BIG DATA MINING AND ANALYTICS
ISSN 2096-0654 01/01 pp1–18
DOI: 10.26599/BDMA.2025.9020069
Volume x, Number x, xxxx xxxx

Multiresolution Taxi Demand Prediction: A Big Data Statistical and Zero-Inflated Spatiotemporal GNN Approach

Yifei Shen†, Wenlong Shi†, Jiaxing Shen*, Hengzhi Wang, Hanqing Wu, and Jiannong Cao

Abstract: Urban taxi demand prediction faces a critical resolution paradox: high-resolution forecasts enable operational agility but suffer from extreme sparsity-induced volatility, while low-resolution predictions sacrifice responsiveness for stability. We present a Scalable SpatioTemporal Zero-Inflated Poisson Graph Neural Network (SSTZIP-GNN), that resolves this paradox through three innovations: (1) Zero-Inflated Poisson (ZIP) integration that explicitly models structural zeros in sparse demand distributions, distinguishing genuine low-demand periods from data artifacts; (2) Adaptive spatiotemporal learning that dynamically adjusts kernel dilation factors and graph diffusion rates across temporal resolutions using Diffusion Graph Convolutional Networks (DGCNs) and Temporal Convolutional Networks (TCNs); (3) Multimodal feature fusion incorporating real-time crowd-sourced mobility data, socioeconomic indicators, and Global Position System (GPS) trajectories for enhanced robustness under variable urban conditions. Extensive evaluation on 130 million real-world mobility records demonstrates superior performance, achieving 34.8% Mean Absolute Error (MAE) reduction over state-of-the-art baselines. The model reduces computational costs by 46.3% compared to ensemble approaches while maintaining high accuracy across resolutions, delivering 33.4%–53.3% Root Mean Square Error (RMSE) reduction across different prediction resolution scenarios. This unified framework enables cities to implement demand-responsive fleet management, dynamic pricing, and sustainable mobility planning across diverse urban landscapes.

Key words: statistical big data analytics; urban transportation; taxi demand prediction; multi-resolution prediction; data sparsity; Zero-Inflated Poisson (ZIP) distribution

1 Introduction

The proliferation of ride-hailing platforms has fundamentally transformed urban mobility, creating unprecedented operational complexity that demands sophisticated prediction capabilities[1]. Accurate taxi demand forecasting serves as the cornerstone of intelligent transportation systems, enabling dynamic pricing, optimal fleet dispatching, and sustainable

• Yifei Shen and Jiaxing Shen are with School of Data Science, Lingnan University, Hong Kong, China. E-mail: yifeishen@ln.hk; jiaxingshen@ln.edu.hk.
• Wenlong Shi and Hengzhi Wang are with College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China. E-mail: wlong122795@gmail.com; hz@szu.edu.cn.
• Hanqing Wu is with Lucky Technology Development Limited, Hong Kong, China. E-mail: hanqing.91.wu@connect.polyu.hk.
• Jiannong Cao is with Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China. E-mail: csjcao@comp.polyu.edu.hk.

† Yifei Shen and Wenlong Shi contribute equally to this work.

∗ To whom correspondence should be addressed.

Manuscript received: 2025-03-02; revised: 2025-05-26; accepted: 2025-06-03

urban planning[2, 3]. However, modern smart cities present a critical challenge: achieving prediction accuracy across multiple temporal resolutions while managing inherent data sparsity.

Contemporary forecasting systems encounter a fundamental resolution paradox. High-resolution predictions (5-min intervals) provide operational agility but suffer from severe data sparsity that compromises statistical reliability[4, 5]. Conversely, low-resolution predictions (60-min intervals) achieve stability through temporal aggregation but sacrifice responsiveness to dynamic demand patterns[6]. This paradox intensifies as cities deploy heterogeneous mobility services requiring simultaneous multi-scale predictions—a capability beyond current single-resolution architectures[7]. Maintaining separate prediction systems for different temporal resolutions imposes prohibitive computational costs on mobility platforms[8].

Recent spatiotemporal deep learning advances leverage graph convolutional networks and temporal attention mechanisms[9−12], with diffusion-based hybrid architectures showing particular promise[13]. However, three critical limitations persist: (1) inadequate integration of real-time crowdsensing data streams[14], (2) insufficient modeling of zero-inflated distributions in sparse demand data[4], and (3) architectural inflexibility preventing dynamic multi-resolution adaptation[15].

These limitations manifest as significant operational deficiencies: reduced responsiveness to real-time demand shifts, inaccurate predictions in low-activity zones, and computational inefficiency from maintaining resolution-specific models. Addressing these challenges requires a unified framework capable of handling sparse, zero-inflated demand distributions across multiple temporal scales.

We present Scalable SpatioTemporal Zero-Inflated Poisson Graph Neural Network (SSTZIP-GNN), a scalable spatiotemporal framework that resolves the resolution paradox through three innovations: (1) Zero-Inflated Poisson (ZIP) distribution modeling for explicit structural zero handling, (2) adaptive mechanisms for dynamic multi-resolution prediction, and (3) multimodal feature fusion integrating real-time crowdsensing, socioeconomic indicators, and GPS trajectories.

Our contributions are threefold:

(1) A novel spatiotemporal architecture combining diffusion graph convolutions with temporal dilated convolutions in a ZIP framework, achieving 34.8% Mean Absolute Error (MAE) improvement over state-of-the-art methods.

(2) An adaptive mechanism enabling unified multi-scale prediction with 46.3% computational cost reduction compared to ensemble approaches.

(3) Comprehensive evaluation on 130 million mobility records demonstrating 33.4%−53.3% Root Mean Square Error (RMSE) reduction across the different prediction time resolution scenarios.

Experimental analysis reveals that high-resolution predictions benefit most from ZIP modeling and real-time crowdsensing data (15.6% F1-score improvement), while low-resolution predictions depend more heavily on historical patterns and socioeconomic factors (12.4% F1-score improvement). This adaptive capability explains the framework’s superior performance across temporal scales.

The remainder of this paper is organized as follows: Section 2 formalizes the multi-resolution prediction problem, Section 3 details the SSTZIP-GNN architecture, Section 4 presents experimental validation, Section 5 reviews related work, Section 6 discusses limitations and future directions, and Section 7 concludes.

2 Preliminary

In this section, we first present key definitions, and then formally formulate the taxi demand prediction problem.

2.1 Definitions

Time resolution. The historical taxi demand data are organized based on different time resolutions, such as 5 min, 15 min, 30 min, and 60 min. These resolutions, denoted as R, allow the model to capture temporal patterns at various granularities, enabling a flexible and robust forecasting process.

Sparsity. Data sparsity reflects the uneven distribution of taxi demand across time and space, which poses challenges to predictive modeling. By incorporating sparsity as a feature, the model accounts for underrepresented regions or times with low activity.

2.2 Problem description

The taxi demand prediction task can be formally defined as predicting the future taxi demand based on historical records and additional contextual information. This task can be expressed as

$$\hat{X}_{r+1:r+N} = \mathcal{Y}(X_{r-N'+1:r}, R, S, \text{CGD}, \text{SED}) \qquad (1)$$

where

- N': Number of historical records used for prediction.
- N: Number of predicted records.
- $X_{r-N'+1:r}$: Taxi demand for different areas over N' historical records.
- R: Time resolution of $X_{r-N'+1:r}$, which include granularities, such as 5 min, 15 min, 30 min, and so on.
- S: Sparsity level of $X_{r-N'+1:r}$, reflecting the proportion of zero demand of taxi demand data across time and space.
- CGD: Crowdsensed geolocation data, providing auxiliary spatial and temporal information.
- SED: Demographic and economic indicators that influence taxi demand.
- $\mathcal{Y}(\cdot)$: Prediction function that maps the input features to the predicted taxi demand.

The objective is to develop a scalable prediction function $\mathcal{Y}(\cdot)$ that ensures the predicted taxi demand $\hat{X}$ across varying resolutions closely aligns with the actual demand X.

3 Methodology

In this section, we introduce the proposed methodologies in detail. We start with introducing the overall workflow of the SSTZIP-GNN model, followed by Diffusion Graph Convolution Networks (DGCNs) and Temporal Convolutional Networks (TCNs). Next, we describe the adaptive mechanism. Finally, we present the ZIP distribution.

3.1 SSTZIP-GNN

The overall framework of SSTZIP-GNN is illustrated in Fig. 1, which primarily consists of five steps.

(1) Input representation

The raw input data for SSTZIP-GNN comprise five components. The most critical input is the observation sequence of taxi demand at historical time steps, denoted as $X_{r-N'+1}, X_{r-N'+2}, \ldots, X_r$. This sequence integrates data with varying temporal resolutions. Additionally, two key external factors are considered: CGD and SED, which help the model capture the spatio-temporal variations in taxi demand dynamics. Finally, to enable multi-resolution prediction, the model also incorporates resolution and sparsity

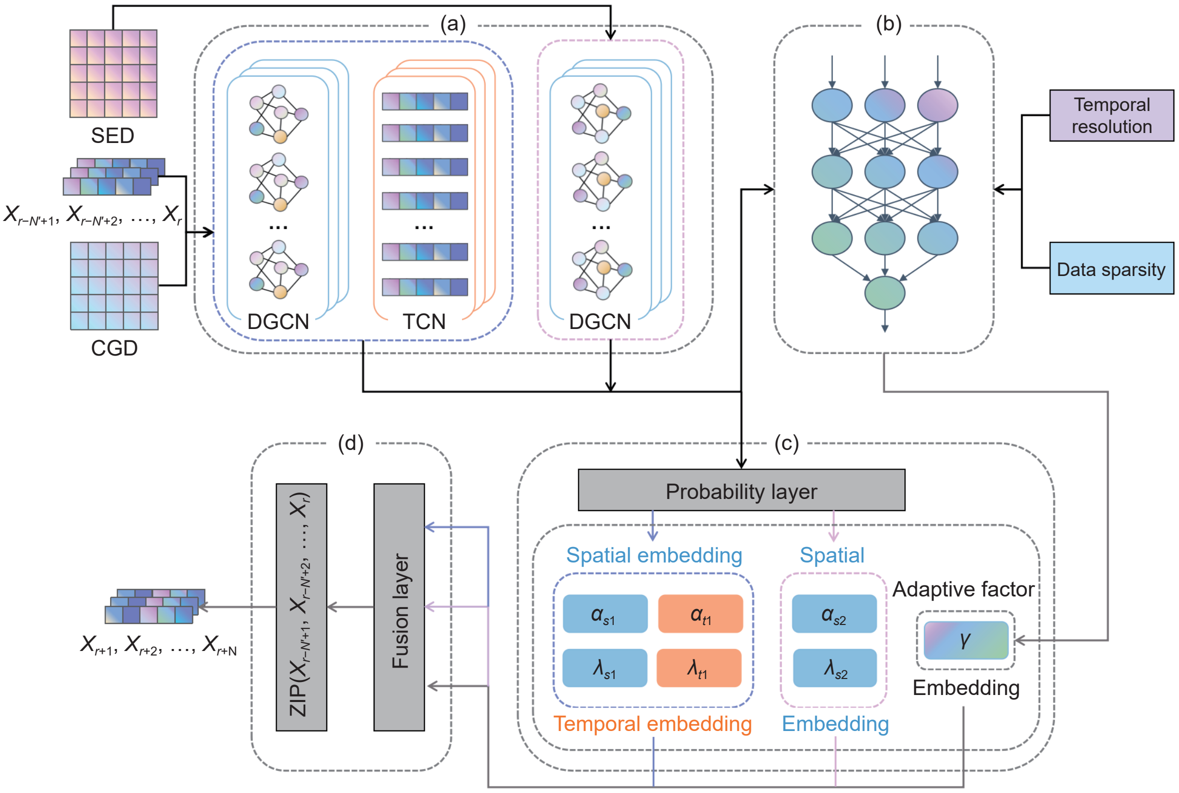


Fig. 1 Framework of SSTZIP-GNN, (a) spatio-temporal neural network, (b) adaptive mechanism, (c) probability estimation layer, and (d) ZIP distribution layer.

information of the data, represented as R and S, respectively, ensuring the model's scalability.

(2) Spatio-temporal feature extraction

To capture the underlying spatio-temporal dependencies in taxi demand, we employ a hybrid learning framework combining DGCN and TCN. This module (Fig. 1a) consists of:

- **Graph-based spatial feature extraction:** The first stacked DGCNs module learns spatial correlations between different locations based on historical observations while capturing real-time crowd mobility characteristics.
- **Temporal dependency modeling:** A stacked TCNs module processes the extracted spatial features to capture temporal dynamics.
- **Enhanced spatio-temporal fusion:** A second stacked DGCNs module refines the learned spatial representations by further incorporating demographic and economic indicators across spatial regions.

(3) Adaptive factor generation

One of the key innovations of SSTZIP-GNN is the adaptive mechanism (Fig. 1b), which dynamically adjusts the probability estimation parameters based on input data characteristics. This module is designed to address the variability in data sparsity and temporal resolution. A deep neural network processes the extracted features and outputs an adaptive factor γ, which modulates the probability distribution in the ZIP.

(4) Independent estimation of ZIP parameters

To parameterize the ZIP distribution, the outputs of DGCN and TCN are fed into a probability estimation layer (Fig. 1c) to obtain spatial and temporal embeddings, denoted as α and λ, respectively. These embeddings provide independent estimates of the ZIP parameters corresponding to their spatial and temporal localities. The independent estimation of the ZIP parameters obtained are as follows:

- **Spatial embedding:** α_{s1} and λ_{s1} (α_{s2} and λ_{s2}), representing independent estimate under the influence of specific dynamic (static) factors at particular spatial positions.
- **Temporal embedding:** α_{t1} and λ_{t1} representing independent estimates under the influence of dynamic temporal variations.

(5) Parameters fusion and prediction

The final step in SSTZIP-GNN involves fusing the independent estimates of the ZIP parameters with the adaptive factors to derive the final ZIP parameters, which are then used to predict taxi demand through the ZIP distribution (Fig. 1d). In this paper, the objective is to predict taxi demand for the next N records, for which the spatial embeddings α_{s1} and α_{s2}, and temporal embeddings α_{t1}, as well as λ_{s1}, λ_{s2} and λ_{t1}, are all N-dimensional vectors. Likewise, the adaptive factor γ is also an N-dimensional vector. To fuse the spatial embeddings, temporal embeddings, and adaptive factors, we apply the Hadamard product, resulting in the parameter set P for the future demand distribution,

$$P = \begin{pmatrix} \alpha \\ \lambda \end{pmatrix} = \begin{pmatrix} \alpha_{s1} \odot \alpha_{s2} \odot \alpha_{t1} \odot \gamma \\ \lambda_{s1} \odot \lambda_{s2} \odot \lambda_{t1} \odot \gamma \end{pmatrix} \tag{2}$$

where α and λ are also N-dimensional vectors, and "$\odot$" denotes the Hadamard product. The final ZIP distribution is shown as follows:

$$f_{\text{ZIP}}(X_{r+1:r+N} | \alpha_{r+1:r+N},\ \lambda_{r+1:r+N}) = f_{\text{ZIP}}(X_{r+1:r+N} | P) \tag{3}$$

Additionally, we use the Negative Log Likelihood (NLL) as our loss function to improve the fit of the distribution to the data. Let y represent the ground truth corresponding to the n-th predicted entry in the matrix, with parameters α_n and λ_n derived from P. The Log Likelihood (LL) of the ZIP distribution is split into two components based on whether y equals 0 or is greater than 0, and can be expressed as

$$\text{LL}_y = \begin{cases} \log \alpha_n + \log(1-\alpha_n)\mathrm{e}^{-\lambda_n}, & \text{if } y = 0; \\ \log(1-\alpha_n) + y \log \lambda_n - \lambda_n - \log y!, & \text{if } y > 0 \end{cases} \tag{4}$$

Accordingly, the final NLL loss function is defined as

$$\text{NLL}_{\text{SSTZIP}} = -\text{LL}_{y=0} - \text{LL}_{y>0} \tag{5}$$

The complete model algorithm pseudocode is outlined in Algorithm 1.

3.2 Diffusion graph convolution network

To effectively capture spatial correlations between different regions, we model these relationships as a diffusion process. This approach facilitates the learning of spatial dependencies, which is essential for accurately predicting regional demand. Due to the Markov property, the diffusion process converges to a stationary distribution after a sufficient number of steps. Each row in this distribution represents the diffusion probability originating from a given node. The stationary distribution is given by[16]

Algorithm 1 SSTZIP-GNN

```
Input: Historical taxi demand records X, real taxi demand $Y$, spatial adjacency matrices CGD and SED, time resolution $R$, sparsity level $S$, and number of layers $L$
Output: Predicted taxi demand $\hat{X}$
// Initialization
$D \leftarrow$ diag (sum (CGD));
$H_0 \leftarrow X, F_0 \leftarrow X$;
$W_{f1}, W_{b1} \leftarrow$ normalize (CGD), normalize (CGD$^{\mathrm{T}}$);
$W_{f2}, W_{b2} \leftarrow$ normalize (SED), normalize (SED$^{\mathrm{T}}$);
Initialize 1D convolution layers $\{W_l\}_{l=1}^{L}$;
while not converged do
    // Spatio-temporal feature extraction
    for $l \leftarrow 0$ to $L-1$ do
        $H_{\mathrm{CGD}}^{l+1} \leftarrow$ DGCN $(H^l, W_{f1}, W_{b1}, \Phi_{f1}^l, \Phi_{b1}^l)$;
        $H_{\mathrm{SED}}^{l+1} \leftarrow$ DGCN $(H^l, W_{f2}, W_{b2}, \Phi_{f2}^l, \Phi_{b2}^l)$;
        $F^{l+1} \leftarrow$ TCN $(F^l, W_l, \beta^l)$;
    end
    $H_{\mathrm{CGD}} \leftarrow H_{\mathrm{CGD}}^L, H_{\mathrm{SED}} \leftarrow H_{\mathrm{SED}}^L, F \leftarrow F^L$;
    // Adaptive factor generation
    $\gamma \leftarrow h\,(H_{\mathrm{CGD}}, F, H_{\mathrm{SED}}, R, S)$;
    // Independent estimation of ZIP parameters
    $\alpha_{s1} \leftarrow \sigma\,(H_{\mathrm{CGD}} W_{s1}^{\alpha} + b_{s1}^{\alpha})$;
    $\lambda_{s1} \leftarrow$ softplus $(H_{\mathrm{CGD}} W_{s1}^{\lambda} + b_{s1}^{\lambda})$;
    $\alpha_{s2} \leftarrow \sigma\,(H_{\mathrm{SED}} W_{s2}^{\alpha} + b_{s2}^{\alpha})$;
    $\lambda_{s2} \leftarrow$ softplus $(H_{\mathrm{SED}} W_{s2}^{\lambda} + b_{s2}^{\lambda})$;
    $\alpha_{t1} \leftarrow \sigma\,(F \times W_{t1}^{\alpha} + b_{t1}^{\alpha})$;
    $\lambda_{t1} \leftarrow$ softplus $(F \times W_{t1}^{\lambda} + b_{t1}^{\lambda})$;
    // Parameters fusion and distribution construction
    $\alpha \leftarrow \alpha_{s1} \odot \alpha_{s2} \odot \alpha_{t1} \odot \gamma$;
    $\lambda \leftarrow \lambda_{s1} \odot \lambda_{s2} \odot \lambda_{t1} \odot \gamma$;
    // Loss computation (NLL)
    NLL $\leftarrow -\log f_{\mathrm{ZIP}}\,(Y|\alpha, \lambda)$;
end
// Final prediction using ZIP expectation
$\hat{X} \leftarrow (1-\alpha) \odot \lambda$;
return $\hat{X}$
```

$$\mathcal{P} = \sum_{k=0}^{+\infty} \eta(1-\eta)^k (D^{-1}B)^k \tag{6}$$

where k denotes the diffusion step, which is typically truncated to a finite value, D is the degree matrix, and B represents the adjacency matrix. The scalar $\eta \in [0, 1]$ is a restart probability that controls the extent of diffusion.

We define the forward diffusion process using the transition matrix $W_f = B/\mathrm{rowsum}(B)$, while the backward diffusion process is characterized by the transition matrix $W_b = B/\mathrm{rowsum}\,(B^{\mathrm{T}})$, where rowsum $(\cdot)$ denotes row-wise summation followed by element-wise division for normalization. Since the adjacency matrix B is symmetric, it follows that $W_f = W_b$. The fundamental operation of a DGCN layer can be expressed as[17]

$$H_{l+1} = \sigma\left(\sum_{p=1}^{K} T_p(W_f) H_l \Phi_f^p + U_p(W_b) H_l \Phi_b^p\right) \tag{7}$$

where H_l denotes the hidden representation at layer l; $T_p(X)$ and $U_p(X)$ are polynomial functions of order p, approximating the convolution operation in DGCN. The learnable parameters Φ_f^p and Φ_b^p regulate the information propagation between nodes in layer l, while $\sigma\,(\cdot)$ represents the activation function, such as ReLU or Linear.

In our model, we employ three stacked DGCN layers to effectively capture spatial dependencies across regions. This hierarchical representation enhances the model's ability to leverage intrinsic spatial correlations within the data, significantly improving the accuracy of regional demand predictions.

3.3 Temporal convolutional network

TCNs offer several advantages over Recurrent Neural Networks (RNNs), as demonstrated by Wu et al.[18]:

- TCNs can accommodate sequences of varying lengths, enhancing adaptability to different temporal scales and resolutions.
- Their simplified architecture facilitates more efficient training compared to RNN-based approaches.

The core idea of TCNs is to leverage shared gated 1D convolutions with a kernel width of k_l in the l-th layer. This structure enables information propagation across k_l neighboring time steps, capturing temporal dependencies effectively. Each TCN layer, denoted as F_l, updates its state based on the preceding layer F_{l-1} according to[19]

$$F_l = f\,(W_l * F_{l-1} + \beta) \tag{8}$$

where W_l represents the convolutional filter, "$*$" denotes the convolution operation, $f\,(\cdot)$ is the activation function, and β is the bias term.

TCNs operate as sequence-to-sequence models, directly forecasting future sequence records. Their receptive field is adjustable by varying the number of layers and kernel sizes, offering flexibility in capturing temporal dependencies at different resolutions. To extract meaningful temporal features, we structure the order data as a time series and process them through multiple TCN layers. This hierarchical feature extraction enhances the model's ability to capture trends and fluctuations over time, improving predictive accuracy. In our model, we employ a stack of three TCN layers to further enhance performance.

3.4 Adaptive mechanism

To improve the model's scalability in handling spatial and temporal sparsity in data with varying temporal resolutions, we propose an adaptive mechanism that dynamically adjusts the parameters of the ZIP distribution. This module leverages lightweight neural networks, particularly Convolutional Neural Networks (CNNs), to generate adaptive adjustment factor. This factor influence the generation of ZIP parameters, enabling the model to effectively respond to different temporal resolutions.

The adaptive mechanism integrates data resolution and sparsity features with the spatio-temporal characteristics extracted by DGCNs and TCNs. It performs adaptive learning to capture the intricate relationships among spatio-temporal embeddings, resolution, and sparsity. Specifically, the mechanism employs CNNs to jointly model these features. This generates an adaptive factor that adjusts the parameters of the ZIP distribution. The CNN thus learns both the spatial-temporal dependencies and the distribution of features under varying levels of sparsity.

Regarding the CNN architecture, it consists of three convolutional layers, each of which plays a key role in learning the relationships between spatio-temporal features and the sparsity characteristics of the data. In the first layer, a 3×3 kernel transforms 16 input channels into 32 output channels. The Leaky ReLU activation function ensures smooth gradient propagation, while BatchNorm accelerates training. A dropout rate of 0.2 effectively prevents overfitting. The second layer uses the same 3×3 kernel to increase the channel depth from 32 to 64. With a dropout rate of 0.3, this layer captures more complex relationships between spatio-temporal features and varying data resolution and sparsity. The final layer increases the channel depth from 64 to 128, utilizing a 1×1 kernel to efficiently fuse features across channels while reducing model complexity. Similar to the previous layers, Leaky ReLU, BatchNorm, and a dropout rate of 0.3 are applied here as well to ensure regularization and activation.

These convolutional layers work together to generate the adaptive factor γ,

$$\gamma = h\left(H_{\text{CGD}}, F, H_{\text{SED}}, R, S\right) \tag{9}$$

where $h(\cdot)$ denotes the adaptive CNN module, H_{CGD} and H_{SHD} denote the hidden spatial features extracted by the DGCNs, and F represents the hidden temporal features obtained from the TCNs.

By incorporating this adaptive mechanism into the SSTZIP-GNN framework, the model's adaptability to data with different temporal resolutions is strengthened, especially in distinguishing zero-demand areas from naturally low-demand regions. As a result, the model becomes more robust in handling multi-resolution data and enhances its scalability in dynamic environments.

3.5 ZIP distribution

The Poisson distribution is widely used for modeling count data, where the Probability Mass Function (PMF) is defined as

$$f_{\text{Poisson}}(x_k;\ \psi) = \Pr\left(X = x_k\right) = \frac{\psi^{x_k} \mathrm{e}^{-\psi}}{x_k!}, \\ x_k = 0,\ 1,\ 2,\ \ldots \tag{10}$$

where ψ is the rate parameter, representing the expected number of occurrences within a fixed interval. The Poisson distribution assumes that the variance equals the mean, making it inadequate for scenarios where data exhibit overdispersion, i.e., where the observed variance exceeds the mean. This limitation becomes particularly evident in sparse datasets, where an excessive number of zeros is present, a phenomenon often referred to as zero inflation.

To address this issue, we employ the ZIP distribution, which introduces an additional parameter α to model the inflation of zeros explicitly. The ZIP distribution is a mixture model that combines a degenerate distribution at zero with a standard Poisson distribution. The corresponding probability mass function can be expressed as

$$f_{\text{ZIP}}(x_k;\theta,\ \psi) = \begin{cases} \theta + (1-\theta) f_{\text{Poisson}}(0;\ \psi), & \text{if } x_k = 0 \\ (1-\theta) f_{\text{Poisson}}(x_k;\ \psi), & \text{if } x_k > 0 \end{cases} \tag{11}$$

where θ represents the probability that an observation is an excess zero, while $1-\theta$ denotes the probability that the count value follows a Poisson distribution with parameter ψ. This formulation allows the ZIP model to flexibly account for both structural zeros (i.e., inherent zeros due to the nature of the data) and sampling zeros (i.e., those generated by a standard Poisson process).

In taxi demand forecasting, the ZIP distribution is particularly suitable for handling spatial and temporal sparsity, where certain regions or time intervals frequently exhibit zero demand. By incorporating zero inflation, the model can better distinguish between areas with genuinely low demand and those where demand is entirely absent due to external factors. This capability enhances the robustness of our predictive framework, leading to more accurate demand estimations in sparse urban environments.

4 Experimental Result

In this section, we first present the dataset and preprocessing. Following that, we introduce the baseline models and outline the evaluation metrics for the task. Finally, we report the experimental results.

4.1 Data description

Taxi Trajectory–HK[‡]: This dataset contains GPS trajectory data collected from Uber drivers in Hong Kong between October 1, 2020, and January 31, 2021. Location coordinates (longitude, latitude), timestamps, anonymized trip identifiers, and binary occupancy status (1 for occupied, 0 for unoccupied) are recorded at one-minute intervals. A summary of the dataset statistics is provided in Table 1.

Taxi Trajectory–MH[§]: This dataset comprises for-hire vehicle trip records in Manhattan from January 2018 to April 2019, collected by the New York City Taxi & Limousine Commission. Each trip includes the pickup and drop-off time, date, and zone location ID. The Manhattan area is divided by ZIP code zones, each with associated demographic and transit metadata.

Table 1 Taxi Trajectory–HK record data.

Attribute	Description	Example
Order IDs	Unique identifier represented as a 32-character hexadecimal string	aaf63e4b38e9b1a 405f20ebe6034d93f
Time	14-digit number format `yyyyMMddHHmmss`	20201018110500
Latitude	2-digit number with 6 decimals, in the degree unit	22.551760
Longitude	3-digit number with 6 decimals, in the degree unit	114.163340
Taxi status	0: Taxi not occupied, 1: Occupied	0, 1

Taxi Zone Map Data[¶]: It utilizes map data aligned with the 2021 town planning framework established by the Hong Kong government. In addition, we incorporate map data from Manhattan, New York, based on the Taxi & Limousine Commission defined taxi zones.

Crowdsensing Geolocation Data (CGD)[‡]: CGD comprises users' geolocation information, including individual latitude and longitude coordinates, as well as residential and workplace locations. This dataset offers a finer granularity compared to traditional taxi Global Position System (GPS) records by capturing a broader range of mobility patterns, such as pedestrian movement and traffic flow. The inclusion of these additional data helps mitigate data sparsity, particularly in regions or time periods with low taxi activity, by providing supplementary contextual information. Moreover, CGD is frequently updated in real time, ensuring its relevance for dynamic urban environments. By integrating CGD with conventional taxi trajectory data, the proposed model can achieve greater robustness, enhance prediction accuracy, and improve the understanding of complex spatiotemporal dependencies in urban traffic systems.

Socioeconomic Data[©]: The socioeconomic dataset encompasses a range of demographic and economic indicators, including population density, salary levels, marital status, household structures, labor force participation, employment rates, income distribution, and housing types. These variables provide valuable insights into the demographic composition and economic conditions of different regions. As depicted in Fig. 2, the taxi demand correlation matrix reveals the relationship between taxi demand and various static socioeconomic factors. Notably, taxi demand exhibits a moderate positive correlation with population density. Suggesting that regions with higher population density, income, and educational attainment tend to show higher taxi demand. These correlations offer critical insights into the influence of socioeconomic conditions on urban mobility patterns, contributing to a more accurate prediction of taxi demand.

4.2 Data preprocessing

To construct the datasets required for our experiments, we focus primarily on processing the GPS trajectory data which involved three main steps:

[‡] The datasets are proprietary and available upon request under confidentiality terms.

[§] https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page

[¶] https://portal.csdi.gov.hk/geoportal/#metadataInfoPanel

[©] https://data.gov.hk/en/

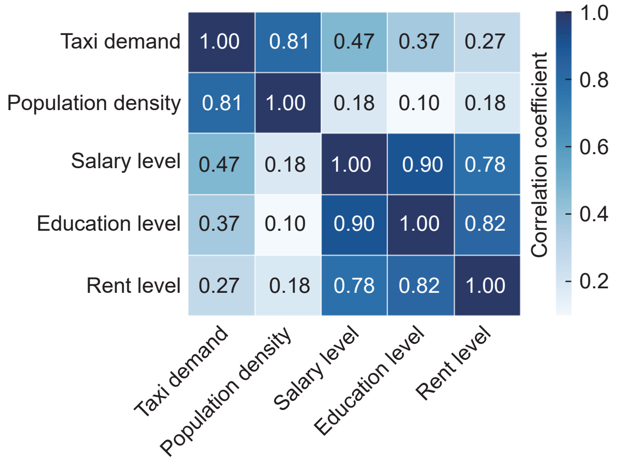


Fig. 2 Correlation matrix between taxi demand and various socioeconomic factors.

(1) Temporal segmentation: The raw GPS trajectory data are segmented based on different temporal resolutions, including 5 min, 15 min, 30 min, 45 min, and 60 min. This step ensures that the data could be analyzed at various resolutions to effectively capture the spatiotemporal dynamics.

(2) Random sampling: After temporal segmentation, the data for each resolution are randomly sampled according to predefined proportions, as shown in Table 2. For instance, in SSTZIP-GNN-I, the data are sampled at 5% for the 5-min resolution, 40% for the 30-min resolution, and 80% for the 60-min resolution. Similar sampling schemes are applied for SSTZIP-GNN-II and SSTZIP-GNN-III, covering time resolutions from 5 to 60 min.

(3) Dataset integration: The data sampled at different time resolutions are integrated into three distinct data structures, each designed to train predictive models with varying scalability capabilities. These models are designed to capture multi-scale temporal dependencies, with their predictive performance varying based on the range of time resolutions included in each data structure.

This preprocessing approach ensures that the constructed datasets effectively represent the spatiotemporal variability, while addressing data sparsity by leveraging a randomized sampling strategy.

Table 2 Proportion of data with different resolutions in the three models.

(%)

Model	5 min	15 min	30 min	45 min	60 min
SSTZIP-GNN-I	5	–	40	–	80
SSTZIP-GNN-II	–	20	40	50	80
SSTZIP-GNN-III	5	20	40	50	80

4.3 Baseline models

In the experiment, we use the following baseline models:

- **Historical Average (HA):** predicts taxi demands at the next time slot in each region by averaging the historical taxi demands at the same time slot.
- **Diffusion Convolution Recurrent Neural Network (DCRNN)[20]:** utilizes diffusion graph convolutional networks and seq2seq to encode spatial information and temporal information, respectively.
- **Spatial-Temporal Graph Convolutional Network (STGCN)[21]:** consists of several ST-Conv blocks, which are built with entirely convolutional layers, to tackle traffic prediction tasks. Specifically, each block is composed of graph convolution and gated temporal convolution, which jointly process graph-structured time series.
- **Spatial-Temporal Zero-Inflated Negative Binomial Graph Neural Network (STZINB-GNN)[4]:** is featured with the uncertainty quantification of the sparse travel demand with diffusion and temporal convolution networks.
- **Spatial-Temporal Guided Multi-graph sandwich-Transformer (STGMT)[22]:** addresses spatial-temporal heterogeneity in traffic demand forecasting using a Sandwich-Transformer architecture. It integrates Multi-head Spatial-Temporal Attention, guided by Node2Vec-based embeddings, to capture spatiotemporal dependencies.
- **SpatioTemporal Zero-Inflated Poisson Graph Neural Network (STZIP-GNN)[23]:** utilizes the zero-inflated poisson distribution to handle the high frequency of zeros in sparse data and integrates CGD and SED mitigate data sparsity.

4.4 Evaluation metrics

To evaluate the prediction accuracy of the expected values, we employ the Mean Absolute Error (MAE), which is defined as

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{12}$$

where y_i and $\hat{y}_i$ represent the ground-truth value and the predicted value, respectively, and N denotes the total number of prediction samples. MAE measures the average magnitude of the errors in a set of predictions, providing a straightforward and interpretable metric for evaluating model performance.

In addition to MAE, we utilize the Root Mean Squared Error (RMSE), defined as

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{13}$$

RMSE penalizes larger errors more heavily than MAE, making it particularly sensitive to outliers in the predictions. It provides a comprehensive evaluation of the model's performance by capturing both the magnitude and variability of prediction errors.

Furthermore, we assess the accuracy of discrete predictions using the F1-score, which evaluates the balance between precision and recall. Although traditionally designed for classification tasks, the F1-score can be adapted to analyze discrete prediction values by treating them as multiple labels and defining precision and recall accordingly. A higher F1-score indicates better performance in terms of discrete value predictions. The F1-score is formulated as

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

where "precision" is the proportion of correctly predicted positive instances among all predicted positives, while "recall" measures the proportion of correctly predicted positives relative to the total actual positives.

4.5 Experimental results

Performance comparison on Taxi Trajectory−HK dataset. The upper half of Table 3 presents model performance on the Taxi Trajectory−HK dataset under three time resolutions (10 min, 30 min, and 50 min). SSTZIP-GNN-III consistently achieves the best results, reporting the lowest MAE (2.421, 1.434, and 3.334) and RMSE (3.514, 2.016, and 4.652), as well as the highest F1-scores (0.830, 0.887, and 0.771). SSTZIP-GNN-II follows closely, maintaining strong performance across all metrics and resolutions.

While SSTZIP-GNN-I remains competitive overall, its effectiveness decreases at lower resolutions. At 30 min, for example, its RMSE (3.995) exceeds those of STGMT (3.581) and STZIP-GNN (3.647); at 50 min, its MAE (3.667) is also higher than that of both models.

Regarding baseline methods, STZIP-GNN performs best in high-resolution settings (10 min), whereas STGMT shows advantages in medium and low

Table 3 Performance comparison of different models on Taxi Trajectory−HK and Taxi Trajectory−MH with different time resolutions. Bold numbers indicate the best results for each metric.

Dataset	Model name	10 min			30 min			50 min		
		MAE	RMSE	F1-score	MAE	RMSE	F1-score	MAE	RMSE	F1-score
Taxi Trajectory−HK	HA	6.476	7.467	0.670	4.223	6.653	0.748	6.658	8.375	0.610
	DCRNN	4.888	5.724	0.712	2.825	4.712	0.760	5.524	7.437	0.642
	STGCN	4.535	5.556	0.720	2.439	4.466	0.800	4.634	6.844	0.670
	STZINB	3.275	5.374	0.744	1.998	3.978	0.811	3.856	6.597	0.740
	STGMT	3.478	5.422	0.721	**1.934**	**3.581**	**0.855**	3.640	6.612	0.736
	STZIP-GNN	**3.256**	**5.275**	**0.746**	1.947	3.647	0.850	**3.625**	**6.427**	**0.740**
	SSTZIP-GNN-I	2.885	4.762	0.763	1.913	3.995	0.876	3.667	6.485	0.730
	SSTZIP-GNN-II	2.562	4.123	0.770	1.850	2.655	0.883	3.485	5.238	0.746
	SSTZIP-GNN-III	**2.421**	**3.514**	**0.830**	**1.434**	**2.016**	**0.887**	**3.334**	**4.652**	**0.771**
Taxi Trajectory−MH	HA	8.371	8.773	0.660	7.611	8.866	0.690	9.212	10.475	0.571
	DCRNN	7.171	7.734	0.710	5.937	7.855	0.722	8.450	9.881	0.590
	STGCN	6.583	6.674	0.721	5.278	7.737	0.741	7.381	8.754	0.647
	STZINB	5.377	6.162	0.730	4.836	6.248	0.760	5.724	7.835	0.677
	STGMT	**4.787**	**5.814**	**0.753**	**4.471**	**5.744**	**0.783**	5.362	6.622	0.701
	STZIP-GNN	5.131	6.126	0.747	4.533	5.991	0.774	**5.078**	**6.227**	**0.723**
	SSTZIP-GNN-I	4.832	6.073	0.750	4.712	6.149	0.771	5.217	6.630	0.720
	SSTZIP-GNN-II	4.774	5.913	0.774	4.579	5.787	0.794	4.833	6.018	0.748
	SSTZIP-GNN-III	**4.113**	**5.347**	**0.793**	**3.957**	**4.875**	**0.807**	**4.661**	**5.844**	**0.762**

resolutions. Other baselines, including STZINB, STGCN, and DCRNN, exhibit higher errors and lower F1-scores, indicating difficulty in modeling sparse spatiotemporal demand. HA yields the weakest performance across all tasks.

Generalization to Taxi Trajectory−MH dataset. To evaluate cross-city generalization, we test the models on the Taxi Trajectory−MH dataset, which corresponds to the public dataset from Manhattan, New York. As shown in the lower half of Table 3, the results largely align with the trends observed on the Taxi Trajectory−HK dataset, confirming the robustness of the proposed framework across diverse urban settings.

SSTZIP-GNN-III once again delivers the strongest overall performance under three time resolutions 10 min, 30 min, and 50 min, achieving the lowest MAE (4.113, 3.957, 4.661) and RMSE (5.347, 4.875, 5.844), along with the highest F1-scores (0.793, 0.807, 0.762). SSTZIP-GNN-II and SSTZIP-GNN-I also perform well, though both are marginally outperformed by STGMT at the 10- and 30-min resolutions. At the 50-min level, however, all SSTZIP-GNN variants surpass the baseline models.

STGMT shows the best generalization among baselines in short to medium horizons, while STZIP-GNN performs slightly better at 50 min. Traditional models, such as DCRNN and HA, produce the highest error rates, highlighting their limited adaptability to cross-domain transfer.

The effectiveness of our proposed models is further illustrated in Fig. 3, where a visual comparison of each model’s average MAE and RMSE is provided. The scatter plot clearly shows SSTZIP-GNN-III positions at the lower-left corner, indicating its superior predictive accuracy with minimal error. SSTZIP-GNN-II and SSTZIP-GNN-I also perform competitively, clustering closely with lower error rates compared to the baseline models. In contrast, baseline models exhibit higher errors, with HA positioned at the far upper-right, reflecting its poor predictive accuracy. Notably, STZIP-GNN and STGMT perform better than other baselines but remain less effective than the proposed SSTZIP-GNN models.

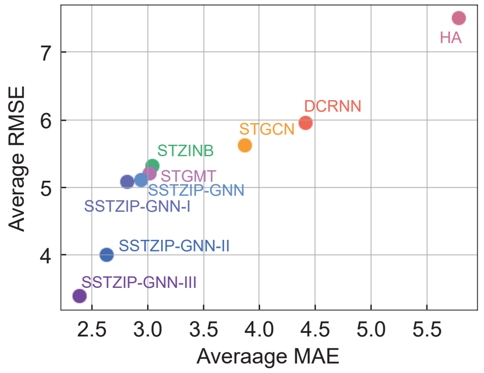


Fig. 3 Comparison of average MAE and RMSE of each model.

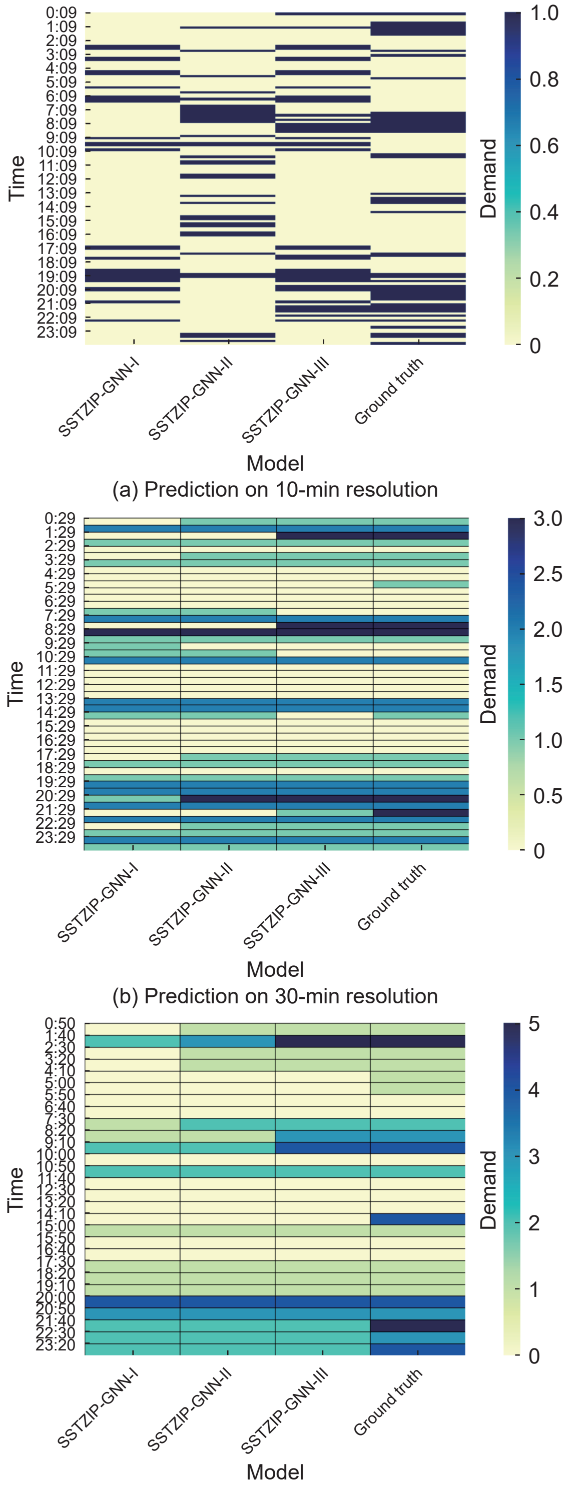
Overall, these results indicate that SSTZIP-GNN-III provides the most reliable predictions across different time resolutions, while SSTZIP-GNN-II and SSTZIP-GNN-I also demonstrate significant improvements over baseline models.

Scalability analysis in specific prediction scenarios. To evaluate the scalability of the proposed SSTZIP-GNN model in taxi demand prediction, we apply its variants to predict demand on a specific day. The predictions are compared with the ground truth, and the performance of each variant is analyzed across different time resolutions (10 min, 30 min, and 50 min). As shown in Fig. 4, all variants perform well, especially in the 30-min resolution, where all models exhibit strong prediction accuracy.

However, scalability varies across prediction scenarios. SSTZIP-GNN-III demonstrates superior scalability compared to SSTZIP-GNN-I and SSTZIP-GNN-II, due to the different time resolution types used in training, which affect the models’ adaptability. As shown in Table 2, SSTZIP-GNN-I is trained on three resolutions (5 min, 30 min, and 60 min), which limits its scalability due to large intervals between them. SSTZIP-GNN-II incorporates 15-min data, improving its ability to capture mid-term demand patterns. SSTZIP-GNN-III, however, integrates time resolution data across a wider range (5 min to 60 min), enabling it to capture both short- and long-term demand patterns, thus improving scalability and prediction accuracy.

When processing the data with a 10-min sampling frequency, taxi demand exhibits greater sparsity, resulting in a significant discrepancy between predicted values and ground truth, as shown in Fig. 4a. SSTZIP-GNN-III performs best, likely due to its use of higher-resolution data during training, allowing it to better capture short-term fluctuations.

In the 30-min resolution, all models show improved performance, with the smallest error between predicted and ground truth values (Fig. 4b). This indicates effective learning of mid-term demand patterns, where the sparsity issue is alleviated by incorporating rich



(a) Prediction on 10-min resolution

(b) Prediction on 30-min resolution

(c) Prediction on 50-min resolution

Fig. 4 Performance comparison of predicted and ground truth taxi demands across different time resolutions.

mid-term data, enabling the models to capture variations more accurately.

At the 50-min resolution, data sparsity is further reduced, and demand trends become more stable, leading to smaller discrepancies between predicted and actual values (Fig. 4c). However, the overall performance is slightly lower than in the 30-min scenario, due to the limited long-term demand data used during training. Nevertheless, SSTZIP-GNN-III still outperforms the other models, benefiting from multi-resolution data fusion, which enhances its ability to capture long-term demand patterns.

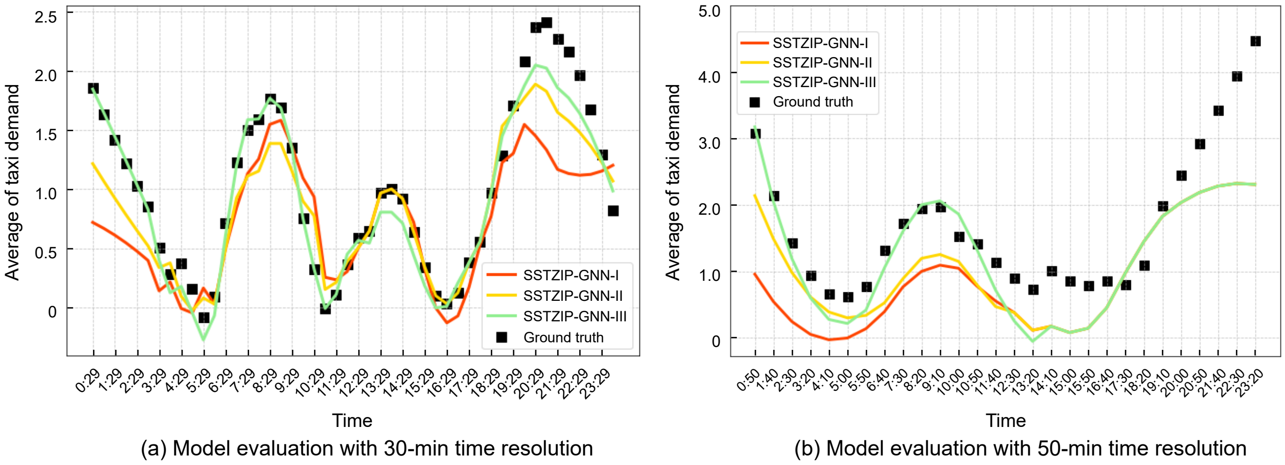
Theoretically, the model's scalability improves by addressing data sparsity and incorporating a wider range of time resolutions during training. However, in practical applications, the model's complexity and training costs must also be considered.

Evaluation of SSTZIP-GNN in capturing dynamic taxi demand patterns. To evaluate SSTZIP-GNN's ability to capture dynamic taxi demand patterns, experiments are conducted on all variants (I, II, and III) in 30-min and 50-min resolution prediction scenarios. Daily average demand fluctuation trends are obtained in Fig. 5, and model predictions are compared with ground-truth data. The results show that all variants effectively capture daily demand fluctuations, reflecting both peak and off-peak variations.

SSTZIP-GNN-III performs best at both resolutions, closely aligning with ground-truth data, demonstrating its strong capability to extract complex spatiotemporal features. However, some deviations occur during peak and off-peak periods. For example, SSTZIP-GNN-I and II tend to underestimate demand during high-demand periods (8:00–9:00 and 18:00–19:00), indicating limited ability to capture rapid fluctuations. In contrast, during low-demand periods (midnight), some models produce overly smoothed predictions, possibly due to insufficient data in low-sample regions. Different time resolutions also impact accuracy. The 30-min resolution scenario shows higher accuracy across all models compared to the 50-min resolution, as the models better capture mid-term fluctuations and are more sensitive to dynamic demand changes. In the 50-min scenario, predictions are smoother, reducing sensitivity to sudden demand shifts.

In summary, SSTZIP-GNN effectively captures dynamic taxi demand characteristics, offering a promising approach for prediction across varying resolutions in dynamic urban environments.

Analysis of computing resource consumption. To explore the advantages of the proposed multi-resolution model in optimizing computational



(a) Model evaluation with 30-min time resolution

(b) Model evaluation with 50-min time resolution

Fig. 5 Evaluation of SSTZIP-GNN model performance for taxi demand prediction at different time resolutions.

resources, we select a subset of data from the original dataset and train both deep learning based baseline models and the SSTZIP-GNN variants. By comparing GPU time consumption for predicting different resolution scenarios, we evaluate the resource performance of each model. As shown in Fig. 6, when predicting a single resolution, all proposed models consume more GPU time than the baseline models, as they require training on multiple resolution datasets, while baseline models use only a single resolution dataset.

However, as the number of predicted scenarios increases to 2, the advantages of SSTZIP-GNN-I and II become more apparent. SSTZIP-GNN-I requires the least GPU time, followed by STGMT and STZIP-GNN, while SSTZIP-GNN-II consumes less GPU time than DCRNN and STGCN. SSTZIP-GNN-III remains the most computationally intensive. When the number of prediction scenarios reaches 3 or more, all baseline models experience a nearly linear increase in GPU time due to the need for multiple parallel prediction systems. In contrast, the three SSTZIP-GNN variants require only a single training session, regardless of the number of scenarios. As the scenario count grows, SSTZIP-GNN's GPU efficiency becomes more pronounced, achieving an average reduction of 46.3% compared to the baseline models.

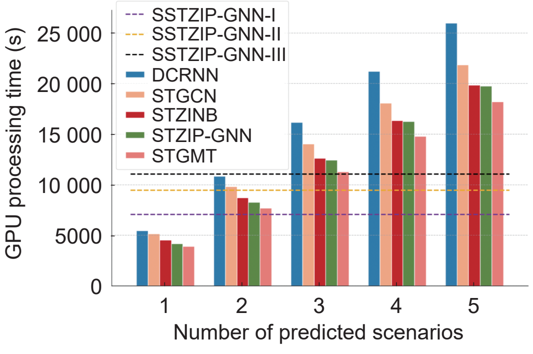
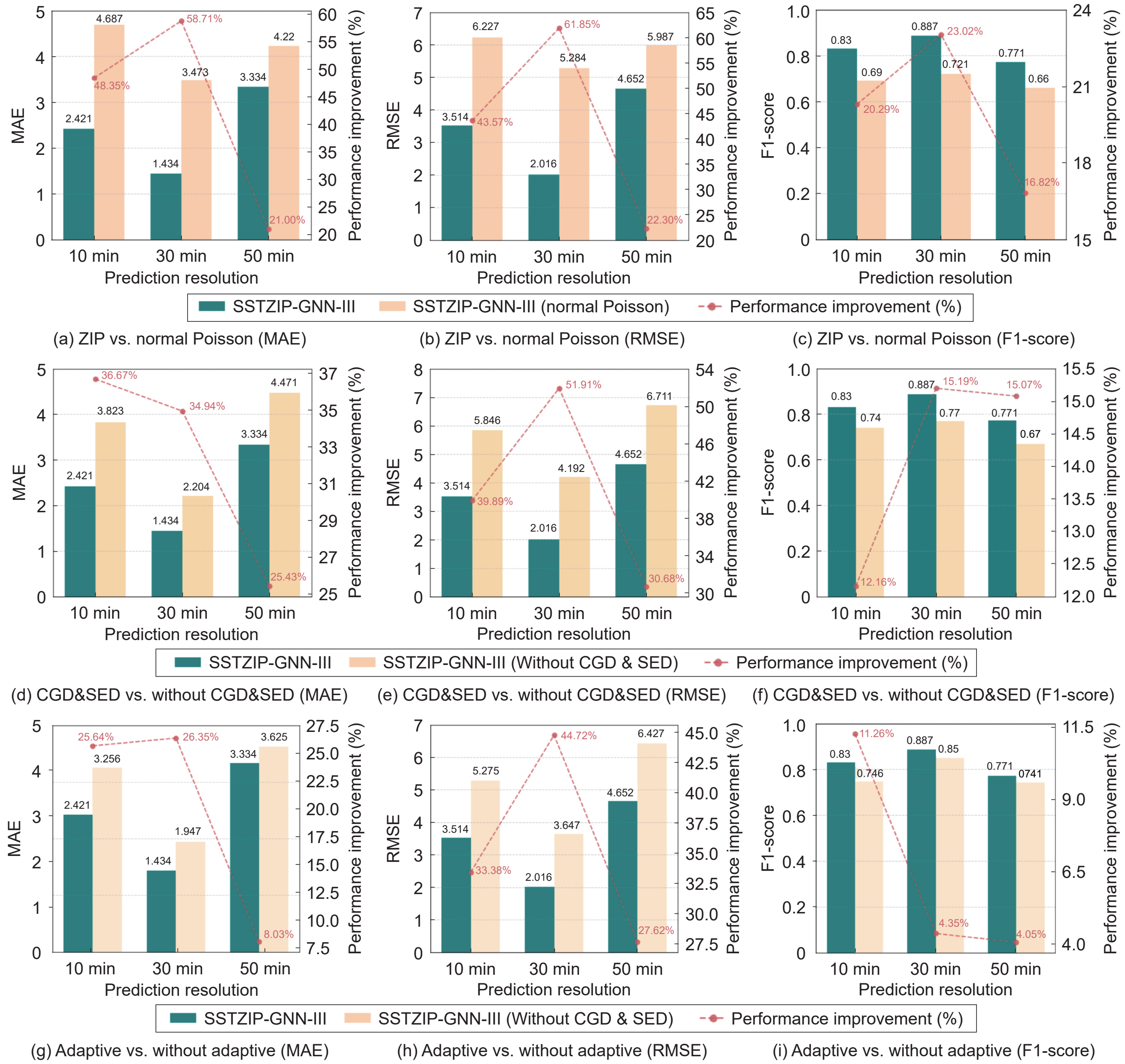


Fig. 6 Comparison of GPU processing time of each model.

In conclusion, the SSTZIP-GNN variants significantly reduce computational resource consumption for multi-resolution prediction tasks. This advantage grows as the number of predicted scenarios increases, and depending on the application, the most suitable SSTZIP-GNN variant can be selected to achieve the best balance between performance and cost.

Ablation experiment. To thoroughly assess the contribution of each key component in the SSTZIP-GNN framework, we conduct ablation studies focusing on three core innovations: the ZIP distribution layer, the adaptive mechanism, and the integration of CGD and SED. By using the SSTZIP-GNN-III model as a baseline, we systematically remove or modify these components to evaluate their individual impact on model performance.

First, we examine the ZIP distribution layer by comparing the performance of SSTZIP-GNN-III with and without this component, evaluating its effect on handling sparse demand. Next, we assess the adaptive mechanism by comparing performance under a fixed learning strategy versus dynamic adaptation based on input data granularity. Finally, we evaluate the impact of external datasets CGD and SED by comparing the model's robustness with and without these additional inputs. The results provide valuable insights into the individual contributions of each module to the overall performance of SSTZIP-GNN.



(a) ZIP vs. normal Poisson (MAE)

(b) ZIP vs. normal Poisson (RMSE)

(c) ZIP vs. normal Poisson (F1-score)

(d) CGD&SED vs. without CGD&SED (MAE)

(e) CGD&SED vs. without CGD&SED (RMSE)

(f) CGD&SED vs. without CGD&SED (F1-score)

(g) Adaptive vs. without adaptive (MAE)

(h) Adaptive vs. without adaptive (RMSE)

(i) Adaptive vs. without adaptive (F1-score)

Fig. 7 Comparison of the results of ablation experiments.

(1) Effectiveness of the ZIP distribution: Figures 7a–7c show the impact of the ZIP distribution on model performance compared to the normal Poisson distribution, evaluated using MAE, RMSE, and F1-score across different prediction resolutions (10 min, 30 min, and 50 min). SSTZIP-GNN-III with the ZIP distribution significantly outperforms the normal Poisson distribution in all metrics. For example, at the 10-min resolution, MAE decreases by 48.35%, and RMSE by 43.57%. At the 30-min resolution, MAE and RMSE decrease by 58.71% and 61.85%, respectively.

Even at a 50-min resolution, where zero-demand occurrences are less frequent, the ZIP distribution significantly enhances model performance. Although zero-inflation is less pronounced at finer temporal resolutions, ZIP improves the model's ability to handle sparse data and low-frequency events, leading to noticeable gains across various performance metrics. These results highlight the effectiveness of the ZIP distribution in addressing zero-inflation, particularly in medium- and short-term prediction scenarios with a higher prevalence of zero-demand instances.

(2) Impact of external data sources on model performance: Figures 7d–7f assess the impact of external data sources, CGD, and SED. The results show that integrating these datasets significantly improves model robustness. SSTZIP-GNN-III with CGD and SED outperforms the model without these inputs across all prediction resolutions. For example, at the 10-min and 30-min resolutions, the model with CGD and SED shows an average improvement of 38.28% in MAE and 43.42% in RMSE. This highlights the importance of CGD and SED in enhancing the model’s accuracy, especially in sparse data scenarios, where CGD helps capture short-term dynamic changes. At the 50-min resolution, the model shows a 15.07% improvement in F1-score, indicating that socioeconomic data enrich the model’s ability to handle low-resolution predictions.

(3) Contribution of the adaptive mechanism: Figures 7g–7i highlight the significant contribution of the adaptive learning mechanism to SSTZIP-GNN-III. Comparing the model with and without the adaptive mechanism underscores the importance of dynamic learning strategies in improving performance. The addition of the adaptive mechanism leads to substantial improvements in MAE, RMSE, and F1-score. It enables the model to better adjust to varying data granularities, improving prediction accuracy. In different resolution scenarios, the model with adaptive learning achieves an average reduction of 20% in MAE and 35.24% in RMSE. These results demonstrate the adaptive mechanism’s ability to handle diverse data granularities and select optimal ZIP parameters, leading to more accurate predictions.

In addition to improving performance, the adaptive module also significantly reduces computational resource consumption. As shown in Fig. 8, when only a single prediction resolution is required, SSTZIP-GNN-III with the adaptive module consumes roughly twice as much GPU time as the model without it. However, as the number of prediction scenarios increases to two, the adaptive model reduces GPU time by about 11%. With three prediction scenarios, the reduction reaches approximately 43.9%. These results highlight the scalability of the adaptive module, especially in dynamic urban environments where multi-resolution predictions are needed. The adaptive mechanism not only reduces resource usage, but also ensures accurate predictions across varying scenarios and data granularities, demonstrating its efficiency in balancing resource consumption while maintaining high accuracy in complex tasks.

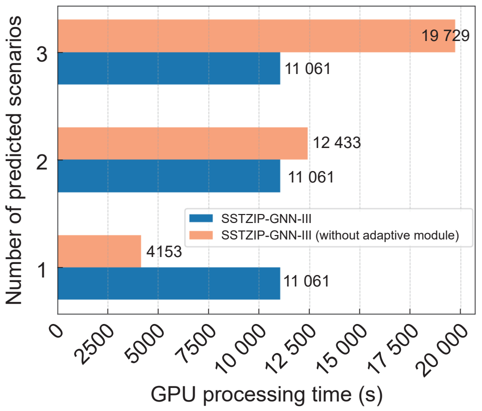


Fig. 8 Adaptive vs. without adaptive (GPU time).

5 Related Work

This section reviews existing research on travel demand prediction, organized into three categories: traditional models, deep learning based spatiotemporal models, and sparse data prediction methods.

5.1 Traditional models

Early efforts in travel demand prediction relied on mathematical and statistical models to capture temporal dependencies in travel data. For example, Li et al.[24] modeled taxi demand as a time series problem using an improved ARIMA method, which demonstrates limited accuracy due to its inability to handle nonlinear dependencies. Similarly, Tong et al.[25] employed a high-dimensional linear regression model for regional taxi demand prediction but faced challenges in modeling complex spatiotemporal interactions. These methods, while foundational, struggle with low accuracy due to their reliance on linear assumptions and inability to capture intricate spatial correlations.

5.2 Deep learning based spatiotemporal models

The advent of deep learning significantly advanced the field by enabling the modeling of complex spatiotemporal dependencies in travel demand data.

Convolutional and recurrent architectures. Yao et al.[26] introduced a hybrid model combining local CNNs for spatial feature extraction with LSTMs for capturing temporal dependencies, demonstrating improved performance over traditional methods. Encoder-decoder frameworks further enhance predictive capabilities by incorporating attention mechanisms to emphasize important spatiotemporal patterns[27, 28]. For instance, Zhou et al.[27] proposed an attention-based encoder-decoder model leveraging ConvLSTM units for citywide passenger demand prediction.

GNNSs. Recent studies have utilized GNNs to address spatial dependencies inherent in transportation networks. He et al.[29] developed a Multi-Graph Convolutional-Recurrent Neural Network (MGC-RNN) to capture inter-station correlations influenced by external factors, such as Points of Interest (PoI). Similarly, Wu et al.[9] proposed a hybrid GCNN-LSTM model to predict urban rail transit passenger flows, integrating inbound-outbound flow dynamics across stations.

While these models excel at capturing spatiotemporal relationships, they are often limited by their reliance on fixed temporal resolutions and lack adaptability to varying data granularities.

5.3 Sparse data prediction methods

Sparse travel demand data presents unique challenges due to the prevalence of zero-demand periods and uneven distribution across time and space.

Several approaches have been developed to address sparsity in travel demand data. Wang et al.[30] introduced a pre-weighted aggregator leveraging grid embeddings to mitigate sparsity at multiple granularities. Zhang et al.[31] proposed a segmentation CNN with masking loss functions to transform sparse traffic data into dense feature representations. To explicitly model zero-inflated distributions, Zhuang et al.[4] developed STZINB-GNN, which combines Zero-Inflated Negative Binomial (ZINB) and Negative Binomial (NB) distributions for sparse OD matrices, incorporating spatiotemporal embeddings for improved predictions. Han et al.[32] extended this concept by introducing layered message-passing modules for virtual clusters to share information with regions.

Recent work has explored hybrid frameworks tailored for sparse scenarios. Lee et al.[33] proposed a multi-task deep learning model for real-time Demand Responsive Transport (DRT) services, incorporating zero-inflated loss functions to simultaneously predict demand probability and volume. Li et al.[34] presented a two-stage framework combining trip generation/ attraction predictions with trip distribution modeling to address sparsity issues effectively.

Despite these advances, existing methods often rely on fixed temporal resolutions (e.g., hourly intervals), limiting their ability to capture multi-time patterns or adapt to varying levels of sparsity across resolutions. By addressing both spatial-temporal dependencies and multiresolution adaptability, SSTZIP-GNN achieves state-of-the-art performance across varying temporal granularities while reducing computational costs compared to ensemble approaches.

6 Limitation and Future Work

While SSTZIP-GNN demonstrates strong performance in multi-resolution taxi demand prediction under data sparsity, several limitations remain. First, the model relies on the availability and quality of multimodal data, including CGD and SED. In cities where such data are unavailable or incomplete, model performance may degrade. Second, the adaptive mechanism, while effective in capturing multi-resolution temporal patterns, introduces additional computational complexity that may pose challenges for real-time deployment in resource-constrained settings. Third, our current evaluation focuses on data from a single metropolitan area. The generalizability of the proposed framework to other cities with different urban topologies, mobility behaviors, and data distributions remains an open question.

In future work, we plan to explore strategies to reduce the model's dependency on auxiliary data by leveraging transfer learning and domain adaptation techniques. Lightweight variants of GNNs and efficient temporal encoding methods will also be investigated to support low-latency, real-time deployment. Furthermore, we intend to evaluate the framework across multiple urban regions to systematically assess its scalability and generalization ability in diverse environments.

7 Conclusion

In this paper, we propose SSTZIP-GNN, that integrates diffusion graph convolution networks and temporal convolutional networks within a ZIP framework. The model effectively captures spatiotemporal dependencies while explicitly modeling structural and sampling zeros, thus addressing data sparsity. Additionally, the adaptive resolution mechanism enhances the scalability of the model across varying temporal granularities. Extensive experiments on real-world datasets demonstrate that SSTZIP-GNN significantly outperforms existing baselines in both accuracy and robustness, particularly in sparse demand scenarios.

Acknowledgment

This research was supported by the Lingnan University (LU) (Nos. DR25F4, DB24C4, and 871242), the Lam Woo Research Fund at LU (No. LWP20021), and the Shenzhen University-Lingnan University Joint Research Programme (No. SZU-LU007/2526).

References

[1] F. R. di Torrepadula, E. V. Napolitano, S. Di Martino, and N. Mazzocca, Machine learning for public transportation demand prediction: A systematic literature review, *Eng. Appl. Artif. Intell.*, vol. 137, p. 109166, 2024.

[2] H. Xu, Y. Chen, C. Li, and X. Chen, Space-time adaptive network for origin-destination passenger demand prediction, *Trans. Res. Part C*: *Emerg. Technol.*, vol. 167, p. 104842, 2024.

[3] S. Guo, B. Deng, C. Chen, J. Ke, J. Wang, S. Long, and K. Xu, Seeking in ride-on-demand service: A reinforcement learning model with dynamic price prediction, *IEEE Internet Things J.*, vol. 11, no. 18, pp. 29890–29910, 2024.

[4] D. Zhuang, S. Wang, H. Koutsopoulos, and J. Zhao, Uncertainty quantification of sparse travel demand prediction with spatial-temporal graph neural networks, in *Proc. 28th ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, Washington, DC, USA, 2022, pp. 4639–4647.

[5] G. Jin, Z. Xi, H. Sha, Y. Feng, and J. Huang, Deep multi-view graph-based network for citywide ride-hailing demand prediction, *Neurocomputing*, vol. 510, pp. 79–94, 2022.

[6] F. Toqué, M. Khouadjia, E. Come, M. Trepanier, and L. Oukhellou, Short & long term forecasting of multimodal transport passenger flows with machine learning methods, in *Proc. 20th Int. Conf. Intelligent Transportation Systems*, Yokohama, Japan, 2017, pp. 560–566.

[7] L. Monje, R. A. Carrasco, C. Rosado, and M. Sánchez-Montañés, Deep learning XAI for bus passenger forecasting: A use case in Spain, *Mathematics*, vol. 10, no. 9, p. 1428, 2022.

[8] N. C. Petersen, F. Rodrigues, and F. C. Pereira, Multi-output bus travel time prediction with convolutional LSTM neural network, *Expert Syst. Appl.*, vol. 120, pp. 426–435, 2019.

[9] J. Wu, X. Li, D. He, Q. Li, and W. Xiang, Learning spatial-temporal dynamics and interactivity for short-term passenger flow prediction in urban rail transit, *Appl. Intell.*, vol. 53, no. 16, pp. 19785–19806, 2023.

[10] J. Zhang, F. Chen, Z. Cui, Y. Guo, and Y. Zhu, Deep learning architecture for short-term passenger flow forecasting in urban rail transit, *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 11, pp. 7004–7014, 2021.

[11] W. Shi, J. Zhang, X. Zhong, X. Chen, and X. Ye, DTS-AdapSTNet: An adaptive spatiotemporal neural networks for traffic prediction with multi-graph fusion, *PeerJ Comput. Sci.*, vol. 10, p. e2527, 2024.

[12] H. Miao, J. Shen, J. Cao, J. Xia, and S. Wang, MBA-STNet: Bayes-enhanced discriminative multi-task learning for flow prediction, *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 7, pp. 7164–7177, 2023.

[13] K. Yao, G. Gao, Y. Liu, X. Ju, and Z. Zhang, A stable passenger flow forecast approach for newly opened metro stations based on multi-source data and random forest regression model, in *Proc. 3rd Int. Conf. Intelligent Design*, Xi’an, China, 2022, pp. 249–254.

[14] A. M. Nagy and V. Simon, Survey on traffic prediction in smart cities, *Pervasive Mob. Comput.*, vol. 50, pp. 148–163, 2018.

[15] C. Ding, J. Duan, Y. Zhang, X. Wu, and G. Yu, Using an ARIMA-GARCH modeling approach to improve subway short-term ridership forecasting accounting for dynamic volatility, *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 4, pp. 1054–1064, 2018.

[16] S. H. Teng, Scalable algorithms for data and network analysis, *Found. Trends*, nos. 1&2, pp. 1–274, 2016.

[17] Y. Wu, D. Zhuang, A. Labbe, and L. Sun, Inductive graph neural networks for spatiotemporal kriging, in *Proc. 35th AAAI Conf. Artificial Intelligence*, Virtual Event, 2021, pp. 4478–4485.

[18] Y. Wu, D. Zhuang, M. Lei, A. Labbe, and L. Sun, Spatial aggregation and temporal convolution networks for real-time kriging, arXiv preprint arXiv: 2109.12144, 2021.

[19] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, Temporal convolutional networks for action segmentation and detection, in *Proc. 2017 IEEE Conf. Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017, pp. 156–165.

[20] Y. Li, R. Yu, C. Shahabi, and Y. Liu, Diffusion convolutional recurrent neural network: Data-driven traffic forecasting, in *Proc. 6th Int. Conf. Learning Representations*, Vancouver, Canada, https://dblp.uni-trier.de/db/conf/iclr/iclr2018.html#LiYS018, 2018.

[21] B. Yu, H. Yin, and Z. Zhu, Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting, in *Proc. 27th Int. Joint Conf. Artificial Intelligence*, Stockholm, Sweden, 2018, pp. 3634–3640.

[22] Y. Wen, Z. Li, X. Wang, and W. Xu, Traffic demand prediction based on spatial-temporal guided multi graph Sandwich-Transformer, *Inf. Sci.*, vol. 643, p. 119269, 2023.

[23] Y. Shen and J. Shen, STZIP-GNN: A robust model for taxi demand prediction in sparse urban environments, in *Proc. 30th Int. Conf. Parallel and Distributed Systems*, Belgrade, Serbia, 2024, pp. 210–217.

[24] X. Li, G. Pan, Z. Wu, G. Qi, S. Li, D. Zhang, W. Zhang, and Z. Wang, Prediction of urban human mobility using large-scale taxi traces and its applications, *Front. Comput. Sci.*, vol. 6, no. 1, pp. 111–121, 2012.

[25] Y. Tong, Y. Chen, Z. Zhou, L. Chen, J. Wang, Q. Yang, J. Ye, and W. Lv, The simpler the better: A unified approach to predicting original taxi demands based on large-scale online platforms, in *Proc. 23rd ACM SIGKDD Int. Conf.*

Knowledge Discovery and Data Mining, Halifax, Canada, 2017, pp. 1653–1662.

[26] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, and Z. Li, Deep multi-view spatial-temporal network for taxi demand prediction, in *Proc. 32nd AAAI Conf. Artificial Intelligence*, New Orleans, LA, USA, 2018, pp. 2588–2595.

[27] X. Zhou, Y. Shen, Y. Zhu, and L. Huang, Predicting multi-step citywide passenger demands using attention-based neural networks, in *Proc. 11th ACM Int. Conf. Web Search and Data Mining*, Marina Del Rey, CA, USA, 2018, pp. 736–744.

[28] A. M. Nayak and N. Chaubey, Predicting passenger flow in BTS and MTS using hybrid stacked auto-encoder and softmax regression, in *Proc. 1st Int. Conf. Computing Science, Communication and Security*, Gujarat, India, 2020, pp. 29–41.

[29] Y. He, L. Li, X. Zhu, and K. L. Tsui, Multi-graph convolutional-recurrent neural network (MGC-RNN) for short-term forecasting of transit passenger flow, *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 18155–18174, 2022.

[30] Y. Wang, H. Yin, H. Chen, T. Wo, J. Xu, and K. Zheng, Origin-destination matrix prediction via graph convolution: A new perspective of passenger demand modeling, in *Proc. 25th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining*, Anchorage, AK, USA, 2019, pp. 1227–1235.

[31] D. Zhang, F. Xiao, M. Shen, and S. Zhong, DNEAT: A novel dynamic node-edge attention network for origin-destination demand prediction, *Transp. Res. Part C: Emerg. Technol.*, vol. 122, p. 102851, 2021.

[32] L. Han, X. Ma, L. Sun, B. Du, Y. Fu, W. Lv, and H. Xiong, Continuous-time and multi-level graph representation learning for origin-destination demand prediction, in *Proc. 28th ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, Washington, DC, USA, 2022, pp. 516–524.

[33] J. Lee, Y. Choi, and J. Kim, A multi-task deep learning framework for forecasting sparse demand of demand responsive transit, *Expert Syst. Appl.*, vol. 250, p. 123833, 2024.

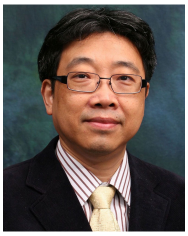
[34] D. Li, W. Wang, and D. Zhao, Designing a novel two-stage fusion framework to predict short-term origin-destination flow, *J. Transp. Eng., Part A: Syst.*, vol. 149, no. 5, p. 04023032, 2023.

Yifei Shen received the BS and MS degrees from University of Sussex, UK in 2018 and 2021, respectively. He is currently an MPhil student at School of Data Science, Lingnan University, Hong Kong, China. and will pursue the PhD degree in data science at Lingnan University starting in 2025. He has published several papers in conferences and journals, including ICPADS, DASFAA, BDAI, and *Frontiers in Public Health*. His research interests include spatiotemporal data mining, deep learning, big data analytics, and their applications in smart transportation.

Jiaxing Shen is an assistant professor at School of Data Science, Lingnan University, China. He received the BEng degree in software engineering from Jilin University, China in 2014, and the PhD degree in computer science from The Hong Kong Polytechnic University, China in 2019. He was a visiting scholar at the Media Lab, Massachusetts Institute of Technology, USA in 2017. His research interests include human-centric computing, mobile computing, and data mining. His research has been published in top-tier journals, such as *IEEE TMC*, *ACM TOIS*, *ACM IMWUT*, and *IEEE TKDE*. He was awarded conference best paper twice including one from IEEE INFOCOM 2020.

Wenlong Shi received the BEng degree in software engineering from Guangzhou University, China in 2022. He is currently a PhD candidate in computer technology at Shenzhen University, China. He has published several papers in journals, such as *The Computer Journal* and *Scientific Reports*. His research interests include smart transportation, data mining, machine learning, and privacy protection.

Hengzhi Wang is currently serving as an assistant professor at College of Computer Science and Software Engineering, Shenzhen University, China. He received the BEng degree in software engineering from Jilin University, China in 2017, and the PhD degree in computer science from the same university in 2023. During his doctoral studies, he also worked as a visiting PhD student at School of Computing Science, Simon Fraser University in British Columbia, Canada. His research interests primarily focus on spatial crowdsourcing, federated learning, and privacy protection. He is a member of the IEEE.



Jiannong Cao received the MEng and PhD degrees in computer science from Washington State University, USA in 1986 and 1990, respectively. He is currently the Otto Poon Charitable Foundation Professor in data science and the chair professor at Department of Computing, The Hong Kong Polytechnic University (PolyU), China, where he is also the dean of the Graduate School, the director of the Research Institute of Artificial Intelligent of Things, and the vice director of the University's Research Facility in Big Data Analytics. His current research interests include wireless sensing and networking, big data and machine learning, and mobile cloud and edge computing. He is a member of the Academia Europaea and an ACM distinguished member. He served as the chair and member of organizing and technical committees for many international conferences, such as IEEE INFOCOM and IEEE PERCOM, and as an associate editor for many international journals, such as *IEEE Transactions on Computers*, *IEEE Transactions on Parallel and Distributed Systems*, and *IEEE Transactions on Big Data*. He is a fellow of IEEE.

Hanqing Wu received the BEng degree in software engineering from Tongji University, China in 2010, and the PhD degree from The Hong Kong Polytechnic University, Hong Kong, China in 2022. He served as a research assistant at The Hong Kong Polytechnic University from May 2005 to December 2015. Since 2021, he has held the position of CEO at Lucky Technology Development Limited, Hong Kong, China. His research interests focus on distributed computing, blockchain, and big data.