Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

International Symposium of Transport Simulation (ISTS'18) and the International Workshop on Traffic Data Collection and its Standardization (IWTDCS'18)

Predicting traffic congestion maps using convolutional long short-term memory

Yan Li[a,b,*], Majid Sarvi[a,b], Kourosh Khoshelham[b], Milad Haghani[a,b]

[a]Transportation engineering group, University of Melbourne, 3052, Victoria, Australia

[b]Department of infrastructure engineering, University of Melbourne, 3052, Victoria, Australia

Abstract

Predicting the traffic congestion over a relatively short period of time is of great importance for traffic management and public safety. Previous methods require data from multiple sources and involve multiple parameters, such as traffic flow, volume, occupancy, speed and ramp flow for the analysis of congestion. Congestion prediction is still a challenge due to the variety of data sources required and the complex calculation of traffic parameters. Little research has examined the congestion prediction problem from the image processing perspective, which considers only images as inputs and outputs. The advantage of image-based congestion prediction is that it is direct and hence no intermediate parameters are needed. In this paper, we propose a deep-learning based approach, called convolutional long short-term memory (ConvLSTM), to predict congestion maps in a specific region, which is more informative and can provide not only the temporal changes of congestion level, but also the spatial distribution of the congestion. The test on freeway I-80 E in San Francisco shows that the traffic congestion can be accurately predicted with an accuracy up to 88%.

Keywords: congestion; short-term traffic prediction; deep learning; convolutional long short-term memory.

1. Introduction

Short-term traffic congestion prediction has been a very important topic as traffic congestion could cause increased travel time and extra harm to the environment. It has been estimated that the costs caused by congestion will in-crease approximately 50% by 2020 (Bureau of Transport and Regional economics (2015)). As the sensors technology development and computational advances enable researchers to collect real-time data and predict at very high temporal resolutions, prediction of traffic congestion has been a vital component of intelligent transportation system (ITS).

* Corresponding author. Tel.: +61-422188658.
E-mail address: li17@student.unimelb.edu.au

Over the past few decades, many research efforts have been made in the area of traffic congestion prediction (Vlahogianni et al. (2004, 2014); Suhas et al. (2017)). The traffic congestion prediction problem, viewed as a type of time series prediction problem, concerns the estimation of the future values of congestion level based on current and past data records. Previous methods can be divided into linear models and nonlinear models (Vlahogianni et al. (2004)). The Auto-Regressive Integrated Moving Average (ARIMA) model assumes that the future value of a variable is a linear combination of previous values and the errors. Furthermore, the model requires stationary time series, where individual values do not fluctuate around the average values with obvious upward and downward trends. The nonstationary time series should be differentiated before the analysis (Box et al. (2015)). Nonlinear models such as neural networks are also frequently used in time-series prediction. Common neural networks are feedforward networks and are often called the sliding window technique as the N-tuple input slides over the full training set to generate a single output as the target of the network (Connor et al. (1994); Frank et al. (2001)). Current traffic data collection techniques include loop detectors, automatic vehicle identification(AVI) system, and simulation. In general, traffic parameters, such as traffic flow, volume, occupancy, speed, and ramp flow are calculated to capture and predict traffic conditions. However, data collection from multiple sources and effective fusion of the data to perform congestion analysis is still a challenge due to the variety of data source and the complex calculation of traffic parameters.

In recent years deep learning (LeCun et al. (2015)) has shown great performance in visual recognition tasks, such as image classification (Krizhevsky et al. (2012)), pedestrian detection and tracking in images (Dollar et al. (2012)). Deep learning is a machine learning method, which is good at extracting inherent features in data from the lowest level to the highest level by using multiple-layer architectures or deep architectures. Furthermore many traffic image database are available online. For example Google map and Here map, which can show real-time traffic congestion on the map, provide comprehensive APIs for customers to access traffic images. The availability of such image sources and the remarkable performance of deep learning methods motivates us to approach the traffic prediction problem based on image processing techniques.

In this paper, we propose an innovative spatio-temporal deep learning method for congestion analysis from an image processing perspective. The proposed method requires only historical congestion maps to predict future congestion maps. This research provides insights to understanding the spatial and temporal patterns from traffic images which could greatly simplify the prediction procedures. The contributions of this research are three-fold:

- firstly, to predict short-term congestion using only traffic images, without requiring other traffic parameters.
- secondly, to introduce a novel spatio-temporal deep learning method to accurately predict traffic congestion up to 15 minutes in the future.
- finally, to convert the images to GIS data for a wider application of congestion maps in congestion analysis.

The rest of this paper is organized as follows. Section 2 reviews current short-term prediction method and deep learning methods related to this research. Section 3 presents our approach of congestion prediction. Section 4 discusses our experiments in San Fransisco . Concluding remarks and discussions are described in Section 5.

2. Related work

2.1. Spatio-temporal traffic prediction with deep learning

The patterns in traffic prediction can be divided into three categories (Zhang et al. (2017)): 1) Spatial relations: the traffic performance, for example traffic flow, is affected by the nearby regions, more than the remote areas. 2) Temporal relations: traffic performance is also related to near time and far in the future. Traffic flow will change gradually in continuous time and also shows periodic patterns. The morning and afternoon peak hours may appear similar patterns on weekdays. 3) External features: Some external factors, such as weather conditions and road works will affect the traffic performance.

The general procedures of spatio-temporal traffic prediction using deep learning is to first extract the patterns, then relate the features to the traffic parameters we intend to predict. Deep learning methods have been applied to extract different features depending on the characteristics of the neural network. Typically convolutional neural network (CNN) is used to extract spatial features and recurrent neural network (RNN) is used to extract temporal features.

Convolutional neural networks use convolutional operators to learn image features and can preserve the spatial relations between pixels, achieving remarkable performance on image recognition (Krizhevsky et al. (2012)). Different structures of CNN have been applied, for example standard CNN (Wu and Tan (2016); Ke et al. (2017); Zhang et al. (2017)), graph CNN (Yu et al. (2017)). Recurrent neural networks are good at handling temporal tasks because they recurrently calculate each element of time series data and memorize the captured information. Gated residual unit (GRU) (Li et al. (2018b)) and long short-term memory (LSTM) (Yu et al. (2017); Wu and Tan (2016)) are two special categories of RNN, which both can learn long-term temporal dependencies. While GRU has two gates (update and reset) and LSTM has three gates (forget, input and output), LSTM has more parameters than GRU and show better performance on big volume of data.

There are different ways of relating the features extracted to the traffic parameter we intend to predict. Some works concatenate the features first and then build a regression model to predict, for example linear regression (Wu and Tan (2016)), logistic regression (Lv et al. (2015)), support vector machine (Li et al. (2018a)) and random forest (Ke et al. (2017)). Other works feed the spatial features directly to recurrent neural network (RNN) and a fully-connected layer is added to generate the final prediction result. ConvLSTM (Xingjian et al. (2015)) is a promising structure that can learn spatial and temporal dependencies simultaneously.

2.2. *Video sequence prediction*

Video sequence prediction can be divided into two tasks: high-level semantic prediction and pixel-level predictionVillegas et al. (2017). High level semantics refer to action, event and motion. Forecasting semantics describes what will happen and provides basis for decision-making. The training for semantic prediction tasks requires a large amount of training data, which involves a tremendous manual labeling work. In contrast, pixel-level prediction (also called frame prediction) provides a dense and direct description of the world. Spatio-temporal correlations in videos provide a self-supervision for frame prediction, which enables purely unsupervised training of a model by observing raw video frames.

3. Methodology

Figure 1 shows the primary steps of our approach. First, we collect the traffic maps every 5 minutes covering the afternoon peak hours for three years based on Google Map API. Second data preprocessing is required to reduce the computational cost of the subsequent processes. The procedures include removing background and gridding the images. Third, we feed the data to the ConvLSTM network to capture spatial and temporal relations between the map images and generate the predicted frames. Then a map is reconstructed by merging the grids and overlaying the basemap. Finally, we vectorize the maps to refine the prediction result and also for further congestion analysis.

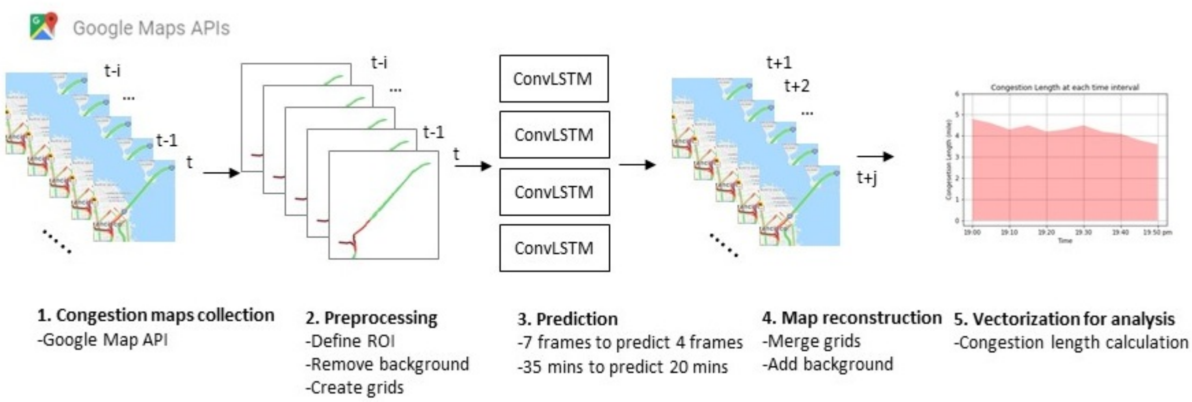

Fig. 1. Prediction procedures

3.1. *Collection of traffic congestion maps*

The dataset of congestion maps used in this paper is collected by Google Maps API. The Google Maps JavaScript API provides access to real-time traffic information using the TrafficLayer object, which can be overlaid on a base map. Traffic congestion is divided into four levels, from slow to fast. Traffic information is refreshed frequently so through rapid consecutive requests for the same area we can retrieve the congestion maps at predefined time intervals. A program based on Javascript was developed to automatically run the capture of Google maps with traffic layers every 5 minutes. The congestion maps were saved using the capture times as the file name so that later analysis can recognize the time sequences.

3.2. *Preprocessing the maps*

Since the aim of this research is to predict the congestion of the roads, other land areas are not of interests. Therefore we extracted the regions of interest by comparing the base maps with congestion maps and finding the difference of these two images, which is our target road area. Another reason for this step is to save the computing resources by excluding the useless pixels from the calculation. As millions of parameters are calculated in deep learning, in order to save computing resources, the maps are dividing into small grids. Based on our experiment, 40*40 pixels grids are the most appropriate image size to be feed into the network.

3.3. *Prediction by ConvLSTM*

We use the convLSTM network, which extends the fully connected LSTM layer to have convolutional structures in both the input-to-state and state-to-state transitions, so that it can learn spatial and temporal relations at the same time (Xingjian et al. (2015)). The convLSTM network applied in this research is composed of four ConvLSTM2D layers and one Conv3D layer, as illustrated in Figure 2. The structure is implemented in Keras with Tensorflow backend (Chollet et al. (2015)). The ConvLSTM2D layer is similar to an LSTM layer, but the input transformations and recurrent transformations are both convolutional. The Batchnormalization layers normalize the values of the previous layer at each batch and apply a transformation that maintains the mean activation close to 0 and the activation standard deviation close to 1. The last Conv3D layer creates a 3D convolution kernel convolved with the layer input over the temporal dimension to produce time series outputs.

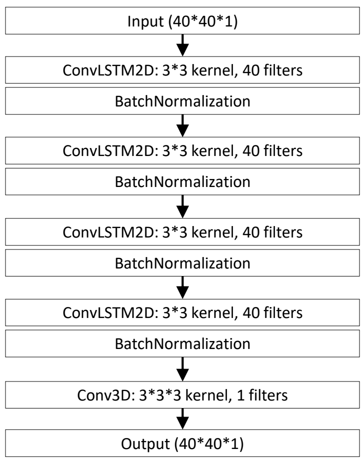

Fig. 2. ConvLSTM structure

3.4. Reconstructing congestion maps

Reconstructing congestion maps is composed of two steps: merging the grids and overlaying basemaps. As mentioned above, the maps are divided into small grids, so we need to merge multiple image grids to a whole map. After obtaining the future congestion maps over the regions of interest, we overlay these maps on the base map, which we used in the second step, so that we can generate the final complete congestion maps.

3.5. Vectorizing maps for congestion analysis

GIS data with accurate geographic locations and attributes can provide a rich source of information for traffic congestion studies such as travel times and delays, congestion levels, and energy emissions (Taylor et al. (2000)). In order to conduct further analysis, we convert the raster congestion maps to GIS data by retrieving the congestion level of each road segment and assigning the values to the attribute of GIS road objects. Another benefit of this procedure is to deal with the inherently blurry problem of video prediction (Mathieu et al. (2015)). After converting to GIS data, we assign one congestion value to each road segment, which can guarantee that each road segments can have only one congestion level thus improving the accuracy of pixel prediction.

4. Experiments

4.1. Test area

An experiment was conducted to evaluate the performance of the method in predicting congestion maps over a study area, as shown in Figure 3 . We chose freeway I80-E in San Francisco County, from Highway 101 to the Treasure Island portion of the Bay Bridge, which is the most congested segment in San Francisco during afternoon peak hours on weekdays. We record congestion maps from 14 October 2014 to 4 Feb 2018 along freeway from 19:00 - 20:00 pm on weekdays covering 3 years. We collected 12 frames of size 240*240 pixels per day and a total of 10,392 frames in 866 days, excluding the days that were missing data.

Table 1 shows the basic information of Freeway I80 eastbound. The total length of this road segment is 8.69 kilometers. This road segment is a 5-lane road, which is 17.98 meters width and each lane is 3.60 meters width.

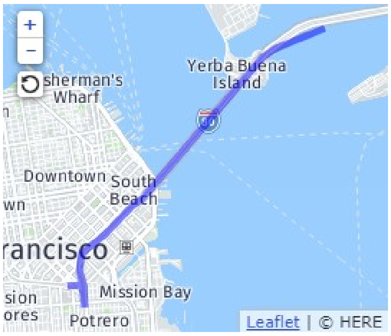

Fig. 3. Freeway I80-E in San Francisco County as test area, which is the most congested segment in San Francisco

Table 1. Road segment information

Name	Freeway I80-E
Length	8.69 km
Road Width	17.98 m
Lane Width	3.60 m
No. of Lanes	5

4.2. Evaluation

We measure the accuracy of the prediction by calculating the percentage of correctly predicted pixels in the road region. As Figure 4 and Table 2 shows, the accuracy of prediction is above 80% in the first 5 minutes but it declines over time. As there is no standard to define a satisfactory level of accuracy, in this research we assume that a prediction accuracy of 80% is acceptable, which means that in this experiment the prediction accuracy is acceptable within 15 minutes (3 frames). And the accuracy can also be highly enhanced by vectorizing the maps, which we can observe from Table 2 that the accuracy can be improved up to 15% .

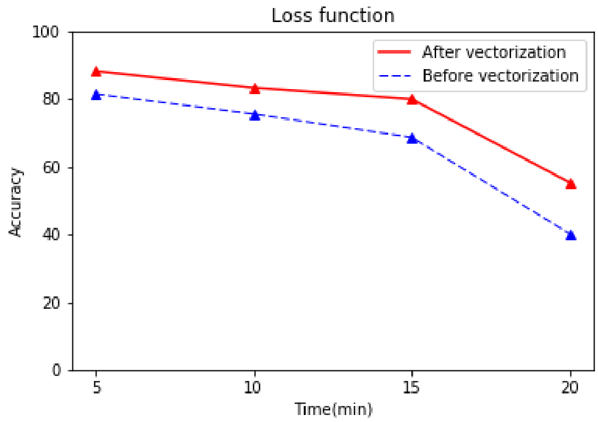

Fig. 4. Accuracy of each predicted frame before and after refinement

Table 2. The accuracy at different time intervals

Accuracy	5 mins	10 mins	15 mins	20 mins
Before vectorization	81.4	75.5	68.6	40.2
After vectorization	88.1	83.2	79.9	55.3

4.3. Congestion length calculation

After converting the raster maps to georeferenced maps, we conduct congestion analysis based on the GIS data. The length of traffic congestion is a typical variable showing the present state of congestion, which is defined as the total length of road segments with congestion level above a certain threshold, as shown in Figure 5 in red and brown color. We filter the roads with high congestion level (red and brown colors) and obtain the length attribute of the filtered road, and finally summarize the length value to get the congestion length of the whole road segment. As it can be observed in Figure 5 that the congestion length decreases after 19:00 pm but decreases after 19:20 with 19:30 pm being the most congested time after 19:00 pm.

5. Conclusion

In this paper, we applied deep learning method to the challenging problem of short-term congestion prediction, which so far has not been studied from an image analysis perspective based on open source congestion maps. We formulate congestion prediction as a spatio-temporal sequence forecasting problem and introduce convLSTM structure to tackle the problem. The advantage of our approach is that unlike the previous methods that involve the laborious work of fusing many traffic parameters, our method requires only historical congestion maps for congestion analysis. Based on our experiment on the most congested road segment of San Francisco, this framework can predict 15 minutes ahead with an acceptable accuracy. For the future work, we will investigate how to apply convLSTM to crowd

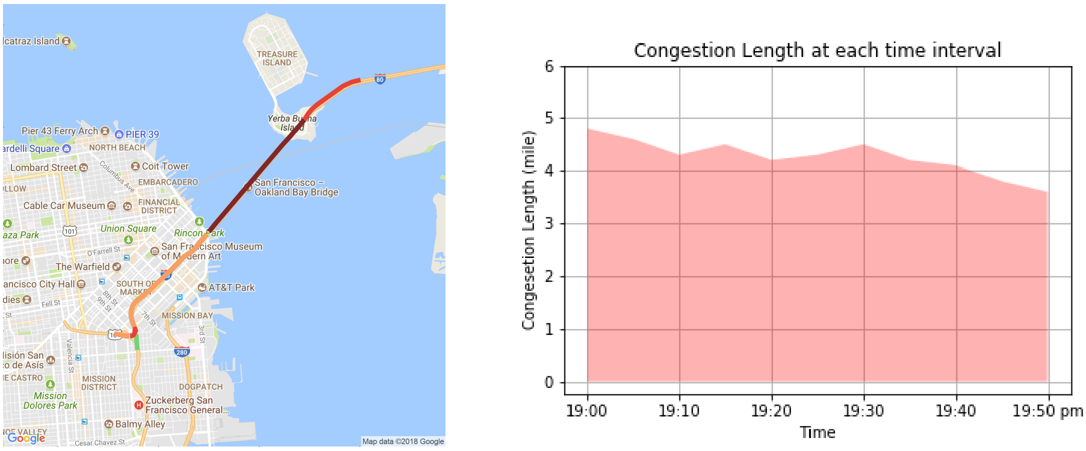

Fig. 5. (a) Congestion length of one road, the total of red and brown parts; (b) Congestion length by time.

congestion prediction in transportation hubs, which is quite similar to traffic congestion prediction problem that has spatio-temporal characteristics.

Acknowledgements

The first author also acknowledges the financial support of China Scholarship Council (CSC).

References

Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2015. Time series analysis: forecasting and control. John Wiley & Sons.

Chollet, F., et al., 2015. Keras. https://github.com/fchollet/keras.

Connor, J.T., Martin, R.D., Atlas, L.E., 1994. Recurrent neural networks and robust time series prediction. IEEE transactions on neural networks 5, 240–254.

Dollar, P., Wojek, C., Schiele, B., Perona, P., 2012. Pedestrian detection: An evaluation of the state of the art. IEEE transactions on pattern analysis and machine intelligence 34, 743–761.

Frank, R.J., Davey, N., Hunt, S.P., 2001. Time series prediction and neural networks. Journal of Intelligent & Robotic Systems 31, 91–103.

Ke, J., Zheng, H., Yang, H., Chen, X.M., 2017. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. Transportation Research Part C: Emerging Technologies 85, 591–608.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, pp. 1097–1105.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436.

Li, Y., Sarvi, M., Khoshelham, K., Haghani, M., 2018a. Real-Time Level-of-Service Maps Generation from CCTV Videos. Technical Report.

Li, Y., Yu, R., Shahabi, C., Liu, Y., 2018b. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting, in: International Conference on Learning Representations.

Lv, Y., Duan, Y., Kang, W., Li, Z., Wang, F.Y., 2015. Traffic flow prediction with big data: a deep learning approach. IEEE Transactions on Intelligent Transportation Systems 16, 865–873.

Mathieu, M., Couprie, C., LeCun, Y., 2015. Deep multi-scale video prediction beyond mean square error. arXiv preprint arXiv:1511.05440 .

Suhas, S., Kalyan, V.V., Katti, M., Prakash, B.A., Naveena, C., 2017. A comprehensive review on traffic prediction for intelligent transport system, in: Recent Advances in Electronics and Communication Technology (ICRAECT), 2017 International Conference on, IEEE. pp. 138–143.

Taylor, M.A., Woolley, J.E., Zito, R., 2000. Integration of the global positioning system and geographical information systems for traffic congestion studies. Transportation Research Part C: Emerging Technologies 8, 257–285.

Bureau of Transport and Regional Economics., 2015. Estimating urban traffic and congestion cost trends for Australian cities - Working Paper No 71. Canberra, Australia: Bureau of Transport and Regional Economics.

Villegas, R., Yang, J., Hong, S., Lin, X., Lee, H., 2017. Decomposing motion and content for natural video sequence prediction. International Conference on Learning Representations 1, 7.

Vlahogianni, E.I., Golias, J.C., Karlaftis, M.G., 2004. Short-term traffic forecasting: Overview of objectives and methods. Transport reviews 24, 533–557.

Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2014. Short-term traffic forecasting: Where we are and where we are going. Transportation Research Part C: Emerging Technologies 43, 3–19.

Wu, Y., Tan, H., 2016. Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework. arXiv preprint arXiv:1612.01022 .

Xingjian, S., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.c., 2015. Convolutional lstm network: A machine learning approach for precipitation nowcasting, in: Advances in neural information processing systems, pp. 802–810.

Yu, B., Yin, H., Zhu, Z., 2017. Spatio-temporal graph convolutional neural network: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875 .

Zhang, J., Zheng, Y., Qi, D., 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), pp. 1655–1661.