

Cellular Traffic Prediction Using Deep Convolutional Neural Network with Attention Mechanism

Zihuan Wang and Vincent W.S. Wong

Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, Canada
email: {zihuanwang, vincentw}@ece.ubc.ca

Abstract—Predictive analysis on cellular traffic is important for the control and monitoring of wireless networks. Cellular traffic prediction is a challenging problem due to the non-stationarity and dynamic spatial-temporal correlation of the traffic. In this paper, we address the problem of accurate traffic prediction in a base station by proposing a deep neural network called RACConv. Its structure includes residual network, attention mechanism, and deep convolutional network. In the proposed architecture, a deep 3D residual convolutional network (ResConv3D) with three residual blocks are employed to learn the local spatial-temporal features. An attention-aided convolutional long short-term memory network (AConvLSTM) is then used to capture the long-term spatial-temporal dependencies. The use of the attention modules enable the network to focus on the most important spatial-temporal information. We evaluate the performance of the proposed RACConv network using a dataset provided by a Canadian wireless service provider. We consider the traffic prediction on two time scales (i.e., hourly and daily), which exhibit different spatial-temporal dependency patterns. Experimental results show that the proposed RACConv network can achieve accurate prediction under both time scales. Results also show that our proposed network provides a lower root-mean-square error (RMSE) than the conventional ConvLSTM baseline scheme.

I. INTRODUCTION

The fifth generation (5G) wireless systems support different use cases with diverse quality of service (QoS) requirement. It is important for the wireless service providers to allocate and utilize the network resources efficiently. Both resource allocation and energy management in wireless communication systems require accurate traffic analysis and prediction. By proactively estimating the future traffic load, wireless service providers can dynamically allocate network resources and improve the spectral and energy efficiencies.

It is challenging to predict cellular traffic at fine granularity due to the time-varying and load-dependent traffic dynamics [1]. In particular, the traffic demand depends on many factors such as time of the day (e.g., rush hour vs. off-peak hour, weekday vs. weekend), special events, and public holidays. The cellular traffic load is also location dependent. The load at different base stations can vary significantly due to user behavior and QoS requirement from different use cases. Moreover, user mobility introduces spatial dependencies in the traffic data between neighboring base stations. The aforementioned factors have collective effect among base stations and complicate the spatial-temporal dependency of cellular traffic.

Various types of data-driven based traffic prediction approaches have recently been proposed in the literature. The works in [2] and [3] use long short-term memory (LSTM) networks for traffic load prediction, where LSTM is able to capture the long-term temporal dependency. The convolutional neural networks (CNNs) are proposed in [4] to analyze the spatial-temporal features for traffic prediction. A hybrid deep learning model is proposed in [5], which includes a stacked autoencoder structure for spatial feature extraction and an LSTM for temporal feature modeling. The convolutional LSTM (ConvLSTM) network has been proposed in [6] to capture spatial and temporal relationships. In [7], a spatial-temporal neural network is proposed for mobile traffic forecasting. The proposed architecture consists of a ConvLSTM and a 3D convolutional (Conv3D) neural network to encode the spatial-temporal features, followed by fully-connected layers to predict future cellular traffic. In [8], generative adversarial network (GAN) is utilized for traffic prediction, where CNN and LSTM are embedded into the GAN to analyze the spatial-temporal correlations. In [9], a transformer network is proposed where two transformer blocks are constructed for spatial and temporal features extraction. In [10] and [11], the cross-domain data, e.g., information from point of interests distribution, are utilized to improve the prediction accuracy. In [12] and [13], side information, such as weather condition, is adopted to enhance traffic prediction performance.

Most of the aforementioned works study the grid-based traffic data, where the coverage area is divided into multiple grids. Although the grid-based approach can simplify the analysis, it may not accurately capture the spatial dependencies of the traffic across all the base stations. In order to predict the dynamic traffic pattern in a base station, it is crucial to capture both the local short-term and long-term spatial-temporal characteristics. In this paper, we propose a Residual network and Attention mechanism embedded deep Convolutional (RACConv) network to efficiently learn the local and long-term spatial-temporal dependencies. The main contributions of this paper are summarized as follows:

- By using a dataset collected from a wireless service provider in Canada, we analyze the spatial-temporal dependency under different time scales (i.e., on hourly and daily basis). To capture the spatial pattern similarity

of cellular traffic in different areas, a spectral clustering algorithm is proposed to assign the base stations into different groups.

- We propose an RAConv network for accurate traffic prediction. The first part of the RAConv network is a deep 3D residual convolutional (ResConv3D) network with three residual blocks to learn the local spatial-temporal features. Then, an attention mechanism [14] aided ConvLSTM network (AConvLSTM) is proposed to learn the long-term spatial-temporal dependencies of the traffic demands. The use of the attention modules enable the network to focus on the most important spatial-temporal information.
- We evaluate the proposed traffic prediction algorithm. Hourly and daily traffic predictions are considered. Experimental results illustrate that the proposed RAConv network can achieve accurate predictions under both time scales. Moreover, results show that the proposed RAConv network has a lower root-mean-square error (RMSE) than the ConvLSTM baseline scheme.

The rest of this paper is organized as follows. Section II describes the dataset and presents an analysis on its spatial and temporal dependencies. Section III presents the proposed RAConv network for accurate traffic prediction. Section IV presents the performance comparison between the proposed network and a baseline scheme. Conclusions are drawn in Section V.

II. DATASET DESCRIPTION AND ANALYSIS

In this section, we introduce the cellular traffic dataset considered for evaluation in this work. The spatial and temporal dependencies of the traffic demands under different time scales are illustrated.

A. Traffic Dataset Description

The cellular traffic data analyzed in this paper was collected by a wireless service provider in Canada, Rogers Communications Inc., from January 1 to June 6, 2020. The traffic data was collected at each base station every 15 minutes. The data was collected from 4,096 base stations in major cities (e.g., Mississauga, Toronto, Ottawa, Montreal, Quebec City) located in the provinces of Ontario and Quebec in Canada, approximately covering an area of 12,000 km², with an average coverage area of 3 km² of each base station. To describe the spatial correlation among base stations, we use a 2D tuple to identify each base station. With a total of 4,096 base stations, they can be presented by 64 × 64 2D tuples. Locations of the base stations are shown in the map given by Fig. 1. In this paper, we consider the traffic prediction time scale to be either on an hourly or daily basis. The collected data was aggregated either in each hour or each day accordingly.

Given a prediction time scale s , where $s \in \{\text{hourly, daily}\}$, the traffic data matrix at the t -th time interval in an area with

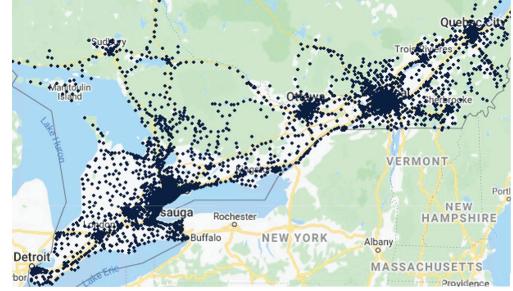


Fig. 1. Some of the locations of the base stations in the provinces of Ontario and Quebec in Canada.

$M \times N$ base stations (where $M = N = 64$ in our study) can be expressed as $\hat{\mathbf{D}}_{s,t} \in \mathbb{R}^{M \times N}$:

$$\hat{\mathbf{D}}_{s,t} = \begin{bmatrix} \hat{d}_{s,t}^{(1,1)} & \hat{d}_{s,t}^{(1,2)} & \dots & \hat{d}_{s,t}^{(1,N)} \\ \vdots & \vdots & & \vdots \\ \hat{d}_{s,t}^{(M,1)} & \hat{d}_{s,t}^{(M,2)} & \dots & \hat{d}_{s,t}^{(M,N)} \end{bmatrix},$$

where $\hat{d}_{s,t}^{(m,n)}$, $m \in \{1, \dots, M\}$, $n \in \{1, \dots, N\}$, denotes the traffic of base station (m, n) at the t -th time interval under the prediction time scale s . Note that the current wireless cellular networks do not have a regular grid-based topology, the tuples (m, n) , $m \in \{1, \dots, M\}$, $n \in \{1, \dots, N\}$, are not the coordinates of the base stations. They only reflect the spatial relationship among base stations (e.g., base stations $(1, 1)$, $(1, 2)$, $(2, 1)$, and $(2, 2)$ are neighboring base stations in practical systems, but the physical topology may not form a grid). The traffic data can be described as a spatial-temporal sequence $\hat{\mathbf{D}}_s = \{\hat{\mathbf{D}}_{s,t} \mid t = 1, 2, \dots, T_s\}$, where T_s is the total number of time intervals (e.g., hours or days) under time scale s . For each base station (m, n) , we normalize the traffic data to be within the range of $[0, 1]$ by using max-min normalization:

$$\mathbf{d}_s^{(m,n)} = \frac{\hat{\mathbf{d}}_s^{(m,n)} - \min(\hat{\mathbf{d}}_s^{(m,n)})}{\max(\hat{\mathbf{d}}_s^{(m,n)}) - \min(\hat{\mathbf{d}}_s^{(m,n)})}, \quad (1)$$

where $\hat{\mathbf{d}}_s^{(m,n)} = [\hat{d}_{s,1}^{(m,n)} \dots \hat{d}_{s,T_s}^{(m,n)}]$, $\mathbf{d}_s^{(m,n)} = [d_{s,1}^{(m,n)} \dots d_{s,T_s}^{(m,n)}]$, $s \in \{\text{hourly, daily}\}$.

B. Data Analysis

In this subsection, we explore the traffic data dependencies in both temporal and spatial domains. Our analysis is based on the number of user requests at each base station. Both the hourly and daily traffic patterns are evaluated.

1) *Temporal dependency analysis*: The sample autocorrelation function [15], as a function of time lag l , is widely used for temporal dependency evaluation. The autocorrelation function for a sequence of data $d_{s,t}^{(m,n)}$, $t = 1, \dots, T_s$, at base station (m, n) under time scale s is given by:

$$r_s^{(m,n)}(l) = \frac{\sum_{t=1}^{T_s-l} (d_{s,t}^{(m,n)} - \bar{d}_s^{(m,n)})(d_{s,t+l}^{(m,n)} - \bar{d}_s^{(m,n)})}{\sum_{t=1}^{T_s} (d_{s,t}^{(m,n)} - \bar{d}_s^{(m,n)})^2}, \quad 0 \leq l < T_s, \quad (2)$$

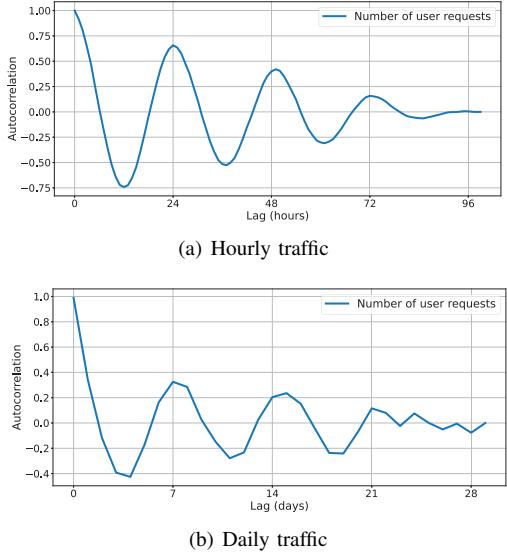


Fig. 2. Temporal autocorrelation of (a) hourly traffic and (b) daily traffic.

where $\bar{d}_s^{(m,n)}$ represents the mean value of the traffic data at base station (m, n) under time scale s . The autocorrelation value is between -1 and 1 .

As an example, we select a base station located in the city of Montreal, Canada, to show the temporal dependency. Fig. 2 illustrates the temporal autocorrelation of the hourly and daily traffic. For hourly traffic in Fig. 2(a), we can observe that the number of user requests has non-zero autocorrelation in the temporal domain. The peak points appear regularly when the time lag is equal to an integer multiple of 24 hours. For daily traffic in Fig. 2(b), it also exhibits non-zero autocorrelation, but with peak points appeared when the time lag is an integer multiple of 7 days. It can be observed that future traffic can be predicted based on previous observations. In order to understand the traffic under different time scales, it is critical to capture the autocorrelation patterns.

2) *Spatial Dependency Analysis:* Given a pair of base stations (m, n) and (m', n') , we use the Pearson correlation coefficient [16] to model the spatial dependency of the traffic data. The Pearson correlation coefficient for the traffic data under time scale s is defined as

$$\rho_s = \frac{\text{cov}(\mathbf{d}_s^{(m,n)}, \mathbf{d}_s^{(m',n')})}{\sigma_{\mathbf{d}_s^{(m,n)}} \sigma_{\mathbf{d}_s^{(m',n')}},} \quad (3)$$

where $\text{cov}(\cdot)$ represents the covariance operation, and σ is the standard deviation. Note that the Pearson correlation coefficient is within the range of $[-1, 1]$. Table I shows the spatial correlation of the number of user requests among a cluster of 8 neighboring base stations in Montreal, Canada. The upper triangular part shows the spatial correlation of daily traffic and the lower triangular part shows the spatial correlation of hourly traffic. Results in Table I show that there exist non-zero correlations in a cluster of neighboring base stations. We can also observe the differences of the spatial correlation pattern between hourly and daily traffic.

TABLE I
SPATIAL CORRELATION OF A RANDOMLY SELECTED CLUSTER OF 8 BASE STATIONS. (UPPER TRIANGULAR PART: DAILY TRAFFIC; LOWER TRIANGULAR PART: HOURLY TRAFFIC)

	1	2	3	4	5	6	7	8
1	1.0	0.53	0.55	0.63	0.64	0.48	0.54	0.36
2	0.51	1.0	0.47	0.41	0.46	0.39	0.59	0.42
3	0.49	0.54	1.0	0.49	0.55	0.48	0.64	0.5
4	0.47	0.46	0.46	1.0	0.65	0.46	0.59	0.5
5	0.49	0.53	0.52	0.46	1.0	0.47	0.59	0.52
6	0.44	0.49	0.5	0.39	0.54	1.0	0.48	0.44
7	0.49	0.54	0.52	0.43	0.54	0.57	1.0	0.49
8	0.56	0.52	0.5	0.52	0.44	0.49	0.5	1.0

Results in Fig. 2 and Table I show that the traffic demands are highly spatial-temporal correlated. Moreover, the correlation varies significantly under different time scales. In the next section, we present a deep convolutional network called RAConv. The proposed RAConv network can accurately capture the local short-term and long-term spatial-temporal dependencies of the traffic demands.

III. RAConv NETWORK FOR TRAFFIC PREDICTION

Consider a sequence of traffic data over a spatial region with $M \times N$ base stations. Given a time scale s , the goal of a traffic forecasting task is to predict the most likely Q -step sequence of data based on the previous P observations which include the current interval:

$$\begin{aligned} & \mathbf{D}_{s,t+1}^*, \dots, \mathbf{D}_{s,t+Q}^* \\ &= \arg \max_{\mathbf{D}_{s,t+1}, \dots, \mathbf{D}_{s,t+Q}} p(\mathbf{D}_{s,t+1}, \dots, \mathbf{D}_{s,t+Q} | \\ & \quad \mathbf{D}_{s,t-P+1}, \dots, \mathbf{D}_{s,t}), \end{aligned} \quad (4)$$

where $p(\mathbf{A} | \mathbf{B})$ denotes the conditional probability of \mathbf{A} given \mathbf{B} . The sequence of observations $\mathbf{D}_{s,t-P+1}, \dots, \mathbf{D}_{s,t}$ can be modeled as a video-like data which has P frames. Similar to an output sequence, a frame can be modeled as an image, which represents a traffic snapshot at one time interval (e.g., one hour or one day).

A. Base Station Clustering

Since the cellular traffic pattern in different areas may differ significantly, we propose to group the base stations by using a spectral clustering algorithm to capture the pattern similarity of different areas. Given time scale s , we aggregate the traffic data over all time intervals at base station (m, n) :

$$d_s^{(m,n)} = \sum_{t=1}^{T_s} d_{s,t}^{(m,n)}, \quad (5)$$

and we have

$$\mathbf{D}_s = \begin{bmatrix} d_s^{(1,1)} & d_s^{(1,2)} & \dots & d_s^{(1,N)} \\ \vdots & \vdots & \vdots & \vdots \\ d_s^{(M,1)} & d_s^{(M,2)} & \dots & d_s^{(M,N)} \end{bmatrix}.$$

Then, we convert the matrix \mathbf{D}_s into an adjacency matrix $\mathbf{G}_s \in \mathbb{R}^{MN \times MN}$, where the weight of an edge corresponds

to the value of the gradient. The Laplacian matrix can be determined as

$$\mathbf{L}_s = \mathbf{Q}_s^{-\frac{1}{2}} \mathbf{G}_s \mathbf{Q}_s^{-\frac{1}{2}}, \quad (6)$$

where \mathbf{Q}_s is a diagonal matrix with the (i, i) -th element equal to the sum of the i -row of matrix \mathbf{G}_s . Next, the K largest eigenvectors of \mathbf{L}_s , i.e., $\mathbf{u}_{s,1}, \dots, \mathbf{u}_{s,K}$, can be calculated. We use $\mathbf{U}_s = [\mathbf{u}_{s,1} \cdots \mathbf{u}_{s,K}] \in \mathbb{R}^{MN \times K}$ to denote the matrix containing the eigenvectors as columns. The rows of \mathbf{U}_s can be treated as features of the base stations, which are input into the K-means clustering algorithm to form K clusters. Finally, the cluster label of each base station can be obtained. After grouping, the total $M \times N$ base stations are clustered into K groups. The spectral clustering algorithm for base stations grouping is presented in Algorithm 1.

Algorithm 1 Spectral Clustering Algorithm

- 1: **Input:** Aggregate traffic dataset $\mathbf{D}_s \in \mathbb{R}^{M \times N}$ under time scale s , number of clusters K .
- 2: Construct an adjacency matrix $\mathbf{G}_s \in \mathbb{R}^{MN \times MN}$. The weights of \mathbf{G}_s are the gradients of \mathbf{D}_s .
- 3: Compute the Laplacian matrix \mathbf{L}_s based on (6).
- 4: Compute the first K eigenvectors of \mathbf{L}_s , i.e., $\mathbf{u}_{s,1}, \mathbf{u}_{s,2}, \dots, \mathbf{u}_{s,K}$. Let $\mathbf{U}_s = [\mathbf{u}_{s,1} \cdots \mathbf{u}_{s,K}] \in \mathbb{R}^{MN \times K}$.
- 5: For $i = 1, \dots, MN$, let $\mathbf{r}_{s,i} \in \mathbb{R}^K$ be the i -th row vector of matrix \mathbf{U}_s . Cluster the points $\{\mathbf{r}_{s,i}\}_{i=1}^{MN}$ based on K -means clustering algorithm, and construct K clusters C_s^1, \dots, C_s^K .
- 6: **Output:** Clusters $\mathbf{D}_s^1, \dots, \mathbf{D}_s^K$ with $\mathbf{D}_s^j = \{\mathbf{r}_{s,i} \in C_s^j, \forall i = 1, \dots, MN\}$, $j \in \{1, \dots, K\}$.

B. Proposed RAConv Network

1) *ResConv3D Module:* 3D convolutions can extract features of the traffic data in the spatial and temporal domains by using a 3D kernel. The Conv3D network has been shown to perform well in capturing local spatial and temporal dependencies [7]. For the plain network shown in Fig. 3(a), there is only one connection between adjacent layers. The output of a Conv3D layer is sent to an activation function layer, i.e., the rectified linear unit (ReLU). As the number of layers increases, training a network becomes more difficult due to the gradient diffusion problems. To ease this issue, we use a residual network and integrate it with the Conv3D network to form a ResConv3D module. A residual network is illustrated in Fig. 3(b), in which there is a shortcut connection that can skip one or more layers. In Fig. 4, we propose and construct a residual block (ResBlock) by stacking five Conv3D layers. We consider an identity mapping for the shortcut connection. The kernel size of each Conv3D layer is $3 \times 3 \times 3$ with stride $1 \times 1 \times 1$. Batch normalization is utilized to optimize the networks. The input and output are 3D tensors with dimension being the length of the traffic data sequence, the height and width of a base station cluster.

2) *AConvLSTM Module:* The structure of an AConvLSTM cell is illustrated in Fig. 5. We use an AConvLSTM module

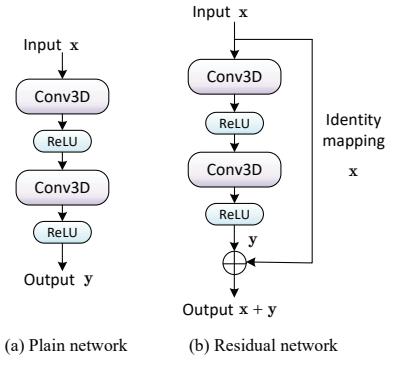


Fig. 3. Illustration of (a) a plain network and (b) a residual network.

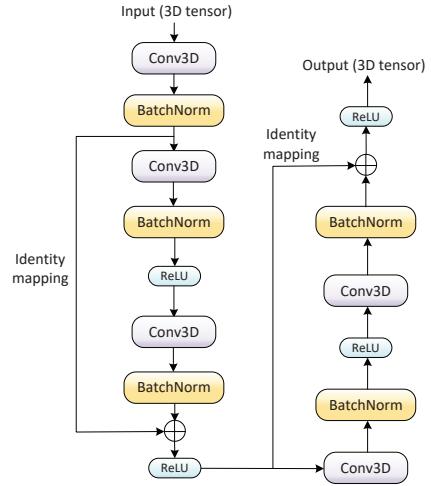


Fig. 4. Structure of the proposed ResBlock with five Conv3D layers.

to extract long-term spatial-temporal features. Conventional ConvLSTM networks replace the inner dense connections in LSTM with convolution operations, which can extract both the long-term temporal and spatial information. Given a T -sequence of 3D input data $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T\}$, the ConvLSTM operations can be expressed as

$$\begin{aligned} \mathbf{i}_t &= \sigma(\mathbf{W}_{xi} * \mathbf{X}_t + \mathbf{W}_{hi} * \mathbf{H}_{t-1} + \mathbf{W}_{ci} \odot \mathbf{C}_{t-1} + \mathbf{b}_i), \\ \mathbf{f}_t &= \sigma(\mathbf{W}_{xf} * \mathbf{X}_t + \mathbf{W}_{hf} * \mathbf{H}_{t-1} + \mathbf{W}_{cf} \odot \mathbf{C}_{t-1} + \mathbf{b}_f), \\ \mathbf{C}_t &= \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot \tanh(\mathbf{W}_{xc} * \mathbf{X}_t + \mathbf{W}_{hc} * \mathbf{H}_{t-1} + \mathbf{b}_c), \\ \mathbf{o}_t &= \sigma(\mathbf{W}_{xo} * \mathbf{X}_t + \mathbf{W}_{ho} * \mathbf{H}_{t-1} + \mathbf{W}_{co} \odot \mathbf{C}_t + \mathbf{b}_o), \\ \mathbf{H}_t &= \mathbf{o}_t \odot \tanh(\mathbf{C}_t), \end{aligned}$$

where $*$ and \odot denote the 2D convolution operator and Hadamard product, respectively. $\sigma(\cdot)$ is the sigmoid function. \mathbf{i}_t , \mathbf{f}_t , \mathbf{C}_t , \mathbf{o}_t , and \mathbf{H}_t denote the input gate, forget gate, cell output, output gate, and hidden state, respectively. They are all 3D tensors. \mathbf{W}_{xi} , \mathbf{W}_{hi} , \mathbf{W}_{ci} , and \mathbf{b}_i are the weights and bias for the input gate, which need to be learned through model training. Similarly, \mathbf{W}_{xf} , \mathbf{W}_{hf} , \mathbf{W}_{cf} , and \mathbf{b}_f are the weights and bias associated with the forget gate. \mathbf{W}_{xc} , \mathbf{W}_{hc} , and \mathbf{b}_c are the weights and bias related to the cell. \mathbf{W}_{xo} , \mathbf{W}_{ho} , \mathbf{W}_{co} , and \mathbf{b}_o are the weights and bias for the output gate. In addition, $\tanh(\cdot)$ is the hyperbolic tangent function. The input-to-state, cell-to-

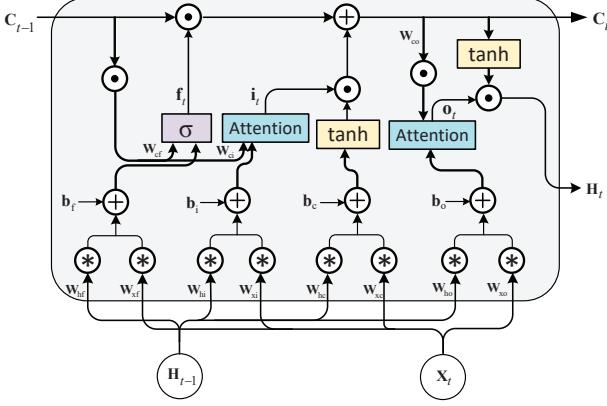


Fig. 5. The structure of an AConvLSTM cell.

state, and cell-to-cell transitions are element-wise controlled by each gate i_t , f_t , and o_t , which allow the model to keep the important information and learn to forget in the spatial-temporal dimension. This property significantly improves the models ability to capture the spatial-temporal trends.

To further improve the ability of capturing important spatial-temporal dependencies, we propose to use an attention mechanism [14] and reconstruct the input and output gates. In particular, we can express the input gate as follows:

$$Z_t = \mathbf{W}_i * \tanh(\mathbf{W}_{xi} * \mathbf{X}_t + \mathbf{W}_{hi} * \mathbf{H}_{t-1} + \mathbf{b}_i), \quad (7)$$

$$\mathbf{A}_t^{ij}(h) = \frac{\exp(Z_t^{ij}(h))}{\max_{i,j} \exp(Z_t^{ij}(h))}, \quad (8)$$

$$i_t = \{\mathbf{A}_t^{ij}(h) \mid (h, i, j) \in \mathbb{R}^{H \times M \times N}\}, \quad (9)$$

where \mathbf{W}_i is a 2D convolutional kernel. The number of channels is equal to H . $\mathbf{A}_t(h) = \{\mathbf{A}_t^{ij}(h) \mid (i, j) \in \mathbb{R}^{M \times N}\}$ denotes a 2D score map for the corresponding channel h , where $h = 1, \dots, H$. The term $\max_{i,j} \exp(Z_t^{ij}(h))$ corresponds to the maximum element chosen within the channel h of Z_t . The division by the maximum value ensures that the attention scores are distributed in the range between zero and one. The output gate of the ConvLSTM cell can be reconstructed in a similar manner as the input gate shown in (7)–(9). By embedding the attention modules into the input and output gates of ConvLSTM, the network can focus on the most important spatial-temporal features of the traffic data.

3) *RACconv network*: By stacking the ResConv3D layers and AConvLSTM layers, the final RACconv network architecture is shown in Fig. 6. The proposed network preserves the advantages of Conv3D as well as ConvLSTM, and can well extract the local and long-term spatial-temporal dependencies. Thus, the proposed RACconv network is suitable for dynamic traffic predictions in wireless cellular systems. During the training process, the inputs are given by the training samples which consist of traffic data sequences. The data sequences will subsequently be processed by a Conv3D layer, three ResConv3D layers, and two AConvLSTM layers. The output of the last layer is the final predicted results. The proposed

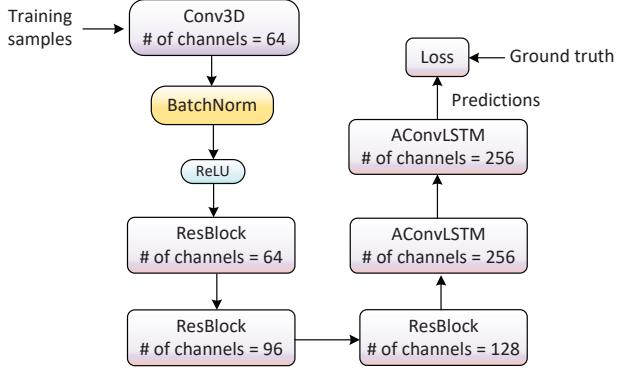


Fig. 6. Structure of the proposed RACconv network.

network is trained to minimize the loss, which is the error between the predictions and the ground truth.

IV. PERFORMANCE EVALUATION

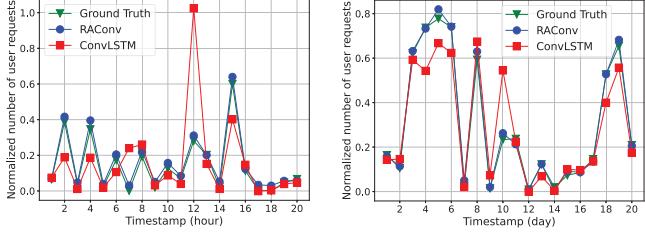
In this section, we present experimental results to evaluate the performance of the proposed RACconv network for traffic prediction. The evaluation is based on the number of user requests collected from a Canadian wireless service provider, Rogers Communications Inc. There are 4,096 base stations, which are clustered into 16 groups using Algorithm 1. Each group consists of 16×16 spatially correlated base stations¹.

We compare the performance of our proposed RACconv network with the ConvLSTM [6] as a baseline scheme. Fig. 7 compares the predicted traffic demands of RACconv and ConvLSTM networks with the ground truth at a randomly selected base station. We consider the number of observations P to be equal to 8. It needs to make $Q = 4$ predictions for future time intervals. The hourly and daily traffic predictions are presented. It can be observed that the proposed RACconv network can predict the cellular traffic more accurately than ConvLSTM network on both hourly and daily basis. Moreover, results show that even when the traffic dynamics change rapidly either every hour or every day, the predicted results from our proposed RACconv network are very close to the ground truth.

In Fig. 8, we present an image comparison at one time interval (i.e., one hour or one day) between the ground truth and predicted results. The results are from a randomly chosen cluster of base stations. The brightness of a pixel indicates the traffic load of the corresponding base station. Results from Fig. 8 show that the proposed RACconv network can make more accurate traffic predictions than the ConvLSTM network.

The root-mean-square error (RMSE) performance is shown in Fig. 9. The number of prediction steps Q is set to 4. Results show that when the number of observations P increases, both RACconv and ConvLSTM networks can utilize the additional historical information and make better prediction. Thus, the values of RMSE in both networks decrease. We also observe

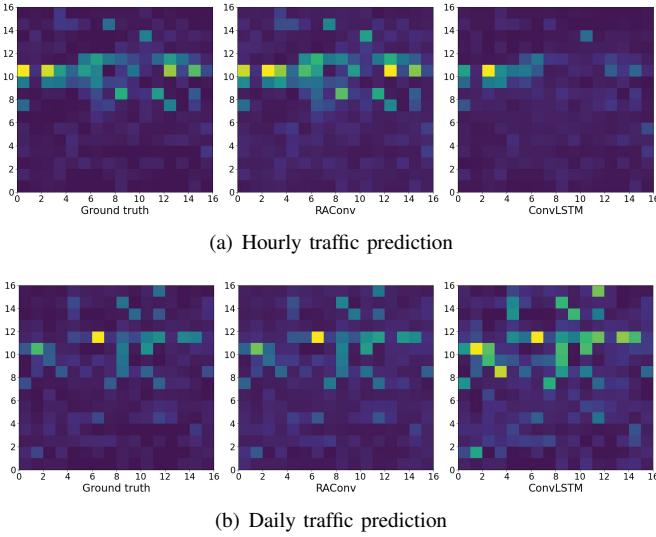
¹Note that a larger group may contain more spatial information but will result in higher computational complexity. Considering this trade-off, we choose the size of each group to be 16×16 .



(a) Hourly traffic prediction

(b) Daily traffic prediction

Fig. 7. Comparison of (a) hourly and (b) daily predicted results of RAConv and ConvLSTM versus the ground truth at a randomly selected base station.

Fig. 8. Performance comparison of the normalized (a) hourly and (b) daily number of user requests on a randomly selected cluster of 16×16 base stations and a frame.

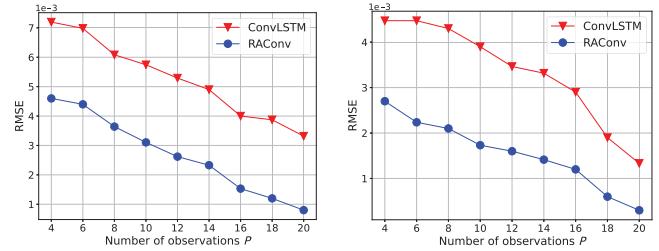
that the proposed RAConv network outperforms the ConvLSTM network under different number of observations.

V. CONCLUSION

In this paper, we studied the problem of cellular traffic prediction in wireless cellular networks. We proposed a RAConv network, which consists of ResConv3D and AConvLSTM modules, to learn the local short-term and long-term spatial-temporal features. We evaluated the proposed RAConv network based on a dataset provided by a Canadian wireless service provider. Experimental results showed that our proposed RAConv network provided more accurate prediction when compared with the ConvLSTM network, and achieved a lower RMSE, on both hourly and daily time scales. In our current work, the hourly and daily traffic predictions are determined by separately training multiple networks, where the knowledge of hourly traffic can actually be utilized for daily traffic forecasting. For future work, traffic similarity between different time scales will be employed to further improve the prediction accuracy.

ACKNOWLEDGEMENT

This work was supported by Rogers Communications Canada Inc.



(a) Hourly traffic prediction

(b) Daily traffic prediction

Fig. 9. RMSE performance comparison of (a) hourly and (b) daily traffic prediction under different number of observations P . The number of prediction steps Q is equal to 4.

REFERENCES

- [1] U. Paul, A. P. Subramanian, M. M. Buddikot, and S. R. Das, "Understanding traffic dynamics in cellular data networks," in *Proc. of IEEE Int. Conf. Computer Commun. (INFOCOM)*, Shanghai, China, Apr. 2011.
- [2] H. D. Trinh, L. Giupponi, and P. Dini, "Mobile traffic prediction from raw data using LSTM networks," in *Proc. IEEE Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Bologna, Italy, Sep. 2018.
- [3] W. Wang, C. Zhou, H. He, W. Wu, W. Zhuang, and X. Shen, "Cellular traffic load prediction with LSTM and Gaussian process regression," in *Proc. of IEEE Int. Conf. Commun. (ICC)*, Jun. 2020.
- [4] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, Apr. 2017.
- [5] J. Wang, J. Tang, Z. Xu, Y. Wang, G. Xue, X. Zhang, and D. Yang, "Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach," in *Proc. IEEE Int. Conf. on Computer Commun. (INFOCOM)*, Atlanta, GA, May 2017.
- [6] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst (NeurIPS)*, Montreal, Canada, Dec. 2015.
- [7] C. Zhang and P. Patras, "Long-term mobile traffic forecasting using deep spatio-temporal neural networks," in *Proc. ACM Int. Symp. on Mobile Ad Hoc Netw. and Comput. (MobiHoc)*, Los Angeles, CA, Jun. 2018.
- [8] Y. Zhang, S. Wang, B. Chen, J. Cao, and Z. Huang, "TrafficGAN: Network-scale deep traffic prediction with generative adversarial nets," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 1, pp. 219–230, Jan. 2021.
- [9] Q. Liu, J. Li, and Z. Lu, "ST-Tran: Spatial-temporal transformer for cellular traffic prediction," *IEEE Commun. Lett.*, vol. 25, no. 10, pp. 3325–3329, Oct. 2021.
- [10] C. Zhang, H. Zhang, J. Qiao, D. Yuan, and M. Zhang, "Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1389–1401, Jun. 2019.
- [11] C. Zheng, X. Fan, C. Wen, L. Chen, C. Wang, and J. Li, "DeepSTD: Mining spatio-temporal disturbances of multiple context factors for citywide traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3744–3755, Sep. 2020.
- [12] S. Fang, X. Pan, S. Xiang, and C. Pan, "Meta-MSNet: Meta-learning based multi-source data fusion for traffic flow prediction," *IEEE Signal Process. Lett.*, vol. 28, pp. 6–10, Nov. 2021.
- [13] T. Kuber, I. Seskar, and N. Mandayam, "Traffic prediction by augmenting cellular data with non-cellular attributes," in *Proc. of IEEE Wireless Commun. and Netw. Conf. (WCNC)*, Nanjing, China, Mar./Apr. 2021.
- [14] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. Int. Conf. Machine Learning (ICML)*, Lille, France, Jul. 2015.
- [15] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting, Third Edition*. Springer, 2016.
- [16] S. Yin, D. Chen, Q. Zhang, M. Liu, and S. Li, "Mining spectrum usage data: A large-scale spectrum measurement study," *IEEE Trans. Mobile Comput.*, vol. 11, no. 6, pp. 1033–1046, Jun. 2012.