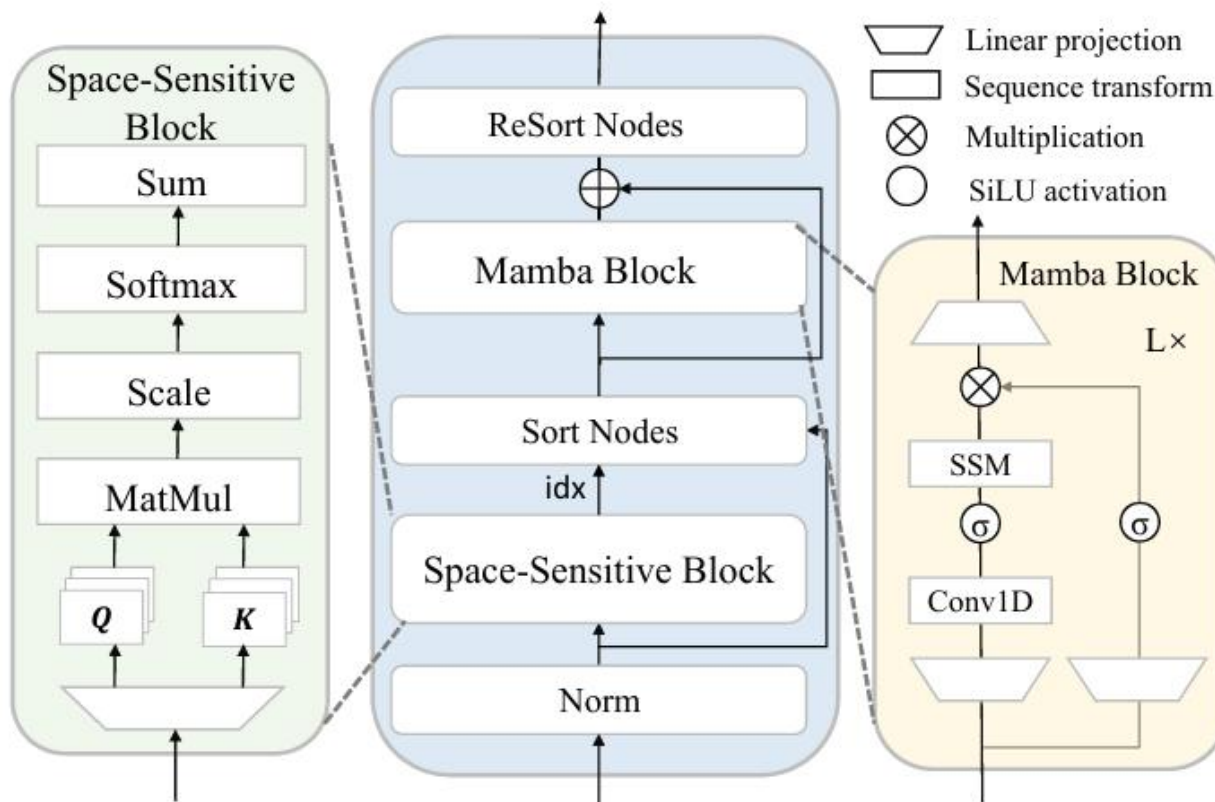
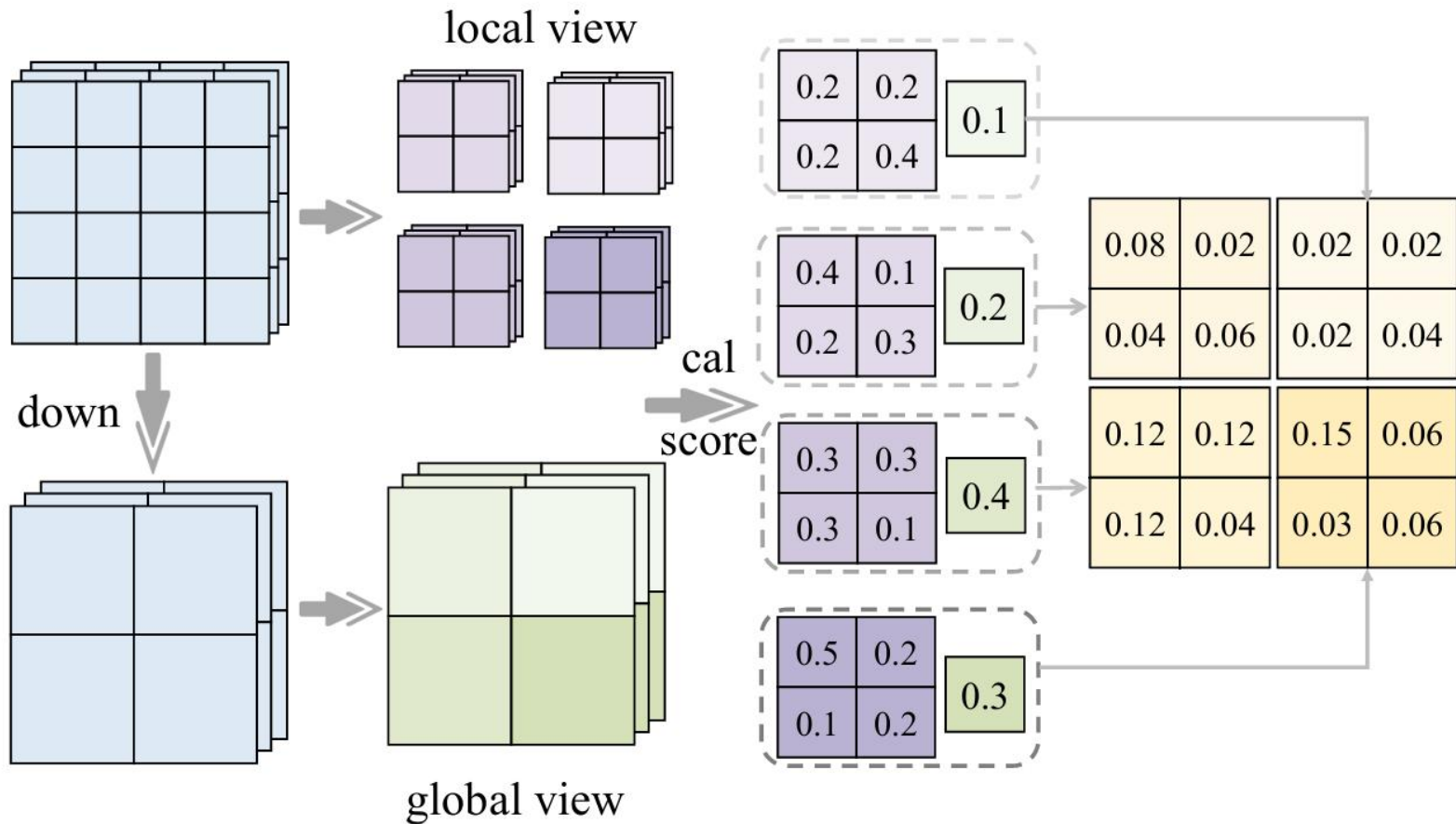
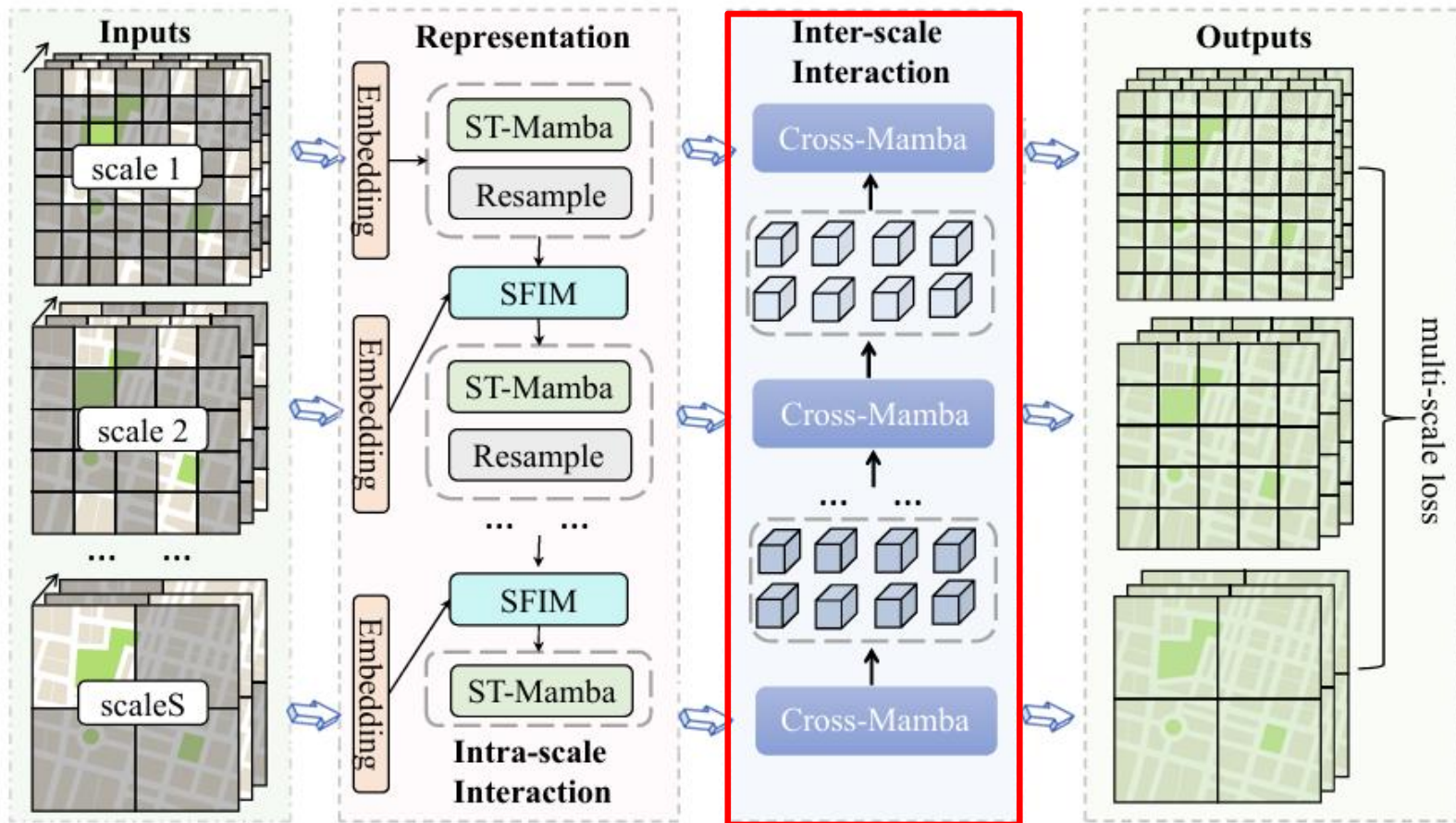
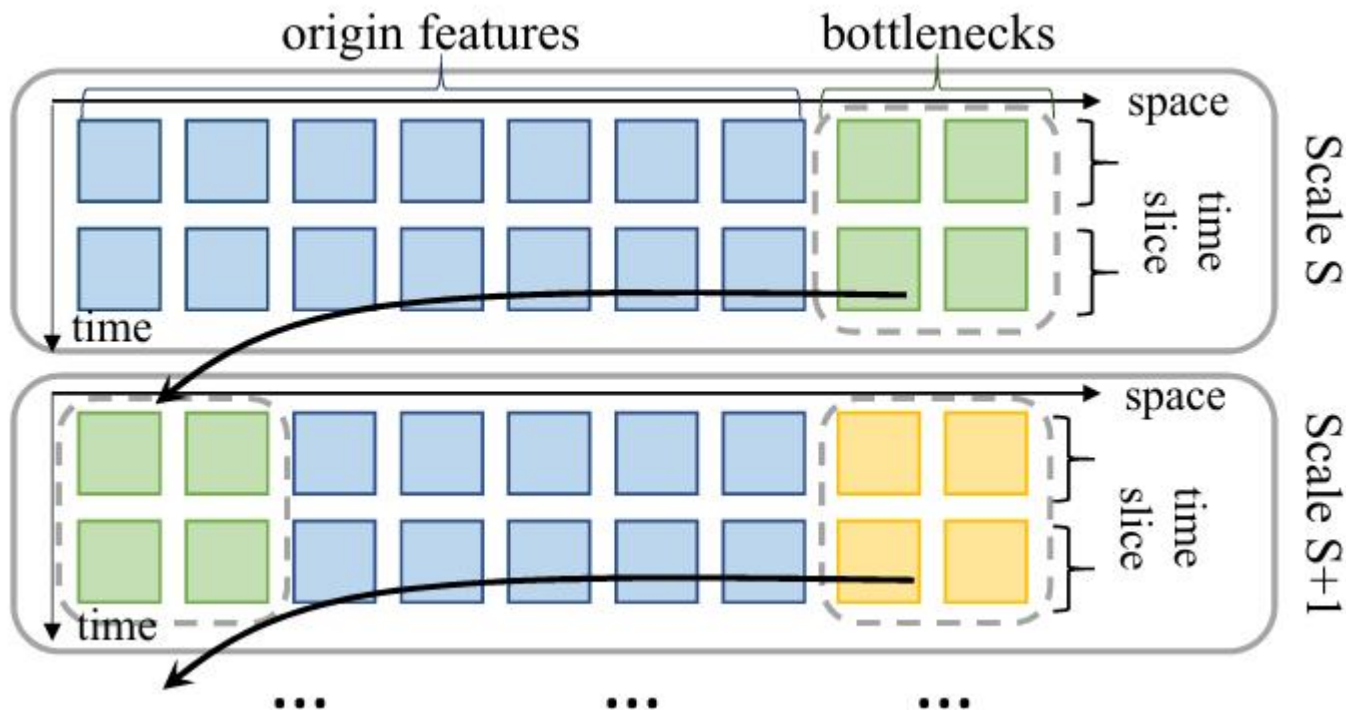


Fig. 6: The structure of ST-Mamba model.

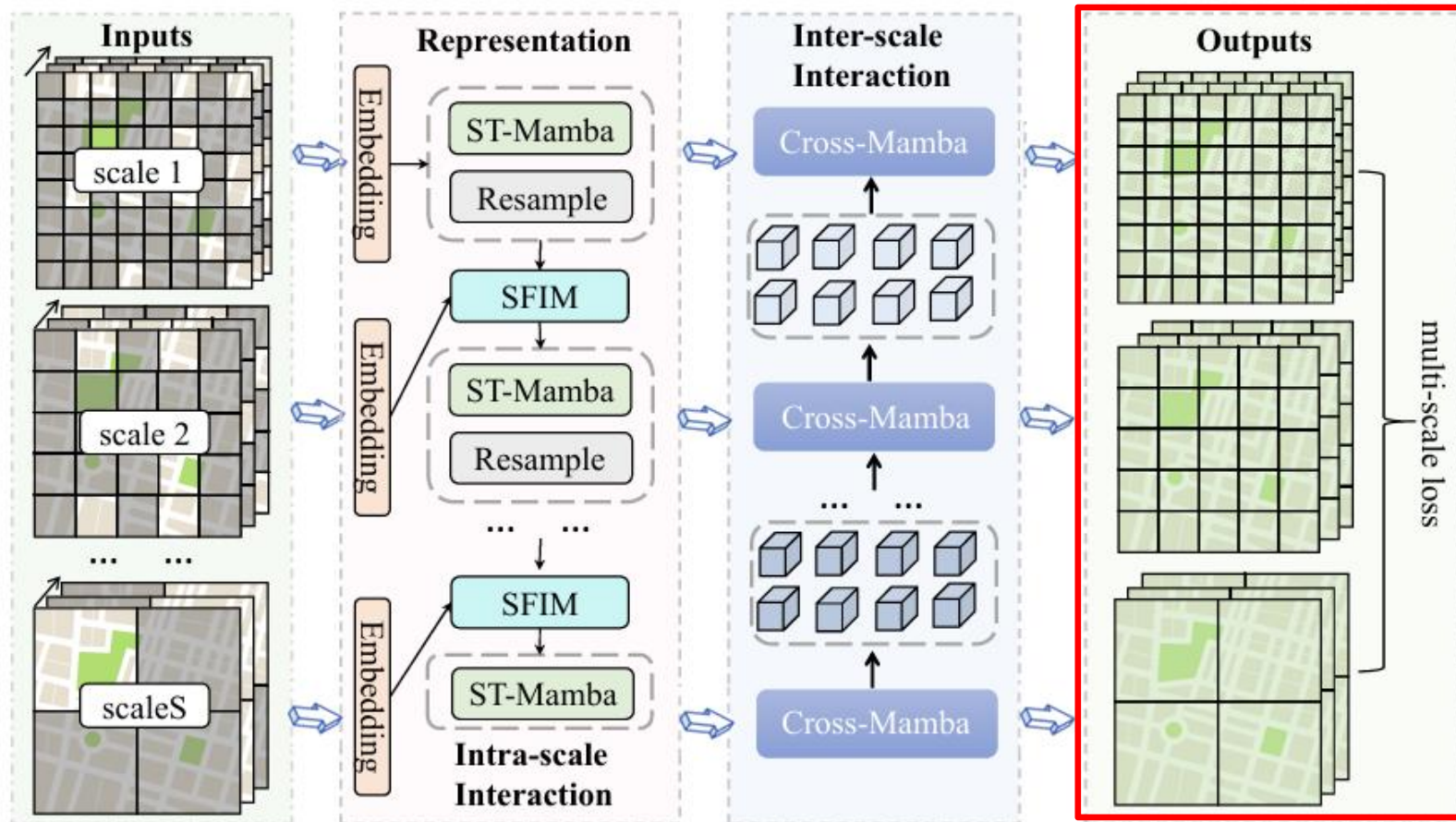
## complexity optimization for ST-Mamba







$$\begin{aligned}
 [\mathbf{X}_{\text{inter}}^{(1)}, \mathbf{E}^{(1)}] &= \text{STM} \left( [\mathbf{X}_{\text{intra}}^{(1)}, \mathbf{B}^{(1)}] \right), \\
 [-, \mathbf{X}_{\text{inter}}^{(s)}, \mathbf{E}^{(s)}] &= \text{STM} \left( [\mathbf{E}^{(s-1)}, \mathbf{X}_{\text{intra}}^{(s)}, \mathbf{B}^{(s)}] \right), \\
 [-, \mathbf{X}_{\text{inter}}^{(S)}] &= \text{STM} \left( [\mathbf{E}^{(S-1)}, \mathbf{X}_{\text{intra}}^{(S)}] \right),
 \end{aligned}$$





Multi-scale output



Multi task

$$\hat{\mathbf{Y}}^{(s)} = \text{Linear}(\mathbf{X}_{\text{inter}}^{(s)}),$$

$$\mathcal{L} = \sum_{s=1}^S C_s \times \text{MSE}(\hat{\mathbf{Y}}^{(s)}, \mathbf{Y}^{(s)}),$$



# Outline



- I. Background and Motivation
- II. Problem and Workflow
- III. Method
- IV. Performance Evaluation**
- V. Conclusion



# Five typical sensing tasks



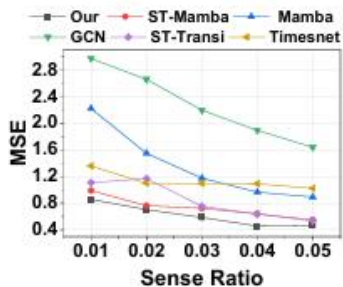
- Air-Quality<sup>1</sup>
  - PM2.5, PM10
- Weather<sup>2</sup>
  - Humidity, Temperature
- Traffic<sup>3</sup>
  - P1, P2

1 <https://nnu.geodata.cn/featureddata.html>

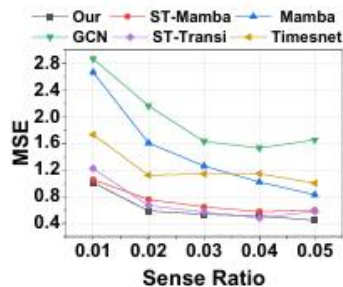
2 <https://www.kdd.org/kdd2018/kdd-cup>

3 <https://gitee.com/arislee/taxi-bj>

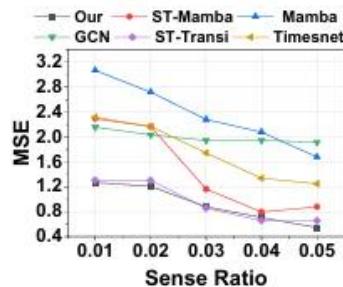
Data Set	Air-Quality				Weather				Traffic			
	PM2.5		PM10		Humidity		Temperature		P1		P2	
Models	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
MF	60.070	7.000	60.146	7.000	33.601	4.465	363.61	16.543	6250.6	42.251	8623.4	49.746
DMF	69.753	6.175	68.866	6.134	57.170	5.738	351.84	13.621	14660.8	69.171	18115.0	76.284
GCN	2.969	1.545	2.865	1.520	2.158	1.064	61.313	5.466	599.94	14.880	938.05	17.785
Mamba	2.220	0.945	2.662	1.008	3.069	1.220	35.419	4.048	491.85	12.914	710.42	15.594
Informer	2.526	1.288	3.374	1.506	2.949	1.294	47.384	4.984	534.42	13.717	627.17	14.207
Timesnet	1.359	0.855	1.729	1.005	2.317	1.111	44.491	4.837	451.28	13.196	568.56	14.799
TS-BGMC	3.589	1.424	3.585	1.426	22.33	3.565	384.365	15.799	9452.02	51.930	12601.53	60.269
ST-Transi	1.108	0.765	1.221	0.802	1.306	0.873	13.782	2.665	425.81	12.214	520.97	13.710
ST-Mamba	0.989	0.703	1.053	0.740	2.294	1.086	30.410	3.812	480.95	12.804	594.77	14.281
<b>Our</b>	<b>0.857</b>	<b>0.653</b>	<b>1.009</b>	<b>0.723</b>	<b>1.269</b>	<b>0.859</b>	<b>13.179</b>	<b>2.616</b>	<b>419.18</b>	<b>12.043</b>	<b>500.97</b>	<b>13.510</b>



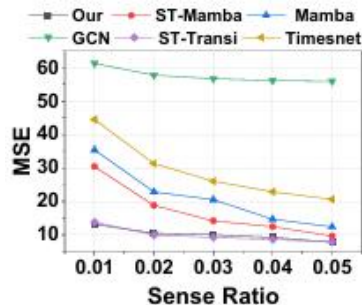
(a) PM2.5



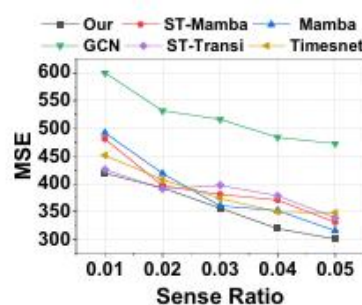
(b) PM10



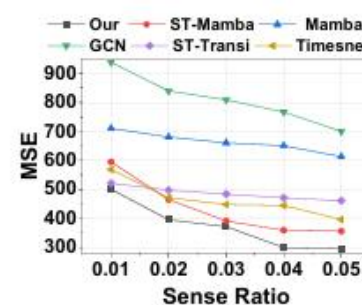
(c) Temperature



(d) Humidity



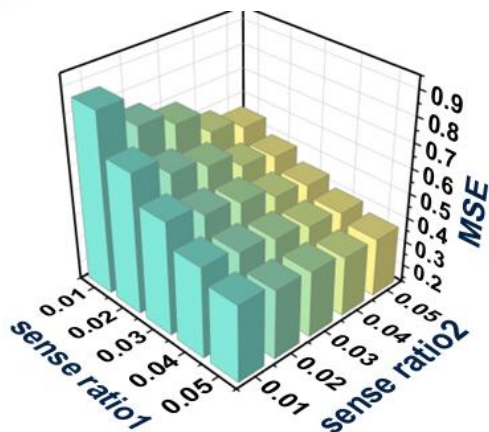
(e) P1



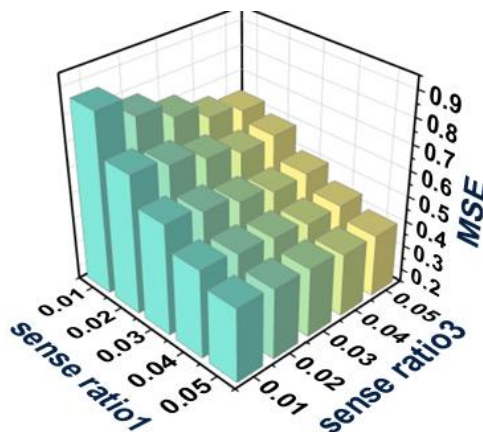
(f) P2

Completion performance

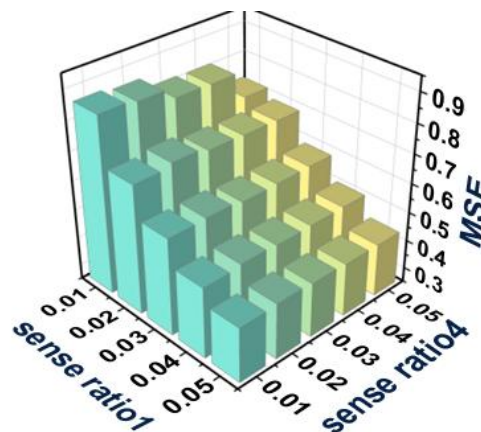
## Completion performance under two different scales



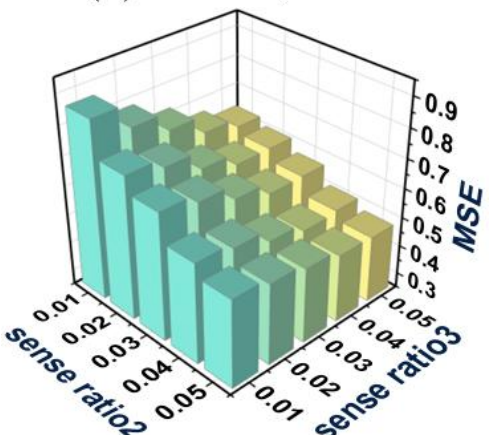
(a) scale1, scale2



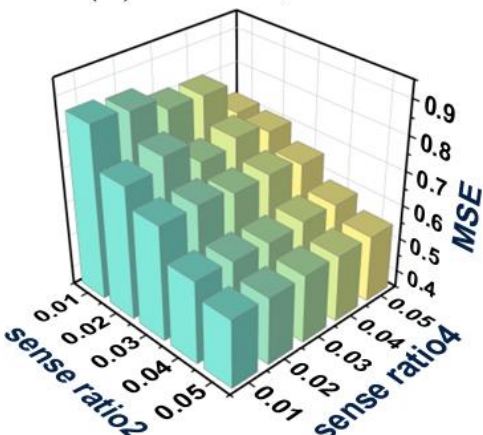
(b) scale1, scale3



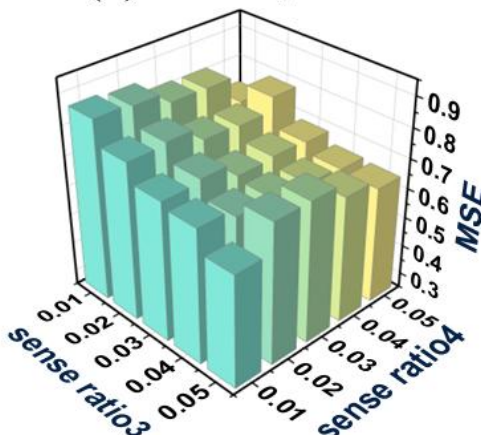
(c) scale1, scale4



(d) scale2, scale3



(e) scale2, scale4



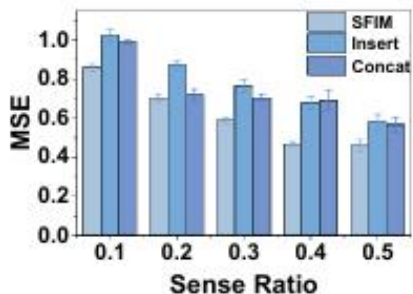
(f) scale3, scale4

## Ablation Study

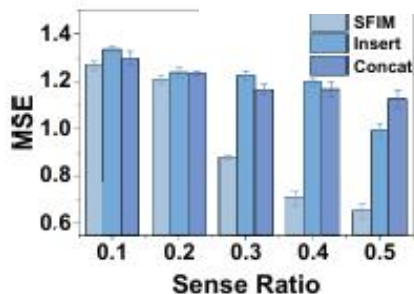
resample method

TABLE II: Performance of different resample methods.

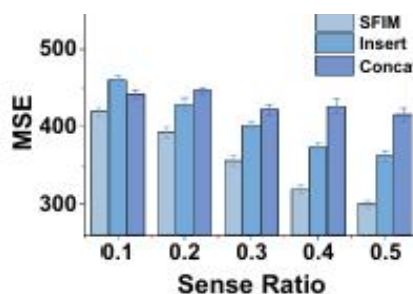
Method	Air		Weather		Traffic	
	MSE	MAE	MSE	MAE	MSE	MAE
scale-adjusted+pool	<b>0.857</b>	<b>0.653</b>	<b>1.269</b>	<b>0.859</b>	<b>419.177</b>	<b>12.043</b>
scale-adjusted	0.877	0.673	1.292	0.873	448.466	12.157
pool	0.975	0.717	1.512	0.949	464.561	12.377
conv	0.974	0.716	1.452	0.930	497.299	12.579



(a) Air



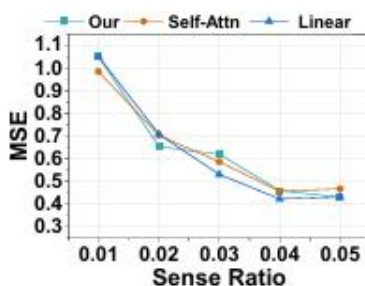
(b) Weather



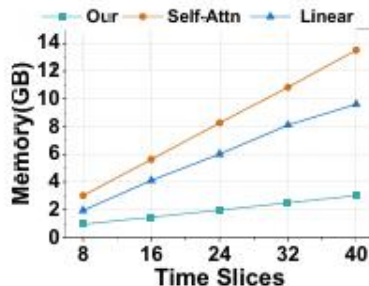
(c) Traffic

feature integration module

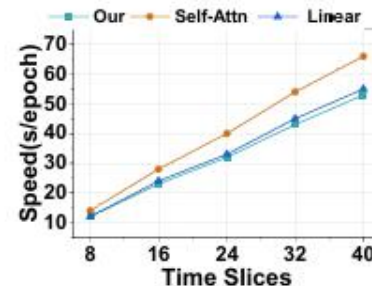
Space-Sensitive Block



(a) Performance



(b) Memory



(c) Speed



# Outline



- I. Background and Motivation
- II. Problem and Workflow
- III. Method
- IV. Performance Evaluation
- V. Conclusion**



# Conclusion



- *non-aligned multi-scale data*
- *Framework :*
  - Multi-scale Data Representation(Resample, SFIM)
  - Intra-scale (ST-Mamba) Interaction and Inter-scale (Cross-Mamba) Interaction
  - Multi-scale viewed as multi-tasking
- *Extensive Evaluation*
  - Six typical sensing tasks (in three types)



# Q&A



# Thank you!

## Q&A

