STeP-UNet: Prediction of Moving and Communication Behaviors of Vehicles

Daojun Liang*‡, Haixia Zhang†*, Xiaotian Zhou†*, Dongyang Li*‡, Dongfeng Yuan*‡ and Minggao Zhang†
*Shandong Key Laboratory of Wireless Communication Technologies,
‡School of Information Science and Engineering, Shandong University, Qingdao, Shandong, China
†School of Control Science and Engineering, Shandong University, Jinan, Shandong, China
liangdaojun@mail.sdu.edu.cn, haixia.zhang@sdu.edu.cn, xtzhou@sdu.edu.cn,
lidongyang@mail.sdu.edu.cn, dfyuan@sdu.edu.cn, zhangmg@cae.cn

2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall) | 978-1-6654-1368-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/VTC2021-FALL52928.2021.9625575

***Abstract*—Wireless traffic prediction has drawn increasing research interests as it can provide guidance to the network optimization. With the predicted information, one can preassign the resources on demand and perform network congestion control adaptively. The network efficiency is therefore enhanced. However, the wireless traffic prediction in the context of mobile scenario, such as Internet of Vehicles (IoVs), is still a challenge issue. The mobile nature of devices, which dynamically changes the topology of network, would brings difficulties to the prediction. This paper focuses on the deep learning based wireless traffic prediction in the IoVs scenario. We first propose a novel method to match up the movement- and communication-behavior of users, by merging two independent datasets on the trajectories of vehicles and communication traffic volumes together. Then a novel STeP-UNet is proposed, in which the SpatioTemporal Partial (STeP) Convolutional Neural Network module is embedded to capture cross-domain features of the wireless traffic pattern, and the UNet structure is utilized to realize the skipping connection from front layer to back layer to fuse different resolutions. Experimental results confirms the promising performance of the proposed model, where 4%∼8% performance improvement over other benchmark methods can be achieved.**

I. Introduction

Wireless traffic prediction plays an important role in developing a more intelligent and greener communication networks [1]. By gathering and analyzing the historical data, one can predict the future trend of wireless traffic in the network and preassign the necessary resources on demands. The various Quality of Service (QoS) from users can be satisfied with high resource utilizing efficiency [2], [3].

The accurate wireless traffic prediction relies on the understanding of the communication behaviors inherit in the historical data, where the spatiotemporal distribution of communication demands is expected to be captured. Deep learning [4]–[7], which is originally used in the areas such as Computer Vision (CV) and Natural Language Processing (NLP), is considered to be a promising tool for wireless traffic prediction [8]–[10]. For instance, the recurrent neural network (RNN) based framework was proposed in [8] to model the nonlinear temporal dependency of the wireless traffic. The authors in [9] proposed to jointly model the spatial and temporal dependencies of wireless traffic, through mixture deep Long Short Term Memory (LSTM) model. In [11], a hybrid deep learning framework which explored the combination of CNN and LSTM (ConvLSTM) to capture the spatial and temporal dependencies of traffic, is reported. Moreover, [12] proposed to predict the BS traffic volumes based on the integration of K-means clustering and wavelet decomposition methods, through digging the spatiotemporal information of cellular traffic flow. [13] investigated the spatial and temporal dependence of wireless traffic among different cells, where the spatiotemporal DenseNet (STDenseNet) [7] based prediction framework is designed. [1] conducted clustering on different Point of Interest (POI) modes, and then uses transfer learning in different communication traffic patterns to improve the prediction accuracy.

Though promising methods have been reported in literatures for communication traffic prediction in traditional wireless network, they are hardly applied to the IoVs scenario directly. In IoVs, the topology of network would change dynamically owning to the movement of the users (now the vehicles). The spatiotemporal distribution of communication traffic is therefore impacted by not only the communication demands, but also the movement behaviors of the vehicles. The correlation between them should be considered, as it is critical to the performance of prediction. In fact, there already exists several research work which consider the vehicle traffic [14], [15]. For example, [14] uses three residual networks to model the closeness, period and trend of the vehicle traffic, where the weather conditions and daily events are also padded as the external features. [15] considered the combination of ConvPlus module and SemanticPlus module, where the former captures long-range spatial dependence and the latter is employed to lower the impacts of location function to the crowd movement. However, to our best knowledge, few efforts have been done to consider the joint prediction of vehicle traffic and wireless communication traffic, in the context of IoVs case.

In this paper, we propose a novel framework for wireless traffic prediction in IoVs, where the movement behavior of vehicles, as well as the communication behavior of users are jointly considered. To achieve that, we first propose a pattern matching approach to merge the wireless traffic dataset with the real vehicle trajectories, The new four dimensional dataset containing the amount and types of wireless data traffic, as well as the vehicle coordinates per each time stamp can be obtained through the proposed approach. With the

978-1-6654-1368-8/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: SHANDONG UNIVERSITY. Downloaded on July 09,2022 at 11:48:02 UTC from IEEE Xplore. Restrictions apply.

dataset in hand, we further design a flexible and embeddable SpatioTemporal Partial (STeP) convolutional network module to accurately model the traffic data of the IoVs with a time series relationship. Moreover, the proposed STeP modules is further embedded into Unet [16] to perform the spatiotemporal traffic prediction. Experimental results confirms the promising performance of the proposed method.

The remaining of the paper is organized as follows. In Section II, the data process modeling methodology will be introduced. The STeP-UNet of the proposed method is included in Section III. Section IV introduces simulation result of the proposed method. Finally, we conclude the paper in Section V.

II. Data Matching and Processing

A. Dataset Merging

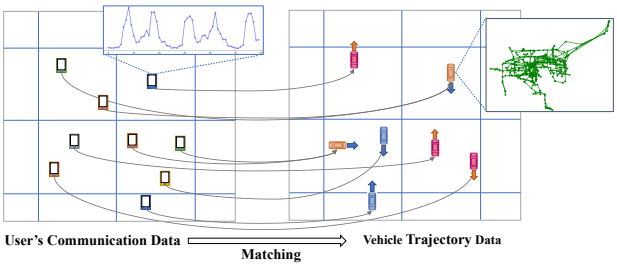

Fig. 1. The user's communication data are used to match the vehicle trajectory data.

As aforementioned, though there are datasets tracking the trajectories of vehicles [17], as well as those for communication behaviors [18] separately, few works have been done to merge them for wireless data traffic prediction in the context of IoVs. As a result, the first mission in this work is to match up the dataset containing the location information of the vehicle with that for communication traffic. Denoting these two datasets as V and U, respectively, they can both be expressed as the three dimensional tensors. To be specific, $V = (N, T, P)$ records all location information of N vehicles in T time stamps, where P contains the location information of each vehicle (such as latitude and longitude information). On the other hand, the communication traffic dataset $U = (N', T', S)$ provides S types of communication traffic of N' users in T' time stamps. Apparently, the main focus here is the match up S with P, where the relationships between N and N', as well as T and T' should be firstly established as:

$$N = \Psi(N'), \quad T = \Phi(\Psi(T')), \tag{1}$$

where $\Psi(\cdot)$ function gives the correspondence between the vehicles and the users. The intuition here is that we can treat the N users in U is the passengers scattered in the N' vehicles. $\Phi(\cdot)$ is the scaling function which forces T to be in the same time scale of T'. With the above two equations in hand, the matching up of these two datasets can be formulated as:

$$\begin{aligned} MS &= (\Psi(\Phi(N')), \Psi(\Phi(T')), P, \Psi(\Phi(S))) \\ &= (N, T, P, \Psi(\Phi(S))). \end{aligned} \tag{2}$$

To better illustrate the above process, Fig. 1 is provided to show more details about the mapping from user communication data to vehicle trajectory data. Note that for easy understanding, here we consider the special case when $N = N'$. In such a case, $\Phi(\cdot)$ degrades to the identity mapping function to bind each vehicle with one certain user, where the communication behavior of each user during the movement of the vehicle can be readily obtained.

B. Traffic Map Division and Data Processing

Fig. 2. Dividing the city map and generating the matrix. The map is covered by $H \times W$ grids, and the total traffic in each grid are calculated by aggregation operation.

With the merged dataset MS in hand, we then focus on tracking the dynamics of wireless data traffic based on the vehicle trajectory in the citywide scale. To be specific, we in this work take the taxi trajectory collected in Beijing City as the basis dataset [17], from which the maximum and minimum latitudes/longitudes can be obtained to determine the moving area of vehicles. As illustrated in Fig. 2 we first divide the corresponding citywide map into $H \times W$ grids with the height and width of each grid denoted as G_H and G_W, respectively. Then in each grid, the amount of wireless traffic during the t^{th} time slot are calculated independently for different communication types, through aggregation the corresponding wireless flows from the vehicles currently positioned within that grid. Such aggregation operation is illustrated in the following **Definition 1** [14].

Definition 1: Let P be a collection of trajectories at the t^{th} time slot. For a $Grid(i, j)$ that lies at the i^{th} row and the j^{th} column, the volume of the traffic at the time slot t are defined as

$$x_t^{i,j} = \sum_{Tr \in P} |\{k \geq 1 | g_{k-1} \notin (i, j) \wedge g_k \in (i, j)\}| \tag{3}$$

where $Tr : g_1 \rightarrow g_2 \rightarrow \cdots \rightarrow g_{Tr}$ is the trajectory from P. g_k denotes the coordinates of one target point in the citywide map. $|\cdot|$ is defined as the cardinality of trajectories set. By performing such aggregation in each grid per time slot, a set of wireless traffic related tensor $\{x_t^{i,j}\}$ with dimension $T \times H \times W$ can be finally obtained. The entire process is concluded in Algorithm 1, where $\{LT_{max}, LT_{min}\}$ and $\{LN_{max}, LN_{min}\}$ represents the maximum/minimum latitude and longitude of the city, respectively.

Through Algorithm 1, the time series of vehicle traffic and its counterpart wireless data traffic can be obtained, which is

Authorized licensed use limited to: SHANDONG UNIVERSITY. Downloaded on July 09,2022 at 11:48:02 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1 Matching algorithm between vehicle traffic data and communication traffic data.

Require: V, U, H, W

Compute $T, S, G_H, G_W, LT_{max}, LT_{min}, LN_{max}, LN_{min}$ according by V and U
Initial $Grid = Zero(shape = [T, X, Y, S])$
for $i = 1 \rightarrow T$ **do**
  $Car_{pos} = V^i$
  for $j = 1 \rightarrow N$ **do**
    $Map_{pos} = Car^j_{pos}$
    if **then** $LT_{max} > Map_{pos} \geq LT_{max}$
      or $LN_{max} > Map_{pos} \geq LN_{min}$
      $P_{LT} = \frac{Map^{LT}_{pos} - LT_{min}}{G_H}$
      $P_{LN} = \frac{Map^{LN}_{pos} - LN_{min}}{G_W}$
      $Grid[i, P_{LT}, P_{LN}]_{Pos} += V[j, i]$
      $Grid[i, P_{LT}, P_{LN}]_U += \Phi(\Psi(U[j, i]))$
    end if
  end for
end for

a 4-dimensional tensor illustrating the spatiotemporal distribution of the wireless traffic volume with different types. With that the fundamental network architecture can be designed to model and predict the traffic patterns.

III. STeP-UNet

A. Review of ResNet and DenseNet Module

To better understanding the proposed STeP module, we would quickly go through the basic idea of ResNet and DenseNet. Denoting the features of the input and the l^{th} layer of the network as X_0 and X_l, respectively. The structure of ResNet can be defined as

$$X_l = X_l + f(X_{l-1}), \tag{4}$$

where $f(\cdot)$ is the standard composite function [6]. It contains the successive operations like Convolution, Batch Normalization and Rectified Linear Unit [19]–[21]. Apparently, X_l in Eq. (4) is updated recursively, in which all the features are treated as the units through $f(\cdot)$ and then aggregated together through summation.

On the other hand, the basic structure of DenseNet is expressed as:

$$X_l = [X_{l-1}, f(X_{l-1})], \tag{5}$$

where $[\cdot, \cdot]$ represents the concatenation operation on the features set. It can be observed that Eq. (5) is quite similar to Eq. (4) except that the summation among features is replaced by the concatenation. Both networks inherits the advantages of achieving efficient gradients in backpropagation by fusing the features of the front and back layers. However, they also have drawbacks. For instance, ResNet adds all the front layer features to the back layer and hence has a higher amount of parameters. While for the DenseNet, it merges only the front layer features but without fusing them. It may reduce the complexity but on the other hand drag down the performance.

In order to inherit the advantages of both network and make it suitable for traffic prediction tasks, the STeP module is proposed in this paper. The basic idea is to divide the features in each layer into multiple groups and appropriately apply summation or concatenation operations in different groups. In this way a lightweight variant network incorporating ResNet and DenseNet can be achieved. The details of such a network module is described in subsection III-B.

B. STeP-Module

In this subsection, we give the STeP-Module in details [21], [22]. As aforementioned, for a deep learning network we can divide the feature channels into G groups. Denoting the features in the $j^{th}(1 \leq j \leq G)$ group of layer l as g^j_l, the features of layer l can be defined as

$$X_l = [g^1_l, \cdots, g^G_l], \tag{6}$$

Consequently, the block structure of STeP-Module can be defined as

$$X_l = X_{l-1} + f(g^{j-1}_{l-1}). \tag{7}$$

Fig. 3 illustrates the structure of one single layer of STeP-Module, which is divided into a number of 3 groups [22]. The number of features of each traffic pattern will be amplified by group convolution to obtain G group features $[g^1_l, \cdots, g^G_l]$[1]. All group features can then be regarded as the input X_0 in the STeP-Module. The purpose of dividing multiple groups is to capture the different traffic patterns in the IoVs, with each group convolution are used to model one corresponding pattern, respectively. STeP-Module makes full use of the advantages of ResNet and DenseNet, so that the gradient can be efficiently back-propagated. Moreover, with the proposed module one can independently apply summation or concatenation operations for different groups, according to the different types of pattern it deals with. It helps reduce the amount of parameters of the whole model and can potentially reach a good trade-off between complexity and performance. It is also noted that though each group only represents one of the multiple traffic patterns. The STeP-Module can still fusion and learn each traffic pattern sequentially as the layers go deeper. It reveals that STeP-Module can model the traffic data with spatiotemporal relationship well.

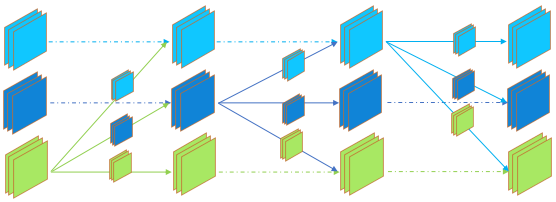

Fig. 3. The structure of STeP-Module [22].

[1]In default setup, $G = S$

Authorized licensed use limited to: SHANDONG UNIVERSITY. Downloaded on July 09,2022 at 11:48:02 UTC from IEEE Xplore. Restrictions apply.

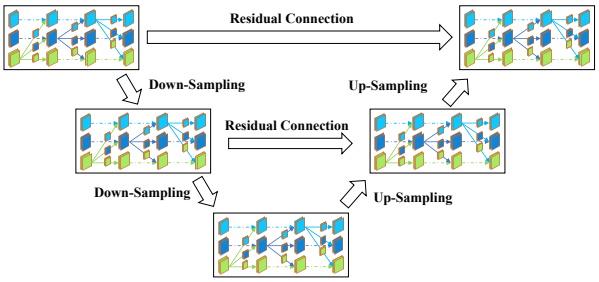

Fig. 4. The Architecture of STeP-UNet.

C. STeP-UNet

In this subsection, we consider to embed the STeP-Module into the UNet architecture to setup the final learning model for prediction. Fig. 4 illustrates the architecture of the proposed STeP-UNet, where each node in UNet is compromised by the STeP-Module. It can be easily found that STeP-UNet has a large number of skip connections from the architecture level to the node level. Such characteristic enables the model to efficiently learn the timing relationship of traffic data and thereby avoid the problem of gradient dispersion.

As for the parameter setting of the model, the summation operations are used for overlapping channels and the concatenation operations are used for the redundant channels. These operations can reduce network parameters, which can be analyzed from two aspects. On the one hand, STeP-Module uses 3×3 and 1×1 convolution sequentially in the amplification function. The number of the group channels are amplified to $\frac{C}{4}$ using 3×3 convolution, and then there are amplified to C using 1×1 convolution. The number of parameters of 1×1 convolution are much less than that of 3×3 convolution. In other words, the network parameters will decrease with the increase of the number of the group compared with other networks. On the other hand, unlike STResNet and STDenseNet that use multiple ResNet or DenseNet Module to capture multiple traffic patterns separately, STeP-Module employ one single module for multi-pattern data learning. That is because STeP-Module uses group convolution to separately model multiple patterns, where each mode corresponds to only one group and each group sequentially performs partial-to-whole feature fusion [22]. Based on these characteristics, STeP-UNet can greatly reduce the amount of network parameters and speed up the network training. It should also be noted that the STeP-Module can be embedded in any network architecture to achieve efficient integration of features. While the purpose of choosing UNet architecture in this work is just to achieve a direct skip connection from the front layer to the back layer.

IV. Experiments and Discussions

A. Datasets

We use two different datasets to model the moving trend of the vehicles and the communication traffic of users, respectively. The vehicle traffic dataset is from [17], which contains the GPS logs of the taxi cars in Beijing City. While the end user communication dataset is from the user communication data in Milan, Italy and the BS traffic data from some city in China. The details of the corresponding datasets are described as follows.

Vehicle Traffic: This dataset contains the GPS trajectories of 10,357 taxis during the period of 02/02/2008 to 02/08/2008 within Beijing [17]. The total number of points in this dataset is about 15 million and the total distance of the trajectories reaches to 9 million kilometers.

Cellular Traffic: Inspired by [1], we employ the cellular traffic dataset from [18], which is provided by Telecom Italia. The dataset is collected from 11/01/2013 to 01/01/2014 with a temporal interval of 10 minutes over the whole city of Milan (62 days, 300 million records, about 19 GB).

BS Traffic: The real communication traffic data is collected from all BS in a certain city in China, including up-link and down-link communication traffic data. The dataset is collected from 08/06/2019 to 11/07/2019 with a temporal interval of 60 minutes over the whole city (94 days, 51.9 million records, about 17.1 GB).

The traffic data are comprised of three major components modeling temporal closeness, period, trend, respectively. With **Definitions 1**, the traffic data of the city during one each time slot can be first converted into a tensor, similar to that of a multi-channel image [21]. The 6-channel Vehicle-Cellular traffic tensors (Vehicle Traffic, In-SMS, Out-SMS, In-Call, Out-Call, Internet) and 3-channel Vehicle-BS traffic tensors (Vehicle Traffic, Up-Link, Down-Link) are fed into the first three components network separately to model three temporal properties: closeness, period and trend. Note that these three component networks in the proposed UNet [16] inherit the same structure thanks to the STeP-Module embedded. Finally, with the help of Sigmoid function $S(x) = \frac{1}{1+e^{-x}}$, the aggregated features are maped into the values within [0,1]. It yields a faster convergence than the standard logistic function in the process of backpropagation learning.

B. Settings

Dataset Settings: For the traffic matching proccess, the whole city is split into 32×32 grids. Then to get the input data ready, we further perform normalization on the traffic volume through the widely adopted Min-max normalization approach [1]. For functions Ψ and Φ, the truncation operation is first taken, and then the identity mapping is used to match the data. Finally, each vehicle trajectory corresponds to a unique communication process. During prediction stage, the prediction value are denormalized and used for evaluation. Note that to reserve not only the temporal correlation but also the spatial one in the original dataset, the sliding window method is employed when sampling. 80% of the samples are randomly chosen for training while the remaining is left for testing [21].

Loss Function: The proposed deep learning model can be easily trained through minimizing the Frobenius norm between

Authorized licensed use limited to: SHANDONG UNIVERSITY. Downloaded on July 09,2022 at 11:48:02 UTC from IEEE Xplore. Restrictions apply.

TABLE I
TRAFFIC PREDICTION PERFORMANCE COMPARISON BETWEEN STEP-UNET, STRESNET AND STDENSENET.

Dataset	Model	Params (M)	MAE	RMSE	R^2
Vehicle-Cellular	STResNet	0.85	0.09	0.24	72.54
	STDenseNet	0.78	0.10	0.25	72.05
	STeP-UNet	0.75	**0.07**	**0.23**	**75.85**
Vehicle-BS	STResNet	0.85	0.56	1.56	94.50
	STDenseNet	0.78	0.87	2.19	88.76
	STeP-UNet	0.75	**0.55**	**1.55**	**94.52**

the predicted value and the ground truth value of the t^{th} slot [13]

$$L(\theta) = \underset{\theta}{argmin} \, ||X_t - \tilde{X}_t||_F^2, \tag{8}$$

where X_t and $\tilde{X}_t$ denotes the true traffic map and the prediction one in time slot t, respectively.

Hyperparameter Settings: The STeP-UNet is trained using one NVIDIA Tesla V100 GPU under the Pytorch framework. The well known Adam optimization method is employed for gradient descent, given the initial learning rate 0.001 [23]. We choose the hyperparameters according to that of the optimal model examined on the validation set.

Network Settings: For the proposed STeP-UNet, we set the initial number of convolution channels to be 4G, where G denotes the number of traffic types and equals to 6 in this work. Moreover, for performance comparison we also exam two other competitive schemes, namely, STResNet and STDenseNet with the similar network settings as that in [21]. More details can be found therein.

C. Performance

We compared the performance of STeP-UNet, STResNet and STDenseNet on the Vehicle-Cellular and Vehicle-BS datasets. The results are concluded in Table I. It can be found that for different indicators, such as Params (the parameters is abbreviated to Params) MAE, RMSE and R^2, STeP-UNet has shown good generalization performance on these indicators.

On the Vehicle-Cellular dataset, STeP-UNet achieves lowest MAE and RMSE compare to STResNet and STDenseNet. In terms of MAE metrics, the performance of STeP-UNet has improved significantly. In terms of RMSE metrics, STeP-UNet has 4% and 8% performance improvements compared to STResNet and STDenseNet.

The proposed STeP-UNet also performs best on the Vehicle-BS dataset. The reason lies in that the STeP module allows the network to capture the information of temporal and spatial domains more effectively, especially in the case when the feature size is large. Moreover, the UNet structure facilitates the sharing of parameters among multiple traffic patterns, which again confirms the promising performance of the proposed model in joint prediction of both communication and movement behaviors.

D. Visualization and Analysis

Fig. 5 and Fig. 6 depict the predictions of the vehicle traffic and communication traffic from the perspective of spatial and temporal domain, respectively. It can be seen from Fig. 5 that the vehicle traffic and its communication traffic pattern can be well matched. It generally obeys the rule that communication traffic increases with vehicle traffic, but there may be exceptions in some areas. For example, the city center. Fig. 6 also shows the diversity of different traffic patterns, but they are all positively correlated with vehicle traffic.

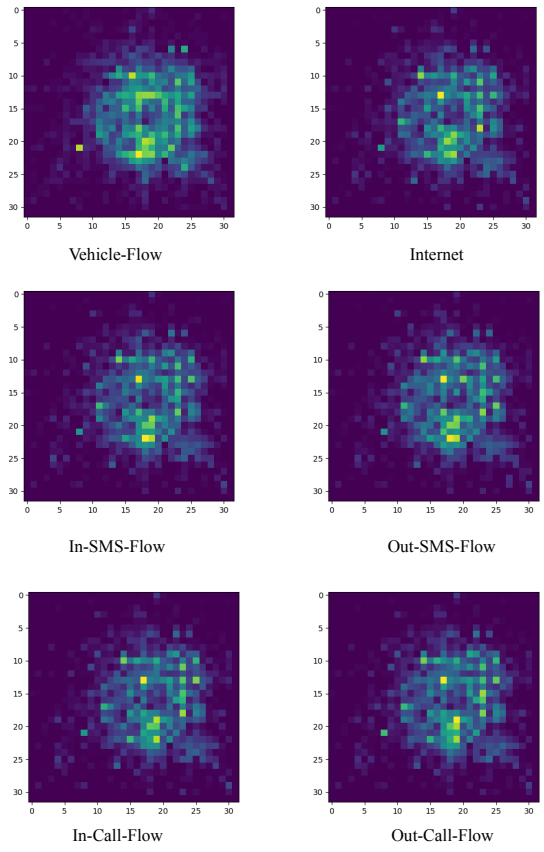

Fig. 5. Spatial pattern of Vehicle-Cellular traffic prediction results.

Fig. 6 describes the time-varying relationship of various traffic patterns. Note that different types of traffic have different magnitudes, but they almost have the same timing relationship. As can be seen from the figure, although there are some local disturbances, STeP-UNet can better predict the traffic pattern in the IoVs scenario.

V. CONCLUSIONS

In this paper, we studied the traffic prediction in the IoVs scenario. In order to collect data from the IoVs scenario, a traffic matching algorithm was proposed to match the vehicle traffic with the real communication traffic, where the statistical analysis and modeling of the vehicle traffic was carried out on a citywide scale. In order to better predict various traffic patterns, the STeP-Module based on spatiotemporal relationship is used. This module models each traffic pattern into a group, and each group can perform partial-to-whole feature fusion

Authorized licensed use limited to: SHANDONG UNIVERSITY. Downloaded on July 09,2022 at 11:48:02 UTC from IEEE Xplore. Restrictions apply.

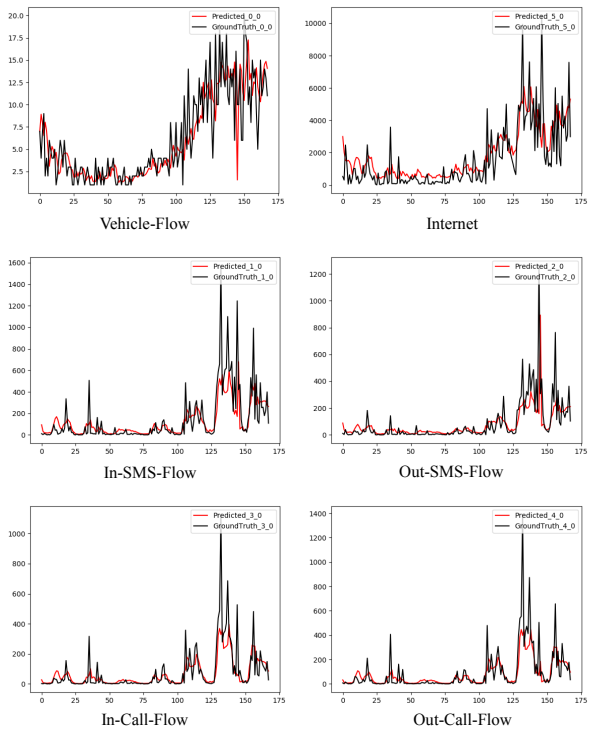

Fig. 6. Temporal pattern of Vehicle-Cellular traffic prediction results.

through expansion operation to predict multiple traffic patterns simultaneously. Furthermore, the STeP-Module is embedded into UNet to realize efficient capture of spatiotemporal relationship of traffic data, thereby alleviate the gradient dispersion problem. Experimental results have proved that STeP-UNet can not only capture a variety of traffic patterns, but also precisely predict the traffic volume in the IoVs scenario.

ACKNOWLEDGMENT

This work is supported in part by the Project of International Cooperation and Exchanges NSFC under Grant No. 61860206005, the National Natural Science Foundation of China under Grant No. 61971270, the Major Scientific and Technological Innovation Project of Shandong Province under Grant No. 2020CXGC010108, and the Shandong Provincial Natural Science Foundation (Grant No.ZR2019QF016).

REFERENCES

[1] 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 Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1389–1401, 2019.

[2] M. Noor-A-Rahim, Z. Liu, H. Lee, G. G. M. N. Ali, D. Pesch, and P. Xiao, "A survey on resource allocation in vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–21, 2020.

[3] H. Peng and X. Shen, "Deep reinforcement learning based resource management for multi-access edge computing in vehicular networks," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 4, pp. 2416–2428, 2020.

[4] Y. Lecun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 770–778.

[7] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely connected convolutional networks," in *CVPR*, 2016.

[8] J. T. Connor, R. D. Martin, and L. E. Atlas, "Recurrent neural networks and robust time series prediction," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 240–254, 1994.

[9] R. Yu, Y. Li, C. Shahabi, U. Demiryurek, and Y. Liu, "Deep learning: A generic approach for extreme condition traffic forecasting," in *Proceedings of the 2017 SIAM international Conference on Data Mining*. SIAM, 2017, pp. 777–785.

[10] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[11] S. Xingjian, 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 *Advances in neural information processing systems*, 2015, pp. 802–810.

[12] X. Chen, Y. Jin, S. Qiang, W. Hu, and K. Jiang, "Analyzing and modeling spatio-temporal dependence of cellular traffic at city scale," in *2015 IEEE International Conference on Communications (ICC)*, 2015, pp. 3585–3591.

[13] C. Zhang, H. Zhang, D. Yuan, and M. Zhang, "Citywide cellular traffic prediction based on densely connected convolutional neural networks," *IEEE Communications Letters*, vol. 22, no. 8, pp. 1656–1659, 2018.

[14] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proceeding of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)*, November 2016.

[15] Z. Lin, J. Feng, Z. Lu, Y. Li, and D. Jin, "Deepstn+: Context-aware spatial-temporal neural network for crowd flow prediction in metropolis," in *Proceeding of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)*, 2019.

[16] O. Ronneberger, P.Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, ser. LNCS, vol. 9351. Springer, 2015, pp. 234–241.

[17] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "Driving with knowledge from the physical world," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2011, pp. 316–324.

[18] G. Barlacchi, G. Barlacchi, R. Larcher, and A. Casella, "A multi-source dataset of urban life in the city of milan and the province of trentino," *Scientific Data*, pp. 2052–4463, 2015.

[19] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *ICML*, 2015, pp. 448–456.

[20] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *International Conference on Artificial Intelligence and Statistics*, 2012.

[21] L. Zhai, Y. Yang, S. Song, S. Ma, X. Zhu, and F. Yang, "Self-supervision spatiotemporal part-whole convolutional neural network for traffic prediction," *Physica A: Statistical Mechanics and its Applications*, vol. 579, p. 126141, 2021.

[22] D. Liang, F. Yang, T. Zhang, J. Tian, and P. Yang, "Wpnets and pwnets: From the perspective of channel fusion," *IEEE Access*, vol. 6, pp. 34 226–34 236, 2018.

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

Authorized licensed use limited to: SHANDONG UNIVERSITY. Downloaded on July 09,2022 at 11:48:02 UTC from IEEE Xplore. Restrictions apply.