



Research on Congestion Prediction for High-Density Crowd Scenes

Yanan liu¹, Fang Wu¹, Di Zhou^{2*}, Fanjie Zong¹, Yu wang¹, Lei Zhou³, Jingyuan Chen⁴ and Yong Ding⁵

¹School of Information Technology and Artificial Intelligence, Zhejiang University of Finance & Economics, Hangzhou, 310018, China;

²Zhejiang Uniview Technologies Co.,Ltd, Hangzhou, 310051, China;

³Huaxin College of Hebei Geo University, Shijiazhang, 050700, China;

⁴Hangzhou Tengrunda Technology Co., Ltd., Hangzhou, 310000, China;

⁵College of Integrated Circuits, Zhejiang University, Hangzhou, 310058, China;

*Correspondence: Di Zhou(zhoudi@uniview.com);

<https://doi.org/10.63138/irp010205>

Abstract: With the rapid development of the tourism industry, tourist attractions often face tremendous pressure from crowds during peak times such as holidays, leading to safety hazards and a decline in visitor experience. Effective pedestrian flow prediction can anticipate changes in the flow of people in public places, allowing for preemptive measures to be taken and effectively preventing congestion issues. This study focuses on the prediction of pedestrian flow in public places, aiming to achieve real-time monitoring and dynamic management of pedestrian flow through advanced information technology, thereby enhancing the safety of public places. Traditional pedestrian flow models are not accurate enough and lack specificity. In this study, a multi-feature integrated pedestrian flow prediction model named the GAT-LSTM model is proposed based on the Graph Attention Network (GAT) and Long Short-Term Memory (LSTM) .

Keywords: congestion prediction; deep learning; attention mechanism

1. Introduction

Under the impetus of globalization and economic integration, the tourism industry has experienced rapid development. With this comes the challenge of surging visitor numbers during peak times such as holidays, which not only negatively impacts the visitor experience but also poses significant safety risks. For instance, the 2022 Halloween crowd crush in Itaewon, South Korea, resulted in hundreds of casualties due to congestion on a narrow street. This incident highlighted the dangers of high-density crowds in confined areas and exposed the deficiencies in visitor flow warning and evacuation management in dense venues. Against this backdrop, accurately predicting changes in visitor flow in public places is crucial for preventing congestion, ensuring visitor safety, and enhancing the tourism experience. Effective visitor flow prediction can help managers deploy resources in advance and devise strategies to be well-prepared before peak visitor flows arrive.

Traditional visitor flow prediction models often lack comprehensive consideration of complex environmental factors, leading to low prediction accuracy and inability to meet practical application needs. To address this issue, this study proposes an innovative multi-feature integrated visitor flow prediction model—the GAT-LSTM model. Visitor flow in public places often exhibits periodicity in time, such as daily, weekly, and monthly cycles. Meanwhile, the internal road networks are intricately connected, with different roads having spatial correlations, and the propagation of congestion states is greatly related to spatial relationships. This study integrates external features affecting visitor flow changes and combines the advantages of Long Short-Term Memory networks (LSTM) and Graph Attention Networks (GAT), aiming to achieve accurate prediction of visitor flow in public places by integrating time series analysis and spatial network structure analysis.

The LSTM model part focuses on capturing the temporal dynamics of visitor flow data, while the GAT model part effectively identifies and utilizes key nodes and connections in the spatial network. Through this multi-feature fusion approach, the GAT-LSTM model can more comprehensively understand and predict trends in visitor flow. This study validates the advantages of the GAT-LSTM model in prediction effectiveness through experimental comparisons with other models. The results of this research not only provide new technical means for visitor flow management in public places but also offer new insights for the fields of intelligent transportation systems and urban safety management.

2. Related Work

In the field of time series analysis, some classic algorithms such as Historical Average (HA), Autoregressive Integrated Moving Average (ARIMA) [1], and Vector Autoregression (VAR) [2] have been widely applied to traffic flow forecasting problems. However, these traditional models are generally only suitable for small-scale stationary time series and are not adequate for dealing with complex, large-scale, and dynamic time series data. Moreover, since these methods only consider the impact of temporal factors, they tend to overlook the spatial information in time series data, leading to less accurate forecasting results. Subsequently, many traditional machine learning methods, such as Support Vector Regression (SVR) [3] and Random Forest Regression (RFR) [4], have been proposed for traffic flow prediction. These methods can effectively handle high-dimensional data and capture complex nonlinear relationships. In recent years, researchers have further modeled some external factors to enhance the predictive accuracy of models. However, the overall nonlinear capability of these models is limited.

With the rapid development of the Internet of Things (IoT) and artificial intelligence technology, conducting precise analysis and prediction of pedestrian flow data has become an essential part of emergency evacuation in public places. As IoT devices become more widespread, multi-dimensional spatiotemporal complex data is growing explosively, and deep learning models can better handle such complex data. Therefore, an increasing number of researchers are applying deep learning models to the field of prediction. RNN[5] consists of input layers, hidden layers, and output layers. RNN can process any input time series with its own memory units, making it suitable for capturing long-term dependencies in dynamically changing sequences[6], but it is prone to issues like vanishing or exploding gradients. To address this issue, Hochreiter et al.[7] proposed the Long Short-Term Memory network, which can not only learn and model long-term correlations in time series but also automatically determine the optimal prediction lag. Although LSTM can alleviate the issues of vanishing and exploding gradients, it has a relatively large number of parameters and is computationally complex. Both LSTM model and Gated Recurrent Units (GRU) use gating mechanisms to control the memory cell's long-term memory of information, but GRU has a simpler structure. Fu et al.[8] compared prediction results using LSTM model and GRU model, and the results showed that under the same experimental configuration, GRU model outperformed LSTM model, with faster training speed and easier convergence. Liang et al.[9] proposed a global relationship module to capture global spatial dependencies and a meta-learner to study the impact of external factors on traffic data. In addition, many studies[10-17] have combined convolutional neural networks with recurrent neural networks to jointly model the spatiotemporal correlations of traffic data. Existing models recognize the importance of temporal and spatial correlations but fail to effectively model this spatiotemporal dynamic correlation, which often leads to unstable prediction results. This study combines the LSTM model and the GAT model, leveraging LSTM's ability to handle time series data and GAT's ability to handle graph-structured data, to address issues that require consideration of both temporal dependencies and spatial (or graph) structures. By integrating LSTM

model and GAT model, the model can not only utilize historical information from time series data but also leverage structural information from graphs, which helps improve the accuracy of multivariate time series forecasting.

3. Congestion Prediction Model Design

Pedestrian flow data exhibits complex spatiotemporal characteristics. In the temporal dimension, the entire road network shows patterns such as daily periodicity, weekly periodicity, and holiday periodicity, and there is also correlation between adjacent time nodes. In the spatial dimension, road sections directly adjacent to a node exhibit direct spatial correlation, while sections nearby but not directly adjacent exhibit indirect spatial correlation, which gradually weakens with increasing distance from the node. Although existing forecasting models have recognized the importance of temporal and spatial correlations, they often fail to effectively capture this dynamic spatiotemporal correlation. Therefore, this paper proposes a multi-feature integrated GAT-LSTM model. The LSTM model is chosen for its ability to capture long-term dependencies in time series to extract temporal correlations; the GAT model is selected for its ability to capture complex spatial relationships between nodes through an attention mechanism to extract spatial correlations. By combining these, we aim to more accurately model the dynamic spatiotemporal characteristics of pedestrian flow and improve the accuracy of predictions.

3.1. Study Area

This study predicts the pedestrian flow in an open scenic area of Huzhou City. The Huzhou Dragon Dream scenic area, located in Huzhou City, Zhejiang Province, is a large-scale comprehensive tourist resort covering an area of about 12,000 mu (approximately 8 million square meters). Among them, Taihu Ancient Town is a tourist area themed on ancient town features. With a total construction area of about 660,000 square meters, Taihu Ancient Town is an open scenic area with various public service functions such as historical relics and leisure tourism, attracting a large number of tourists during peak travel seasons, making it highly susceptible to congestion and posing significant safety issues. Therefore, there is a considerable need for pedestrian flow prediction and management in this area.

As shown in Figure 1, the road network in this area is quite complex, including many intersections. The scenic area has a large number of snack shops, and the roads along the lakeshore are suitable for sightseeing, attracting a large number of tourists. The high-density crowds in these areas are likely to lead to stampede incidents. Therefore, cameras and detectors are densely deployed within the scenic area to capture the dynamic distribution of pedestrians.



Figure 1. Map of the scenic area interior.

3.2. Experimental Data

A detector was set up every 10 meters within the road network. To facilitate calculations, all pedestrian flow data recorded by the detectors on each road were summarized to represent the flow data for that road. From January 1 to October 31, 2023, we collected all data from all detectors at one-minute intervals, as detailed in Table 1. In this experiment, the dataset was divided into training, validation, and test sets with a ratio of 7:2:1. The model was trained in an end-to-end manner using the Adam optimizer, with an initial learning rate of 0.001. The training process was conducted over 50 epochs, with a dropout rate of 0.2 and a batch size of 64. The model was implemented using Python 3.11 and PyTorch 2.2.2, and it was executed on a laptop equipped with an Intel Core i7-13900HK CPU and an NVIDIA GeForce RTX 4060 GPU.

Table 1. Pedestrian data.

| Time | Roads | | | |
|------------------|-------|-------|-------|--------|
| | Road1 | Road2 | | Road48 |
| 2023-01-01 00:00 | 0 | 0 | | 0 |
| | | | | |
| 2023-01-01 11:00 | 267 | 432 | | 436 |
| | | | | |
| 2023-01-01 18:00 | 298 | 311 | | 582 |
| | | | | |
| 2023-01-01 21:00 | 489 | 517 | | 677 |
| | | | | |

3.3. Model Design and Implementation

Pedestrian flow prediction is the process of forecasting future pedestrian flow based on already generated pedestrian flow data. By using p spatiotemporal sequences of pedestrian flow from t time steps ago, the pedestrian flow data for t time steps ahead is predicted. As shown in Equation :

$$(x_{t+1}, \dots, x_{t+p}) = f_{\theta} \cdot (x_{t-p+1}, \dots, x_t) \quad (1)$$

Where θ represents the model's learnable parameters.

(1) Feature Fusion

Modeling Holiday Information. Holidays often lead to significant changes in pedestrian flow, which may manifest as either an increase or decrease in pedestrian flow. For instance, during holidays, the number of visitors in tourist attractions may rise noticeably, while on working days it may drop significantly. Therefore, modeling these special dates and analyzing their impact on daily pedestrian flow patterns is crucial. This paper refers to the Prophet model [18] for modeling holiday information. By using the 2023 calendar information, detailed data on weekends, holidays, vacations, and more can be obtained. First, a mapping table like Table 2 is established to distinguish each date as a working day or a holiday. Let B_i represent whether the date is a holiday, thus obtaining an indicator function for holidays.

$$Z_t = [1 (t \in B_i)] \quad (2)$$

$$h_t = Z_t \cdot k_i \quad (3)$$

Where $k \in N(0, V^2)$ represents the parameter corresponding to holidays, indicating the impact of holidays on pedestrian flow; B_i denotes the set of all holidays. The holiday representation h_t is obtained by performing a dot product operation between the weight k_i composed of all holiday weights and the indicator vector of time Z_t .

The holiday information $X_{holiday}$ is aligned with the pedestrian flow data through a feedforward layer and dimensionality transformation, thereby obtaining the holiday information feature embedding $Y_{holiday}$:

$$Y_{holiday} = \sigma \cdot (X_{holiday} W_h + b_h) \quad (4)$$

Where W_h , b_h are learnable parameters, and σ is the activation function.

Table 2. Holiday information mapping table.

| Data | Holiday |
|------------|------------------------|
| 2023-01-01 | New Year's Day Holiday |
| 2023-01-02 | New Year's Day Holiday |
| 2023-01-03 | Working Day |
| | |

Modeling weather information. To model weather information, we employ the Gated Recurrent Unit (GRU) to capture the dynamic implicit relationship between weather data $X_{weather}$ and passenger flow X_{flow} , ultimately obtaining the weather features. The passenger flow data and weather data are concatenated on the feature dimension to form a unified input matrix X .

$$X = \text{concat}(X_{flow}, X_{weather}) \quad (5)$$

The input to the GRU is the matrix X , and its objective is to extract dynamic features from the time series. The feature representation $Y_{weather}$ is obtained after X passes through the GRU.

(2) Dynamic Spatial Relationship Extraction

Align and concatenate $Y_{weather}$, $Y_{holiday}$ and X_{flow} in the spatiotemporal dimensions to obtain the new output C . Input C together with the spatial relationship data into the Graph Attention Network(GAT) model to achieve the extraction of dynamic spatial relationships.

$$e_{ij} = a(w \cdot h_i, W \cdot h_j) \quad (6)$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in N(i)} \exp(e_{ik})} \quad (7)$$

$$h'_i = \sigma \left(\sum_{j \in N(i)} \alpha_{ij} W h_j \right) \quad (8)$$

Where h_i and h_j are the features of nodes i and j , respectively. e_{ij} is the attention score between nodes i and j . α_{ij} is the normalized attention weight. $N(i)$ is the set of neighbors of node i . W is the learnable weight matrix. σ is the activation function. The output after processing by the GAT is used as part of the input to the LSTM.

(3) Time Series Prediction

LSTM is particularly adept at handling time series data and can capture long-term dependencies, which is difficult to achieve in traditional RNNs. By controlling the flow of information through gating mechanisms (forget gate, input gate, output gate), it effectively avoids the vanishing and exploding gradient problems. LSTM allows for customization of the complexity of gating mechanisms and cell states, making it adaptable to a variety of sequence prediction problems, and thus it is widely used in time series forecasting.

The LSTM model is used to capture the temporal dependencies of pedestrian flow. The updated node feature from the GAT are fed into LSTM model. The LSTM model update equation are:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (9)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (10)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (11)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (12)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (13)$$

$$h_t = o_t * \tanh(C_t) \quad (14)$$

Where f_t is the forget gate, controlling the retention of the previous memory. i_t is the input gate, controlling the update of the current input. \tilde{C}_t is the candidate memory cell. C_t is the current memory cell. o_t is the output gate, controlling the current output. h_t is the hidden state at time step t . Finally, the hidden state h_t of the LSTM model is passed through a fully connected layer to obtain the predicted pedestrian flow:

$$\hat{y}_t = W_y * h_t + b_y \quad (15)$$

Where $\hat{y}_t \in \mathbb{R}$ is the predicted pedestrian flow at time step t . W_y and b_y are learnable weights and biases. The GAT-LSTM model is a hybrid model that combines Graph Attention Networks (GAT) and Long Short-Term Memory networks (LSTM), designed to handle data with graph structural features and sequential dependencies. By leveraging the strengths of both GAT model and LSTM model, the GAT-LSTM model can

capture spatial relationships in graph data and temporal dependencies in sequential data simultaneously. The GAT layer is responsible for assigning attention weights in graph-structured data, while the LSTM layer models the temporal dependencies in sequential data. Figure 2 is the model architecture diagram of the GAT-LSTM model.

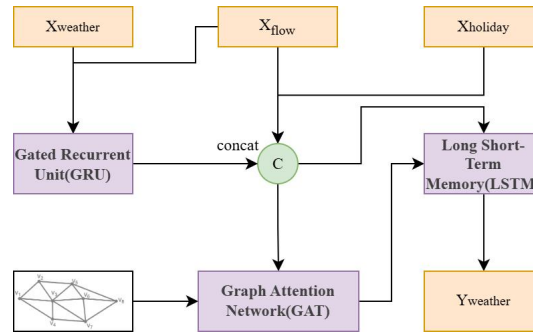


Figure 2. GAT-LSTM Model architecture.

4. Experimental and Results

This paper uses four metrics — Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2) — to evaluate the LSTM, GCN, and GAT-LSTM models. The smaller the metric values, the higher the predictive accuracy of the model.

(1) Mean Squared Error (MSE). MSE is a commonly used loss function for assessing the difference between a model's predicted values and the actual values. It calculates the average of the squares of the differences between the predicted values and the actual values.

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (16)$$

(2) Root Mean Squared Error (RMSE). RMSE is the square root of MSE, measuring the standard deviation of the differences between predicted values and actual values, providing an error measure in the same units as the original data.

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (17)$$

(3) Mean Absolute Error (MAE). MAE measures the average absolute difference between predicted values and actual values, calculating the mean of the absolute values of prediction errors.

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (18)$$

(4) Coefficient of Determination (R^2). R^2 is a statistical measure that assesses the goodness of fit of a regression model, representing the percentage of the dependent variable's variation that can be explained by the independent variables. The value of R^2 ranges from 0 to 1, with values closer to 1 indicating a better fit of the model, meaning that the model can explain more of the variation in the dependent variable.

$$R^2 = 1 - \frac{SSE}{SST} \quad (19)$$

Where SSE is the sum of squared residuals, and SST is the total sum of squares.

By conducting experiments with the same data and configuration using LSTM, GCN, and GAT-LSTM models respectively, the experimental results are obtained as shown in Table 3.

Table 3. Comparison of model prediction results.

| Model | MSE | RMSE | MAE | R^2 |
|-------|-----------|----------|---------|--------|
| LSTM | 6167.8978 | 78.53596 | 62.7659 | 0.8356 |

| | | | | |
|----------|------------|----------|---------|--------|
| GCN | 11643.9042 | 107.9069 | 80.7350 | 0.8077 |
| GAT-LSTM | 3314.9153 | 57.5753 | 24.4944 | 0.9476 |

5. Conclusion

The time-series data of tourist attraction pedestrian flow is often influenced by external factors such as holidays, which exhibit deep implicit relationships with pedestrian flow. By modeling the correlations of these external factors, feature enhancement can be achieved, thereby improving prediction performance. By analyzing historical data and related features, precise pedestrian flow prediction is achieved using deep learning, thereby enabling congestion prediction. This paper proposes a feature fusion-based pedestrian flow prediction model, incorporating holiday information as a key feature to enhance prediction accuracy and robustness through feature fusion modeling. The model combines Graph Attention Network (GAT) and Long Short-Term Memory (LSTM), leveraging GAT model to capture spatial dependencies in the traffic network and LSTM model to capture temporal correlations. The GAT-LSTM model can adapt to new graph structures, such as predicting unseen graphs in inductive learning tasks, thereby enhancing its capability to handle graph-structured data. Furthermore, the model does not rely on complete graph information for learning, demonstrating high flexibility and effectiveness in processing dynamically changing or partially observable graph data. Experimental results show that the GAT-LSTM model exhibits significant advantages in pedestrian flow prediction, providing high-precision predictions and adapting to various traffic networks and conditions.

Funding: This research was supported by the Public Welfare Project of Zhejiang Provincial Science and Technology Department No.LGF22F020034.

References

- [1] Li, M.; Zhu, Z. Spatial-temporal fusion graph neural networks for traffic flow forecasting, In Proceedings of the AAAI conference on artificial intelligence, 2-9 February 2021.
- [2] Zivot, E.; Wang, J. Vector autoregressive models for multivariate time series. In *Modeling Financial Time Series with S-PLUS®*; Zivot, E., Wang, J., Eds.; Springer: New York, USA, 2006; pp. 385-429.
- [3] Chen, R.; Liang, C.Y.; Hong, W.C. et al. Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm. *Appl. Soft Comput.* **2015**, *26*, 435-443.
- [4] Johansson, U.; Boström, H.; Löfström, T. et al. Regression conformal prediction with random forests. *Mach. Learn.* **2014**, *97*, 155-176.
- [5] Jain, A.; Zamir, A.R.; Savarese, S. et al. Structural-rnn: Deep learning on spatio-temporal graphs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27-30 June 2016.
- [6] Guo, K.; Hu, Y.; Qian, Z. et al. Optimized graph convolution recurrent neural network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*(2), 1138-1149.
- [7] Hochreiter, S. Long Short-term Memory. *Neural Comput.* **1997**.
- [8] Fu, R.; Zhang, Z.; Li, L. Using LSTM and GRU neural network methods for traffic flow prediction. In Proceedings of the 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Beijing, China, 10-12 June 2016.
- [9] Liang, Y.; Ouyang, K.; Sun, J. et al. Fine-grained urban flow prediction. In Proceedings of the Web Conference 2021, Ljubljana, Slovenia, 19-23 April 2021.
- [10] Zeng, H.; Peng, Z.; Huang, X.H. et al. Deep spatio-temporal neural network based on interactive attention for traffic flow prediction. *Appl. Intell.* **2022**, 1-12.
- [11] Tian, Z. Approach for short-term traffic flow prediction based on empirical mode decomposition and combination model fusion. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*(9), 5566-5576.
- [12] Lv, M.; Hong, Z.; Chen, L. et al. Temporal multi-graph convolutional network for traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*(6), 3337-3348.
- [13] Yu, L.; Du, B.; Hu, X.; et al. Deep spatio-temporal graph convolutional network for traffic accident prediction. *Neurocomputing* **2021**, *423*, 135-147.

- [14] Zheng, H.; Lin, F.; Feng, X. et al. A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.***2020**, *22*(11), 6910-6920.
- [15] Wang, L.; Guo, D.; Wu, H. et al. TC-GCN: Triple cross-attention and graph convolutional network for traffic forecasting. *INFORM FUSION***2024**, *105*, 102229.
- [16] Yang, H.; Li, Z.; Qi, Y. Predicting traffic propagation flow in urban road network with multi-graph convolutional network. *Complex Intell. Syst.***2024**, *10*(1), 23-35.
- [17] Peng, D.; Zhang, Y. MA-GCN: a memory augmented graph convolutional network for traffic prediction. *Eng. Appl. Artif. Intell.***2023**, *121*, 106046.
- [18] Zaraket, K.; Harb, H.; Bennis, I. et al. Hyper-Flophet: A neural Prophet-based model for traffic flow forecasting in transportation systems. *SIMUL MODEL PRACT TH***2024**, *134*, 102954.