2020城市计算夏令营

京东城市

京 东 数 科 旗 下

时空深度学习

Deep Learning for Spatio-Temporal Data

张钧波

京东智能城市研究院

京东城市 时空AI产品部

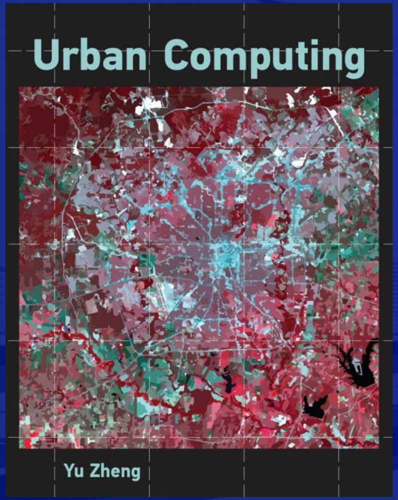
2020年7月22日

城市计算(Urban Computing)

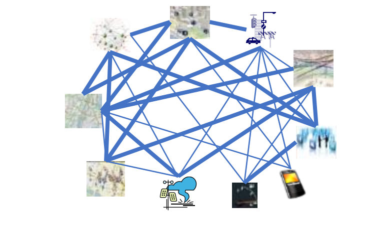
城市数据的采集、管理、分析挖掘和服务提供

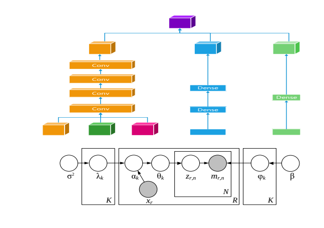
数据 + 计算

解决交通、规划、环境、能耗、公共安全、商业、医疗等痛点



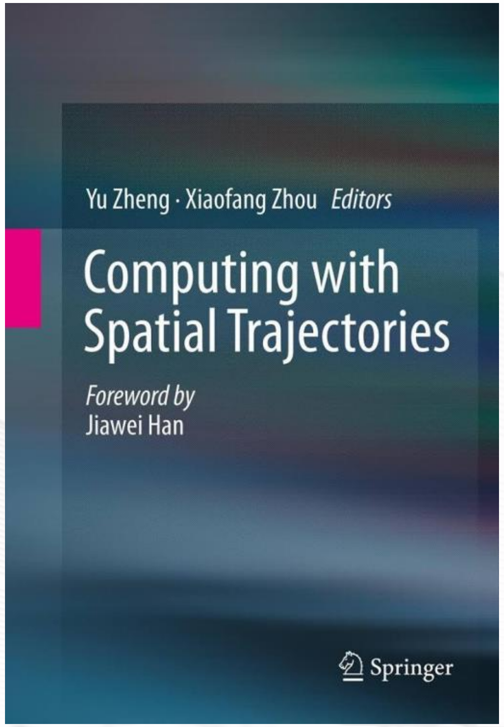
云计算 + 大数据 + AI + 城市场景

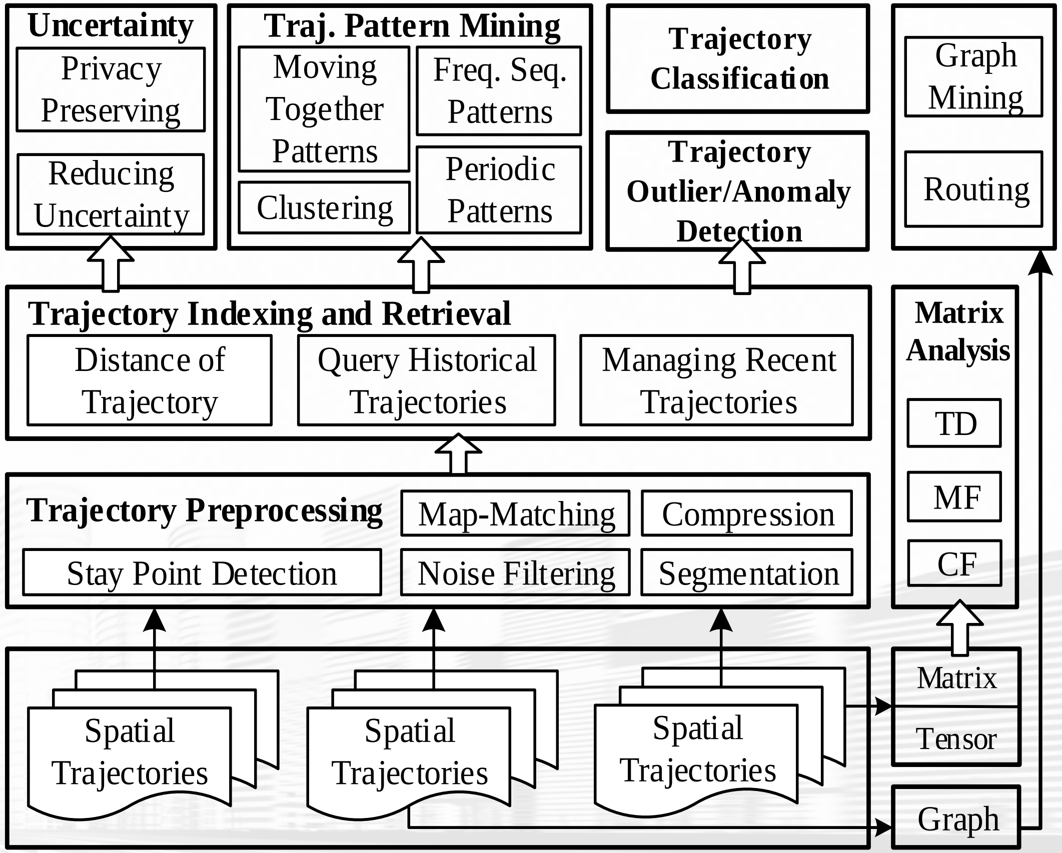




Yu Zheng. Urban Computing. MIT Press, 2019

Trajectory Data Management and Mining





Yu Zheng, Xiaofang Zhou. Computing with Spatial Trajectories, Springer Press 2011

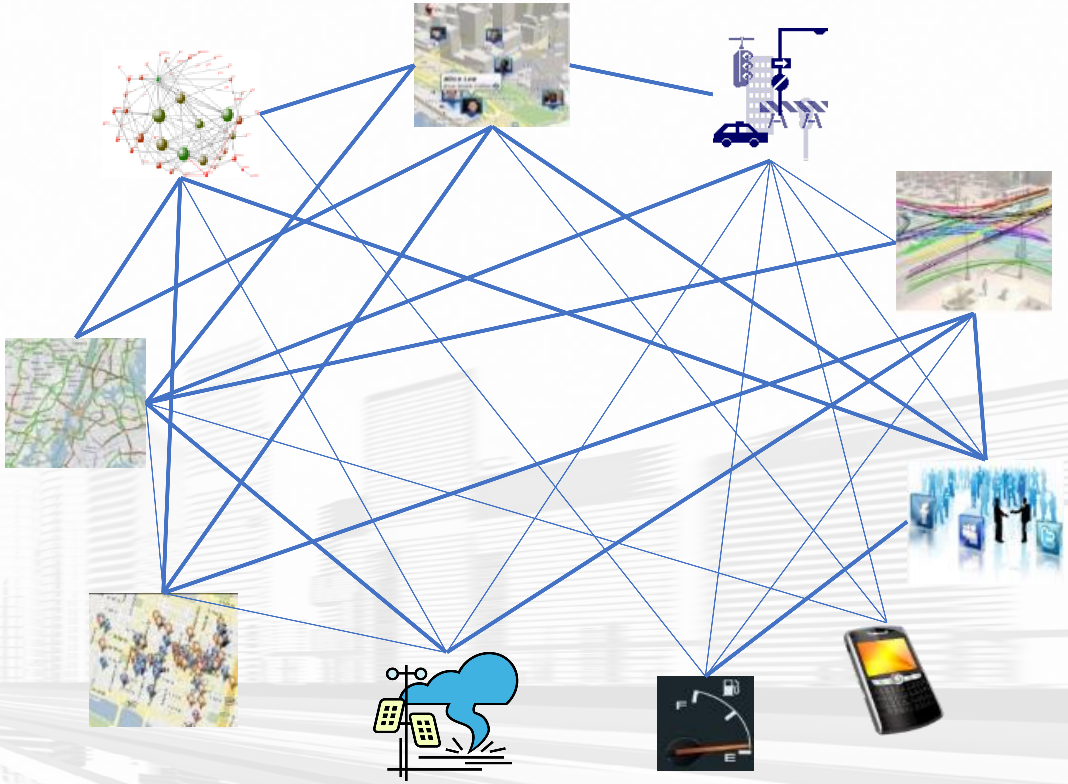
Yu Zheng. Trajectory Data Mining: An Overview. ACM Transactions on Intelligent Systems and Technology. 2015

Agenda

1. Why Deep Learning meets ST data
2. Deep Neural Networks for ST data
   - Spatio-Temporal Neural Networks
     - ST Point Data, ST Gridded Data, ST Networks, ST Sequence Data
   - Advanced ST Neural Networks
     - ST Meta Learning, ST Network Architecture Search
3. More Spatio-Temporal AI Applications

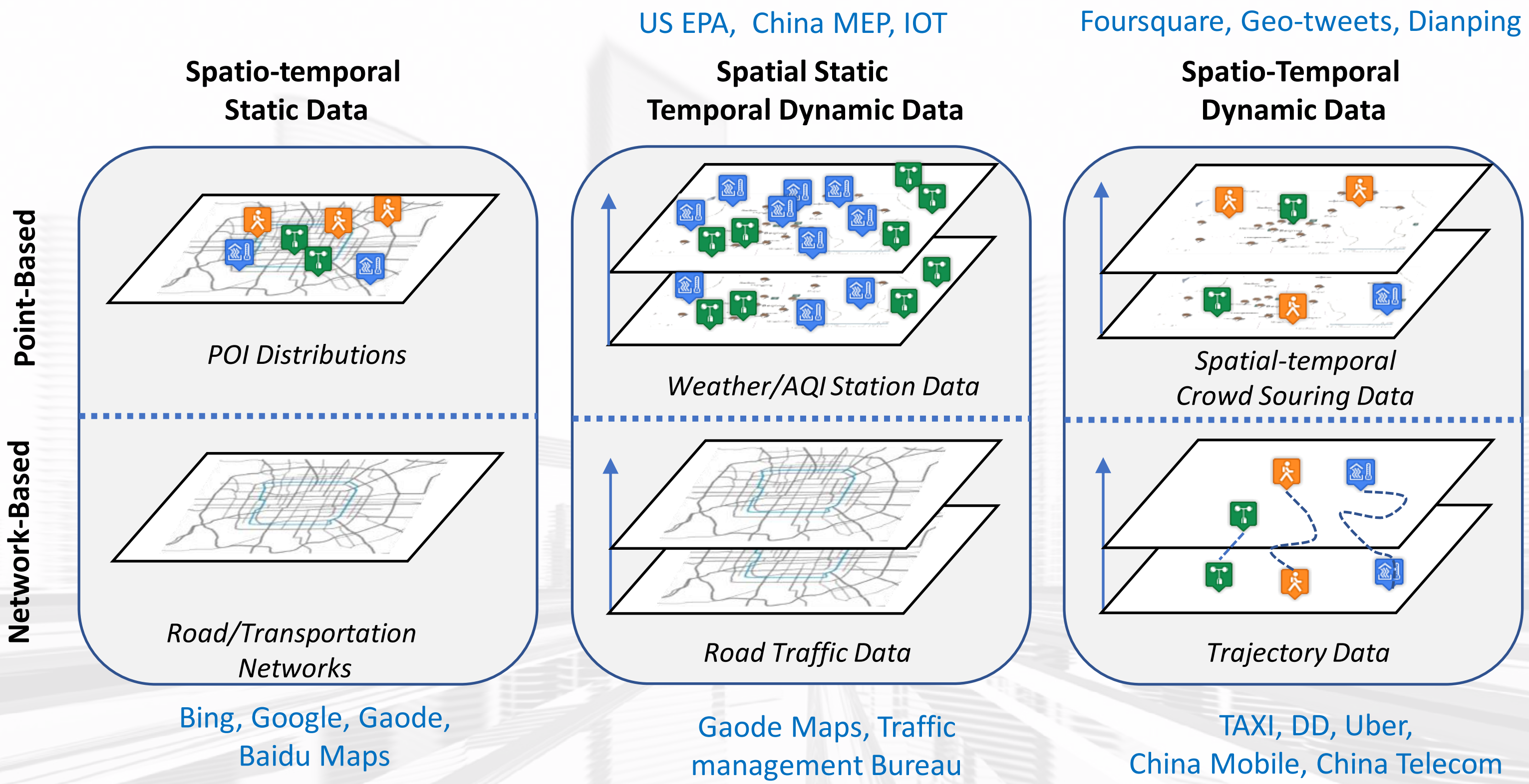
Part 1. Why Deep Learning meets *Spatio-Temporal* data

Big *Spatio-Temporal* Data in Cities



Taxonomy of Spatio-Temporal (ST) Data

- Data Structures
- Spatio-temporal (ST) Properties



京东城市

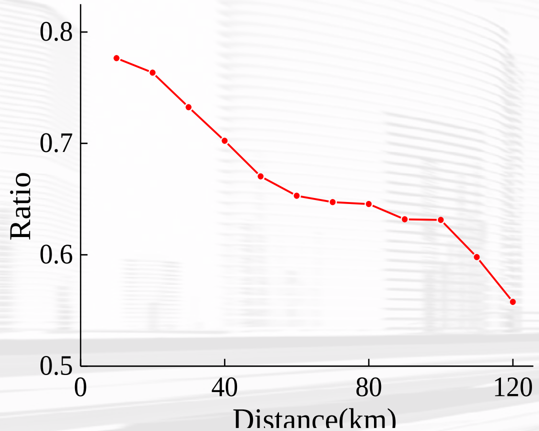
京 东 数 科 旗 下

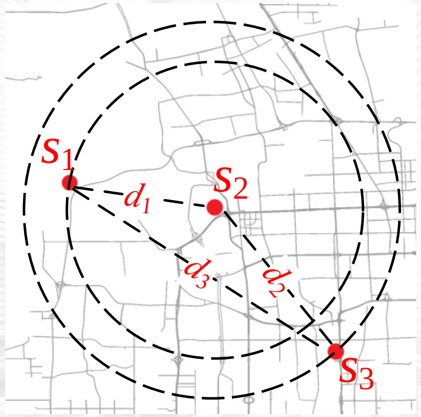
Why Spatio-Temporal Data Is Unique

Spatial Properties

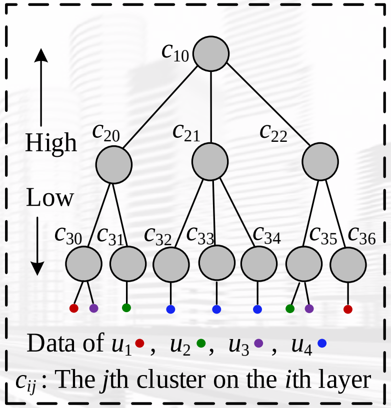
- Distance
  - Spatial closeness
  - Triangle inequality: $|d_1 - d_2| \le d_3 \le |d_1 + d_2|$

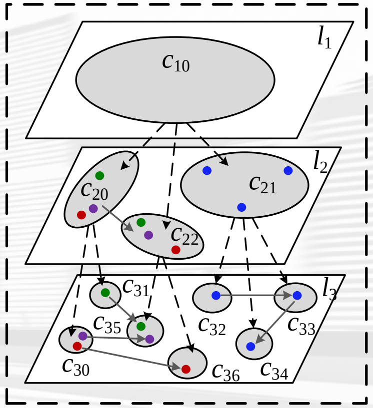
- Hierarchy
  - Different spatial granularities
  - City structures





$\boldsymbol{u_1}: u_2 > u_4$



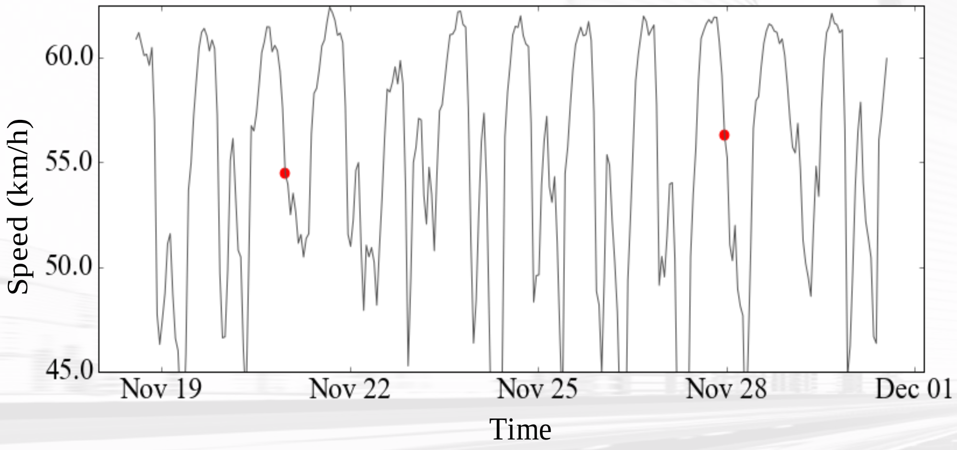


$\boldsymbol{u_1}: u_3 > u_2$

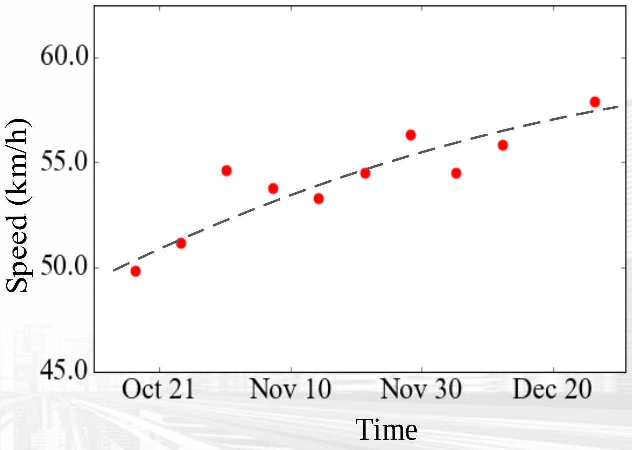
Junbo Zhang, Yu Zheng, et al. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction, AAAI 2017

Why Spatio-Temporal Data Is Unique

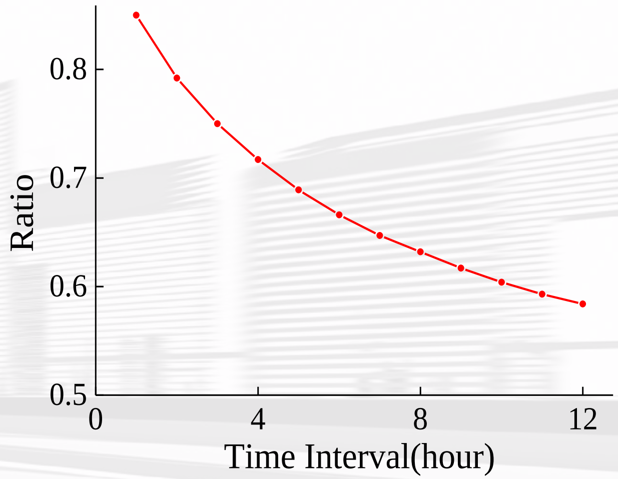
- Temporal properties
  - Temporal closeness
  - Period
  - Trend



A) Hourly traffic speed on consecutive days



B) Traffic speed at 9-10am on consecutive weekends



京东城市

京 东 数 科 旗 下

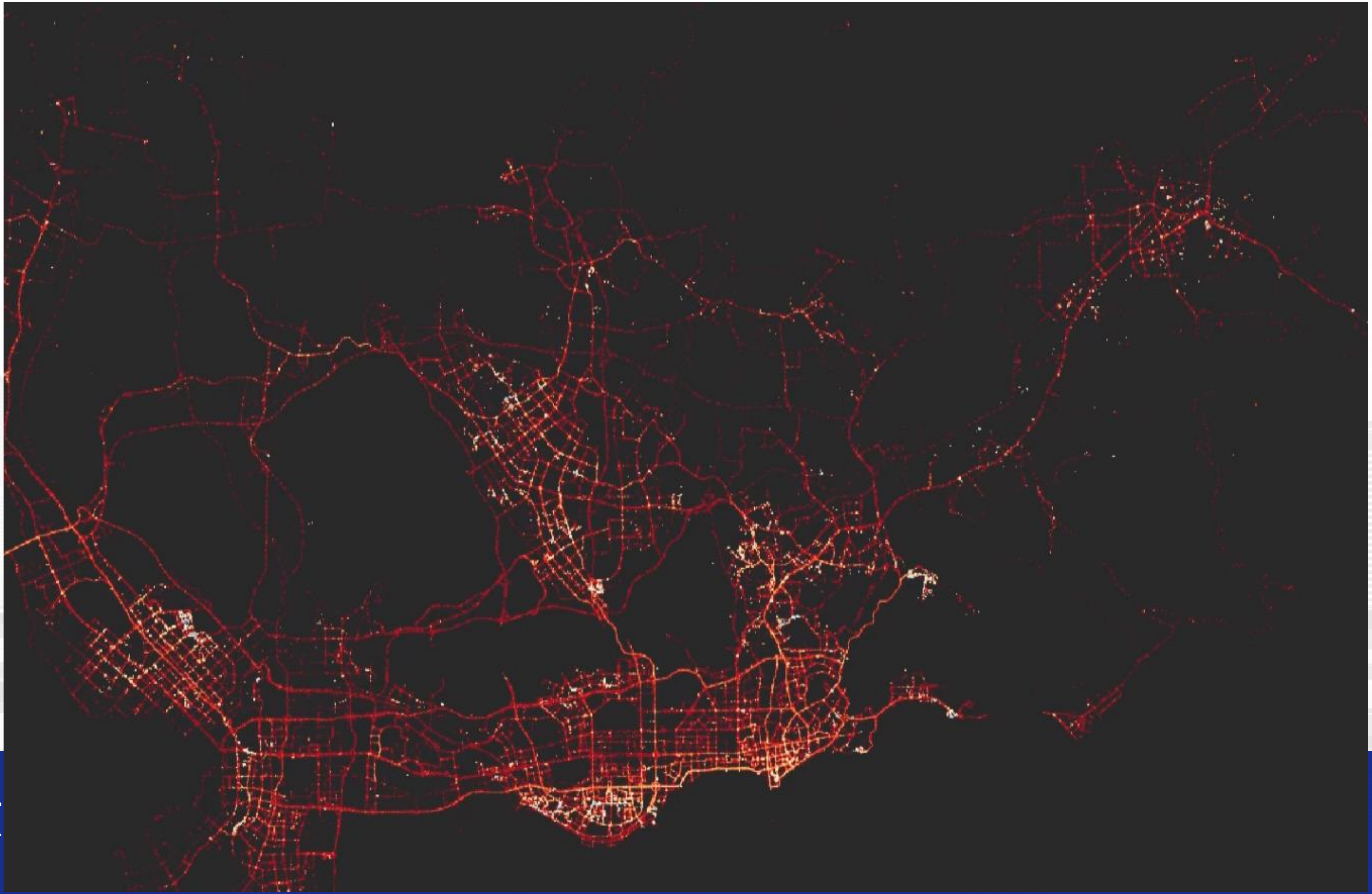
Junbo Zhang, Yu Zheng, et al. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction, AAAI 2017

Why Deep Learning meets ST Data

- What Deep Learning can do for ST Data
  - Encoding a (single) ST dataset
  - Fusing multiple ST datasets

- What ST data can provide to Deep Learning
  - Massive and diverse Data
  - Computing infrastructures are ready
  - Application scenarios requiring
    - Instantaneous responses at large spaces
    - Collective computing
    - (traditional machine learning models many not be able to handle)

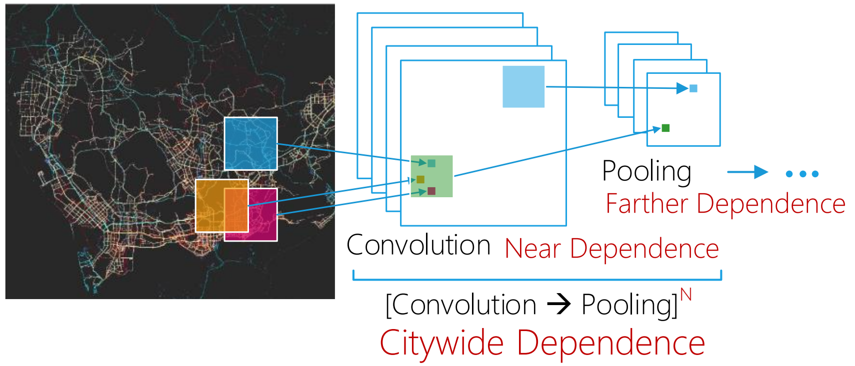
Taxi Trajectory Data of Shenzhen

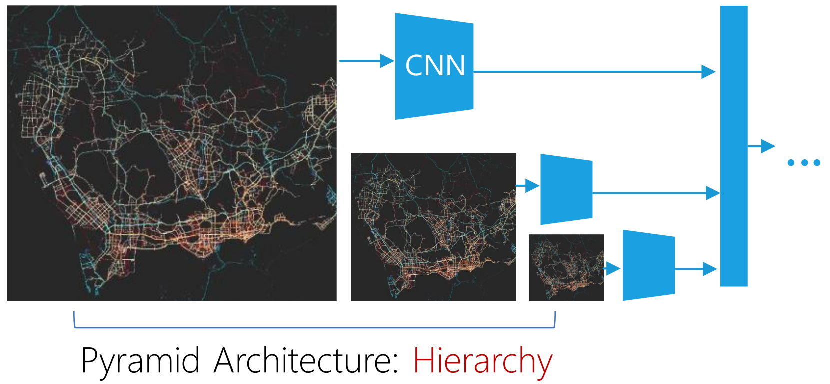


京东
京 东 数 科

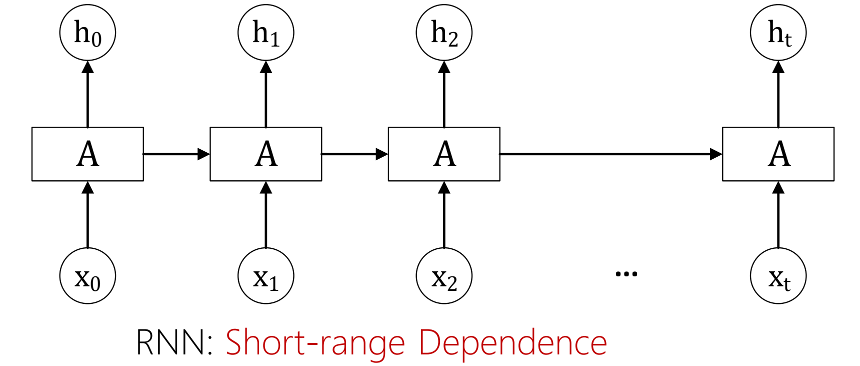
Encoding Spatio-Temporal Properties

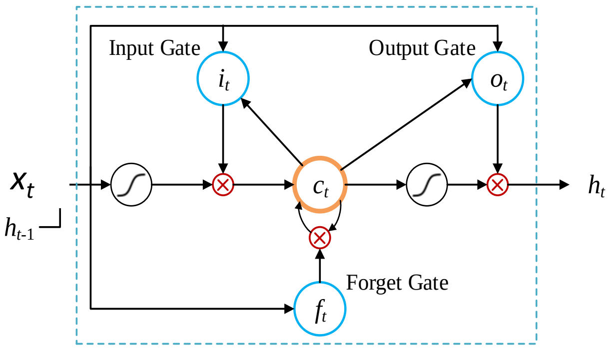
CNN is able to model spatial properties





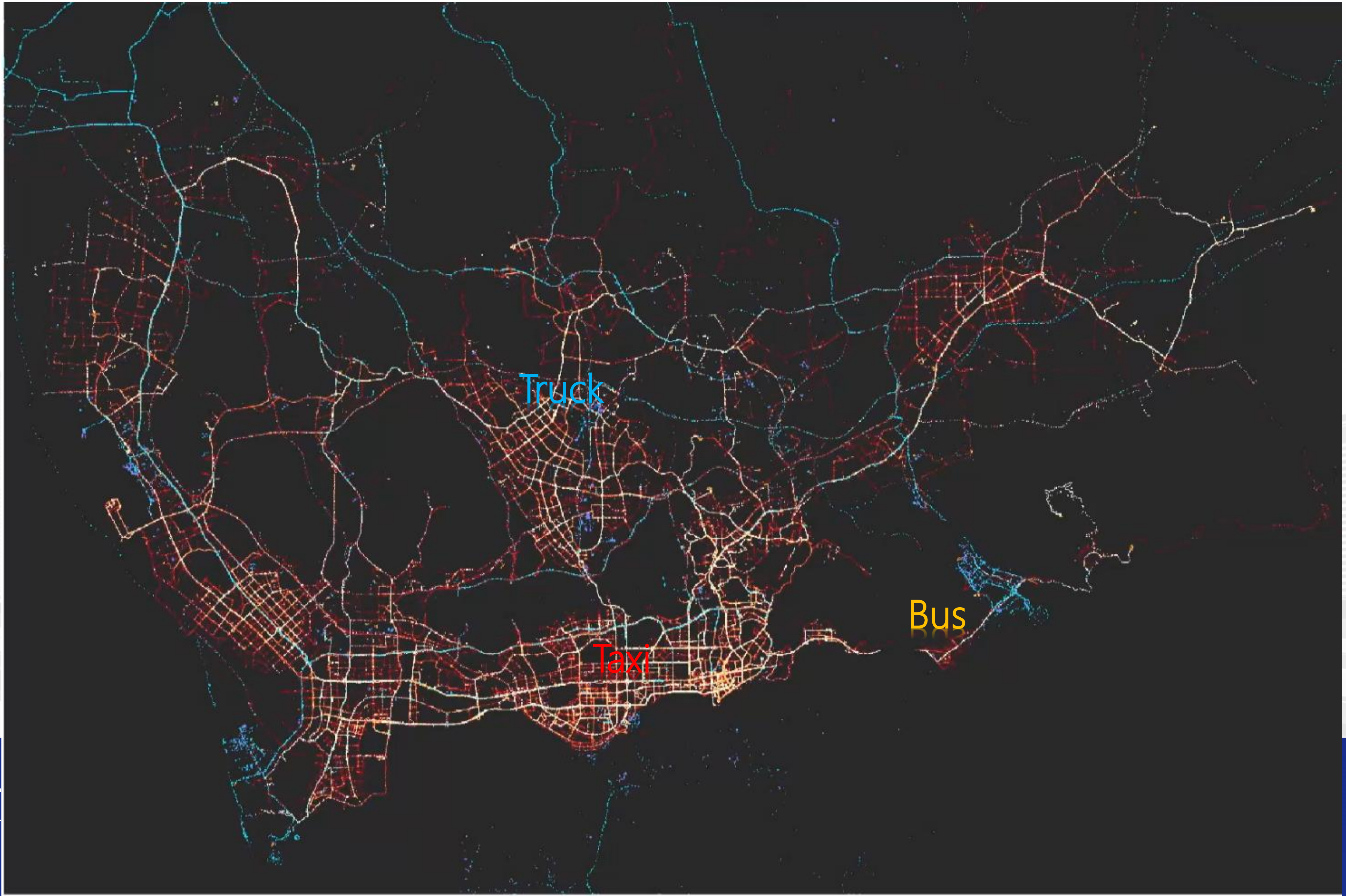
RNN/LSTM is able to model temporal properties





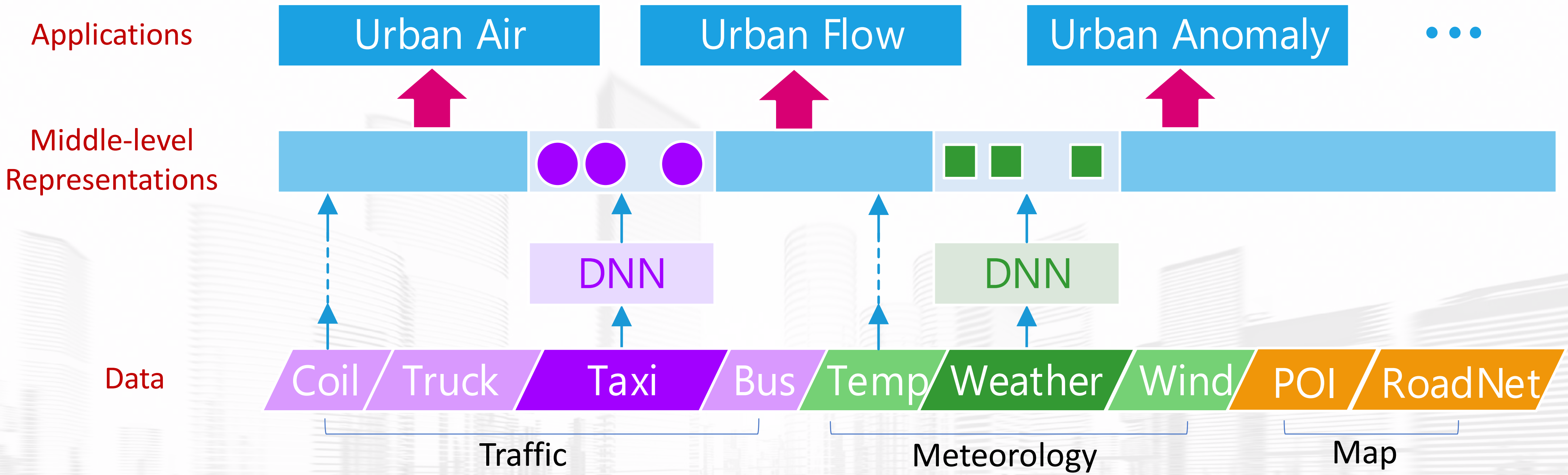
LSTM: Long-range Dependence

Trajectories of taxis, trucks and buses



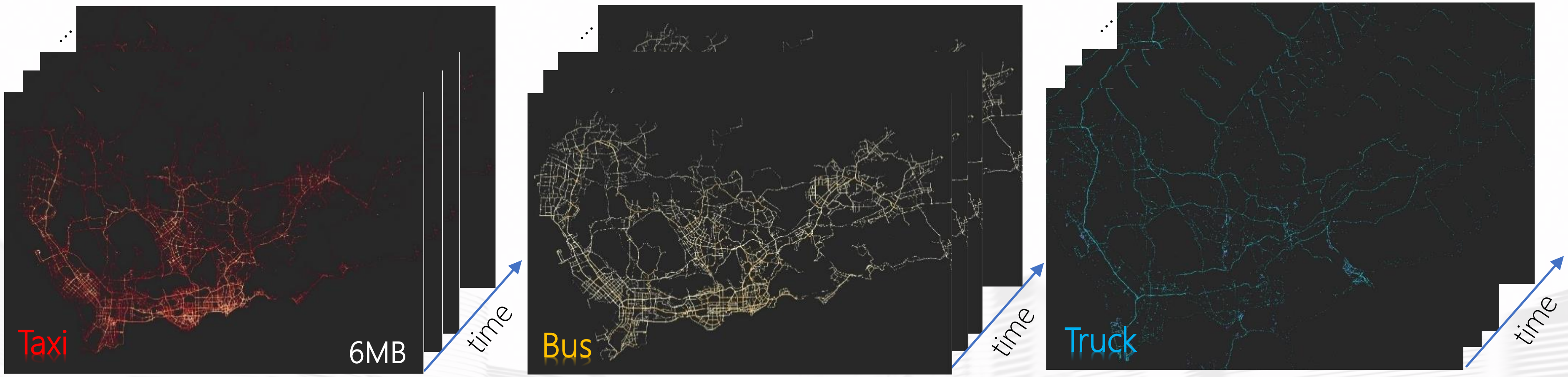
京东
京 东 数 科

Fusing Multiple ST-Datasets



Why Deep Learning for ST Data

- Big ST-Data （**5G + IoT**）



1 frame/min, one year: 3TB

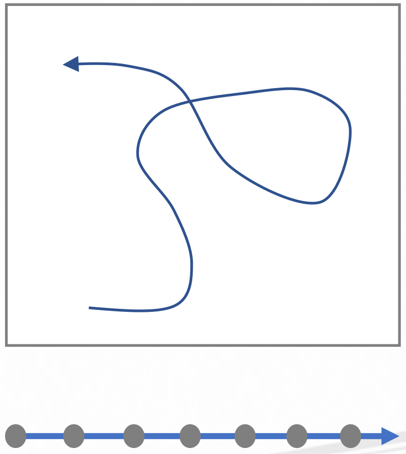
1 City: 9TB

120 Cities: 1PB

Traditional ML algorithms cannot model spatial and temporal properties of such a live and large-scale data

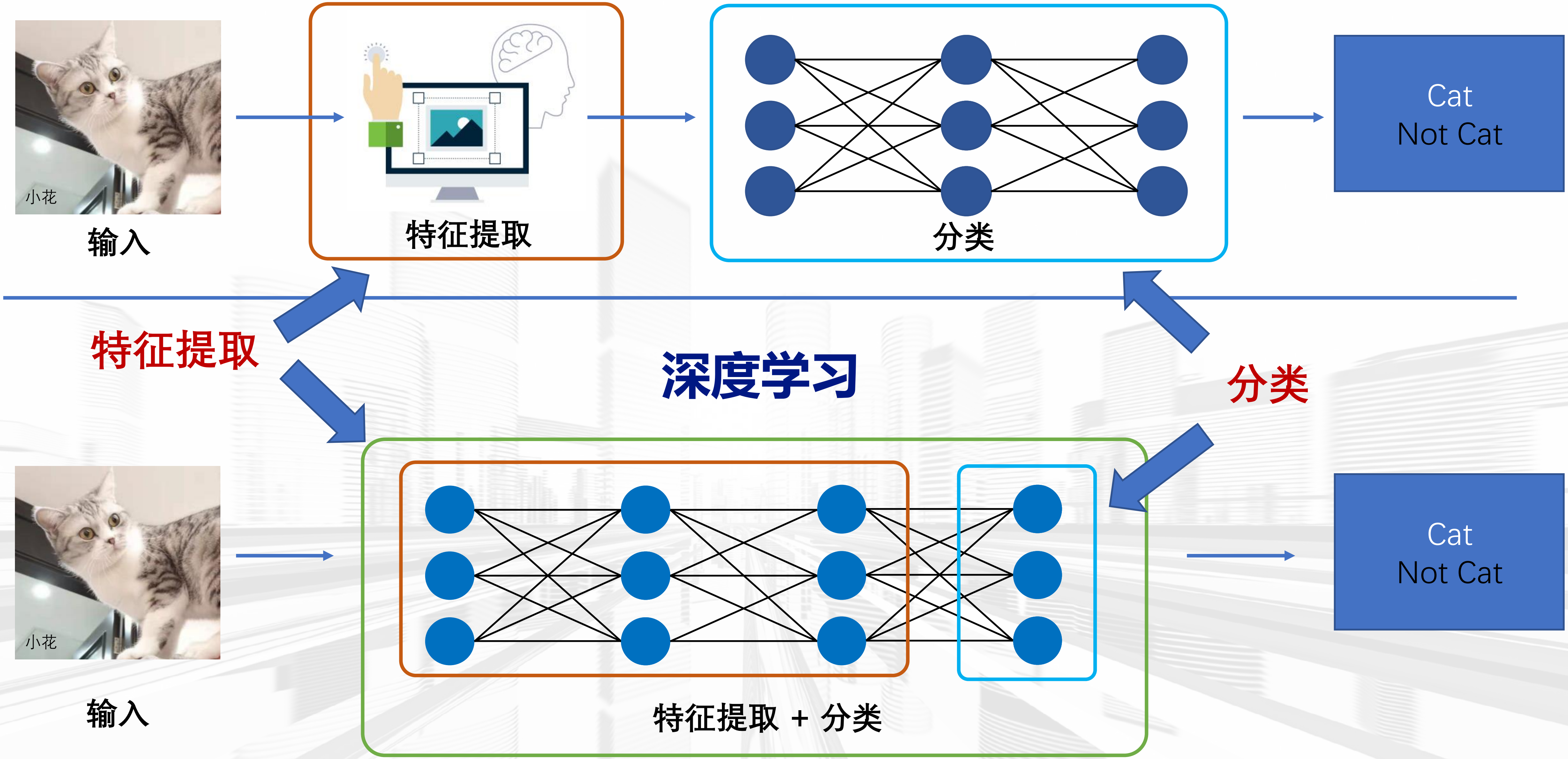
Challenges

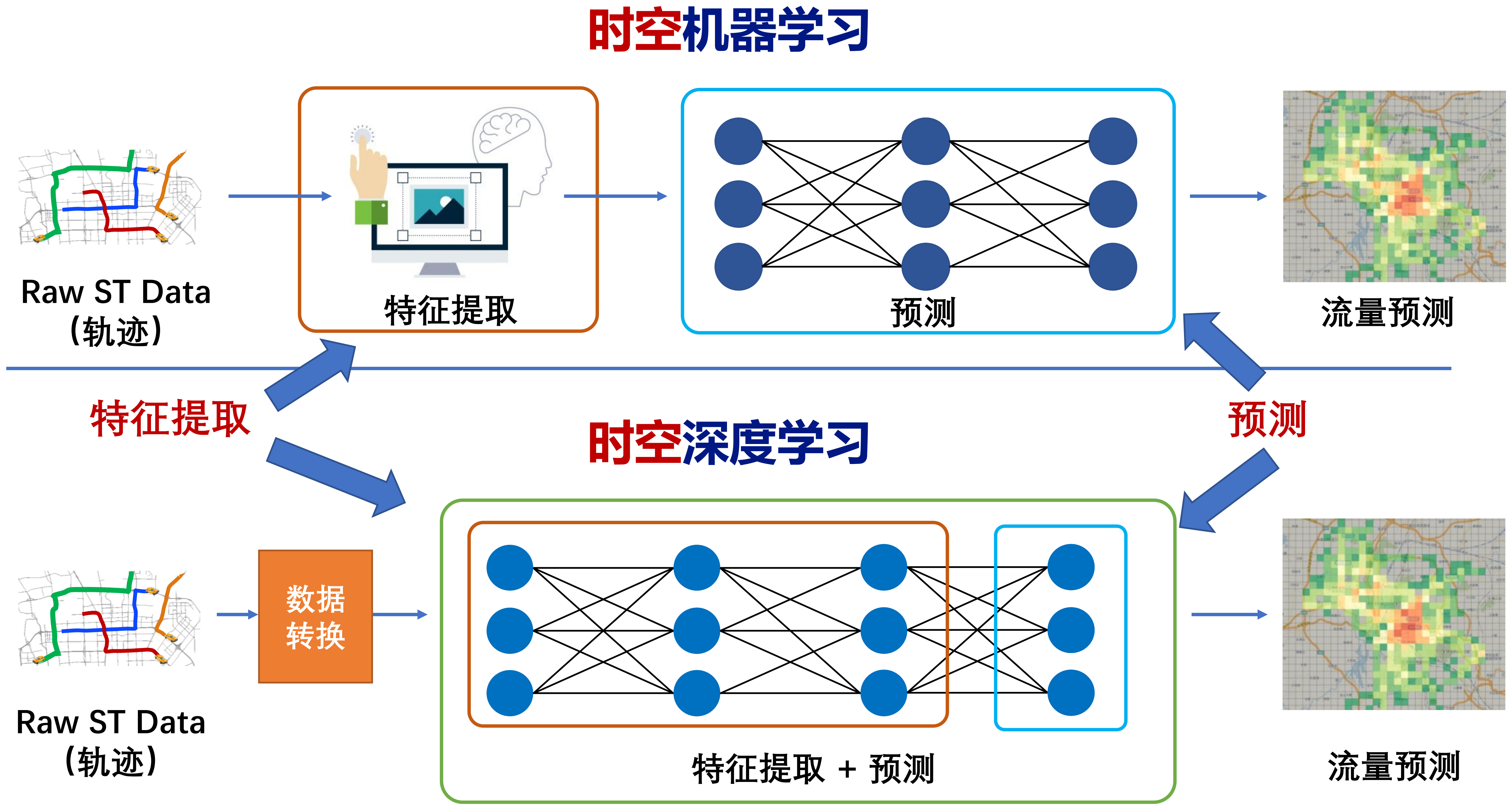
- Deep Learning for ST-Data
  - Cannot fit raw spatiotemporal data into a deep learning model → **Data transformation**
  - Texts and images → spatial and spatiotemporal data; (**Encoding spatiotemporal properties**)
  - Mining a single data source → **Mining data across different domains**



Data transformation

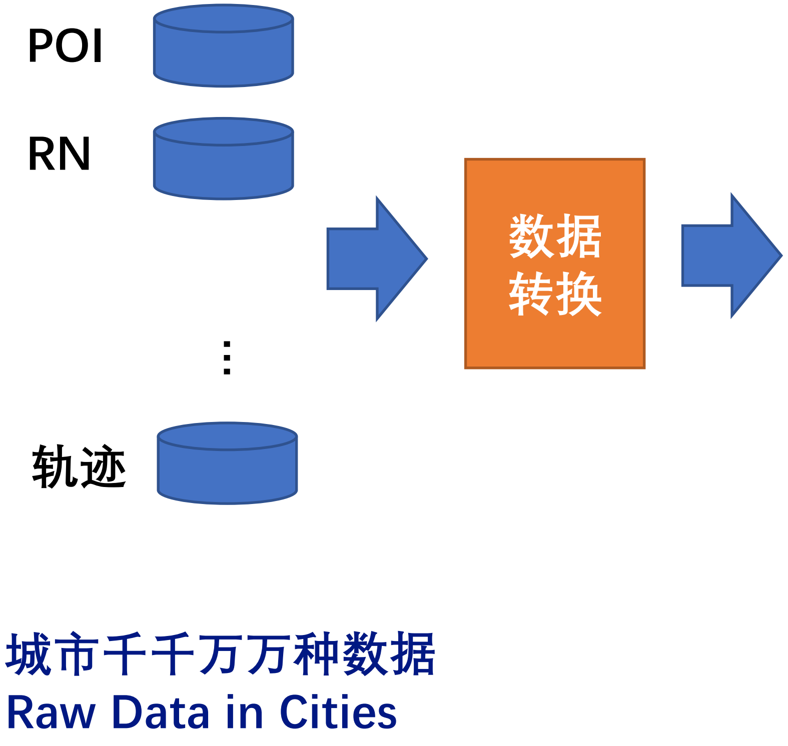
机器学习

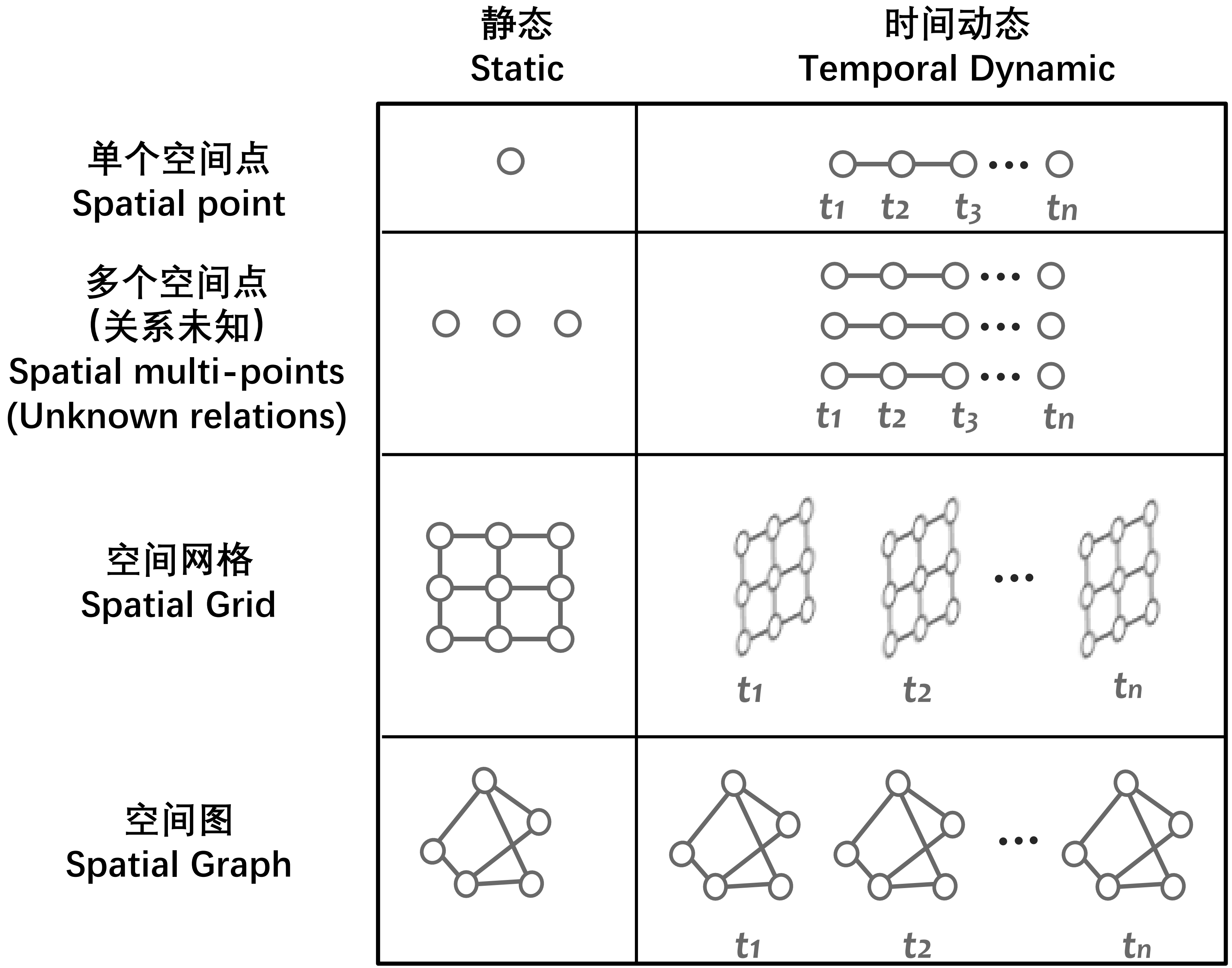




时空机器学习
Raw ST Data
（轨迹）
特征提取
预测
流量预测
特征提取
时空深度学习
预测
数据
转换
Raw ST Data
（轨迹）
特征提取 + 预测
流量预测

数据转换为深度学习模型输入





Part 2. Deep Neural Networks for *Spatio-Temporal* data

2.1 Spatio-Temporal Neural Networks

Spatio-Temporal Neural Networks

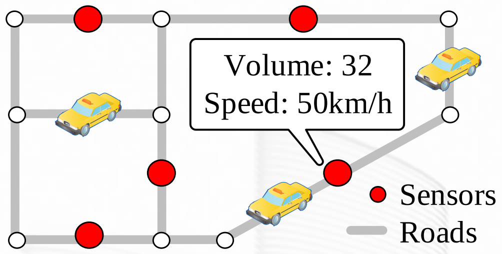
- ST Point Data: GeoMAN
- ST Gridded Data: ST-ResNet
- ST Network (Graph) Data: MVGCN, MDL
- ST Sequence Data: DeepTTE

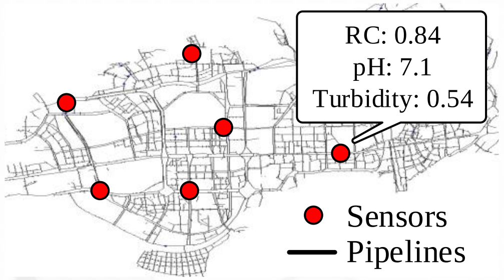
Spatio-Temporal Neural Networks

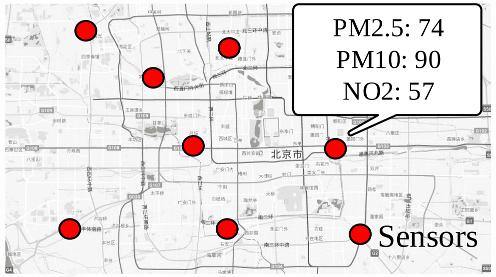
- ST Point Data: GeoMAN
- ST Gridded Data: ST-ResNet
- ST Network (Graph) Data: MVGCN, MDL
- ST Sequence Data: DeepTTE

ST Point Data (Geo-sensory Time Series)

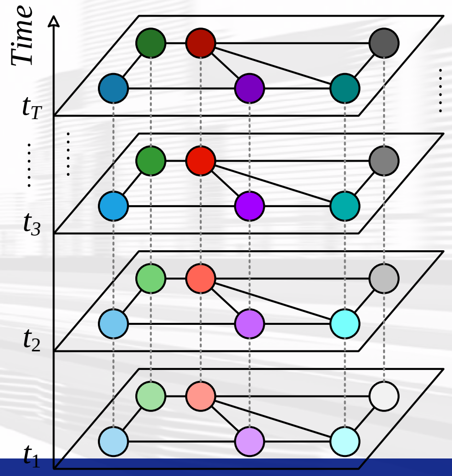
- There are massive sensors deployed in physical world





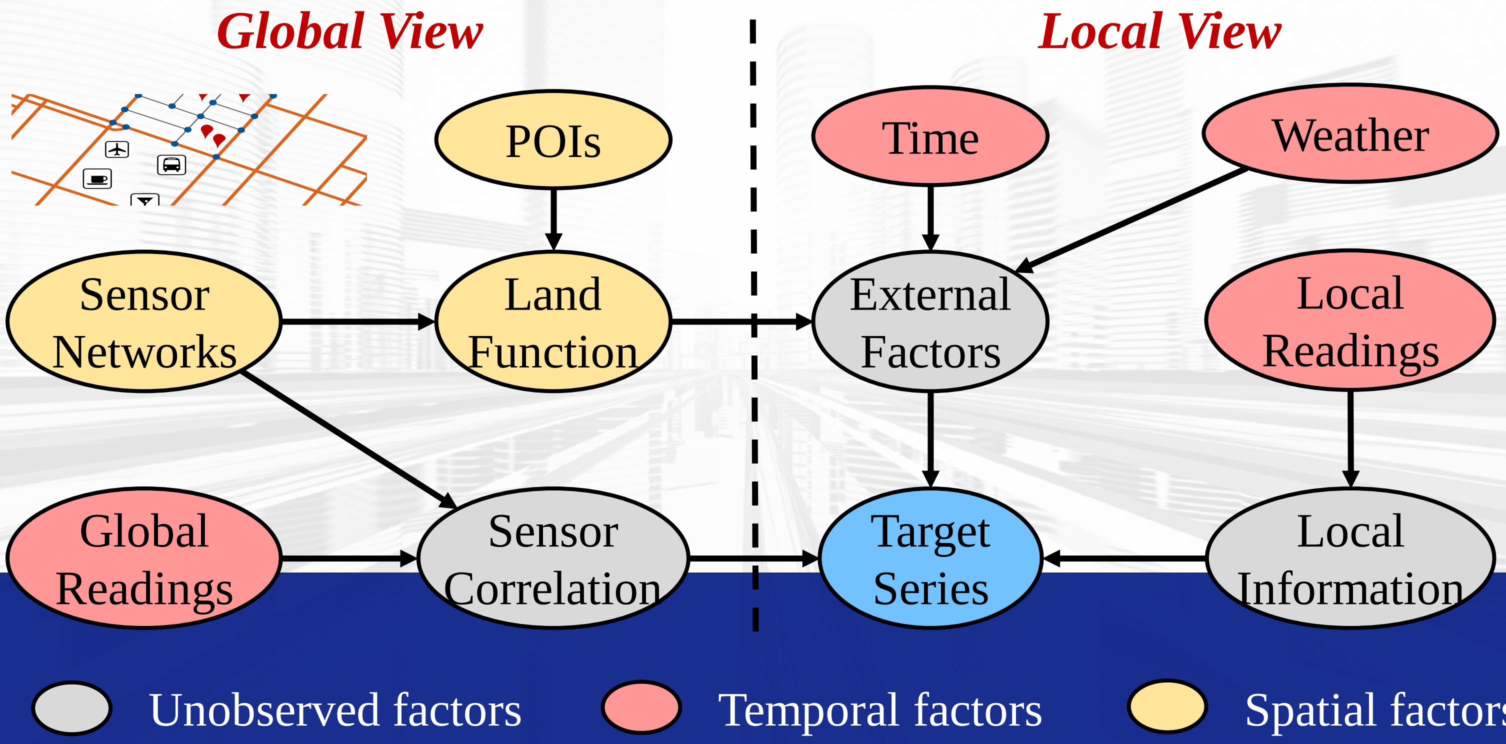


- Properties
  - Each sensor has a unique geospatial location
  - Constantly reporting time series readings about different measurements
  - With geospatial correlation between their readings



Insight

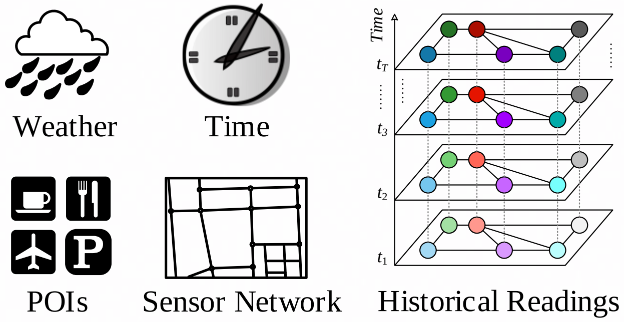
- Affected by many factors
  - Local information
  - Sensor correlations
  - External factors, *e.g.*, weather, time and land use

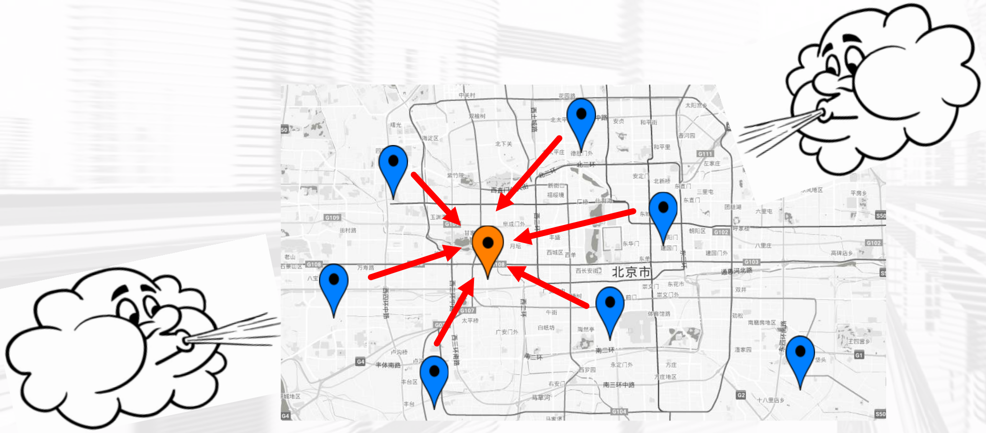


京东城市
京 东 数 科 旗 下

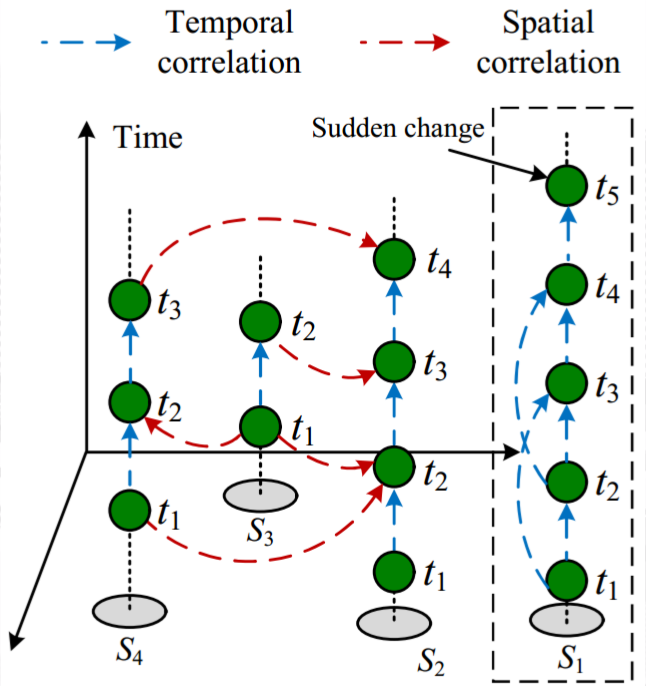
Challenges

- Affected by many factors
  - Readings of previous time interval
  - Readings of other sensors in nearby regions
  - External factors: weather, time and land use
- Dynamic Inter-sensor correlations
- Dynamic temporal correlation

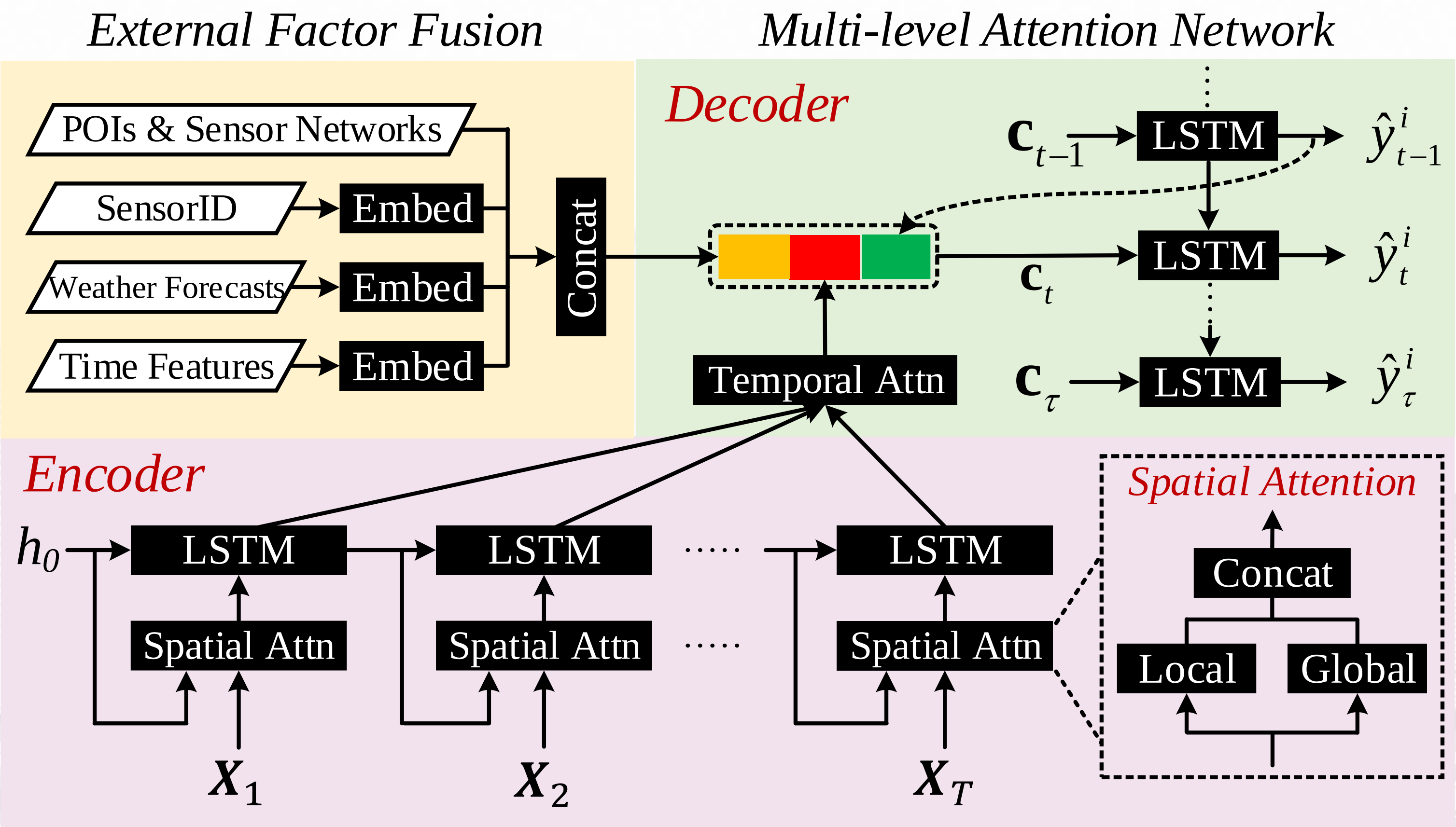




DAY 2

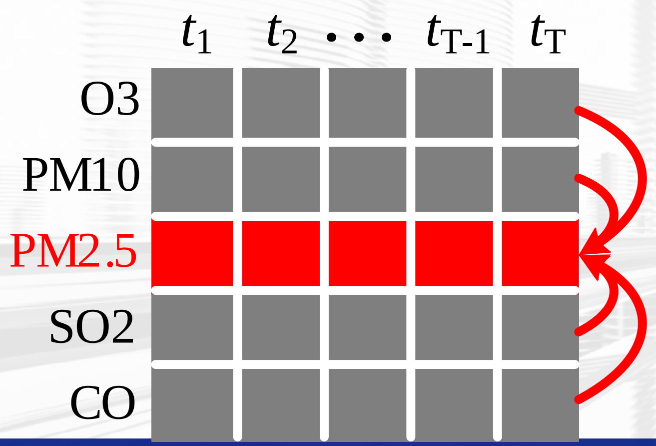


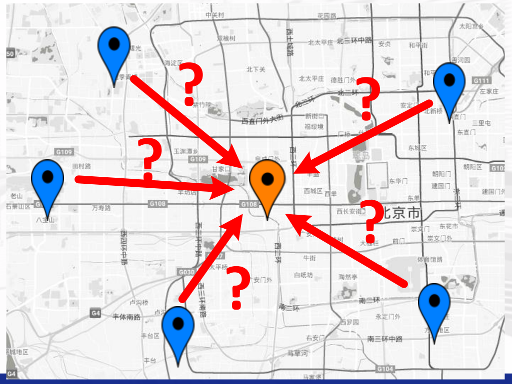
Framework

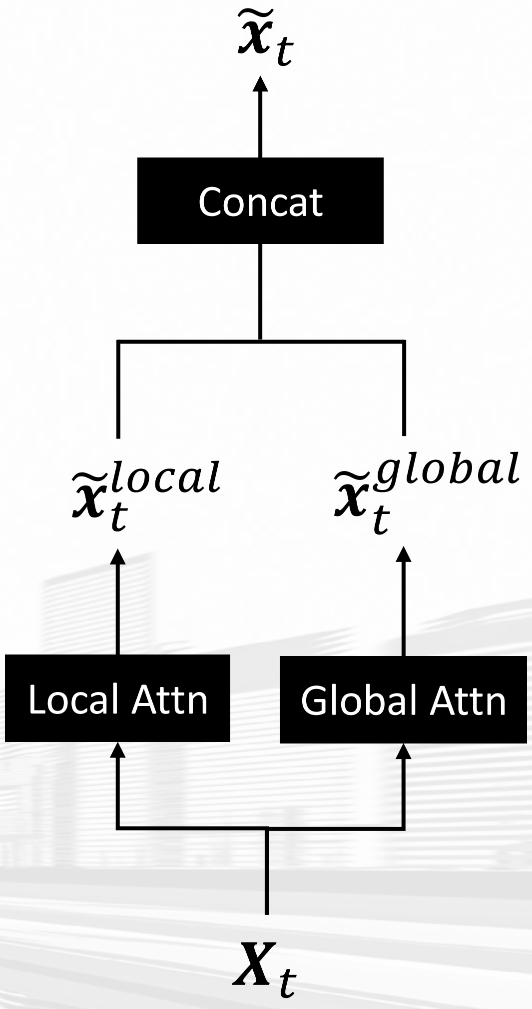

External Factor Fusion
Multi-level Attention Network
POIs & Sensor Networks
SensorID
Weather Forecasts
Time Features
Embed
Embed
Embed
Concat
Decoder
LSTM
LSTM
LSTM
Temporal Attn
Encoder
h_0
LSTM
LSTM
LSTM
Spatial Attn
Spatial Attn
Spatial Attn
X_1
X_2
X_T
Spatial Attention
Concat
Local
Global

Spatial Attention

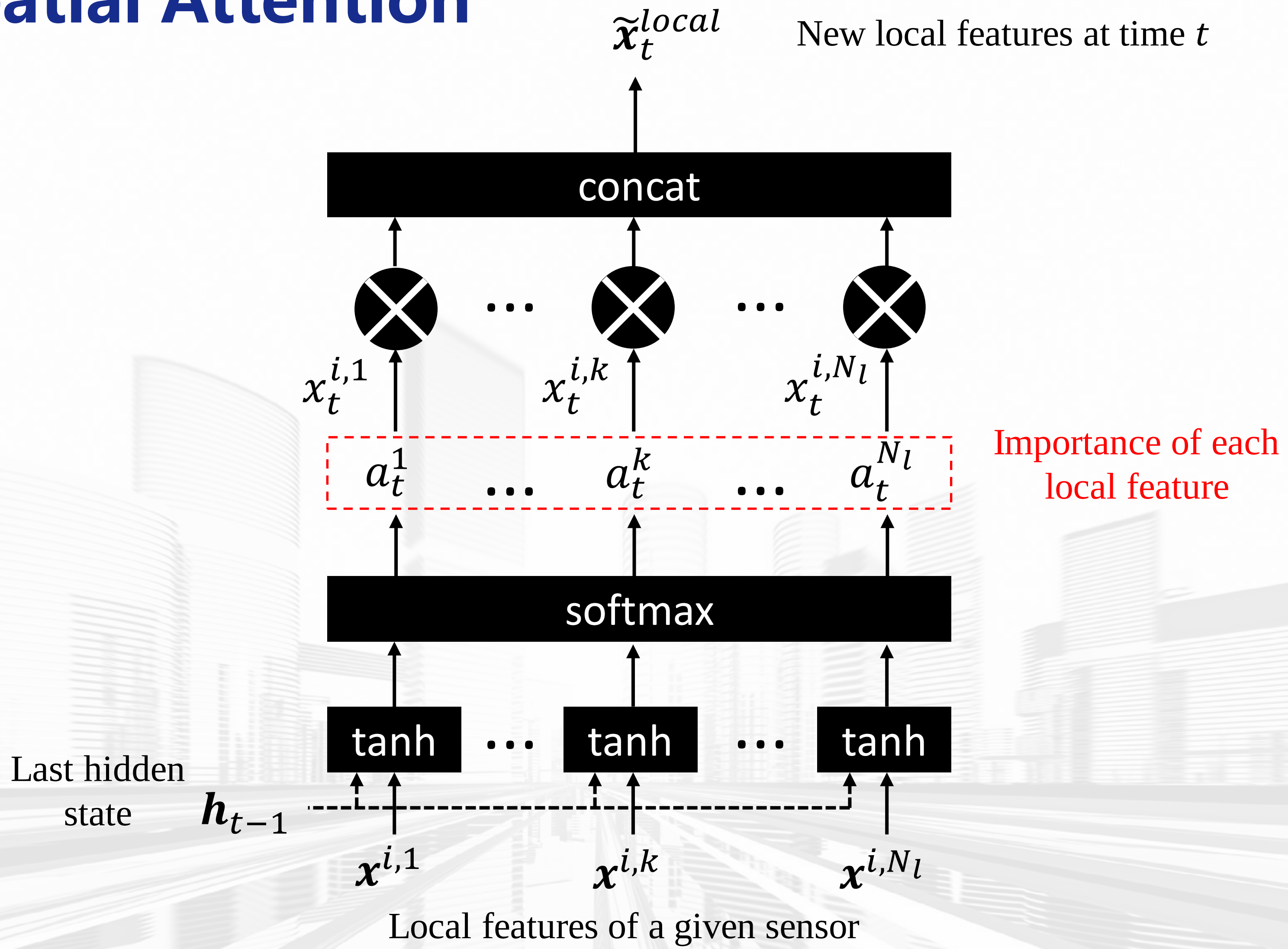
- Capture the dynamic inter-sensor correlation
- Local spatial attention
  - Adaptively capture the dynamic correlation between target series and each local feature
- Global spatial attention
  - Adaptively select relevant sensors to make predictions



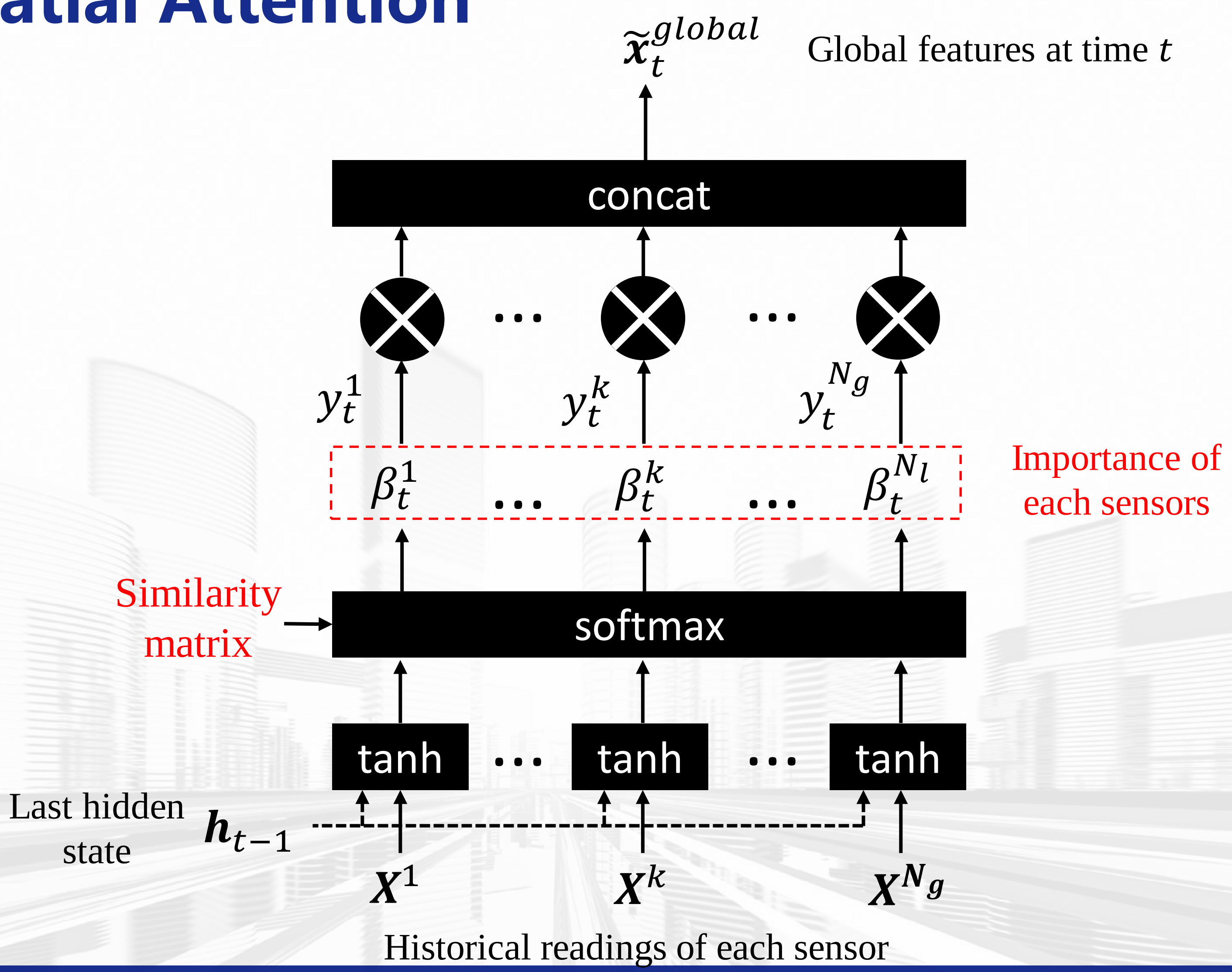




Local Spatial Attention

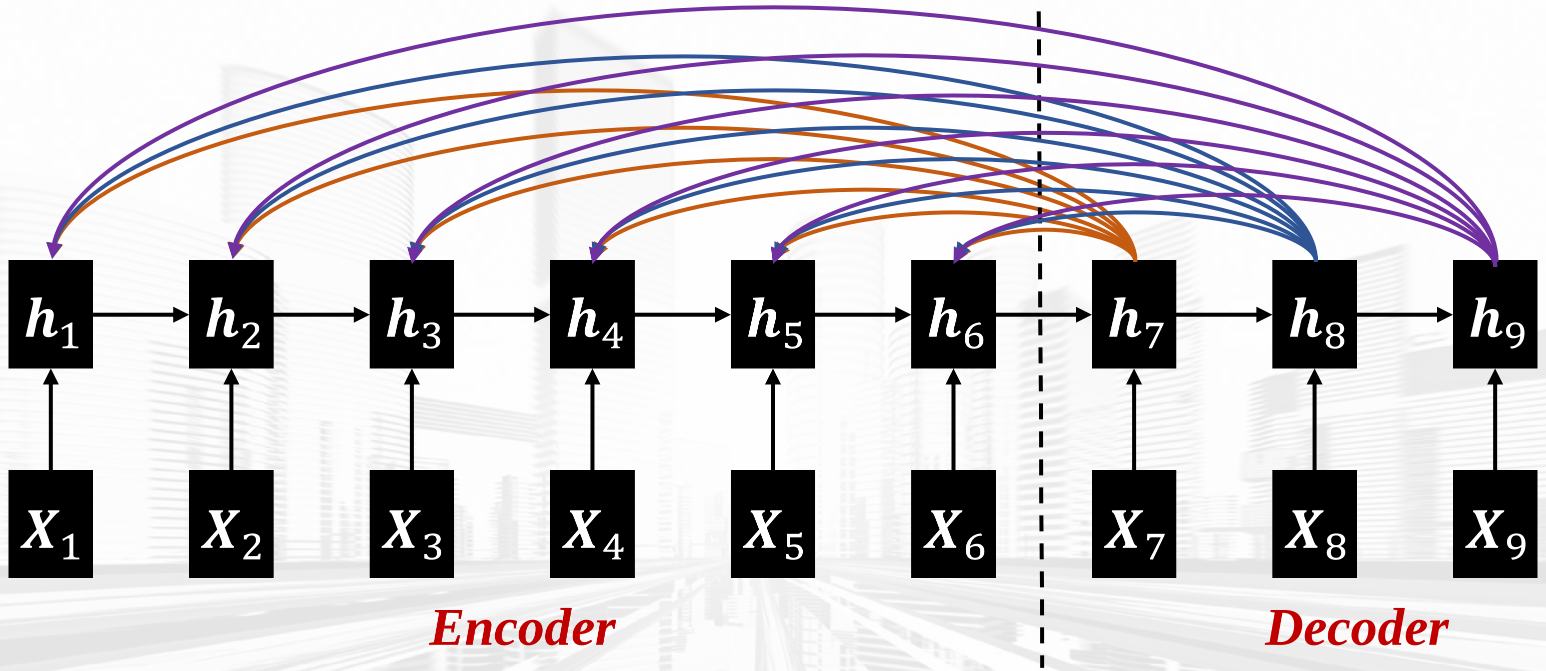


Global Spatial Attention



Temporal Attention

- Select relevant previous time slots to make predictions



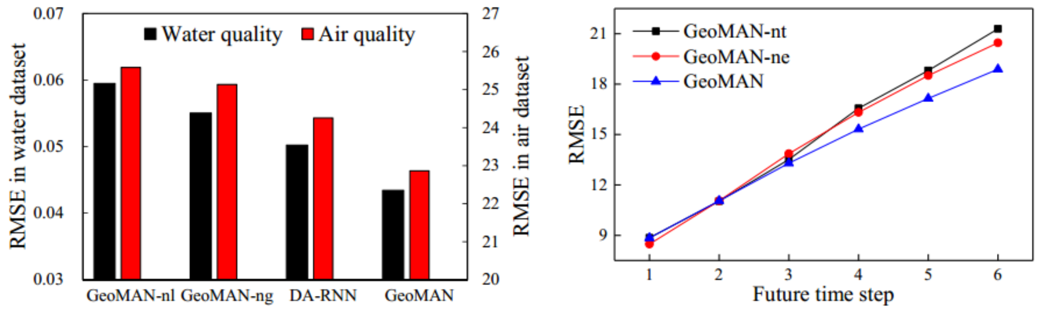
Datasets

- Water quality data
  - Residual chlorine, turbidity, pH, flow, *etc.*
  - From 14 sensors in Shenzhen
  - Update each 5 minutes
- Air quality data
  - PM2.5, PM10, NO2, SO2, O3, CO, *etc.*
  - From 35 sensors in Beijing
  - Hourly updates
- Weather forecasts
- POIs data
- Sensor networks

Dataset		Water Quality	Air Quality
Target series		RC	PM2.5
#Sensors		14	35
#Attributes		10	19
Time Spans		1/1/2012-2014/12/31	8/20/2014-2017/11/30
Time Intervals		5 minutes	1 hour
#Instances		4,415,040	920,640
Meteorology	#Sensors	8	16
	#Attributes	6	13
POIs	#POIs	185,841	651,016
	#Categories	20	20

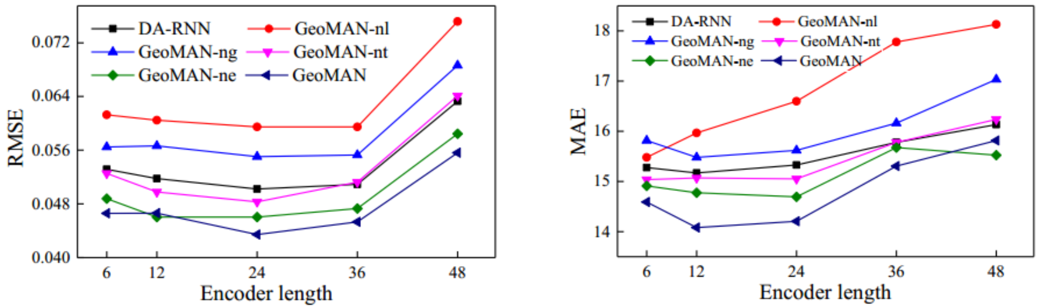
Results

Method	Water Quality		Air Quality	
	RMSE	MAE	RMSE	MAE
ARIMA	8.61E-02	7.97E-02	31.07	20.58
VAR	5.02E-02	4.42E-02	24.60	16.17
GBRT	5.17E-02	3.30E-02	24.00	15.03
FFA	6.04E-02	4.10E-02	23.83	15.75
stMTMVL	6.07E-02	4.16E-02	29.72	19.26
stDNN	5.77E-02	3.99E-02	25.64	16.49
LSTM	6.89E-02	5.04E-02	24.62	16.70
Seq2seq	5.80E-02	4.03E-02	24.55	15.09
DA-RNN	5.02E-02	3.52E-02	24.25	15.17
GeoMAN	**4.34E-02**	**3.02E-02**	**22.86**	**14.08**



(a) Evaluation on spatial attention (b) Future time step vs. RMSE

Figure 3: Performance comparison among different vairants.

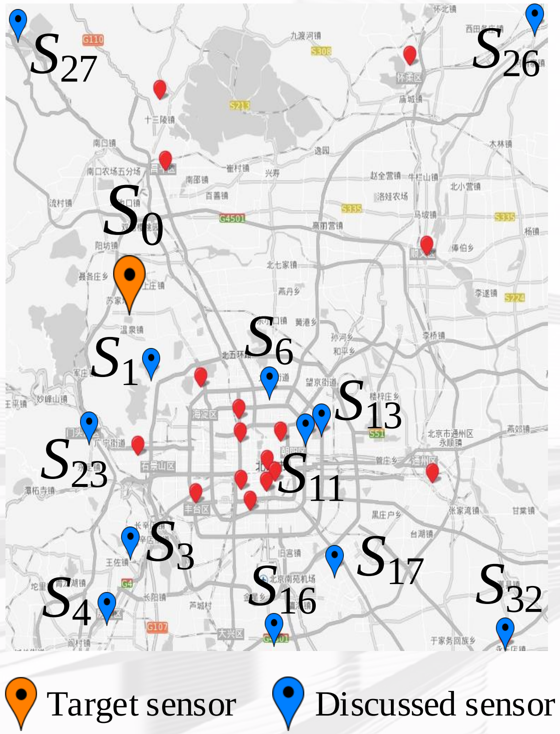


(a) Results on water quality. (b) Results on air quality.

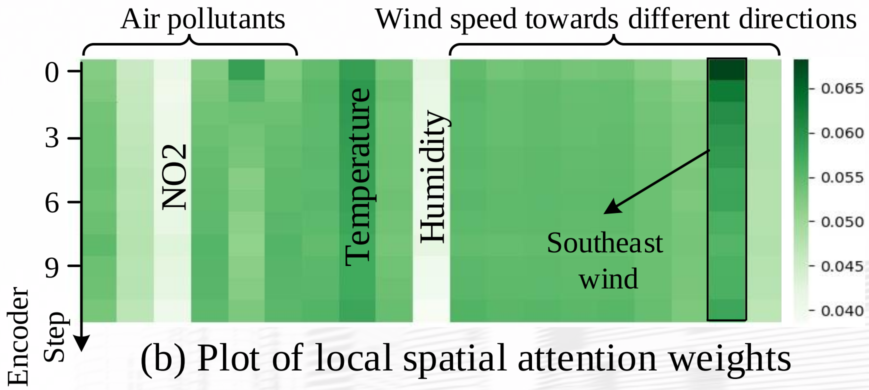
Figure 4: Encoder length vs. metrics over the two datasets.

Attention Visualization

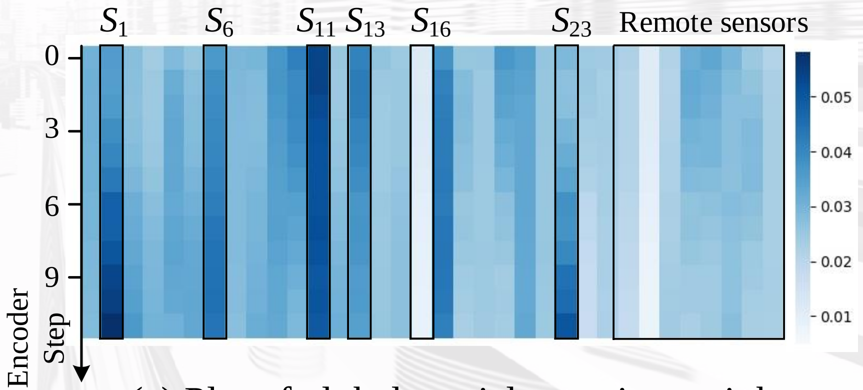
- Case study over air quality dataset
  - Discuss on sensor S_0
  - 4:00 to 16:00 on Feb. 28, 2017



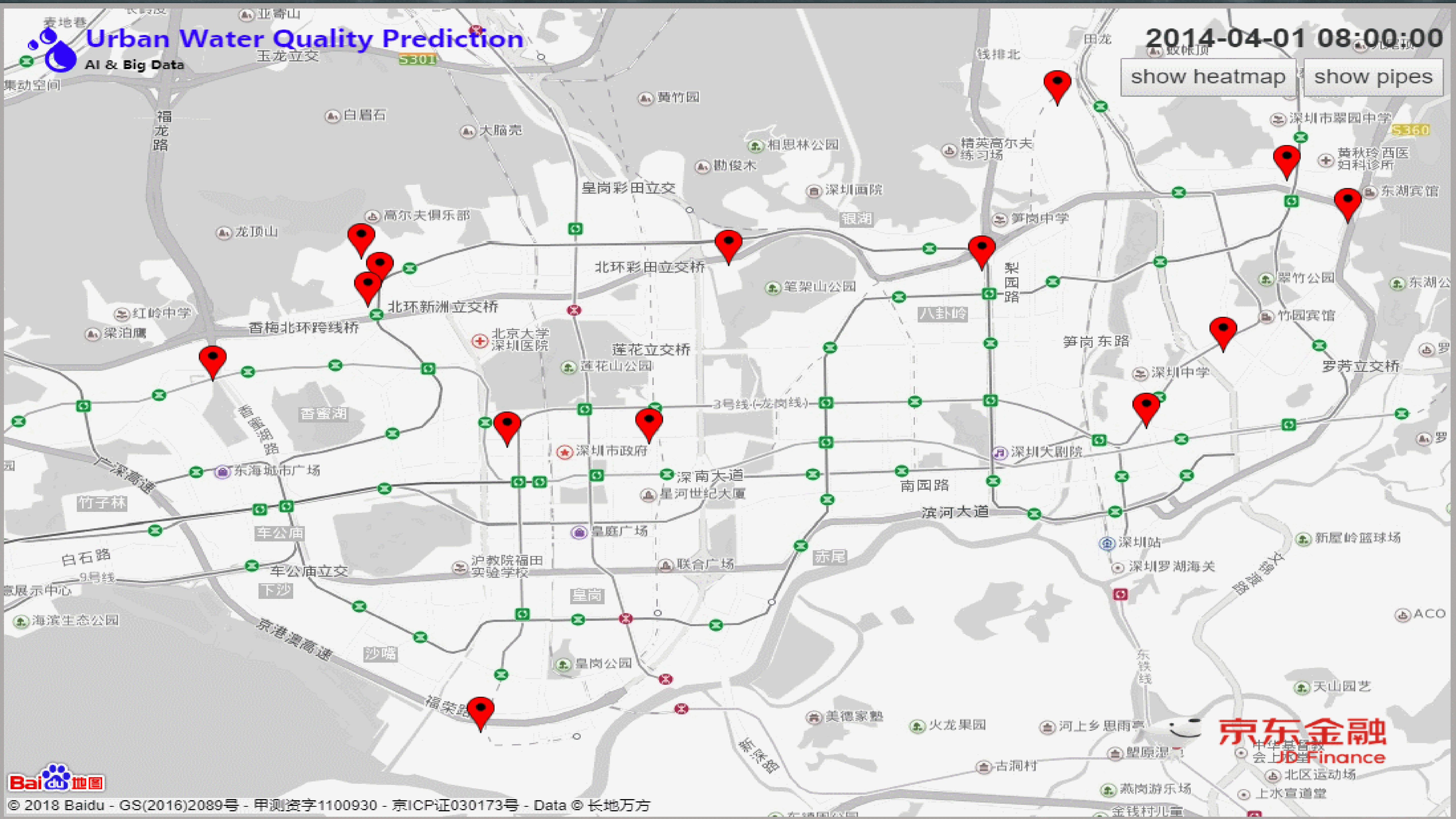
(a) Air quality stations in Beijing



(b) Plot of local spatial attention weights



(c) Plot of global spatial attention weights

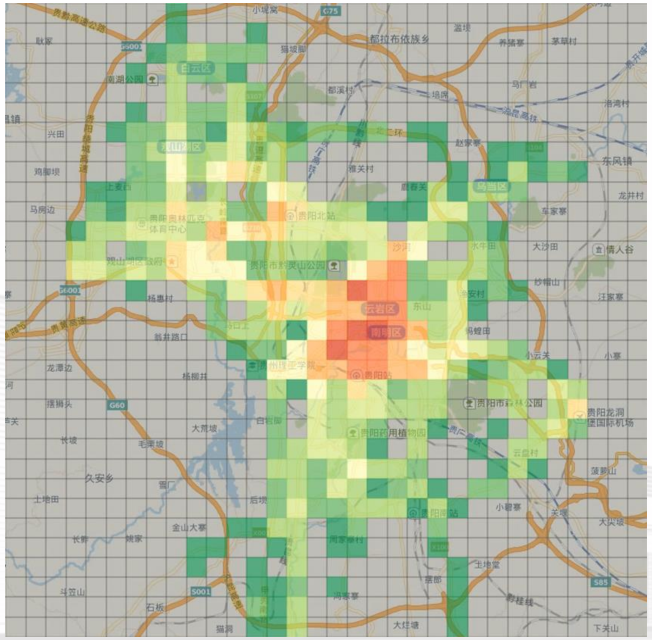


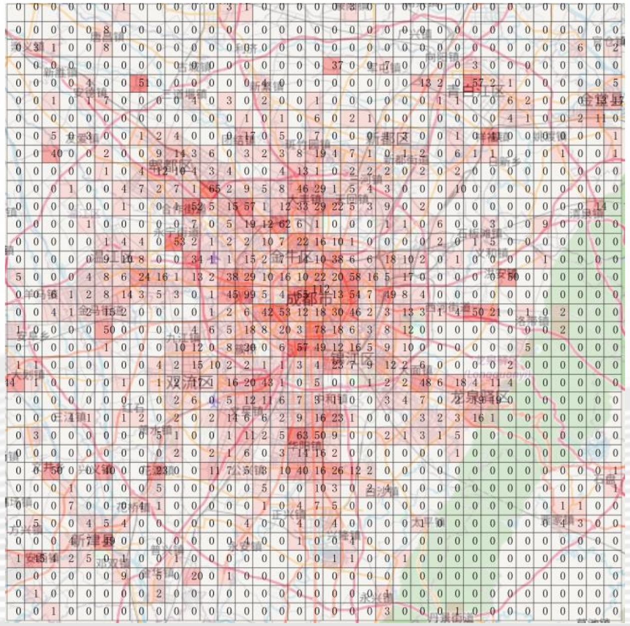
Urban Water Quality Prediction
AI & Big Data
2014-04-01 08:00:00
show heatmap
show pipes
黄竹园
白眉石
大脑壳
相思林公园
勘俊木
精英高尔夫练习场
深圳市翠园中学
黄秋玲西医妇科诊所
东湖宾馆
皇岗彩田立交
深圳画院
银湖
笋岗中学
高尔夫俱乐部
龙顶山
北环彩田立交桥
笔架山公园
梨园路
翠竹公园
东湖公
北环新洲立交桥
八卦岭
竹园宾馆
红岭中学
香梅北环跨线桥
北京大学深圳医院
笋岗东路
梁泊鹰
莲花立交桥
莲花山公园
深圳中学
罗芳立交桥
3号线(龙岗线)
香蜜湖
深圳市政府
深圳大剧院
东海城市广场
深南大道
南园路
星河世纪大厦
滨河大道
竹子林
车公庙
皇庭广场
深圳站
新屋岭篮球场
白石路
沪教院福田实验学校
联合广场
赤尾
深圳罗湖海关
车公庙立交
9号线
下沙
皇岗
海滨生态公园
京港澳高速
沙嘴
皇岗公园
天山园艺
福荣路
美德家塾
火龙果园
河上乡思雨亭
古洞村
北区运动场
燕岗游乐场
上水宣道堂
京东金融
JD Finance
Bai地图
© 2018 Baidu - GS(2016)2089号 - 甲测资字1100930 - 京ICP证030173号 - Data © 长地万方

Spatio-Temporal Neural Networks

- ST Point Data: GeoMAN
- ST Gridded Data: ST-ResNet
- ST Network (Graph) Data: MVGCN, MDL
- ST Sequence Data: DeepTTE

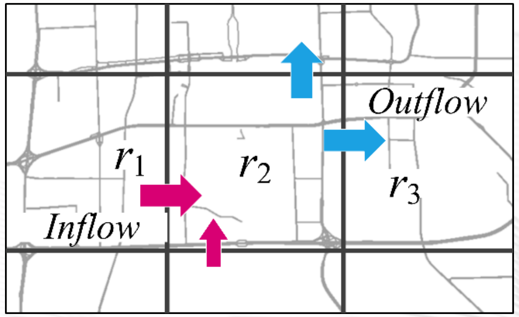
ST Gridded Data

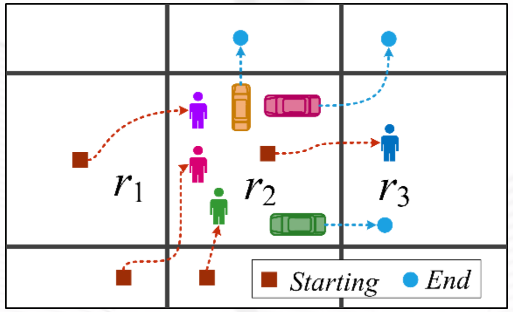


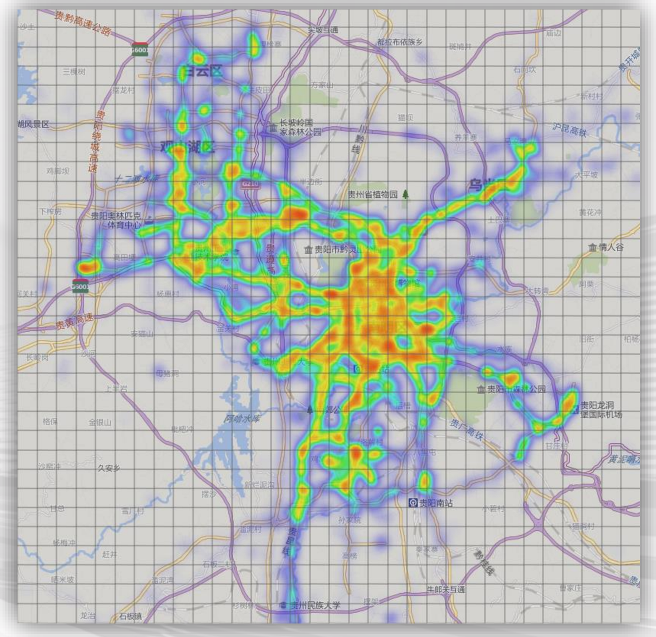


DNN-Based Urban Flow Prediction

Predict In-flow and out-flow of crowds in each region at next time interval throughout a city



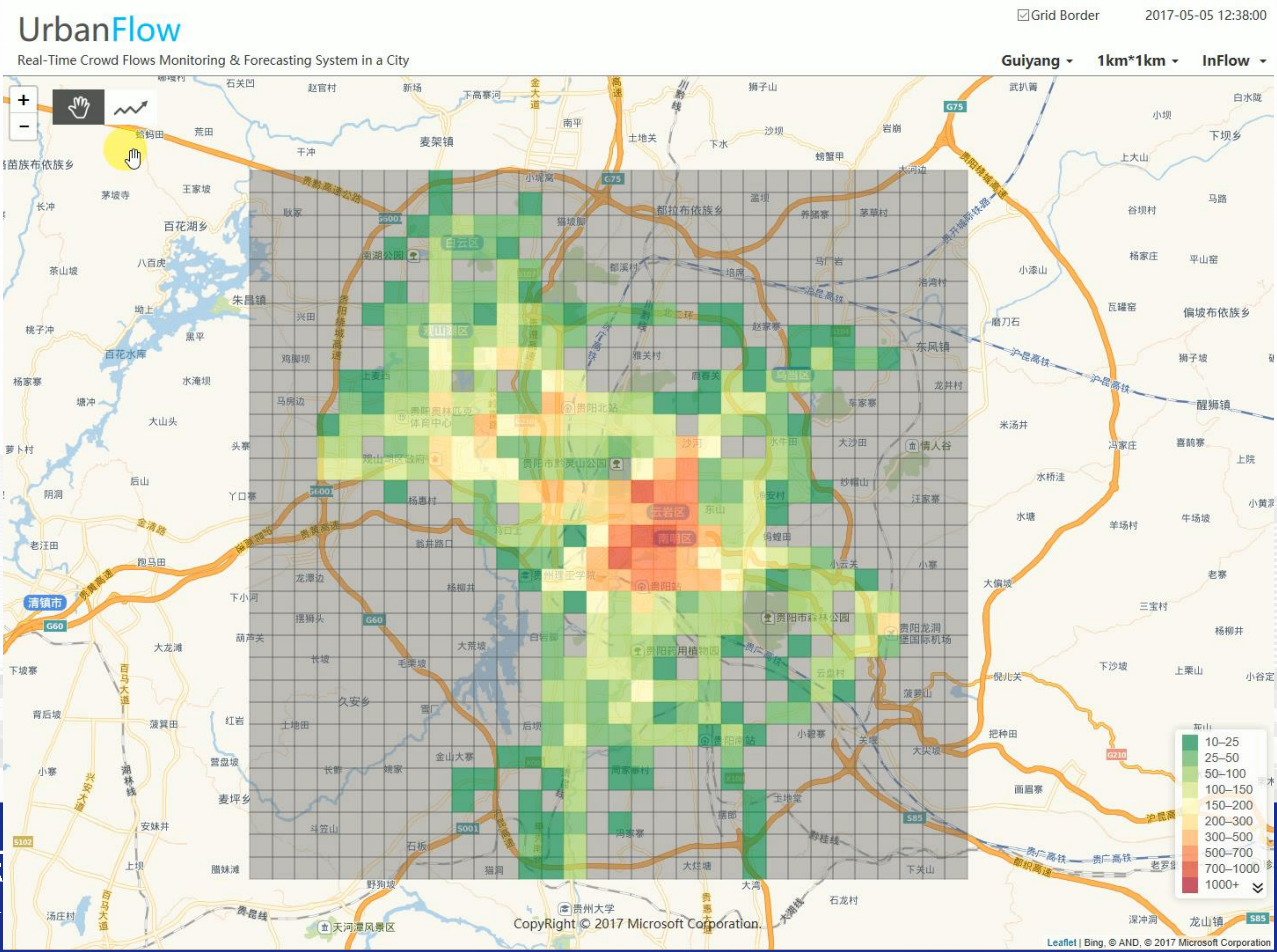




- Important for:
  - Traffic management
  - Risk assessment
  - Public safety

Junbo Zhang, Yu Zheng, et al. DNN-Based Prediction Model for Spatial-Temporal Data. ACM SIGSPAITAL 2016

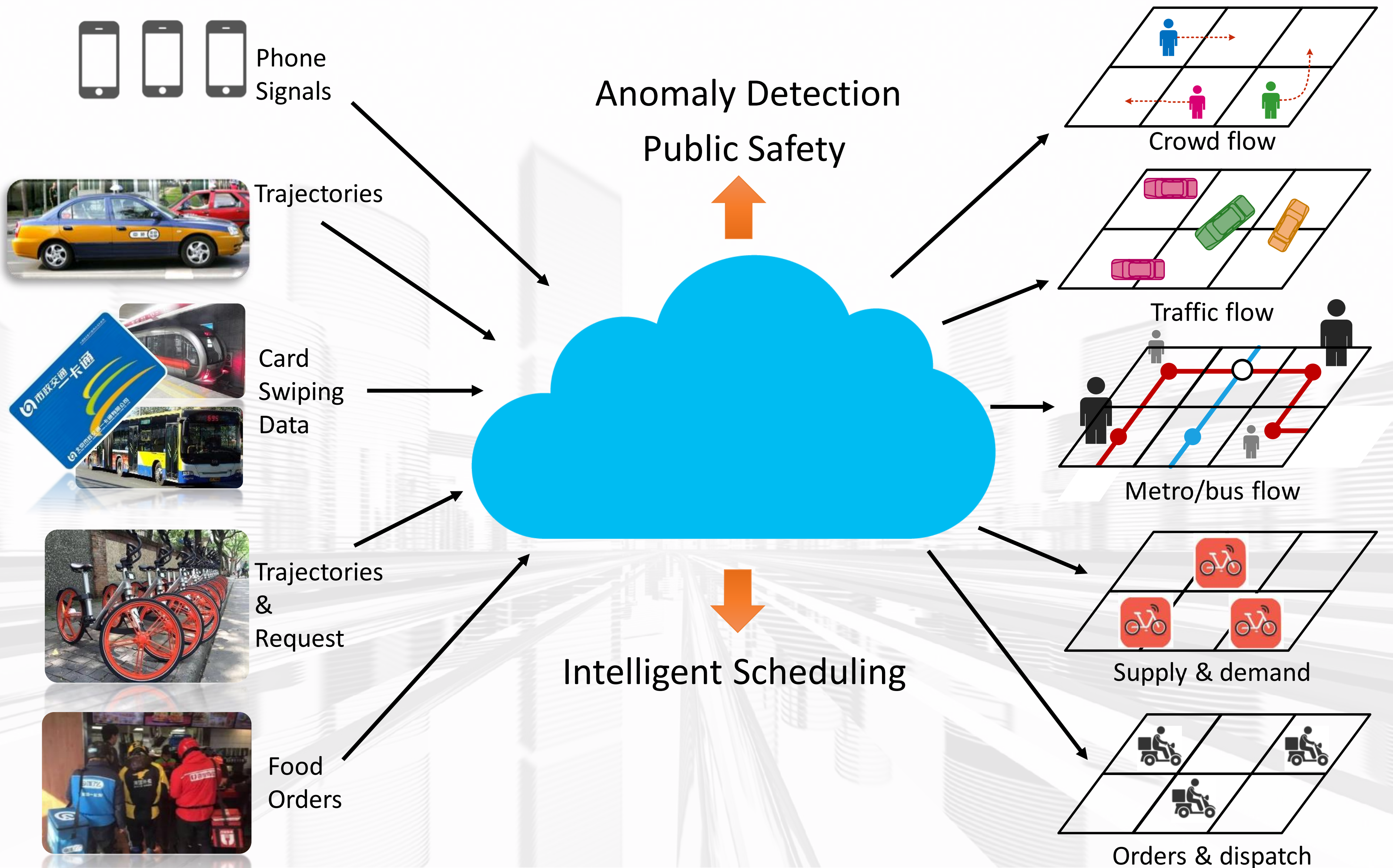
Junbo Zhang, Yu Zheng, et al. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction, AAAI 2017



UrbanFlow
Real-Time Crowd Flows Monitoring & Forecasting System in a City
Grid Border
2017-05-05 12:38:00
Guiyang
1km*1km
InFlow
10–25
25–50
50–100
100–150
150–200
200–300
300–500
500–700
700–1000
1000+
CopyRight © 2017 Microsoft Corporation.
Leaflet | Bing, © AND, © 2017 Microsoft Corporation

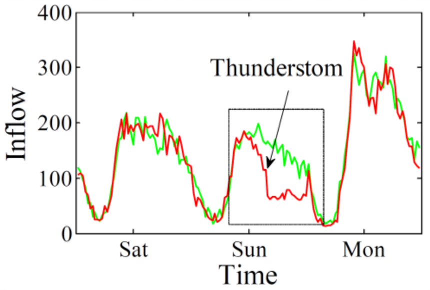
京 东 数 科

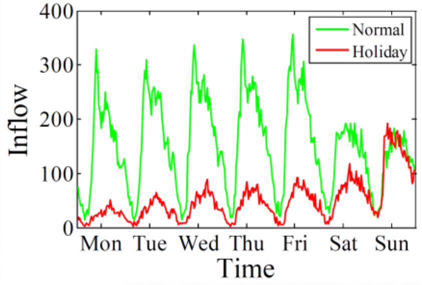
Many Applications

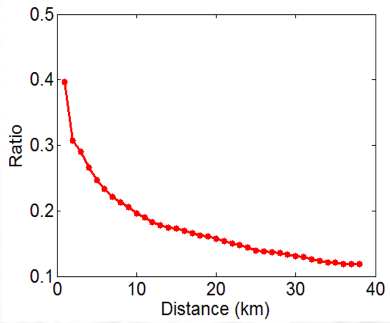


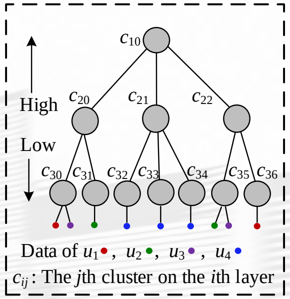
Challenges

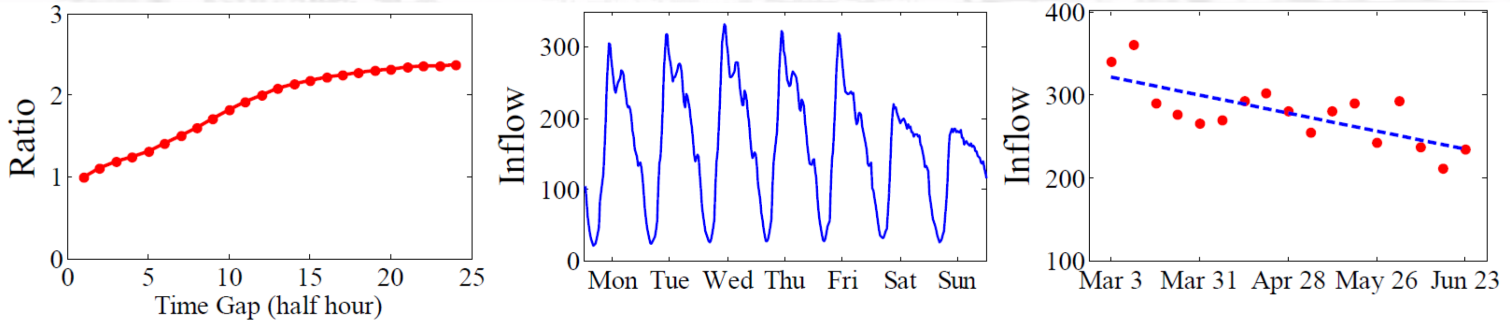
- Urban crowd flow depends on many factors
  - Flows of previous time interval
  - Flows of nearby regions and distant regions
  - Weather, traffic control and events
- Capturing spatial properties
  - Spatial distance and hierarchy
- Capturing temporal properties
  - Temporal closeness
  - Period and trend





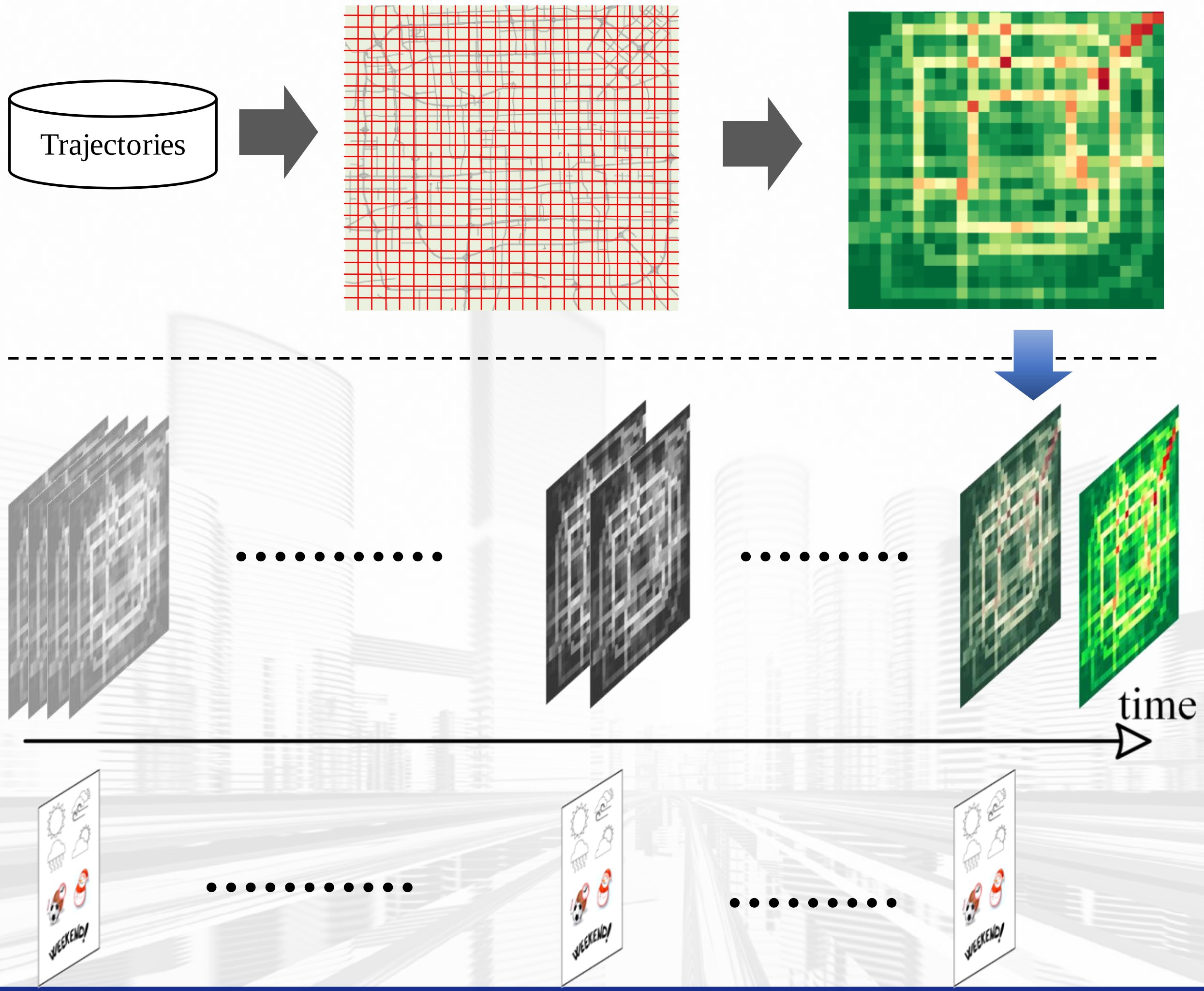






(a) Closeness of Office Area (b) Period of Office Area (c) Trend of Office Area

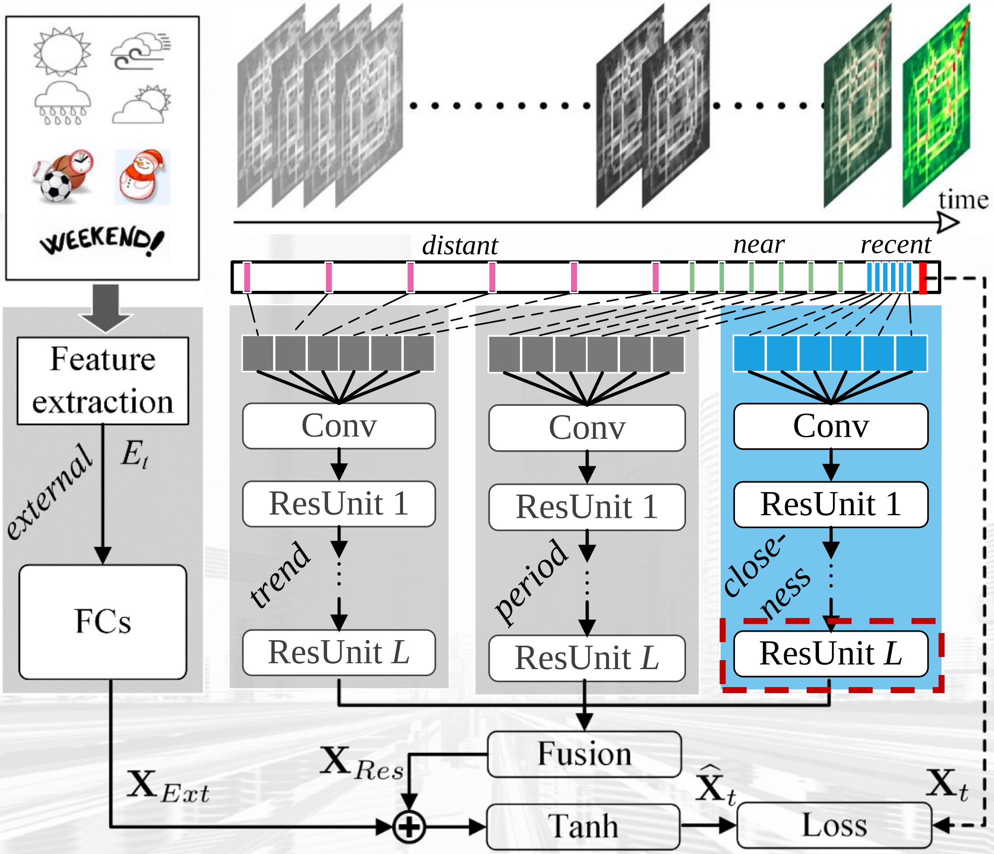
Converting Trajectories into Video-like Data



京东城市
京 东 数 科 旗 下

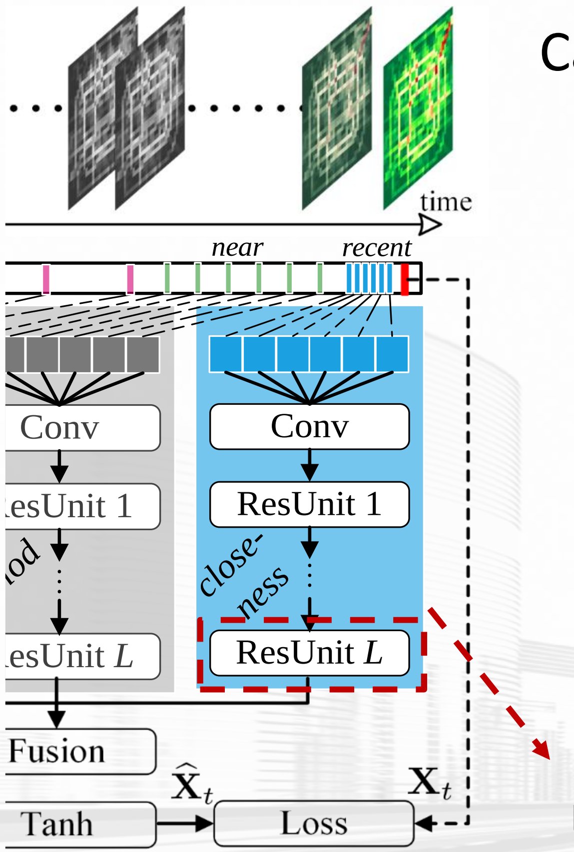
Junbo Zhang, Yu Zheng, et al. [Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction](), AAAI 2017

ST-ResNet Architecture: A Collective Prediction

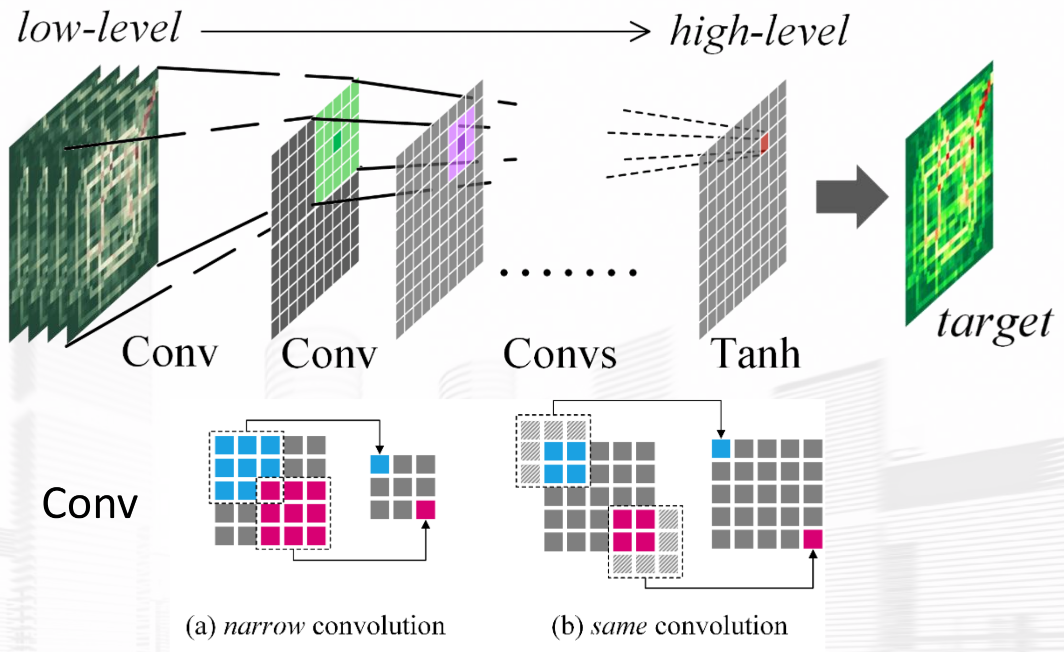


Junbo Zhang, Yu Zheng, et al. [Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction](), AAAI 2017

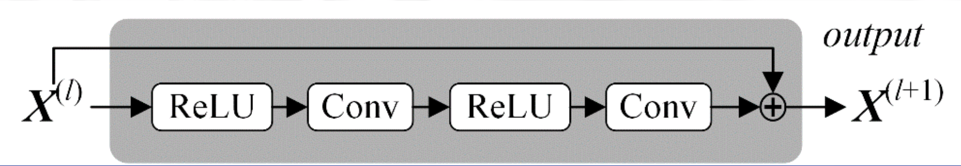
Residual Deep Convolutional Neural Network



Capturing spatial correlation of both near and far



Using residual network framework to help training



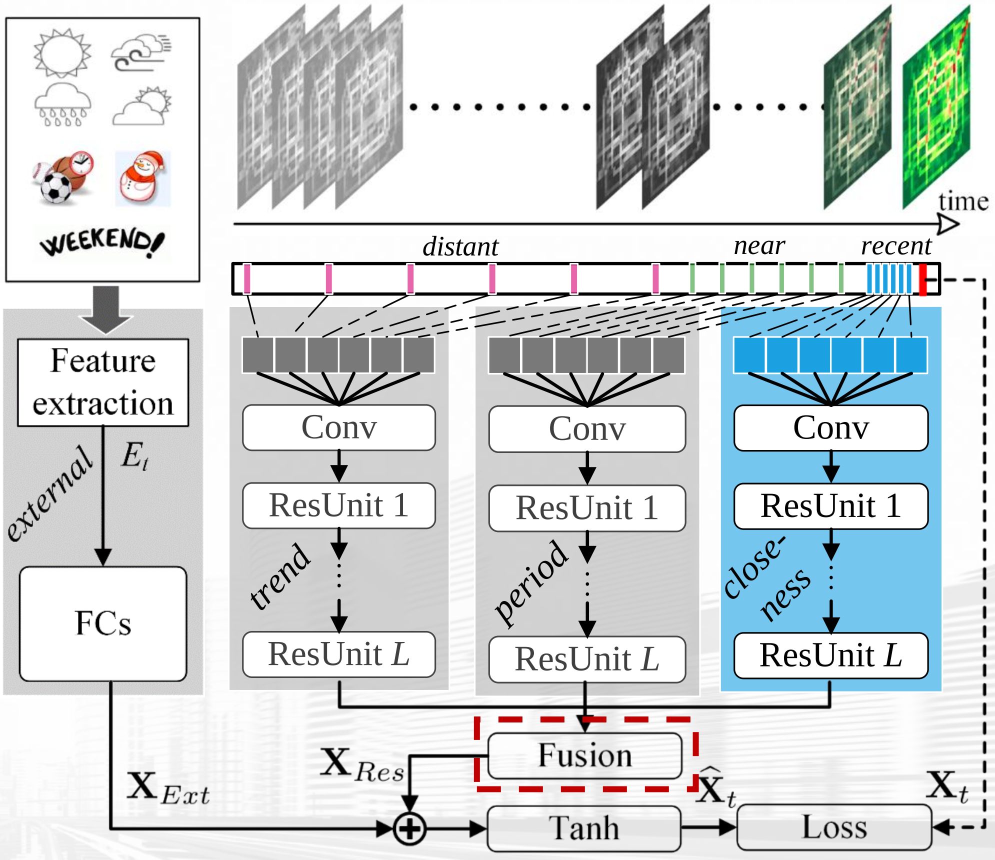
京东城市
京 东 数 科 旗 下

Junbo Zhang, Yu Zheng, et al. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction, AAAI 2017

ST-ResNet Architecture

- A Collective Prediction
- Capture temporal closeness, period and trend
- Capture external factors
- Capture spatial correlation of both near and far distances
- Fusing factors differently in different regions

$$\begin{bmatrix} (\omega_{c,1},\omega_{p,1},\omega_{q,1}) & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & \omega_{c,n},\omega_{p,n},\omega_{q,n} \end{bmatrix}$$



$$\mathbf{X}_{Res} = \mathbf{W}_c \circ \mathbf{X}_c^{(L+2)} + \mathbf{W}_p \circ \mathbf{X}_p^{(L+2)} + \mathbf{W}_q \circ \mathbf{X}_q^{(L+2)}$$

Junbo Zhang, Yu Zheng, et al. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction, AAAI 2017

Experiments

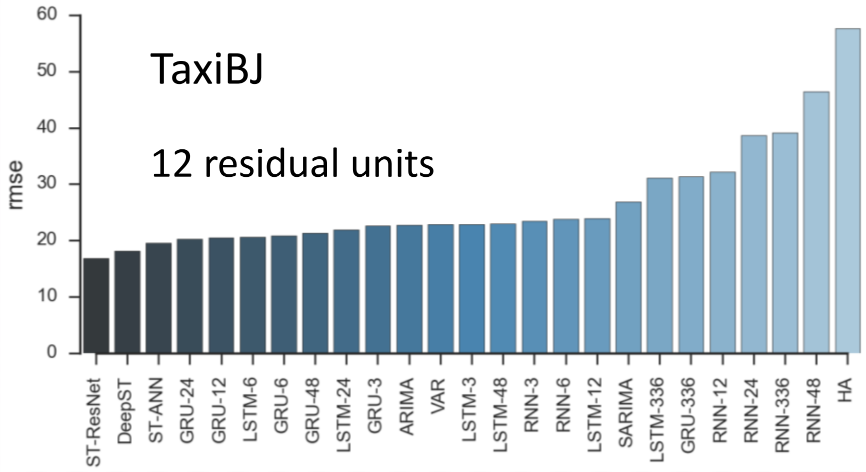
Datasets

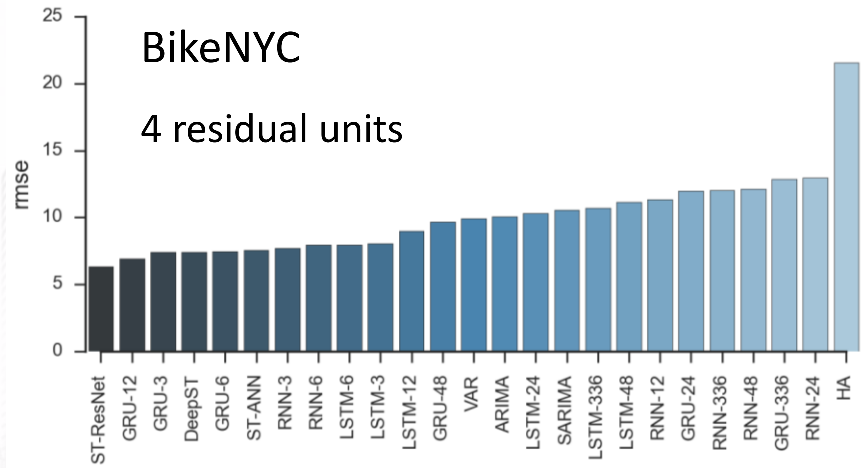
> 700GB

Dataset	TaxiBJ	BikeNYC
Data type	Taxi GPS	Bike rent
Location	Beijing	New York
Time Span	7/1/2013 - 10/30/2013 3/1/2014 - 6/30/2014 3/1/2015 - 6/30/2015 11/1/2015 - 4/10/2016	4/1/2014 - 9/30/2014
Time interval	30 minutes	1 hour
Gird map size	(32, 32)	(16, 8)
Trajectory data		
Average sampling rate (s)	~ 60	\
# taxis/bikes	34,000+	6,800+
# available time interval	22,459	4,392
External factors (holidays and meteorology)		
# holidays	41	20
Weather conditions	16 types (*e.g.*, Sunny, Rainy)	\
Temperature / °C	$[-24.6, 41.0]$	\
Wind speed / mph	$[0, 48.6]$	\

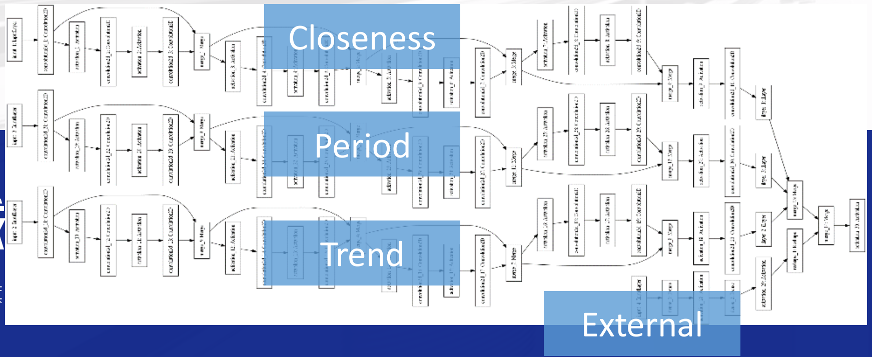
Results

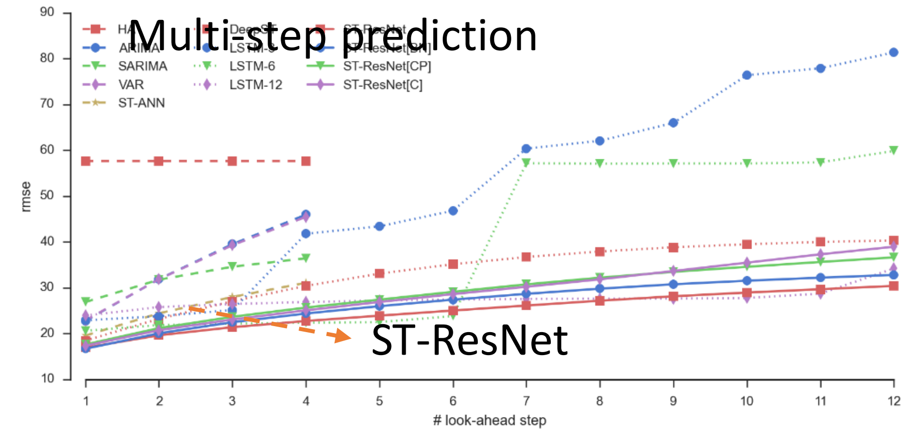
Baseline	Spatial	Temporal	Intra-region dependence	External Factors
HA		✓		
ARIMA		✓		
SARIMA		✓		
VAR		✓	✓	
RNN/ LSTM/ GRU (3, 6, 12, 24, 48, 336)		✓	✓	
ST-ANN	✓	✓		
DeepST	✓	✓	✓	✓
ST-ResNet	✓	✓	✓	✓





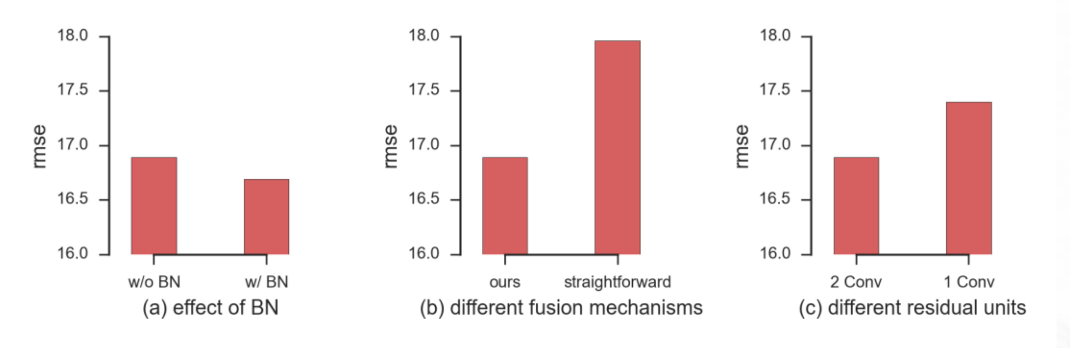
ST-ResNet [with 4 residual units]



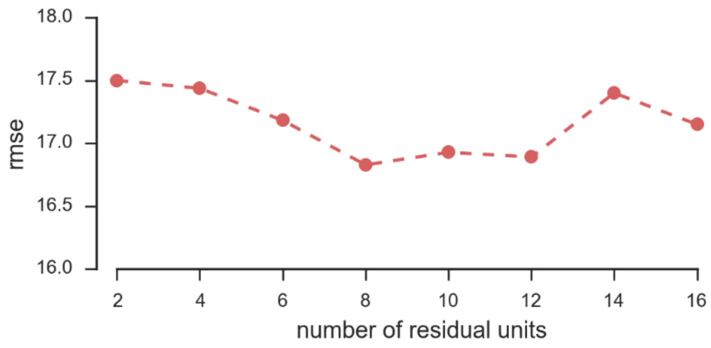


Experiments

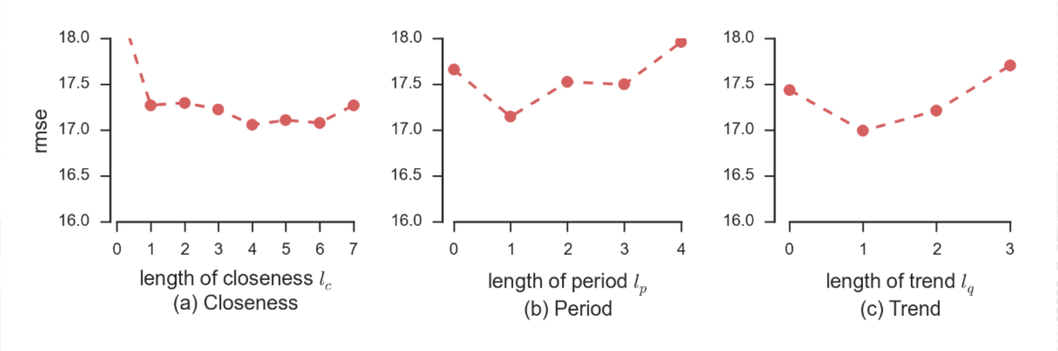
(a) Different network architectures

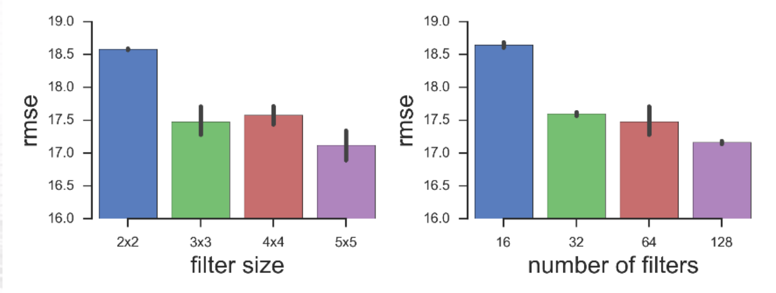


(c) Parameter sensitivity



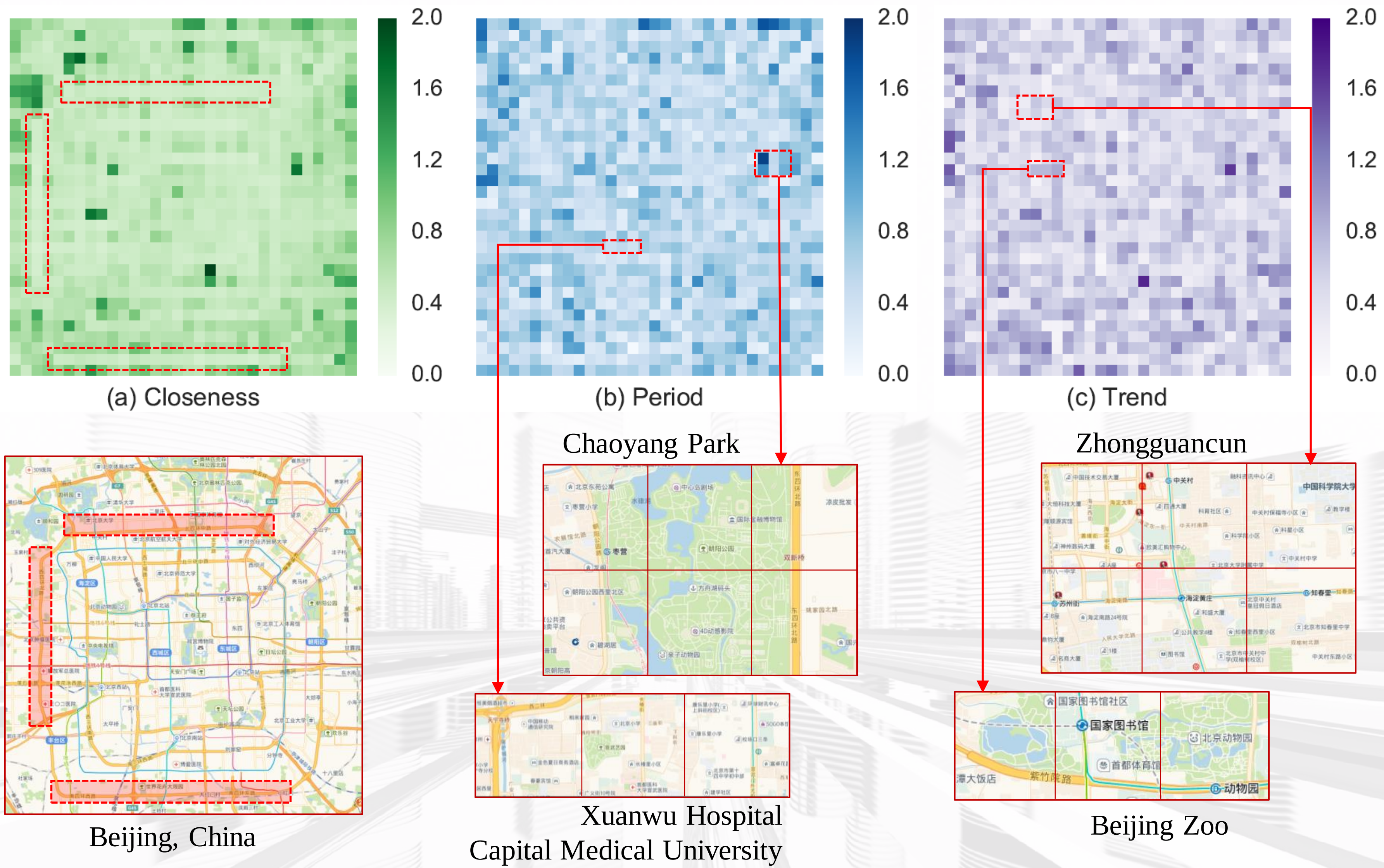
(b) Different components





Junbo Zhang, Yu Zheng, et al. Predicting Citywide Crowd Flows Using Deep Spatio-Temporal Residual Networks. AI Journal, 2018

Visualization of the Fusion Component

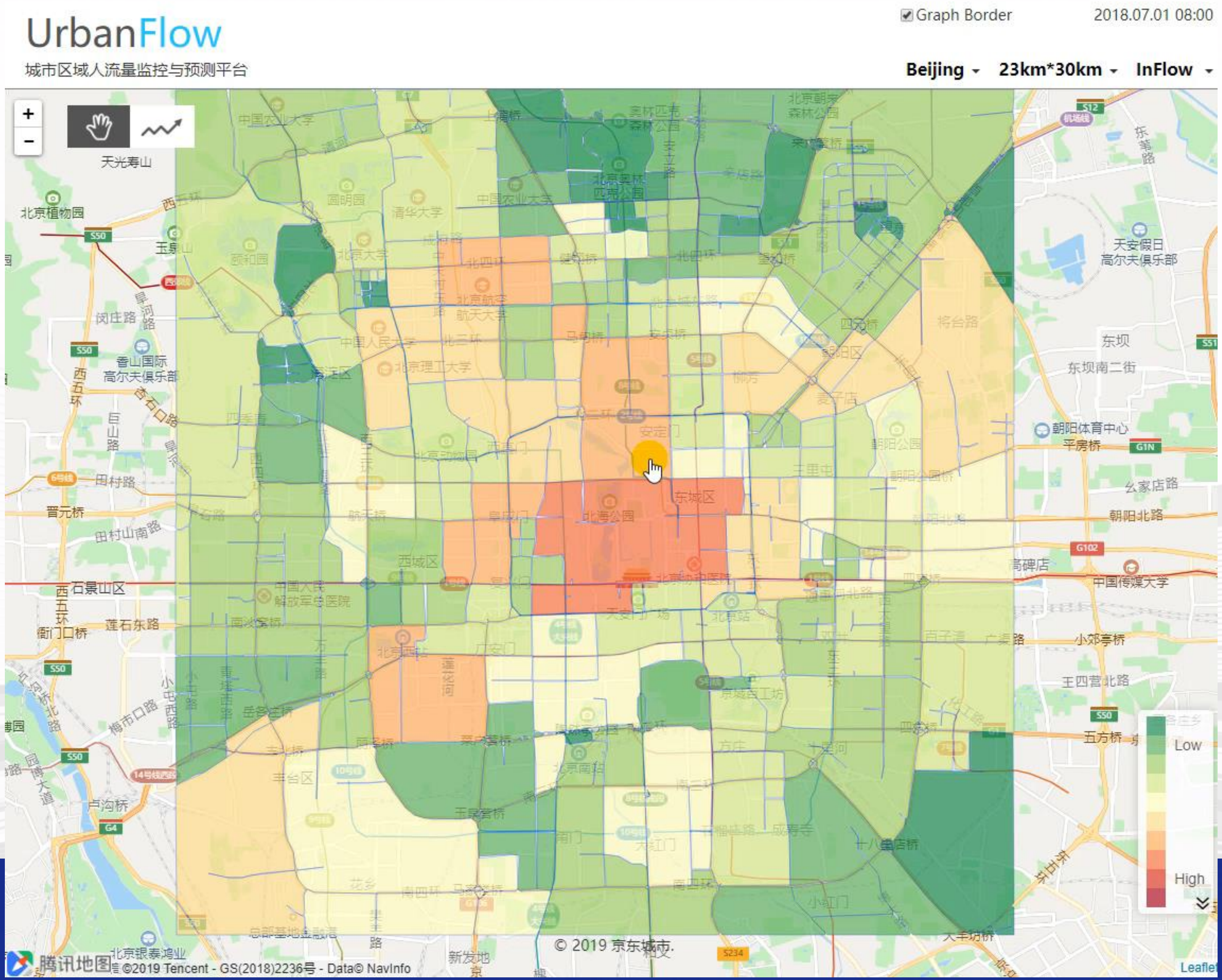


Junbo Zhang, Yu Zheng, et al. Predicting Citywide Crowd Flows Using Deep Spatio-Temporal Residual Networks. AI Journal, 2018

Spatio-Temporal Neural Networks

- ST Point Data: GeoMAN
- ST Gridded Data: ST-ResNet
- ST Network (Graph) Data: MVGCN, MDL
- ST Sequence Data: DeepTTE

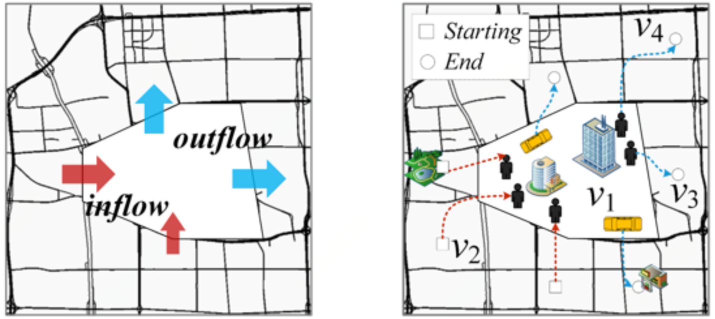
AI预测城市不规则区域人流量



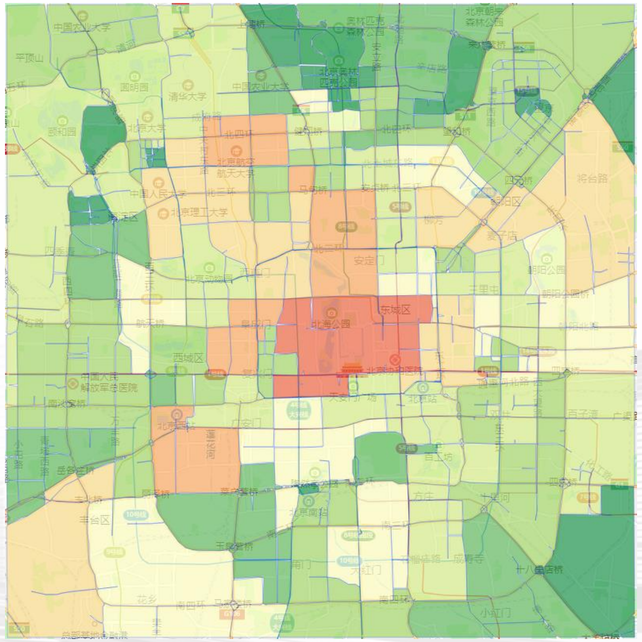
京东城市
京 东 数 科 旗 下

Urban Flow Prediction In Irregular Regions

Predict Inflow and outflow of crowds in each irregular region at next time interval throughout a city



(a) Inflow and outflow (b) Measurement of flows

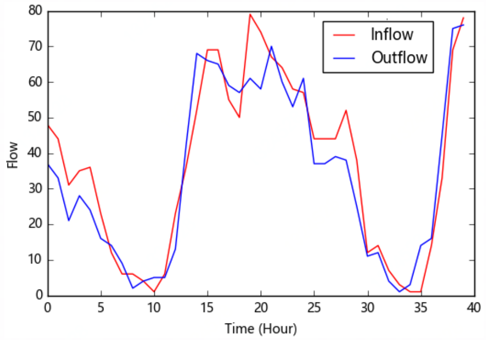


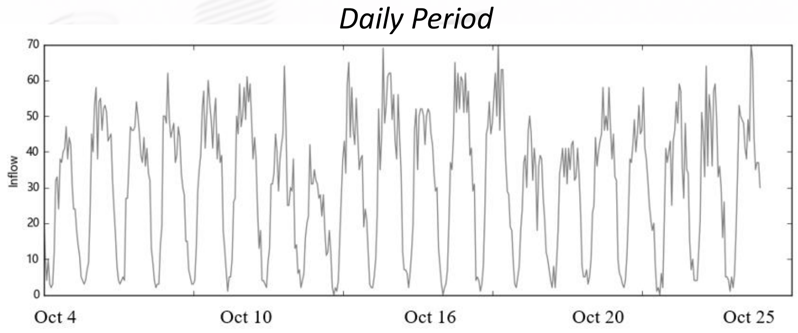
- Important for:
  - Traffic management
  - Risk assessment
  - Public safety

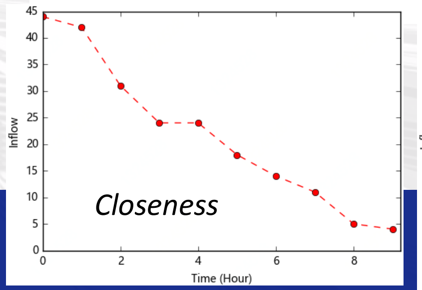
Junkai Sun, Junbo Zhang, et al. Predicting Citywide Crowd Flows in Irregular Regions Using Multi-View Graph Convolutional Networks. *IEEE TKDE 2020*

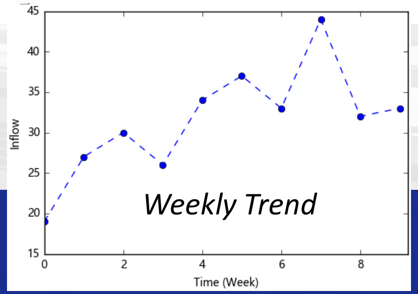
Challenges

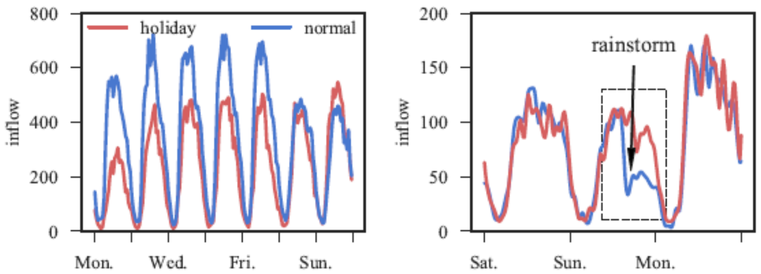
- Interactions and spatial correlations between different regions
  - Flows of adjacent (1-hop) regions
  - Flows of distant (multi-hop) regions
  - Region's inflow and outflow interact with each other
- Temporal correlations among different time intervals
  - Flows of recent previous time intervals (*Closeness*)
  - Flows of daily periodic time intervals (*Periods*)
  - Flows of weekly, monthly, quarterly time intervals (*Trends*)
- External factors and meta features
  - Holidays, weekdays, weekends
  - Weather information (rainstorm etc.)





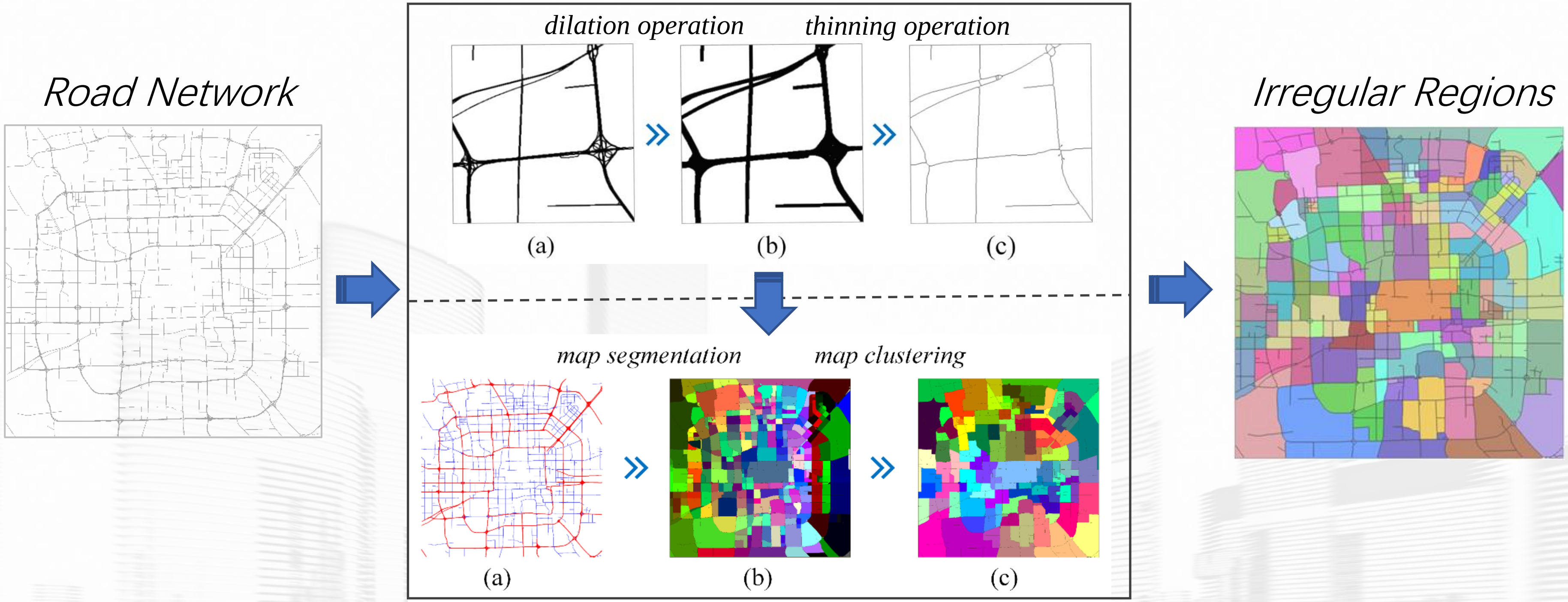






(a) holiday influence (b) weather influence

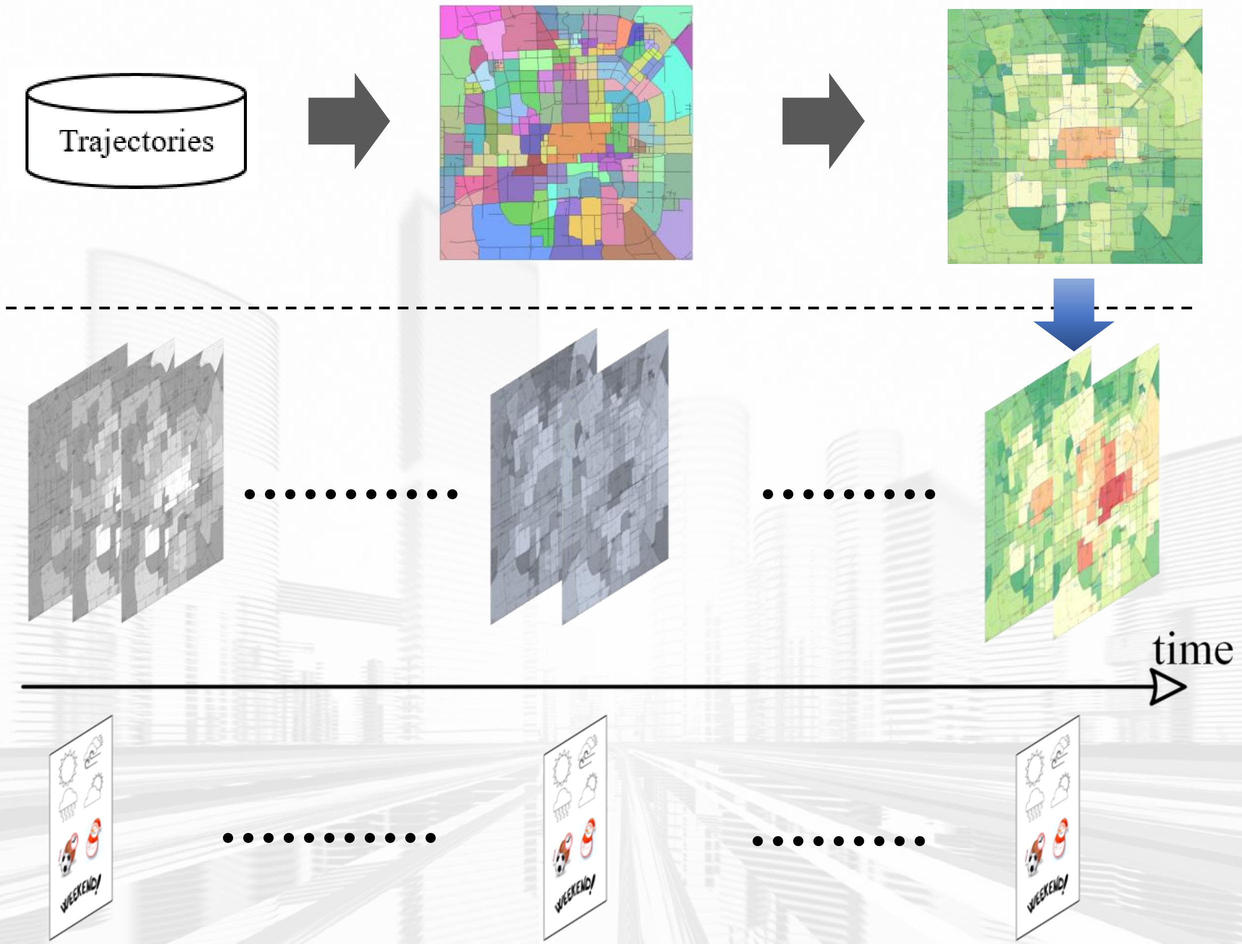
Irregular Regions Construction



Cluster low-level regions into high-level regions that:

1. Are adjacent on the geographical map
2. Have similar crowd flow patterns

Mapping Trajectories Into Irregular Regions



Junkai Sun, Junbo Zhang, et al. Predicting Citywide Crowd Flows in Irregular Regions Using Multi-View Graph Convolutional Networks. *IEEE TKDE 2020*

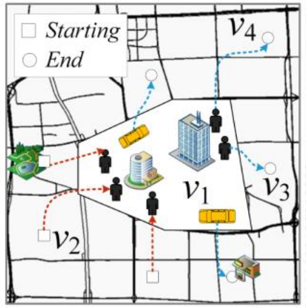
Spatial GCN

Adjacency with self-connections

Degree matrix of $\tilde{A}$ *Inputs* *Parameters*

$$\mathbf{Y} = \tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{X}\mathbf{W}$$

GCN Thomas Kipf et.al. in ICLR 2017



Transition Flow → $\mathbf{A}$ → *Spatial Distance* → ω

$$\mathbf{A} = \begin{pmatrix} 1, & 1, & 0, & 1, & 0, & 0, & 1 \\ 1, & 1, & 0, & 1, & 1, & 1, & 0 \\ 0, & 0, & 1, & 1, & 1, & 1, & 1 \\ 1, & 1, & 1, & 1, & 1, & 0, & 0 \\ 0, & 1, & 1, & 1, & 1, & 0, & 0 \\ 0, & 1, & 1, & 0, & 0, & 1, & 0 \end{pmatrix}$$

$$\omega = \begin{pmatrix} 1. & , 0.06, & 0. & , 0.05, & 0. & , 0. \\ 0.06, & 1. & , 0. & , 0.14, & 0.44, & 0.19 \\ 0. & , 0. & , 1. & , 0.47, & 0.4 & , 0.13 \\ 0.05, & 0.14, & 0.47, & 1. & , 0.37, & 0. \\ 0. & , 0.44, & 0.4 & , 0.37, & 1. & , 0. \\ 0. & , 0.19, & 0.13, & 0. & , 0. & , 1. \end{pmatrix}$$

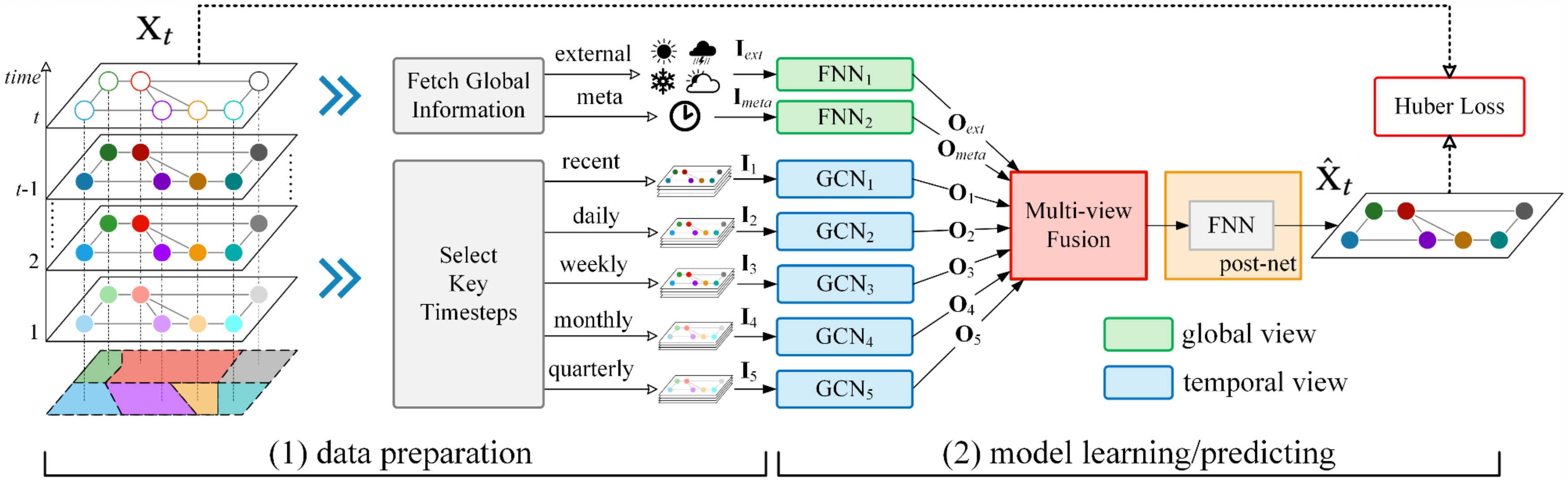
$$\mathbf{S} = \mathbf{A} \odot \omega$$

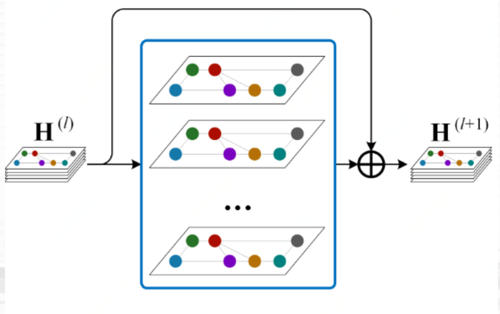
$$\omega_{ij} = \begin{cases} \exp\left(-\frac{[dist(p_i,p_j)]^2}{2\theta^2}\right) & \text{if } \mathrm{dist}(p_i,p_j) \leq \kappa \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{H}^{(l+1)} = f\left(\mathbf{Q}^{-\frac{1}{2}}\tilde{\mathbf{S}}\mathbf{Q}^{-\frac{1}{2}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right)$$

Where $\tilde{S} = S + I_N$, $Q_{ii} = \sum_j \tilde{S}_{ij}$

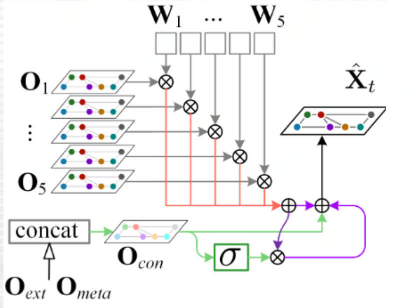
MVGCN Architecture: A Multi-View Framework





(a) GCN-based residual unit

$$\mathbf{H}^{(l+1)} = \mathbf{H}^{(l)} + f\left(\mathbf{Q}^{-\frac{1}{2}}\tilde{\mathbf{S}}\mathbf{Q}^{-\frac{1}{2}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right)$$



(b) Multi-view fusion module

$$\mathbf{O} = \mathbf{W}_1 \odot \mathbf{O}_1 + \mathbf{W}_2 \odot \mathbf{O}_2 + \cdots + \mathbf{W}_5 \odot \mathbf{O}_5$$

$$\hat{\mathbf{X}}_t = f_o\left(\mathbf{O} + \mathbf{O}_{con} + \sigma(\mathbf{O}_{con}) \odot \mathbf{O}\right)$$

Junkai Sun, Junbo Zhang, et al. Predicting Citywide Crowd Flows in Irregular Regions Using Multi-View Graph Convolutional Networks. *IEEE TKDE 2020*

Experiments

4 Datasets

Dataset	TaxiNYC	TaxiBJ	BikeDC	BikeNYC
Data trype	Taxi trip	Taxi GPS	Bike rent	Bike rent
Location	NYC	Beijing	D.C.	NYC
Start time	1/1/2011	7/1/2013	1/1/2011	7/1/2013
End time	6/30/2016	4/10/2016	12/31/2016	12/31/2016
Time interval	1 hour	1 hour	1 hour	1 hour
# timesteps	48192	12336	52608	30720
# regions (stations)	100	100	120 (472)	120 (416)
# holidays	627	105	686	401
Weather	\	16 types	\	\
Temp. / °C	\	[-24.6,41]	\	\
WS / mph	\	[0,48.6]	\	\

Performance Comparison

Dataset	Metric	HA	VAR	GBRT	FC-LSTM	GCN	DCRNN	FCCFnoTrans	FCCF	ST-MGCN	MVGCN
TaxiNYC	RMSE	101.54	30.78	83.71	27.82	26.52	25.50	26.02	26.00	23.53	**23.15**
	MAE	33.02	11.21	23.46	11.25	11.12	11.20	9.25	**9.24**	9.52	9.40
TaxiBJ	RMSE	38.77	18.79	33.89	19.04	17.38	16.44	18.70	18.42	16.30	**14.37**
	MAE	22.89	11.38	20.34	11.86	10.60	9.68	10.74	10.44	10.18	**9.11**
BikeDC	RMSE	2.61	1.95	3.46	1.88	1.88	1.90	2.22	2.14	-	**1.72**
	MAE	1.48	1.20	1.98	1.10	1.08	1.20	1.34	1.27	-	**1.00**
BikeNYC	RMSE	6.77	4.21	8.57	4.66	5.06	4.35	4.41	4.19	-	**4.15**
	MAE	4.00	2.71	5.17	2.78	2.85	2.90	2.79	2.65	-	**2.60**

京东城市

京 东 数 科 旗 下

Junkai Sun, Junbo Zhang, et al. Predicting Citywide Crowd Flows in Irregular Regions Using Multi-View Graph Convolutional Networks. *IEEE TKDE 2020*

Experiments

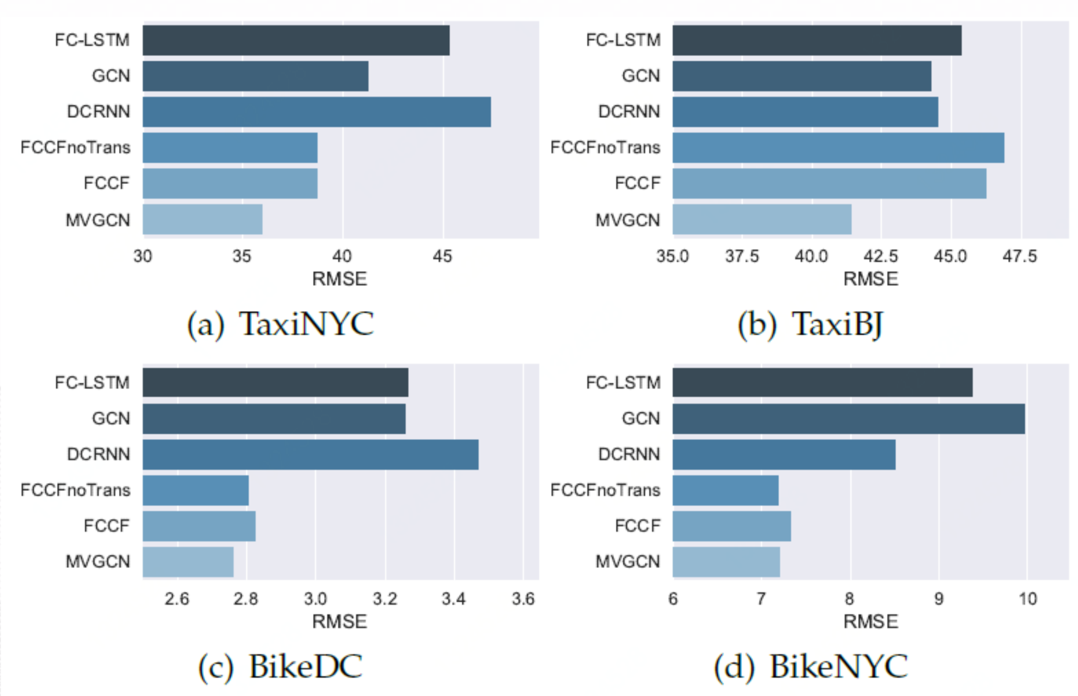


Fig. 7. RMSE comparisons on *sudden changes* in the four datasets.

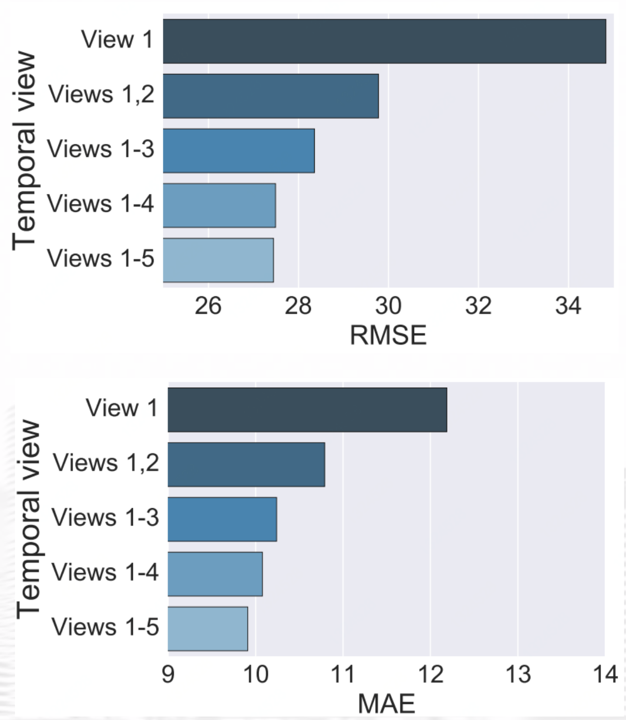


Fig. 10. Effect of temporal views using TaxiNYC.

Sudden changes

Temporal views

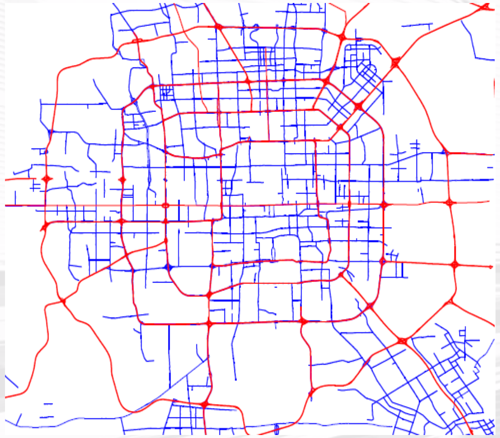
Junkai Sun, Junbo Zhang, et al. Predicting Citywide Crowd Flows in Irregular Regions Using Multi-View Graph Convolutional Networks. *IEEE TKDE 2020*

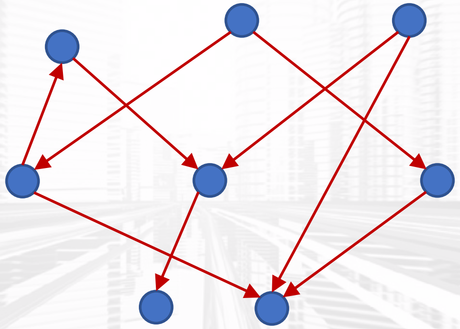
Spatio-Temporal Neural Networks

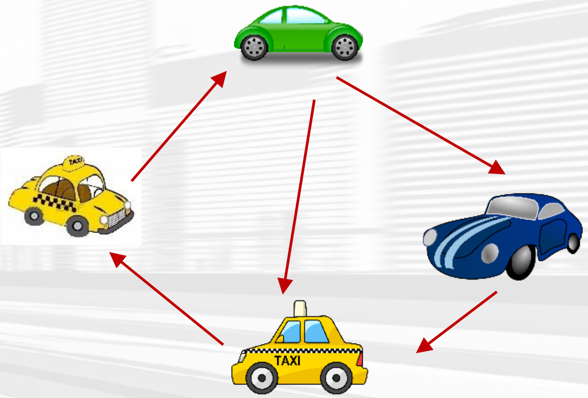
- ☐ST Point Data
- ☐ST Gridded Data
- ☐ST Network (Graph) Data: MVGCN, MDL
- ☐ST Sequence Data: DeepTTE

Spatio-Temporal Networks (Graphs)

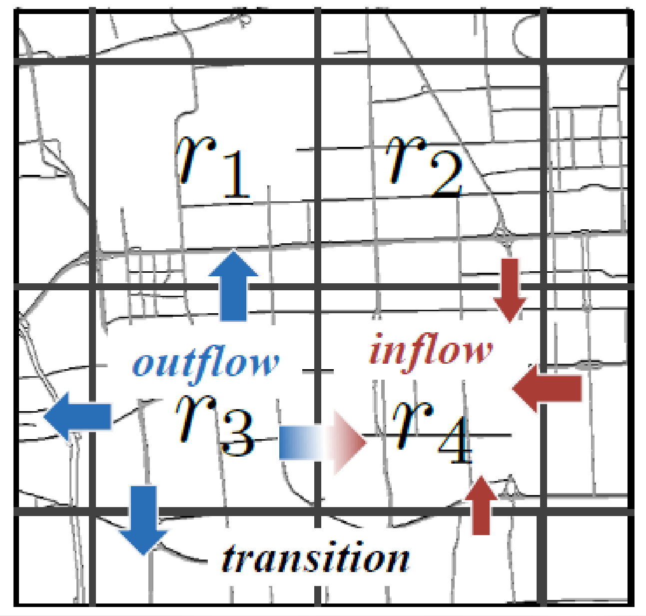
- Nodes
  - Spatial coordinates
  - Distance between nodes
  - Moving over time
- Edges
  - Temporal dynamic properties
  - Dynamic structures

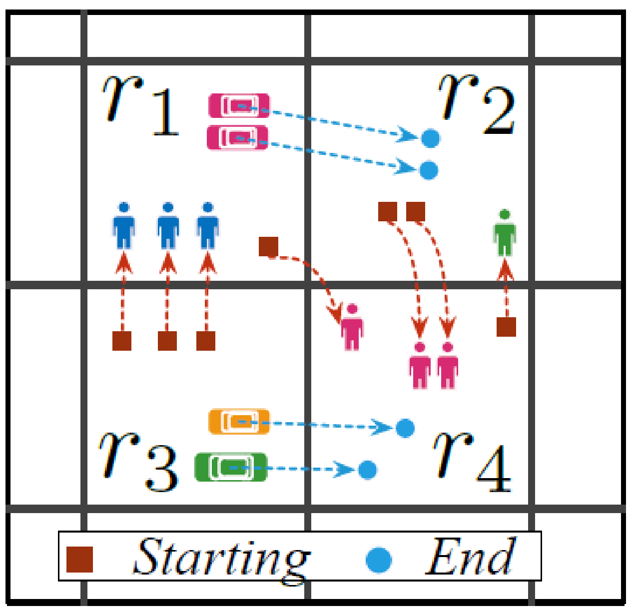


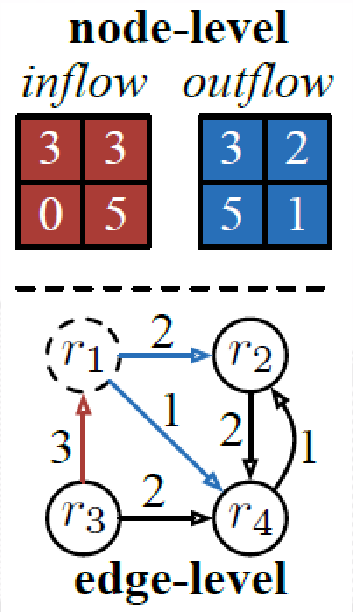




Predicting Transition and In/out Flows in a ST-Network








9:00 - Transitions Between Regions of Beijing
海淀区
西城区
东城区
朝阳区
丰台区

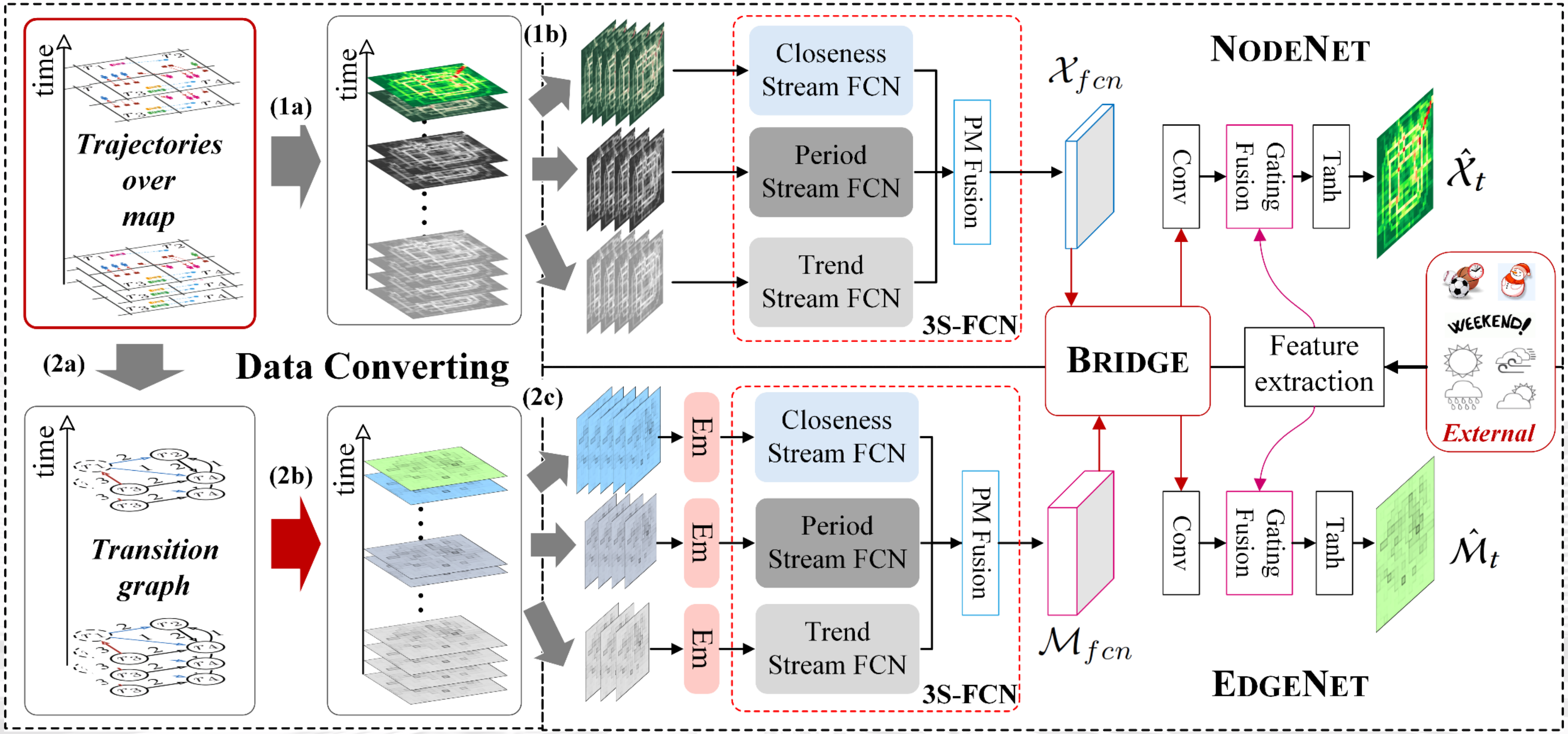
京东城市

京 东 数 科 旗 下

Challenges

- Scale and complexity
  - Dimension is very high
    - In/out-flow: $m * m$
    - Transition matrix: $m^2 * m^2$
- Model multiple correlations
  - Spatio-temporal dependencies
  - In/out flows and transition flow are highly correlated and mutually reinforced
  - External factors: Events, weather, accidents
- Dynamics and sparsity
  - Transition changes over time much more tremendously than in/out flows
  - Transition that will really occur at the next time interval may be a very small portion of the $m^2 * m^2$ possibilities (i.e. very sparse)

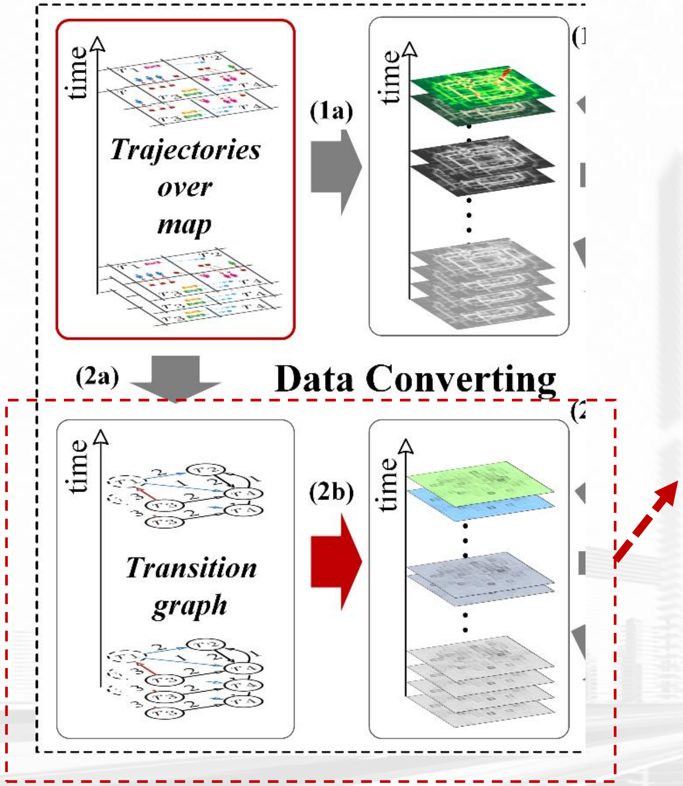
Multitask Deep Learning (MDL) Framework

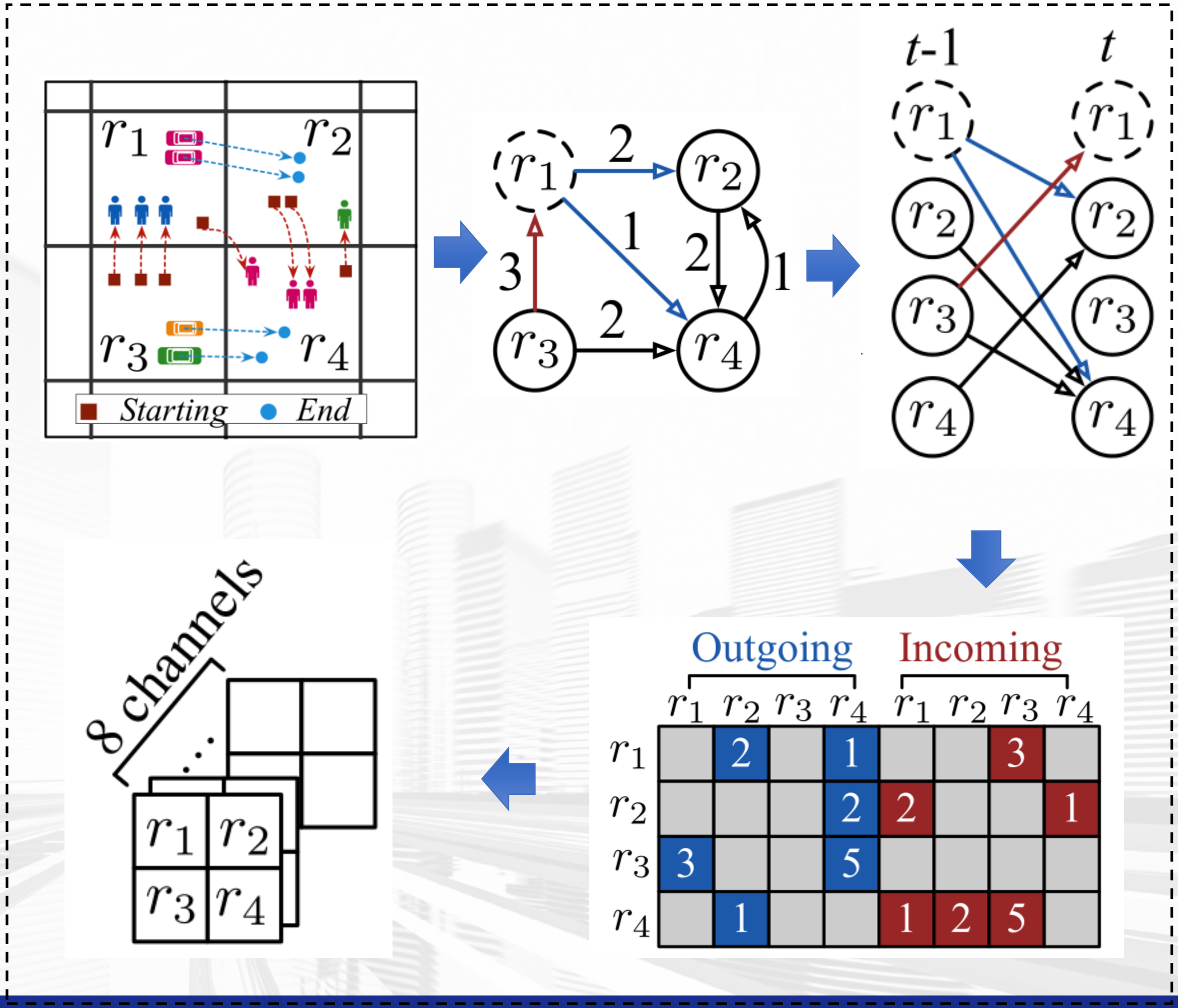


京东城市
京 东 数 科 旗 下

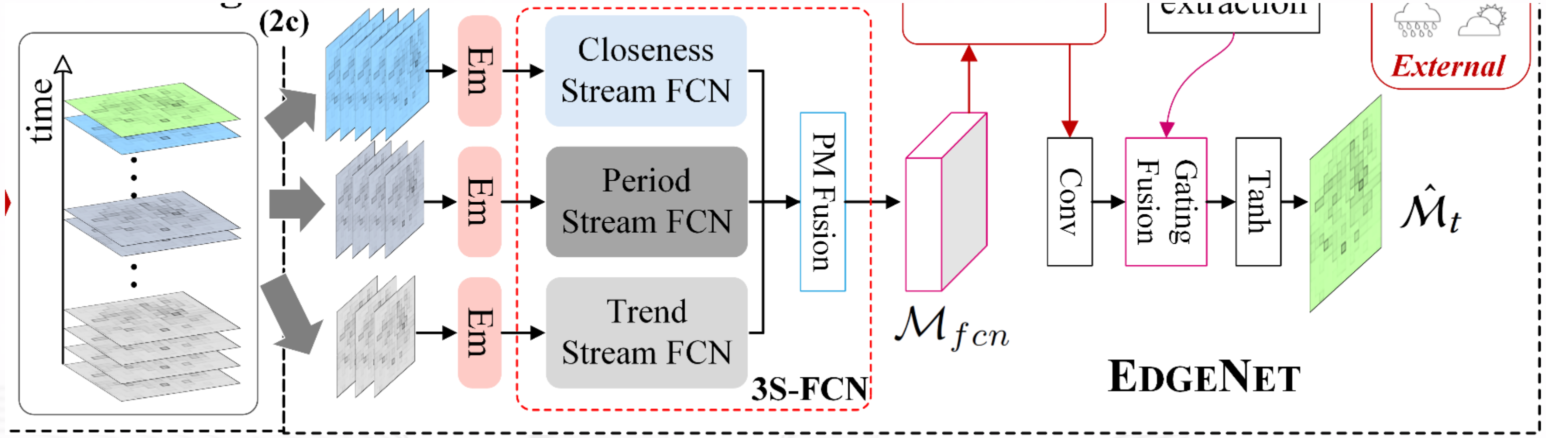
Junbo Zhang et al. Flow Prediction in Spatio-Temporal Networks Based on Multitask Deep Learning. IEEE TKDE, 2019

Data Converting: Graph → Tensor





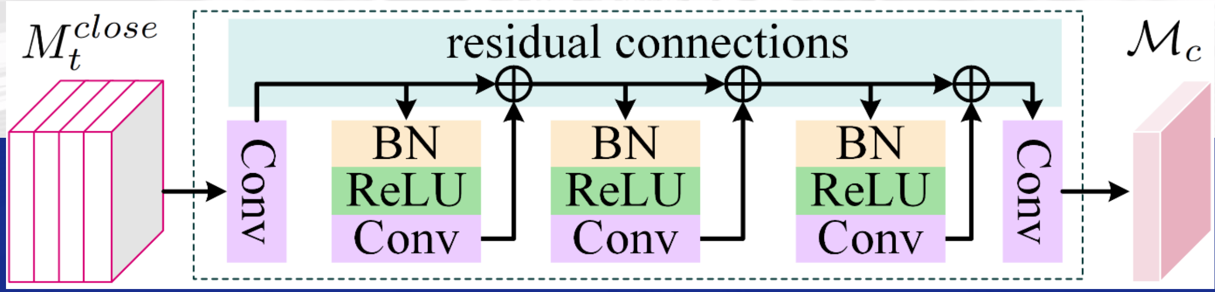
EdgeNet for Transition Prediction



- Tackle the sparse and high-dimensional transitions using a *spatial embedding* layer:

$$\mathcal{Z}_t(:,i,j) = \mathbf{W}_m \mathcal{M}_t(:,i,j) + \mathbf{b}_m, 1 \le i \le I, 1 \le j \le J$$

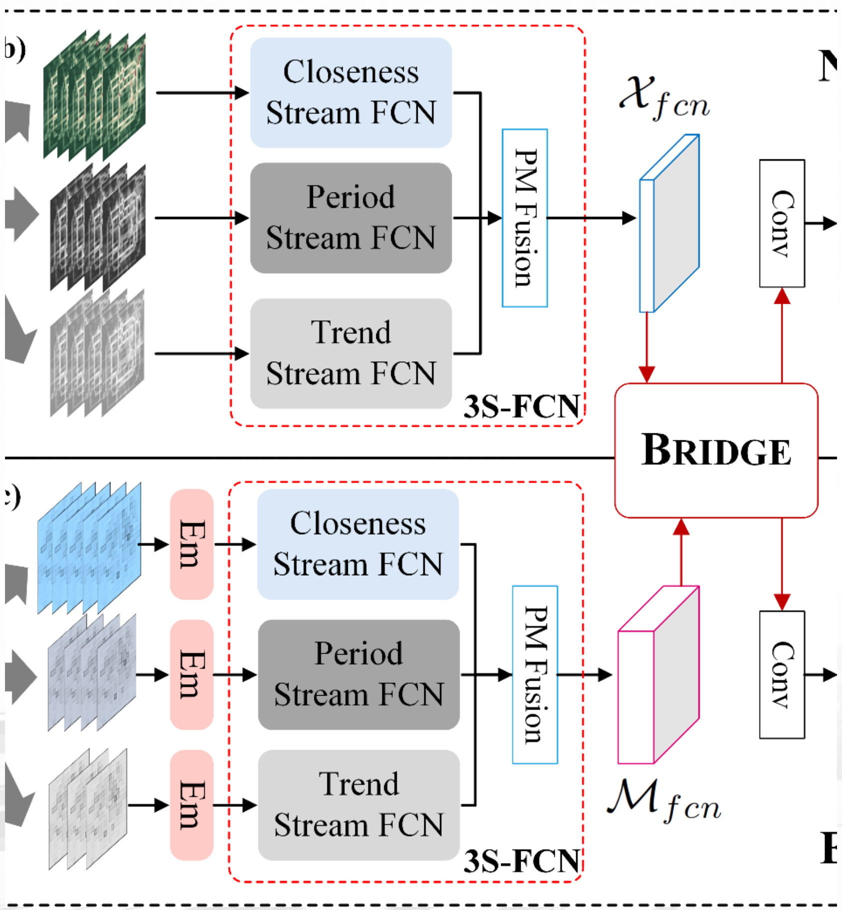
- Fully convolutional networks with residual connections:

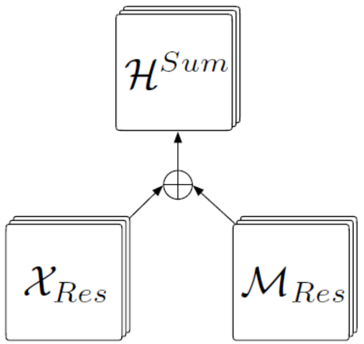


京东城市
京 东 数 科 旗 下

Couple EdgeNet and NodeNet

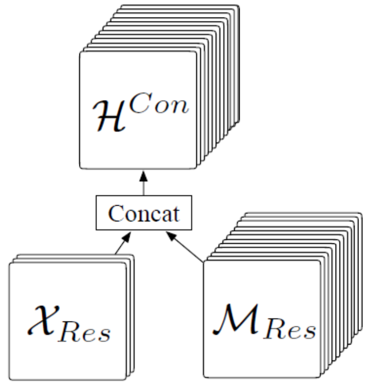
NodeNet is also a 3S-FCN, but no embedding





Sum Fusion

- Must have the same shape
- Easily harm the information contained in each of them



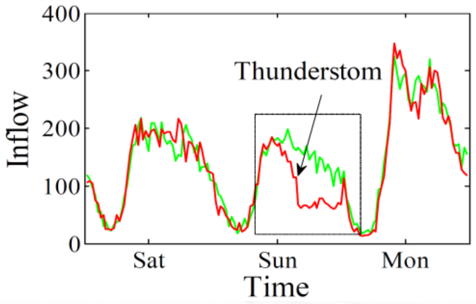
Concat Fusion

- Shapes can be different
- Integrate two levels of node and edge flows by mutually reinforcing

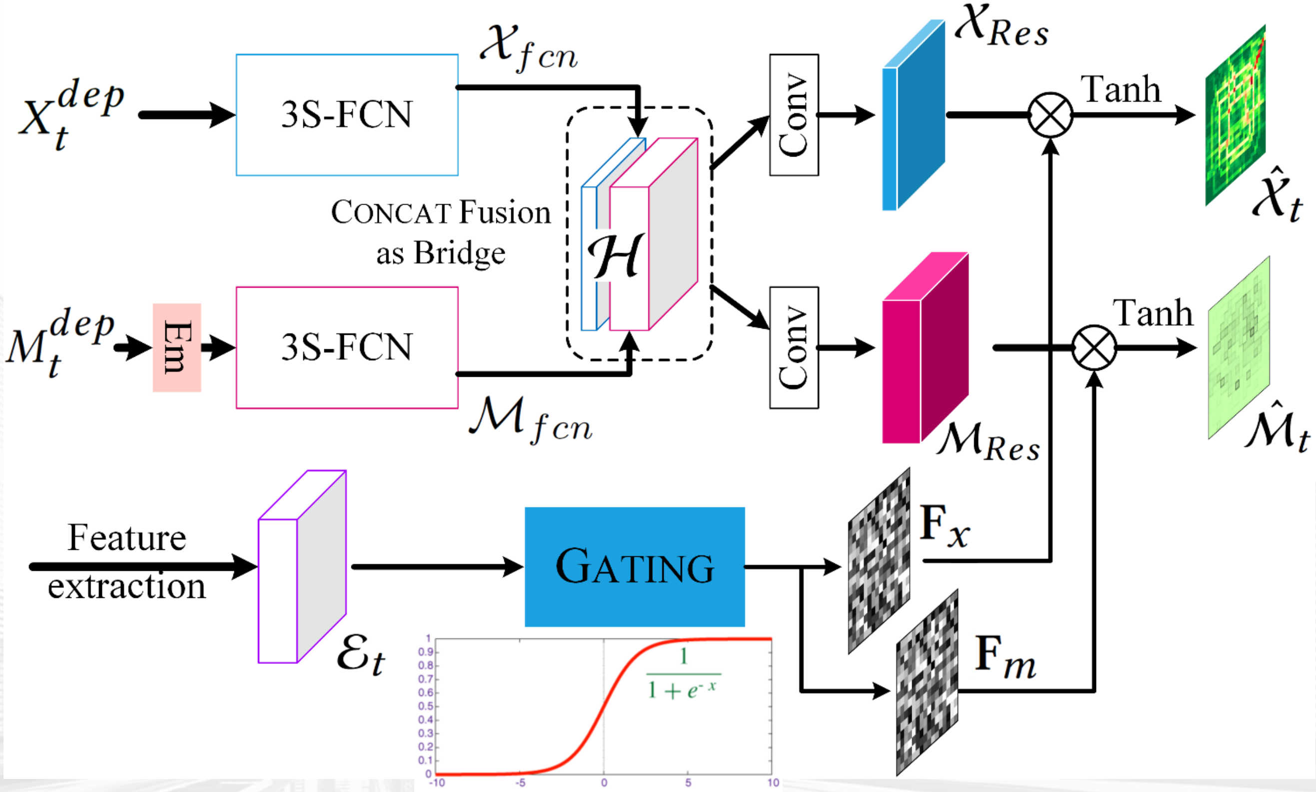
Other fusion methods?

- Multiply Fusion
- Kullback–Leibler (KD) divergence

Fusing External Factors

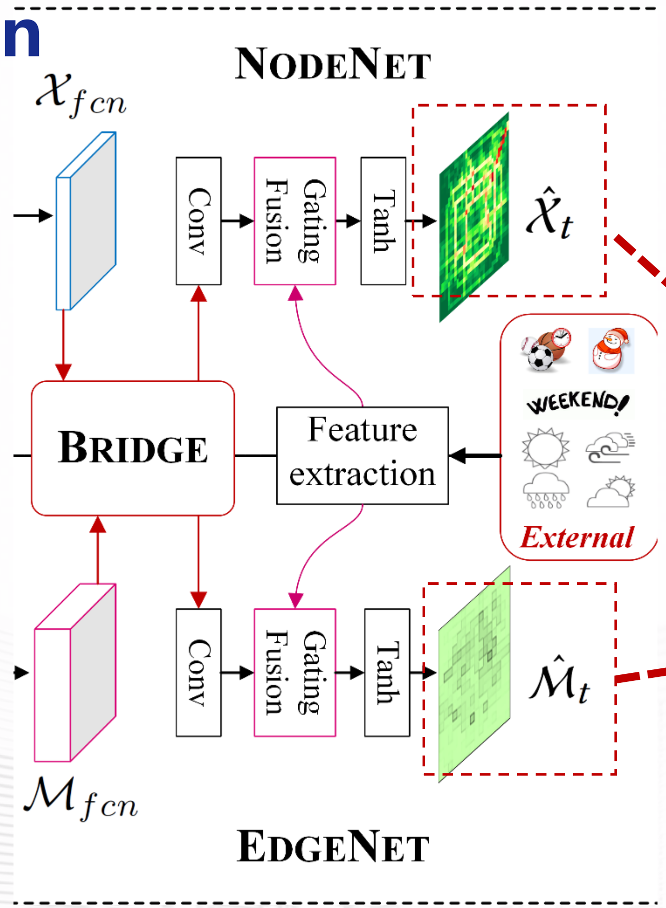


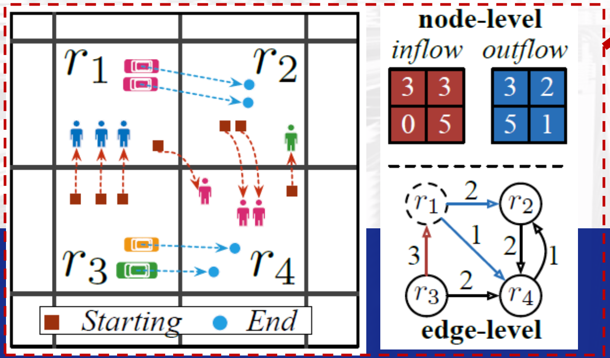
Just like a switch



$$\mathbf{F}_m(i,j) = \sigma\left(\mathbf{W}_e(:,i,j)\cdot\mathcal{E}_t(:,i,j) + \mathbf{b}_e(i,j)\right), 1 \le i \le I, 1 \le j \le J$$

Optimization objective





Node flow:

$$\mathcal{J}_{node} = \sum_{t\in\mathcal{T}}\sum_{c=0}^{1} \|P_t^c \odot \left(\hat{\mathcal{X}}_t(c,:,:) - \mathcal{X}_t(c,:,:)\right)\|_F^2$$

Edge flow:

$$\mathcal{J}_{edge} = \sum_{t\in\mathcal{T}}\sum_{c=0}^{2N-1} \|Q_t^c \odot \left(\hat{\mathcal{M}}_t(c,:,:) - \mathcal{M}_t(c,:,:)\right)\|_F^2$$

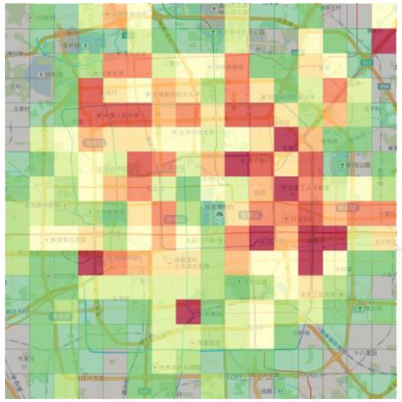
Mutual constraints:

$$\mathcal{J}_{mdl} = \sum_{t\in\mathcal{T}}\left(\|\underbrace{\hat{\mathcal{X}}_t(0,:,:)}_{\text{outflow}} - \underbrace{\sum_{c=0}^{N-1}\hat{\mathcal{M}}_t(c,:,:)}_{\text{outgoing transitions}}\|_F^2 + \|\underbrace{\hat{\mathcal{X}}_t(1,:,:)}_{\text{inflow}} - \underbrace{\sum_{c=N}^{2N-1}\hat{\mathcal{M}}_t(c,:,:)}_{\text{incoming transitions}}\|_F^2 \right)$$

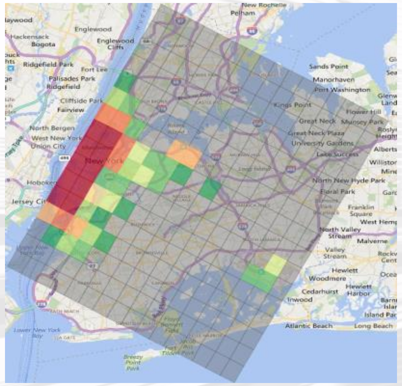
Final:

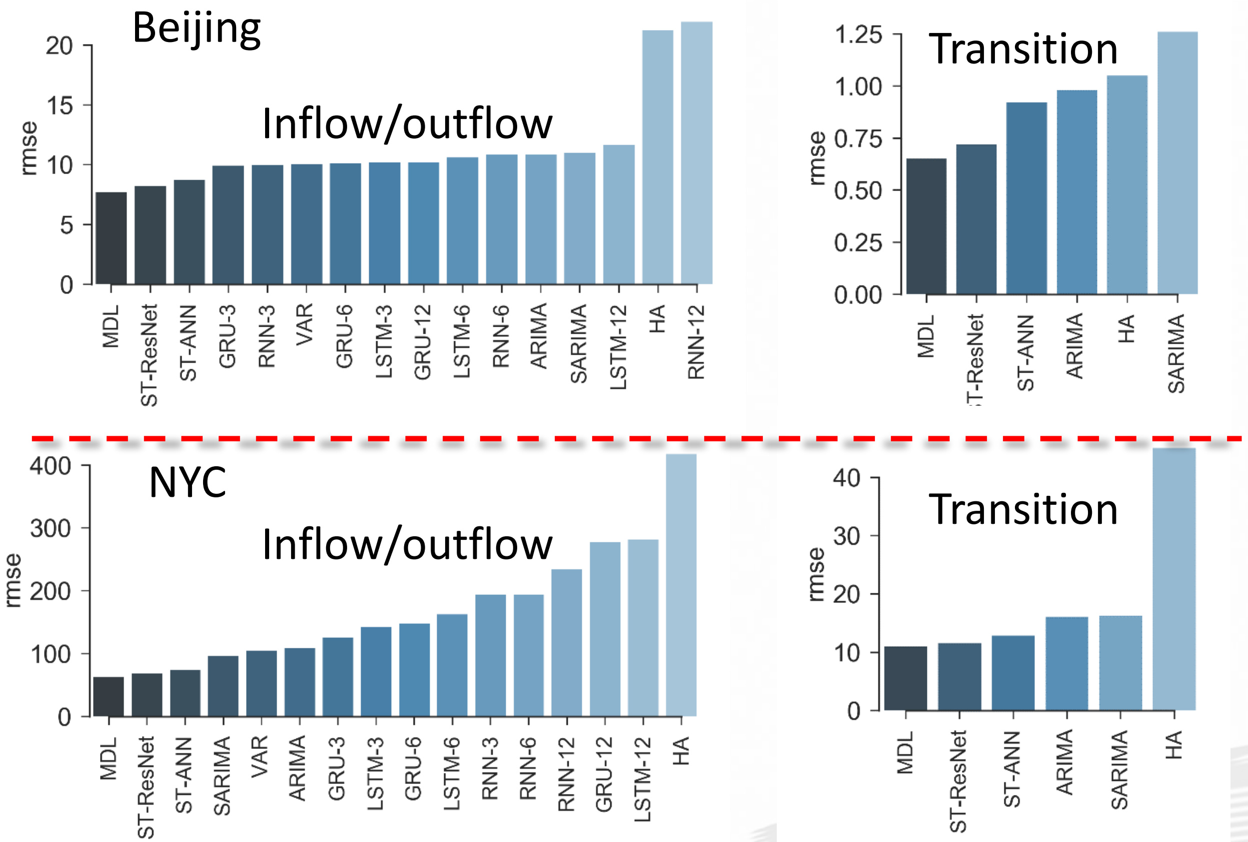
$$\arg\min_{\theta,\phi} \lambda_{node}\mathcal{J}_{node} + \lambda_{edge}\mathcal{J}_{edge} + \lambda_{mdl}\mathcal{J}_{mdl}$$

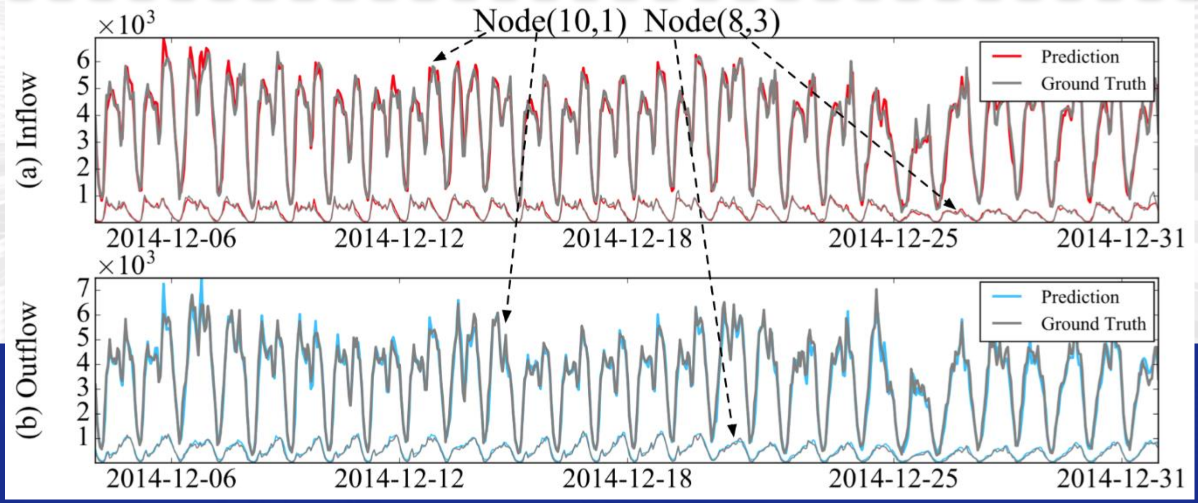
Beijing (2013~2016)



NYC (2011~2014)

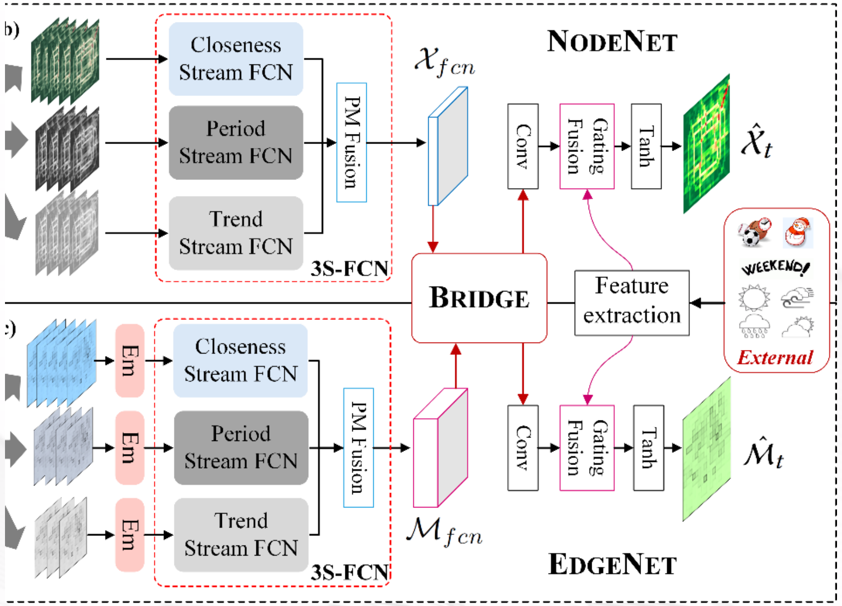






Inflow/outflow prediction of two regions, New York City

Evaluation on Fusing Mechanisms



Bridge
- Concat fusion
- Sum fusion

Fusing external features
- Gating
- Simple (*i.e.* sum fusion used in ST-ResNet)
- without

Fusing type: Bridge	Fusing type: External	RMSE/ MAE: inflow	RMSE/ MAE: outflow	RMSE/ MAE: transition
Concat	Gating	**65.30/ 17.27**	**55.29/ 17.66**	**11.68/ 3.70**
Concat	Simple	68.51/ 17.90	58.51/ 18.61	11.91/ 3.78
Concat	w/o	75.79/ 18.78	61.60/ 19.19	11.87/ 3.74
Sum	Gating	67.82/ 17.87	65.18/ 19.80	12.67/ 3.88
Sum	Simple	71.65/ 18.44	67.20/ 20.34	12.77/ 3.96
Sum	w/o	80.50/ 20.07	61.33/ 19.54	12.66/ 3.99

Spatio-Temporal Neural Networks

- ST Point Data: GeoMAN
- ST Gridded Data: ST-ResNet
- ST Network (Graph) Data: MVGCN, MDL
- ST Sequence Data: DeepTTE

Travel Time Estimation (TTE)

TTE is a long-standing and critically important topic in the area of Intelligent Transportation Systems

- Dispatch taxis to passengers in shortest time
- Better planning the routes, avoiding congested roads
- Help to alleviate urban traffic congestion

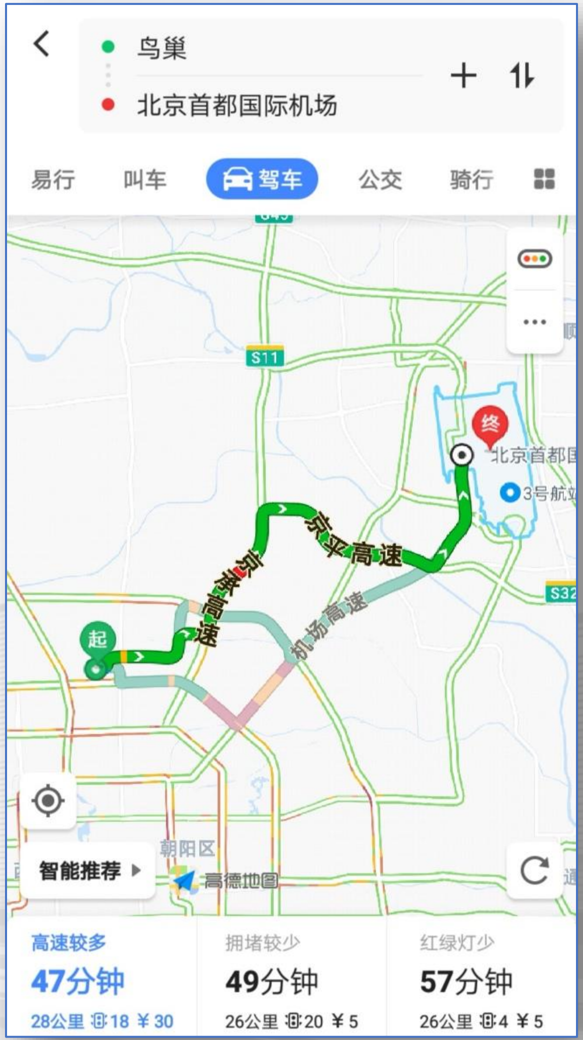
Bing maps

高德地图 amap.com

滴滴

Google Maps

UBER

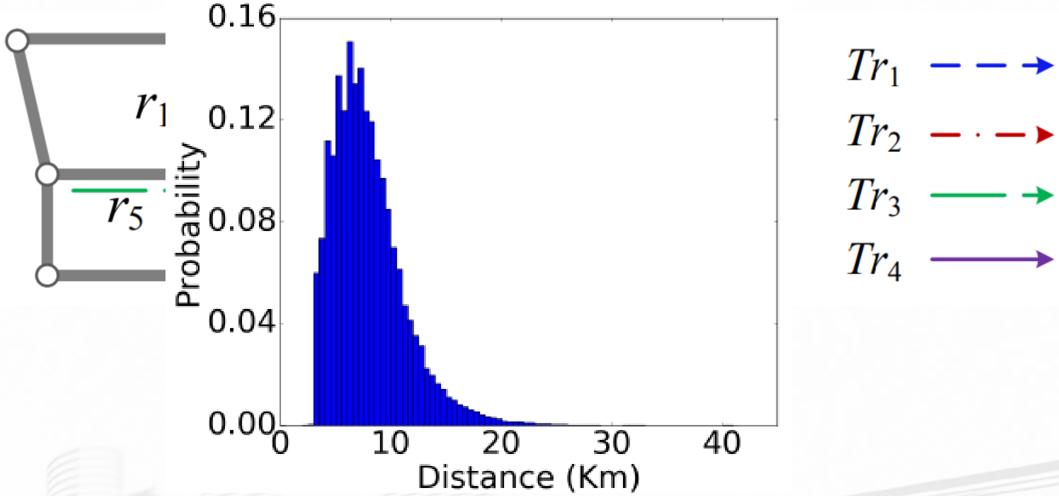


京东城市
京 东 数 科 旗 下

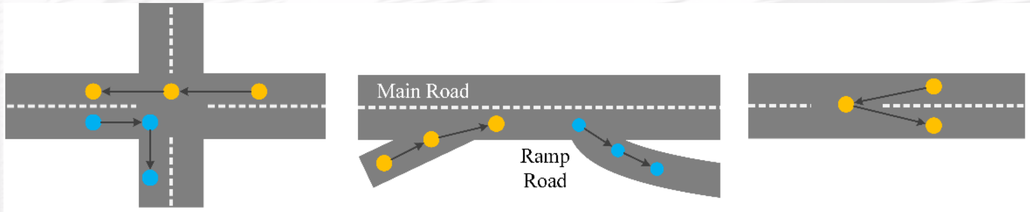
Dong Wang, **Junbo Zhang**, et al. When Will You Arrive Estimating Travel Time Based on Deep Neural Networks. AAAI 2018

Challenges

- Individual vs collective
  - Estimate local paths: Cannot handle intersections, traffic lights and direction turns
  - Estimate entire path: uneven data distribution



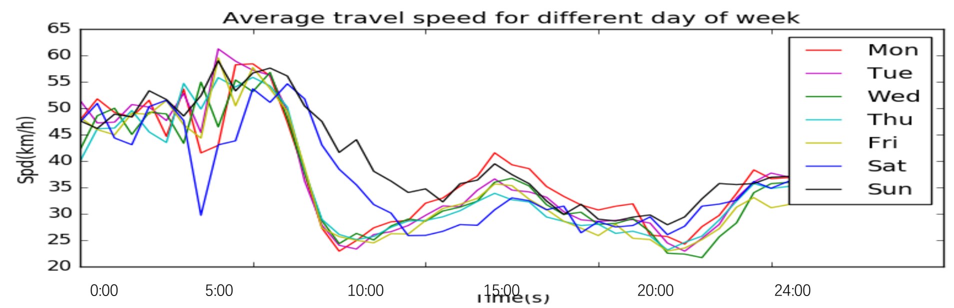
- Diverse influences
  - Spatial correlations: various & complex
  - Temporal dependencies
  - External factors (day of the week, starting time, driver, distance)



(a) Go straight (yellow) or turn right (blue)

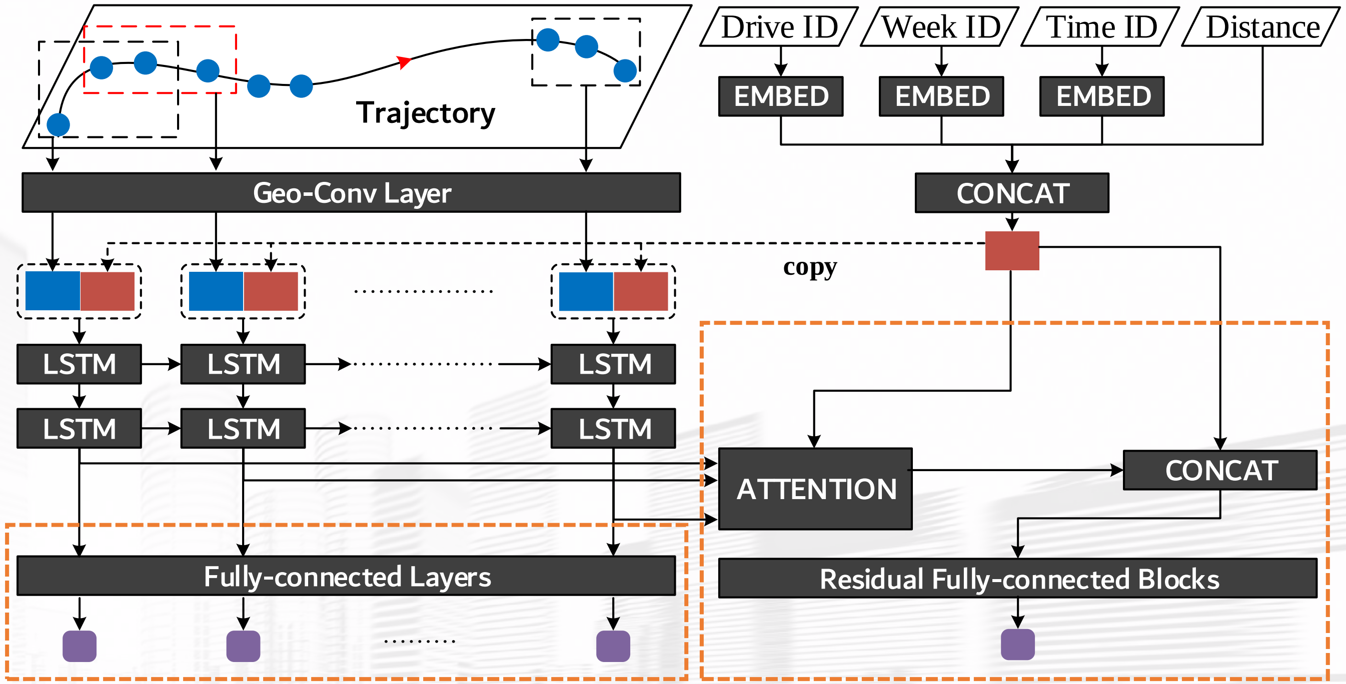
(b) Drive into main road (yellow) or ramp road (blue)

(c) Turn around



DeepTTE: Deep Learning + Multitask Learning

- Capture spatial dependencies
- Handle external factors & share similar pattern
- Learn temporal dependencies
- Address imbalance data problem
- Attention module to learn weights for different local path
- Help train deeper network for better result



TTE for each local path

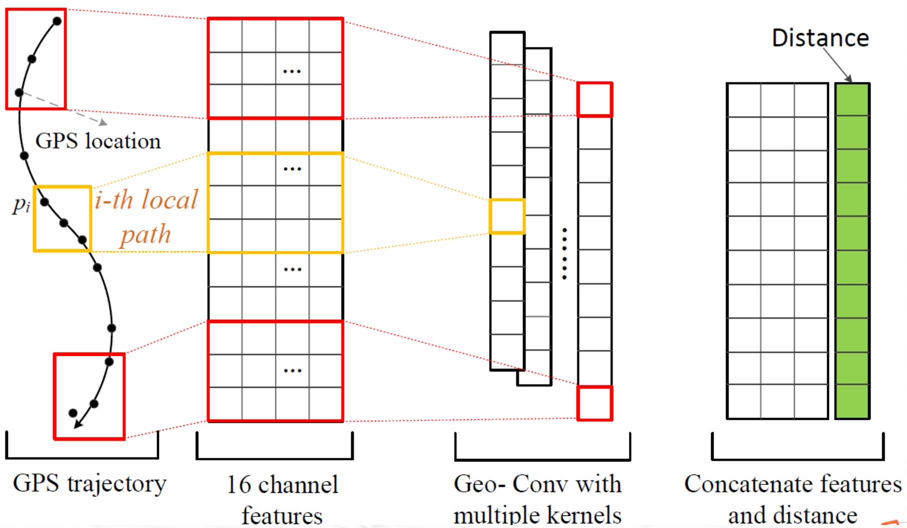
TTE for the whole path

Loss: $\beta \cdot L_{local} + (1 - \beta) \cdot L_{en}$

京东城市
京 东 数 科 旗 下

Dong Wang, **Junbo Zhang**, et al. When Will You Arrive Estimating Travel Time Based on Deep Neural Networks. AAAI 2018

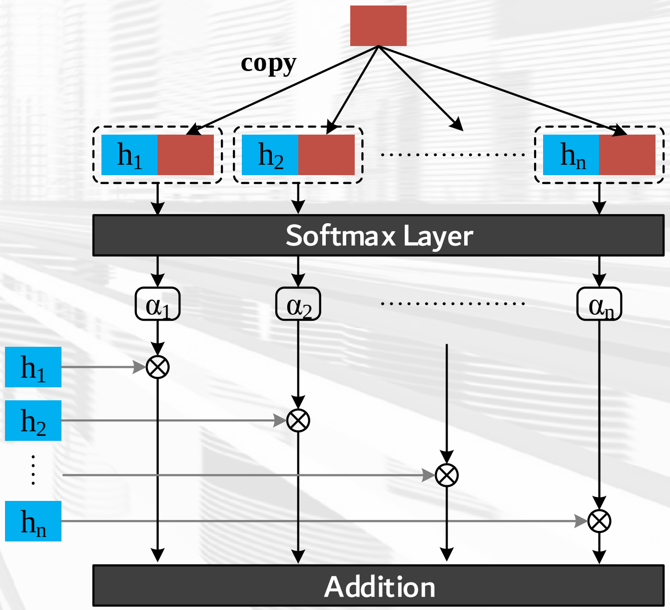
Geo-Conv Layer



- Transforms the raw GPS sequence to a series of feature maps.
- Capture spatial correlations of the local paths
- Remain the information in a fine granularity.

Attention Component

Local path with more intersections or in extremely congested need more attention.



Experiments

Chengdu Dataset: Chengdu Dataset consists of 9, 737, 557 trajectories (1.4 billion GPS records) of 14864 taxis in August 2014 in Chengdu, China. The shortest trajectory contains only 11 GPS records (2km) and the longest trajectory contains 128 GPS records (41km).

Beijing Dataset: Beijing Dataset consists of 3, 149, 023 trajectories (0.45 billion GPS records) of 20442 taxis in April 2015 in Beijing, China. The shortest trajectory contains 15 GPS records (3.5km) and the longest trajectory contains 128 GPS records (50km).

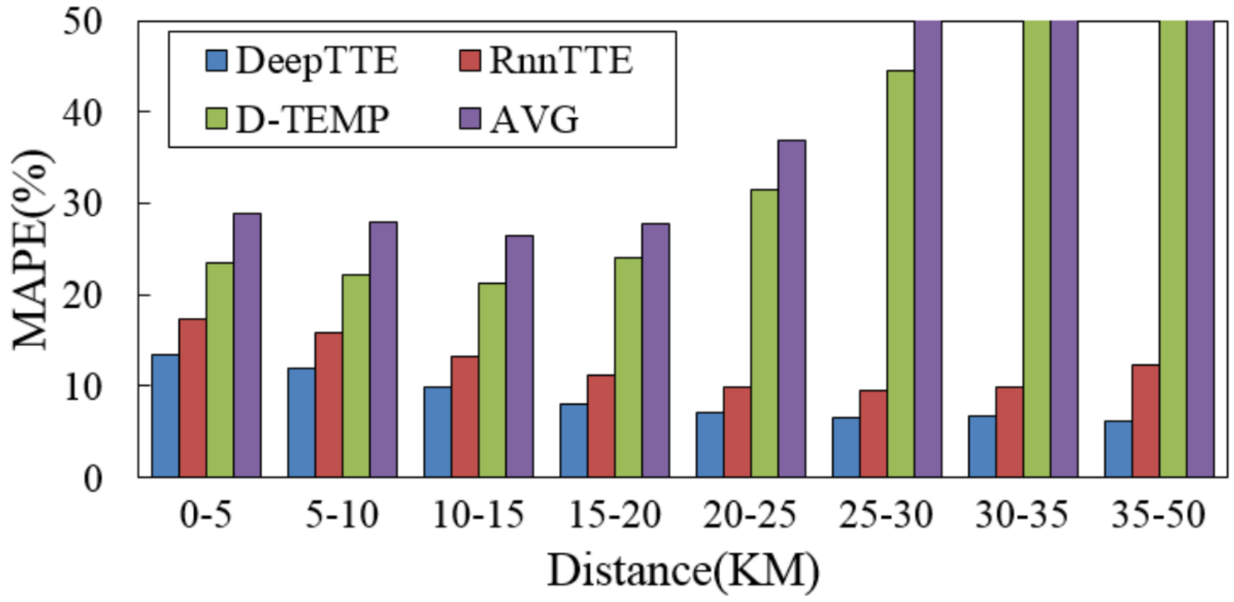
Table 1: Performance Comparison

	Chengdu			**Beijing**		
	MAPE (%)	RMSE (sec)	MAE (sec)	MAPE (%)	RMSE (sec)	MAE (sec)
AVG	28.1	533.57	403.71	24.78	703.17	501.23
D-TEMP	22.82	441.50	323.37	19.63	606.76	402.50
GBDT	19.32 ± 0.04	357.09 ± 2.44	266.15 ± 2.24	19.98 ± 0.02	512.96 ± 3.96	393.98 ± 2.99
MlpTTE	16.90 ± 0.06	379.39 ± 1.94	265.47 ± 1.53	23.73 ± 0.14	701.61 ± 1.82	489.54 ± 1.61
RnnTTE	15.65 ± 0.06	358.74 ± 2.02	246.52 ± 1.65	13.73 ± 0.05	408.33 ± 1.83	275.07 ± 1.48
DeepTTE	$\mathbf{11.89 \pm 0.04}$	$\mathbf{282.55 \pm 1.32}$	$\mathbf{186.93 \pm 1.01}$	$\mathbf{10.92 \pm 0.06}$	$\mathbf{329.65 \pm 2.17}$	$\mathbf{218.29 \pm 1.63}$

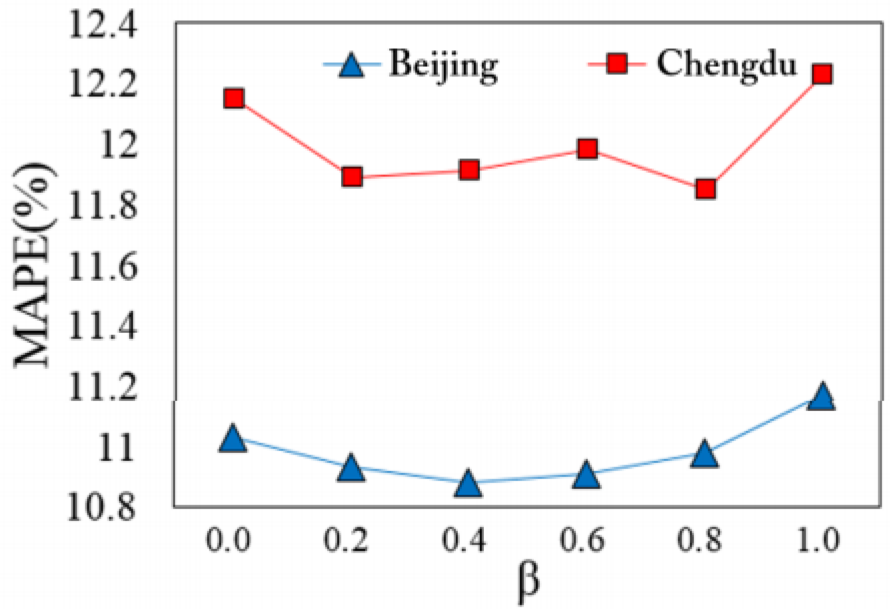
京东城市
京 东 数 科 旗 下

Dong Wang, **Junbo Zhang**, et al. When Will You Arrive Estimating Travel Time Based on Deep Neural Networks. AAAI 2018

Experiments



Error rates for trajectories with different lengths.



Error rates for different β.

Loss: $\beta \cdot L_{local} + (1-\beta) \cdot L_{en}$

2.1 ST Neural Networks 小结

1. 原始时空数据
2. 空间特性
3. 时间特性
4. 外部因素

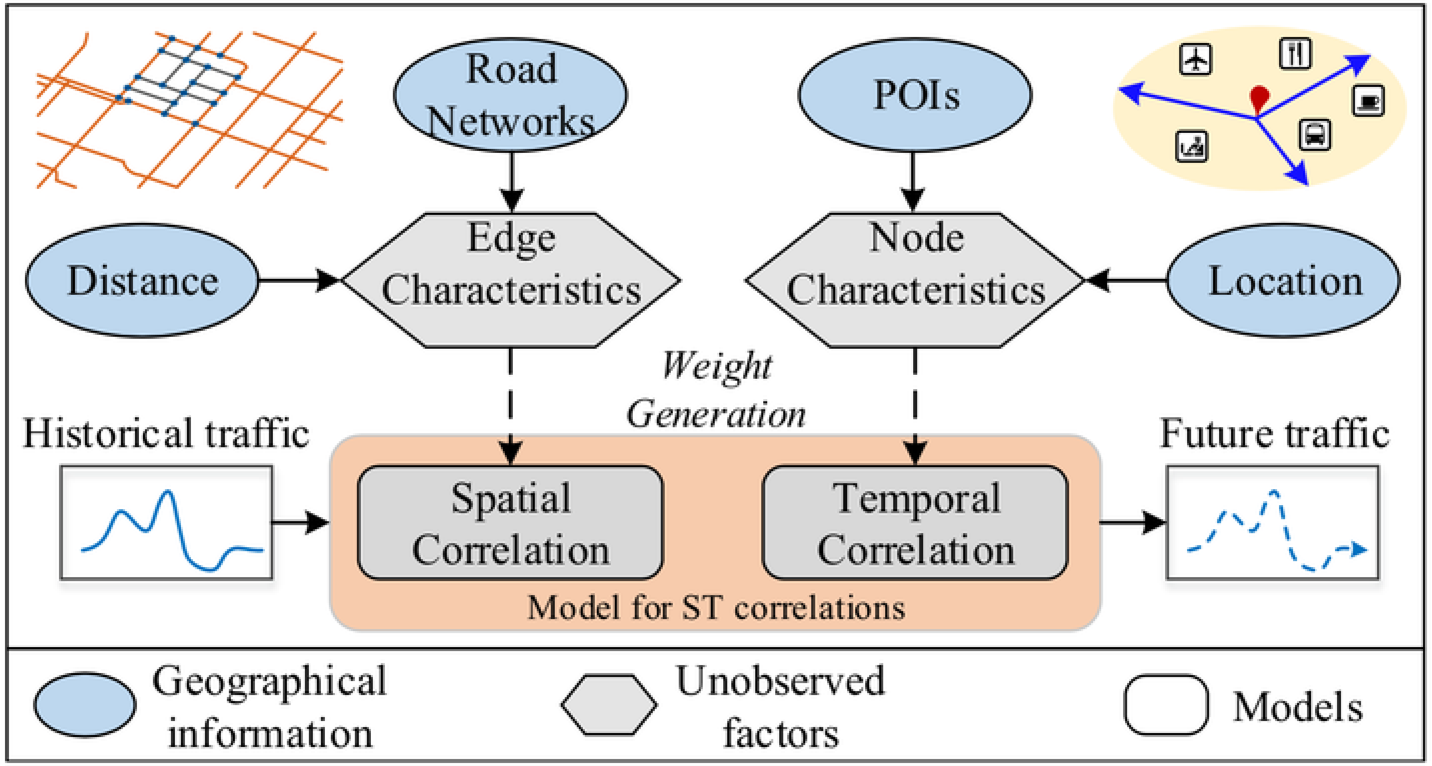
- 数据转换
- CNN
- RNN
- GCN
- Attention
- …

Part 2.2 Advanced ST Neural Networks

- ST-MetaNet: Learning parameters → Generating parameters
- AutoST: Manually-designed → Automated Network Architecture Search

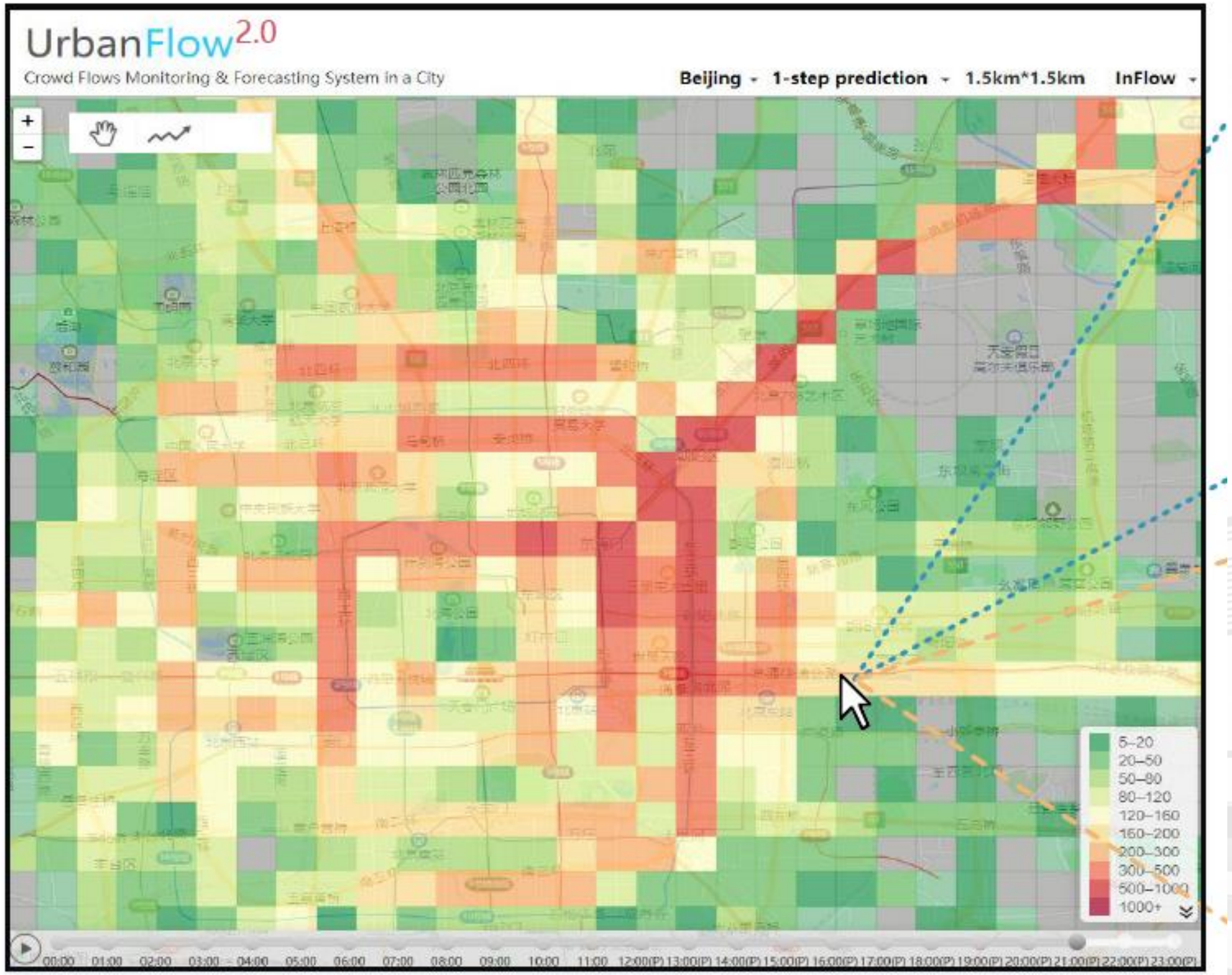
Urban Traffic Prediction from Spatio-Temporal Data using Deep Meta Learning

KDD 2019



京东城市
京 东 数 科 旗 下

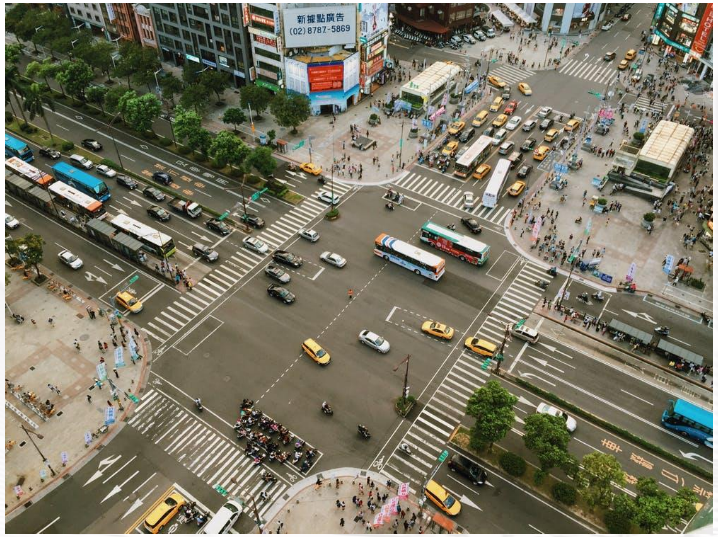
Intro. to Urban Traffic



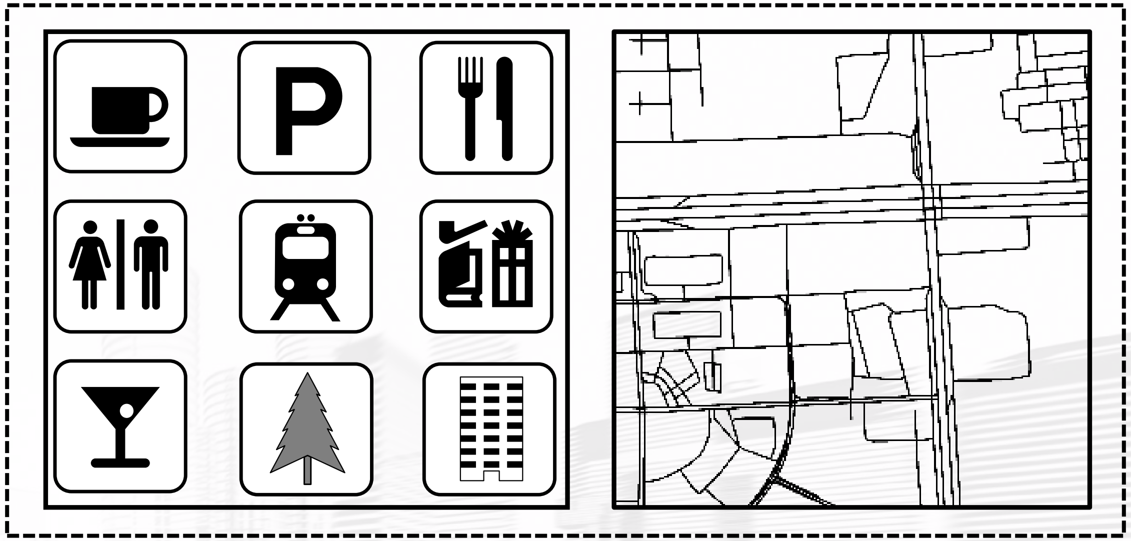
- Traffic data includes:
  - Speed data
  - Flow data
  - Regional demand data
- Important for:
  - Traffic management
  - Risk assessment
  - Service provide

Intro. to Urban Traffic Prediction

Traffic data



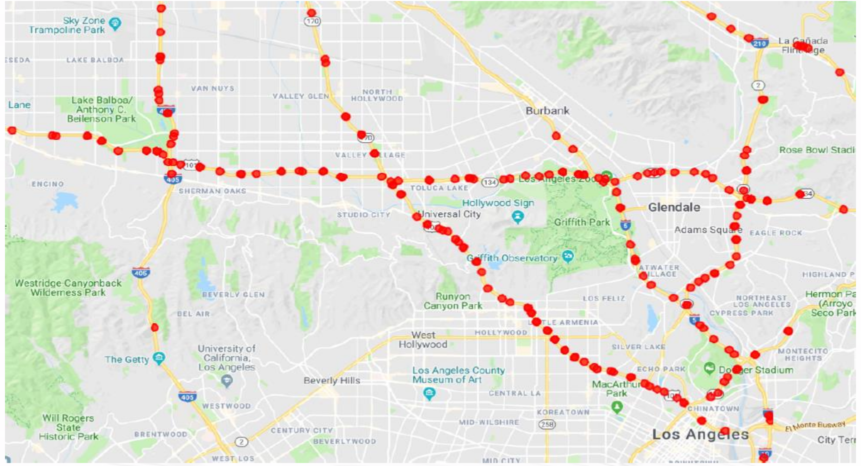
Geo-attributes



Predict urban traffic on each location at next time interval throughout a city by using ***historical traffic data*** and ***geo-attributes*** (e.g., points of interests and road networks)

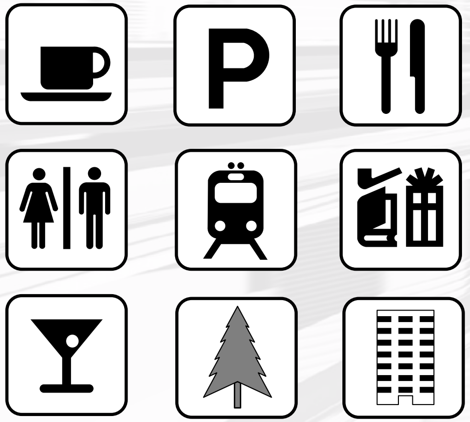
Intro. to Urban Traffic Prediction

Traffic data collected from loop detectors



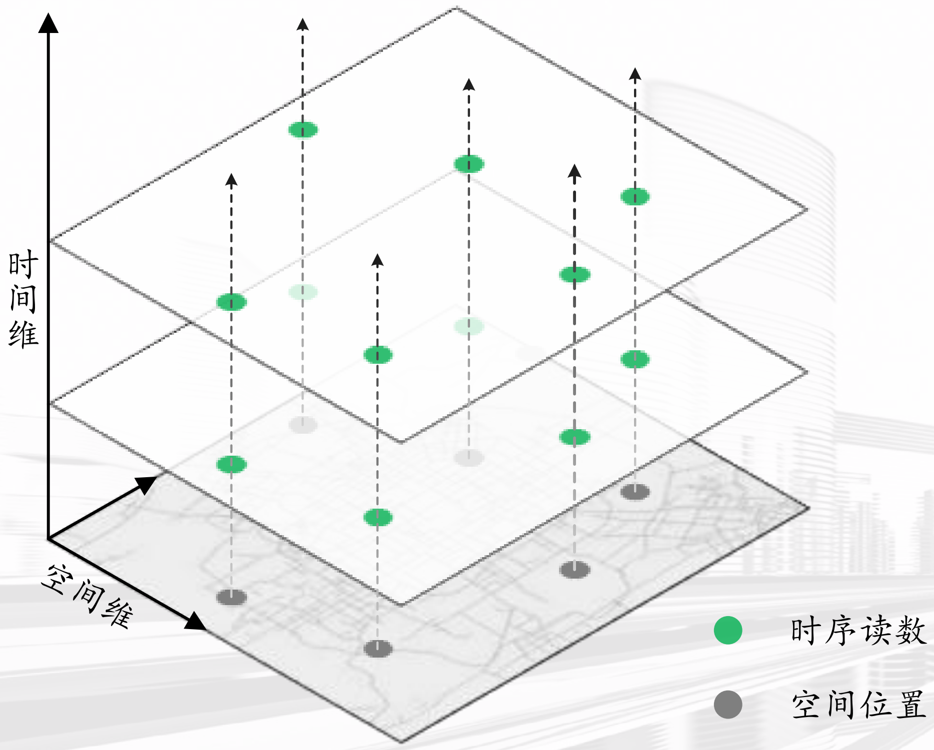
	sensor_0	sensor_1	sensor_2	sensor_n
2018/01/01 00:00:00	60.0	65.0	70.0	...
2018/01/01 00:05:00	61.0	64.0	65.0	...
2018/01/01 00:10:00	63.0	65.0	60.0	...
...	...	...	...	...

Points of interests



Road networks

Challenges – ST correlations



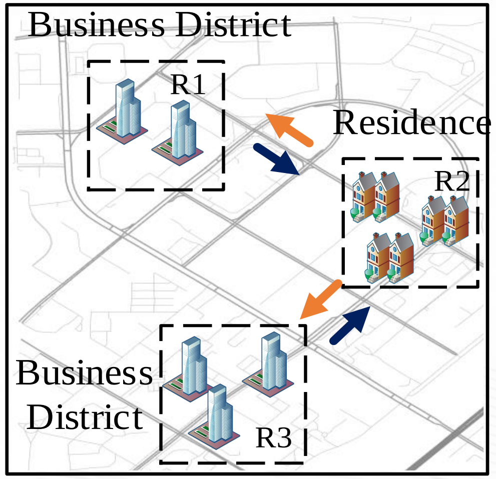
❑ Spatial correlations

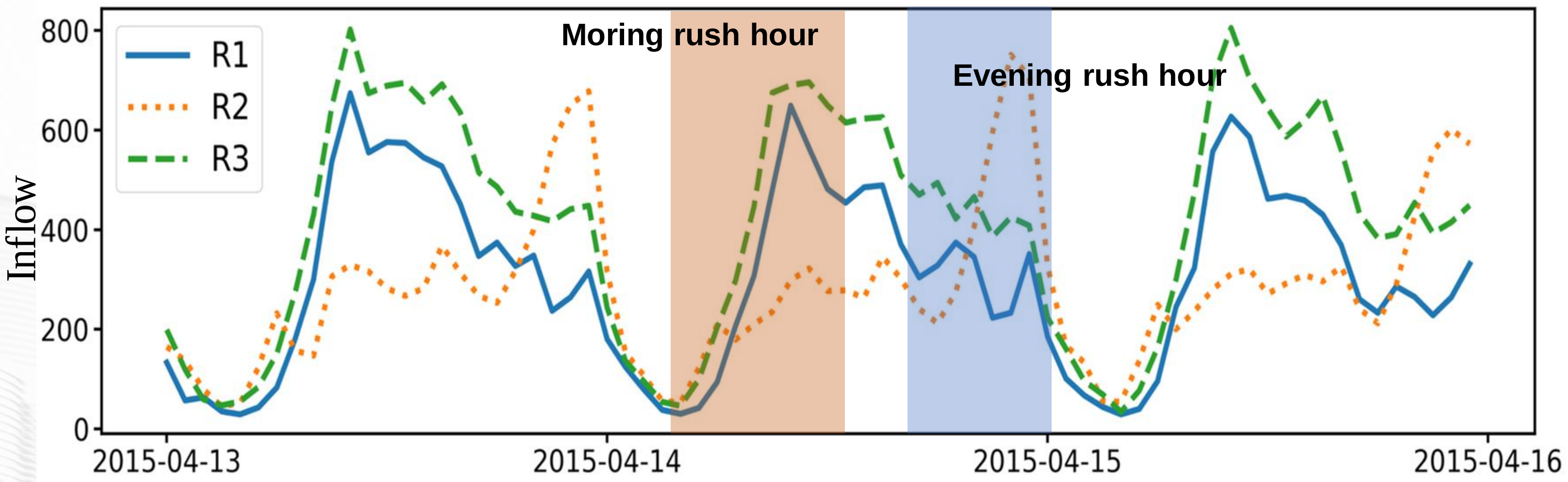
Traffic is moving on road networks, so the state of a location can broadcast to other locations.

❑ Temporal correlations

The state of a location can impact its latter states

Challenges – Diversity of ST Correlations





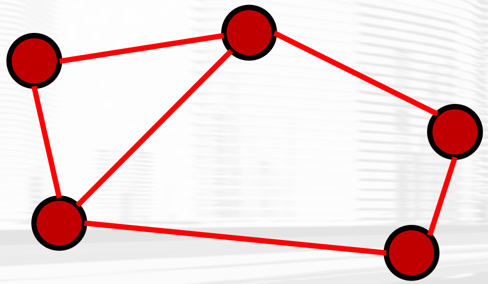
- ❑ Characteristics of locations and their mutual relationship are diverse, depending on their own ***geo-attributes***.
- ❑ Locations with similar combinations of ***geo-attributes*** lead to similar characteristics of locations and analogous types of ST correlations.

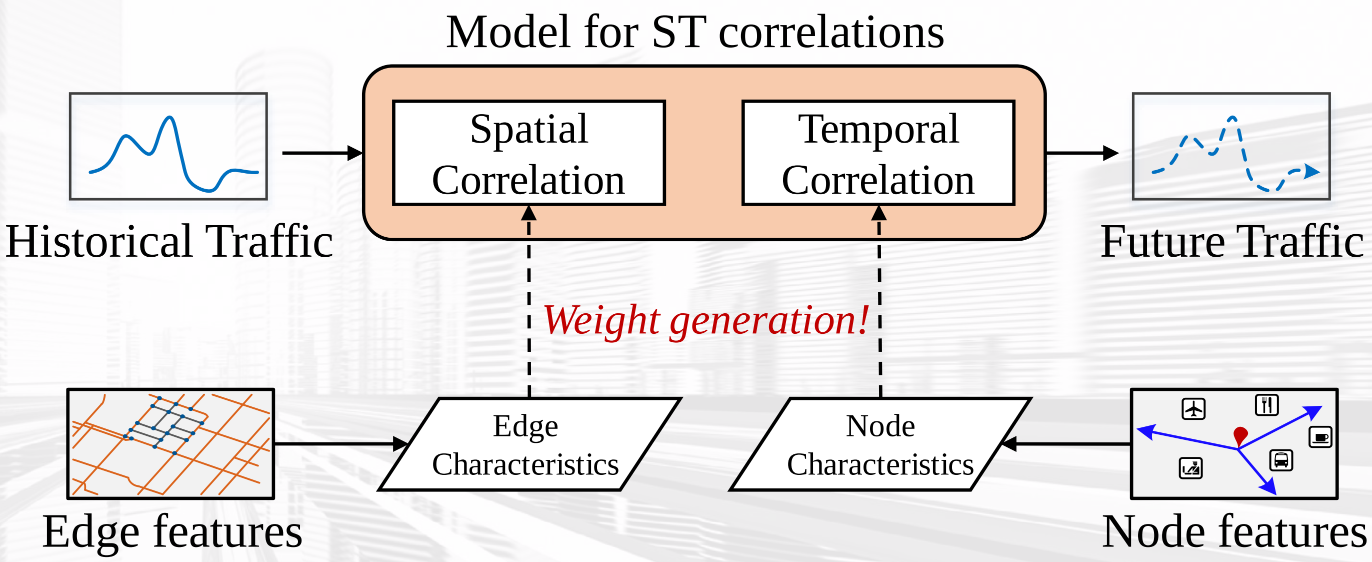
Insights

We build a geo-graph to describe spatial structures of road networks

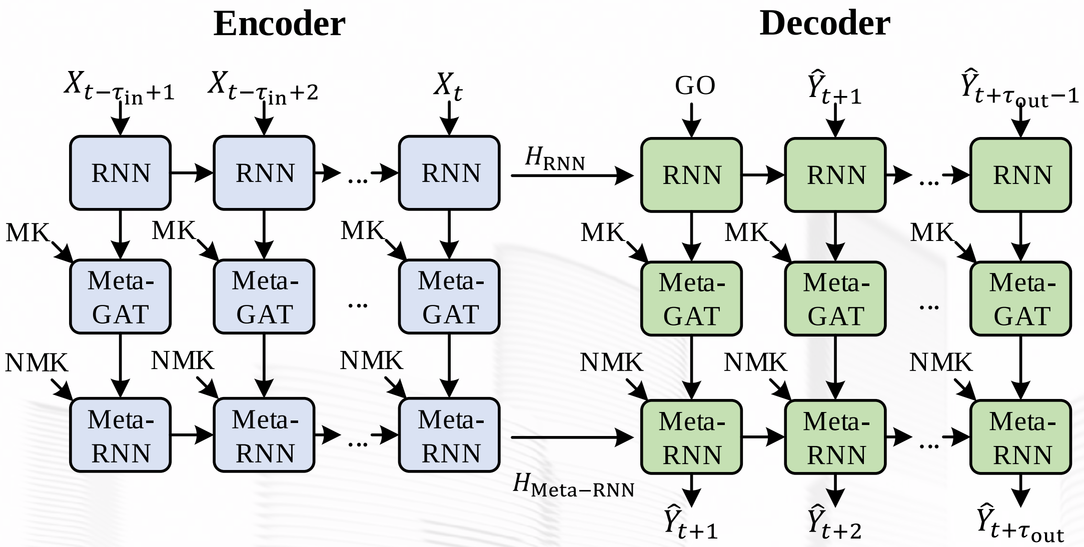
- Nodes – locations
- Edges – relation between locations

Geographical attributes ***reveals*** characteristics of nodes and edges & ***impacts*** different types of ST correlations.





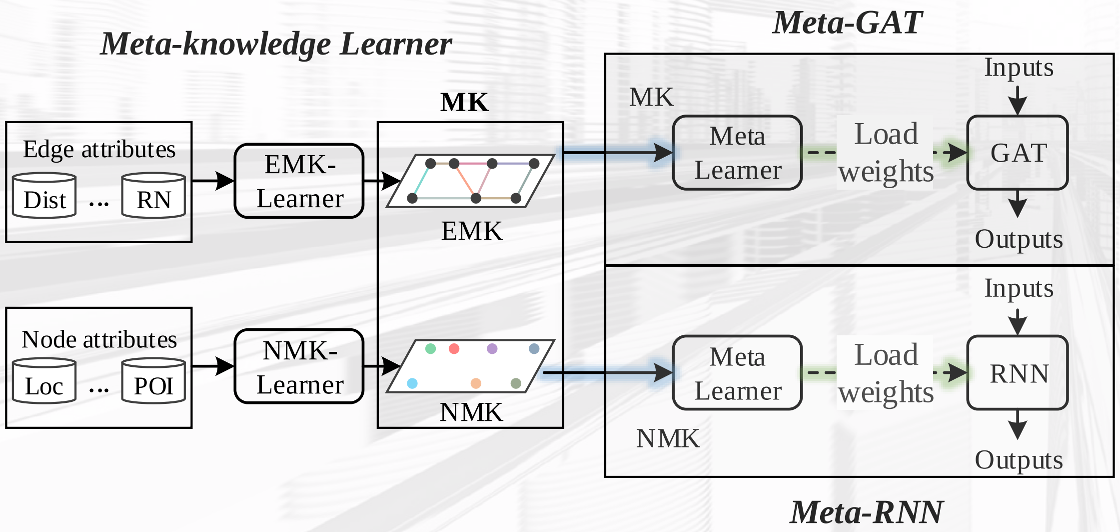
Framework of ST-MetaNet



Recurrent Neural Network (RNN)
❑ Embedding the sequence of urban traffic.

Meta Graph Attention Network (Meta-GAT)
❑ Modeling diverse spatial correlations.

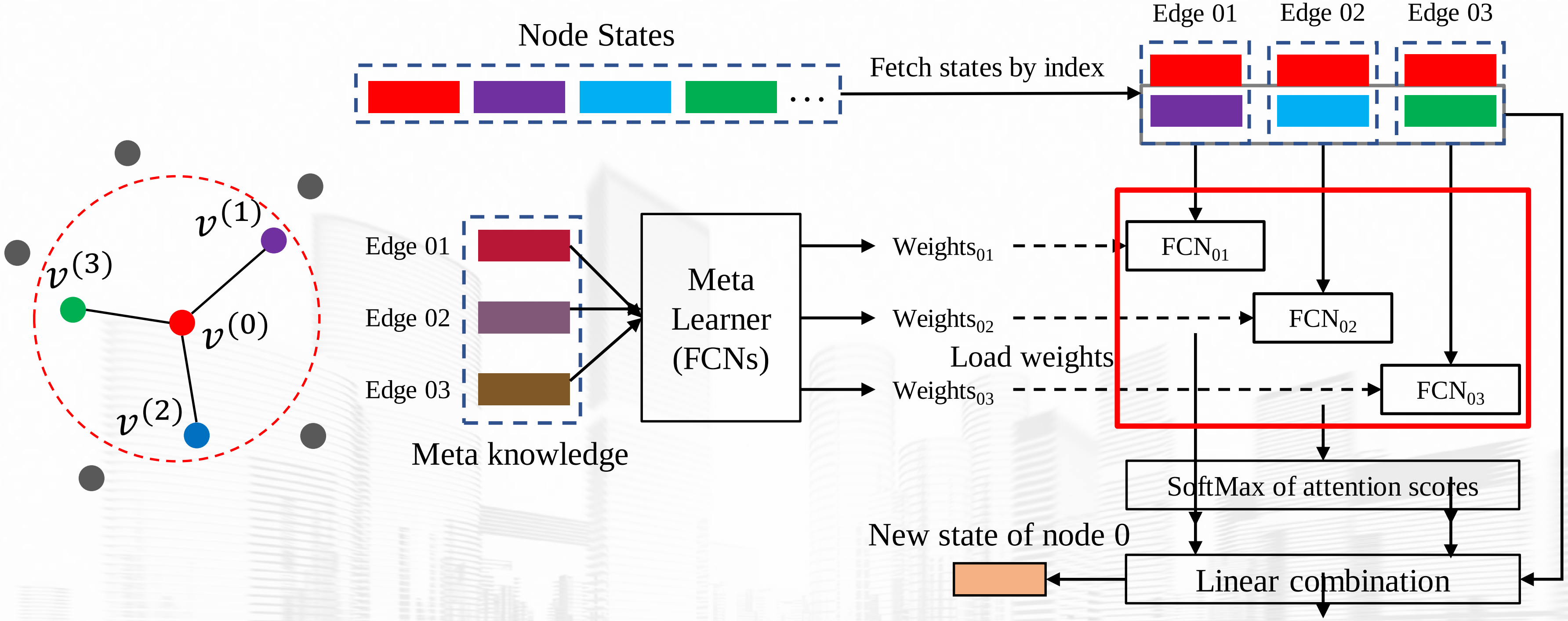
Meta Recurrent Neural Network (Meta-RNN)
❑ Modeling diverse temporal correlations.



Meta-knowledge Learner
❑ Learning node & edge characteristics from geo-attributes.

Meta Learner
❑ Generating parameter weights in GAT and RNN.

Meta Graph Attention Network



Compute attention scores

$$w^{(ij)} = \text{LeakyReLU}\left(\boxed{W^{(ij)}}\left[h^{(i)} \parallel h^{(j)}\right] + \boxed{b^{(ij)}}\right)$$

Get edge meta knowledge

$$MK^{(ij)} = NMK^{(i)} \parallel NMK^{(j)} \parallel EMK^{(ij)}$$

- Outputs of meta knowledge learners

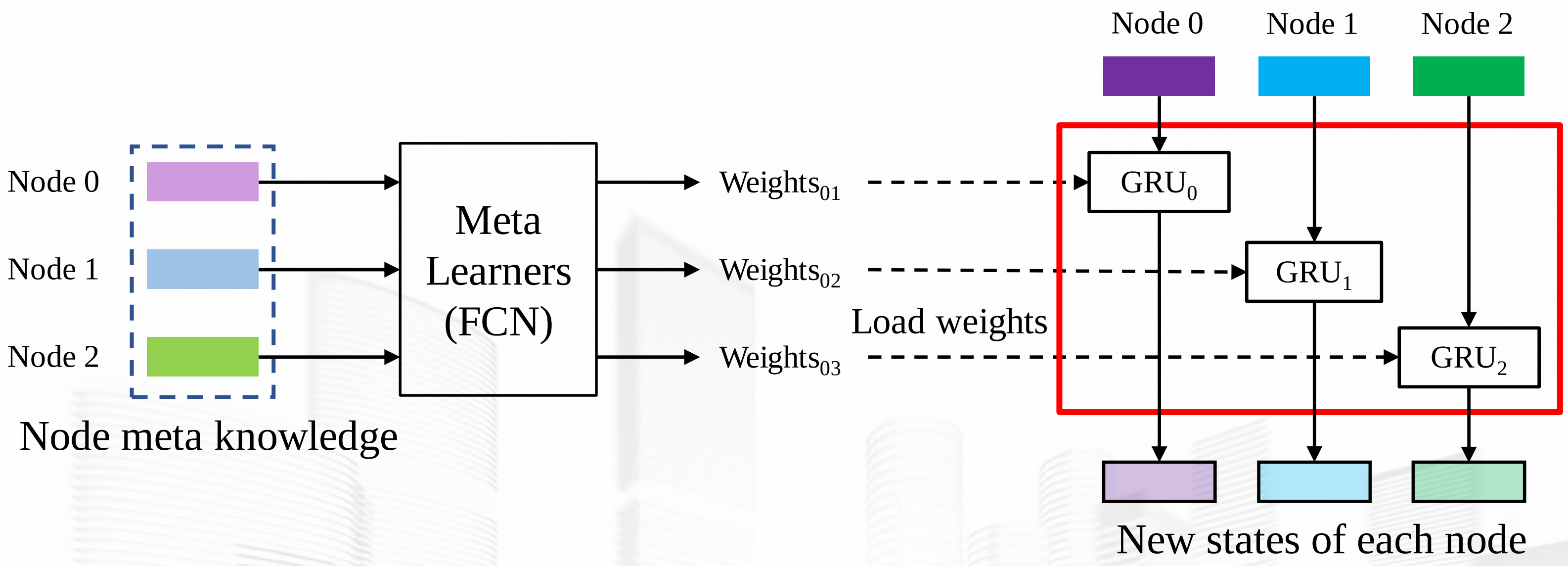
Weight generation

$$w^{(ij)} = g_W\left(MK^{(ij)}\right), b^{(ij)} = g_b\left(MK^{(ij)}\right)$$

Aggregate hidden states by softmax function

$$\bar{h}^{(i)} = \left(1 - \lambda^{(i)}\right)h^{(i)} + \lambda^{(i)} \sum_j \frac{\exp\left(w^{(ij)}\right)}{\sum_k \exp\left(w^{(ik)}\right)} h^{(j)}$$

Meta Recurrent Neural Network



$$u = \mathrm{sigmoid}\left(\boxed{W_u^{(i)}} z_t^{(i)} + \boxed{U_u^{(i)}} h_{t-1}^{(i)} + \boxed{b_u^{(i)}}\right),$$

$$r = \mathrm{sigmoid}\left(\boxed{W_r^{(i)}} z_t^{(i)} + \boxed{U_r^{(i)}} h_{t-1}^{(i)} + \boxed{b_r^{(i)}}\right),$$

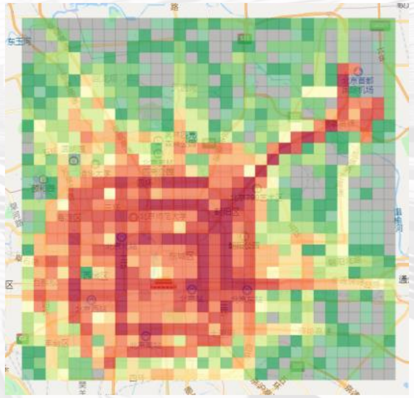
$$h_t^{(i)} = u \circ h_{t-1}^{(i)} + (1-u) \circ \tanh\left(\boxed{W_h^{(i)}} z_t^{(i)} + \boxed{U_h^{(i)}}\left(r \circ h_{t-1}^{(i)}\right) + \boxed{b_h^{(i)}}\right).$$

Evaluation

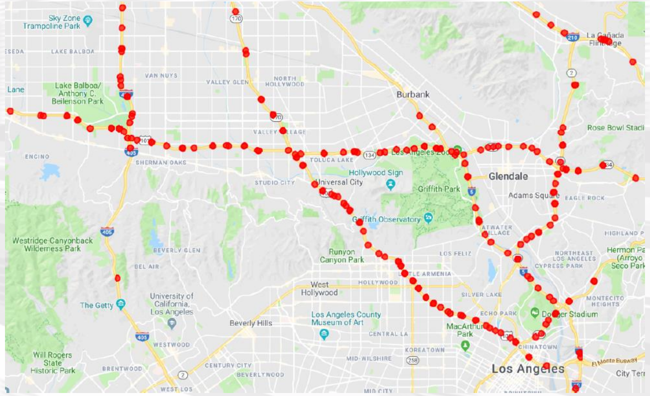
- **Datasets:** *TaxiBJ & METR-LA*
- **Metrics:**
  - $MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$
  - $MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i-\hat{y}_i|}{y_i}$
- **Baselines:**
  - **Statistics models:** Historical Average, ARIMA
  - **Tree model:** GBRT
  - **Deep models:** Seq2Seq, GAT-Seq2Seq, DCRNN, ST-ResNet

Tasks	Taxi flow prediction	Speed prediction
Prediction target	inflow & outflow	traffic speed
Timespan	2/1/2015 - 6/2/2015	3/1/2012 - 6/30/2012
Time interval	1 hour	5 minutes
# timestamps	3600	34272
# nodes	1024	207
# edges	4114	3726
# node features	989	20
# edge features	32	1

Statistics of datasets



Flow data



Traffic speed data

Evaluation on Prediction Accuracy

Learning parameters

Generating parameters

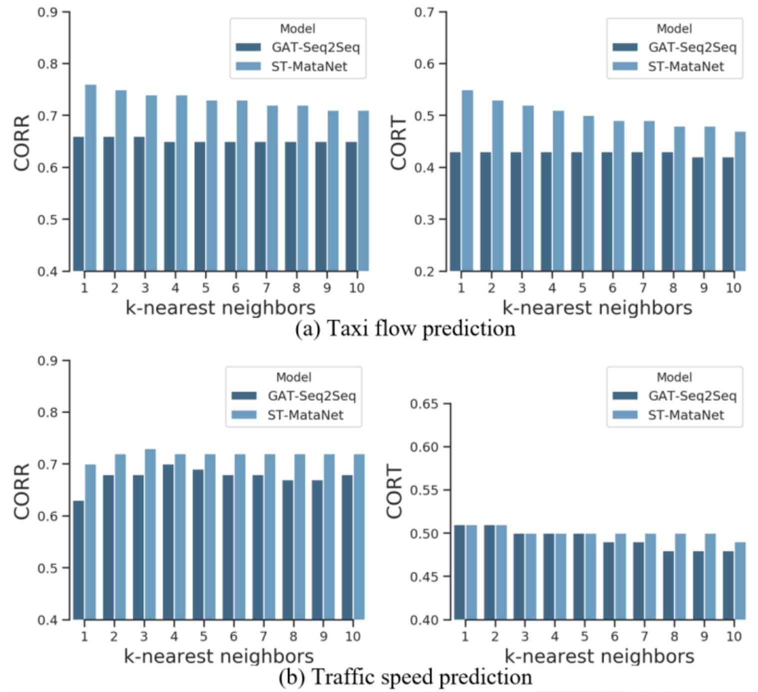
		MAE (↓)				RMSE (↓)			
城市区域流量预测		overall	1hour	2hour	3hour	overall	1hour	2hour	3hour
HA		26.2	26.2	26.2	26.2	56.5	56.5	56.5	56.5
ARIMA		40.0	27.1	41.2	51.8	86.8	58.3	77.0	108.0
GBRT		28.8	22.3	29.8	34.2	60.9	47.7	62.6	70.3
Seq2Seq	[333k]	21.3±0.06	17.8±0.05	22.0±0.06	24.2±0.09	42.6±0.14	35.1±0.07	43.6±0.16	48.1±0.20
GAT-Seq2Seq	[407k]	18.3±0.13	16.3±0.12	18.7±0.12	19.9±0.14	35.6±0.23	31.9±0.21	36.3±0.20	38.4±0.30
ST-ResNet	[445k]	18.7±0.53	16.8±0.50	18.9±0.57	20.3±0.52	36.1±0.59	31.9±0.69	36.4±0.71	39.5±0.46
ST-MetaNet	[268k]	**16.9±0.13**	**15.0±0.14**	**17.3±0.14**	**18.4±0.10**	**34.0±0.25**	**29.9±0.08**	**34.7±0.25**	**37.1±0.41**
城市路网车辆速度预测		overall	15min	30min	60min	overall	15min	30min	60min
HA		4.79	4.79	4.79	4.79	8.72	8.72	8.72	8.72
ARIMA		4.03	3.27	3.99	5.18	7.94	6.14	7.78	10.10
GBRT		3.85	3.16	3.85	4.85	7.48	6.05	7.50	9.08
Seq2Seq	[81k]	3.55±0.01	2.98±0.01	3.57±0.01	4.38±0.01	7.27±0.01	5.88±0.01	7.26±0.01	8.88±0.02
GAT-Seq2Seq	[113k]	3.28±0.00	2.83±0.01	3.31±0.00	3.93±0.01	6.66±0.01	5.47±0.01	6.68±0.00	8.03±0.02
DCRNN	[373k]	3.10±0.01	2.75±0.01	3.14±0.01	**3.60±0.02**	6.31±0.03	5.33±0.02	6.45±0.04	7.65±0.06
ST-MetaNet	[85k]	**3.05±0.02**	**2.68±0.02**	**3.09±0.03**	**3.60±0.04**	**6.25±0.02**	**5.15±0.02**	**6.25±0.05**	**7.52±0.01**

Evaluation on Meta Knowledge

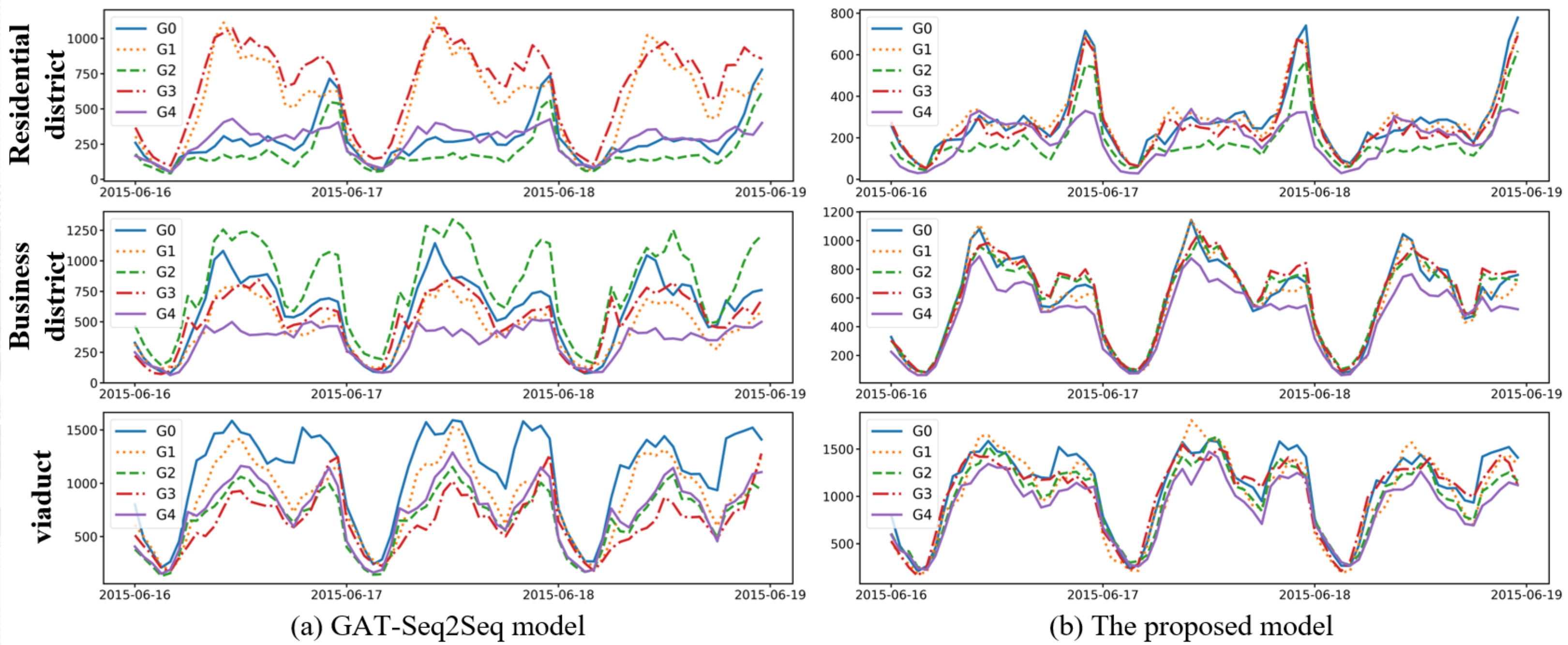
Validate that meta knowledge can reveal the similarity of ST correlations on nodes.

- ❑ For each node in the embedding space, find its k-nearest neighborhoods
- ❑ Calculate the similarity between node & its neighbor based on the test dataset

$$\mathrm{CORR}(\mathbf{x},\mathbf{y}) = \frac{\sum_i (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_i (x_i-\bar{x})^2}\sqrt{\sum_i (y_i-\bar{y})^2}}, \qquad \mathrm{CORT}(\mathbf{x},\mathbf{y}) = \frac{\sum_i (x_i-x_{i-1})(y_i-y_{i-1})}{\sqrt{\sum_i (x_i-x_{i-1})^2}\sqrt{\sum_i (y_i-y_{i-1})^2}},$$



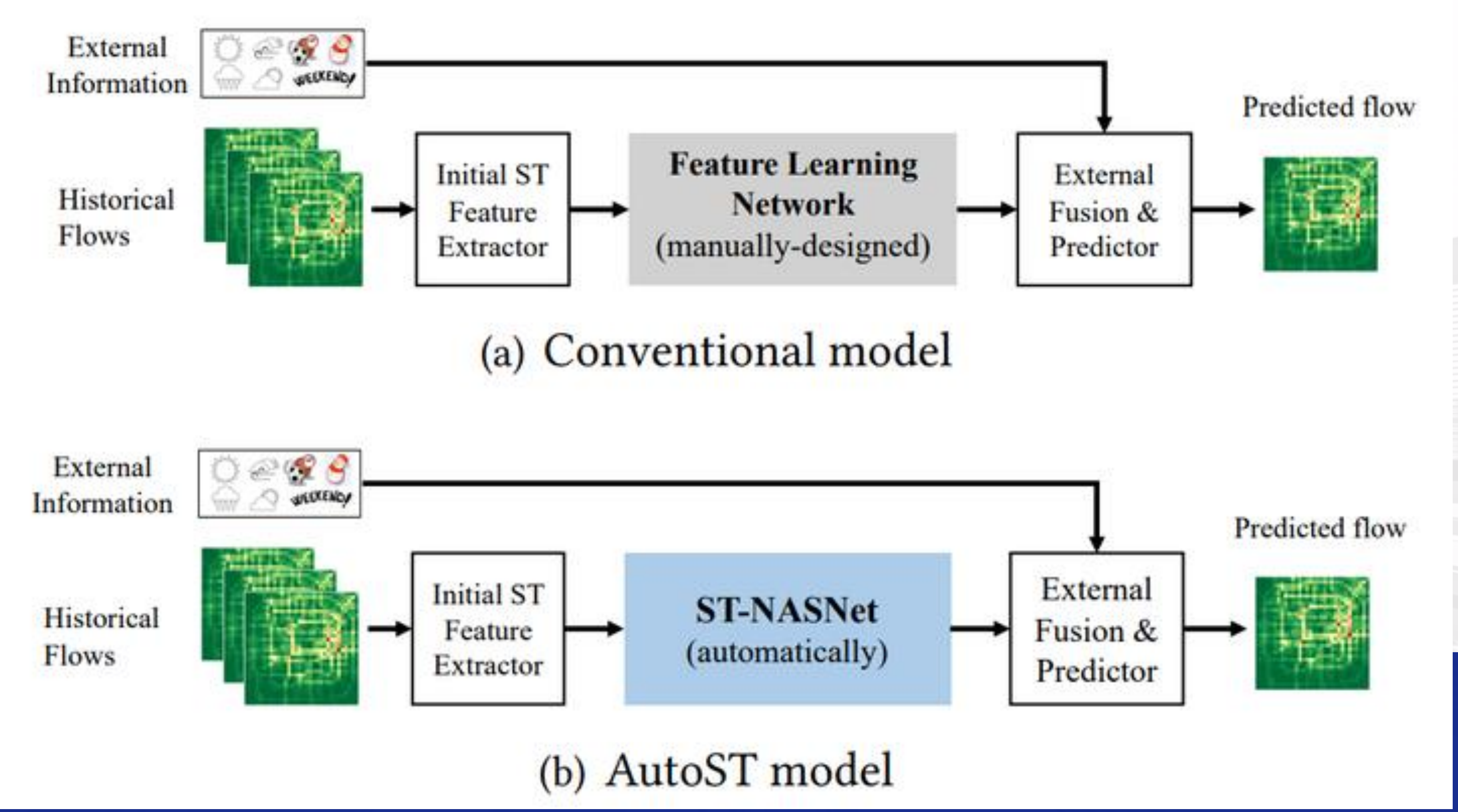
(a) Taxi flow prediction
(b) Traffic speed prediction



(a) GAT-Seq2Seq model
(b) The proposed model

AutoST: Efficient Neural Architecture Search for Spatio-Temporal Prediction

KDD 2020



Motivation

- The spatio-temporal correlation is：
  - Heterogeneous to different tasks varying from bike flow to taxi flow.
  - Diverse to traffic conditions from core city to small city.

Bike Flow

Taxi Flow

Guiyang

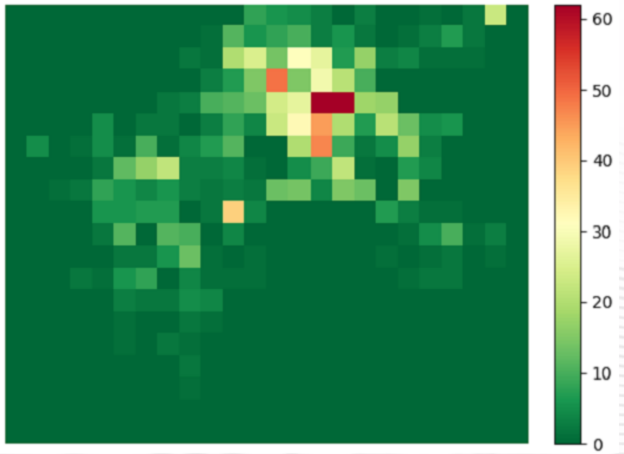
Beijing

Is there a network suitable for all scenarios?

Observations

- **Two observations** illustrating the optimal architecture is distinct among different cities:
  - Cities may have different spatial range preferences
  - Low and high-level features do not contribute equally in all cases.

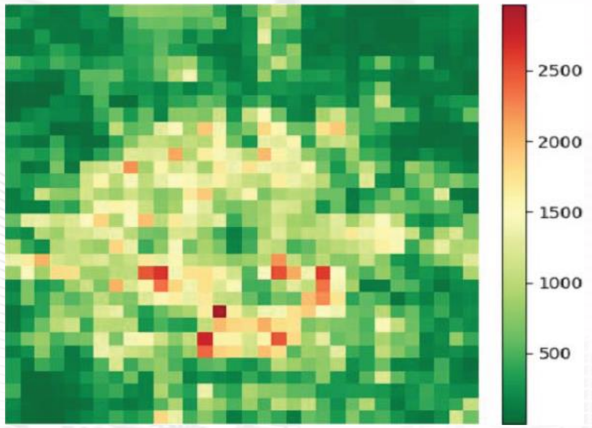
20 x 24



Taxi flow in Guiyang

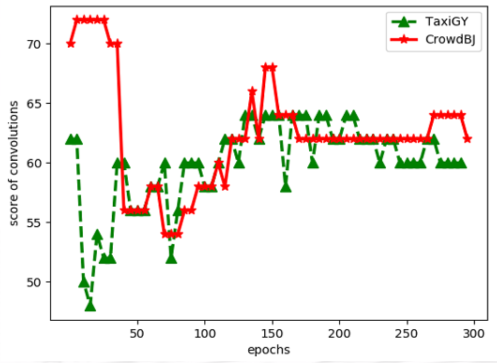
Local Dependency

32 x 32



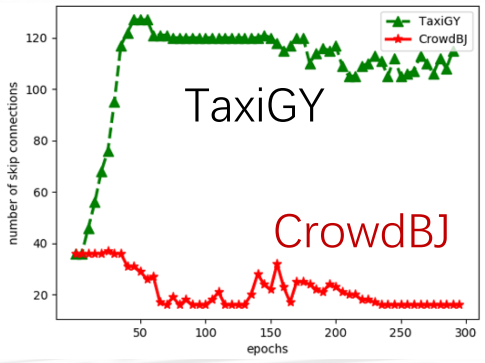
Crowd flow in Beijing

Global Dependency



Convolutions

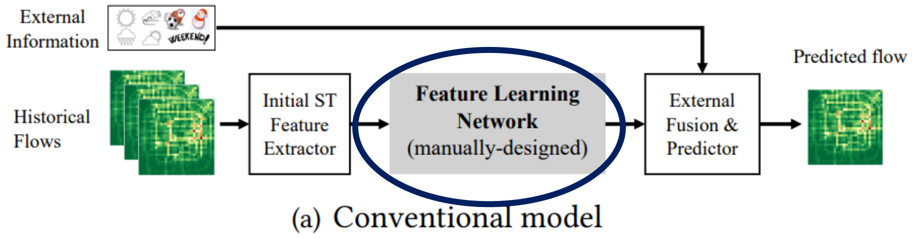
CrowdBJ needs larger convolution kernels

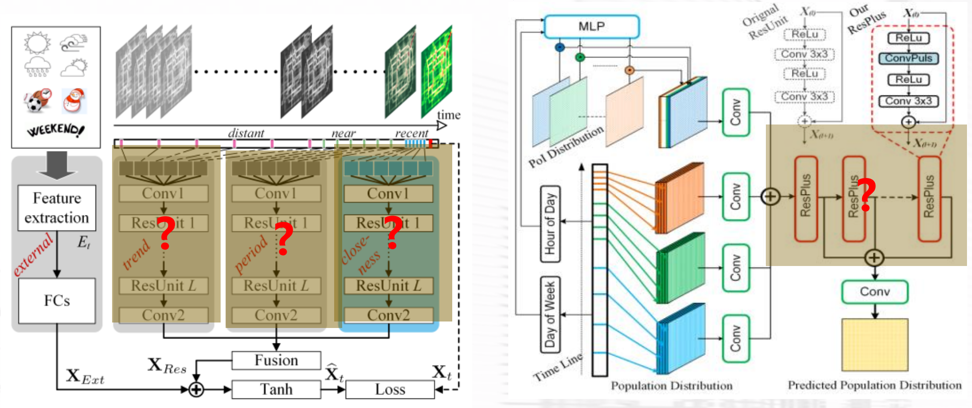


Skip Connections

TaxiGY requires more skip connections

Goal



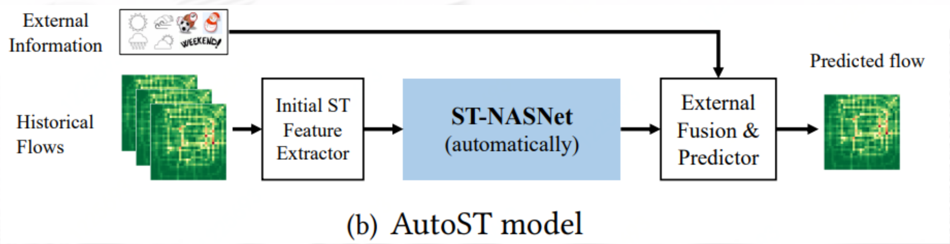


ST-ResNet (AAAI' 17)   DeepSTN+(AAAI' 19)

Conventional model designs the feature network manually！

- AutoST model：
  - *Multi-range(long-range/short-range)* correlations automatically.
  - *Fusing multi-level* features dynamically.

Goal



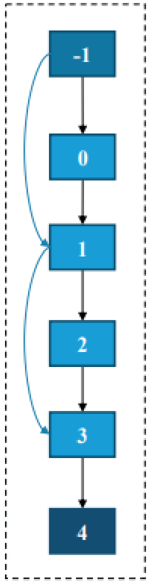
Oriented to *ST data* rather than specific task！

Methodology-Search Space

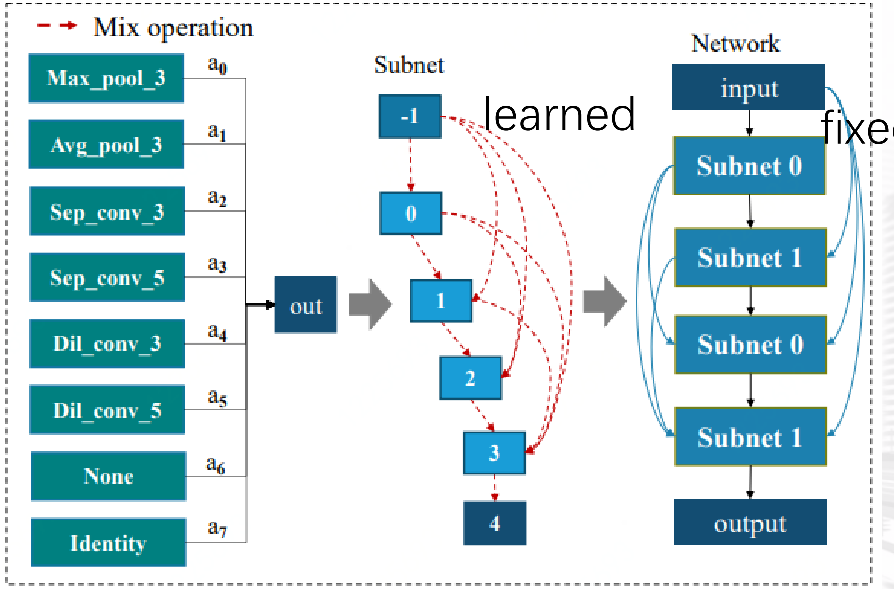
- ResNet
  - Fixed Convs
  - Fixed Conns

- Search space of *Darts* (NAS for CV) :
  - mix op block includes all candidate ops (convs, pools and conns).
  - subnet (automatic) and outer net (fixed).

- Search space of *ST-NASNet*:
  - mix convolution block
  - mix connection block
  - NAS network (automatic)



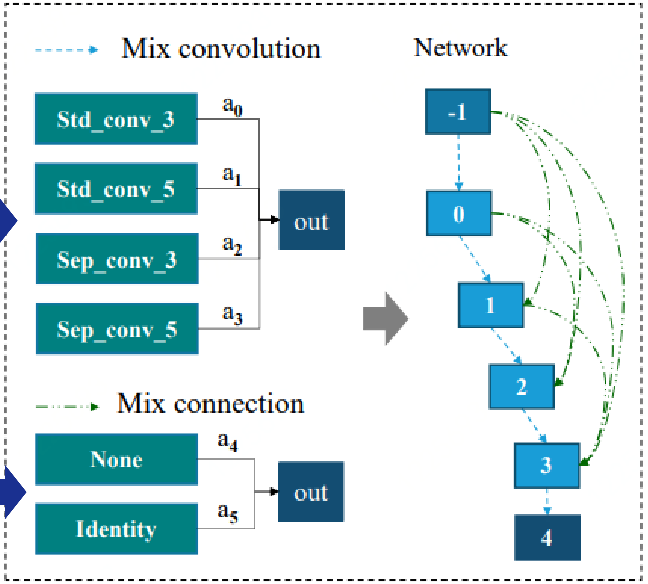
(a) ResNet



(b) Search space of Darts

Multi-ranges

Multi-levels



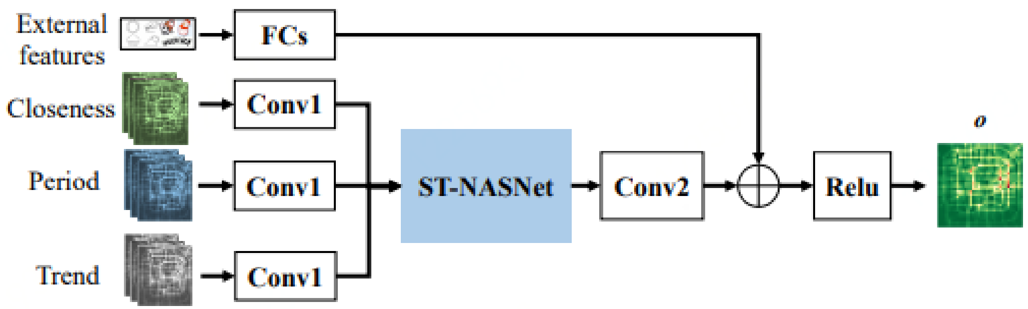
(c) Search space of ST-NASNet

Experiment-Settings

- **Dataset Description:**
  - Four datasets including TaxiBJ, CrowdBJ, TaxiJN and TaxiGY.
- **Baseline algorithms:**
  - ST models: ST-ResNet, ST-3DNet and DeepSTNPlus.
  - NAS models: ENAS and DARTS.
- ST-ResNet+, ST-3DNet+ and DeepSTNPlus+ represent the AutoST enhanced models.

Dataset	Time spans	Grid size	# Ins
TaxiBJ	7/1/2013-10/30/2013 3/1/2014-6/30/2014 3/1/2015-6/30/2015 11/1/2015-4/10/2016	(32,32)	15072
CrowdBJ	9/1/2017-11/30/2017	(32, 32)	2016
TaxiJN	9/1/2017-1/31/2018	(32,16)	3323
TaxiGY	10/1/2018-5/26/2019	(20, 24)	5270

Dataset



DeepSTNPlus+

Experiments-Overall Performances

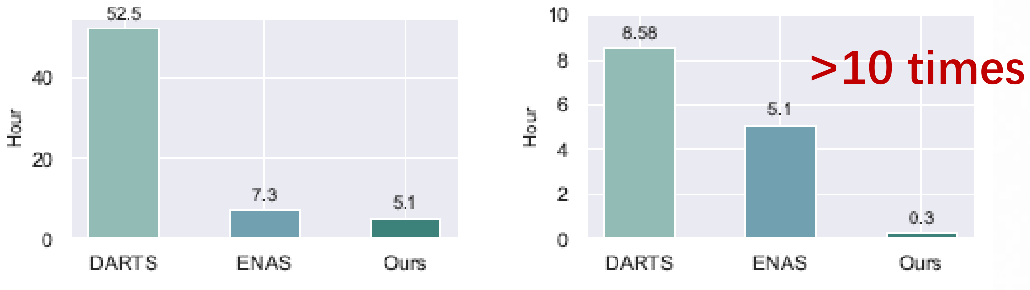
- Effectiveness perspective:
  - AutoST can be applied to a *wide-range* of ST models and steadily improves the performances.
  - DeepSTNPlus+ achieves *state-of-the-arts* results at all scenarios.

Models	Param	RMSE	MAPE
ST-ResNet	0.92M	17.51 ± 0.05	33.92% ± 0.41%
ST-ResNet+	3.38M	17.47 ± 0.05	33.52% ± 0.40%
ST-3DNet	0.54M	17.82 ± 0.36	31.04% ± 0.02%
ST-3DNet+	1.36M	17.37 ± 0.20	27.77% ± 0.02%
DeepSTN-ne	0.42M	16.09 ± 0.02	27.05% ± 0.15%
DeepSTN-ne+	1.24M	15.97 ± 0.06	27.72% ± 0.14%
DeepSTNPlus	0.44M	15.98 ± 0.05	26.52% ± 0.64%
DeepSTNPlus+	1.26M	15.88 ± 0.19	25.97% ± 0.65%

Models	CrowdBJ		TaxiJN		TaxiGY	
	RMSE	MAPE	RMSE	MAPE	RMAE	MAPE
ST-ResNet	92.27 ± 4.42	74.24% ± 4.53%	5.876 ± 0.26	62.22% ± 0.80%	2.773 ± 0.10	56.95% ± 1.09%
ST-ResNet+	87.35 ± 4.42	63.17% ± 4.53%	5.624 ± 0.06	72.30% ± 1.92%	2.521 ± 0.07	51.69% ± 0.59%
ST-3DNet	76.13 ± 2.14	55.51% ± 1.18%	5.458 ± 0.19	58.71% ± 2.71%	2.574 ± 0.08	52.71% ± 2.27%
ST-3DNet+	62.28 ± 2.68	36.56% ± 3.49%	5.103 ± 0.04	57.11% ± 1.54%	2.488 ± 0.04	51.32% ± 0.78%
DeepSTN-ne	52.49 ± 0.37	32.17% ± 1.94%	4.664 ± 0.05	45.69% ± 0.97%	2.175 ± 0.02	50.81% ± 0.20%
DeepSTN-ne+	51.38 ± 0.61	28.43% ± 1.98%	4.653 ± 0.20	46.58% ± 0.65%	2.169 ± 0.03	47.61% ± 0.06%
DeepSTNPlus	49.76 ± 0.57	28.60% ± 2.75%	4.653 ± 0.01	54.52% ± 0.30%	2.172 ± 0.06	50.01% ± 0.71%
DeepSTNPlus+	49.09 ± 0.61	29.08% ± 5.80%	4.602 ± 0.00	44.35% ± 0.87%	2.157 ± 0.01	49.55% ± 0.87%

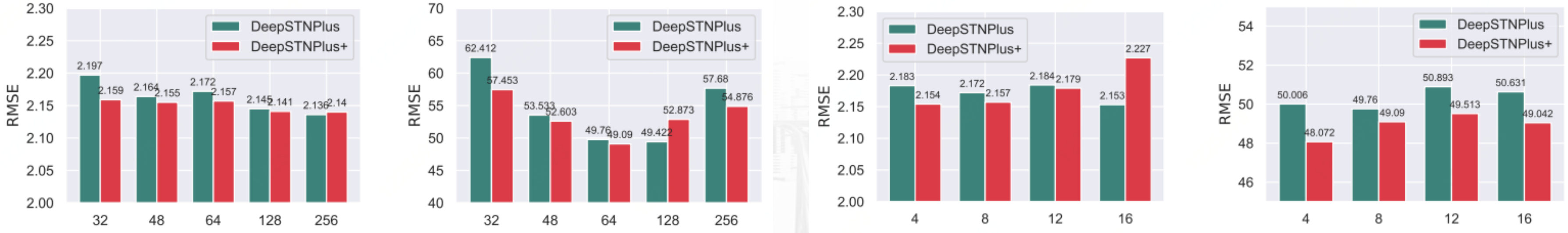
Experiments-Efficiency and Robustness

- Efficiency perspective: AutoST searches faster than existing NAS algorithms



(c) Time on TaxiBJ (d) Time on TaxiGY

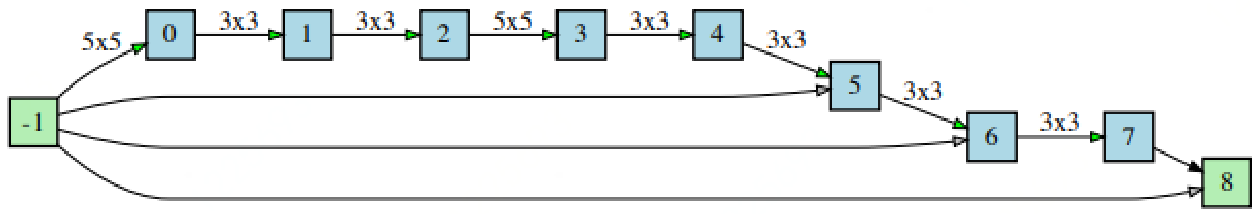
- Robustness perspective: the proposed model outperforms baselines with most settings.



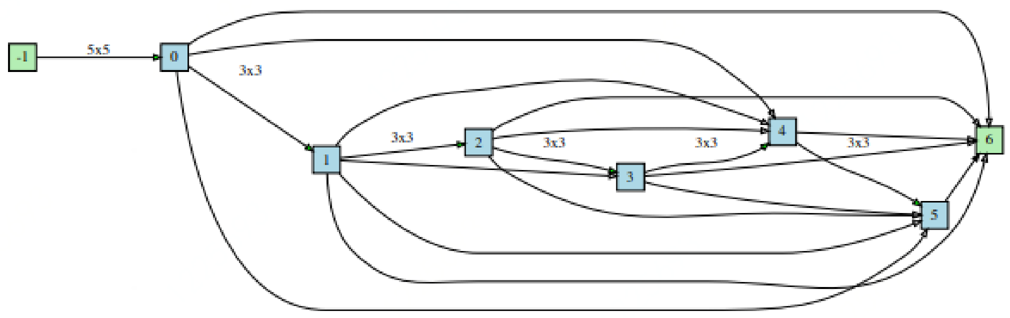
(a) # channels on TaxiGY (b) # channels on CrowdBJ (c) # layers on TaxiGY (d) # layers on CrowdBJ

Experiments-Case Study

- The optimal architecture on CrowdBJ has:
  - *no connections* at first four layers showing the long-range correlation captured by stacked multi-layer convolutions is important.
  - *connections* at last four layers to fuse the spatial features with neighborhood information.
- The optimal architecture on TaxiGY has:
  - *a large amount of skip connections* among layers indicating that the short-range neighborhood dependency contributes more than global features.

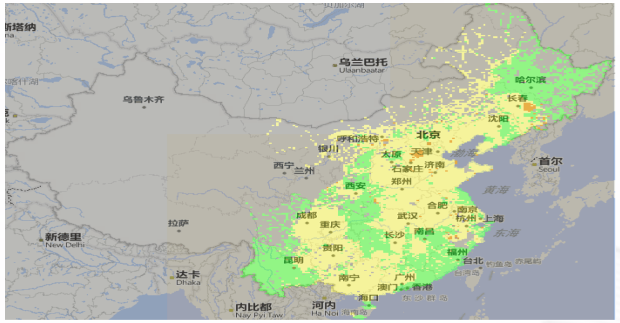


(a) CrowdBJ

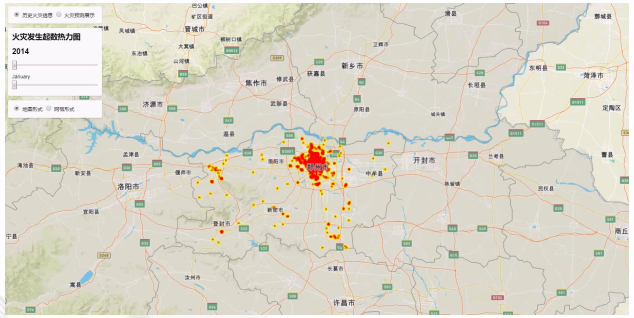


(b) TaxiGY

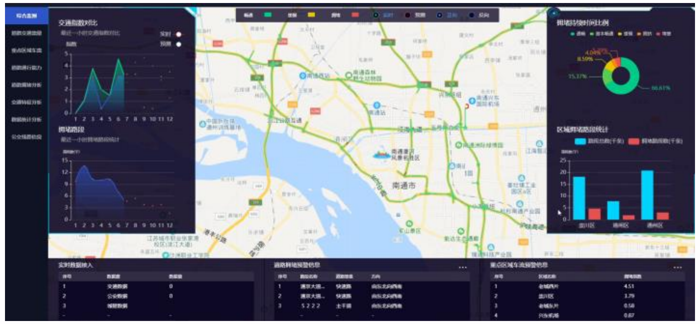
Part 3. More *Spatio-Temporal* AI Applications



城市空气质量预测系统



城市火灾监控与预测系统



城市交通流量预测系统

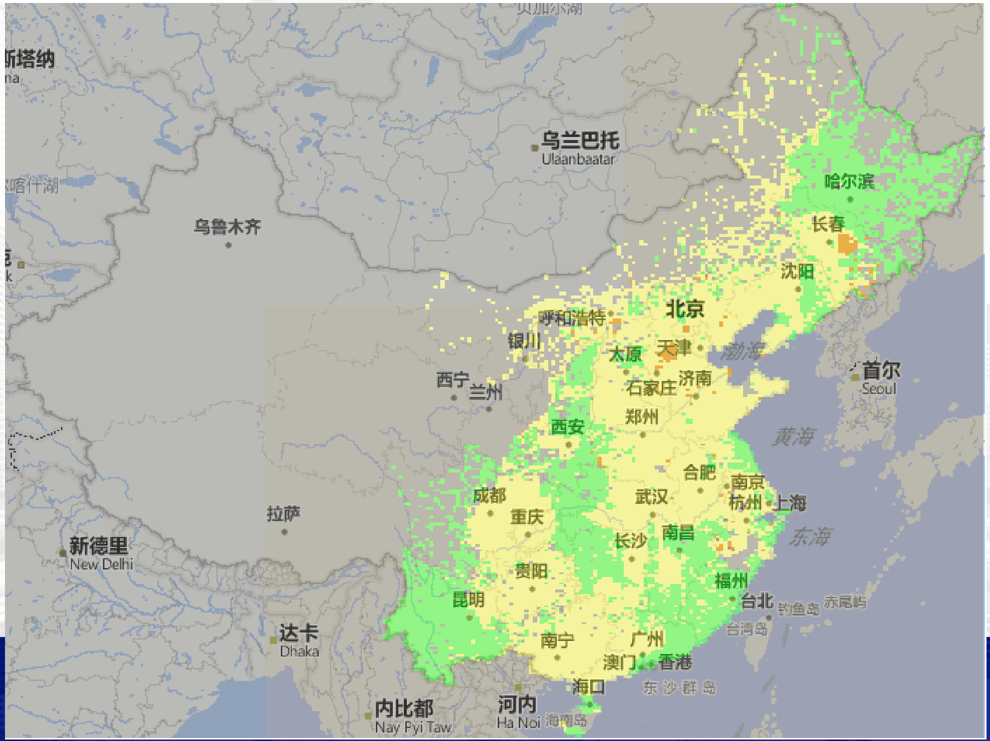
大数据智能选址平台

城市地铁客群分析系统

城市画像系统

Deep Distributed Fusion Network for Air Quality Prediction

KDD 2018

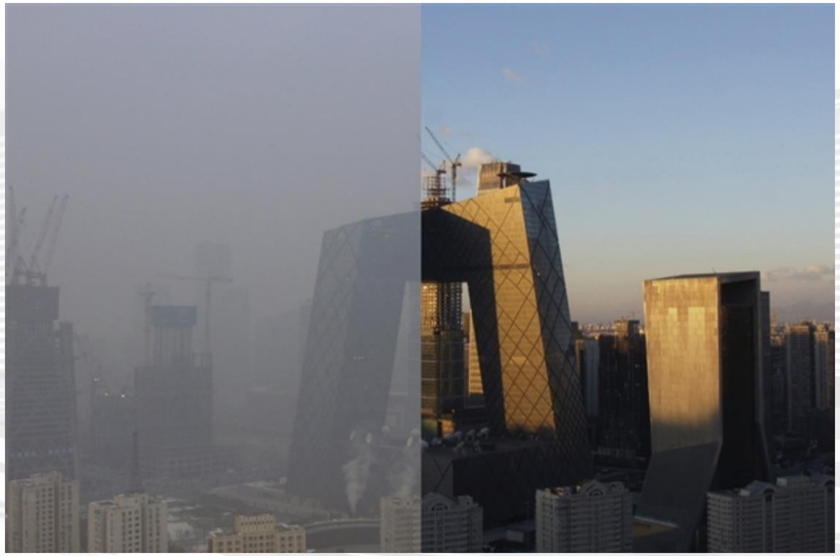


京东城市

京 东 数 科 旗 下

Motivation

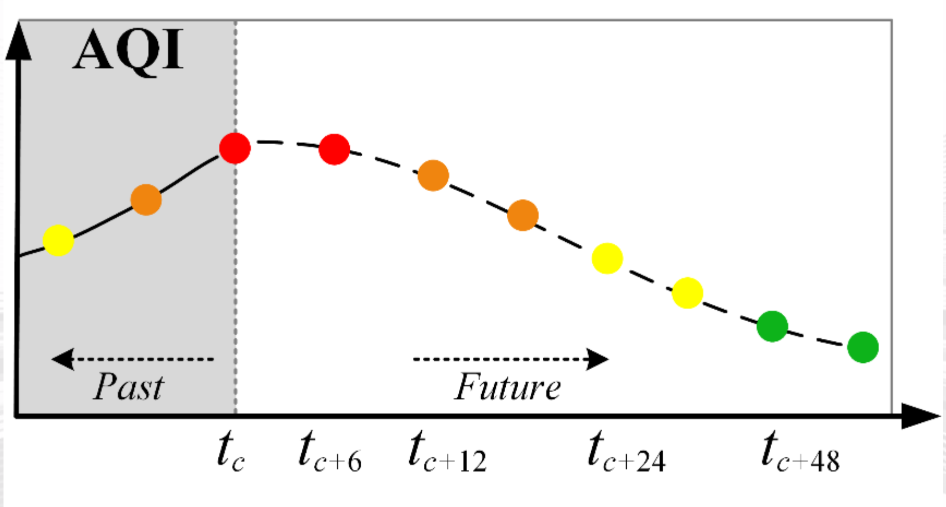
- Background
  - Developing countries are suffering from seriously air pollution problem
  - Air pollution consists of a mixture of particulate matter ($PM_{2.5}$, PM_{10}) and gaseous species (NO_2, CO, O_3, SO_2)
  - Health alert to young and elderly for breathing problems



- Prediction demand
  - Support government’ s policy making
  - Inform people decision making

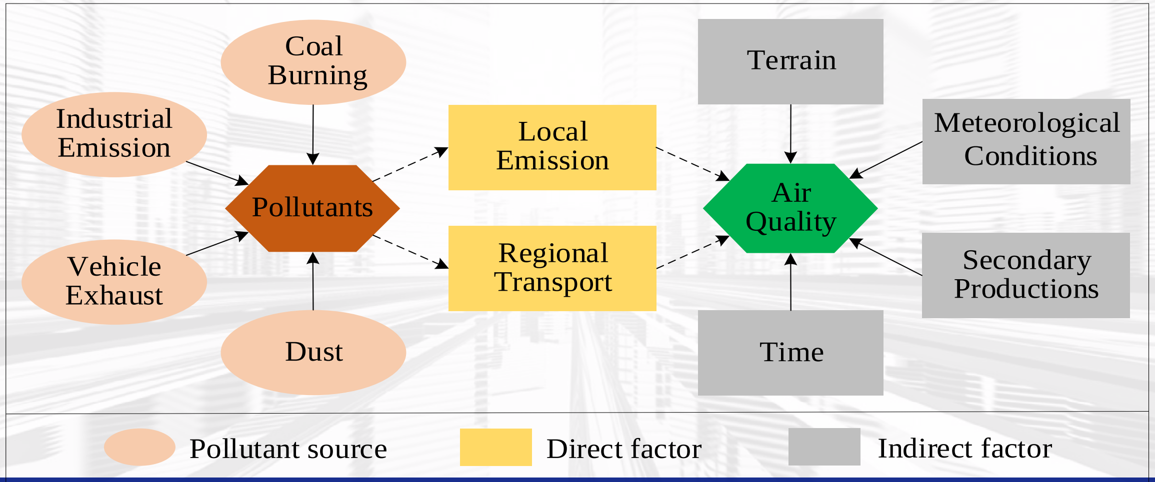
Goal

- Predicting Fine-Grained Air Quality ($PM_{2.5}$)
  - Spatial granularity
    - For each air quality monitoring station
  - Temporal granularity
    - For each hour over the next 48 hours



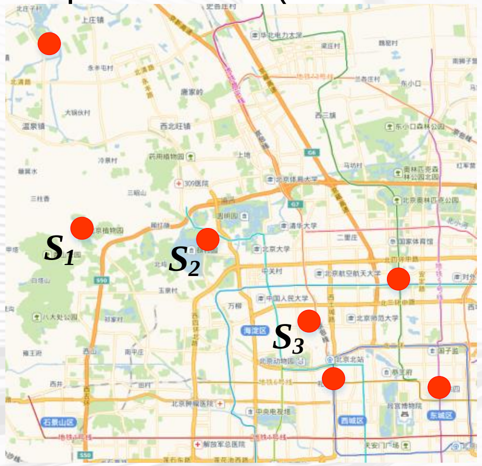
Challenge

- Multiple influential factors with complex interactions
  - Pollution sources, direct factors and indirect factors
    - Insufficient and inaccurate data
  - Affected by multiply factors simultaneously
    - Effect from one specific factor is not absolute
    - Hard to decide the weight for each factor
    - Difficult to capture the dispersion patterns of air pollution

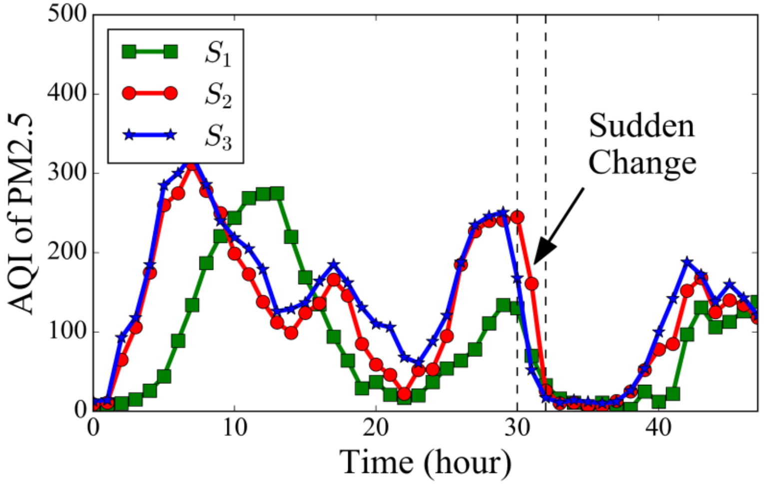


Challenge

- Dynamic spatio-temporal correlation and sudden changes
  - Urban air changes over location and time significantly
  - AQI drops very sharply in a very short time span
    - Caused by some specific factors (e.g., wind)
    - People pay more attention to special cases than normal cases
    - Data imbalance problem (2.3% within 3 years data in Beijing)

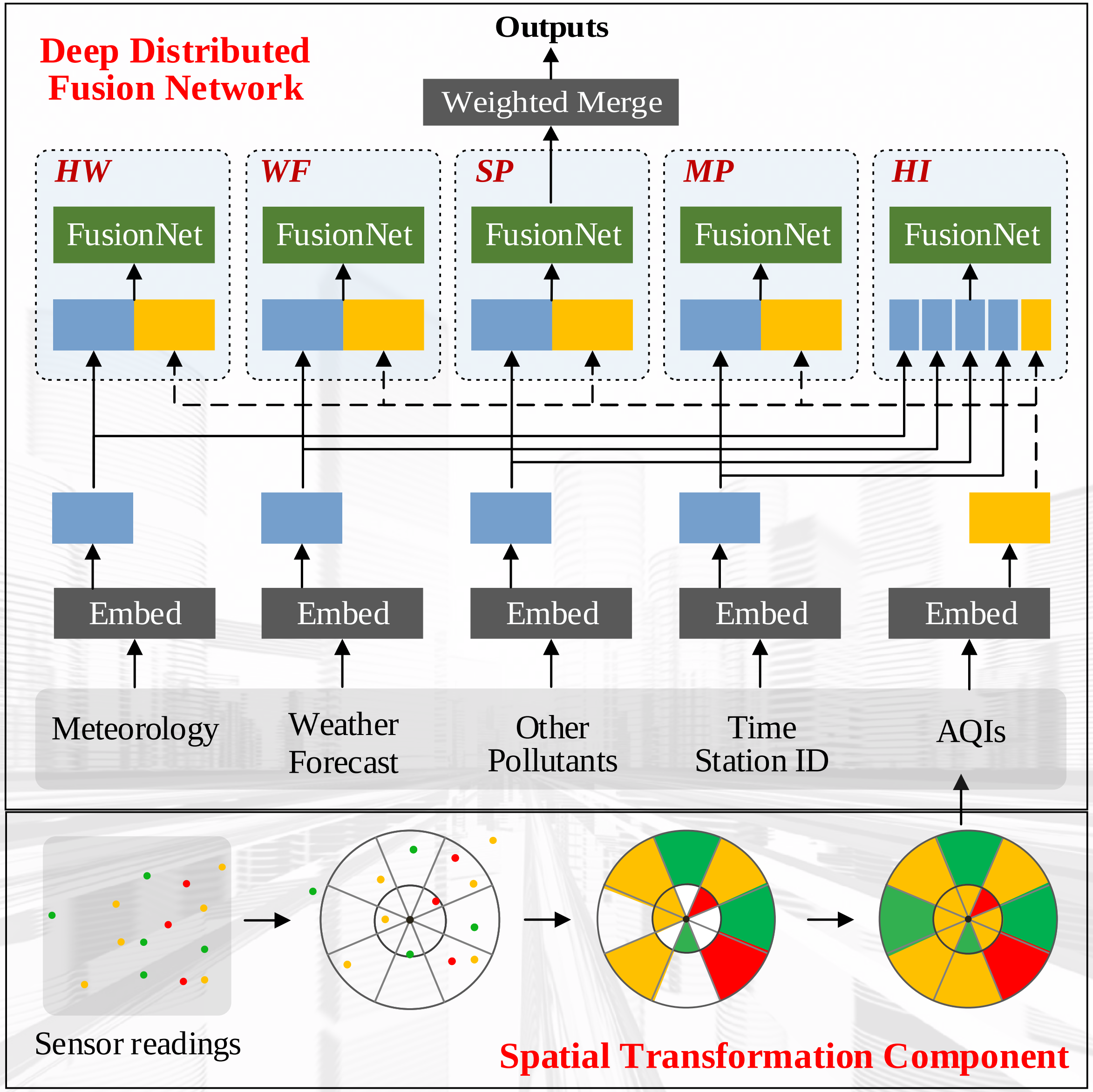


A) Monitoring stations



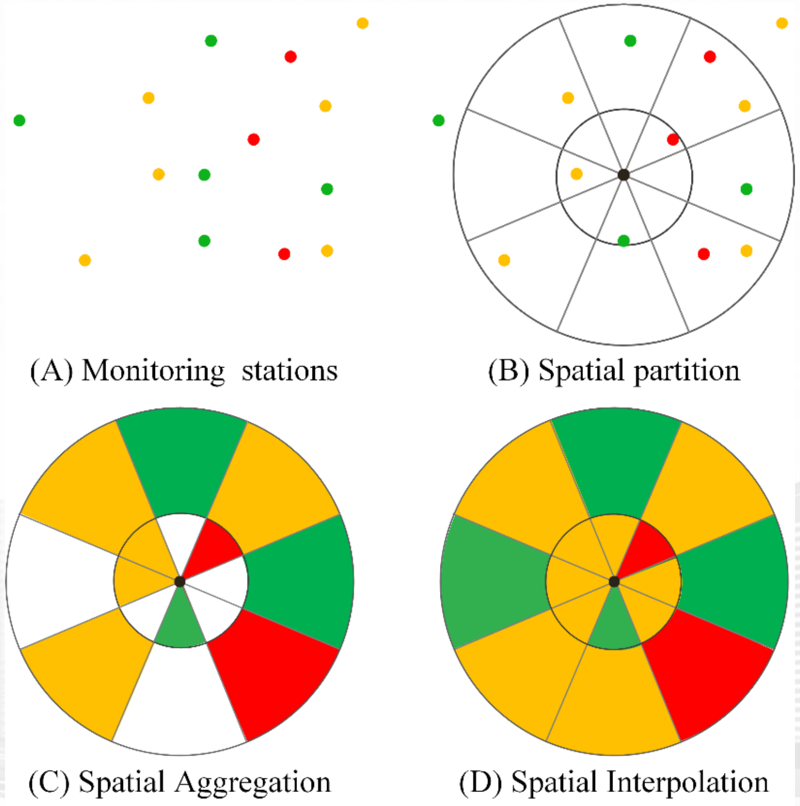
B) AQI change over time

Framework

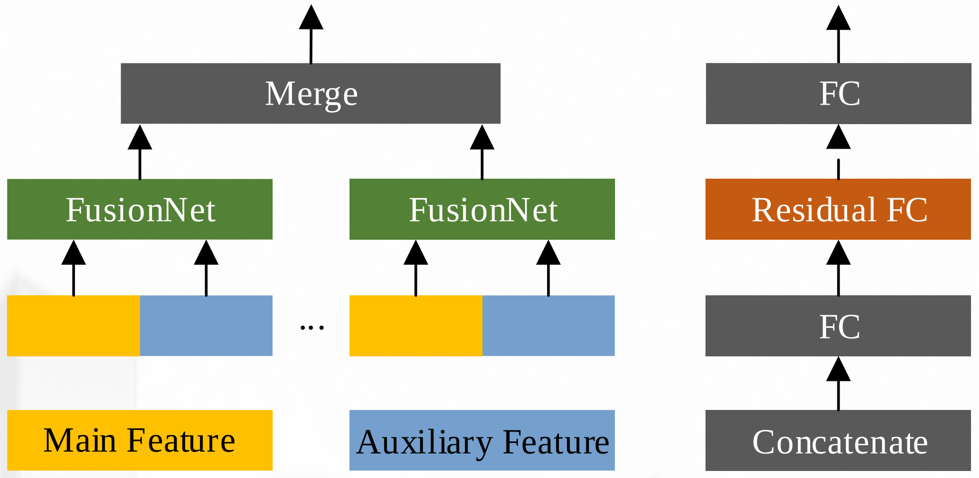

Outputs
Deep Distributed
Fusion Network
Weighted Merge
HW
WF
SP
MP
HI
FusionNet
FusionNet
FusionNet
FusionNet
FusionNet
Embed
Embed
Embed
Embed
Embed
Meteorology
Weather
Forecast
Other
Pollutants
Time
Station ID
AQIs
Sensor readings
Spatial Transformation Component

Spatial Transformation

- Air pollution dispersion
  - AQI recorded by monitoring stations can be regarded as second-hand pollution sources
- Spatial correlation
  - Regions with different distance show different impacts varying by distance
  - Closer regions have a finer granularity and farther regions have a coarser granularity
- Scalability
  - Set number of regions as upper bound
  - Overcoming spatial sparsity
  - Train one model with all stations' data



Distributed Fusion



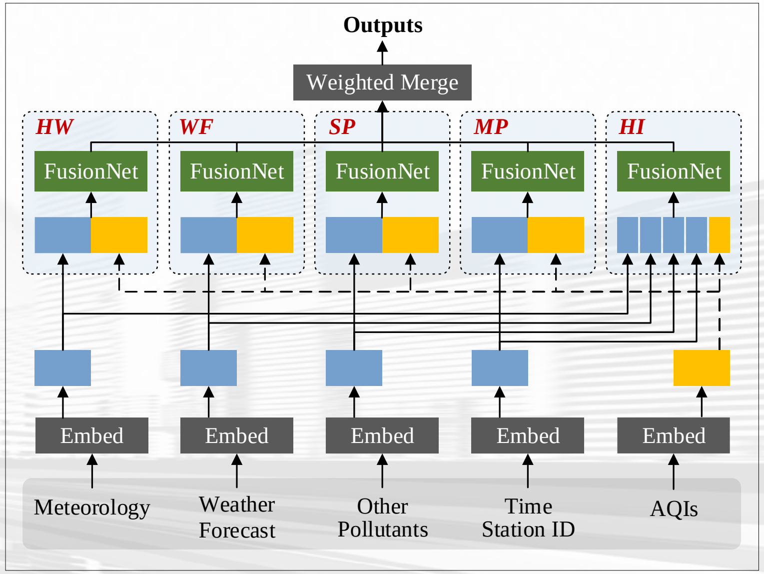
(a) Distributed Fusion (b) FusionNet

- Main Feature vs auxiliary feature
  - Main Feature and prediction target come from same domain
- Distributed fusion
  - Main feature fuses each auxiliary feature in a parallel manner
  - Highlight main feature and capture each auxiliary feature's effect

Subnet

- Individual influence subnets (FusionNet)
  - HW: capture the influence of historical weather
  - WF: learn the impact from future weather conditions
  - SP: simulate secondary chemical production
  - MP: model the effect of time and terrain
- Holistic influence subnet (FusionNet)
  - Learn holistic influence from all factors simultaneously
- Weighted Merge
  - Determine the weight for each factor

$$\hat{y} = Sigmoid(\, y_{hw} \circ w_{hw} + y_{wf} \circ w_{wf} + y_{sp} \circ w_{sp} + y_{mp} \circ w_{mp} + y_{hi} \circ w_{hi})$$



Embedding

- Transform raw features to a low-dimensional space
  - Learn intra-dynamics of each domain data
  - Capture temporal information

<table>
<tr><th colspan="2">Domain</th><th>Feature</th><th>Encoding</th><th>Embedding</th></tr>
<tr><td colspan="2">AQIs</td><td>PM2.5</td><td>6*17</td><td>36</td></tr>
<tr><td colspan="2" rowspan="6">Other Pollutants</td><td>PM_{10}</td><td>6*1</td><td rowspan="6">6</td></tr>
<tr><td>NO_2</td><td>6*1</td></tr>
<tr><td>SO_2</td><td>6*1</td></tr>
<tr><td>CO</td><td>6*1</td></tr>
<tr><td>O_3</td><td>6*1</td></tr>
<tr><td colspan="2" rowspan="5">Meteorology</td><td>Weather</td><td>6*8</td><td rowspan="5">6</td></tr>
<tr><td>Wind direction</td><td>6*4</td></tr>
<tr><td>Wind speed</td><td>6*1</td></tr>
<tr><td>Humidity</td><td>6*1</td></tr>
<tr><td>Pressure</td><td>6*1</td></tr>
<tr><td colspan="2" rowspan="3">Weather Forecast</td><td>Weather</td><td>k*8</td><td rowspan="3">6</td></tr>
<tr><td>Wind direction</td><td>k*4</td></tr>
<tr><td>Wind Strength</td><td>k*4</td></tr>
<tr><td rowspan="4">Meta Property</td><td>Station ID</td><td>Beijing</td><td>36</td><td rowspan="4">6</td></tr>
<tr><td rowspan="3">Time</td><td>Season</td><td>4</td></tr>
<tr><td>isWorkday</td><td>2</td></tr>
<tr><td>Hour</td><td>4</td></tr>
</table>

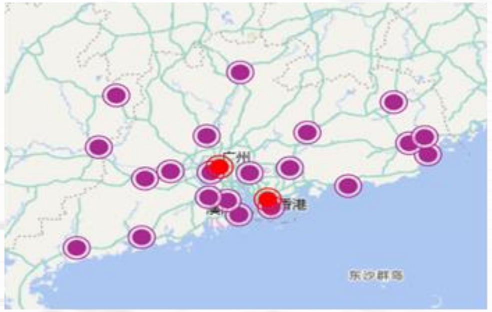
Data in Urban Air

- Air quality data
  - From 2,296 stations in 302 Chinese cities
  - Hourly updates
  - Convert concentrations into corresponding AQI based on Chinese AQI standards
- Meteorological data
  - 3,514 cities/districts/stations
  - a district-level (or even finer) granularity
  - Hourly update
- Weather forecasts
  - 2,612 cities/districts
  - Next three days forecast (3-hour segment)
  - Updating frequency: 12-hour

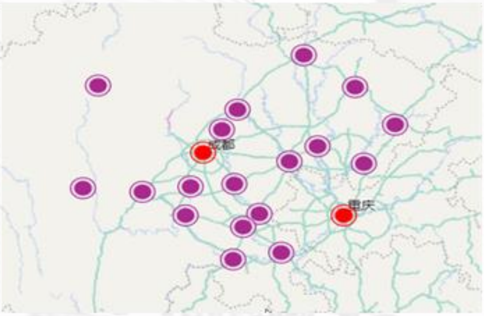
Datasets

Beijing, Tianjin

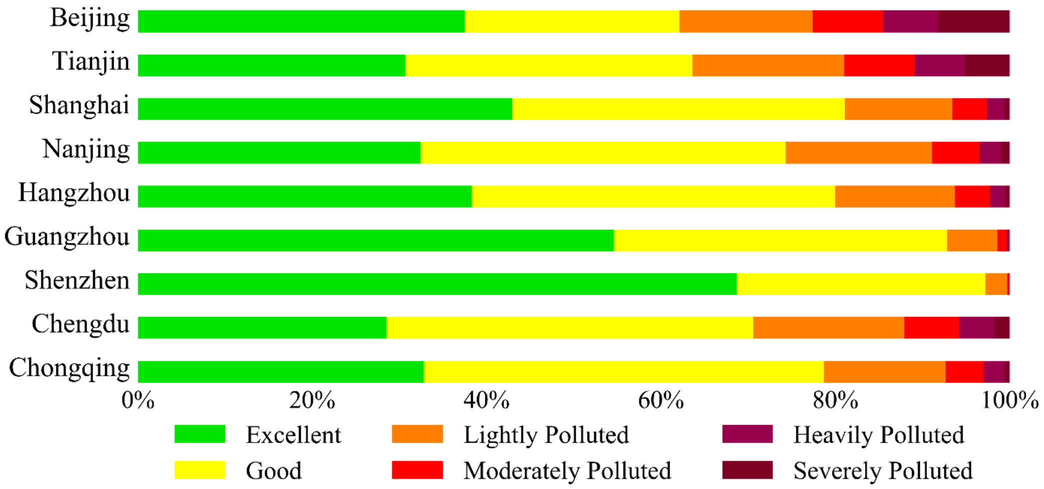
Shanghai, Nanjing, Hangzhou



Guangzhou, Shenzhen



Chengdu, Chongqing



Datasets		Beijing
Time span	2014/5/1-2017/4/30	
AQI	In-city stations	36
	In-city instances	875,394
	Sudden changes	20,540
	Average PM2.5	108.2
	Neighbor stations	74
Meteorology	In-city sources	17
	In-city instances	327,514
Weather Forecast	In-city sources	17
	In-city instances	1,282,918

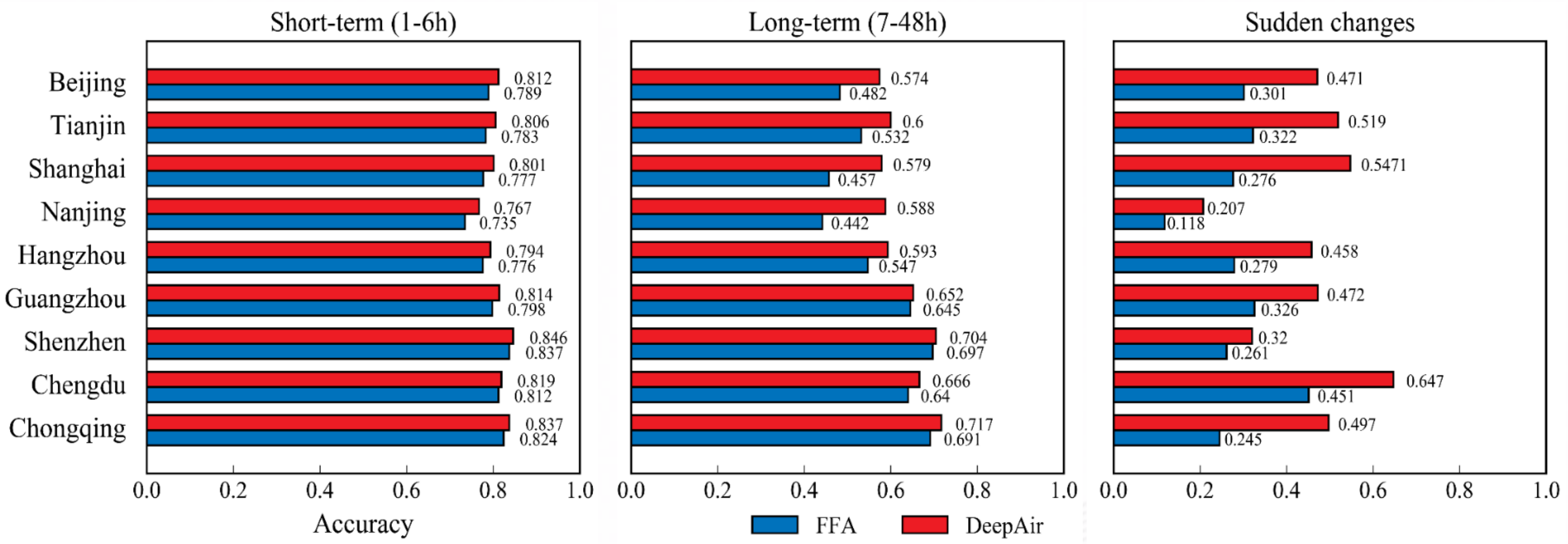
Overall Results

Method	1-6h		7-12h		13-24h		24-48h		Sudden Change	
	acc	*mae*	*acc*	*mae*	*acc*	*mae*	*acc*	*mae*	*acc*	*mae*
ARIMA	0.751	28.3	0.576	52.1	0.458	65.4	0.307	74.6	0.066	112.9
LASSO	0.790	21.9	0.620	39.7	0.534	48.9	0.452	57.1	0.273	87.2
GBDT	0.792	21.8	0.629	38.8	0.540	48.0	0.458	56.5	0.321	21.8
LSTM	0.780	23.1	0.606	41.2	0.491	53.2	0.380	64.8	0.240	90.1
LSTM-STC	0.794	21.6	0.622	39.6	0.508	51.4	0.396	63.0	0.314	82.5
DeepST	0.806	20.4	0.633	38.1	0.545	47.5	0.466	55.7	0. 38	79.5
DJMVST-Net	0.806	20.4	0.638	37.8	0.550	47.4	0.481	53.9	0.419	70.4
DeepFM	0.808	20.1	0.643	37.3	0.549	47.2	0.474	54.9	0.396	72.3
DeepSD	0.811	19.7	0.645	37.1	0.551	46.8	0.479	54.3	0.428	69.5
DeepAir	**0.812**	**19.5**	**0.656**	**36.1**	**0.569**	**45.1**	**0.5**	**52.1**	**0.471**	**63.8**

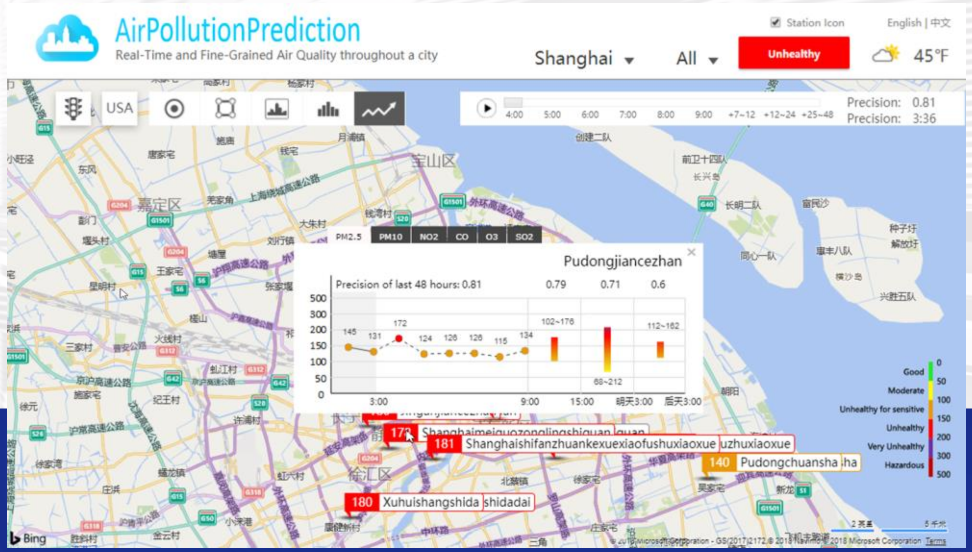
$$\mathrm{acc} = 1 - \frac{\sum_i |\hat{y}_i - y_i|}{\sum_i y_i}$$

$$mae = \frac{\sum_i |\hat{y}_i - y_i|}{n}$$

Online Performance



2.4%, 12.2%, 63.2% relative accuracy improvements on **short-term, long-term and sudden** changes



京东城市

京 东 数 科 旗 下

Official Prediction

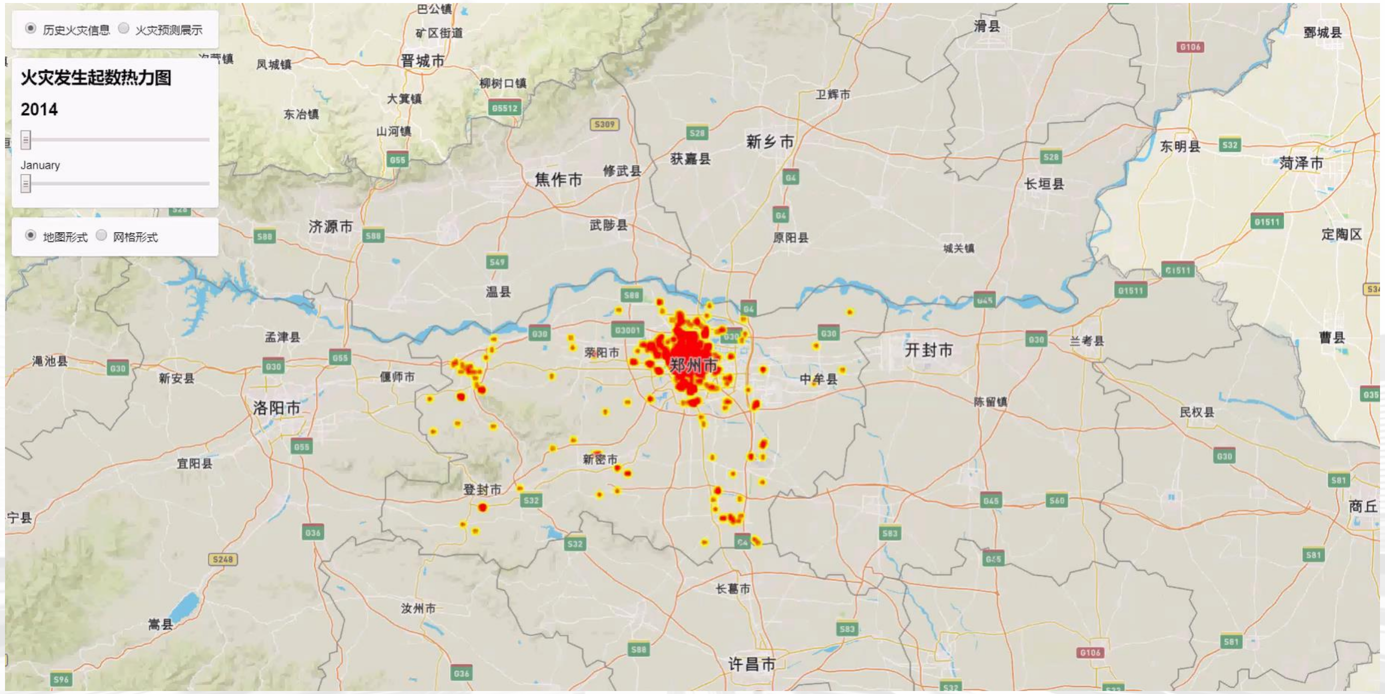
- Advantages beyond Weather-Forecast-Based Method (WFM)
  - Spatial granularity: station vs district
  - Farther predictive capability: 48 vs 12 hours
  - Updating frequency: 1 hour vs 12 hours
  - Need less data sources
  - More accurate

10/1/2014 to 12/30/2016.
Beijing Municipal Environmental Monitoring Center (using *WFM*)

Method	Station Level		District Level		Update	Grained
	acc	*mae*	*acc*	*mae*	Hours	level
WFM	0.54	54.5	0.64	46.1	12	District
DeepAir	0.77	26.7	0.86	17.9	1	Station

CityGuard: Citywide Fire Risk Forecasting Using A Machine Learning Approach

UbiComp 2020



京东城市

京　东　数　科　旗　下

Background

Urban fire cause financial loss, injuries and even deaths.

Center of Fire Statistics[1] shows fire statistics from 34 countries in 2017:

- 3,115,061 fires cause 16,808 deaths and 47,948 injuries.

- Examples
  - 15, Apr., 2019, a major fire has engulfed the Notre Dame de Paris, destroying priceless treasures.[2]
  - 12, Aug., 2015, a series of explosions killed 173 people and injured hundreds of others at a container storage station at the Port of Tianjin.[3]

[1] www.ctif.org
[2] https://www.bbc.com/news/world-europe-47941794
[3] https://en.wikipedia.org/wiki/2015_Tianjin_explosions

Goal

Fire forecasting is a necessary work to control the risks of fire

Traditional fire prevention deployment

1. Specialist
2. Based on statistics

} Coarse-grained in temporal and spatial

help

Fire Risk Rank

For fire prevention, fire-fighting department would like to learn the region with top risks (more fires) rather than whether fire or the number of fire in each area when deploying fire prevention.

Our Goal:

We need to rank fire risk of regions in a city at each timestamp

Challenges

To address the problem of ranking areas by fire risks, we face 2 main challenges：

1. Temporal

- Internal effects (e.g., historical fire risk).
- External effects
  - Immediate impact (e.g., temperature)
  - Delayed impact (e.g., electronic order)

2. Spatial

- Local spatial attributes (e.g., POIs, human activities, population of area)
- Global spatial dependencies

Architecture

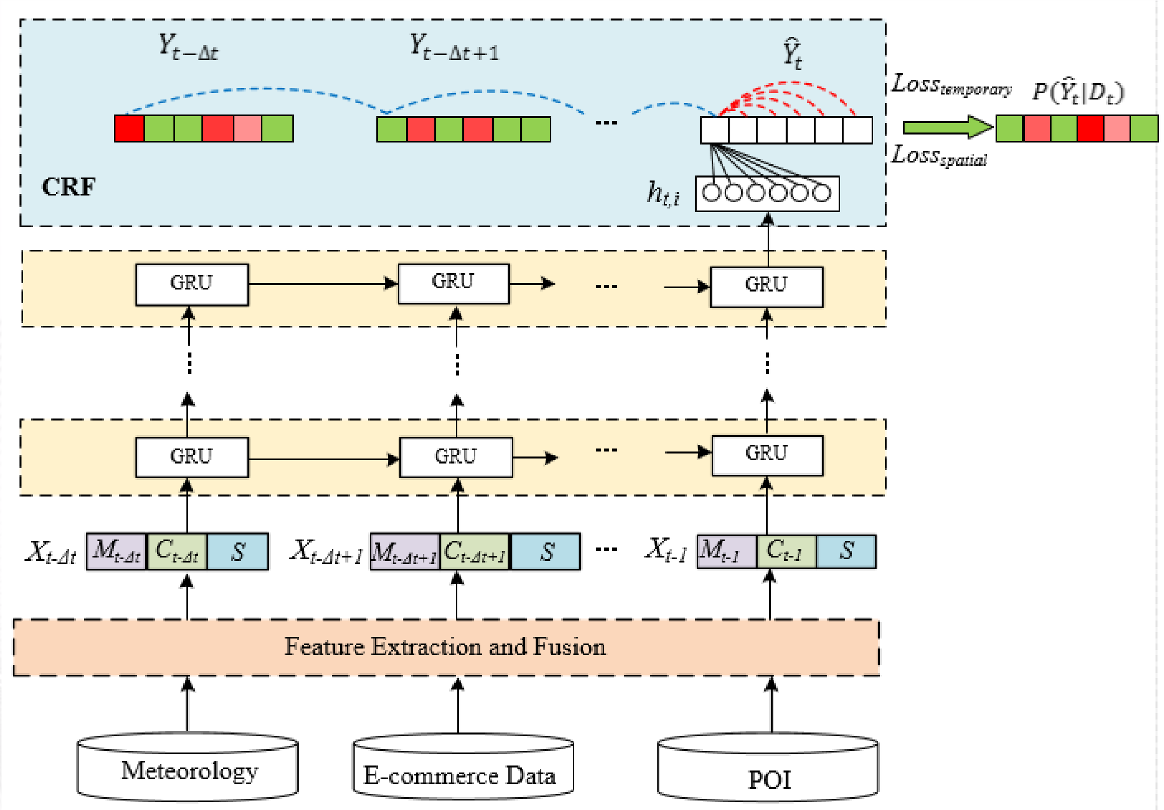
We propose a machine learning model named **NeuroFire** to integrate the temporal and spatial which consists of two-step:

- ☐ **Temporal fire classification**

  We use GRU-CRF to combine internal effects (historical risks) and external effects (weather,…)

- ☐ **Spatial fire risk forecasting**

  To lean the rank of regions at one timestamp, we use S-BPR to compare risks of regions



Method

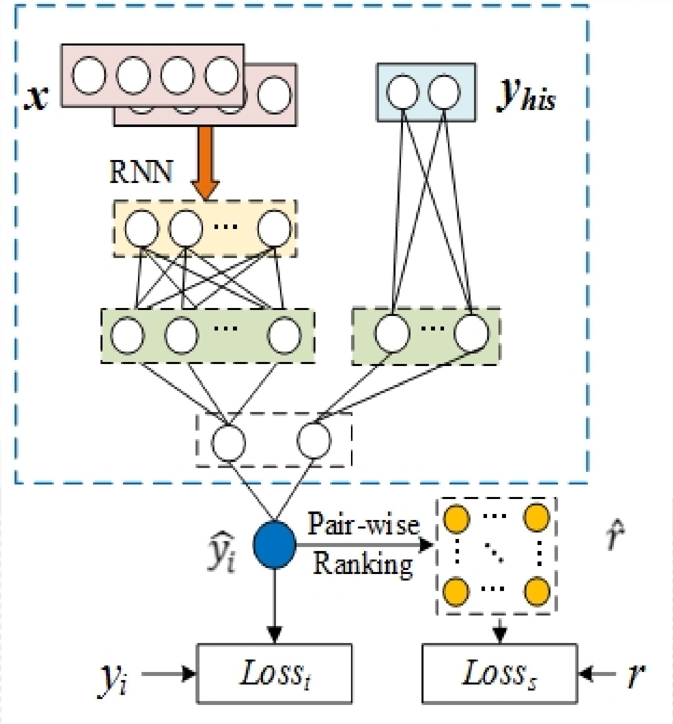
- Leverage RNN to learn temporal representations of urban data
- Integrate the temporal representations into fire risk sequences by CRF.

$$p(\mathbf{y}|\mathbf{X}) = \frac{1}{Z(\mathbf{X})} exp\{\underbrace{\alpha \sum f(\mathbf{y}, \mathbf{X})}_{Temporal} + \underbrace{\beta \sum g(y_i, y_j)}_{Spatial}\}$$

Temporal → RNN → $\hat{\mathbf{y}}$

Spatial → Pair-wise Rank → BPR

GRU-CRF



$y_{t,1}$	$y_{t,2}$	$y_{t,3}$	$y_{t,4}$	$y_{t,5}$
0	0	1	1	0

Y_t →

	$y_{t,1}$	$y_{t,2}$	$y_{t,3}$	$y_{t,4}$	$y_{t,5}$
$y_{t,1}$	0	0	0	0	0
$y_{t,2}$	0	0	0	0	0
$y_{t,3}$	1	1	0	0	1
$y_{t,4}$	1	1	0	0	1
$y_{t,5}$	0	0	0	0	0

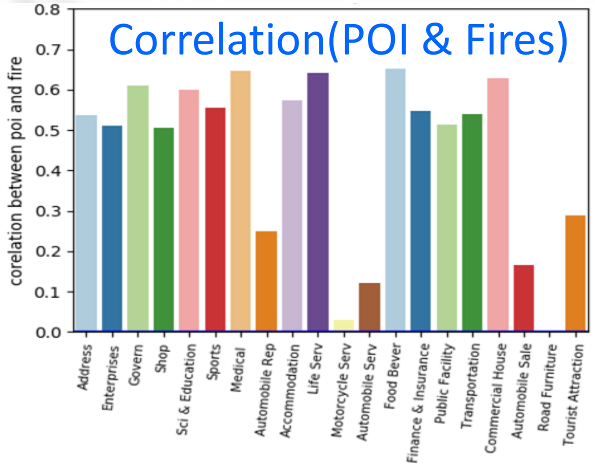
r_t

transform the metric

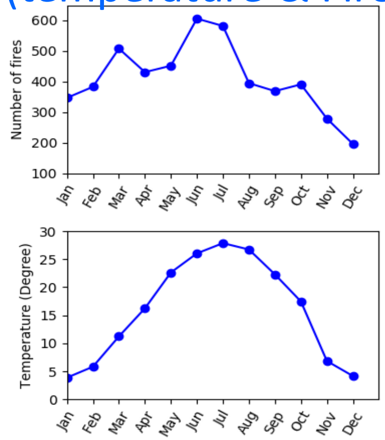
Experiments

Datasets

- *Fires dataset* is collected in a city, China during a 59-month period (Jan-2014~Nov-2018)：
    - Location (latitude, longitude)
    - Time
- *POI*
- *Weather*
    - Temperature
    - Humidity



Changings (temperature & Fires)



Metrics: Recall & F-1 score & AUC & MAE

Experiments

Results

Our model performs better than other 9 baselines:

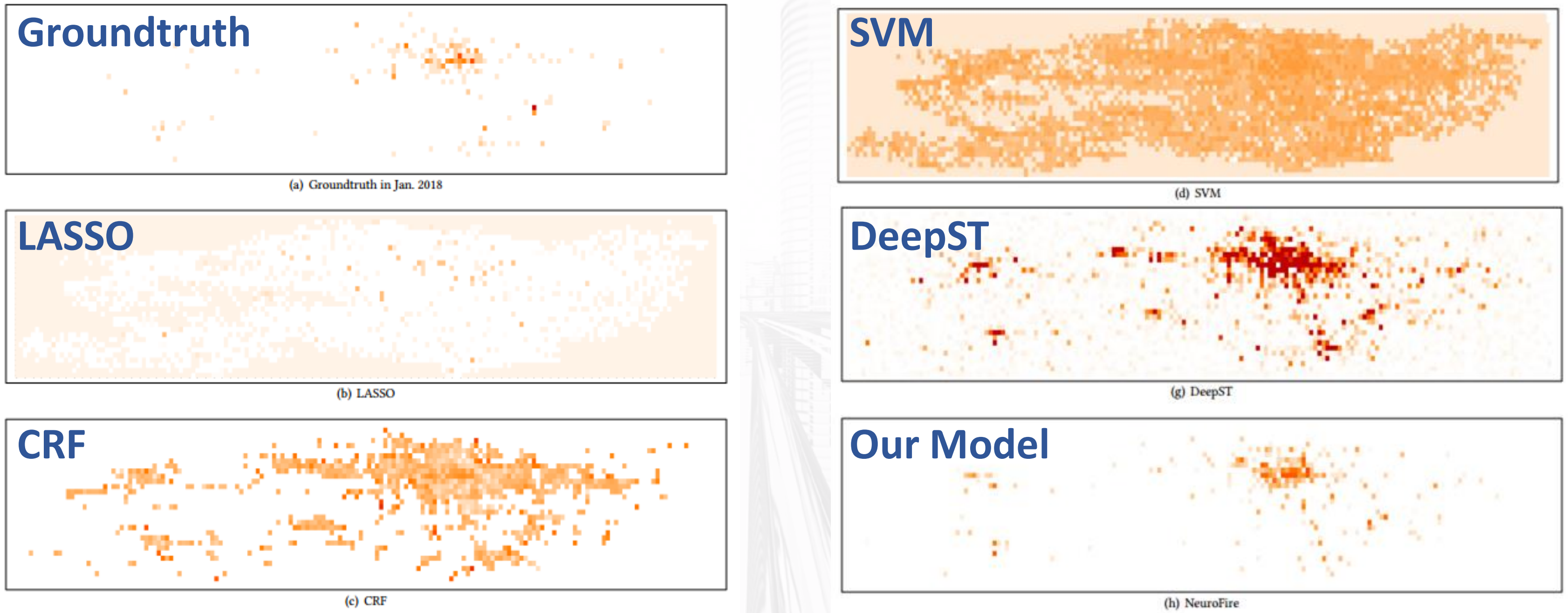
- **The effectiveness of temporal factors**: *Methods using GRU perform better than LASSO, CRF, LR and SVM.*
- **The effectiveness of historical sequence of fire risks**: *GRU-CRF performs better than GRU-LR*
- **The importance of spatial dependence**: *our model using S-BPR performs better than DeepST due to the sparsity of fires*

Model	Recall	F1-score	AUC	MAE
LASSO	0.202±0.306	0.025±0.014	0.454±0.075	0.411±0.265
CRF	0.263±0.075	0.349±0.071	0.629±0.037	0.807±0.005
LR	0.333±0.59	0.324±0.059	0.659±0.052	0.249± 0.036
LR-P	0.179 ±0.188	0.121±0.058	0.573±0.072	0.162±0.019
SVM	0.540±0.273	0.087±0.038	0.56±0.096	0.498±0.224
SVM-P	0.308±0.066	0.227±0.085	0.629±0.031	0.267 ±0.097
GRU-LR	0.342±0.101	0.365±0.102	0.666±0.076	0.146± 0.0139
GRU-CRF	0.284±0.111	0.387±0.071	0.673±0.06	0.100±0.003
DeepST	0.532±0.070	0.310±0.080	0.740±0.037	0.193±0.092
NeuroFire	**0.558±0.134**	**0.400±0.067**	**0.763±0.045**	**0.094±0.01**

Experiments

Visualization

- LASSO, CRF and SVM didn't perform well on the ranking problem.
- DeepST seems discriminating in spatial, it performs worse in forecasting.



The deeper color in a grid, the higher risk is forecasted.

Summary

1. Massive ST data and application scenarios → ST Deep Learning/AI
2. ST Neural Networks
   - Data transform + CNN/RNN/GNN/Attention
   - ST Point Data, ST Gridded Data, ST Networks(Graphs), ST Sequence Data
3. Advanced ST Neural Networks
   - ST-MetaNet, AutoST
4. More Spatio-Temporal AI Applications: Air, Fire, ...

References

- Ting Li, **Junbo Zhang**, et al. AutoST: Efficient Neural Architecture Search for Spatio-Temporal Prediction, KDD 2020
- Zheyi Pan, Yuxuan Liang, Weifeng Wang, Yong Yu, Yu Zheng, **Junbo Zhang**. Urban Traffic Prediction from Spatio-Temporal Data using Deep Meta Learning. KDD 2019
- Junkai Sun, **Junbo Zhang**, et al. Predicting Citywide Crowd Flows in Irregular Regions Using Multi-View Graph Convolutional Networks, ***IEEE TKDE***, 2020 (CCF A类)
- Zheyi Pan, Wentao Zhang, Yuxuan Liang, Weinan Zhang, Yong Yu, **Junbo Zhang**, Yu Zheng. Spatio-Temporal Meta Learning for Urban Traffic Prediction, ***IEEE TKDE***, 2020 (CCF A类)
- **Junbo Zhang**, et al. Flow Prediction in Spatio-Temporal Networks Based on Multitask Deep Learning. ***IEEE TKDE***, 2019. (CCF A类)
- Zheyi Pan, Zhaoyuan Wang, Weifeng Wang, Yong Yu, **Junbo Zhang**, Yu Zheng, Matrix Factorization for Spatio-Temporal Neural Networks with Applications to Urban Flow Prediction, CIKM 2019
- **Junbo Zhang**, et al. Predicting citywide crowd flows using deep spatio-temporal residual networks. ***Artificial Intelligence***, 2018. (CCF A类)
- **Junbo Zhang**, et al. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction. AAAI, 2017.
- **Junbo Zhang**, et al. DNN-Based Prediction Model for Spatial-Temporal Data. ACM SIGSPATIAL, 2016.

References

- Qianru Wang, **Junbo Zhang**, et al. CityGuard: Citywide Fire Risk Forecasting Using A Machine Learning Approach. **ACM UbiComp 2020**
- Yuxuan Liang, Kun Ouyang, Yiwei Wang, Ye Liu, **Junbo Zhang**, et al. Revisiting Convolutional Neural Networks for Citywide Crowd Flow Analytics. ECML-PKDD 2020
- Yuxuan Liang, Kun Ouyang, Lin Jing, Sijie Ruan, Ye Liu, **Junbo Zhang**, et al. UrbanFM: Inferring Fine-Grained Urban Flows. **KDD 2019**
- Xiuwen Yi, Zhewen Duan, Ting Li, **Junbo Zhang**, et al. CityTraffic: Modeling Citywide Traffic via Neural Memorization and Generalization Approach, **CIKM 2019**
- Xiuwen Yi, **Junbo Zhang**, et al. Deep Distributed Fusion Network for Air Quality Prediction, **KDD 2018**
- Chao Huang, **Junbo Zhang**, et al.. DeepCrime: Attentive Hierarchical Recurrent Networks for Crime Prediction. **CIKM 2018**.
- Yuxuan Liang, Songyu Ke, **Junbo Zhang**, et al. GeoMAN: Multi-level Attention Networks for Geo-sensory Time Series Prediction. **IJCAI 2018**.
- Dong Wang, **Junbo Zhang**, et al. When Will You Arrive Estimating Travel Time Based on Deep Neural Networks. **AAAI 2018**
- Xiuwen Yi, Yu Zheng, **Junbo Zhang**, et al. ST-MVL: Filling Missing Values in Geo-Sensory Time Series Data. **IJCAI 2016**

Call For Papers

https://dl.acm.org/journal/tist/cfp

ACM Transactions on Intelligent Systems and Technology Special Issue on "Deep Learning for Spatio-Temporal Data"

Topics of interest include but not limited to:

- Heterogeneous spatio-temporal data fusion with deep learning
- Deep spatio-temporal data representation learning
- rajectory data mining with deep learning
- Anomaly detection in spatio-temporal data with deep learning
- Deep learning based urban traffic prediction models
- Spatio-temporal crowdsourcing with deep learning
- Interpretable deep learning models for spatio-temporal data mining
- Novel deep learning models for mining noisy and sparse spatio-temporal data
- Deep learning models for novel applications based on spatio-temporal data
- Deep learning based spatio-temporal data mining for smart city
- Deep learning for spatio-temporal control and optimization
- Spatio-temporal data management with deep learning
- Spatio-temporal privacy and security with deep learning
- Spatio-temporal knowledge guided deep learning
- Spatio-temporal reasoning, uncertainty, and causality with deep learning

Tentative submission deadline of the Special Issue.

- Otc 30, 2020: Deadline for paper submissions

https://dl.acm.org/pb-assets/static_journal_pages/tist/cfps/tist-si-cfp-08-2020-spatio-temporal-data-1598567859617.pdf

THANKS!

Q & A

京东城市

京 东 数 科 旗 下

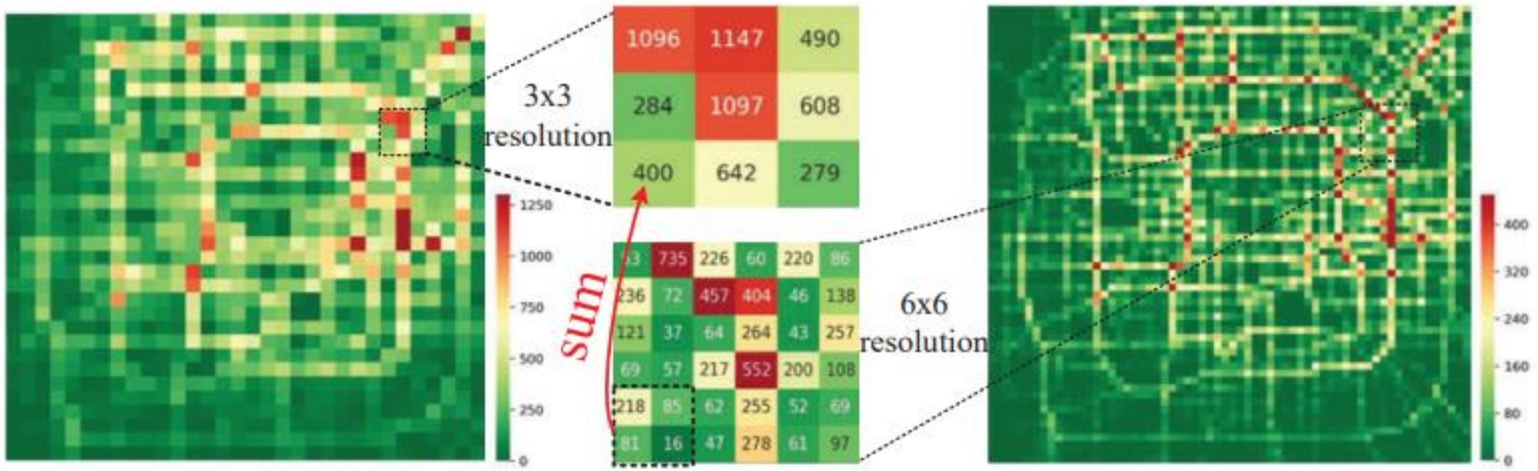
时空数据高分辨技术

时空数据也存在数据稀疏的问题，而细粒度的监控系统需要部署大规模的设备和传感器，意味着系统维护大量的资金支持。

一种基于深度神经网络的模型UrbanFM (Urban Flow Magnifier)，能够利用粗粒度城市人流量数据准确地还原细粒度人流量数据。

UrbanFM

粗粒度城市流量图 → 细粒度城市流量图



Yuxuan Liang, et.al. UrbanFM: Inferring Fine-Grained Urban Flows. KDD2019

Challenges

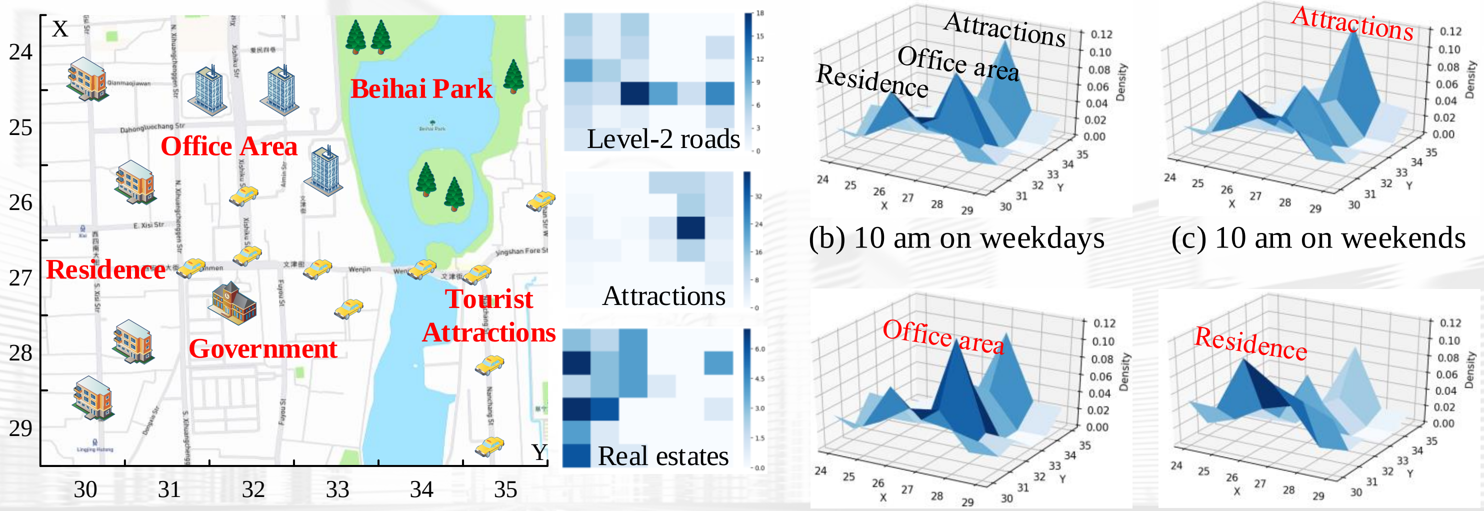
- Spatial correlations
  - Spatial hierarchy
  - Remote influence

(a) Coarse-grained crowd flows (32x32) (b) Fine-grained crowd flows (64x64)

Y. Liang, K. Ouyang, L. Jing, S. Ruan, Y. Liu, **J. Zhang**, et al. UrbanFM: Inferring Fine-Grained Urban Flows, KDD 2019

Challenges

- External factors
  - Meteorology
  - Time
  - Event



(a) A core area of Beijing and heatmaps of several geospatial attributes

(b) 10 am on weekdays

(c) 10 am on weekends

(d) 10 am on weekdays with thunderstorm

(e) 8 pm on weekdays

Framework

- Inference network
- External factor fusion