Transportation Research Record

A Pseudo-3D Convolutional Neural Network based Framework for Short-term Mixed Passenger Flow Prediction in Large-scale Public Transit

--Manuscript Draft--

Full Title:	A Pseudo-3D Convolutional Neural Network based Framework for Short-term Mixed Passenger Flow Prediction in Large-scale Public Transit
Abstract:	Globally, many metropolitan areas tend to be promoting the multi-modal public transit which interconnects different modes to move large quantities of commuters in the urban area. Therefore, the accurate short-term passenger flow prediction can contribute to improve the reliability, responsiveness and the service quality of the multi-modal public transit. In this paper, we propose an end-to-end deep learning framework, called ST-Pseudo3D Net to collectively predict different types of public transit passenger flow at each region of city in the near future. The framework employs the deep Pseudo-3D residual architecture to model the network-wide spatial-temporal correlations among different types of passenger flow. Based on smart card data collected from Singapore's multi-modal public transit, considering metro passenger flow, bus passenger flow as well as the transfer passenger flow between these metro and bus, totally we construct 7 different types of passenger flow to be collectively predicted, where the experiments results demonstrate that the proposed model synthetically outperforms other baselines. To the best knowledge of the authors, this is the first attempt to investigate the integrated prediction for multi-modal passenger flow leveraging deep learning techniques.
Manuscript Classifications:	Data and Information Technology; Artificial Intelligence and Advanced Computing Applications ABJ70; Artificial Intelligence; Machine Learning (Artificial Intelligence); Neural Networks; Transformative Transit Data AP000; Smartcard; Transportation Issues in Major Cities and Urban Mobility ABE30; Urban Mobility; Public Transportation; Urban Transportation Data and Information Systems ABJ30; Multimodal Analysis; General
Manuscript Number:	
Article Type:	Presentation
Order of Authors:	Siyu Hao
	Dingyi Zhuang
	De Zhao
	Der-Horng Lee

A Pseudo-3D Convolutional Neural Network based Framework for Short-term Mixed Passenger Flow Prediction in Large-scale Public Transit

Siyu Hao
(Corresponding author)
Department of Civil & Environmental Engineering,
National Unversity of Singapore
Singapore 117576, Singapore
Email: siyuhao@u.nus.edu

Dingyi Zhuang
Department of Civil Engineering and Applied Mechanics
McGill University
Montreal, Quebec H3A 0C3, Canada
Email: dingyi.zhuang@mail.mcgill.ca

De Zhao
Department of Civil & Environmental Engineering
National University of Singapore
Singapore 117576, Singapore
Email: ceezde@nus.edu.sg

Der-Horng Lee
Department of Civil & Environmental Engineering,
National Unversity of Singapore
Singapore 117576, Singapore
Email: dhl@nus.edu.sg

Word count: 4516 words + 4 table $\times$ 250 + 5 figures = 5516 words

Submission Date: 1st Aug 2019

ABSTRACT

Globally, many metropolitan areas tend to be promoting the multi-modal public transit which interconnects different modes to move large quantities of commuters in the urban area. Therefore, the accurate short-term passenger flow prediction can contribute to improve the reliability, responsiveness and the service quality of the multi-modal public transit. In this paper, we propose an end-to-end deep learning framework, called *ST-Pseudo3D Net* to collectively predict different types of public transit passenger flow at each region of city in the near future. The framework employs the deep Pseudo-3D residual architecture to model the network-wide spatial-temporal correlations among different types of passenger flow. Based on smart card data collected from Singapore's multi-modal public transit, considering metro passenger flow, bus passenger flow as well as the transfer passenger flow between these metro and bus, totally we construct 7 different types of passenger flow to be collectively predicted, where the experiments results demonstrate that the proposed model synthetically outperforms other baselines. To the best knowledge of the authors, this is the first attempt to investigate the integrated prediction for multi-modal passenger flow leveraging deep learning techniques.

Keywords: public transit, deep learning, 3DCNN, passenger flow, multi-modal

INTRODUCTION

The core concept of a multi-modal public transit system is to provide passengers with reliable, convenient, highly connective and integrated services using different effective transportation modes. Normally, an integrated multi-modal public transport system mainly consists of an urban metro system and a bus system, where these two subsystems interconnect collaboratively to guarantee that the urban is functioning smoothly and properly. However, many public transport systems in global metropolitan areas are still facing a lot of challenges, especially when the urban dynamics are becoming more and more active. The accurate short-term passenger flow prediction is one of the critical solutions that can provide significant improvements at the operation and management level. Specifically, the operators might pointedly adjust the real-time operation scheme and allocate the resources on demand according to the prediction for the passenger demand in the near future. Additionally, if any anomaly pattern was detected from the prediction, then relevant agencies could timely implement the responses, which contributes is of great significance to public safety.

In fact, there are many related research that are focusing on predicting short-term passenger/traffic flow in the past years, where the methodologies that have been utilized range from some traditional statistical method to a batch of advanced deep learning models. Especially, leveraging deep learning techniques to solve transportation related problems has become the recent main stream, which has also demonstrated that the deep learning based techniques do have a number of superiorities in modelling transportation data.

Among these related research, the prediction for traffic flow or traffic speed on road using vehicle trajectory data or traffic flow data accounts for a relatively larger proportion. For instance, Ma et al. (*1*) proposed a deep learning framework to predict the traffic congestion using GPS trajectory data from taxi. The framework is formed by combining Restricted Boltzmann Machine (RBM) and Recurrent Neural Network (RNN), which shows effectiveness in the predicting traffic congestion in large-scale network. Lv et al. (*2*) firstly utilized a stacked autoencoder (SAE) based model to forecast the short-term traffic flow, through which the complex spatial-temporal features of traffic data could be well captured and abstracted. Polson and Sokolov (*3*) developed a hybrid deep learning architecture for short-term traffic prediction using road sensor data. This framework contains a liner vector autoregressive model for detecting the spatial-temporal correlations among predictors, of which the findings will be fed into another neural network with a deep structure to further model the nonlinear dependencies. Cui et al. (*4*), Zhao et al. (*5*), Fu et al. (*6*), Yu et al. (*7*) and Wang et al. (*8*) investigated the use of Long Short Term Memory (LSTM) based models to predict short-term traffic flow/speed, as LSTM based models can effectively learn relatively long-range temporal dependencies.

Compared with the short-term prediction for link/network traffic based on trajectory data, leveraging deep learning techniques to forecast passenger flow/demand in public transit system has received less attention from researchers. For bus system, Liu and Chen (*9*) combined SAE and Deep Neural Network (DNN) forming a hierarchical hybrid framework to predict short-term bus passenger flow. Bai et al. (*10*) proposed a multi-pattern deep fusion model to predict short-term bus passenger flow, where the Affinity Propagation (AP) algorithm is applied to identify different patterns from passenger flow and then the Deep Belief Network (DBN) is fused in each pattern to obtain the abstract representation. When it comes to metro system, Ma et al. (*11*) developed a hybrid architecture to predict large-scale metro ridership by taking advantages of both CNN and Bi-directional LSTM. Liu et al. (*12*) proposed a deep learning based framework that incorporates both LSTM module and manually designed features which is able to model the spatial-temporal

characteristics of metro passenger flow. Furthermore, Zhang et al. (*13*) designed a deep residual CNN based framework to predict crowd flows (inflow and outflow) at each region in the city area, which has been validated on taxi trajectory data from Beijing and bike sharing data from New York City.

As reviewed and discussed above, although the existing research on the related topics have already gain some achievements in both theory and practice, some limitations and unexplored gaps are still observed. According to our investigation, for the short-term passenger/traffic flow prediction related tasks, all the proposed frameworks can only make prediction for a single type of flow (e.g. metro passenger, bus passenger, taxi), without considering the mixed passenger flows in the multi-modal transport system. The motivations of collectively predicting the mixed passenger flow in a large-scale multi-modal public transit are summarized as follows:

- Many multi-modal public transit systems around the world bring together both buses and metro to move people from place to place in the city area. Therefore, the investigating merely on a single mode (e.g. bus or metro) could not comprehensively characterize the overall pattern of the multi-modal public transit.

- Normally, the metro system in the urban area forms the backbone network, which is mainly served for mass and long distance transit while the bus system can further complement the metro system and cover the area devoid of metro service as a local auxiliary feeder solution. Thus, incorporating the bus passenger prediction into metro passenger prediction would not only broaden the scope of the network to larger-scale, but also increase the density of the network that allow the prediction to be made in a finer resolution.

- The highly integration of different modes of transportation in a multi-modal public transit is imperative for attracting passengers and providing them with seamless services. Many commuters use more than one mode of transportation to complete the journey, where the whole journey consists of several different boarding, alighting and transfer activities. Therefore, if different types of passenger flow (e.g. bus boarding demand, metro alighting demand, etc.) at each region in the city could be collectively predicted by a single integrated framework, it would provide comprehensive decision supports for the relevant agencies and operators.

- Many cities are promoting the use of public transport to embrace the sustainability and safety. We have witnessed that the share of usage of public transport have been increasing over the past years in many cities. In some cases, such as Hongkong, Singapore, London, the trips carried by public transport account for a larger proportion. Therefore, compared with other types of traffic flow, we believe that the mixed passenger flows in large-scale multi-modal public transit might be a better reflection of the citywide mobility dynamics.

Taking advantages of Singapore's distance-based AFC system, the smart card data record both passengers' boarding and alighting activities and more importantly it integrates both bus and metro trips in the same data frame, which allows us to put metro passenger flow and bus passenger flow together to conduct analysis. In this paper, we propose an end-to-end Pseudo-3D Convolutional Neural Network based framework that can collectively predict totally 7 different types of passenger flow at each region of the city area in the near future. The key contributions of this research summarized as follows:

- We propose a Pseudo-3D Convolutional Neural Network (Pseudo-3DCNN) based model to predict the public transport passenger flow in a network-wide region level.
- We take metro passenger flow, bus passenger flow as well as the transfer flow between metro system and bus system together into consideration instead of merely predicting a single type of passenger flow.
- The feasibility and effectiveness of our proposed model has been demonstrated on real-world data collected in Singapore.

The remainder of this paper is organized as follows. Section 3 elaborates all the preliminaries and methodologies. Section 4 describes the experiment settings and analyzes the results. Lastly, section 5 summarizes the main findings and discusses the future directions.

METHODOLOGY

In this section, we first give a detailed statement for all relevant definitions and concepts in this mixed passenger flow prediction task, and then we introduce the core methodologies in this research.

Prediction Scope

In this research, we predict the mixed passenger flow at each region of the city, which means that the prediction is performed in region level(we term as network-wide region level) instead of stop/station level. We first uniformly partition the entire city area into grid-like form with shape of $I \times J$, hence each of the grid refers to a region in this study. Then, we assign the bus stops and metro stations to corresponding regions based on the longitude and latitude coordinates of the stops/stations. The assignment procedure is illustrated as follows:

$$G_{i,j} = \{s_{bus} \in BS | s^c_{bus} \in (i,j)\} \cup \{s_{metro} \in MS | s^c_{metro} \in (i,j)\} \tag{1}$$

Where BS and MS denote the set of all the bus stops and metro stations, respectively. s_{bus} refers to an element of set BS, a particular bus stop, and s^c_{bus} is the coordinate of bs. Similarly, s_{metro} represents a metro station in the set of BM, of which the coordinate is denoted by s^c_{metro}. Therefore, $G_{i,j}$ is the collection of all bus stops and metro stations that are located in grid (i,j).

Mixed Passenger Flow

Totally, we construct 7 different types of passenger flow/demand by simultaneously taking both metro and bus system into consideration. The first 4 types of passenger flows are relatively straightforward, which are *bus boarding* demand, *bus alighting* demand, *metro boarding* demand and *metro alighting* demand, respectively.

We first obtain 4 sets that contain the corresponding feasible trips during time interval t, which can be defined as:

$$D^{bus-board}_t = \{\forall d \in D | d^{bus-board-time} \in t\} \tag{2}$$

$$D^{bus-alight}_t = \{\forall d \in D | d^{bus-alight-time} \in t\} \tag{3}$$

$$D^{metro-board}_t = \{\forall d \in D | d^{metro-board-time} \in t\} \tag{4}$$

$$D_t^{metro-alight} = \{\forall d \in D | d^{metro-alight-time} \in t\} \tag{5}$$

Where D is the collection of all the trips. d refers to a record in D representing a complete trip made by a passenger and this trip might consists of a sequence of subtrips (e.g. a bus trip followed by a metro trip). Specifically, in a trip d, if there is a bus boarding activity that takes place during time interval t, then d will be added to set $D_t^{bus-board}$ (see Equation 2) or if there is a metro alighting activity that takes place during time interval t', then d will be also added to set $D_{t'}^{metro-alight}$ (see Equation 5). The same rules applies for the other sets. Subsequently, the *bus boarding* demand, *bus alighting* demand, *metro boarding* demand and *metro alighting* demand can be obtained according to the following logic:

$$x_{t,i,j}^{bus-board} = Card(\{\forall d_t \in D_t^{bus-board} | d_t^{bus-board-stop} \in G_{i,j}\}) \tag{6}$$

$$x_{t,i,j}^{bus-alight} = Card(\{\forall d_t \in D_t^{bus-alight} | d_t^{bus-alight-stop} \in G_{i,j}\}) \tag{7}$$

$$x_{t,i,j}^{metro-board} = Card(\{\forall d_t \in D_t^{metro-board} | d_t^{metro-board-station} \in G_{i,j}\}) \tag{8}$$

$$x_{t,i,j}^{metro-alight} = Card(\{\forall d_t \in D_t^{metro-alight} | d_t^{metro-alight-station} \in G_{i,j}\}) \tag{9}$$

Take bus passenger for instance (see Equation 6 and Equation 7), $D_t^{bus-bboard}$ represents the set of all the trips that have a bus boarding activity during time interval t and similarly $D_t^{bus-alight}$ denotes the collection of all trips that have an alighting activity during time interval t. d_t is a satisfied trip in $D_t^{bus_board}$, of which the bus boarding stop is denoted by $d_t^{bus-board-stop}$. Hence, $\{d_t \in D_t | d_t^{bus-board-stop} \in G_{i,j}\}$ is the set consisting of all the trips that have a boarding activity in grid (i, j) during time interval t. Lastly, we take the cardinality of this set to obtain the exact bus boarding demand (total number of bus boarding passengers) in grid (i, j) during time interval t, which is denoted by $x_{t,i,j}^{bus-board}$. Thus, $x_{t,i,j}^{bus-board}$ is a 2D flow matrix, shown in in Fig 1. The same rule applies for the other types of passenger flow (see Equation 7, 8 and 9).

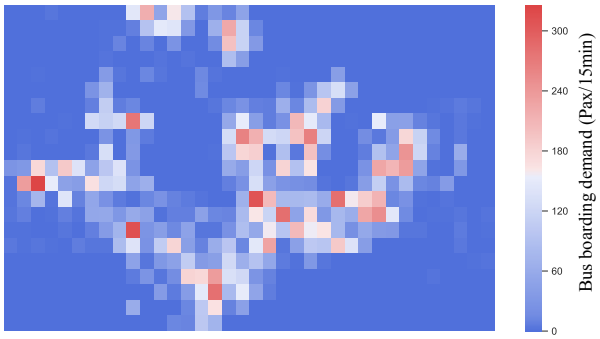


FIGURE 1 : Flow matrix (bus boarding)

In fact, the transfer activities are very common in a mature multi-modal public transit, where there are lots of commuters that would take more than 1 mode of transportation to complete the journey. Therefore, in addition to the 4 types of passenger flow mentioned above, we construct another 3 different types of passenger flow by considering the transfer between different modes,

which are *bus-to-metro* demand, *metro-to-bus* demand and *bus-to-bus* demand, respectively. The transfer related flows can be obtained by performing the following operations:

$$x_{t,i,j}^{bus-metro} = Card(\{\forall d_t \in \{D_t^{bus-alight} \cap D_t^{metro-board}\} \mid d_t^{bus-alight-station} \in G_{i,j} \wedge d_t^{metro-board-station} \in G_{i,j}\}) \quad (10)$$

$$x_{t,i,j}^{metro-bus} = Card(\{\forall d_t \in \{D_t^{metro-alight} \cap D_t^{bus-board}\} \mid d_t^{metro-alight-station} \in G_{i,j} \wedge d_t^{bus-board-station} \in G_{i,j}\}) \quad (11)$$

$$x_{t,i,j}^{bus-bus} = Card(\{\forall d_t \in \{D_t^{bus-alight} \cap D_t^{bus-board}\} \mid d_t^{bus-alight-station} \in G_{i,j} \wedge d_t^{bus-board-station} \in G_{i,j}\}) \quad (12)$$

Where $x_{t,i,j}^{bus-metro}$ refers to the *bus-to-metro* demand in grid (i, j) during time t. To be specific, $x_{t,i,j}^{bus-metro}$ is the total number of passengers who alight from a bus in grid (i, j) during time interval t and then make a transfer to board a metro in the same grid during the same time interval. Similarly, $x_{t,i,j}^{metro-bus}$ and $x_{t,i,j}^{bus-bus}$ represent the corresponding *metro-to-bus* demand and *bus-to-bus* demand, respectively.

At a specific time interval t, the mixed passenger flow observation in all regions $I \times J$ could be structured as a tensor $X_t \in \mathbb{R}^{7\times I\times J}$ by concatenating different types of passenger flow matrix along the channel dimension, where $(X_t)_{0,i,j} = x_{t,i,j}^{bus-board}$, $(X_t)_{1,i,j} = x_{t,i,j}^{bus-alight}$, $(X_t)_{2,i,j} = x_{t,i,j}^{metro-board}$,..., $(X_t)_{6,i,j} = x_{t,i,j}^{bus-bus}$. The objective of this framework is to simultaneously predict different types of passenger flow at each region in the near future according to the mixed flow tensors from the last few short-term periods. The input of the model is a tensor by concatenating a sequence of historical observations, $[X_{t-T}, .., X_{t-1}, X_t] \in \mathbb{R}^{7\times T\times I\times J}$, where T refers to the number of historical time intervals that are taken into consideration. Besides, the predicted output is denoted by $\hat{Y} \in \mathbb{R}^{7\times T'\times I\times J}$, where T' is the number of predicted steps ahead.

3D Convolutional Neural Network

Convolutional Neural Network (CNN) have been widely applied for many tasks and shown a couple of superiorities, owing to its great ability of modelling spatial structures. In transportation related prediction tasks, not only the spatial dependencies should be effectively captured, but also the temporal correlations are of great importance. Unlike traditional 2DCNN that the convolution operations are performed only spatially, 3DCNN can perform the convolution operations spatial-temporally, which have exhibited significant effectiveness in some video/motion prediction tasks (*14*); (*15*). Fig 2 demonstrates how 3D convolution performs on the tensor.

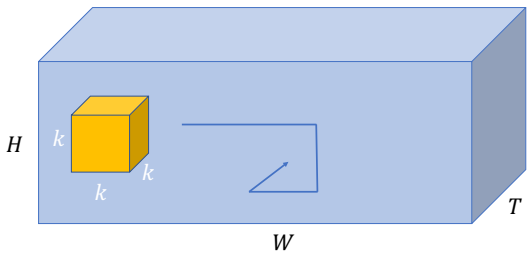


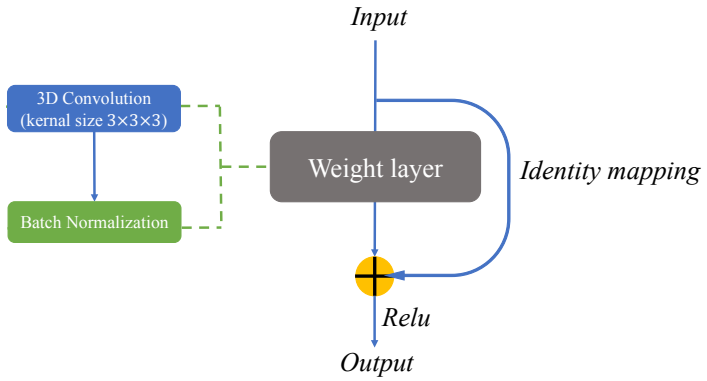
FIGURE 2 : **The illustration of 3D convolution**

As illustrated in Fig 2, 3DCNN applies a 3 dimensional filter to the tensor and the filter moves in 3 dimensions to aggregate the nearby spatial-temporal features.

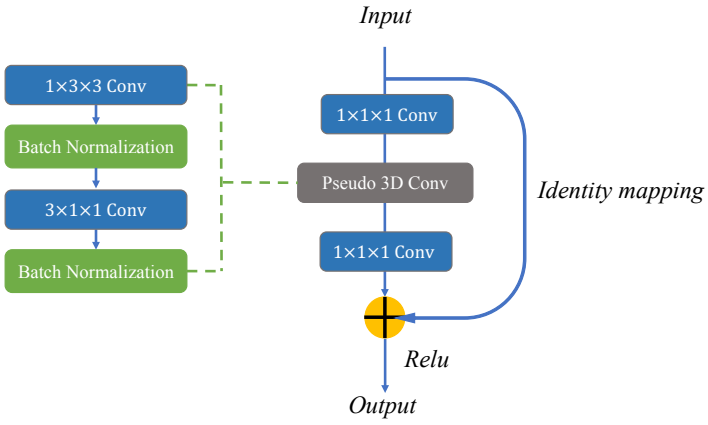
Pseudo-3DCNN Residual Learning

3D Residual Learning

The local spatial patterns can be continuously captured through the sliding convolution operations. Since many commuters are likely to take public transport for relatively longer-distance trips, consequently the non-local or long-range spatial-temporal patterns might be more critical. Normally there are several different ways to capture the long-range dependencies, such as downsampling (e.g. pooling), applying fully connected layers as well as stacking multiple convolutional layers. Given that applying downsampling strategies (e.g. pooling) might lead to the loss of information and using fully connected layers would not only significantly increase the number of parameters but also miss the spatial information. Hence, inspired by the deep residual learning mechanism (*16*), we stack multiple residual units forming a deeper network in order to model the long-range citywide dependencies. Fig 3a illustrates the structure of a residual unit.



(a) **The structure of a 3DCNN residual unit**



(b) **The structure of a Pseudo-3D residual unit**

FIGURE 3 : The structure of the residual unit

The identity mapping in the residual unit, defined in Equation 13, is to directly transmit the activation from previous layers to deeper layers which contributes to ease the vanishing and

exploding gradient problems when training a deep network.

$$X^{l+1} = X^l + F(X^l) \tag{13}$$

Where X^{l+1} and X^l are the output and input of the l-th residul unit. Besides, in this research, F refers to the operation of 1 layer 3DCNN.

Pseudo-3D Residual Learning

However, compared with 2DCNN, training 3DCNN is very computationally expensive meanwhile the model size also has a quadratic growth (*17*). Thus, a relatively light weight framework with great prediction performance is of better efficiency and scalability, especially when being applied or deployed in practice. Fortunately, Qiu et al. (*17*) proposed a Pseudo-3DCNN residual architecture that can learn spatial-temporal video representation with greater efficiency by simplifying the normal 3DCNN which is computational expensive. Hence, we adopt the concept of the Pseudo-3DCNN to construct our prediction framework. In pseudo-3DCNN (*17*), a normal 3DCNN operation with kernal size $3 \times 3 \times 3$ could be decomposed to a combination of a 2D spatial CNN with kernal size $1 \times 3 \times 3$ and a 1D temporal CNN with kernal size $3 \times 1 \times 1$. Therefore, we first replace the normal 3DCNN in our original residual units (Fig 3a) with two consecutive convolution operations, which are Spatial-CNN ($1 \times 3 \times 3$) and Temporal-CNN ($3 \times 1 \times 1$), respectively. Besides, we add another two $1 \times 1 \times 1$ convolutions at both ends of the unit for reducing and rescaling the tensor dimensions, which helps to further reduce the computational costs. The revised residual unit is illustrated in Fig 3b.

Fig 4 presents the schematic architecture of our proposed model which we term as *ST-Pseudo3D Net*.

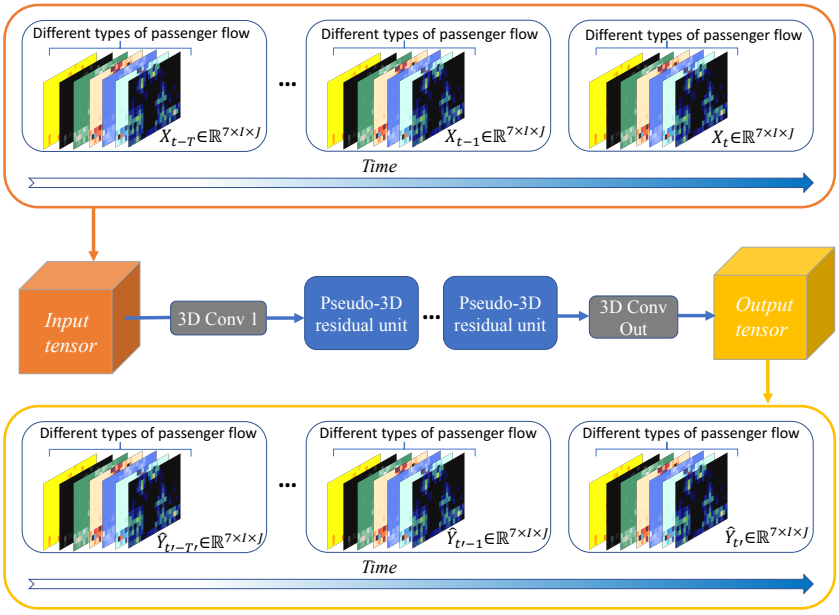


FIGURE 4 : The schematic architecture of the proposed framework

In the framework, the input tensor $X \in \mathbb{R}^{7\times T\times I\times J}$ is first transformed by *conv-1* (3DCNN with kernal size $3 \times 3 \times 3$) to generate the internal feature maps which will be subsequently fed into

the stacked residual units followed by a final output 3DCNN layer to obtain the predicted result $\hat{Y} \in \mathbb{R}^{7\times\ \mathrm{T}'\times \mathrm{I}\times \mathrm{J}}$. The entire computation flow could be written as:

$$a^l = ReLU(BN(W^l * X + b^l)),\ \ l = 1 \tag{14}$$

$$a^l = a^{l-1} + F(a^{l-1};\theta^l),\ \ \forall\, l = 2,..,L-1 \tag{15}$$

$$\hat{Y} = W^l * a^{l-1} + b^l,\ \ l = L \tag{16}$$

Equation 14 represents the computation in the first 3DCNN layer $conv-1$, where $*$ denotes the 3D convolution operation, a^l is the output of layer l. Both W^l and b^l are trainable parameters at layer l. *BN* refers to the batch normalization operation (*18*) which helps to accelerate the training by reducing internal covariateshift and *ReLU* is the non-linear activation function, $ReLU(x) = max(0,x)$. Equation 15 outlines the computation flow through the stacked residual units, where a^l and a^{l-1} are the activation from layer l and $l-1$, respectively. F is the residual function, which is a sequence of convolution operations in our research (see Fig 3b). Finally, the output layer is illustrated in Equation 16, where $*$ denotes the 3D convolution operation, W^l and b^l are both learnable parameters. Notably, we just use the direct linear output from $W^l * a^{l-1} + b^l$ as the final predicted output without applying any non-linear activation at the last layer.

EXPERIMENTS AND RESULTS

In this section, we first introduce the experimental settings and then use Singapore as the case to evaluate the performance of the proposed framework.

Data Preparation

The data set used for training and evaluation is collected from Singapore's AFC system (19th March 2012 to 25th March 2012). Singapore adopts the distance-based AFC system and integrates both bus trips and metro trips in the same data frame which allows us to conduct research on the multi-modal related issues. The exact study area of this particular research starts from 103.678 to 104.014 in longitude and from 1.255 to 1.448 in latitude, which is uniformly partitioned into 36×21 grids. Thus, the size of each grid equals to $0.0093\ lng \times 0.992\ lat$, which is about $1.03\ km \times 1.02\ km$. Additionally, in the experiment, we use the historical observations in the last 90 minutes by a 15-min interval to predict the passenger flows in the next 1 hour by a 15-min. Notably, the different types of passenger flow might have significant differences in magnitude. For example, the metro boarding demand tend to be much larger than the bus-to-bus transfer demand. Therefore, we first apply the Min-Max normalization (illustrated in Equation 17) to scale the different types of passenger flow into the same range ([0,1]) in order to improve training efficiency.

$$z = \frac{x - min(x)}{max(x) - min(x)} \tag{17}$$

Training Settings

The randomly split is applied on the data set where 85% of the data are used for training while another 15% are used for evaluation. The batch size is 32 and the maximum training epochs is fixed to 500. Additionally, the early-stopping mechanism is introduced to continuously monitor the loss on the validation set in order to avoid overfitting issue. Root Mean Square Error (RMSE) is the evaluation metric for the proposed model, given by:

$$z = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(Y^{(n)} - \hat{Y}^{(n)})^2} \tag{18}$$

The model is optimized by Adam optimization algorithm (*19*) with learning rate of 0.001. All experiments are conducted on our Google Cloud Platform, of which the detail configuration are listed in Table 1.

TABLE 1 : **Experiment environment**

Item	Configuration
OS	GNU/Linux server
CPU version	Intel (R) Xeon (R) CPU @2.20GHz
RAM	15GB
GPU	2 × NVIDIA Tesla T4
Pytorch version	1.1.0
CUDA version	9.0

Results Analysis

We compare our proposed model *ST-Pseudo3D Net* with a group of state-of-the-art benchmarks, including DNN, RNN, LSTM (*20*), GRU (*21*), 2DCNN with residual connection (Res2DCNN) and 3DCNN with residual connection (Res3DCNN). In order to show the results in a more meticulous manner, we decompose the overall prediction results into several groups with respect to the type of the passenger flow, which are presented in Table 2-4. The best score in each passenger flow type is highlighted in bold.

TABLE 2 : **Results comparison for bus passenger flow**

Model	*RMSE*	
	Bus boarding demand	***Bus alighting demand***
DNN	12.51	12.80
RNN	14.33	14.71
LSTM	13.65	14.35
GRU	13.84	14.51
Res2DCNN	10.56	11.93
Res3DCNN	9.57	10.73
ST-Pseudo3D	**9.15**	**10.55**

TABLE 3 : **Results comparison for metro passenger flow**

Model	*RMSE*	
	Metro boarding demand	***Metro alighting demand***
DNN	15.93	**19.62**
RNN	16.49	20.84
LSTM	**15.46**	20.10
GRU	15.84	20.29
Res2DCNN	17.91	21.82
Res3DCNN	15.82	20.26
ST-Pseudo3D	15.98	20.33

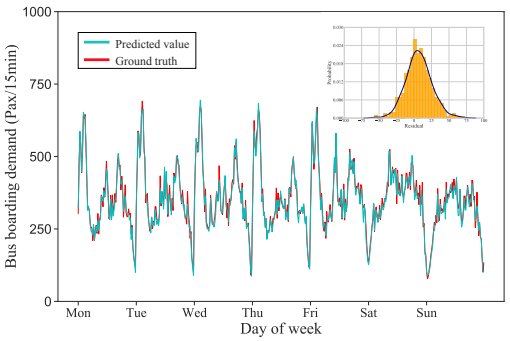
TABLE 4 : **Results comparison for transfer passenger flow**

Model	*RMSE*		
	Bus-to-bus demand	***Bus-to-metro demand***	***Metro-to-bus demand***
DNN	7.87	10.24	9.93
RNN	7.99	10.98	10.46
LSTM	7.76	10.69	10.07
GRU	7.81	10.78	10.26
Res2DCNN	7.00	9.68	9.03
Res3DCNN	6.41	**7.81**	7.45
ST-Pseudo3D	**6.33**	7.89	**7.38**

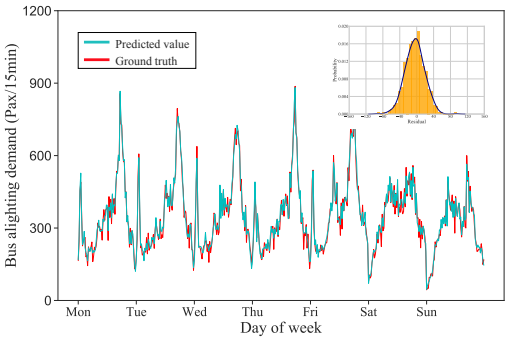
As shown in Table 2 - 4, our proposed model *ST-Pseudo3D Net* outperforms other baselines in most of the above-listed prediction tasks. Specifically, for bus passenger flow prediction, CNN based models perform better than DNN and other RNN based models, owing to the effectiveness in extracting spatial features, where the proposed *ST-Pseudo Net* achieve the best results for both *bus boarding demand* prediction and *bus alighting demand* prediction, which are relatively 4.4% up to 36.1% better than other baselines in predicting *bus boarding demand* and 1.7% up to 28.3% better than other baselines in predicting *bus alighting demand* at each region of the city. When it comes to metro passenger flow, we observe that the DNN and RNN based models generally perform better than CNN based models, where our proposed model slightly lags behind LSTM by 3.3% and DNN by 3.5% in predicting *metro boarding demand* and *metro alighting demand*, respectively. However, for transfer passenger flow, the CNN based models, especially 3DCNN based models show significant superiorities again, where 3DCNN based models are relatively 8.4% up to 19.7%, 18.5% up to 28.9% and 18.3% up to 29.4% better than other baselines in predicting *bus-to-bus demand*, *bus-to-metro demand*, *metro-to-bus demand*, respectively. Notably, although Res3DCNN almost perform as good as the proposed ST-Pseudo3D, the computation cost and memory demand of Res3DNN are much higher than ST-Pseudo3D. Specifically, in the experiment, we keep Res3DCNN and ST-Pseudo3D with the same depth (stacking 16 residual units) and the total number of trainable parameters of ST-Pseudo3D is 46.1% less than Res3DCNN. Therefore, we believe that the proposed framework is superior to Res3DCNN in terms of efficiency and practicability.

Case Analysis

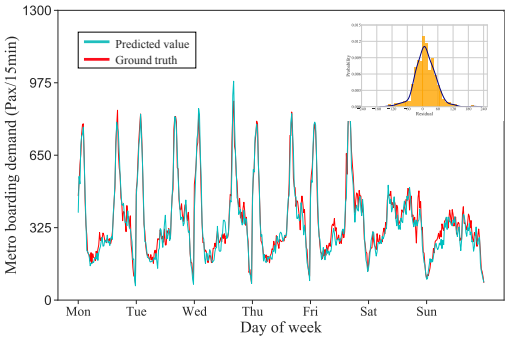
In order to show the prediction results of the proposed framework intuitively, a specific region is selected for detailed performance demonstration. The selected region is located in Clementi area which is one of the major residential towns in Singapore. The selected region covers 1 metro station (Clementi station), 1 integrated bus interchange hub and several bus stops, which makes it an ideal spot for results analysis. Both the true and predicted demand are plotted in Fig 5 according to different types of passenger flow.



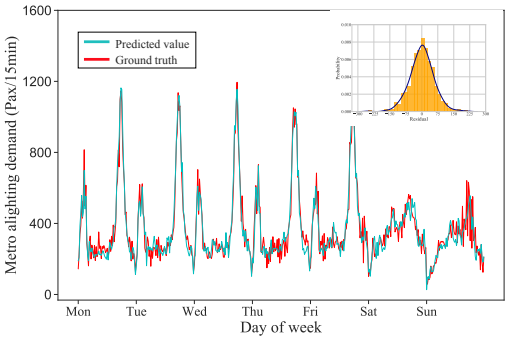
(a) **Bus boarding demand comparison**



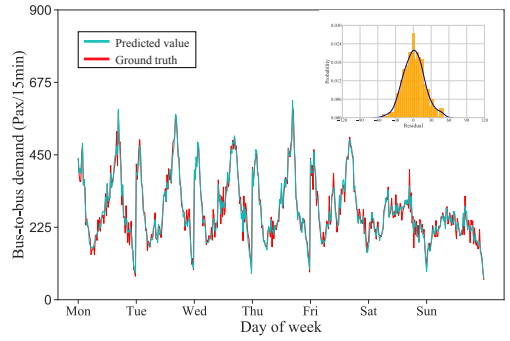
(b) **Bus alighting demand comparison**



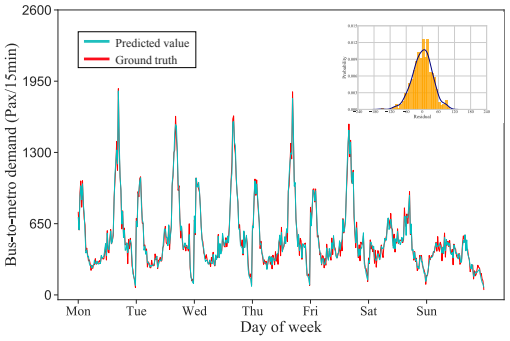
(c) **Metro boarding demand comparison**



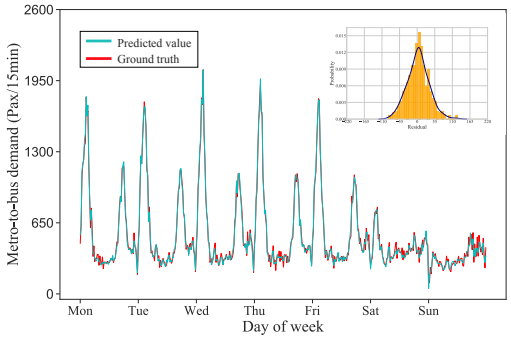
(d) **Metro alighting demand comparison**



(e) **Bus-to-bus demand comparison**



(f) **Bus-to-metro demand comparison**



(g) **Metro-to-bus demand comparison**

FIGURE 5 : **Comparisons of the ground truth and predicted passenger flow**

As illustrated in Fig 5a - 5g, the proposed model can make relatively reliable prediction for different types of passenger flow. Furthermore, since different prediction tasks have different numerical scales, so a small *RMSE* or *MAE* value does not necessarily mean that the model is good enough. Therefore, we further analyze the residual of the prediction, which are plotted in the upper right corner of each figure. We observe that the residuals of our prediction for different types of passenger flow basically follow normal distributions with zero mean, which indicates that there are no significant non-random patterns in the residuals.

CONCLUSION

In this paper, we propose an end-to-end deep learning based framework to collectively predict network-wide multi-modal passenger flows in a large-scale public transit. Considering metro passenger flow, bus passenger flow as well as transfer flow between metro and bus system, we totally construct 7 different types of passenger flow to be collectively predicted. The framework is based on deep 3DCNN residual architecture which can effectively extract spatial-temporal features. Moreover, inspired by Pseudo-3D, we replace the traditional 3D convolution operation in the residual units with two consecutive convolution operations (a Spatial-CNN followed by a Temporal-CNN), which significantly reduce the number of parameters and the computational cost. Additionally, the model is validated by real-world data collected from Singapore's public transit and the experiment results demonstrate that the proposed framework is superior to other baselines in terms of the accuracy and practicability.

In the future, we plan to train the model on some larger data sets in order to further improve its generalization ability. In addition, combining with the theories from graph learning domain, we will also try to explore the movement of passenger flows in the network.

ACKNOWLEDGEMENT

We would like to show our gratitude to Land Transport Authority of Singapore for providing the data.

REFERENCES

[1] Ma, X., H. Yu, Y. Wang, and Y. Wang, Large-scale transportation network congestion evolution prediction using deep learning theory. *PloS one*, Vol. 10, No. 3, 2015, p. e0119044.

[2] Lv, Y., Y. Duan, W. Kang, Z. Li, F.-Y. Wang, et al., Traffic flow prediction with big data: A deep learning approach. *IEEE Trans. Intelligent Transportation Systems*, Vol. 16, No. 2, 2015, pp. 865–873.

[3] Polson, N. G. and V. O. Sokolov, Deep learning for short-term traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, Vol. 79, 2017, pp. 1–17.

[4] Cui, Z., R. Ke, and Y. Wang, Deep Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction. *CoRR*, Vol. abs/1801.02143, 2018.

[5] Zhao, Z., W. Chen, X. Wu, P. C. Chen, and J. Liu, LSTM network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, Vol. 11, No. 2, 2017, pp. 68–75.

[6] Fu, R., Z. Zhang, and L. Li, Using LSTM and GRU neural network methods for traffic flow prediction. In *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, IEEE, 2016, pp. 324–328.

[7] Yu, H., Z. Wu, S. Wang, Y. Wang, and X. Ma, Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors*, Vol. 17, No. 7, 2017, p. 1501.

[8] Wang, J., R. Chen, and Z. He, Traffic speed prediction for urban transportation network: A path based deep learning approach. *Transportation Research Part C: Emerging Technologies*, Vol. 100, 2019, pp. 372–385.

[9] Liu, L. and R.-C. Chen, A novel passenger flow prediction model using deep learning methods. *Transportation Research Part C: Emerging Technologies*, Vol. 84, 2017, pp. 74–91.

[10] Bai, Y., Z. Sun, B. Zeng, J. Deng, and C. Li, A multi-pattern deep fusion model for short-term bus passenger flow forecasting. *Applied Soft Computing*, Vol. 58, 2017, pp. 669–680.

[11] Ma, X., J. Zhang, B. Du, C. Ding, and L. Sun, Parallel Architecture of Convolutional Bi-Directional LSTM Neural Networks for Network-Wide Metro Ridership Prediction. *IEEE Transactions on Intelligent Transportation Systems*, 2018.

[12] Liu, Y., Z. Liu, and R. Jia, DeepPF: A deep learning based architecture for metro passenger flow prediction. *Transportation Research Part C: Emerging Technologies*, Vol. 101, 2019, pp. 18–34.

[13] Zhang, J., Y. Zheng, and D. Qi, Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction. In *AAAI*, 2017, pp. 1655–1661.

[14] Ji, S., W. Xu, M. Yang, and K. Yu, 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, Vol. 35, No. 1, 2012, pp. 221–231.

[15] Tran, D., L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4489–4497.

[16] He, K., X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[17] Qiu, Z., T. Yao, and T. Mei, Learning spatio-temporal representation with pseudo-3d residual networks. In *proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 5533–5541.

[18] Ioffe, S. and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

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

[20] Hochreiter, S. and J. Schmidhuber, Long short-term memory. *Neural computation*, Vol. 9, No. 8, 1997, pp. 1735–1780.

[21] Cho, K., B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.