Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction

J.Zhang, Y. Zheng, D. Qi

David Albert

Korea Advanced Institute of Science and Technology

November 28, 2019

Main Goal

Predicting crowd flows in a city.

Why ?

- To be able to manage traffic.
- Lack of safety when massive crowds of people streamed into a strip region.
  - $\rightarrow$ Stampede at Shanghai New Year's celebration kills 36

Figure 1: Massive crowds of people in Paris during World Cup (2018)

Proposed method

Deep-learning-based approach to collectively forecast the inflow and outflow of crowds in each and every region of a city.

→ method called *ST-ResNet* (Spatio-Temporal Residual Networks)

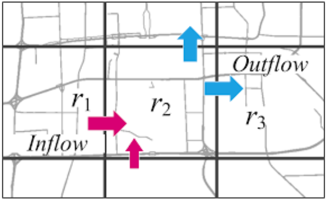


Figure 2: Inflow and outflow

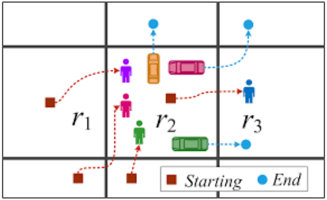


Figure 3: Measurement of flows

Introduction - Method

Proposed method

Deep-learning-based approach to collectively forecast the inflow and outflow of crowds in each and every region of a city.

$\rightarrow$ method called *ST-ResNet* (Spatio-Temporal Residual Networks)

- What is include ?
  - Number of pedestrian
  - Number of car
  - Number of people traveling on public transportation systems (e.g. metro, bus)

Introduction - Method

Proposed method

Deep-learning-based approach to collectively forecast the inflow and outflow of crowds in each and every region of a city.

$\rightarrow$ method called *ST-ResNet* (Spatio-Temporal Residual Networks)

- What is include ?
  - Number of pedestrian
  - Number of car
  - Number of people traveling on public transportation systems (e.g. metro, bus)
- How to get the necessary data ?
  - Mobile phone signals (for pedestrians)
  - GPS trajectories (for vehicles)
- Thus, we can obtain a set of trajectories $\mathbb{P}$ where each trajectory $Tr \in \mathbb{P}$, $Tr = (g_1, g_2, ..., g_k)$ *with* $g_k \in \mathbb{R}^2$ is a list of coordinates (longitude and latitude).

Introduction - Difficulties

1. Spatial dependencies

Inflow of a region is affected by outflows of nearby regions.

- After an event occured in a specific place, the inflow of nearby regions will increase.

2. Temporal dependencies

The flow of crowds in a region is affected by more or less recent time intervals.

- Traffic congestion at 8 AM will affect that of 9 AM.
- Traffic conditions during morning rush hours may be similar on consecutive workday.

3. External influence

Some external factors may change the flow of crowds.

- Weather, events, etc.

Partition the city

Many ways to partition a city. In this method, the city is partitioned into an $I \times J$ grid map based on longitude and latitude.

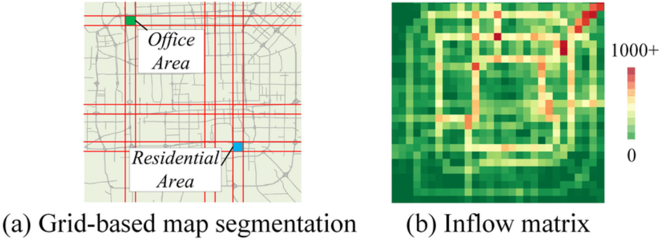


Figure 4: Regions in Beijing

Formulation of Crowd Flows Problem

Definition (Inflow & Outflow)

For a cell (i, j) that lies at the i^{th} row and j^{th} column of the grid, the inflow and outflow of the crowds at the time interval t are defined as:

$$x_t^{in,i,j} = \sum_{Tr \in \mathbb{P}} |\{k > 1 | g_{k-1} \notin (i,j) \wedge g_k \in (i,j)\}|$$

$$x_t^{out,i,j} = \sum_{Tr \in \mathbb{P}} |\{k \geq 1 | g_k \in (i,j) \wedge g_{k+1} \notin (i,j)\}|$$

where

- $\mathbb{P}$: a collection of trajectories at the t^{th} time interval
- $Tr : g_1 \rightarrow g_2 \rightarrow ... \rightarrow g_{|Tr|}$: a specific trajectory in $\mathbb{P}$
- g_k : the geospatial coordinate (i.e. longitude and latitude)
- $|.|$: the cardinal

Thus, at the t^{th} time interval, inflow and outflow in all the city can be denoted as a tensor $\mathbf{X}_t \in \mathbb{R}^{2 \times I \times J}$ where $(\mathbf{X}_t)_{0,i,j} = x_t^{in,i,j}$ and $(\mathbf{X}_t)_{1,i,j} = x_t^{out,i,j}$.

Formulation of Crowd Flows Problem

Thus, at the t^{th} time interval, inflow and outflow in all the city can be denoted as a tensor $\mathbf{X}_t \in \mathbb{R}^{2\times I\times J}$ where $(\mathbf{X}_t)_{0,i,j} = x_t^{in,i,j}$ and $(\mathbf{X}_t)_{1,i,j} = x_t^{out,i,j}$.

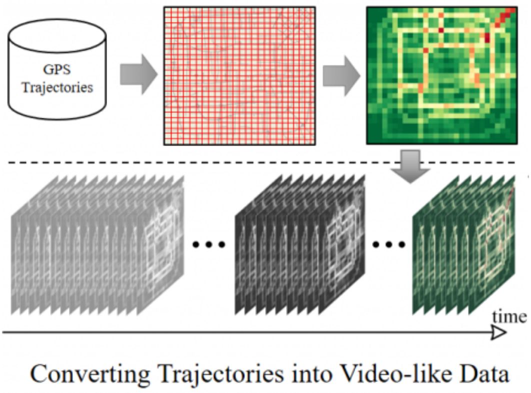


Figure 5: Processing data

Definition (Deep residual network)

Convolution neural networks whose structure is composed of residual units:

$$\mathbf{X}^{(l+1)} = \mathbf{X}^{(l)} + \mathcal{F}(\mathbf{X}^{(l)})$$

where $\mathbf{X}^{(l)}$ and $\mathbf{X}^{(l+1)}$ are the input and output af the l^{th} residual unit.

Such networks can be trained even with more than 100 layers.

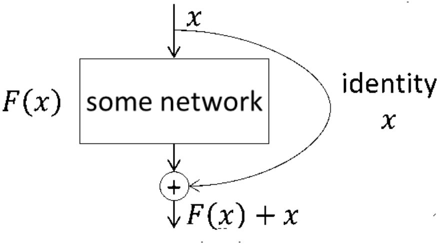


Figure 6: Residual Unit

ST-ResNet - Model Details

- Recall the objective:
  - Find a function f which predict X_n given $(X_{n-1}, X_{n-2}, ..., X_0) = H_t$
    - we will use neural networks to approximate f
    - learn $f_\theta : \mathbb{R}^{2n \times I \times J} \to \mathbb{R}^{2 \times I \times J}$

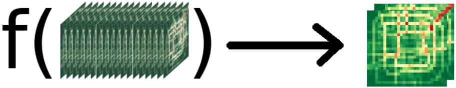


Figure 7: Objective

- Recall the difficulties :
  - temporal dependencies
  - spatial dependencies
  - external influences

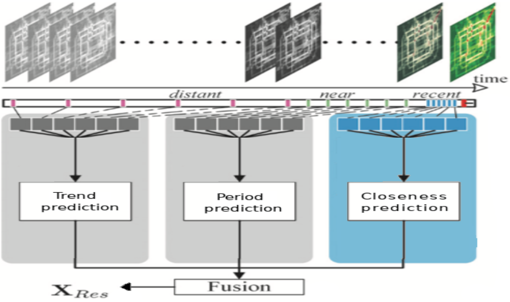
We will build our model f_θ to cope with these difficulties.

Dealing with the temporal dependencies difficulty

ST-ResNet summarize the temporal properties of crowds flows into three categories:

- Temporal **closeness** $\mathbf{X}_c$ $\rightarrow$ function of $\left[\mathbf{X}_{t-l_c}, \mathbf{X}_{t-(l_c-1)}, \ldots, \mathbf{X}_{t-1}\right]$
- Temporal **period** $\mathbf{X}_p$ $\rightarrow$ function of $\left[\mathbf{X}_{t-l_p.p}, \mathbf{X}_{t-(l_p-1).p}, \ldots, \mathbf{X}_{t-p}\right]$
- Temporal **trend** $\mathbf{X}_q$ $\rightarrow$ function of $\left[\mathbf{X}_{t-l_q.q}, \mathbf{X}_{t-(l_q-1).q}, \ldots, \mathbf{X}_{t-q}\right]$

Use three networks to model these properties and dynamically aggregates the three outputs (as shown on the figure).



Dealing with the spatial dependencies difficulty

ST-ResNet employs convolution-based residual networks to model nearby and distant spatial dependencies between any two regions of the city.

- Need very large citywide dependencies (each node in output depends on all nodes in input)
- For 32×32 grid and 3×3 kernel convolution $\rightarrow$ at least 15 layers
- Use residual networks to this end (with 26 layers in practice)

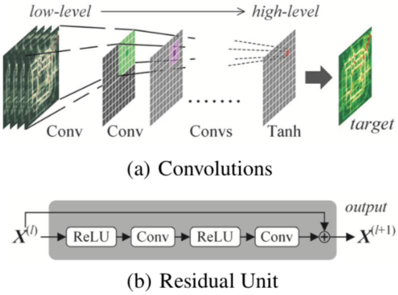


Figure 8: Convolutional and residual units

Dealing with the external influences difficulty

Adding a fourth component : the *external influence* component. This component is approximate by a simple two-fully-connected neural network where inputs are some features manually extracted from external datasets.

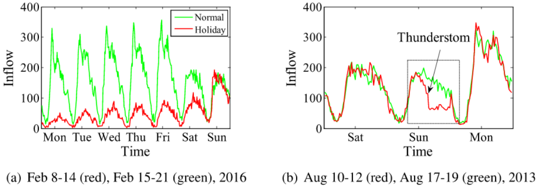


Figure 9: Effect of holidays and weather on Office Area of Beijing

Figure 10: External component

Global architecture

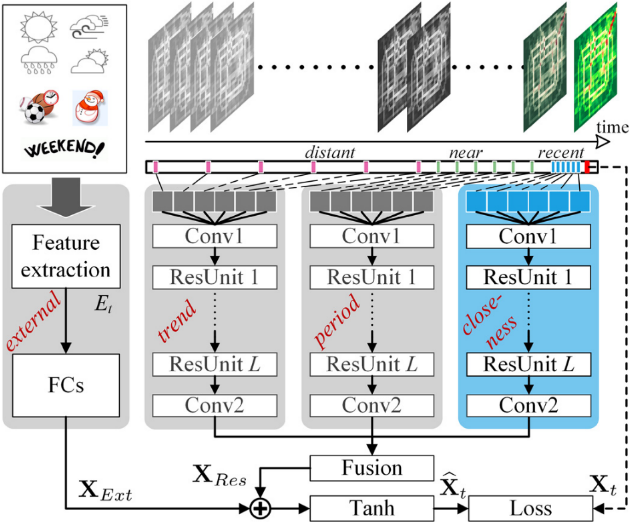


Figure 11: ST-ResNet Architecture

Experiment

Experiments are made on two differents datasets:

- **TaxiBJ** - Taxicab GPS trajectories
  - Time interval : $30min$
  - Grid map size : $(32, 32)$
  - Number of taxi: 34000+
  - Available time interval: 22459
  - External influence data : holidays, weather, temperature, wind speed
- **BikeNYC** - Bike trajectories
  - Time interval : $60min$
  - Grid map size : $(16, 8)$
  - Number of bikes: 6800+
  - Available time interval: 4392
  - External influence data : holidays

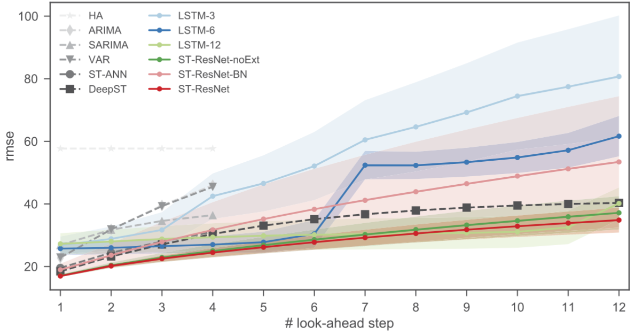


Figure 12: Multi-step ahead prediction on TaxiBJ dataset

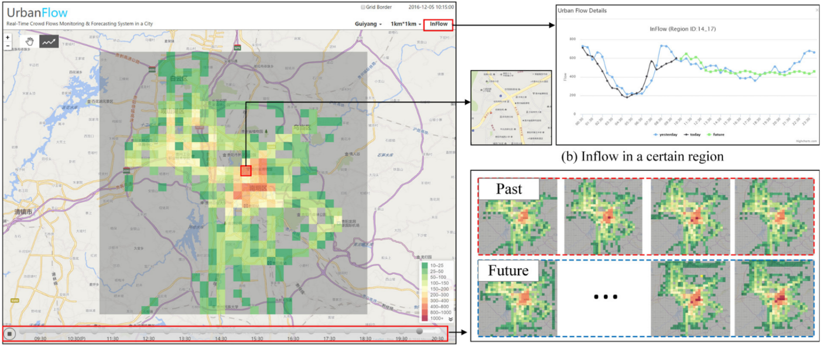
Experiment

Table 6
Comparisons with baselines on TaxiBJ and BikeNYC. Results of ARIMA, SARIMA, VAR and DeepST on BikeNYC are taken from [3]. For RNN-based models and ST-ResNet variants, we run each of them 10 times and show "mean ± standard deviation".

Model	RMSE	
	TaxiBY	BikeNYC
HA	57.69	21.578
ARIMA	22.78	10.07
SARIMA	26.88	10.56
VAR	22.88	9.92
ST-ANN	19.57	7.58
DeepST	18.18	7.43
RNN-3	26.68 ± 3.41	9.01 ± 0.50
RNN-6	30.03 ± 1.60	8.61 ± 0.42
RNN-12	45.51 ± 2.01	12.41 ± 0.84
RNN-24	51.12 ± 1.99	12.53 ± 0.14
RNN-48	43.42 ± 1.20	12.76 ± 0.18
RNN-336	39.61 ± 0.77	10.74 ± 0.35
LSTM-3	26.81 ± 2.80	8.67 ± 0.33
LSTM-6	26.07 ± 1.87	9.56 ± 0.64
LSTM-12	27.59 ± 3.69	9.90 ± 0.67
LSTM-24	25.69 ± 2.25	11.34 ± 0.44
LSTM-48	27.80 ± 2.87	12.36 ± 0.87
LSTM-336	40.68 ± 1.08	10.54 ± 0.13
GRU-3	22.97 ± 1.11	8.76 ± 0.37
GRU-6	23.64 ± 1.14	8.57 ± 0.22
GRU-12	27.40 ± 3.72	9.68 ± 0.51
GRU-24	27.01 ± 1.58	12.27 ± 0.77
GRU-48	28.56 ± 3.71	12.71 ± 0.93
GRU-336	40.27 ± 2.30	10.76 ± 0.33
ST-ResNet	**17.17** ± **0.21** (12 residual units)	6.32 ± 0.13 (4 residual units)
ST-ResNet-noExt	17.26 ± 0.38 (12 residual units)	\
ST-ResNet-AVG1	19.07 ± 0.30 (12 residual units)	6.44 ± 0.19 (4 residual units)
ST-ResNet-AVG2	18.04 ± 0.15 (12 residual units)	**5.99** ± **0.09** (4 residual units)

Figure 13: Results for TaxiBJ dataset (left) and BikeNYC dataset (right)

Visualization



(a) Web user interface of UrbanFlow

(b) Inflow in a certain region

(c) Heatmaps in past & future

Figure 14: Web interface of the UrbanFlow

Conclusion

- Forecast the flow of crowds collectively in each and every region of a city
- Performances beyond at least 6 baseline methods
- Future works :
  - Model to predict with irregular regions
  - Consider other types of flows

    $\rightarrow$ bus, phone signals data, metro card swiping data
- Since, other methods have been developped to improve citywide crowd flows prediction
  - MVGCC
  - ST-DCCNAL (overcome ST-ResNet)