Data Science for Transportation (2024) 6:27
https://doi.org/10.1007/s42421-024-00112-2

RESEARCH

Deep Learning for Traffic Prediction and Trend Deviation Identification: A Case Study in Hong Kong

Xiexin Zou[1] · Edward Chung[1] · Hongbo Ye[1] · Haolin Zhang[2]

Received: 9 June 2024 / Revised: 24 September 2024 / Accepted: 24 September 2024

Abstract
This paper introduces a robust methodology for predicting traffic volume and speed on major strategic routes in Hong Kong by leveraging data from data.gov.hk and utilizing deep learning models. The approach offers predictions from 6 min to 1 h, considering detector reliability. By extracting hidden deep features from historical detector data to establish detector profiles and grouping detectors into clusters based on profile similarities, the method employs a CNN-LSTM prediction model for each cluster. The study demonstrates the model's resilience to detector failures, with tests conducted across failure rates from 1% to 20%, highlighting its ability to maintain accurate predictions despite random failures. In scenarios without failed detectors, the method achieves favorable performance metrics: MAE, RMSE, and MAPE for traffic volume prediction over the next 6 min stand at 5.17 vehicles/6 min, 7.64 vehicles/6 min, and 14.07%, respectively, while for traffic speed prediction, the values are 3.70 km/h, 6.32 km/h, and 6.33%. Considering a failure rate of approximately 6% in the Hong Kong dataset, in simulated scenarios with 6% failures, the model maintains its predictive accuracy, with average MAE, RMSE, and MAPE for traffic volume prediction at 5.24 vehicles/6 min, 7.81 vehicles/6 min, and 14.21%, and for traffic speed prediction at 3.87 km/h, 6.55 km/h, and 6.68%. However, the limitation of the proposed method is its potential to underperform when predicting rare or unseen scenarios, indicating the need for future research to incorporate additional data sources and methods to enhance predictive performance.

Keywords Traffic volume prediction · Traffic speed prediction · Deep learning · Clustering method · Detector profile

Introduction

Traffic congestion is a prominent issue in urban transportation systems, especially in densely populated cities. With its limited land and large population, Hong Kong has been dedicated to implementing various measures to optimize the transportation system and alleviate traffic congestion. One practical and cost-efficient method is intelligent traffic signal control systems, which enable dynamic adjustment of signal phases based on real-time traffic flow, reducing congestion (Zhao et al. 2011; Mandhare et al. 2018; Qadri et al. 2020). By incorporating short-term traffic variables prediction, such as traffic volume and speed, signal phases can be more accurately adjusted to adapt to near-future traffic changes. This proactive approach, such as extending the green light duration when an increase in traffic volume is predicted, effectively reduces waiting time, improves overall road capacity, and alleviates traffic pressure.

To collect data on traffic volume, speed, and occupancy, traffic detectors have been installed on all strategic routes in Hong Kong, with the installation completed by the end of 2020. Data-driven analysis and deep learning models can be used to leverage the vast traffic data collected to understand underlying traffic patterns and predict short-term traffic volume and speed (Lu et al. 2015; Kong et al. 2019). In recent years, deep learning (DL) methods have demonstrated significant advantages in predicting short-term traffic volume

✉ Edward Chung
edward.cs.chung@polyu.edu.hk

Xiexin Zou
xiexin.zou@connect.polyu.hk

Hongbo Ye
hongbo.ye@polyu.edu.hk

Haolin Zhang
haolin2002.zhang@connect.polyu.hk

[1] Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong, China

[2] Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China

Published online: 04 October 2024

and traffic speed (Tedjopurnomo et al. 2022; Lee et al. 2021; Yin et al. 2021). These methods excel in modeling complex nonlinear relationships. Some works focus on predicting traffic for individual stations or road sections. For instance, Xu et al. (2023) proposed a hybrid model that combines autoregressive fractionally integrated moving average and nonlinear autoregressive neural network models for traffic flow prediction. Lv et al. (2015) employed stacked autoencoders to predict short-term road traffic flow. Qu et al. (2021) utilized a stacked recurrent neural network (RNN) to extract features from historical time series and employed an autoencoder to capture contextual features like dates and weeks to assist in traffic speed prediction. Han et al. (2019) proposed DeepCluster, a method that applies k-means clustering to group all detectors and developed individual detector prediction models for each cluster. For the daily flow prediction of a single detector, Song et al. (2018) divided the historical daily data into groups using density-based clustering and trained a feedforward neural network model, GMDH, for each group. Long Short-Term Memory (LSTM) models and their variants, known for effectively capturing temporal dependencies and patterns in time series data, have been widely applied in traffic prediction studies (Khan et al. 2023; Jiang et al. 2021; Yuan and Li 2021). Abduljabbar and Dia (2021); Abduljabbar et al. (2021) developed LSTMs for traffic volume/speed prediction on a single highway and validated their performance on multiple highways using a public dataset. The results showcased the outstanding performance of LSTM in this task. The results also highlighted the challenges of traffic volume prediction due to its noisier nature compared to speed measurements.

However, such single-detector input–output prediction models did not consider the accessibility between adjacent road sections or stations, thus overlooking spatial dependencies. Hence, Ma et al. (2022) utilized a hybrid spatial-temporal feature selection algorithm (STFSA) to identify the four monitoring points with the highest correlation to the prediction point, constructing a spatiotemporal matrix of traffic speed. This matrix was used as input features for a prediction model based on Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) to forecast the traffic speed at a single prediction point. Numerous studies provided a prediction model outputting predictions for all monitoring points along a single road or a few connected roads. Tan et al. (2016) proposed a new short-term traffic flow prediction method based on Dynamic Tensor Completion (DTC), representing traffic data as dynamic tensor patterns and estimating future traffic using weekly, daily, and close historical data under low-rank assumptions. However, this approach is not suitable for larger-scale traffic networks due to the large size of tensors and the amount of missing data. Zhang et al. (2020) decomposed traffic speed time series data into different frequency components using Ensemble Empirical Mode Decomposition (EEMD) and constructed a three-dimensional tensor $X \in R^{(n_1,n_2,n_3)}$, where n_1, n_2, n_3 represent the number of historical time slices, decomposed components, and detectors, respectively. They then employed 3D CNN for speed prediction.

Combining CNN and LSTM has emerged as a popular solution in traffic prediction (Zang et al. 2019; Rajalakshmi and Ganesh Vaidyanathan 2022; Ma et al. 2020). This combination allows CNN to capture spatial dependencies and LSTM to capture temporal relationships. Cao et al. (2020) proposed a CNN-LSTM model that utilized historical traffic speed from upstream and downstream and the predicted road to predict traffic speed in the Cross-Harbour Tunnel from Hong Kong Island to Kowloon. Ke et al. (2020) considered the volume impact in speed prediction and used CNN with historical traffic volume and speed data to predict traffic speed. Furthermore, attention mechanisms are incorporated to capture crucial contextual information and improve predictions. Liu et al. (2018) introduced attention CNN, which used a three-dimensional data matrix constructed from traffic flow, speed, and occupancy to predict traffic speed. Wu et al. (2018) proposed a traffic flow prediction model that integrated an attention mechanism, CNN, and RNN. This model utilized historical traffic speed and flow data while incorporating recent traffic flow and the preceding day and week to account for the multi-periodicity of traffic flow. Zheng et al. (2020) developed a Conv-LSTM module with the attention mechanism to extract spatial and short-term temporal features and a bidirectional LSTM module to capture daily and weekly cyclical features for short-term traffic flow prediction.

However, CNN is designed for fixed-size image inputs, which poses a challenge when dealing with irregular traffic data. When CNN is applied to traffic prediction, the preprocessing of traffic data is required to transform it into inputs of fixed dimensions. Lv et al. (2018) proposed a speed prediction method based on a road network embedded convolution method and RNN. They constructed a matrix that provided information about the connectivity between adjacent road sections and utilized a combination of search operations and convolution to capture spatial relationships among connected road sections. However, preprocessing becomes challenging for large-scale traffic networks and may result in information loss or introduce noise, thus impacting the model's performance. Thus, Graph Neural Networks (GNN) have been introduced to handle irregular inputs and learn spatial relationships (Diao et al. 2019; Song et al. 2020; Zhou et al. 2022).

For network-level predictions, various studies constructed traffic graphs based on traffic topology. For instance, Xie et al. (2019) emphasized that the connectivity of road sections is the fundamental cause of traffic congestion propagation. Consequently, they represented multiple road sections

as graphs based on connectivity and proposed SeqGNN to handle graph sequences for prediction. Wang et al. (2023) proposed PFNet, which utilized graph embedding techniques to capture spatial relationships and employed a Deep Multi-View Sequence Encoder (DMVSE) to capture time correlation from monthly periodicity, weekly periodicity, and closeness. The spatial correlations among different traffic network nodes are local and non-local. To address this, Fang et al. (2019) proposed the Global Spatial-Temporal Network (GSTNet), which considered both local spatial correlations among adjacent stations and global spatial correlations. Lv et al. (2021) considered road connectivity and incorporated historical traffic pattern correlations and local area functionality similarity to construct traffic graphs. They proposed a traffic flow prediction model based on multi-graph convolution and GRU. Moreover, the interaction between road sections is influenced by factors such as the natural environment and socio-economic factors. Xia et al. (2024) extracted static features such as spatial distance and dynamic features from the most recent traffic data using a multi-head attention network and dynamic node embedding to capture these dynamics. They integrated dynamic graph convolution modules and LSTM to achieve accurate traffic flow prediction. Chen and Chen (2022) introduced a graph learning model based on reinforcement learning. They initially constructed a traffic graph based on the accessibility of road sections. Then, they utilized reinforcement learning to extract nonlinear spatiotemporal dependencies between stations, adaptively generating the graph adjacency matrix. However, GNN-based approaches need to address challenges such as handling large-scale graph structures, high-dimensional features, and the issue of gradient vanishing during information propagation. The computational cost of processing large-scale graph structures is high and may not satisfy real-time requirements.

To predict traffic volume and speed for detectors on major strategy routes in Hong Kong, it is essential to consider the detectors' reliability, given the average failure rate of 6% in Hong Kong. Handling the resulting missing data is of utmost importance. This paper proposes an approach that utilizes data from data.gov.hk to develop a robust traffic volume/speed prediction model resilient to detector failures. Hidden deep features within each detector's historical daily traffic volume data are extracted to derive the corresponding detector profiles. These detectors are then grouped into clusters based on profile similarities. Subsequently, for each cluster, a prediction model based on CNN-LSTM is developed to forecast traffic volume/speed for all detectors within the cluster. Moreover, the extracted profiles can aid in identifying unusual daily traffic patterns for each detector, facilitating investigations into potential equipment issues or unrecorded traffic incidents.

The remaining sections of this paper are organized as follows. Section "Methodology" presents the proposed method and DL model architectures. Section "Results" discusses the experimental results. Finally, Sect. "Conclusion" provides a summary of this paper.

Data

Description of the Data Source

The traffic data used in this study is sourced from data.gov.hk and extracted from fixed-location traffic detectors. The dataset includes timestamps, detector IDs, traffic volume, speed, and occupancy. These detectors are strategically placed on major roads and routes in Hong Kong. A total of 1,912 lane detectors corresponding to 714 sites are surveyed. The locations of these sites are depicted in Fig. 1. By examining the traffic flow-speed graph of the dataset, noteworthy changes in the relationships of most detectors are observed in May 2022, which are attributed to detector recalibration. Consequently, data from May 2022 to November 2022 are selected for training, validation, and testing. Each time slice in this paper represents a 6-minute time interval.

Due to communication interruptions or detector failures, the collected raw traffic flow data inevitably contains missing values. Figure 2 depicts the traffic volume measured by all detectors over six months. The x-axis corresponds to time, while the y-axis represents detector IDs arranged by name. Each data point signifies the traffic volume a detector measures at a specific time interval. The color intensity indicates the magnitude of the traffic volume, with lighter colors representing more vehicles passing by the detector during that time interval. Blank spaces in the figure denote missing data. The overall average failure rate of the dataset is 5.94%. Data imputation is necessary for this dataset to facilitate the training of prediction models.

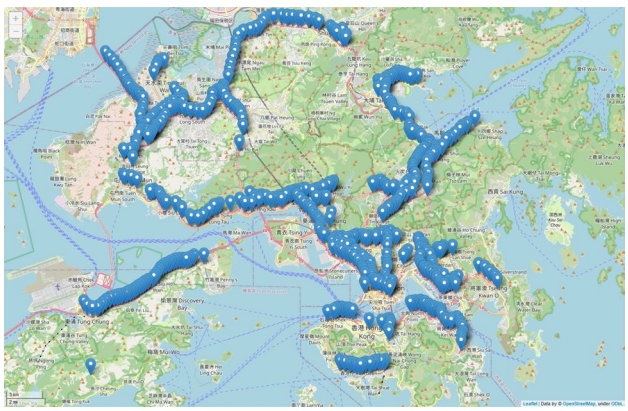

Fig. 1 Locations of investigated traffic detectors in HK

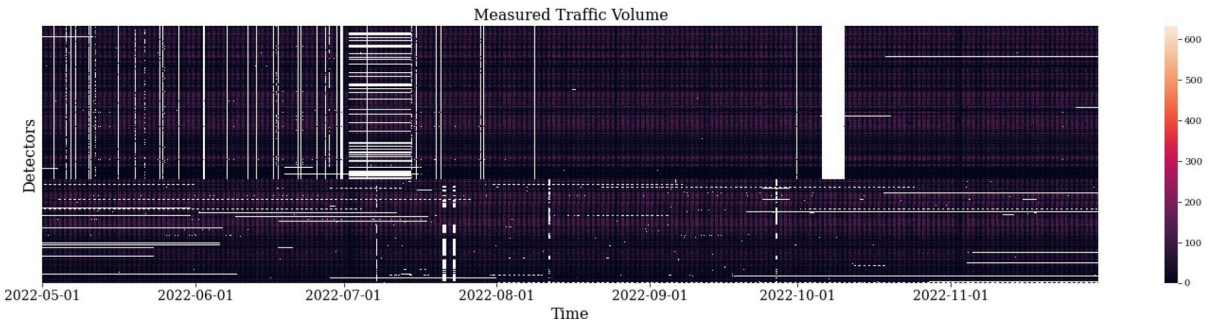

Fig. 2 Measured traffic volume of traffic detectors over six months

Data Imputation

Figure 2 provides insights into the types of missing data, which can be categorized into three groups based on the duration of continuous detector failures. Each category requires a different data imputation method:

(1) For cases where a detector's continuous failure duration is no more than 6 min, the most recent observed traffic value can be used for imputation.
(2) For cases where a detector's continuous failure duration is more than 6 min but less than 6 h, the K-Nearest Neighbors (KNN) method is utilized for data imputation. KNN's similarity-based, nonparametric approach is well-suited for interpolating missing values in complex datasets without prior modeling or training. Within our context, KNN computes the average Euclidean distance between the failed detector's available measurements and those of other detectors at corresponding time intervals to identify detectors sharing similar traffic patterns for that day. Subsequently, a weighted average of these similar detectors' measurements is utilized for data imputation. Smaller average Euclidean distances indicate higher similarity (weight) between detectors.
(3) For cases where a detector's continuous failure duration exceeds 6 h, using KNN to find similar detectors may not be accurate. In such instances, the approach involves identifying the K most similar dates for data imputation. The distance between two dates is determined by the average Euclidean distance of measured values provided by working detectors within their corresponding time slices. If the number of points with values between two dates is less than 60% of the total number of points in a day, the distance between these two dates is set to infinity. Missing values can be filled with the average values from the K nearest dates whose distances are smaller than a threshold value.

Data Normalization

To expedite the search for the optimal solution for the DL models, perform min-max normalization to scale the data to the range of [0, 1]:

$$X^* = \frac{X - x_{min}}{x_{max} - x_{min}}, \tag{1}$$

where X represents the input traffic variable, x_{min} and x_{max} represent the minimum and maximum values of that traffic variable in the training dataset.

For the model's output, perform the following operation to restore it to the original scale:

$$Y' = Y^* \times (x_{max} - x_{min}) + x_{min}, \tag{2}$$

where Y^* is the output of the model, and Y' is the traffic variable prediction after being restored to the original scale.

Methodology

Overview of the Proposed Method

Taking traffic volume prediction as an example, the proposed method consists of the following steps:

(1) Extracting deep features with detector signatures from each daily traffic volume sequence.
(2) Generating representative profiles for each detector based on their daily features and grouping the detectors based on profile similarity.
(3) Developing a traffic volume prediction model for each group.

Following these steps, detectors across the Intelligent Road Network (IRN) are grouped into multiple clusters, each with

a dedicated prediction model that operates independently. This division allows each model to process inputs from specific sub-segments, reducing input complexity and enhancing prediction speed. Furthermore, detectors within the same cluster can provide mutual support. In the event of detector failure, real-time imputation and effective prediction can be achieved by leveraging information from similar detectors within the cluster. This approach also minimizes the number of prediction models for the entire network, reducing development and maintenance costs. Subsequent sections will elaborate on these three pivotal steps.

Daily Feature Extraction Model

The traffic volume measured by each detector shows periodic changes, such as similar peak and off-peak periods on workdays. When making traffic predictions, it is beneficial to consider the temporal similarity for each location and draw insights from locations with similar periodic patterns, especially in cases of detector failure. These similarities may arise from factors like points of interest (POI), geographical location, and functionality. To capture these similarities, we analyze each detector's daily trends in historical traffic volume. DL models possess strong pattern analysis capabilities and are utilized to extract features from the daily traffic volumes of each detector. The periodic variations in traffic volume manifest in the similarity of the extracted daily deep features, aiding in creating a distinct profile for each detector. Notably, the profile extraction process focuses on daily features rather than 6-minute intervals, as our detector clustering strategy centers on similarities in periodic trends rather than transient traffic fluctuations caused by weather conditions or special events.

The proposed feature extraction model, as illustrated in Fig. 3, consists of several CNN layers and an FC layer. It takes a daily traffic volume data of a detector $v_{x_i}^{d_j}$ as input and outputs a distribution $\hat{L}_{x_i}^{d_j}$ representing the ID of the detector corresponding to the input daily data. CNN applies convolutional operations with kernels of a certain size to local regions of the input data, computing the inner product between different data windows and the convolutional kernel to capture local relationships. Hence, CNNs are utilized to extract features from time series and reduce feature dimensionality. Initially, we perform convolution operations on the input using kernels of different sizes while maintaining the output feature map size through padding. In Fig. 3, a, b, and c represent kernel sizes of 3, 5, and 7, respectively. The convolution results are then summed together to achieve multi-scale feature extraction and fusion. Subsequently, several stacked convolutional layers are used, with the sliding stride of the convolutional kernels set to $(a-2)$. By sliding at different positions, the convolutional kernels can traverse the entire input. The stride setting reduces the output size of the convolutional layer by half, decreasing the feature map size in each convolutional layer. After passing through multiple CNN layers and flattening, the features $y_{x_i}^{d_j}$ are obtained, representing the features of detector x_i throughout day d_j. Next, through a fully-connected layer, $y_{x_i}^{d_j}$ is mapped to an output vector $\hat{L}_{x_i}^{d_j}$, with dimensions equal to the number of detectors. Each dimension of $\hat{L}_{x_i}^{d_j}$ represents the relevance to the corresponding detector ID.

$$y_{x_i}^{d_j}, \hat{L}_{x_i}^{d_j} = f_{CLU}(v_{x_i}^{d_j}), \tag{3}$$

where f_{CLU} indicate the proposed daily feature extraction model. $v_{x_i}^{d_j}$ is the traffic volume sequence of detector $x_i$on day d_j, encompassing the entire day.

Profile Obtain

Deep features associated with individual detectors are gathered. However, certain days for a detector may exhibit atypical traffic patterns, potentially leading to inaccuracies in identifying the detector ID using the deep features from these days. To better extract the general characteristics of detectors, only deep features extracted from typical days are used to calculate the profile for each detector, encompassing both weekdays and weekends. The excluded atypical days vary across different detectors. Nevertheless, no atypical days are excluded during the prediction model's learning and testing phases.

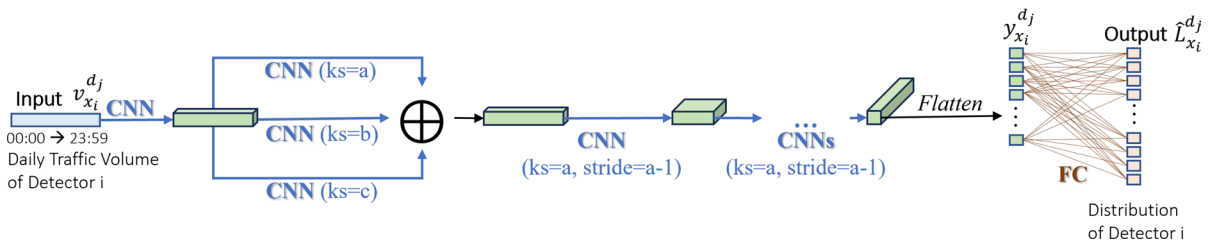

Fig. 3 Proposed daily feature extraction model

For detector x_i on the j^{th} day, the predicted distribution $\hat{L}_{x_i}^{d_j}$ is as follows:

$$\hat{L}_{x_i}^{d_j} = \{p_{x_1}^{d_j}, p_{x_2}^{d_j}, \cdots, p_{x_i}^{d_j}, \cdots, p_{x_N}^{d_j}\} \tag{4}$$

where $p_{x_i}^{d_j}$ represents the model's estimation of the probability that the input daily data belongs to the i^{th} detector, and N is the total number of detector IDs.

Sort $\hat{L}_{x_i}^{d_j}$ in descending order based on probabilities, resulting in:

$$\hat{L'}_{x_i}^{d_j} = \{p_{x_{i_1}}^{d_j}, p_{x_{i_2}}^{d_j}, \cdots, p_{x_{i_N}}^{d_j}\}(p_{x_{i_1}}^{d_j} \geq p_{x_{i_2}}^{d_j} \geq \cdots \geq p_{x_{i_N}}^{d_j}) \tag{5}$$

Denote the top 1% of the sorted probabilities in $\hat{L'}_{x_i}^{d_j}$ as $P'^{d_j}_{x_i} = \{p_{x_{i_1}}^{d_j}, p_{x_{i_2}}^{d_j}, \cdots p_{x_{i_{N/100}}}^{d_j}\}$

$$\bar{y}_{x_i} = \frac{1}{k}\sum_{j=1}^{k} y_{x_i}^{d_j}, if p_{x_i}^{d_j} \in P'^{d_j}_{x_i}, \tag{6}$$

where $\bar{y}_{x_i}$ is the obtained profile of detector x_i.

The Euclidean distance between each pair of detector profiles is computed, and the hierarchical clustering algorithm f_{HA} is used to divide the detectors into multiple clusters $(C_1, C_2, ..., C_M)$:

$$\begin{aligned} &C_1, C_2, ..., C_M = f_{HA}(\bar{y}_{x_1}, \bar{y}_{x_2}, \cdots, \bar{y}_{x_N}) \\ &C_1 \cup C_2 \cup \cdots \cup C_M = \{x_1, x_2, \cdots, x_N\} \end{aligned} \tag{7}$$

Proposed Traffic Volume Prediction Model Architecture

After clustering, each cluster comprises detectors with similar profiles. A separate traffic volume prediction model is constructed for each cluster, as illustrated in Fig. 4. For a given Cluster $C_J (J = 1, 2, .., M)$, the LSTM and CNN are employed to extract spatial-temporal features. The LSTM captures global information for the cluster, while the CNN utilizes the individual detector's time series to obtain each detector's deep features separately. The obtained deep features are then fused to generate traffic volume predictions for the cluster for the next hour. For example, for Cluster C_J composed of n detectors $(x_{J_1}, x_{J_2}, \cdots, x_{J_n})$:

$$C_J = \{x_{J_1}, x_{J_2}, \cdots, x_{J_n}\}, v_{C_J}^{\tau-\Delta} = \{v_{x_{J_1}}^{\tau-\Delta}, v_{x_{J_2}}^{\tau-\Delta}, \cdots v_{x_{J_n}}^{\tau-\Delta}\} \tag{8}$$

$$\begin{aligned} y_{C_J}^{\tau} = &f_{FC} f_{LSTM} f_{FC}(v_{C_J}^{\tau-10\Delta}, v_{C_J}^{\tau-9\Delta}, \cdots, v_{C_J}^{\tau-\Delta}) \\ &\oplus f_{CNNs}(v_{C_J}^{\tau-10\Delta}, v_{C_J}^{\tau-9\Delta}, \cdots, v_{C_J}^{\tau-\Delta}) \end{aligned} \tag{9}$$

$$(\hat{v}_{C_J}^{\tau}, \hat{v}_{C_J}^{\tau+\Delta}, \cdots, \hat{v}_{C_J}^{\tau+9\Delta}) = f_{CNNs}(y_{C_J}^{\tau}) \tag{10}$$

where $y_{C_J}^{\tau}$ represents the concatenated deep features extracted from the cluster using the two neural networks. $\hat{v}_{C_J}^{\tau}$ is an n-dimensional vector representing the predicted traffic volume at time τ for each detector in the cluster.

In Eq. (10), f_{CNNs} utilizes multiple convolutional layers to extract temporal features from the input data. At each layer, neighboring k_t^l features are combined along the temporal dimension. The convolutional kernels at each layer have a size of $(k_t^l, 1)$, where 1 indicates that no feature fusion is performed in the spatial (detector) dimension. By passing through multiple convolutional layers, the temporal dimension is compressed from the input T_h to 1. Since the desired output is the prediction for the next T_p time steps, the number of convolutional kernels in the final layer is set to T_p. The resulting output from the right-hand side of the equation is a tensor of size $(T_p, 1, N)$.

On the left-hand side of Eq. (10), cluster information is extracted. Firstly, for each time step in the data, spatial features are extracted using a fully connected layer (f_{FC}). The input information from N detectors at each time step is connected to a new set of N units. The information from different time steps is independent of each other. Subsequently, the new N-dimensional features from each time step are fed into an LSTM (f_{LSTM}) to extract temporal features and obtain the feature prediction for the next time step, resulting in a

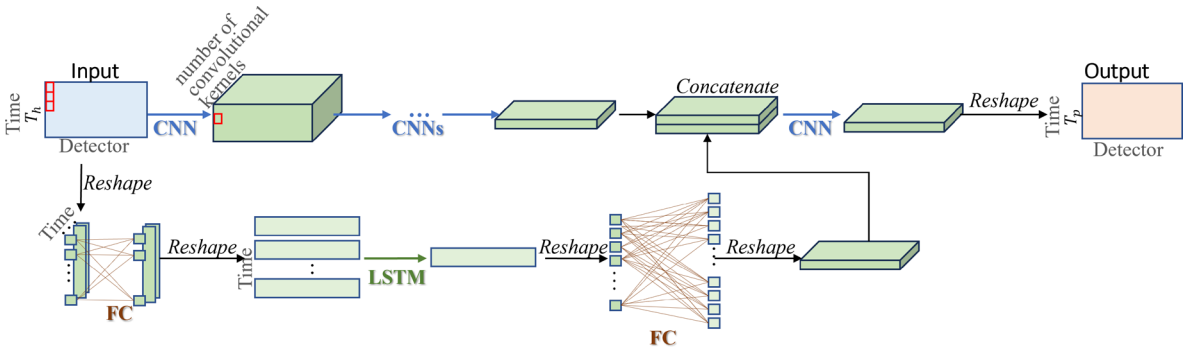

Fig. 4 Proposed DL traffic volume prediction model

tensor with the size of (1, N). Afterward, a fully connected layer is used to obtain a ($T_P \times N$) -dimensional vector, which is then transformed into a tensor of size (T_P, 1, N).

Finally, feature concatenation is performed. Each detector at each time step has a 2-dimensional feature. The final prediction is obtained using a convolutional layer with a kernel size of (2, 1). The proposed prediction model not only considers the temporal features of each detector but also incorporates cluster information to assist in the prediction process.

Training of DL Models

The daily feature extraction model, which functions as a multi-classification neural network, utilizes the cross-entropy loss function to minimize the disparity between the predicted probability distribution and the one-hot encoded representation of the actual detector ID. The input daily traffic volume is standardized to ensure uniformity. Model parameters are updated through backpropagation and gradient descent optimization. The maximum number of training iterations for the model is capped at 300.

Regarding the Traffic Volume Prediction Model, the mean squared error (MSE) loss function measures the discrepancy between the predicted traffic volume and the actual values. During training, the standardized historical 1-hour traffic volume data for each detector within a cluster is used as input, while the corresponding standardized ground truth values (i.e., the future 1-hour traffic volume) serve as the target for prediction.

All DL models are trained iteratively on a training dataset, and the performance is evaluated on a validation dataset. Hyperparameters, including learning rate, batch size, and network architecture, are fine-tuned for optimal performance. Adam optimization is employed as the optimizer for all deep models. The learning rate, a critical parameter affecting convergence, is selected through multiple experiments within the range of $[10^{-3}, 10^{-4}, 10^{-5}]$. A dynamic learning rate scheme is implemented, reducing the learning rate by a factor of 0.1 if the model's evaluation metric fails to improve after 30 training iterations until it reaches the minimum value of 10^{-6}. Additionally, an early stopping strategy is utilized to prevent overfitting.

Results

The proposed method involved training and validating traffic volume and speed prediction models using datasets collected from major roads in Hong Kong between May 2022 and October 2022. Subsequently, we evaluate the proposed method using the corresponding traffic volume and speed datasets from November 2022.

The testing included the following aspects:

(1) Evaluating the predictive performance of the prediction models when all detectors are working.
(2) Simulating 1%-20% detector failures to assess the resilience of the prediction models to detector failures.
(3) Assessment of the proposed method's performance in offline data imputation for addressing missing data in the dataset.
(4) Additionally, we evaluated the accuracy of the proposed daily feature extraction model in identifying detector IDs.

These tests enabled a comprehensive understanding of the proposed method's performance in various scenarios, including its predictive capabilities under both normal and detector failure conditions, its performance in data imputation, and the accuracy of the daily feature extraction model in identifying detector IDs.

Evaluation Metrics Used for Testing

The predictive model is evaluated using the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE).

$$\mathrm{MAE}(Y, \hat{Y}) = \frac{1}{N}\sum_{i=1}^{N} |Y_i - \hat{Y}_i|, \tag{11}$$

$$\mathrm{RMSE}(Y, \hat{Y}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2}, \tag{12}$$

$$\mathrm{MAPE}(Y, \hat{Y}) = \frac{1}{N}\sum_{i=1}^{N} \frac{\left|\hat{Y}_i - Y_i\right|}{Y_i}, \tag{13}$$

where $\hat{Y}$ and Y represent the prediction and the observation, respectively. The predictive performance improves as MAE, RMSE, and MAPE decrease. All MAPE calculations for traffic volume/speed prediction only consider observations with a traffic volume greater than 10 vehicles per 6 min or 1 km per hour to avoid division by zero and biases in error for low observations. For example, if the observed traffic volume is four vehicles and the predicted traffic volume is two, the MAPE is 50%, although the error is only two vehicles.

Predictive Performance of the Prediction Model without Failed Detectors

This paper introduces a method for forecasting traffic volume and speed for 6 min to 1 h. These forecasts can aid road traffic management departments in promptly identifying congested areas. Additionally, travel time can be derived from traffic volume and speed predictions, allowing individual travelers to plan their journeys effectively.

Traffic Volume Prediction Without Failures

The proposed method enables a one-step traffic volume prediction from 6 min to 1 h. To evaluate the overall performance, we compared the proposed method with simple baseline and advanced models, including: (1) Naive: The simple method for prediction using the most recent observed traffic volume. (2) STGCN (Spatio-Temporal Graph Convolutional Network) (Yu et al. 2018): The DL method utilizes Graph Convolutional Networks and one-dimensional temporal convolutions to capture spatial and temporal correlations. Following the STGCN paper, we constructed edges based on the geographical distances between stations. Additionally, we calculated the Pearson correlation coefficients between the historical time series of detectors. Detectors with high coefficients were also connected by edges, even if their geographical locations were potentially distant. The model's construction is based on publicly available code at https://github.com/FelixOpolka/STGCN-PyTorch/blob/master/stgcn.py. (3) Transformer (Vaswani et al. 2017): The DL method uses attention mechanisms to capture spatio-temporal relationships. The Transformer is employed to learn temporal relationships, and our position encoding for the input historical time series follows the absolute positional encoding based on sine and cosine functions from the paper. Fully connected layers are utilized to learn the global spatial relationships from historical data of all detectors to make predictions. (4) MC_Transformer: Building upon the Transformer's promising performance, we further tested using our clustering results with a separate Transformer for each cluster. Post-clustering, each detector extracts global spatial relationships from its highly correlated subset for prediction purposes.

For the DL models (STGCN, Transformer) used for comparison, we utilized trained models to conduct multi-step forecasting, evaluating performance across prediction horizons from 6 min to 1 h. These results were then quantitatively compared with those generated by our proposed method. Besides the single-step Transformer generating predictions for one time interval at a time, we also attempted training the Transformer to provide one-step predictions for the next 6 min to 1 h. However, we found that although its performance for 1-hour predictions was comparable to that of the single-step Transformer, it exhibited larger errors in predicting the subsequent 6 min. In practical applications, the emphasis on forecasting the next 6 or 12 min outweighs that of the following hour. Hence, we used the single-step Transformer for subsequent comparisons. All DL models were implemented on the same platform with the same resources as our method.

Figure 5 illustrates the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) for each method at each time interval. For predictions of the next 6 min, the proposed method achieved MAE, RMSE, and MAPE values of 5.17 vehicles/6 min, 7.64 vehicles/6 min, and 14.07%, respectively. For forecasting traffic volume for the next 1 h, the obtained MAE, RMSE, and MAPE were 5.69 vehicles/6 min, 8.96 vehicles/6 min, and 15.26%, respectively. However, for the 6-minute predictions, the MC_Transformer shows a very slight advantage over the proposed method, with corresponding MAE, RMSE, and MAPE values of 5.16 vehicles/6 min, 7.60 vehicles/6 min, and 14.06%, respectively. Nevertheless, the proposed method outperforms other approaches. Over time, though, the proposed method exhibits superior forecasting compared to the MC_Transformer. We attribute this to the exceptional performance of the attention mechanism in capturing short-term relationships. However, LSTMs are more suitable for capturing longer temporal variations in this task.

Furthermore, as depicted in Fig. 5, the superior performance of the MC_Transformer over a standalone

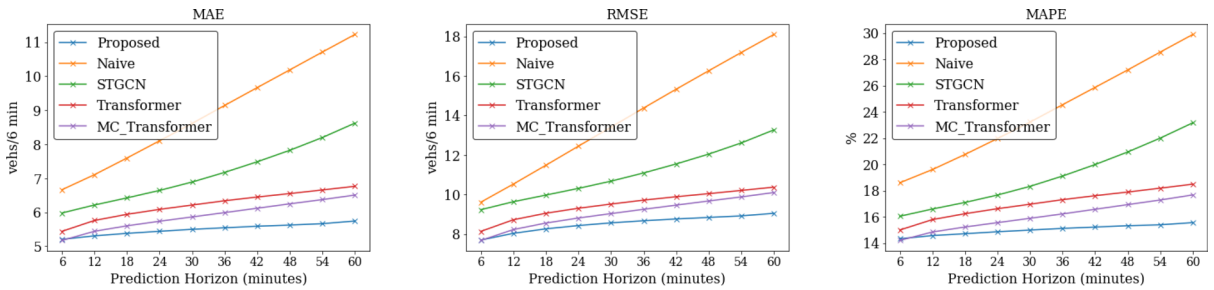

Fig. 5 Traffic volume predictive performance without detector failures

Transformer indicates that our profiling cluster enhances the coherence within each cluster, aiding in learning more accurate traffic patterns for prediction. While STGCN exhibits comparatively lower performance than other DL methods in this context, it is crucial to note that the graphs utilized are based on the geographical distance between stations and the historical data correlations of lane detectors. If alternative graphs, such as those based on functional similarities, were incorporated, we believe STGCN could demonstrate improved efficacy. Additionally, learning graphs is more challenging and demands a substantial volume of data, which may be another reason STGCN is less effective than the Transformer and the proposed method in this task.

As the forecasting time horizon increases, accuracy decreases for longer-term (1 h) predictions compared to short-term ones. Still, it provides relatively accurate forecasts, as shown in Fig. 6. However, Fig. 6 shows that the proposed method can capture traffic volume trends but may underestimate traffic volume under rarely seen congested conditions.

Traffic Speed Prediction Without Failures

Accurate traffic speed prediction models were developed by adapting the prediction model from traffic volume data to traffic speed data and retraining the models. The evaluation metrics, including MAE, RMSE, and MAPE, are illustrated in Fig. 7, showcasing the speed predictions generated by the proposed approach for the subsequent 6 min to 1 h. The 6-minute predictions yielded MAE/RMSE/MAPE values of 3.70 km/h, 6.32 km/h, and 6.33%, respectively. For the

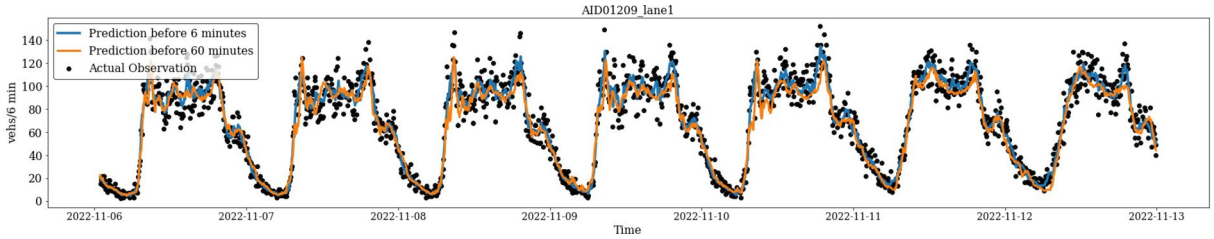

Fig. 6 Visualization of predicted traffic volume and actual observation

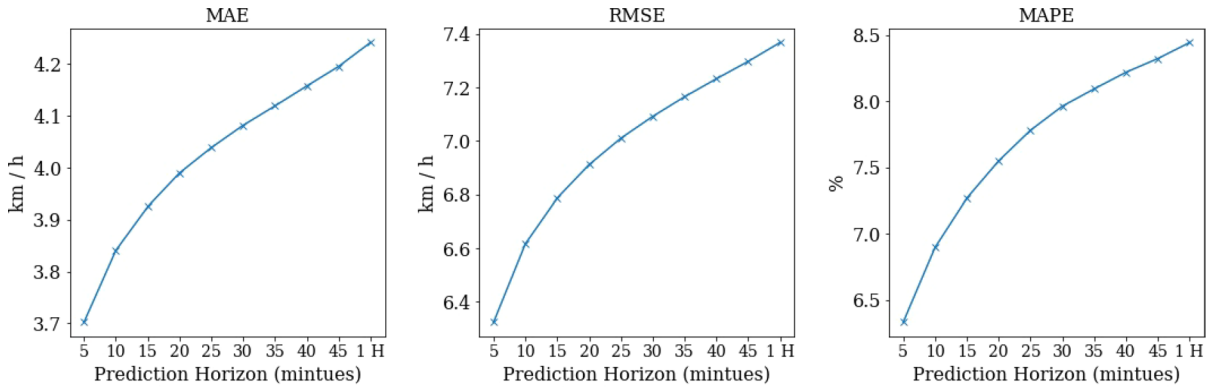

Fig. 7 Traffic speed predictive performance without detector failures

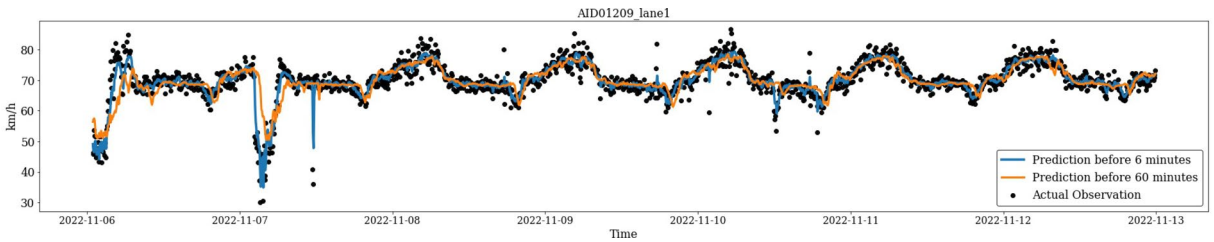

Fig. 8 Visualization of predicted traffic speed and actual observation

Table 1 Traffic Speed Prediction performance of different speed ranges

Speed range	Proportion %	MAE	RMSE	MAPE%
$0 \leq x < 15$ km/h	0.58	9.9 km/h	15.6 km/h	149
$15 \leq x < 30$ km/h	1.42	7.3 km/h	11.8 km/h	33
$x \geq 30$ km/h	98	3.7 km/h	6.1 km/h	5

following 1 h, the corresponding values were 4.24 km/h, 7.37 km/h, and 8.44% for MAE/RMSE/MAPE.

Figure 8 visually compares the predicted speeds obtained 6 min and 1 h ahead using the proposed method and the actual measurements. It demonstrates the method's proficiency in predicting traffic speeds under normal conditions and its ability to capture trends of sudden speed drops.

The proposed method faces challenges in extreme or rare low-speed (0–15 km/h) traffic scenarios (< 1% of the data, see Table 1). Figure 9 illustrates the counts of detectors measuring low speeds at different time intervals, shows that low-speed events are not confined to specific days, with prevalence during peak hours and rarity in the early morning. Additionally, Fig. 10 visualizes the spatial distribution of low-speed data within the dataset timeframe, showing variability; regions with darker colors indicate higher frequencies of low-speed events. For improved clarity, Figs. 9 and 10 display data within the dataset for one month, with similar patterns for other months. Note that Figures 5–8 present prediction errors across all speed ranges, including low speeds. Table 1 details the sample proportions within different speed ranges and their respective MAE, RMSE, and MAPE values. Due to limited available samples for learning, the proposed model might not be suitable for non-recurrent congestion scenarios with low traffic speeds. Innovative methods are required to predict non-recurrent congestion speeds effectively.

Table 1 shows that the proposed method yields relatively large values for all the metrics when the speed is less than 15 km/h, indicating a significant distance between the predictions and the actual observed values. Hence, we visualized such data, as shown in Fig. 11. The 'Median' in the figure represents the median values of the training dataset

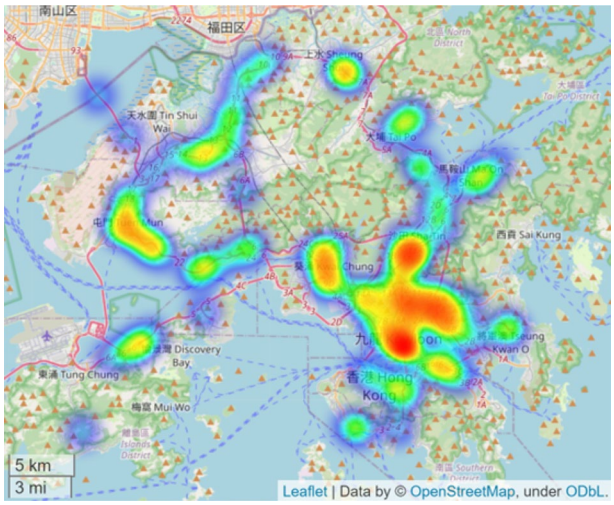

Fig. 10 Heatmap displaying the distribution of detectors measuring very low speeds (0–15 km/h). The increasing intensity of red signifies a higher frequency of low-speed occurrences

for each corresponding time slot from Monday to Sunday. 'Mean' indicates the mean values. The blue line represents the visualization of the actual observed values for one week from the test set of this detector, while the orange line represents the predictions obtained from the proposed method. It can be observed that the actual observed values exhibit significant fluctuations, while the predictions are only able to fit a general trend. However, such substantial fluctuations in speed observations throughout the week are highly abnormal and likely caused by erroneous reports due to detector malfunctions. Hence, we removed these exceptional values where the difference between the current speed and the speed measured in the previous time slot exceeds 30 km/h, with a frequency greater than 50% within a day.

The predictive performance after removing these exceptional values is shown in Table 2. It can be observed that the predictive performance for low-speed traffic scenarios improves after the removal. However, it is still not particularly satisfactory. We conducted visualizations to investigate possible reasons, as shown in Fig. 12. It can be observed that the variation trend of traffic speed for this

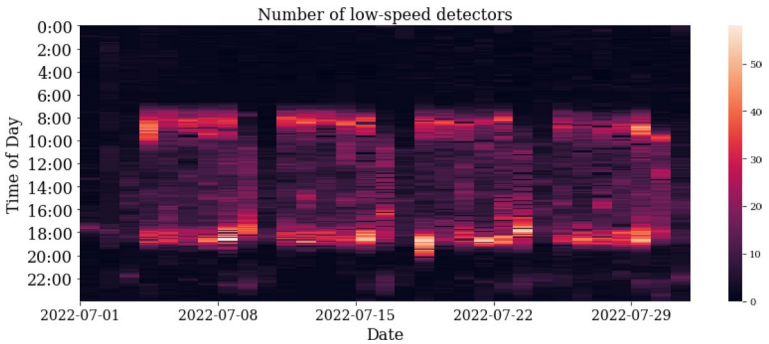

Fig. 9 The count of detectors detecting very low speeds (0–15 km/h) at different times of the day. Brighter colors indicate more detectors capturing low speeds

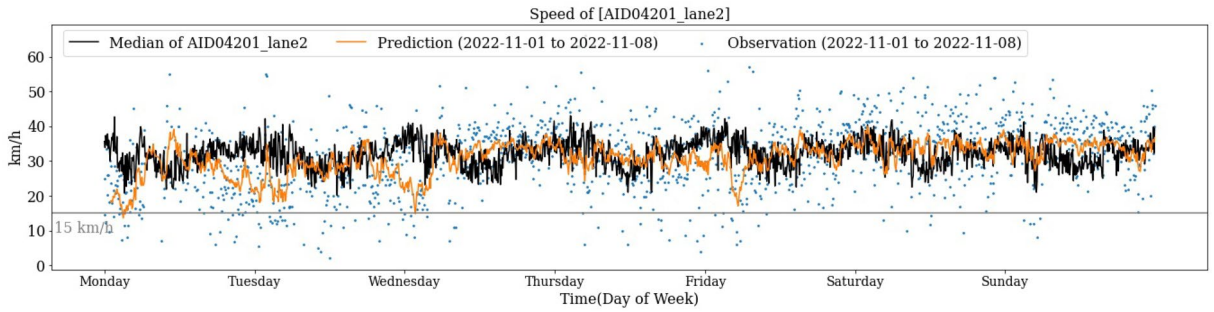

Fig. 11 Visualization of predicted traffic speed and actual observation (speed < 15 km/h) (Abnormal Speed)

detector has changed, differing from the training data used to train the prediction model. The results also highlight that DL models rely on the data used for learning. If traffic trends change, it is necessary to train the prediction model using new data to adapt to the new circumstances.

Predictive Performance of the Prediction Model with Detector Failures

In the real world, the failures in traffic detectors are random in terms of location and timing. Nevertheless, the proposed method demonstrates minimal performance degradation even under such scenarios of random failures.

Traffic volume prediction with failures

The proposed method can generate accurate traffic volume predictions even with detector failures. Failures were simulated by randomly designating some detectors in each cluster as 'failed'. When a detector is marked as failed, all its traffic data during testing is set to missing. Failure rates range from 1% to 20%, meaning that 1% to 20% of detectors in each cluster are randomly selected to fail in each simulation. Detector malfunctions usually occur randomly, although it is possible for a group of detectors in a specific area to malfunction due to network connectivity issues. However, our clustering is based on profile similarity rather than geographical proximity, leading to diverse detector locations within a cluster. In this paper, we only consider random failures.

When a detector fails, the model's input contains missing values corresponding to the failed detectors. We filled in the missing inputs to make predictions based on the current available observations of other working detectors within the same cluster as the failed detector and their historical average weekly traffic volume (AWT) derived from the training dataset.

For a detector x_j in cluster C_J, if it fails at time τ (corresponding to day of the week D_a and time of the day T_b), the imputed value $\tilde{v}^{\tau}_{x_j}$ is obtained as follows:

$$\tilde{v}^{\tau}_{x_j} = AWT^{D_a,T_b}_{x_j} \times Avg_{x_l \in C_J} \frac{v^{\tau}_{x_l}}{AWT^{D_a,T_b}_{x_l}}, \tag{14}$$

Table 2 Traffic Speed Prediction performance of different speed ranges after removing abnormal speed

Speed range	MAE	RMSE	MAPE%
$0 \leq x < 15$ km/h	6.8 km/h	10.5 km/h	86
$15 \leq x < 30$ km/h	6.4 km/h	9.7 km/h	29
$x \geq 30$ km/h	3.3 km/h	5.0 km/h	5

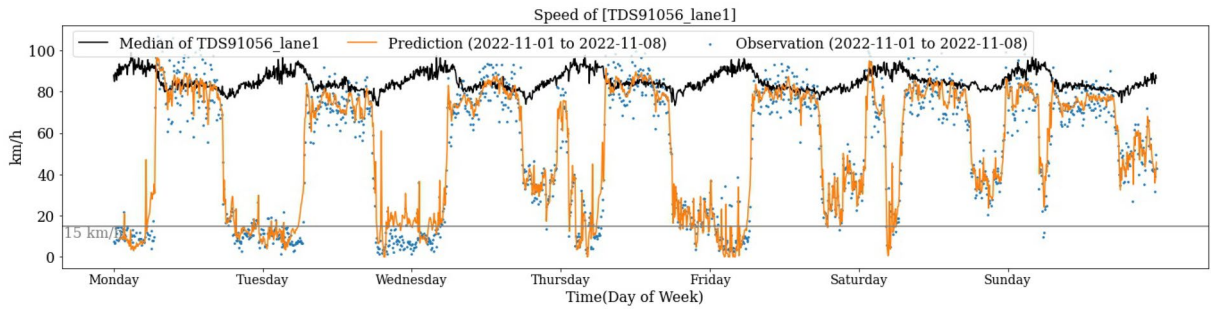

Fig. 12 Visualization of predicted traffic speed and actual observation (speed < 15 km/h) (Pattern Changed)

where $\tilde{v}^{\tau}_{x_j}$ represents the real-time measured traffic volume of the working detector x_l in the cluster. *Avg* refers to the average function.

The average values of MAE/RMSE/MAPE obtained are visualized in Fig. 13. Each data point in the figure represents the average of 30 experiments, where different detectors are randomly designated as failures. The results in Fig. 13 reveal a gradual decline in the model's performance with an increasing number of failed detectors, aligning with expectations. Notably, the proposed method demonstrates minimal degradation in performance in the presence of failures. For the next 6-minute prediction, when the failure rate of detectors reaches 20%, the average performance degradation results in an increase of 0.29 vehs/6min in MAE, 0.72 vehs/6min in RMSE, and 0.56% in MAPE (as indicated in the bottom line of Fig. 13). For the prediction of the next 1 h (as displayed in the top line of Fig. 13), the degradation leads to an average increase of 0.11 vehs/6min in MAE, 0.27 vehs/6min in RMSE, and 0.21% in MAPE.

As a reference, the average failure rate in the dataset is 5.94%. With 6% failed detectors, the average MAE, RMSE, and MAPE obtained for the 6-minute prediction are 5.24 vehs/6min, 7.81 vehs/6min, and 14.21%, respectively. Compared to the scenario without detector failures, this represents increases of 0.07 vehs/6min, 0.17 vehs/6min, and 0.14%, respectively. For the 1-hour prediction with 6% failed detectors, the average MAE, RMSE, and MAPE are 5.70 vehs/6min, 8.99 vehs/6min, and 15.29%, respectively. Compared to the scenario without detector failures, this represents increases of 0.02 vehs/6min, 0.03 vehs/6min, and 0.03%, respectively. Moreover, Fig. 14 presents a box plot illustrating the results for all test cases, offering a comprehensive visualization of the proposed method's predictive performance with different failed detectors. Despite fluctuations in predictive performance with different failed detectors, the proposed method consistently exhibits resilience to detector failures, maintaining satisfactory prediction accuracy.

These findings underscore the model's resilience, highlighting its ability to maintain relatively stable performance despite failures. This resilience is valuable for

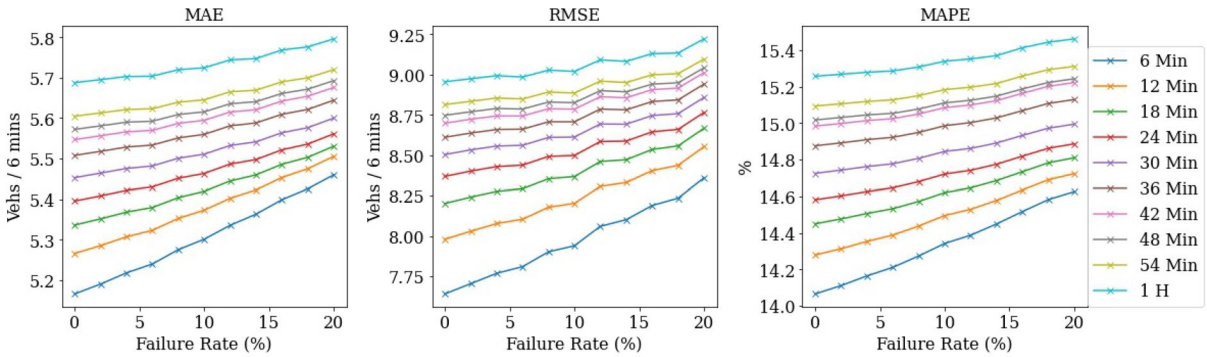

Fig. 13 Traffic volume predictive performance with detector failures

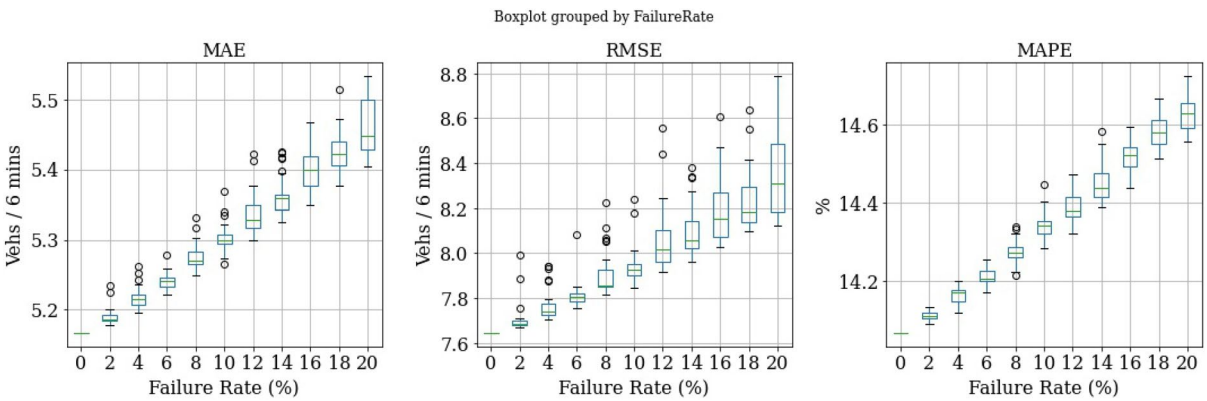

Fig. 14 Traffic volume predictive performance of each case for next6 min with detector failures

real-world applications where sensor malfunctions or failures are common, emphasizing the practical utility of the proposed method.

Traffic Speed Prediction with Failures

The performance of traffic speed prediction with failed detectors is shown in Fig. 15. For the 6-minute predictions, when there is a 6% failure rate in the detectors, the MAE, RMSE, and MAPE values are 3.87 km/h, 6.55 km/h, and 6.68%, respectively. Compared to the predictions without failures, the average performance decreases by 0.17 km/h, 0.23 km/h, and 0.34% in MAE, RMSE, and MAPE, respectively. Figure 16 shows box plots illustrating the MAE, RMSE, and MAPE values for each failure case in the next 6 min. For the predictions of the next 1 h, with a 6% failure rate in the detectors, the MAE, RMSE, and MAPE values are 4.30 km/h, 7.43 km/h, and 8.55%, respectively. On average, there is a decrease of 0.06 km/h, 0.06 km/h, and 0.10% in MAE, RMSE, and MAPE, respectively, compared to the predictions without failures.

Similar to traffic volume prediction, these results offer valuable insights into the model's performance under diverse failure scenarios concerning traffic speed, showcasing its robustness and adaptability for traffic speed prediction in failure scenarios.

Imputation Performance of the Prediction Model for Offline Dataset

Traffic Volume Imputation

Furthermore, the proposed prediction model can be utilized for offline missing data imputation. The results of utilizing the proposed method for data imputation are presented in Fig. 17. Each data point in the figure represents the average of 30 experiments, with different detectors randomly selected as targets for imputation in each experiment. The results demonstrate that the proposed method's imputation yields MAE, RMSE, and MAPE values ranging from approximately 5.1 to 5.3 vehs/6min, 7.6 to 7.8 vehs/6min, and 14% to 14.2%, respectively. These findings underscore the effectiveness of the proposed method in accurately filling in missing data.

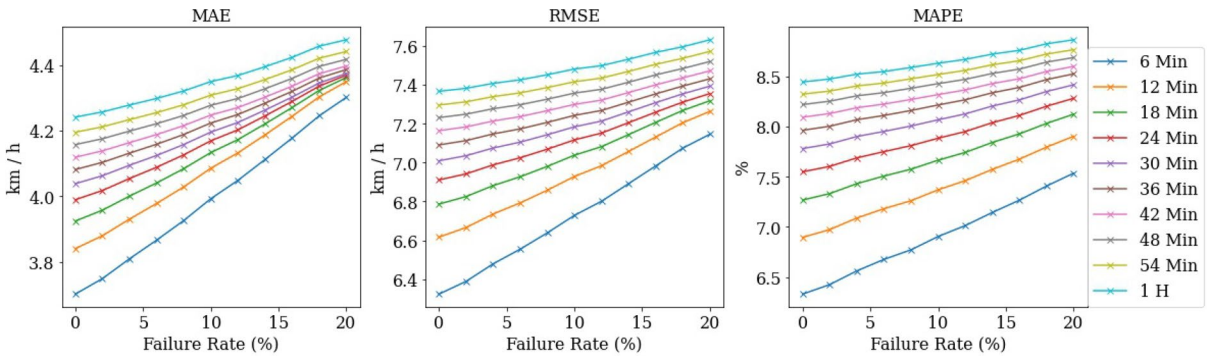

Fig. 15 Traffic speed predictive performance with detector failures

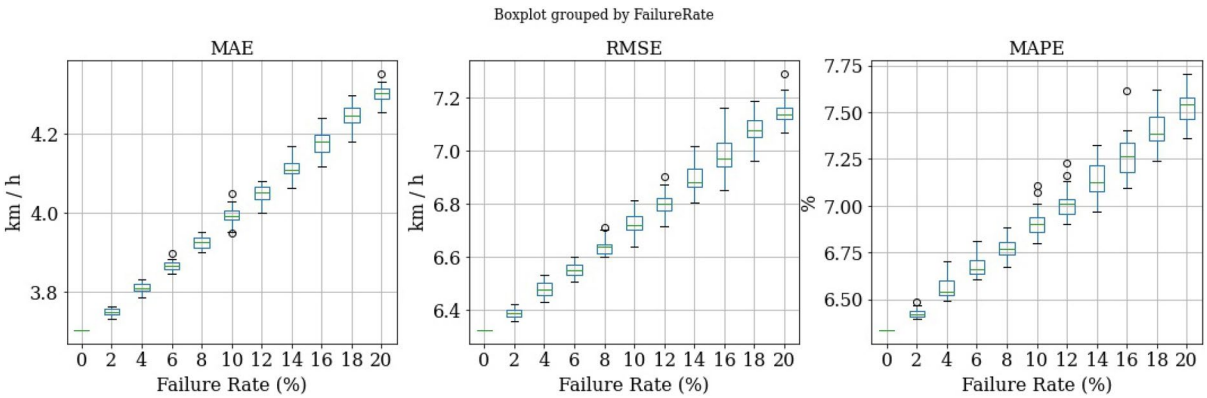

Fig. 16 Traffic speed predictive performance of each case for next 6 min with detector failures

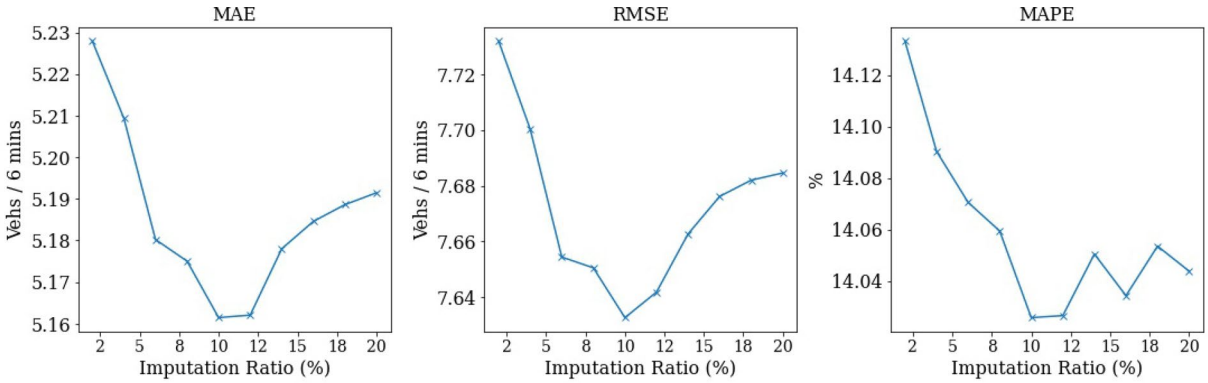

Fig. 17 Performance of offline traffic volume data imputation

Traffic Speed Imputation

The offline imputation performance of traffic speed data is depicted in Fig. 18. The results demonstrate that the imputation method yields a range of MAE, RMSE, and MAPE values ranging from 3.7 to 3.8 km/h, 6.2 to 6.4 km/h, and 6.5% to 6.9%, respectively. These findings emphasize the capability of the proposed method in effectively filling missing traffic speed data.

Recognition Performance of the Daily Feature Extraction Model for Detector IDs

The proposed daily feature extraction model can identify the detector ID corresponding to each daily traffic volume in the dataset. The proposed daily feature extraction model produces a probability distribution, with each element indicating the likelihood of the input daily volume originating from a specific detector. As a result, it enables the calculation of the probability of each detector being correctly identified daily, as well as the percentage of all detectors that can be accurately recognized each day.

Detectors can be ranked based on the probabilities provided by the model's output. In Figs. 19 and 20, the label 'Top 1' indicates that the detector ID with the highest probability predicted by the model is the correct ID. Similarly, 'Top 1%' signifies that the correct ID corresponding to the input's daily volume is among the top 1% of the ranked detectors. It should be noted that different detectors may exhibit similar profiles, thereby providing a basis for clustering them based on their similarities. Our focus lies on the top 1% ranking results.

Figure 19 illustrates that the daily feature extraction model achieves a correct identification rate of exceeding 95% on most days. Moreover, Fig. 20 reveals that only a minimal 1.4% of detectors exhibit identification accuracy below 90% but still above 70%.

The daily feature extraction model can also identify abnormal situations in the dataset where the traffic pattern deviates from the normal. For instance, on November 2, 2022, as depicted in Fig. 19, the identification accuracy (Top 1%) drops significantly to only 32%, while it surpasses 90% for other days. Figure 21 visualizes the average traffic volume of all detectors over time for that day and the previous and following weeks. It is evident that the traffic

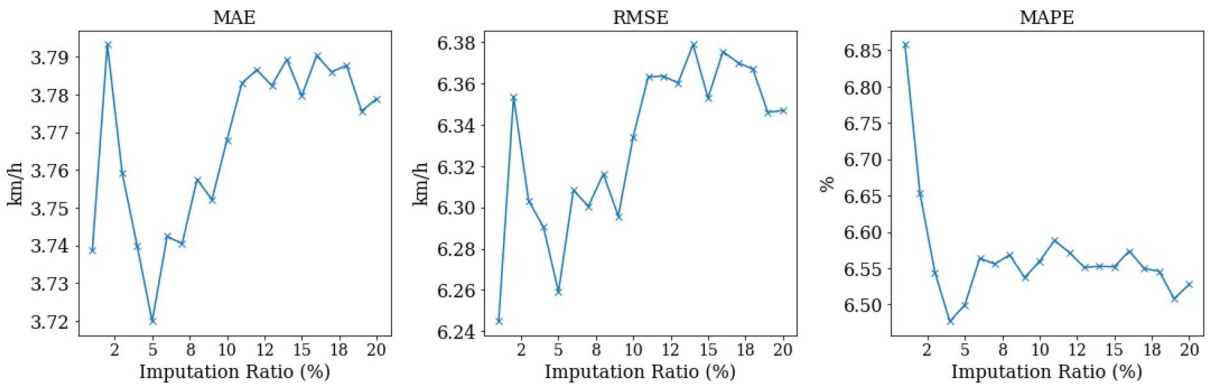

Fig. 18 Performance of offline traffic speed data imputation

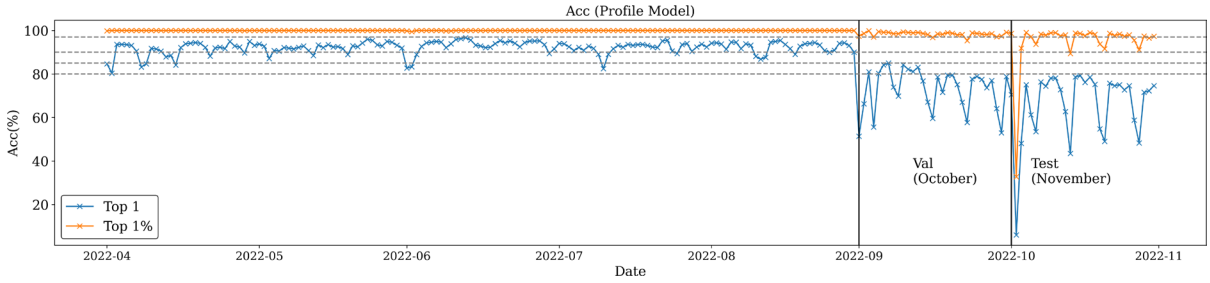

Fig. 19 Identification accuracy of each day of proposed DL daily feature extraction model

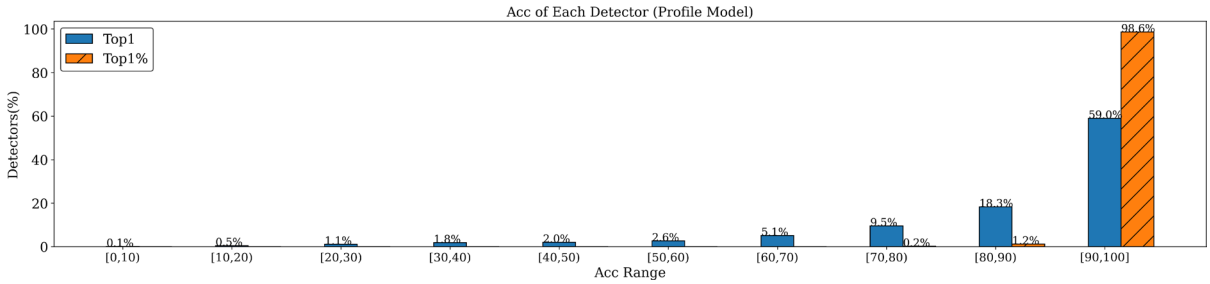

Fig. 20 Identification accuracy of each detector of proposed DL daily feature extraction model

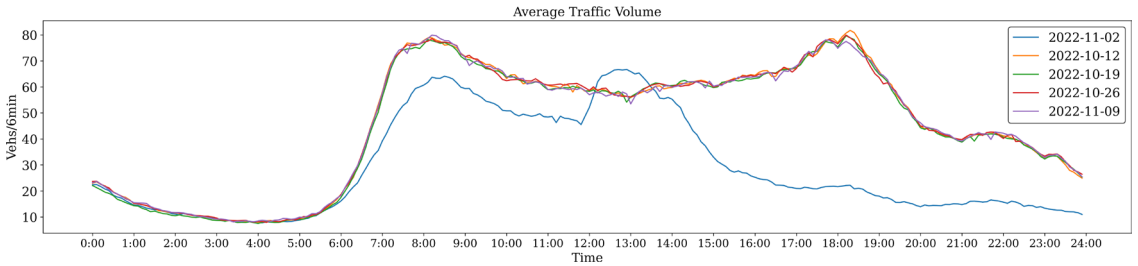

Fig. 21 Visualization of average traffic volumes of all detectors for few days

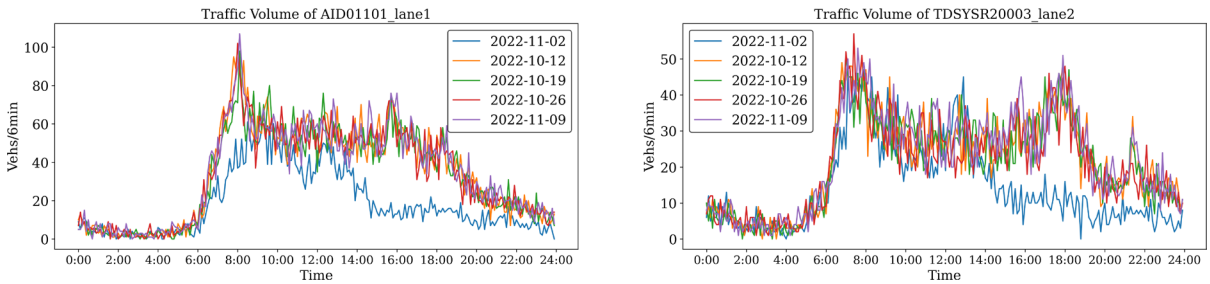

Fig. 22 Visualization of traffic volumes of two detectors of few days

volume changes follow a periodic pattern, but on November 2, 2022, a different trend emerged. Figure 22 shows the traffic volume trends for two randomly selected detectors on these days. Research on relevant news revealed that the Hong Kong Observatory issued a Typhoon Signal No. 8 Northwest Gale or Storm warning at 1:40 PM that day. This explains why the traffic volume deviated from the regular pattern, with a decrease in the number of vehicles on the road. It can be seen that the daily feature extraction model can identify special days exhibiting divergent traffic patterns

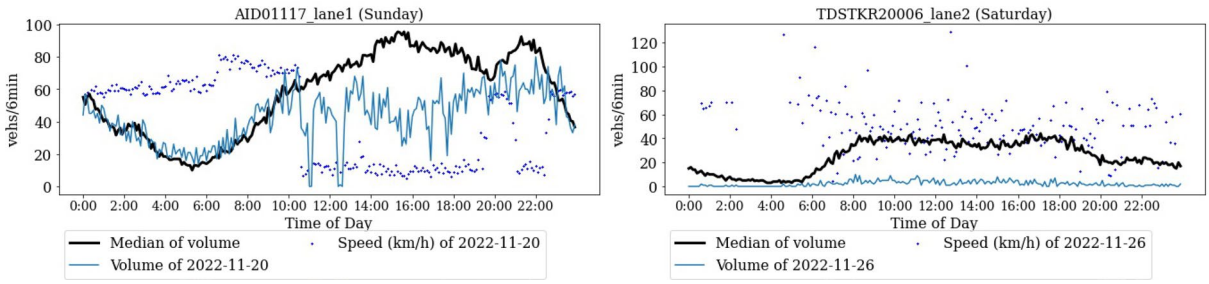

Fig. 23 Days diverging from typical traffic volume

compared to regular days. In the future, integrating the proposed daily feature extraction model with reports on different events would enhance the analysis of traffic changes in various scenarios.

Furthermore, the proposed profile model can serve as one of the metrics to identify drifting traffic volumes for specific detectors. If a daily volume is inputted into the profile model and the correct detector ID does not rank within the top 1% of the output probability ranking, this daily volume likely deviates from its regular traffic pattern. For example, for detector 'AID01117 lane1', there are two days in the test set where it cannot be correctly identified. As mentioned earlier, one of those days is November 2nd, and the other is November 20th, 2022, as shown in Fig. 23. The abnormal days the profile model identifies differ from the usual traffic volume. Similarly, another detector identified different traffic patterns than usual on November 26th, 2022. Hence, the profile can be utilized as one of the indicators for identifying anomalies in future work. The deviation of traffic patterns from the norm can be attributed to various factors, such as events and detector malfunctions. For instance, on the left of Fig. 23, besides abnormal traffic volume, there is an extended period of abnormally low speed, which could result from detector malfunction. By examining the subsequent day, it was observed that the traffic speed in that lane returned to the normal range, suggesting that an event likely caused the anomaly on that particular day. In future work, the proposed profiles, along with traffic incidents, weather data, and other datasets, can be combined to provide a better analysis of the causes of traffic anomalies.

Conclusion

This paper proposes an approach for predicting traffic volume and speed on major strategic routes in Hong Kong. The proposed approach considers the reliability of detectors, allowing for accurate predictions with and without failed detectors.

The approach incorporates two deep model structures: the daily feature extraction model and the traffic volume/speed prediction model, leveraging a hierarchical clustering method to enhance performance. The daily feature extraction model extracts deep features from daily traffic volume to establish detector profiles, facilitating the formation of clusters based on similar profiles for tailored prediction models.

The proposed method achieves promising results under ideal conditions with all detectors operational. For traffic volume prediction over the next 6 min, the MAE, RMSE, and MAPE are 5.17 vehicles/6 min, 7.64 vehicles/6 min, and 14.07%, respectively. Similarly, for traffic speed prediction, the values are 3.70 km/h, 6.32 km/h, and 6.33%.

Moreover, the approach demonstrates resilience to detector failures, as evidenced by performance metrics with a 6% detector failure rate (referring to a dataset loss rate of 5.9%). With 6% failures, the MAE, RMSE, and MAPE for traffic volume prediction are 5.24 vehicles/6 min, 7.81 vehicles/6 min, and 14.21%, respectively, while for traffic speed prediction, the values are 3.87 km/h, 6.55 km/h, and 6.68%. Furthermore, the proposed daily feature extraction model is instrumental in clustering detectors and can be a valuable metric for detecting abnormalities in traffic data.

However, it is essential to note that the proposed method may not perform optimally in predicting traffic volume and speed in rare or unseen situations, such as non-recurrent congestion. To enhance the predictive capabilities of our method across diverse scenarios, future research will focus on integrating multiple data sources, including traffic events, weather conditions, holiday schedules, and other relevant factors. By incorporating these additional dimensions, we aim to refine our approach further and improve its performance in handling varied and complex traffic conditions.

Acknowledgements The work described in this paper was supported by grants from the Smart Traffic Fund (Project No. PSRI/21/2111/PR).

Funding Open access funding provided by The Hong Kong Polytechnic University.

Data Availability The dataset can be retrieved from data.gov.hk.

Declarations

Conflict of Interest The authors declare that there is no conflict of interest.

References

Abduljabbar R, Dia H (2021) Short-term traffic forecasting: an LSTM network for spatial-temporal speed prediction. Future Transp 1(1):21–37

Abduljabbar RL, Dia H (2021) Tsai PW (2021) Unidirectional and bidirectional LSTM models for short-term traffic prediction. J Adv Transp 1:5589075

Cao M, Li VO, Chan VW (2020) A CNN-LSTM model for traffic speed prediction. In: 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), IEEE, pp 1–5

Chen Y, Chen XM (2022) A novel reinforced dynamic graph convolutional network model with data imputation for network-wide traffic flow prediction. Transp Res Part C: Emerg Technol 143:103820

Diao Z, Wang X, Zhang D, et al (2019) Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp 890–897

Fang S, Zhang Q, Meng G, et al (2019) GSTNet: Global spatial-temporal network for traffic flow prediction. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, pp 2286–2293

Han L, Zheng K, Zhao L et al (2019) Short-term traffic prediction based on DeepCluster in large-scale road networks. IEEE Trans Veh Technol 68(12):12301–12313

Jiang R, Yin D, Wang Z, et al (2021) DL-Traff: Survey and benchmark of deep learning models for urban traffic prediction. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management. Association for Computing Machinery, New York, NY, USA, CIKM '21, pp 4515–4525

Ke R, Li W, Cui Z et al (2020) Two-stream multi-channel convolutional neural network for multi-lane traffic speed prediction considering traffic volume impact. Transp Res Rec 2674(4):459–470

Khan A, Fouda MM, Do DT et al (2023) Short-term traffic prediction using deep learning long short-term memory: taxonomy, applications, challenges, and future trends. IEEE Access 11:94371–91

Kong F, Li J, Jiang B et al (2019) Big data-driven machine learning-enabled traffic flow prediction. Trans Emerg Telecommun Technol 30(9):e3482

Lee K, Eo M, Jung E et al (2021) Short-term traffic prediction with deep neural networks: A survey. IEEE Access 9:54739–54756

Liu Q, Wang B, Zhu Y (2018) Short-term traffic speed forecasting based on attention convolutional neural network for arterials. Comput-Aided Civ Infrastruct Eng 33(11):999–1016

Lu HP, Sun ZY, Qu WC et al (2015) Big data-driven based real-time traffic flow state identification and prediction. Discret Dyn Nat Soc 2015(1):284906

Lv Y, Duan Y, Kang W et al (2015) Traffic flow prediction with big data: a deep learning approach. IEEE Trans Intell Transp Syst 16(2):865–873

Lv M, Hong Z, Chen L et al (2021) Temporal multi-graph convolutional network for traffic flow prediction. IEEE Trans Intell Transp Syst 22(6):3337–3348

Lv Z, Xu J, Zheng K, et al (2018) LC-RNN: A deep learning model for traffic speed prediction. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18, pp 3470–3476

Ma Y, Zhang Z, Ihler A (2020) Multi-lane short-term traffic forecasting with convolutional LSTM network. IEEE Access 8:34629–34643. https://doi.org/10.1109/ACCESS.2020.2974575

Ma C, Zhao Y, Dai G et al (2022) A novel STFSA-CNN-GRU hybrid model for short-term traffic speed prediction. IEEE Trans Intell Transp Syst 24(4):3728–3737

Mandhare PA, Kharat V, Patil C (2018) Intelligent road traffic control system for traffic congestion a perspective. Int J Comput Sci Eng 6(7):2018

Qadri SSSM, Gökçe MA, Öner E (2020) State-of-art review of traffic signal control methods: challenges and opportunities. Eur Transp Res Rev 12:1–23

Qu L, Lyu J, Li W et al (2021) Features injected recurrent neural networks for short-term traffic speed prediction. Neurocomputing 451:290–304

Rajalakshmi V, Ganesh Vaidyanathan S (2022) Hybrid CNN-LSTM for traffic flow forecasting. In: Mathur G, Bundele M, Lalwani M, et al (eds) Proceedings of 2nd International Conference on Artificial Intelligence: Advances and Applications. Springer Nature Singapore, Singapore, pp 407–414

Song X, Li W, Ma D et al (2018) A match-then-predict method for daily traffic flow forecasting based on group method of data handling. Comput-Aided Civ Infrastruct Eng 33(11):982–998

Song Q, Ming R, Hu J, et al (2020) Graph attention convolutional network: Spatiotemporal modeling for urban traffic prediction. In: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), IEEE, pp 1–6

Tan H, Wu Y, Shen B et al (2016) Short-term traffic prediction based on dynamic tensor completion. IEEE Trans Intell Transp Syst 17(8):2123–2133

Tedjopurnomo DA, Bao Z, Zheng B et al (2022) A survey on modern deep neural network for traffic prediction: trends, methods and challenges. IEEE Trans Knowl Data Eng 34(4):1544–1561. https://doi.org/10.1109/TKDE.2020.3001195

Vaswani A, Shazeer N, Parmar N et al (2017) Attention is all you need. Adv Neural inform Process Syst 30:140036

Wang C, Zuo K, Zhang S et al (2023) PFNet: large-scale traffic forecasting with progressive spatio-temporal fusion. IEEE Trans Intell Transp Syst 24(12):14580–14597

Wu Y, Tan H, Qin L et al (2018) A hybrid deep learning based traffic flow prediction method and its understanding. Transport Res Part C: Emerg Technol 90:166–180

Xia Z, Zhang Y, Yang J et al (2024) Dynamic spatial-temporal graph convolutional recurrent networks for traffic flow forecasting. Expert Syst Appl 240:122381

Xie Z, Lv W, Huang S et al (2019) Sequential graph neural network for urban road traffic speed prediction. IEEE Access 8:63349–63358

Xu X, Jin X, Xiao D et al (2023) A hybrid autoregressive fractionally integrated moving average and nonlinear autoregressive neural

network model for short-term traffic flow prediction. J Intell Transport Syst 27(1):1–18

Yin X, Wu G, Wei J et al (2021) Deep learning on traffic prediction: methods, analysis and future directions. IEEE Trans Intell Transp Syst 23(6):4927–43

Yuan H, Li G (2021) A survey of traffic prediction: from spatio-temporal data to intelligent transportation. Data Sci Eng 6:63–85

Yu B, Yin H, Zhu Z (2018) Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In: International Joint Conferences on Artificial Intelligence Organization, pp 3634–3640

Zang D, Ling J, Wei Z et al (2019) Long-term traffic speed prediction based on multiscale spatio-temporal feature learning network. IEEE Trans Intell Transp Syst 20(10):3700–3709

Zhang S, Zhou L, Chen X et al (2020) Network-wide traffic speed forecasting: 3D convolutional neural network with ensemble empirical mode decomposition. Comput-Aided Civ Infrastruct Eng 35(10):1132–1147

Zhao D, Dai Y, Zhang Z (2011) Computational intelligence in urban traffic signal control: a survey. IEEE Trans Syst, Man, Cybern Part C (Appl Rev) 42(4):485–494

Zheng H, Lin F, Feng X et al (2020) A hybrid deep learning model with attention-based Conv-LSTM networks for short-term traffic flow prediction. IEEE Trans Intell Transp Syst 22(11):6910–6920

Zhou J, Shuai S, Wang L et al (2022) Lane-level traffic flow prediction with heterogeneous data and dynamic graphs. Appl Sci 12(11):5340