U3 Singapore City-wide Crowd Flow Prediction using Deep Learning

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Abstract

A deep-neural-network-based approach is proposed as an end-to-end predictive model to forecast the flow of crowds in an urban environment. Fine-grained crowd flow prediction within a city is valuable for traffic control and management and could potentially improve travelling experience and public safety. However, getting accurate forecasts is challenging due to a mix of complex factors, such as the presence of convoluted spatial and temporal correlations, weather and special events. To address these challenges, a deep neural network is proposed which includes convolution neural networks coupled with residual units to model the complex spatio-temporal patterns. The aggregation of these patterns is further fused with an external component which models the exogenous factors such as weather and events, to predict the aggregated inflows and outflows in the next time frame for each region. An experiment is conducted in Singapore city to evaluate the performance of this model. Incorporating real-time news information from the Internet into the prediction model to extract crowd-influencing events is also explored.

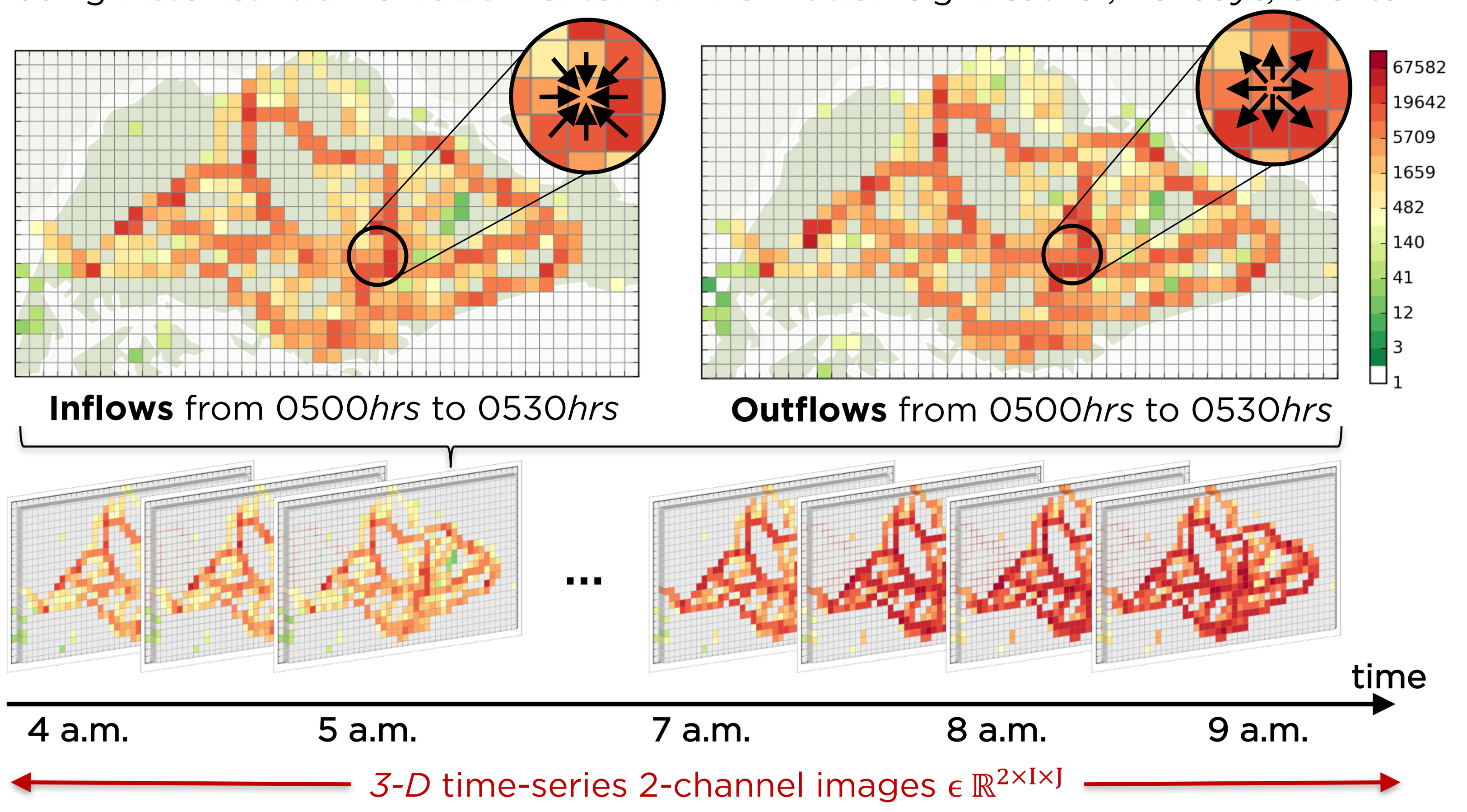
Overview

Objective:

Predict *future* **aggregated traffic flows** (**crowd flows**) that are:

i. in a multi-dimensional space subjected to temporal changes;
ii. aggregated spatially (**not** individual flows); represented by a grid map
iii. aggregated temporally (**not** at instantaneous time); in 30 mins / 1 hr intervals
iv. subjected to complex system behaviors; e.g. holidays, weather, events

using **historical traffic flows** + **external information** e.g. *weather, holidays, events.*

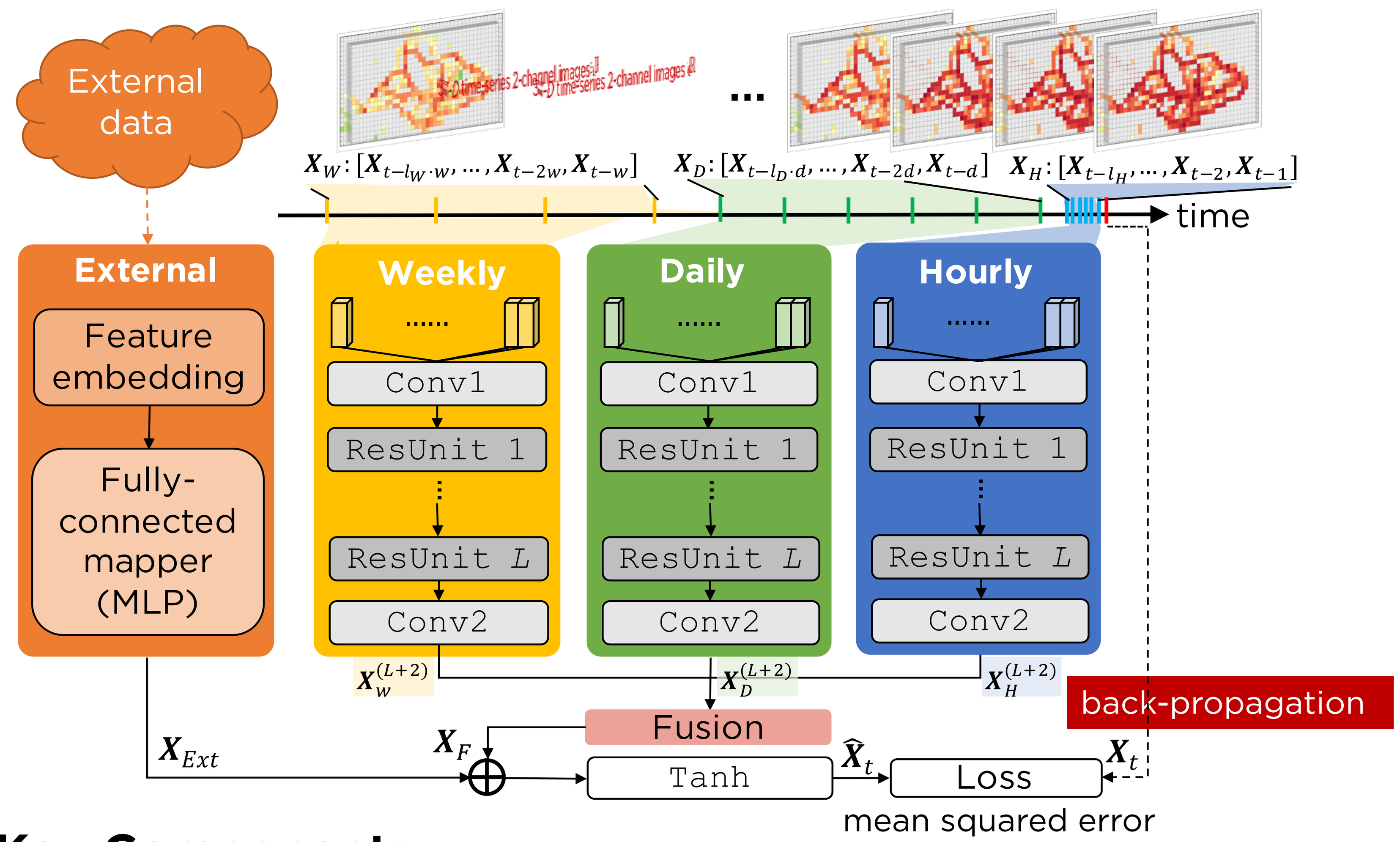


Motivations:

✓ Improve travel time estimation
✓ Optimize route selection
✓ Minimize traffic congestion
✓ Better urban city planning
✓ Facilitate traffic control measures
✓ **Ensure public safety**
✓ **Improve traveling experience**

Deep Learning Model

ST-ResNet Model Architecture:

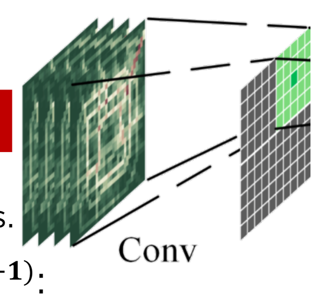


Key Components:

- `Conv` is a convolutional layer which maps input $X^{(i)}$ to output $X^{(i+1)}$:

$$X^{(i+1)} = f\left(W^{(i)} * X^{(i)} + b^{(i)}\right)$$ capture spatial correlations

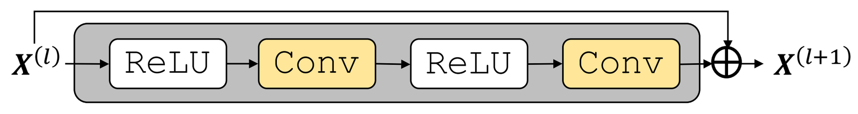
where f is the `ReLU` activation function $f(z) := \max(0, z)$;
$*$ denotes discrete convolution operation; and $\mathbf{W}^i$ and $\mathbf{b}^i$ are learnable parameters.



- `ResUnit` is a residual unit layer which maps input $X^{(l)}$ to output $X^{(l+1)}$:

$$X^{(l+1)} = X^{(l)} + \mathcal{F}\left(X^{(l)}; \theta^{(l)}\right), \qquad l = 1, \dots, L$$

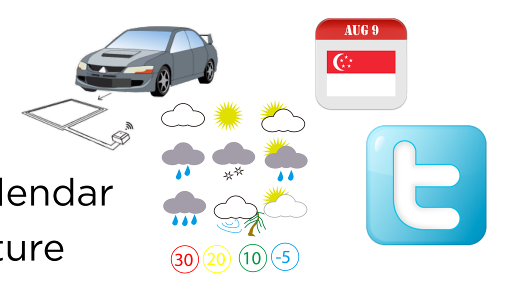
where $\mathcal{F}$ is a composite function *(see grey box →)*; and $\theta^{(l)}$ include all the learnable parameters.



minimize training degradation

Experiment Results & Discussion

Input Data:



1. **Traffic Flows:** Vehicular Loop Detectors
2. **Holidays:** Public Holidays + Weekends Calendar
3. **Weather:** State + Wind Speed + Temperature
4. **Crowd-influencing Events:** Twitter tweets
(keywords used: MRT stations' names, large places names, events names)

Experiment Setup:

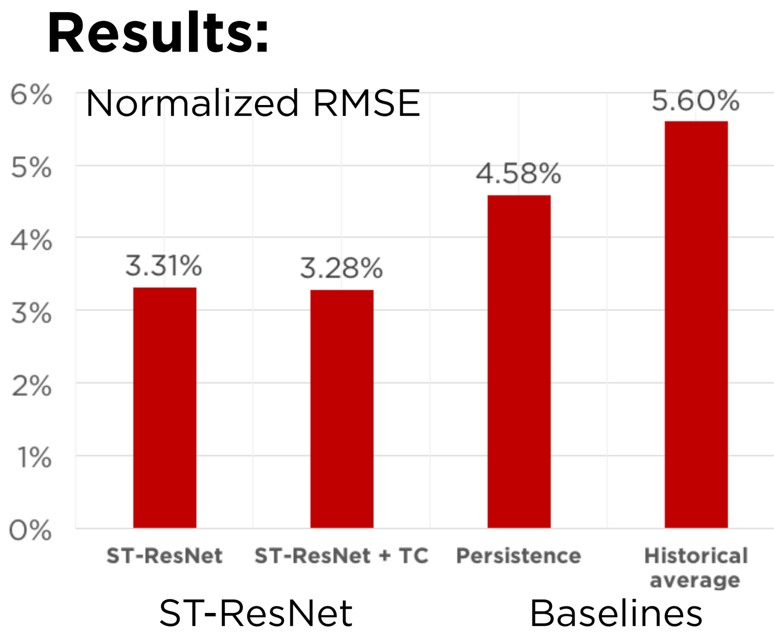
➢ **Validation datasets:**
i. 1st Mar 2013 – 30th Jun 2013
ii. 1st Sep 2014 – 31st Dec 2014
iii. 1st Jun 2015 – 30th Sep 2015
iv. 1st Dec 2015 – 31st Mar 2016

➢ **Datasets size**: 122 days (last 4 weeks test set)
➢ **Map grid size** (I × J): 89 x 49 (500m x 500m)
➢ **Time interval unit**: 30 minutes
➢ **Hyper-parameters** (l_H, l_D, l_W, L): (4,1,1,2)
➢ **Learning rate** (α): 0.002

Models & Baselines:

- **ST-ResNet:** standard model (holidays + weather)
- **ST-ResNet + TC:** add previous time frame's tweet counts as the **3rd input image channel** to the flow matrices
- **Historical Average:** use *moving average* of flows from preceding time frames and day of week as current prediction values
- **Persistence:** use the last time frame's flow values as current prediction values

Results:



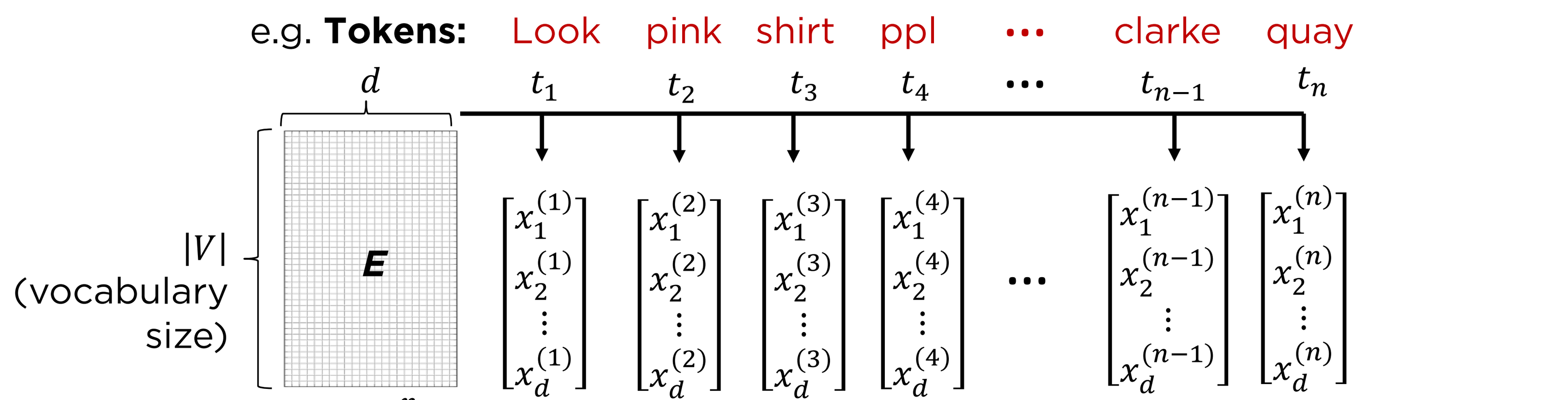
Discussion:

- ST-ResNet predictive model is effective as it **outperforms baseline** prediction models
- Usage of tweet counts have both **positive** and **negative** effects; largely due to ***presence of noise*** (e.g. spam and inaccurate spatio-temporal references)
- Potential to further process tweets' contents to ***detect hot/cold spots*** information across the city

Ongoing & Future Work

Shared Tweet Tokens Semantic Representation:

- Encode each tweet token (t_i) as a vector of size d **using an learnable embedding matrix** $(\boldsymbol{E})$ *initialized with pre-trained GloVE Twitter word vectors*
- Perform **softmax weighted sum** of all unique token vectors in each grid cell
- Append the d-sized summed vector $(\boldsymbol{S})$ as additional image channels to the flow matrixes



$$\boldsymbol{V} = \sum_{i=1}^{n} \boldsymbol{x}_i = \boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_3 + \boldsymbol{x}_4 + \dots + \boldsymbol{x}_{n-1} + \boldsymbol{x}_n$$

$$\boldsymbol{b} = [\boldsymbol{V}\cdot\boldsymbol{x}_1 \quad \boldsymbol{V}\cdot\boldsymbol{x}_2 \quad \boldsymbol{V}\cdot\boldsymbol{x}_3 \quad \boldsymbol{V}\cdot\boldsymbol{x}_4 \quad \dots \quad \boldsymbol{V}\cdot\boldsymbol{x}_{n-1} \quad \boldsymbol{V}\cdot\boldsymbol{x}_n]^T$$

$$\boldsymbol{c} = softmax(\boldsymbol{b}); \qquad \boldsymbol{S} = \sum_{i=1}^{n} \boldsymbol{c}_i \cdot \boldsymbol{x}_i$$

Future Work:

- **NLP Feature Extraction**: employ natural language processing methods to extract deeper meaning (e.g. +ve/-ve sentiment, past/future/present tenses)
- **Tweet Filtering**: Prior to processing of tweet, classify tweets as relevant or irrelevant to the right time and place
- **Event Detection**: Perform online clustering of tweets to detect event clusters

References:

- Goh, G., Koh, J. Y., Zhang, Y. (2018). Twitter-informed Crowd Flow Prediction. *Manuscript submitted for publication in ICKM 2018.*
- Zhang, J., Zheng, Y., & Qi, D. (2017). Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction. In *AAAI* (pp. 1655-1661).
- Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).

Gary Goh Shing Wee
SUTD PhD Programme (September 2016 Intake)

Information Systems Technology and Design
Supervisors: Assistant Prof. Zhang Yue & Prof. Patrick Jaillet