TRAFFIC FLOW PREDICTION USING DEEP SPATIO-TEMPORAL RESIDUAL NETWORKS

Arizona State University

ARPITH REDDY SINGAREDDY
MANISH AAKARAM
NISHANT WASHISTH
SHARAD SAXENA
ABHAY SRINIVAS SARASWATHULA

TABLE OF CONTENTS

INTRODUCTION

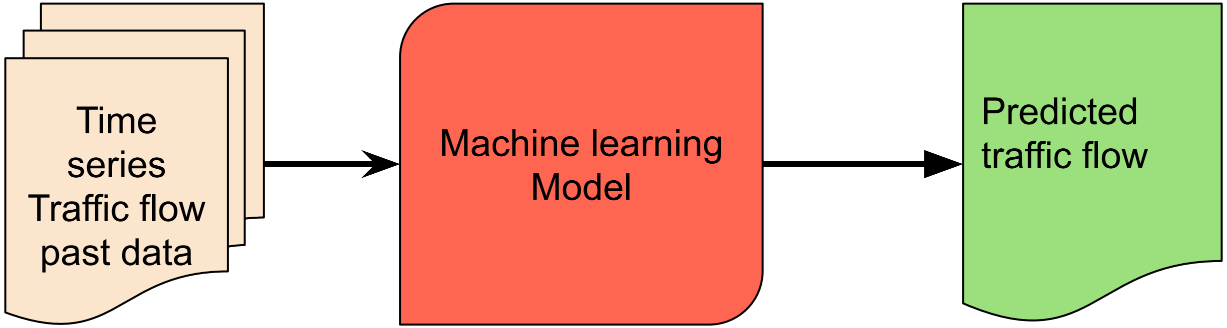
- Metropolitan cities like Los Angeles to Shanghai experience significant traffic flow as well as crowd flow.
- Public interest requires the management of traffic flow and crowd flow by predicting and forecasting them on basis of data collected such as climate and date.
- The problem is complicated since we have to deal with spatial dependencies, temporal dependencies, and stochastic events.

PROBLEM STATEMENT

- The goal is to develop a robust system that uses machine learning techniques to predict the traffic flow in a city for a specified time

- Given historical information of traffic flow of a city, we predict the inflow and outflow of traffic in the given region by capturing the spatial and temporal dependencies using end-to-end machine learning models.

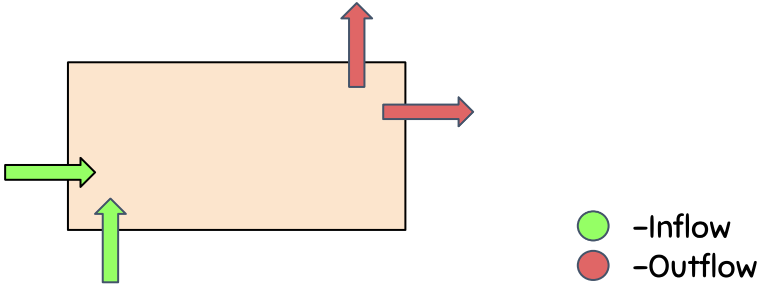
FLOW



Time series Traffic flow past data
Machine learning Model
Predicted traffic flow

TRAFFIC INFLOW AND OUTFLOW

- Inflow is defined as the number of vehicles or people entering a given region.
- Similarly Traffic outflow is the number of vehicles and people leaving a given region.



Spatial Dependencies

- Dependencies of traffic flow exist between different spaces of a given region. There is also high amount of traffic flow in areas such as schools or offices.

- If a person travels from region A to B. There is a spatial dependency between region A and B. There is an outflow of traffic from region A and inflow of traffic in region B.

- How do we capture this relationship...?

TEMPORAL DEPENDENCIES

- We are very familiar with the dependencies of Traffic flow on time. Common example can be the existence of rush hours.
- Traffic flow also depends on scenarios occurring in the past. Such as traffic congestion during a certain time will affect the traffic flow of the coming hours.

EXTERNAL FACTORS

- External factors such as rain, storm or snow affect the flow of traffic.

HOW TO ACCOUNT FOR ALL THESE DEPENDENCIES USING A MACHINE LEARNING MODEL?

ST-ResNet!

ST-ResNet

- An End-to-End Deep learning model to capture the spatial and temporal dependencies.
- Convolutional Neural networks to capture Spatial Dependencies
- Three residual networks to capture temporal dependencies in the series data

DATA AND PRE PROCESSING

- For input data, we have used the Taxi trip data from NYC government site that catalogues details such as trip pick up location, drop off location, dates and times for taxi cabs in NYC.
- A square region from New York city map was selected and is divided into 32x32 grid.
- The pickup point data was used to calculate the outflow map and drop off point data was used to calculate the inflow map.

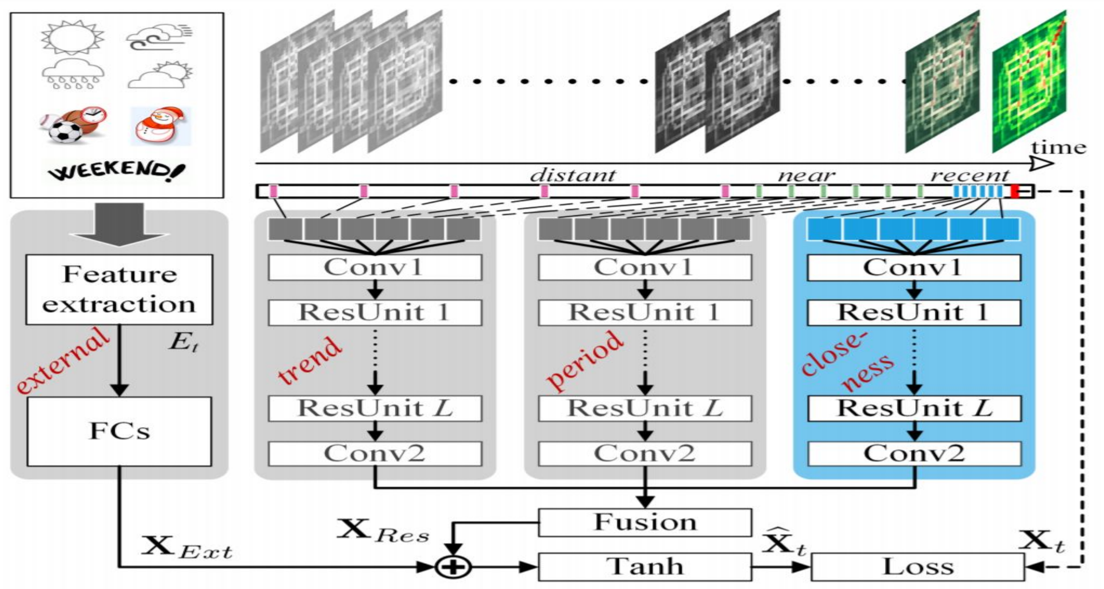
Data and Pre Processing

- Maps are generated for 30-minute time intervals. The pickup and drop off times are used to assign data points to these time intervals.
- Since we generate 2 maps (inflow and outflow) for every 30 minutes, the data can be considered as a 2 channel image.
- Therefore, convolutional neural networks are a good choice to identify the spatial dependencies in the data.

METHODOLOGY

- We use ST-ResNet to capture the spatial and temporal dependencies.
- ST-ResNet contains 3 components to capture the temporal properties of closeness, period and trend.
- The 3 components are implemented using standard Residual networks that contain a series of convolutional units with skip connections (that help with the problem of vanishing gradients).
- Batch Normalization is also done after each residual unit to make the network faster and more stable.

ST-ResNet Architecture

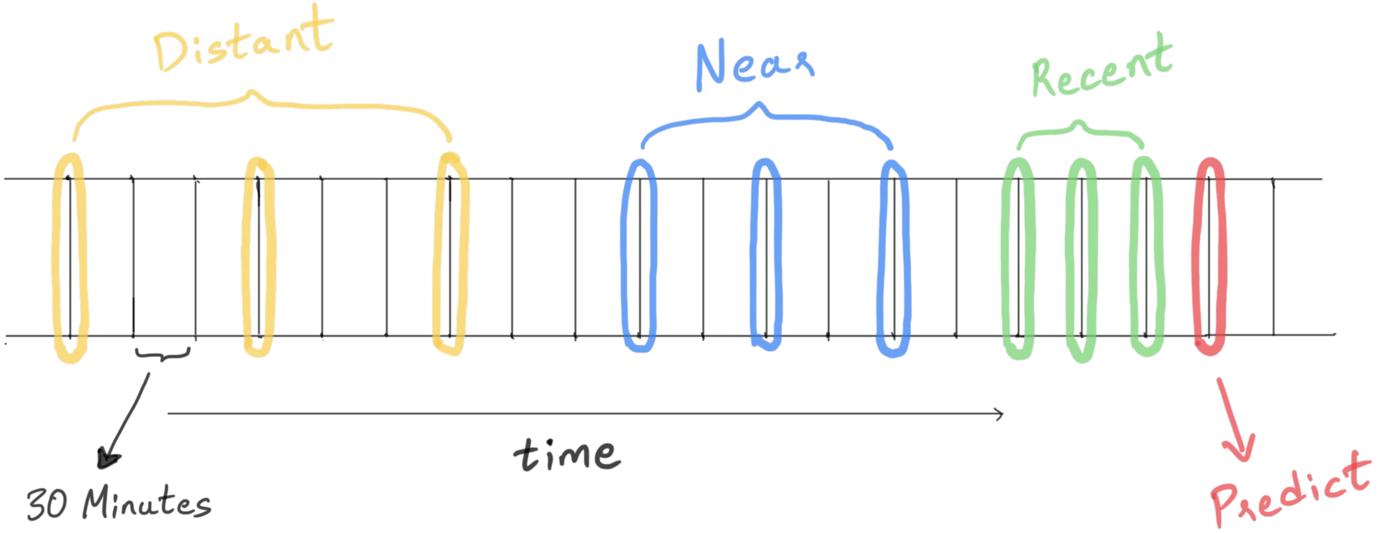


https://arxiv.org/pdf/1610.00081.pdf

METHODOLOGY

- The image like input data from preprocessing is used to select recent, near and distant time trajectories with a certain sequence length.
- This will generate a video like input data that is passed to each of the 3 components of the ST-ResNet.
- Each component will generate a 2 channel image after processing the video input through the residual network.
- The outputs from the 3 residual networks are fused together by addition layers with trainable weights to generate a single 2 channel image.

INPUT SEQUENCE AND PREDICTION

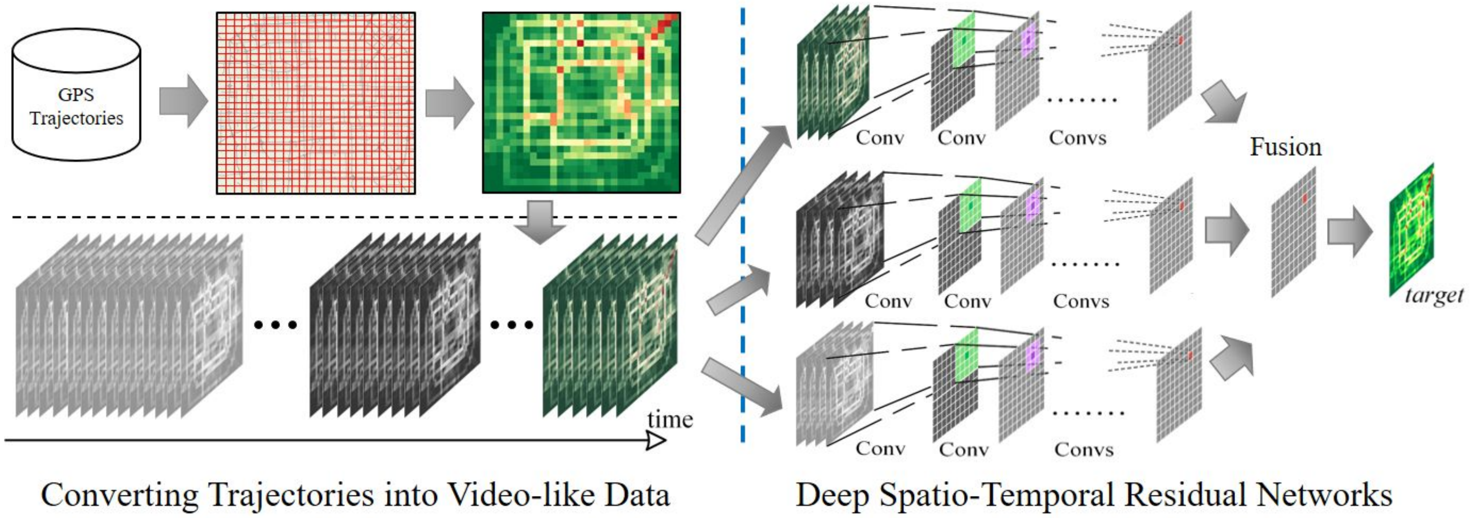


METHODOLOGY

- This fused output is then passed to the Tanh activation layer.
- The output from Tanh layer (range = [-1,1]) is rescaled back using min-max normalization values of training data.
- This output is 2 channel image representing the inflow and outflow prediction values for the given input representing the past.
- During training, loss is calculated using mean-squared error between predicted output and ground truth.

$$\mathcal{L}(\theta) = \|\mathbf{X}_t - \widehat{\mathbf{X}}_t\|_2^2$$

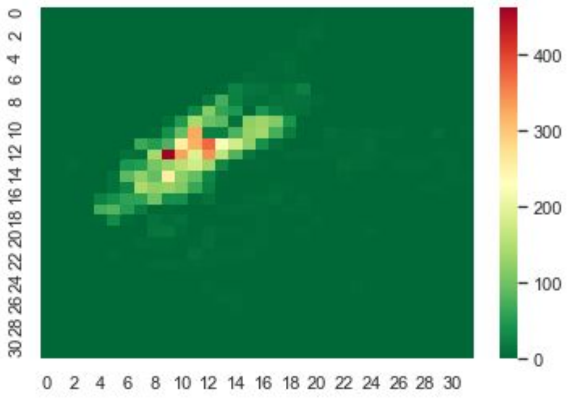
METHODOLOGY



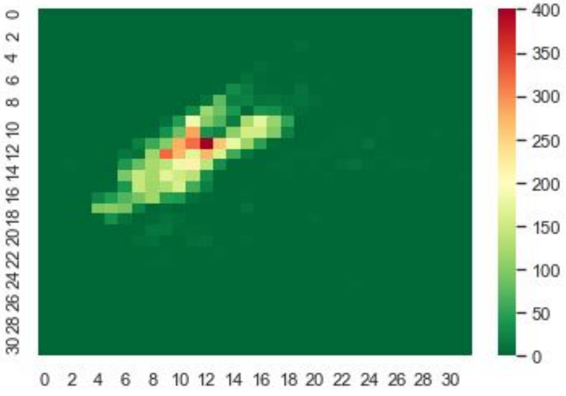
https://www.microsoft.com/en-us/research/wp-content/uploads/2016/11/AAAI2017_overview.png

RESULTS

- Testing accuracy of the model at this stage is 91% on the NY Taxi dataset.



Predicted Output



Ground Truth

FUTURE WORK

- Recurrent neural networks and LSTM networks can be explored to handle the temporal dependencies with better performance.

REFERENCES

- Zhang, Junbo, Yu Zheng, and Dekang Qi. "Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction." AAAI. 2017. https://arxiv.org/pdf/1610.00081.pdf
- Kaiming He and Xiangyu Zhang and Shaoqing Ren and Jian Sun. "Deep Residual Learning for Image Recognition". 2015 https://arxiv.org/abs/1512.03385
- NYC Taxi trip dataset https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page
- ST-Resnet Data flow image https://www.microsoft.com/en-us/research/wp-content/uploads/2016/11/AAAI2017_overview.png

THANKS!