Urban Traffic Flow Congestion Prediction Based on a Data-Driven Model

Kai Zhang¹,², Zixuan Chu³, Jiping Xing⁴, Honggang Zhang¹,² and Qixiu Cheng¹,⁵,∗

¹ School of Transportation, Southeast University, Nanjing 211189, China; zhangk2019@seu.edu.cn (K.Z.);
zheng_honggang@seu.edu.cn (H.Z.)
² Key Laboratory of Transport Industry of Comprehensive Transportation Theory (Nanjing Modern
Multimodal Transportation Laboratory), Ministry of Transport, Nanjing 211135, China
³ School of AI and Advanced Computing, Xi’an Jiaotong-Liverpool University, Suzhou 215400, China;
zixuan.chu23@student.xjtlu.edu.cn
⁴ Xingmin Intelligent Transportation Systems (Group) Co., Ltd., Yantai 265716, China; xingjiping@seu.edu.cn
⁵ University of Bristol Business School, University of Bristol, Bristol BS8 1PY, UK
∗ Correspondence: qixiu.cheng@bristol.ac.uk
† These authors contributed equally to this work.

Abstract: Intelligent transportation systems need to realize accurate traffic congestion prediction. The spatio-temporal features of traffic flow are essential to analyze and predict congestion. Our study proposes a data-driven model to predict the traffic congested flow. Firstly, the traffic zone/grid method is used to store the local area roads’ average speed of the vehicles. Second, the discrete snapshot set is proposed to characterize traffic flow’s spatial and temporal features over a continuous period. Third, the evolution of traffic congested flow in various time dimensions (weekly days, weekend days, and one week) is examined by transforming the global urban transportation network into traffic zones. Finally, the data-driven model is constructed to predict urban road traffic congestion by using the extracted spatio-temporal characteristics of traffic zones’ traffic flow, the snapshot set of which serves as inputs for this model. The model adopts the convolutional LSTM network to learn the temporal and local spatial features of traffic flow, while utilizing a convolutional neural network to effectively capture the global spatial features inherent in traffic flow. The numerical experiments are conducted on two cities’ transportation networks, and the results demonstrate that the performance of the proposed model outperforms traditional traffic flow prediction models.

Keywords: congestion analysis; traffic flow prediction; spatial-temporal model; deep learning model

MSC: 68T07; 37M10

1. Introduction

With the advancement of society and the growing number of vehicles, traffic congestion emerged as a ubiquitous issue, significantly affecting people’s productivity and societal progress. Predicting city-wide future traffic speed distribution propagation is crucial for developing proactive strategies to decrease traffic congestion. In recent years, the transportation industry witnessed the rapid development of new technologies such as artificial intelligence, big data, and the Internet of Things. These technologies generated vast amounts of data from various devices. Given the proliferation of extensive traffic flow data and relative environmental data, predicting traffic flow and roadway congestion within the transportation domain garnered extensive attention in recent years [1–3]. Researchers generally estimated the short-term traffic flow by studying historical traffic data [4]. Existing methods for predicting traffic flow primarily fall into model-driven and data-driven approaches [5]. The model-driven approach involves establishing mathematical models of traffic flow parameters to estimate the traffic flow state [6]. On the other hand, data-driven
approaches utilize intelligent mining of hidden information in historical data to estimate the future state of traffic flow iteratively, often referring to the machine learning and deep learning models.

The urban transportation network system exhibits a complex nonlinear structure influenced by multiple factors, such as road capacity, drivers, passengers, weather, and holidays [7]. Model-driven prediction algorithms struggle to utilize these factors’ data to effectively analyze and predict traffic flow. However, data-driven technologies such as artificial intelligence and big data offer new engines for technological advancement. Although still in the early stages of application exploration, they have much potential for having transformative impacts on traffic flow prediction, decision-making, and traffic system simulations [8,9]. The data-driven approaches such as artificial neural networks (ANNs) gained attention for traffic flow forecasting. The ANNs method does not require constructing a physical model, and it could utilize abundant historical data to study optimal features for prediction [10].

The data-driven traffic prediction models technique offers several advantages. Firstly, data-driven models use abundant historical traffic data from various sources, such as traffic surveillance cameras, loop detectors, and GPS devices. By harnessing this wealth of data, data-driven models can more accurately capture the complex patterns and relationships in traffic flow [11,12]. Secondly, these models can adapt and learn from new data, continuously improving their predictive performance, which is especially vital in dynamic traffic environments where conditions can change rapidly [13]. Additionally, data-driven models can incorporate diverse factors influencing traffic flow, such as weather conditions, time of day, and road network characteristics [14]. By considering these factors, data-driven models can provide more precise and reliable predictions than traditional approaches relying solely on historical traffic data.

1.1. Literature Review

Model-driven prediction approaches were widely used in traffic flow analysis and forecasting, including moving averages, exponential smoothing, the Kalman filter, and autoregressive moving averages (ARIMA) models [15–19]. For instance, Williams et al. [15] employed the exponential smoothing method to analyze and forecast traffic flow in the city transportation network. Ding et al. [16] extended a spatio-temporal ARIMA model to forecast traffic flow in urban areas at five-minute intervals. Li et al. [17] combined ARIMA and support vector regression to enhance traffic prediction performance. Then, Xia et al. [18] introduced a non-parametric K-nearest neighbor (KNN) model for urban road traffic speed and flow prediction. Chang et al. [19] developed a dynamic multi-interval traffic flow prediction method based on KNN non-parametric regression models.

Compared with these above model-driven approaches, data-driven approaches utilizing neural networks demonstrated their effectiveness in traffic flow forecasting [13,14,20–27]. Zhu et al. [24] proposed a regression prediction traffic flow model based on multi-layer perceptron (MLP) neural networks. Kumar et al. [20] employed an ANN model to predict short-term traffic flow using traffic volume, speed, density, and time as input variables. Lv et al. [21] applied deep learning models to capture the complex traffic flow characteristics without prior knowledge. Then, Ma et al. [22] introduced a convolutional neural network (CNN) method to estimate traffic flow speed in large-scale networks. They proved that CNN utilizes convolutional filter layers to extract local features and simulate the spatial dependencies within the city using sliding windows. Zhang et al. [14] developed a deep learning framework combining residual and convolutional networks. They integrated traffic flow data and weather data as inputs to capture temporal similarity in various road areas and optimize traffic flow prediction. Moreover, Luo et al. [23] and Wu et al. [25] presented the short-term traffic flow prediction model based on the hybrid deep learning approach. In addition, Zhao et al. [26] introduced a long short-term memory (LSTM) neural network to estimate traffic flow within the scope of a traffic network. Furthermore, Duan et al. [27] applied the LSTM network to study and estimate commuter travel time, and Kim et al. [13]
combined linear regression with LSTM to forecast travel demand for New York City. The above methods cannot effectively consider both the spatial and temporal characteristics of city-scale traffic flow, which greatly impacts the model’s prediction accuracy.

It is well known that CNN focuses on capturing local spatial features, while LSTM specializes in modeling long-term dependencies in sequential data [11,28–32]. Over the past few years, the coronavirus disease (COVID-19) pandemic significantly impacted people’s travel behavior [28]. In light of this, Islam et al. [29] employed a combination network of CNN and LSTM to train X-ray images of COVID-19, assisting doctors in diagnosing and treating patients with this disease. Furthermore, Koller et al. [30] utilized a combined CNN and LSTM network to recognize the meanings of different hand shapes and lip-reading representations in sign language videos. Chen et al. [31] successfully applied a CNN and LSTM network to predict the remaining useful life of lithium-ion batteries. In the context of traffic problems, Bogaerts et al. [11] utilized a hybrid CNN and LSTM network to train travel demand data for online car-hailing services, enabling effective prediction of traffic demand within a timeframe ranging from the next five minutes to two hours. Additionally, Archana and Sanjay used this hybrid model to study the traffic flow during rainfall [32].

Moreover, the dynamic graph neural network is utilized to study traffic flow’s spatial and temporal features [5,12]. Inspired by the findings of these studies, our study proposes a tailored data-driven model to study the spatial and temporal features of city-scale traffic flow and predict the short-term traffic flow.

1.2. Objective and Contribution

This paper introduces a novel approach to analyzing and predicting urban traffic congested flow. This work aims to analyze the spatial-temporal feature of traffic flow and develop a tailored data-driven model to estimate the short-term traffic congested flow. The main contributions of the paper are as follows:

1. the discrete snapshot set is proposed to store the spatial and temporal features of traffic flow over a continuous period;
2. the evolution of traffic flow is analyzed in various time dimensions (weekly days, weekend days, and one week);
3. the data-driven model was constructed to predict urban traffic congestion, combining ConvLSTM, batch normalization (BN) and CNN network to study the traffic flow’s spatial and temporal features;
4. the numerical experiments are conducted on two Chinese cities’ transportation networks, and the proposed model’s performance outperforms traditional statistical and machine learning methods.

The remainder of this paper is organized as follows: Section 2 presents the mathematical formulations of the clustering algorithm and traffic zone. Section 3 introduces the traffic congested flow prediction data-driven model. We present the numerical experiments in Section 4. Finally, we conclude this study in Section 5.

2. Methods Description

2.1. Clustering the Congested Regions

Within the framework of the Urban Traffic Management Evaluation Index System [33], the average velocity of motorized vehicles was employed to quantitatively assess the congestion levels prevalent on urban roadways. The congestion classification for urban roads is shown in Table 1.

According to Table 1, the average vehicle speed below 30 km/h is classified as a congested road. To identify congestion events, this study first extracted clustered GPS trajectory sequences from the floating vehicle tracks (continuous slow-speed travel lasting for a minimum of 5 min with a speed below 30 km/h). Subsequently, abnormal behaviors were filtered based on vehicle engine state information to obtain the trajectory sequences [34]. These extracted congestion trajectories were aggregated to determine the road congestion state.
Table 1. The congestion classification for urban roads.

<table>
<thead>
<tr>
<th>Speed (km/h)</th>
<th>Traffic Status</th>
<th>Congestion Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v &gt; 50$</td>
<td>Very smooth</td>
<td>0</td>
</tr>
<tr>
<td>$v &gt; 30$</td>
<td>unblocked</td>
<td>1</td>
</tr>
<tr>
<td>$20 &lt; v \leq 30$</td>
<td>Light congestion</td>
<td>2</td>
</tr>
<tr>
<td>$10 &lt; v \leq 20$</td>
<td>Moderate congestion</td>
<td>3</td>
</tr>
<tr>
<td>$v \leq 10$</td>
<td>Severe congestion</td>
<td>4</td>
</tr>
</tbody>
</table>

The congested events’ spatial distribution can be obtained by clustering the travelers’ pick-up and drop-off points [35,36]. Our study uses these points and employs the graph-based spatial clustering of applications with a noise (G-DBSCAN) algorithm to find the congested regions in the study area (Xi’an city, China), as illustrated in Figure 1. The G-DBSCAN algorithm has a graph-based index structure and can be efficient for the nearby search operations [37].

The silhouette coefficient is used as an evaluation indicator for the G-DBSCAN algorithm. $A$ denotes the average distance from it to other samples in the same cluster, and $B$ denotes the average distance from it to samples in the nearest different cluster, and its contour coefficient is:

$$S(i) = \frac{B(i) - A(i)}{\max\{A(i), B(i)\}}$$

(1)

where the range of $s(i)$ is $[-1, 1]$. When $s(i)$ is close to 1, it means that the clustering of sample $i$ is reasonable. If $s(i)$ is close to $-1$, it means that the sample $i$ should be classified into another cluster. If $s(i)$ is approximately 0, then the sample $i$ is on the boundary between the two clusters.

The numerical values accompanying the icons in Figure 1 represent the count of drop-off points within the current hotspot. These congestion zones are primarily concentrated within the city’s central region, with a significant portion of them spatially aligning near the main arteries of the second ring road. The most congested area has an impressive 234,308 travel flow. The underlying cause for this occurrence that lies in the periphery of the second ring road is dotted with numerous workplaces, commercial districts, and educa-
tional institutions. The surge in traffic demand is attributed to the extensive commuting of employees between residential areas and workplaces and students traveling back and forth between homes and schools. This traffic surge is particularly prominent during morning and evening rush hours, resulting in heightened traffic volumes along the second ring road.

2.2. Partitioning Traffic Zones

The study area’s traffic network structure can be gridded to traffic zones established according to the road spatial location information, as shown in Figure 2a.

![Figure 2](image-url)

(a) The grid representation of city transport network. (b) The representation of SnapShotSet.

The city-scale traffic zone’s flow spatial information can be represented using a matrix, as shown:

$$X_t = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1J} \\ x_{21} & x_{22} & \ldots & x_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ x_{I1} & x_{I2} & \ldots & x_{IJ} \end{bmatrix}$$

(2)

where $I, J$ denotes the traffic zone’s latitude and longitude range, and $x_{ij}$ represents the average traffic flow speed (ATFS) in the region $(i, j)$ at timeslot $t$. Note that the traffic flow data recorded by floating cars equipped with GPS devices include the vehicle’s speed, position, and recording time.

The matrix is constructed to effectively capture the spatio-temporal features of the traffic flow. The matrix accurately portrays the spatial relationship between traffic zones, with adjacent values representing traffic flow between adjacent zones [27]. In essence, traffic flow over a continuous period can be represented by discrete snapshots as follows:

$$\text{SnapShotSet} = \bigcup_{t=1}^{T} X_t$$

(3)

where $T$ represents the number of the time slot. Therefore, SnapShotSet stores the spatio-temporal information of traffic flow.

Figure 2b provides the snapshot set of the entire traffic network. The city-scale traffic network is partitioned into $I \times J$ zones, each zone with one or more roads. The value of each zone denotes the traffic flow average speed at timeslot $t$. The detailed process is as follows:

Firstly, the study area is divided into distinct traffic zones, and the ATFS of each zone during each period $t$ is calculated. Subsequently, these average speeds are mapped onto the corresponding grids. If multiple links traverse the same zone, their speeds are averaged to determine the zone’s speed. Figure 3a visually represents this process, emphasizing its application to arterial roads. The green curve shows the main road with an average...
speed of 60 km/h, while the red curve shows the secondary road with an average speed of 40 km/h. At the intersection of the main road and the secondary road, the average speed of the main road is 50 km/h.

(1) Road vehicles average speed  
(2) Partition traffic zones  
  (a) The average speed of traffic zones.  
(3) Average speed of each zone

(10 a.m.  11 a.m.  12 a.m.)

(b) The traffic flow evolutionary in weekend day’s noon peak.

Figure 3. The zone/grid representation of spatio-temporal trajectory data.

We aggregate the traffic speed data from floating vehicle tracks within each zone on each timeslot and employ color variations to visually depict the ATFS, and a grid-based depiction of urban road traffic flow is attained, as shown in Figure 3b. This figure represents the average traffic flow speed in the study area during the weekend, specifically from 10 a.m. to 12 a.m. In this representation, darker colors indicate slower vehicle speeds and more congested traffic flow on the corresponding road, while lighter colors indicate higher vehicle speeds and smoother traffic conditions. As time progresses from 10 a.m. to 12 a.m., the traffic flow gradually increases, experiencing significant growth on arterial roads. Consequently, the next section aims to study the traffic flow features by utilizing the traffic flow snapshot set from the current and preceding moments.

Furthermore, feature extraction is crucial for traffic prediction in machine learning/deep learning. High-quality data features enable ordinary models to achieve accurate predictions [14]. In our research, we use the insights gained from data analysis to extract relevant features that are then incorporated into the model training process. The features analysis conducted is instrumental in helping us develop a more accurate traffic prediction model. By incorporating the spatio-temporal characteristics of traffic flow identified in our study, we can train a model that achieves superior accuracy in predicting traffic flow in the next section.
3. Traffic Congested Flow Prediction Data-Driven Model

In order to accurately predict traffic congestion across the entire transport network, it is essential to simultaneously incorporate the temporal and spatial dimensions of traffic flow [38]. Therefore, this paper presents the traffic congested flow prediction model, which investigates the traffic flow spatio-temporal features from the snapshot set.

3.1. Prediction Data-Driven Model

Traffic flow evolution exhibits significant temporal dependencies, wherein flow conditions from previous hours can exert a lasting influence on the present state. LSTM models can store the traffic flow’s temporal features by using the memory units that enable them to decide whether to retain and update past hidden states, rendering them particularly effective in time series prediction [26]. Given the time-related attributes of traffic flow, LSTM is apt for capturing the temporal dependencies in traffic flow evolution. To enhance the handling of spatio-temporal traffic flow data within the data-driven model, we draw inspiration from convolutional LSTM (ConvLSTM), an extended LSTM model detailed in the literature [39]. ConvLSTM effectively leverages both temporal features and prior local traffic flow states to predict future states. Deep learning models commonly employ batch normalization (BN) layers to expedite convergence, enhance generalization, and mitigate gradient vanishing issues. Moreover, congestion in a particular traffic zone not only impacts neighboring zones, but can also affect remote zones. The convolutional neural network (CNN) has inherent local connectivity characteristics, making it adept at effectively addressing spatially correlated problems [22].

Therefore, our study proposes a data-driven model to predict traffic congested flow. This model comprises four combined ConvLSTM and BN layers, and a CNN layer, as illustrated in Figure 4. This model integrates time series features and spatial relationships within a hybrid deep learning architecture to study the spatio-temporal characteristics of traffic flow evolution.

![Figure 4](image)

**Figure 4.** The structure of a data-driven prediction mode.

3.2. The Process of Prediction Model

Step 1: Input data.

The input is a series of discrete snapshots that represent the traffic flow evolution over a continuous period. Each snapshot stores the average traffic flow speed of city-wide traffic zones. The relationships of these zones can represent the spatial relationships of different roads’ traffic flow.

Step 2: Four ConvLSTM-BN layers.

The ConvLSTM layers can study the temporal features and explore the local spatial characteristics of traffic flow [39]. Moreover, including the normalization layer improves training accuracy by re-normalizing activation values within each batch.
In the ConvLSTM layer, each time series element is processed separately after convolution and merging, and the output becomes the time series vector. The network structure of ConvLSTM is shown in Figure 5, which combines the convolution layer and LSTM layer. For more details, please see the literature [39]. Each element of the snapshot is a spatial correlation of traffic flow among multiple roads in the local region. The time series snapshot set is then used as the input to the LSTM layer. In the LSTM layer of the ConvLSTM, the LSTM module contains the input gate, forget gate, and output gate. To explore the temporal features of traffic flow, the LSTM layer takes the first 80 percent of time series data $\bigcup_{t=1}^{0.8T} X_t$ as input. The output of hidden state matrix is denoted as $H_t$ in time $t$.

![Figure 5. The network structure of ConvLSTM.](image)

To study the temporal characteristics and local spatial characteristics of traffic flow, we set the following equations:

**Input gate:**

\[
in_t = \sigma(W^i_x * X_t + W^i_h * H_{t-1} + b^i).\]  

**Forget gate:**

\[
f_t = \sigma(W^f_x * X_t + W^f_h * H_{t-1} + b^f).\]  

**Memory unit:**

\[
g_t = f_t \circ g_{t-1} + in_t \circ \tanh(W^g_x * X_t + W^g_h * H_{t-1} + b^g).\]  

**Output gate:**

\[
out_t = \sigma(W^o_x * X_t + W^o_h * H_{t-1} + b^o).\]  

The output of hidden state matrix in time $t$:

\[
H_t = out_t \circ \tanh(g_t)\]  

where $W^i_x, W^i_h, W^o_x,$ and $W^o_h$ are the weight matrices connecting the input $X_t$ to three gates and memory units, while $W^f_x, W^f_h, W^g_x,$ and $W^g_h$ represent the weight matrices linking the hidden state matrix $H_{t-1}$ to these gates and input units. Additionally, $W^i_x, W^g_x,$ and $W^o_h$ correspond to the weight matrices connecting the memory unit to three gates. The deviations among three gates and memory units are denoted by $b^i, b^f, b^o, b^g$. Moreover, the
symbol ∗ denotes the convolution operation; ◦ signifies the scalar product of two vectors; and σ() and tanh() denote the nonlinear activation function, expressed as follows:

\[ \sigma = \frac{1}{1+e^x} \]  
\[ \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \]

We add the BN layer following each ConvLSTM layer to enhance the model’s generalization and mitigate gradient vanishing. The BN layer typically requires configuring the scaling coefficient \( \gamma \) and translation coefficient \( \beta \), which are utilized to normalize each batch’s samples by computing the batch mean \( \mu \) and variance \( \sigma^2 \). The BN layer’s output can be expressed as:

\[ \tilde{X}_i = \gamma \left( \frac{X_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \right) + \beta \]  
\[ \mu = \frac{1}{N} \sum_{i=1}^{N} \tilde{X}_i \]  
\[ \sigma^2 = \frac{1}{T} \sum_{i=1}^{T} (\tilde{X}_i - \mu)^2 \]

where \( N \) represents the batch size, and \( \tilde{X}_i \) denotes each sample within the batch. In the context of the traffic prediction model input, the output of the convolutional part in the ConvLSTM preceding the BN layer is denoted as \( (N, P, I, J) \), where \( P \) signifies the number of output channels, and \( I \) and \( J \) represent the width and height of the output feature map. Consequentially, each sample within the batch can be defined as \( \tilde{X}_{p,i,j} \). The BN layer effectively normalizes each \( \tilde{X}_{p,i,j} \) individually, leading to the calculation of \( \mu \), which corresponds to \( P \times I \times J \).

Step 3: CNN layer.

The last layer, CNN, is utilized to capture the global spatial features of traffic flow. CNN could extract traffic flow spatial features by convolving the result of the preceding ConvLSTM-BN layer with the filter. \( H_t \) denotes the input of the CNN layer. \( H_{l,r} \) denotes the output of the \( l \)th layer of the \( r \)th filter, \( H_{l-1,k} \) is the preceding layer output of the \( k \)th filter, and \( O_{l} \) can be obtained by Equation (14),

\[ H_{l,r} = f \left( \sum_k w_{l,r} \ast H_{l-1,k} + b_{l,r} \right) \]  

where \( f \) represents the nonlinear activation function; \( K \) is the depth of CNN layer; \( w_{l,r} \) and \( b_{l,r} \) represent weights and deviations.

Step 4: Output the future short-term traffic flow.

The trained model \( f_{w,b}(\cdot) \) is obtained from the CNN layer. We can determine each traffic zone’s estimated traffic flow average speed values based on the predicted result \( Y_t \).

\[ Y_t = f_{w,b} \left( \cup_{i=1}^{T} X_i \right), t \in \left( [0.8 * T], T \right] \]

The model stops until the acceptable error \( \varepsilon_1 \) and \( \varepsilon_2 \) are reached; otherwise, repeat the above iterative process. The mean absolute error (MAE) and root mean square error (RMSE) are employed as evaluation criteria to evaluate the efficacy of the proposed data-driven model. A lower value for these metrics signifies enhanced prediction accuracy and a more robust feature expression capability of the model.

\[ MAE \leq \varepsilon_1 \cap RMSE \leq \varepsilon_2 \]
\[ MAE = \frac{1}{T} \sum_{i=1}^{T} |Y_i - X_i| \]  
\[ RMSE = \sqrt{\frac{1}{T} \sum_{i=1}^{T} (Y_i - X_i)^2} \]

where \( X_i \) and \( Y_i \) denote the observed and forecasted traffic flow matrix in time \( t \).

### 3.3. Model Parameters and Experimental Platform

The data-driven model in this study utilizes empirical values for its parameters. The initial four ConvLSTM layers have \( 2 \times 2 \) filters, while the CNN layer has \( 3 \times 3 \) filters. To capture short-term temporal features, the feature dimension at each time step is set to 25, and the regularization norm for the convolutional layer weights is set to 0.002. The model uses a learning rate of \( 5.0 \times 10^{-3} \) for 160 iterations and an attenuation parameter of 0.9. A batch size of 6 is utilized. Dropout layers are implemented to prevent overfitting in the CNN. Python is the programming language used for model implementation and training. The experiments are conducted on a Windows platform with an Intel i7 CPU, 16 GB of random access memory, and a TITAN XP GPU.

### 4. Numerical Experiments

#### 4.1. Data Description

The evaluation uses GPS vehicle trajectory data collected from the transportation networks of two Chinese cities (Xi’an and Shenzhen), as shown in Figure 6. Table 2 provides the description of two datasets. The first dataset is the three-month float vehicles’ GPS trajectory in Xi’an city, China. The second dataset consists of the one-week float vehicles’ GPS trajectory data [40] in Shenzhen, China. Table 3 provides details of the data format, including license plate numbers, recording times, longitude, latitude, vehicle instantaneous speeds, speeds, and direction angles. The “**” indicates the province where the license plate belongs. The first eighty percent of the dataset is used as a training dataset to train the proposed model, while the remaining twenty percent is utilized as the test dataset.
Table 2. The description of two datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Data Scale</th>
<th>Study Area</th>
<th>Data Type</th>
<th>Storage Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xi’an city, China</td>
<td>Three months (1 September to 30 November 2016)</td>
<td>396.23 km(^2)</td>
<td>float vehicles’ GPS trajectory</td>
<td>80.8 GB</td>
</tr>
<tr>
<td>Shenzhen city, China</td>
<td>One week (25–31 March 2018)</td>
<td>16.2 km(^2)</td>
<td>float vehicles’ GPS trajectory</td>
<td>199.2 MB</td>
</tr>
</tbody>
</table>

Table 3. The trajectory information of vehicles.

<table>
<thead>
<tr>
<th>Vehicles ID</th>
<th>Timestamp</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Speed (Km/h)</th>
<th>Direction Angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>*AXXXXX</td>
<td>20160901 15:12:01</td>
<td>34.094</td>
<td>108.089 270</td>
<td>45</td>
<td>106</td>
</tr>
<tr>
<td>*AXXXXX</td>
<td>20160901 15:14:01</td>
<td>34.094</td>
<td>108.082 270</td>
<td>50</td>
<td>106</td>
</tr>
<tr>
<td>*BXXXXX</td>
<td>20180325 16:08:00</td>
<td>22.544</td>
<td>114.003 635</td>
<td>48</td>
<td>75</td>
</tr>
<tr>
<td>*BXXXXX</td>
<td>20180325 16:10:0</td>
<td>22.544</td>
<td>114.003 632</td>
<td>50</td>
<td>78</td>
</tr>
</tbody>
</table>

4.2. Analysis the Evolutionary Process of Urban Road Congestion

4.2.1. Analysis of the Most Congested Main Road in Xi’an City

This study conducts macro-level congestion analysis within the area of the Xi’an third ring road, and the traffic zones were utilized to analyze the congested areas. We use the Python library to visualize the tested area’s traffic flow (average speed) for the whole day, as shown in Figure 7.

Figure 7a represents the distribution of traffic flow speeds within the third ring road area on the work day (4 September 2016, Tuesday). The darker colors indicate higher speeds on road segments, with white areas representing non-road sections such as buildings, parks, and lakes. It was discovered that the second ring road and its surrounding areas are the most congested areas where the location of the second loop is marked by a bold red line in Figure 6a. Further, we analyzed the temporal patterns of congestion appearance and dissipation. Figure 7a reveals that during the midnight period (from 2 a.m. to 6 a.m.), the second ring road displays higher speeds than other areas. However, in the morning, noon, and evening peak hours, the vehicles’ speed on the second ring road will decrease significantly, and the low-speed status could last for a long time. This implies that during the three peak periods of the day, the second ring road experiences lower speeds, indirectly indicating higher traffic volume. From 8 p.m. to 11 p.m., the speed on the second ring road gradually increases, as reflected by the brighter color in Figure 7a, and indicates a gradual decrease in citizens’ travel frequency. Moreover, citizen activities are less frequent between 2 a.m. and 6 a.m., with most citizens resting.

Furthermore, we analyze the specific congestion conditions of the second ring road. Figure 7b visually represents the traffic speed of each section of the second ring road over one day. The brighter the color of each segment on the second ring road, the higher the traffic speed. Each image is labeled with a specific time, such as “hours2.5” for 2:30 a.m. It is evident that the traffic speed on the second ring road is higher from 12 a.m. to 6 a.m., whereas it decreases during the morning and evening peak hours. It reflects the travel patterns of commuters, who travel during the morning and evening peak hours.

Due to the periodic feature of the vehicle trajectory, we analyze the traffic flow evolution on the second ring road from early morning peak time (5 a.m. to 7 a.m.) for one week. As shown in Figure 8a, each row consists of seven small graphs representing each day of the week, and each column represents the speed distribution on the second ring road from 5 a.m. to 7 a.m. Above each small image is a time tag, such as “day1b ‘2016111405’”, indicating Monday (14 November 2016) at 5 a.m. It shows that the traffic flow is periodic during the morning peak hours, and the traffic speed at the same moment on the two adjacent days (except Sunday) is similar. Because most people rest/sleep during Sunday morning rush hour and hardly travel during this period, we obtain that the traffic flow speed is near zero.
Figure 7. Evolution of traffic flow throughout the day in Xi’an city.

(a) Speed distribution in the third ring road of Xi’an city.

(b) 4 h congestion change in the second ring road.
Due to the periodic feature of the vehicle trajectory, we analyze the traffic evolution on the second ring road from early morning peak time (5 a.m. to 7 a.m.) for one week. As shown in Figure 8a, each row consists of seven small graphs representing each day of the week, and each column represents the speed distribution on the second ring road from 5 a.m. to 7 a.m. Above each small image is a time tag, such as "day1b2016111405", indicating Monday (14 November 2016) at 5 a.m. It shows that the traffic flow is periodic during the morning peak hours, and the traffic speed at the same moment on the two adjacent days (except Sunday) is similar. Because most people rest/sleep during Sunday morning rush hour and hardly travel during this period, we obtain that the traffic flow speed is near zero.

Furthermore, we counted the number of congestion events lasting 300 s, including moderate and severe congestion, as described in Figure 8b. It is found that congestion events are periodic. The congestion was more obvious during the morning and evening peaks during the five weekly days. However, the congested events continued from 10 a.m. to 8 p.m. on weekends, especially during the evening rush hours, resulting in lower travel speeds on the second ring road.

4.2.2. The Impact of Residents’ Travel on the Second Ring Road of Xi’an on Weekdays and Weekends

In Figure 9, the x-label represents time slots for a whole day, while the y-label denotes the positions of the fourth section of the second ring road, encompassing its north, west, south, and east roads. By employing a time-space matrix as the channel of an image and encapsulating the daily traffic speeds of the network within images, each image

Figure 8. Analyze morning rush hour traffic flow in the second ring road over the course of a week.
corresponds to a distinct day for the network. These images offer comprehensive traffic insights, revealing congested areas highlighted in red regions and congestion propagation patterns. Notable patterns include oscillating congested traffic and localized clusters exhibiting pinned behavior. An in-depth exploration is available in the literature for a more comprehensive understanding of these traffic patterns [41].

![Speed diagram of a traffic cell w.r.t. space-time dimension.](image)

**Figure 9.** Speed diagram of a traffic cell w.r.t. space-time dimension.

During weekday morning and evening peak hours, as shown in Figure 9a, the second ring road experienced heavily congested events. However, on weekends, as shown in Figure 9b, congested events are only observed during the morning and evening peak hours. Furthermore, following the congestion classification outlined in Table 1, the speed data for the second ring road were categorized into five distinct traffic congestion levels. Figure 10 is the three-dimensional (3D) extension of Figure 9, adding the new z-labels representing speed values used for classifying congestion. Figure 10a provides the 3D traffic flow of the weekly day. During morning and evening peak hours, congestion persists for extended durations, reaching severe levels. Notably, the western part of the second ring road has congested events lasting up to 2 h. However, a distinct pattern emerges on weekends, shown in Figure 10b, with congestion intensifying progressively from noon to evening, except on the north side of the second ring road.

By analyzing the historical traffic data, we identify key patterns and trends that influence traffic flow, such as the impact of peak time and weekly/weekend days on traffic flow. The data analysis in this section plays a crucial role in our research as it enables us to identify the underlying characteristics of traffic flow and use these features in model training in the next section.

4.3. Model Prediction Results and Analysis

4.3.1. Sensitivity Analysis

The proposed model is trained using vehicle trajectory data from Shenzhen city to predict the short-term traffic flow for the next two hours. To decrease/avoid overfitting, the batch normalization (BN) layers are incorporated into our model architecture. A comparative analysis of different ConvLSTM-BN models with varying layers indicates that the four-layer ConvLSTM-BN model achieves the lowest value of RMSE, as depicted in Figure 11. Therefore, our data-driven model combines four ConvLSTM layer models and a CNN layer for traffic flow prediction.
By analyzing the historical traffic data, we identify key patterns and trends that influence traffic flow, such as the impact of peak time and weekly/weekend days on traffic flow. The data analysis in this section plays a crucial role in our research as it enables us to identify the underlying characteristics of traffic flow and use these features in model training in the next section.

### 4.3. Model Prediction Results and Analysis

#### 4.3.1. Sensitivity Analysis

The proposed model is trained using vehicle trajectory data from Shenzhen city to predict the short-term traffic flow for the next two hours. To decrease/avoid overfitting, the batch normalization (BN) layers are incorporated into our model architecture. A comparative analysis of different ConvLSTM-BN models with varying layers indicates that the four-layer ConvLSTM-BN model achieves the lowest value of RMSE, as depicted in Figure 11. Therefore, our data-driven model combines four ConvLSTM layer models and a CNN layer for traffic flow prediction.

Figure 11. Optimal number of ConvLSTM layers in the proposed model.

Figure 12 provides the RMSE and MAE results of the proposed model for different short-term timeliness. It displays the pre-training results of Xi’an city, indicating that the model achieves its lowest RMSE and MAE values of 7.26 and 3.87, respectively, when predicting traffic flow in the next two hours. However, the RMSE and MAE values are relatively high when predicting traffic flow in the next four and eight hours. Then, we use the pre-trained model to train and test the traffic flow in Shenzhen. When forecasting road congestion status for the next two hours in Shenzhen, the model achieves its minimum RMSE and MAE values of 11.70 and 7.68, respectively.

Figure 13 visualizes the prediction result of the proposed data-driven model to analyze the congestion traffic flow on both weekdays and the weekends. In Figure 13, the label "predict:2016052316" represents the predicted traffic flow results from 4 p.m. to 5 p.m. on a weekend day (23 May 2018), while "GroundTruth:2016052316" represents the actual traffic flow during the same period. Figure 13a predicts the future two-hour (evening peak) traffic flow in Xi’an city. It reveals that traffic congestion is mainly observed in the second ring road and the roads within it. Similar prediction results are obtained for Shenzhen city, as depicted in Figure 13b, where traffic congestion is observed on the primary roads. Therefore, Figures 12 and 13 demonstrate the model’s effectiveness in short-term traffic flow prediction.
the pre-trained model to train and test the traffic flow in Shenzhen. When forecasting road congestion status for the next two hours in Shenzhen, the model achieves its minimum RMSE and MAE values of 11.70 and 7.68, respectively.

Figure 12. The performance evaluation of proposed model predicts the short-term traffic flow.

Figure 13 visualizes the prediction result of the proposed data-driven model to analyze the congestion traffic flow on both weekdays and the weekends. In Figure 13, the label “predict:2016052316” represents the predicted traffic flow results from 4 p.m. to 5 p.m. on a weekend day (23 May 2018), while “GroundTruth:2016052316” represents the actual traffic flow during the same period. Figure 13a predicts the future two-hour (evening peak) traffic flow in Xi’an city. It reveals that traffic congestion is mainly observed in the second ring road and the roads within it. Similar prediction results are obtained for Shenzhen city, as depicted in Figure 13b, where traffic congestion is observed on the primary roads. Therefore, Figures 12 and 13 demonstrate the model’s effectiveness in short-term traffic flow prediction.

Figure 13. Forecast of traffic flow speed in the next two hours.
4.3.2. Models Comparison

To further validate the effectiveness of the proposed data-driven model, it was compared against common statistical forecasting methods, including the statistical models, such as the simple average method, ARIMA, and exponential smoothing method (ESM), and neural network algorithms such as MLP, CNN, and LSTM. The RMSE and MAE are used to test the performance of the proposed method. The experimental analysis is conducted with 160 iterations. The parameters of these compared are explained as follows:

The simple average method is the time series-based approach that predicts future traffic flow by averaging historical data over a specific observation period.

The exponential smoothing method (ESM) is derived from the moving average method and assigns weights to past observations, giving more importance to recent data [15]. Two and three exponential smoothing techniques, named two ESM and three ESM, are compared with the proposed method.

Autoregressive integrated moving average (ARIMA) is a general forecasting method designed explicitly for non-stationary time series data in transportation [16].

Multi-layer perceptron (MLP) is a popular neural network model for traffic prediction. It learns from historical data to establish an optimal structure between input and expected output. This MLP has two hidden layers, with the first layer consisting of 12 neurons as the test set model and the second with 8 neurons as the training set model [24].

Convolutional neural network (CNN) method utilizes two convolutional neural layers for the prediction of traffic flow. It employs stacked convolutional and activation layers to capture traffic patterns [42].

Convolutional long short-term memory method (ConvLSTM) is a convolutional recurrent neural network structure widely used for time series analysis. This method has two ConvLSTM layers [39].

Firstly, Figure 14a presents the RMSE values of the comparative methods. Traditional forecasting methods (simple average method, ESM, and ARIMA) can only capture the temporal characteristics of traffic flow, resulting in higher RMSE values. The MLP model is a simple neural network that focuses on studying the temporal characteristics, while the CNN method with two convolutional layers can learn the global spatial characteristics of traffic flow. Although ConvLSTM can learn both temporal and local spatial features, its RMSE performance is lower than that of CNN. The proposed model leverages the convolutional LSTM network to learn the temporal and regional spatial features of traffic flow, while effectively capturing the global spatial features using a convolutional neural network. This data-driven model achieves the minimum RMSE compared to the traditional models, and its evaluation metrics surpass those of the CNN and LSTM models.

Secondly, we added the MAE performance of all algorithms in Figure 14b. Upon further analysis, it is evident that our proposed method achieves the smallest MAE value, which is 4.28. The MAE values are less than 10 among all algorithms except for ESM and ARIMA. Specifically, the MAE values of ConvLSTM and CNN are 6.78 and 5.41, respectively. Moreover, the model demonstrates a stable error rate, making it suitable for large-scale transportation networks. The training process is also completed within a reasonable timeframe. These results highlight the superior short-term traffic flow predictive performance of our method using ConvLSTM and CNN.
Firstly, Figure 14a presents the RMSE values of the comparative methods. Traditional forecasting methods (simple average method, ESM, and ARIMA) can only capture the temporal characteristics of traffic flow, resulting in higher RMSE values. The MLP model is a simple neural network that focuses on studying the temporal characteristics, while the CNN method with two convolutional layers can learn the global spatial characteristics of traffic flow. Although ConvLSTM can learn both temporal and local spatial features, its RMSE performance is lower than that of CNN. The proposed model leverages the convolutional LSTM network to learn the temporal and regional spatial features of traffic flow, while effectively capturing the global spatial features using a convolutional neural network. This data-driven model achieves the minimum RMSE compared to the traditional models, and its evaluation metrics surpass those of the CNN and LSTM models.

Secondly, we added the MAE performance of all algorithms in Figure 14b. Upon further analysis, it is evident that our proposed method achieves the smallest MAE value, which is 4.28. The MAE values are less than 10 among all algorithms except for ESM and ARIMA. Specifically, the MAE values of ConvLSTM and CNN are 6.78 and 5.41, respectively. Moreover, the model demonstrates a stable error rate, making it suitable for large-scale transportation networks. The training process is also completed within a reasonable timeframe. These results highlight the superior short-term traffic flow predictive performance of our method using ConvLSTM and CNN.

Figure 14. The performance comparison of predicting the future two-hour traffic flow in Shenzhen city, China.

5. Conclusions

This paper applies the G-DBSCAN clustering algorithm to obtain the traffic congestion state. The traffic network is decomposed into multiple traffic zones by the gridded method, and the discrete snapshot set is then proposed to store traffic flow’s spatial and temporal features over a continuous period. Moreover, we analyze the evolution of traffic flow in various time dimensions (weekly day, weekend day, and one continued week). Further, the data-driven model was constructed to predict urban traffic congestion, combining the ConvLSTM and CNN network to study the traffic flow’s spatial and temporal features. The numerical experiments are conducted on two cities’ transportation networks, and the proposed model’s performance outperforms both the traditional statistical and machine learning methods.

This study primarily analyzes traffic flow evolution through a statistical lens and utilizes black box deep learning models for prediction, thus this work lacks interpretable algorithms for traffic modeling. The limitation of this study is that it focuses on gathering local road traffic flow data within plane traffic zones as the training and prediction object. As a result, our method may not be able to effectively analyze traffic flow with multiple overlapping layers in space, such as overpasses or complex interchanges [27].

In future work, we not only expand our efforts by incorporating methodologies such as the Gaussian process [43] and Bayesian optimization [44] to enhance the internal
interoperability of our models, but also use the predicted result of the proposed model to study urban road congestion pricing [45].

Author Contributions: Writing—original draft preparation, K.Z. and Z.C.; supervision, Q.C.; Software, K.Z. and J.X.; visualization, Z.C.; writing—review and editing, J.X. and H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant No. 71971060; Laboratory for Traffic & Transport Planning Digitalization Program, No. 2022HFO5; Fundamental Research Funds for the Central Universities, China No. 2242022R40025; Start-up Research Fund of Southeast University, China No. RF1028623057.

Data Availability Statement: The original contributions presented in the study are included in the article, and further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References
3. Jafari, S.; Shahbazi, Z.; Byun, Y. Improving the road and traffic control prediction based on fuzzy logic approach in multiple intersections. Mathematics 2022, 10, 2832. [CrossRef]


35. Xia, Z.; Li, H.; Chen, Y.; Liao, W. Identify and delimitate urban hotspot areas using a network-based spatiotemporal field clustering method. *ISPRS Int. J. Geo-Inf.* 2019, 8, 344. [CrossRef]


42. Ma, X.; Dai, Z.; He, Z.; Ma, J.; Wang, Y. Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors* 2017, 17, 818. [CrossRef]


45. Fall, A.N. Analysis of social acceptability in the implementation of a congestion pricing area in senegal. *Multimodal Transp.* 2022, 1, 100036. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.