Attention-based Spatial-Temporal Graph Convolutional Network for Urban Traffic Volume Estimation

Anh Phuong NGUYEN [a], Tzu-Chang LEE [b]

[a,b] *Department of Urban Planning, National Cheng Kung University, Tainan, 701401, Taiwan*
[a] *E-mail: p28127015@gs.ncku.edu.tw*
[b] *E-mail: jtclee@mail.ncku.edu.tw*

Abstract: This study aims to estimate traffic volume for each road segment in the Taichung Metropolitan area, Taiwan, using a limited number of traffic sensors as ground truth data. To achieve this, we applied the Attention-based Spatial-Temporal Graph Convolutional Network (ASTGCN), a deep learning model that integrates multi-temporal traffic speed data with external factors, such as road attributes and temporal features. By effectively capturing spatial and temporal dependencies, ASTGCN enhances prediction accuracy over traditional forecasting methods. Sensitivity analysis shows that incorporating traffic speed patterns from both the previous and following five-minute intervals improves accuracy. Furthermore, models that integrate additional features, such as time and road characteristics, further improve traffic volume estimation accuracy. These findings offer valuable insights for optimizing urban traffic management, infrastructure planning, and transportation policies.

Keywords: Traffic volume forecasting, traffic patterns, deep learning, traffic flow theory.

1. INTRODUCTION

Traffic volume prediction is a critical component of intelligent transportation systems and plays a key role in supporting effective traffic management. Accurate volume estimation facilitates traffic flow optimization, improves road network efficiency, mitigates congestion, and provides information for decision-making. However, traffic flow prediction remains a challenging problem in transportation planning due to the dynamic and complex nature of urban traffic systems.

Traffic volume estimation methods often rely on historical data collected from fixed sensors, such as loop detectors and surveillance camera (Abadi et al., 2014). While these systems provide reliable data, their deployment is often limited to major roads and highways due to high installation and maintenance costs. Consequently, large portions of urban road networks, particularly local streets and secondary roads in densely populated areas, lack sufficient traffic monitoring infrastructure. The lack of complete data limits the ability to develop a full understanding of traffic patterns across the network, presenting a major barrier to efficient and adaptive traffic management.

With the growing demand for traffic state information on every road segment, traditional sensor-based data collection methods are increasingly insufficient. The emergence of new data collection technologies, coupled with advancements in computational capabilities, has led to alternative approaches to traffic monitoring. Traffic data from probe sources, including mobile devices, navigation apps, and vehicle tracking systems has become a valuable source for obtaining speed information across large-scale road networks. However, while this data provides comprehensive speed coverage, it lacks traffic volume measurements. As a result, developing methods to estimate traffic volume from speed data has become a crucial task for enabling large-scale traffic volume estimation.

This study aims to address this gap by proposing a novel approach to estimate traffic volume across the entire road network of city using multitemporal traffic speed data. We explore the dynamic relationship between traffic volume and speed, while also considering external factors such as time-related features and road topography. Our proposed method employed the Attention-based Spatial-Temporal Graph Convolutional Network (ASTGCN) model, which combines both spatial and temporal dependencies to capture the complex interactions between different road segments.

The main contributions of this work are as follows:

• A method for estimating traffic volume for each road segment based on multiple traffic data sources.

• An approach that incorporates multitemporal traffic speed data, time-related features, and road segment characteristics to improve the accuracy of traffic volume estimation.

• The use of a gravity model to define the connectivity between road segments, which enhances the ability of models to capture traffic flow patterns across the network.

• A framework that does not require extensive or expensive sensor data, making it particularly useful for cities with limited resources or for urban environments where sensor deployment is sparse.

The structure of this study is organized as follows: Section 2 provides a review of related research on traffic volume estimation. Section 3 defines the problem, followed by Section 4, which introduces the proposed methodology. Section 5 outlines the experimental setup, including the datasets and evaluation metrics used. Section 6 presents the results and findings of our research.

2. LITERATURE REVIEW

In this section, we review existing studies on traffic volume forecasting, focusing on two key aspects: model development traffic volume estimation, and factors influencing traffic volume.

2.1 Deep learning for traffic volume forecasting

With the increasing availability of traffic data and rapid advancements in deep learning (DL), researchers have developed a variety of neural network models to enhance the accuracy of traffic prediction. Long Short-Term Memory (LSTM) networks, Graph Convolutional Networks (GCNs), and their spatiotemporal extensions have demonstrated notable effectiveness in capturing the complex temporal and spatial dependencies inherent in traffic systems. LSTM networks excel at modeling temporal sequences, leveraging historical traffic patterns to forecast future traffic volumes. However, traffic flow is also influenced by spatial dependencies among road segments. To address this, GCNs represent road networks as graphs, effectively modeling the spatial relationships between interconnected road segments. Building on this, Spatio-Temporal Graph Convolutional Networks (STGCNs) integrate GCNs with temporal sequence models to jointly learn spatiotemporal dependencies (Yu et al., 2017). More recently, the Attention-based Spatio-Temporal Graph Convolutional Network (ASTGCN) has introduced attention mechanisms to adaptively capture the varying importance of spatial and temporal correlations, leading to improved predictive performance (Guo et al., 2019).

2.2 Factors influence traffic volume

In traffic flow theory, speed and volume are interrelated through a nonlinear relationship, often depicted in fundamental diagrams (Greenshields, 1935). Recognizing this, several studies have investigated the potential of using speed data to estimate traffic volume. For instance, Kwon et al. (2023) proposed an imputation method that leverages speed information to estimate missing traffic volume values. Similarly, Kashyap et al. (2022) discussed the application of deep learning models, such as Long Short-Term Memory (LSTM) networks and GCNs, in traffic volume prediction, highlighting their capability to capture spatiotemporal dependencies and speed variations effectively.

Additionally, several researchers have incorporated external factors into traffic volume estimation models. Temporal features, such as time of day and day of the week, play a critical role due to their significant influence on traffic patterns. Peak hours during weekdays typically exhibit higher traffic volumes compared to weekends. Road characteristics, including road type and the number of lanes, also significantly affect traffic flow, as different road types and capacities either facilitate or constrain vehicle movement (Ganji et al., 2022) .Land use attributes and socioeconomic factors further impact traffic demand; areas with high commercial activity or dense populations often experience increased traffic volumes.

The review of method and feature on estimate traffic volume shows some limitation on knowledge about traffic volume forecasting. First, most studies rely heavily on historical traffic volume data for estimation. However, this approach is not feasible for the entire road network, as traffic volume data is often unavailable for many road segments. Second, there is a lack of research focusing on the relationship between traffic volume and traffic speed, despite their well-established connection in traffic flow theory. Finally, there is a limited number of studies that integrate multiple data sources, such as speed, road characteristics, and external contextual factors, to improve the accuracy of traffic volume prediction.

In this study, we use speed as a key predictor for traffic volume estimation. We apply the Attention-based Spatiotemporal Graph Convolutional Network (ASTGCN) to effectively capture complex spatiotemporal patterns between road segments. Additionally, we investigate the relationship between traffic flow and speed to identify optimal speed patterns that can improve volume estimation accuracy. To further enhance model performance, we incorporate temporal features and road characteristics as an additional feature, improving the representation of interactions between road segments within the prediction framework.

3. PROBLEM DEFINITION

In this study, the road network is defined as undirected graph $G = (V, E, A)$, where V is the set of N nodes, each representing a road segment in the study area. The set E includes edges in graph G that represent the connections between different road segments.

At each timestep, each node in graph G holds a vector of features. The input to model consists of a time series of features. Given dynamic features, and N nodes in the graph model of road segments, all the features over the T timesteps (including n past, current and m future steps) form $X = (x_1, x_2, \dots, x_t, \dots x_T)^T$ as the input, where x_t includes all the features for all the nodes at timestep t. The target variable is the current total traffic volume, denoted as, y_{t}which represents the traffic volume at the timestep t. These temporal features are aggregated and preprocessed to match the sampling frequency x_1. Figure 1 illustrates the dynamic relationship between traffic speed and volume.

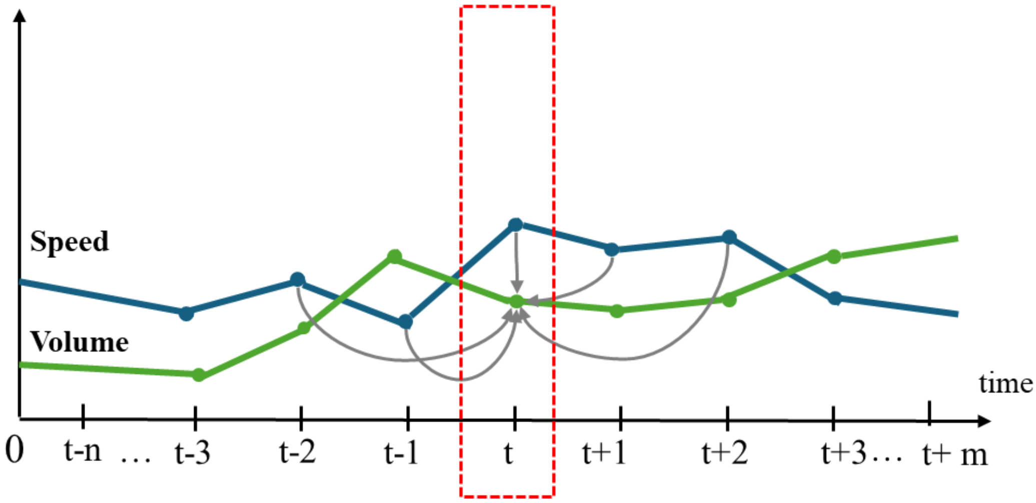

Figure 1. Time series pattern relationship between speed and volume.

4. METHODS

The methodology in this study consists of three key steps: First, dynamic traffic patterns are extracted to construct the speed matrix as input features and the volume matrix as ground truth. Second, the adjacency matrix is constructed using the gravity model to define connections between road segments and form the road network graph. Finally, the structure and mechanism of the ASTGCN model are presented, which estimates traffic volume based on multiple features, including traffic speed, temporal patterns, and road characteristics.

4.1 Time Series Pattern Extraction for Traffic Volume and Traffic Speed

Traffic speed is a dynamic indicator of road conditions that influences traffic volume. The same traffic volume can correspond to different speeds, making predictions based solely on current speed data insufficient. In this study, we analyzed traffic speed from adjacent time steps, following the approach of Gao et al. (2022). Using time-series traffic data, we extracted continuous patterns by considering n previous and m subsequent time steps. The sliding window technique created traffic volume and speed pattern vectors, resulting in traffic volume matrix and speed matrix. The detailed process is illustrated in Figure 2.

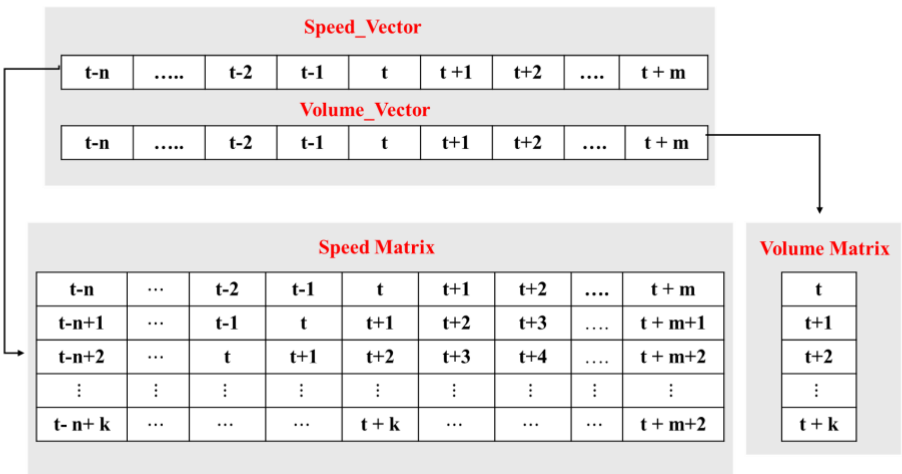

Figure 2. Traffic pattern extraction (Gao et al., 2022)

4.2 Adjacency Matrix Construction Based on Gravity Model

The matrix $A_{N\times N}$ represents the connection strength between road segments. Traditionally, it is a binary matrix with values of 1 indicating a connection between nodes and 0 indicating no connection. However, this approach overlooks the unique traffic characteristics of each road. Therefore, our model constructs the adjacency matrix using the gravity model to effectively capture both spatial and functional dependencies:

$$A_{ij} = K\frac{p_i p_j}{d_{ij}^{\beta}} \tag{1}$$

Where d_{ij} is distance between road segment i and j, p_i and p_j is the number of POI around road segment i and j, reflecting their potential influence on traffic volume, β is a distance decay parameter, and K is a normalization constant.
The distance between road segment we calculate is the shortest driving distance between nodes i and j, obtained using the Open-Source Routing Machine (OSRM) API.

4.3 Attention Based Spatial-Temporal Graph Convolutional Networks (ASTGCN)

In the real world, traffic volume is interconnected across different road segments, and traffic conditions can fluctuate over time. Therefore, effectively capturing both spatial and temporal dependencies is crucial for modeling traffic volume. To achieve this, we employed the ASTGCN (Attention-based Spatiotemporal Graph Convolutional Network) proposed by Guo et al. (2019) for traffic volume forecasting. The model structure consists of spatial temporal blocks and a fully connected. Each spatial-temporal block contains both a spatial-temporal attention module and a spatial-temporal convolution module. These attention mechanisms help capture spatial and temporal correlations among the dynamic features in our traffic volume predictions, allowing the model to weigh the importance of different features and data points effectively. The output from the attention modules feeds into the spatial-temporal convolution module, which captures dependencies between various road segments based on the adjacency matrix and the time series of input features. Figure 3 illustrates the ASTGCN models architecture.

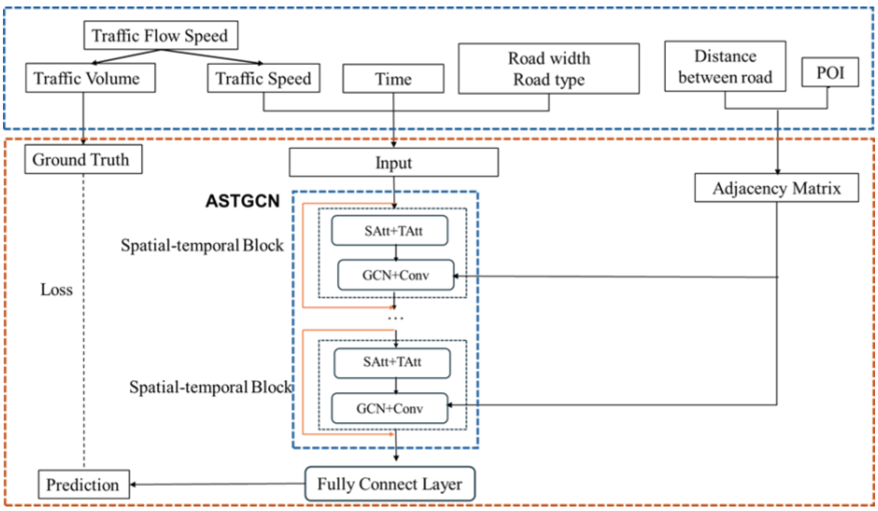

Figure 3. Overview of the Traffic Flow Prediction Model Based on Attention-Enhanced Spatial-Temporal Graph Convolutional Networks (ASTGCN)

4.3.1 Spatial-Temporal Attention Mechanism

The model includes L spatial-temporal blocks, where the input for $(l+1)$th block is:

$$X^l = (x_1, x_2, \dots, x_t, \dots, x_{Tl}) \in R^{N \times C_l \times \tau_l} \tag{1}$$

where C_l denotes features of the input data in the (l + 1) th layer, τ_l denotes the length of the temporal dimension in the l th layer.

Spatial attention focuses on learning the influence of spatial characteristics between nodes and the different influences of each node on graph representations. The spatial attention is then determined as follows:

$$S_{\text{Att}} = P_s \cdot \sigma(X^l W_1) W_2 (W_3 X^l)^T + b_s \tag{2}$$

where P_s and b_s are $N \times N$ learnable parameters, and W_1, W_2, and W_3 are also learnable parameters that are fed into sigmoid function σ as the activation function.
In a similar way, temporal attention procedure to learn the influence on time-series characteristics of data and past traffic information that affects the time point to predict based on the time slice t.

4.3.2 Spatial-Temporal Convolution Module

After the attention modules have processed the data, it becomes better suited for the convolution layer, which captures and extracts both spatial and temporal dependencies. The processed data is then passed to the spatial-temporal convolution module, which operates across both the spatial and temporal dimensions. The convolution structure of the network utilizes spectral graph theory to analyze spatial correlations at each timesteps based on the adjacency matrix. The Laplacian matrix (L) is defined as $L = D - A$, where D is the degree matrix and A is the adjacency matrix. The normalized Laplacian is then applied for graph convolution, represented as follows:

$$g_\theta *_G x = g_\theta(L)\, x = g_\theta(U \Lambda U^T) x = U\, g_\theta(\Lambda)\, U^T x \tag{3}$$

where $*_G\ x$ operates a convolution on the graph G given the signal x and U is Fourier basis.

Next, the convolution process is completed through the inverse Fourier transform to derive the final output $x = U^T x$. However, owing to the large scale of the graph structure, Chebyshev polynomials are applied for operational computation efficiency. The model captures information of neighborhood of 0 to K-1 order of each node. This is achieved by approximately the eigenvalue decomposition on the Laplacian matrix, which is:

$$g_\theta *_G x = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{L}) x \tag{4}$$

where θ consist of K polynomial coefficients, T_k is Chebyshev polynomials,

$T_k(x) = 2x T_{k-1}(x) - T_{k-2}(x)$, where $T_0(x) = 1$, $T_1(x) = x$, and $\tilde{L} = \frac{2}{\lambda_{max}} L - I_N$, $\lambda_{\max}$ represents the maximum eigenvalue of the Laplacian matrix.

The Hadamard product of $T_k(\tilde{L})$ and $SAtt'$ is used to incorporate spatial attention, allowing the model to apply filters for each node at every time step while capturing neighboring information in the spatial dimension. Next, we utilize standard temporal convolution to update the information based on previous time steps. For the lth layer, we have:

$$X^l = \mathrm{ReLU}\left(\Phi * \left(\mathrm{ReLU}(g_\theta *_G X^{l-1})\right)\right) \tag{5}$$

$*$ denotes the standard convolution operation, while Φ represents the parameters of the temporal kernel. ReLU stands for the rectified linear unit activation function. The model used in this study consists of three spatial temporal blocks, which are followed by a fully connected layer with a ReLU activation function to estimate the value of the dependent variable

4.3.3 Estimate traffic volume for the entire network

To estimate traffic volumes across the entire road network, we use traffic speed which covers the entire city, along with external factors such as road attributes and temporal features. First, a global adjacency matrix is created using the gravity model to represent the connectivity between all road segments in the city. The global network is then divided into subgraphs, each containing 16 road segments. During this partitioning, the elements of the adjacency matrix are preserved to accurately reflect traffic characteristics.

Next, a pre-trained model is applied to each subgraph to estimate traffic volumes. Speed-related features extracted from Google_Speed, along with external factors, are incorporated into the model for each subgraph. This approach allows the predictions to capture both local and global traffic patterns effectively. Finally, the traffic volume predictions from all subgraphs are combined to reconstruct the overall traffic conditions across the entire road network. This partitioned modeling approach ensures computational efficiency while maintaining high predictive accuracy across the network.

5. EXPERIMENT

5.1 Study area

The study area for this experiment is the metropolitan area of Taichung City, Taiwan. As a bustling urban center with a diverse road network, Taichung offers a rich environment to capture the complex relationships between traffic volume and its influencing factors. Transportation infrastructure of city, including surveillance cameras and traffic sensors, provides a reliable dataset for modeling. With its varied road types, real-time traffic data, and other contextual factors, Taichung serves as a representative and suitable location for investigating the efficacy of traffic volume forecasting methods that account for multiple features.

5.2 Dataset

This study aims to estimate traffic volume across all road segments by employing the ASTGCN model, which captures the relationship between speed and traffic volume. To facilitate this analysis , traffic data, Points of Interest (POIs), and the road network were utilized.

Traffic data were collected from three sources. The first dataset, named Camera_TDV, used for model development, was obtained from surveillance cameras in Taichung City (https://tcps.nchc.org.tw/). This dataset covers 43 weeks, from January 1 to October 28, 2024, at 5-minute intervals across 16 road segments. Camera_TDV data includes traffic volume, speed, road ID, and timestamps. The camera locations are shown in Figure 4. The locations of the cameras are illustrated in Figure 4.

Figure 4. Study area and location of surveillance cameras

The second dataset, Google_Speed, was collected from Google Map API (Application Programming Interface). Traffic data was retrieved by using the start and end points of each road segment as the departure and destination. This provided real-time travel distance and time for each road segment. Traffic speed was calculated by dividing the travel distance by the travel time. Speed data was collected for each road segment at 5-minute intervals over one week, from October 21 to 27, 2024. This dataset covers 1576 road segments across the city. Google_Speed dataset includes traffic speed, road IDs, and timestamps. This dataset was used to estimate traffic volume across the network.

The third dataset, Sensor_TDV, was obtained from Vehicle Detector Sensors (https://e-traffic.taichung.gov.tw/ATIS_TCC/). The dataset covers 1 week from October 21 to 27, 2024, at 1-hour intervals. It includes key variables such as datetime, road ID, location, and traffic volume. This dataset was utilized to validate the estimated traffic volume across the entire road network.

The road network and Points of Interest (POI) for the study area were collected from OpenStreetMap (OSM).

5.3 Data processing

The data were processed based on road segments, with various features extracted to capture traffic dynamics and road characteristics.

Traffic data from the Camera_TDV dataset were used to extract speed and volume information, as detailed in Section 4.1.

Road attributes, including road type, width, were extracted from OSM for each road segment. These features provide essential information on road infrastructure.

Temporal features were derived from the timestamps of traffic data. Each timestamp was assigned a Week ID and a Min ID. The Week ID was obtained by extracting the day of the week from the timestamp and encoding it as a binary vector. The Min ID represents 5-minute intervals within each day, with a total of 288 intervals per day. The first interval (00:00) is assigned Min

ID = 1, the second interval (00:05) Min ID = 2, and so on.

Points of Interest (POIs) were obtained from OpenStreetMap (OSM) and assigned to road segments within a 50-meter radius. This distance is considered appropriate for urban environments, as it allows drivers or motorized two-wheeler riders to take necessary actions, such as reducing speed or searching for parking, before reaching their destination.

Finally, the input variables used to estimate traffic volume in this study include traffic speed from the Camera_TVD dataset, road characteristics such as road width and type, along with temporal features. The target variable is traffic volume from the same dataset. By adding road type information, we aimed to enable the model to capture the structural characteristics of the road, thus enhancing prediction accuracy. The temporal variables were designed to reflect traffic flow patterns over time. These features were aggregated for each road segment and compiled into the final dataset. Table 1 presents the input features, and the target variable used in the model.

Table 1. Feature used for traffic volume estimation

Variable		Description
Target	TotalVol_Camera	Total number of cars that passed through the road segment in 5 min
Feature	AVG Speed_Camera	Average speed of vehicles on road segment
	MinID	Minute of the day when the data was recorded, ranging from 0 to 288
	WeekID	Day of the week when the data was recorded, ranging from 1 (Monday) to 7 (Sunday)
	Road_width	Width of the road segment
	Road_Type	Classification of the road segment 1: primary road, 2: secondary road, 3: residential road, 4: tertiary road

5.4 Baseline method

Two baseline traffic volume prediction models were adopted to verify the prediction performance:

Long Short-Term Memory (LSTM): An RNN variant designed to model sequential data with long-range dependencies. It captures temporal features in traffic volume but lacks spatial awareness, making it suitable for modeling traffic flow patterns over time but not for considering spatial relationships between road segments.

Spatial-Temporal Graph Convolutional Network (STGCN): This model effectively captures both spatial and temporal dependencies by incorporating graph-based methods. The STGCN, adopted from Yu et al. (2018), enables traffic volume prediction by modeling the spatial structure of the road network and temporal patterns in the data. This model is capable of handling the dynamic nature of traffic data, providing a more comprehensive approach to prediction

5.5 Evaluated Metric

The proposed model was evaluated for accuracy and generalization in two stages. In the first validation, the model was tested on the Camera_TDV dataset on testing subsets. Traffic volumes for road segments in both subsets were predicted, and the predictions were compared with actual values to assess accuracy. In this stage, the performance of model was assessed using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), as shown in Equation (6) and Equation (7), respectively. Smaller values for these metrics indicate better model performance.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \widehat{y_n})^2} \tag{6}$$

$$\text{MAE} = \frac{1}{N}\sum_{n=1}^{N}|y_n - \widehat{y_n}| \tag{7}$$

where N denotes the total number of time intervals, and y_n and $\widehat{y_n}$ correspond to the observed and predicted traffic volumes at the n detector, respectively.

For the second validation, the trained model estimated traffic volumes across the urban road network. Predictions were compared with ground truth data from the Sensor_TDV dataset for overlapping road segments to evaluate the generalization of model to unseen road. Correlation analysis was conducted to compare the predicted and actual traffic volumes.

5.6 Implement detail

Experiments were conducted using an PyTorch framework. For the graph convolution process, the order K was set to 3 to balance computational efficiency and predictive performance. Other standard parameters for the proposed model were optimized using the Ray Tune library's hyperparameter search algorithm. The final settings included 64 convolution kernels for both graph convolution and temporal convolution, a batch size of 32, and a learning rate of 0.001. The model was trained using the Adam optimizer. Camera_TDV dataset was divided into training, validation, and test sets with scale 6:2:2.The number timestep each of dataset is show in table

Although we employed the ASTGCN model proposed by Guo et al. (2019), we extensively modified its structure to better suit this study's objectives. The modifications included:

- Refining the Adjacency Matrix: The graph adjacency matrix was refined using a gravity model, incorporating both the interaction between road segment distances and the built environment, such as Points of Interest (POIs). This approach enhances the representation of spatial relationships by considering both physical proximity and built environment factors.
- Integrating Multiple Data Types: The model was enhanced by integrating multiple data types, including road type, speed features, and time-based information. This approach allows the model to better capture diverse time patterns and improve its ability to model the complex spatiotemporal dynamics of traffic flow. The inclusion of these features provides a richer, more accurate representation of the factors influencing traffic volume.

6. RESULTS

6.1 Model performance comparison

Table 2 presents a comparison of the performance of various models for traffic volume prediction. The LSTM model shows the highest error with an RMSE of 19.92. In contrast, the STGCN model improves accuracy, reducing the RMSE to 14.53 by leveraging spatial relationships. The ASTGCN model further reduces errors by incorporating an attention mechanism, achieving the lowest RMSE of 11.49. Among the models, ASTGCN_Grav, which establishes connections between roads based on the gravity model, performs better than ASTGCN_Dis, which only considers distance between road segments. With an RMSE of 11.33, ASTGCN_Grav highlights the significance of using a gravity-based adjacency matrix.

Table 2. Performance comparison of different models

	LSTM	STGCN	ASTGCN_Dis	ASTGCN_Grav
RMSE	19.92	14.53	11.49	11.33
MAE	14.76	10.75	7.77	7.18

Note: ASTGCN_Dis with adjacency matrix based on driving distance between road segments. ASTGCN_Grav with adjacency matrix base on gravity model

6.2 Best traffic pattern

The relationship between traffic patterns, speed, and volume is investigated by varying the values of n and m to extract traffic speed information for establishing input features. Based on research by Ma et al. (2020), the optimal time window for examining the impact of traffic volume and speed is between 5 and 30 minutes. Therefore, the maximum values for n and m are set to 5. In the first approach, we use the historical speed patterns ad current time step to predict traffic volume ($m = 0$). In the second approach, we extract both past, current time and future speed patterns with equal window sizes. As a result, we have 12 models, each based on different combinations of n and m, which represent different speed patterns extracted. Table 3 presents the model performance with different time patterns extracted for the speed feature.

Table 3. ASTGCN model performance with different time pattern extraction

Model	n,m	MAE	RMSE
Model 1	(n=1, m=1)	7.021	11.16
Model 2	(n=2, m=2)	7.077	11.374
Model 3	(n=3, m=3)	7.479	11.774
Model 4	(n=4, m=4)	7.184	11.432
Model 5	(n=5, m=5)	7.424	11.863
Model 7	(n=0, m=0)	7.861	12.214
Model 8	(n=1, m=0)	7.821	12.114
Model 9	(n=2, m=0)	7.691	11.744
Model 10	(n=3, m=0)	7.511	11.564
Model 11	(n=4, m=0)	7.621	11.554
Model 12	(n=5, m=0)	7.501	11.554

The effect of speed pattern on traffic volume is varied. The model increases the value in n and m it so slightly increases MAE and RMSE. For instance, model 1 with n=1 and m=1 have best performance (RMSE=11.16 and MAE=7.021) while the model with n=5 and m=5 have lower performance. This suggests that while expanding the time window for both past and future traffic patterns captures more data, it does not necessarily lead to improved accuracy and may even introduce unnecessary complexity that negatively impacts performance.

The model that uses historical data with m=0 shows improved performance when compared to models using only current speed. Model 12 (n=5, m=0) performs better than Model 8 (n=1, m=0), indicating that historical speed data is valuable for estimating traffic volume. This demonstrates that relying solely on current speed for volume estimation leads to higher errors. Additionally, models that only consider historical speed data (m=0) perform worse than those that incorporate past, current, and future speed data. Based on these results, we choose n=1 and m=1 as the optimal pattern.

6.3 Impact of Variables on Traffic Volume

To investigate the impact of each feature type on traffic volume estimation, we input different feature combinations into the ASTGCN model. First, we used only the speed feature (ASTGCN I), then we included both speed and time-based features (ASTGCN II), and finally, we incorporated speed, time-based features, and road type (ASTGCN III).

Table 4: ASTGCN model with different input features

Model	ASTGCN I	ASTGCN II	ASTGCN III
RMSE	11.33	11.21	11.16
MAE	7.18	7.15	7.02

From this experiment, we observed that model performance varied based on the input features. Table 4 presents the performance of the ASTGCN model with different input features for estimating traffic speed. The model using only the speed feature (ASTGCN I) showed the lowest performance across all metrics with RMSE= 11.33. However, adding time-based features (ASTGCN II) allowed the model to better capture traffic volume patterns over time, leading to improvement, with RMSE decreasing from 11.33 to 11.21. The best performance was achieved by the ASTGCN model that combined speed, time-based features, and road characteristics such as road type and width (ASTGCN III) with RMSE= 11.16 and MAE=7.02. This improvement is attributed to the inclusion of road characteristics, which enabled the model to account for variations in traffic volume based on road type.

6.4 Traffic volume estimation for entire network

Through experimentation, we obtained a optimize parameter model with a speed-time pattern set to n=1, m=1. This model was applied to estimate traffic volume for entire city base on using Google_speed data along with time base feature and road characteristic. To validate the estimated results, we compared the predictions with the Sensor_TDV data, which provides total volume collected at hourly intervals. Senven road segments selected for validation were those present in both Sensor_TDV data and road segment estimate from model but do not include in training dataset. The result of estimate traffic volume base on Google_Speed and location of sensor shows in the Figure 5.

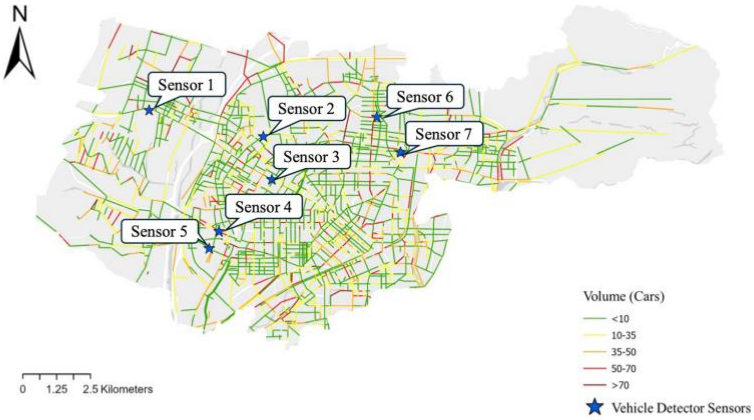

Figure 5. Traffic volume estimate and location of sensor road to validation

To compare the estimated and actual data from the sensor, we first aligned the time intervals in the estimates. We aggregated 12 time points (each representing a 5-minute interval) in traffic volume estimate by model to obtain the traffic volume for one hour. The comparison was then made using correlation analysis. The results are shown in Figure 6 and Figure 7. The correlation between the two datasets was 0.85, which demonstrates that the model can accurately capture the traffic volume on road segments that were not part of the training dataset. From this result, the model could be estimate the traffic volume in unseen road segment .

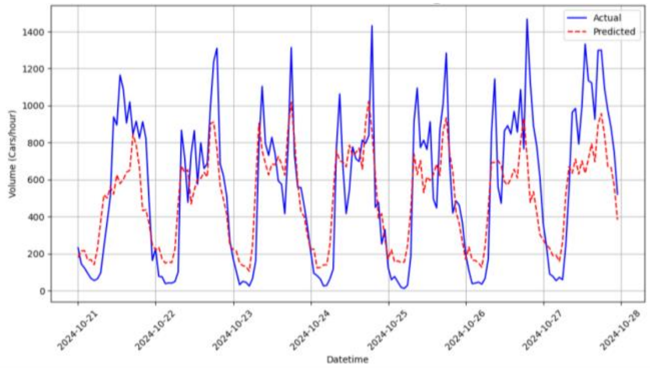

Figure 6. Prediction results and actual volume for Section 2, Wuquan W Rd, at vehicle detector sensor 4

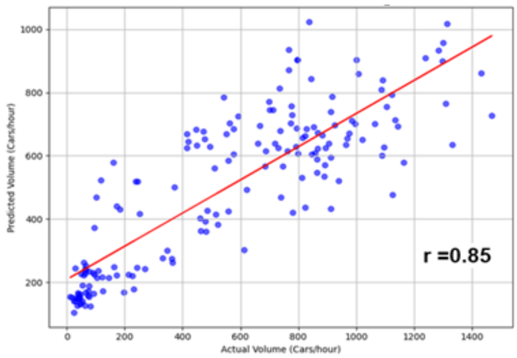

Figure 7. Comparing observed and predicted traffic volume

7. CONCLUSION

This research aims to estimate traffic volume for each road segment in Taichung City based on limited ground truth data. To achieve this, we employed the ASTGCN model to capture the spatial-temporal relationships between road segments. By using gravity models to construct the graph and combining multiple features such as dynamic traffic speed, time-based features, and road characteristics, the model effectively estimates traffic volume. Through experimentation, the relationship between speed and traffic volume was further explored. The results indicate that relying solely on historical speed data is insufficient to accurately estimate traffic volume. The optimal model pattern, which includes data from the previous 5 minutes, the current time step, and the next 5 minutes, significantly enhances the accuracy of model. Regarding feature importance, time-based features play a crucial role in improving model performance. Additionally, road characteristics, such as road type, influence traffic volume predictions.

REFERENCES

Abadi, A., Rajabioun, T., Ioannou, P. A. (2014) Traffic Flow Prediction for Road Transportation Networks With Limited Traffic Data. *IEEE Transactions on Intelligent Transportation Systems*1-10.

Ganji, A., Zhang, M., Hatzopoulou, M. (2022) Traffic volume prediction using aerial imagery and sparse data from road counts. *Transportation Research Part C: Emerging Technologies*, *141*.

Gao, Y., Zhou, C., Rong, J., Wang, Y., Liu, S. (2022) Short-Term Traffic Speed Forecasting Using a Deep Learning Method Based on Multitemporal Traffic Flow Volume. *IEEE Access*, *10*, 82384-82395.

Guo, S., Lin, Y., Feng, N., Song, C., Wan, H. (2019). Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. Proceedings of the AAAI conference on artificial intelligence, Kashyap, A. A., Raviraj, S., Devarakonda, A., Nayak K, S. R., KV, S., Bhat, S. J. (2022) Traffic flow prediction models–A review of deep learning techniques. *Cogent Engineering*, *9*, 2010510.

Kwon, D., Lee, C., Kang, H., Kim, I. (2023) Large-Scale Network Imputation and Prediction of Traffic Volume Based on Multi-Source Data Collection System. *Transportation Research Record: Journal of the Transportation Research Board*, *2677*, 30-42.

Ma, D., Song, X., Li, P. (2020) Daily traffic flow forecasting through a contextual convolutional recurrent neural network modeling inter-and intra-day traffic patterns. *IEEE Transactions on Intelligent Transportation Systems*, *22*, 2627-2636.

Yu, B., Yin, H., Zhu, Z. (2017) Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*.