Spatio-Temporal Heterogeneous Learning For Traffic Flow Forecasting In Smart Transportation

Kaibin Wei[1*], Jianqiang Jing[1], Haifeng Li[2*], Xiannian Xie[1], and Furong Li[1]

[1] School of Electronic Information and Electrical Engineering, Tianshui Normal University, Tianshui, 741000, China
[2] School of Mathematics and Statistics, Fuyang Normal University, Fuyang 236037, China
[*] Corresponding author. E-mail: weikaibin@tsnu.edu.cn; lihaifengdlut@163.com

Received: Aug. 23, 2025; Accepted: Sep. 27, 2025

Accurate traffic flow forecasting is crucial for enhancing urban transportation efficiency and travel experiences. However, existing methods face challenges in capturing the complex spatio-temporal heterogeneity of traffic data. This paper introduces a novel Spatio-Temporal Heterogeneous Learning (STHL) framework for traffic flow forecasting. The framework encompasses three key components: dual spatio-temporal feature extraction, cluster-invariant spatial heterogeneity learning, and information-driven temporal heterogeneity learning. Dual spatio-temporal feature extraction employs semantic and structural augmentations to enrich traffic flow representation learning, capturing spatial and temporal dependencies comprehensively. Cluster-invariant spatial heterogeneity learning distinguishes traffic patterns across urban regions, while information-driven temporal heterogeneity learning injects time - aware heterogeneity into node representations. Experiments on four real - world traffic flow datasets demonstrate that our method outperforms existing state-of-the-art approaches in terms of MAE and MAPE metrics, showcasing its effectiveness in capturing spatio-temporal heterogeneity for enhanced traffic flow prediction accuracy.

Keywords: Traffic flow forecasting; Spatio-temporal heterogeneous learning; graph contrastive learning

http://dx.doi.org/10.6180/jase.202606_29(6).0011

1. Introduction

With the acceleration of global urbanization, the surge in urban populations and the number of motor vehicles has led to a significant increase in traffic flow [1]. This has resulted in problems such as traffic congestion, frequent accidents, and environmental pollution, severely impacting travel experiences and causing substantial economic losses [2, 3]. To address these challenges, many cities are actively constructing Intelligent Transportation Systems (ITS). Traffic flow prediction, as one of the core functions of ITS, provides real-time traffic data and trend forecasts. This enables optimized traffic management, such as traffic light scheduling, dynamic guidance, and emergency response. Accurate predictions also support the development of technologies like shared mobility, intelligent navigation, and autonomous driving, helping management authorities alleviate traffic pressure and enhance travel experiences [4–6].

Existing traffic flow prediction methods are mainly categorized into traditional statistical models, machine learning methods, and deep learning methods. Traditional statistical models like ARIMA rely on the linear characteristics of time series and struggle to handle non-linear data [7]. Long short-term memory can capture long-term temporal dependencies but lacks sufficient spatial feature modeling [8]. Convolutional neural network excels at extracting spatial features but has difficulty modeling complex traffic networks [9]. Graph convolutional network captures node relationships but has limited ability to handle dynamic temporal changes [10]. Spatio-temporal graph convolutional networks and graph attention networks combine

temporal and spatial modeling with attention mechanisms [11]. However, existing models often fail to fully exploit the spatial-temporal heterogeneity inherent in urban traffic flow, leading to suboptimal performance in long-term or real-time predictions. Spatial heterogeneity refers to the significant differences in traffic flow patterns across different urban areas [12]. For instance, areas with different functions such as residential zones,
commercial districts, and transportation hubs exhibit distinct traffic flow patterns [13]. Temporal heterogeneity refers to the variation of traffic flow patterns over time, where patterns differ significantly across different time periods [14]. Traffic flow patterns on weekdays differ from those on weekends; patterns during morning and evening rush hours differ from those during off-peak hours. Even within the same area, traffic flow patterns may vary considerably across different time periods, e.g., morning, afternoon, evening [15].

To address the challenges in traffic flow forecasting, this paper proposes a novel Spatio-Temporal Heterogeneous Learning (STHL) framework that effectively captures the complex spatial and temporal heterogeneity in traffic flow data. STHL integrates three key components: dual spatio-temporal feature extraction, cluster-invariant spatial heterogeneity learning, and information-driven temporal heterogeneity learning. First, STHL employs dual spatio-temporal feature extraction by combining semantic and structural augmentations with an iterative encoding mechanism to enhance the representation learning of traffic flow data, enabling the model to capture spatial and temporal dependencies more comprehensively and effectively. Second, STHL applies cluster-invariant spatial heterogeneity learning to distinguish traffic patterns across different urban regions by clustering nodes into different groups based on their traffic characteristics, thereby capturing the unique traffic patterns of various urban areas. Finally, STHL implements information-driven temporal heterogeneity learning to inject time-aware heterogeneity into node representations by maximizing the mutual information between city representations and node representations at different time steps, thereby strengthening the model's ability to handle dynamic traffic conditions and improving the accuracy of traffic flow forecasting.

Our paper makes the following three main contributions:

- We propose a novel spatio-temporal heterogeneous learning framework that effectively captures the complex spatial and temporal heterogeneity in traffic flow data. This framework integrates dual spatio-temporal feature extraction, cluster-invariant spatial heterogeneity learning, and information-driven temporal heterogeneity learning, addressing the limitations of existing methods in modeling the diverse patterns of urban traffic flow.
- We introduce a dual spatio-temporal feature extraction module that combines semantic and structural augmentations with an iterative encoding mechanism. This module enhances the representation learning of traffic flow data, enabling the model to capture spatial and temporal dependencies more comprehensively and effectively than previous approaches.
- We design an information-driven temporal heterogeneity learning approach that injects time-aware heterogeneity into node representations. By maximizing the mutual information between city representations and node representations at different time steps, this approach strengthens the model's ability to handle dynamic traffic conditions and improves the accuracy of traffic flow forecasting.

The framework is structured as follows: Section 2 delves into the proposed STHL, detailing its core components and their operational mechanisms. Section 3 rigorously evaluates the efficacy of STHL on benchmark datasets and conducts an extensive comparative analysis against state-of-the-art methods. Finally, Section 4 encapsulates the key advancements achieved in STHL and outlines potential directions for further research and development.

2. Method

2.1. Problem Definition

In this paper, the traffic flow forecasting task is regarded as a time series prediction problem. The objective is to predict future traffic flow by utilizing the structural information of the traffic network and historical observation data. Specifically, the traffic network can be represented as a graph $G(V, E, A)$, where V is the set of nodes, i.e., the locations of monitoring nodes or sensors; E is the set of edges, describing the connections between nodes; and $A \in R^{N \times N}$ is the adjacency matrix, containing the connection weights or distance information between nodes. At time T, each monitoring node generates a feature vector. Therefore, the observation data of the entire network can be represented as a graph signal matrix $X_T \in R^{N \times F}$, where N is the number of nodes and F is the feature dimension. Suppose there are i historical time intervals, and each time T has a feature matrix X_T, forming the historical sequence $H = \{X_{T-i+1}, X_{T-i+2}, \ldots, X_T\}$. Based on this historical data sequence, the goal is to predict the traffic

flow for the future T' time steps, i.e., the future sequence $P = \{X_{T+1}, X_{T+2}, \ldots, X_{T+T'}\}$. In summary, this problem can be formalized as a mapping relationship:

$$f : \{X_{T-i+1}, X_{T-i+2}, \ldots, X_T\} \rightarrow \{X_{T+1}, X_{T+2}, \ldots, X_{T+T'}\} \tag{1}$$

where f is the forecasting function from historical traffic flow data to future flow data.

To implement the above goal, a spatio-temporal traffic flow heterogeneous learning method is proposed, which consists of dual spatio-temporal feature extraction, cluster-invariant spatial heterogeneity learning, and information-driven temporal heterogeneity learning, as shown in Fig. 1.

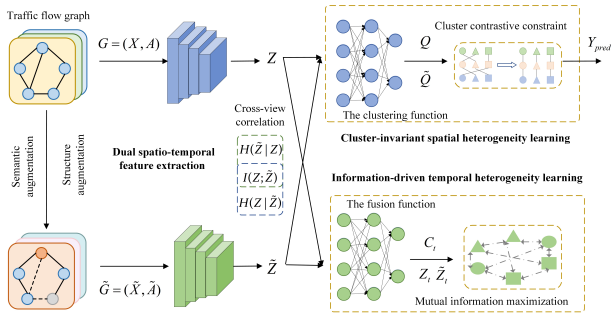


Fig. 1. The illustration of STHL, consisting of dual spatio-temporal feature extraction, cluster-invariant spatial heterogeneity learning, and information-driven temporal heterogeneity learning

2.2. Dual spatio-temporal feature extraction

Dual spatio-temporal feature extraction aims to learn a compact discriminative representation of traffic flow data. To achieve this, dual spatio-temporal feature extraction conducts an adaptive graph augmentation based on structure and semantic perturbations, and then leverages an iterative encoding mechanism to endow representations with fruitful spatial-temporal information.

Semantic augmentation. Given the traffic flow data $X_{T-i+1:T} = \{x_{\tau,n} \mid \tau \in [T-i+1, T], n \in [1, N]\}$, indicating the traffic flow of each node from time step $T-i+1$ to T. For the traffic flow $x_{\tau,n}$ of node n at time step τ, we decide whether to mask it based on the masking probability $\rho_{\tau,n}$. The masking probability $p_{\tau,n}$ follows a Bernoulli distribution, i.e., $\rho_{\tau,n} : \mathrm{Ber}\, n\,(1 - p_{\tau,n})$, where $\rho_{\tau,n} = w_0 x_{\tau,n}$ represents the correlation between the traffic pattern of node \(n\) at time step \(\ tau \) and the overall traffic pattern. w_0 denotes the learnable parameter. The data after traffic-level
augmentation is denoted as $X^0_{T-i+1:T}$, and the calculation is as follows:

$$\tilde{x}_{\tau,n} = \begin{cases} 0, & \text{if } \rho_{\tau,n} = 1 \\ x_{\tau,n} & \text{if } \rho_{\tau,n} = 0 \end{cases} \tag{2}$$

Here, $\rho_{\tau,n}$ can take a value of 1 or 0 , representing masking or not masking, respectively.

Structure augmentation. For two spatially adjacent nodes m and n, if their traffic patterns are not strongly dependent, that is, if the heterogeneity degree $q_{m,n}$ is high, we mask the connecting edge $e_{m,n} \in E$. The masking probability $\rho_{m,n}$ follows a Bernoulli distribution, i.e., $\rho_{m,n} : \mathrm{Ber}\, n\,(1 - q_{m,n})$. The masked set of edges $\tilde{\varsigma}'$ is:

$$\begin{aligned} \tilde{E} &= \{e_{m,n} \in E \mid \rho_{m,n} = 0\} \\ q_{m,n} &= \frac{x_{:,m} \cdot x_{:,n}}{|x_{:,m}|\,|x_{:,n}|} \end{aligned} \tag{3}$$

For two non-adjacent nodes m and n, if their traffic patterns are strongly dependent, that is, if the heterogeneity degree $q_{m,n}$ is low, we add an edge between them. The probability of adding an edge also follows a Bernoulli distribution, i.e., $\rho_{m,n} : \mathrm{Ber}\, n\,(q_{m,n})$. The set of edges after adding edges E' (is:

$$\tilde{E} = \{e_{m,n} \notin E \mid \rho_{m,n} = 1\} \tag{4}$$

After the semantic augmentation and the structure augmentation, we obtain the new adjacency matrix $\tilde{A}$, where the element $\tilde{A}_{m,n}$ indicates whether there is a connection between node m and n . Then, we get the augmented traffic flow graph $\tilde{G} = (V, \tilde{A}, \tilde{E})$.

Iterative encoding mechanism is proposed to capture both temporal and spatial dependencies in traffic flow data. First, a temporal convolution netwrok $TC = (\cdot)$ is applied to the traffic flow data, to generate time-aware representations that reflect the dynamic nature of traffic over time:

$$(M_{T-i+1}, \ldots, M_T) = TC\,(X_{T-i+1}, \ldots, X_T) \tag{5}$$

where $M_T \in R^{N \times D}$ denotes node representations at time step T , with D representing the representation dimension. Subsequently, a graph convolution (SC) network is employed to capture spatial correlations among different node, which leverages the adjacency matrix A of the traffic graph G to propagate information across spatially related regions:

$$B_T = GC\,(M_T, A) \tag{6}$$

Then, we stack multiple such blocks to refine representations. Finally, we obtain ultimate final representations $Z \in R^{T \times N \times D}$ and augmentation representations

$\tilde{Z} \in R^{T\times N\times D}$. Meanwhile, to ensure semantic representations across views, the conditional entropy is minimized as follows:

$$\min H(\tilde{Z} \mid Z) = \min -E_{P_{Z,Z}}[\log(P(\tilde{Z} \mid Z))] \tag{7}$$

Then, a variational distribution $Q(\tilde{Z} \mid Z)$ is used to approximate true distribution $P(\tilde{Z} \mid Z)$ for computing the conditional entropy via maximizing the lower bound:

$$\max E_{P_{Z,Z}}[\log(Q(\tilde{Z} \mid Z))] \tag{8}$$

Next, a Gaussian distribution $N(\tilde{Z} \mid \tilde{G}(Z), \sigma \mathrm{I})$ is used to define Q :

$$\begin{gathered} \max E_{P_{Z,Z}}[\log(Q(\tilde{Z} \mid Z))] = \\ E_{P_{Z,Z}}\left[-\frac{(\tilde{Z}-\tilde{G}(Z))}{2\sigma I}+\log\frac{1}{\sqrt{2\pi\sigma I}}\right] \end{gathered} \tag{9}$$

where $\tilde{G}(\cdot)$ and σI denote the inter-view correlation function and a variance matrix, respectively.
By ignoring the constant, the inter-view correlation loss is derived as follows:

$$L_e = \|\tilde{Z} - \tilde{G}(Z)\|_2^2 + \|Z - G(\tilde{Z})\|_2^2 \tag{10}$$

2.3. Cluster-invariant spatial heterogeneity learning

In traffic flow prediction, spatial heterogeneity poses a significant challenge as different urban regions exhibit distinct traffic patterns influenced by diverse functionalities like residential, commercial, and transportation areas. To tackle this issue, we propose a cluster-invariant spatial heterogeneity learning approach, which captures the unique traffic characteristics of various regions.

Specifically, we partition n region nodes into k clusters via the clustering function $f(\cdot)$

$$\begin{aligned} f\{c_1,\ldots,c_K\} : \{Z_n\}_{n=1}^N \rightarrow \{Q_n\}_{n=1}^N \\ f\{\tilde{c}_1,\ldots,\tilde{c}_K\} : \{\tilde{Z}_n\}_{n=1}^N \rightarrow \{\tilde{Q}_n\}_{n=1}^N \end{aligned} \tag{11}$$

where $\{Q_n\}_{n=1}^N$ and $\{\tilde{Q}_n\}_{n=1}^N$ denote a soft assignment probability with a sum of one, respectively. In generally, data augmentation do not change the original semantics. Thus, we consider Q^j and $\tilde{Q}^j$ as the j -th column of Q and $\tilde{Q}$ where each element denotes a soft cluster assignment. Then, a cluster-invariant spatial heterogeneity learning loss is designed as follows:

$$L_{csh} = -\frac{1}{K}\sum_{k=1}^{K}\log\frac{e^{s\left(Q^k,\tilde{Q}^k\right)/\varpi}}{\sum_{j=1}^{K}e^{s(Q^j,Q^k)/\varpi}+\sum_{j=1}^{K}e^{s\left(Q^j,\tilde{Q}^k\right)/\varpi}} \tag{12}$$

where $s(\cdot)$ denotes the similarity function. ϖ is set as 0.1 . By contrasting positive and negative cluster assignment pairs, the loss function encourages the model to distinguish between different clusters more effectively. This helps in capturing the unique traffic characteristics of various urban regions. Meanwhile, a regularization loss is devised to avoid partitioning all nodes into a cluster:

$$L_a = \sum_{j=1}^{K}\frac{\sum_{i=1}^{N}q_{ij}}{N}\log\frac{\sum_{i=1}^{N}q_{ij}}{N}+\frac{\sum_{i=1}^{N}\tilde{q}_{ij}}{N}\log\frac{\sum_{i=1}^{N}\tilde{q}_{ij}}{N} \tag{13}$$

2.4. Information-driven temporal heterogeneity learning

In this section, a information-driven temporal heterogeneity learning is designed to inject time-aware heterogeneity into node representations by enhancing the divergence among time-step representations.

Specifically, the city representations can be obtained the fusion function as follows:

$$c_t = \frac{1}{N}\sum_{n=1}^{N} w_o z_{t,n} + w_a \tilde{z}_{t,n} \tag{14}$$

where w_o and w_a denotes fusion weights. Then, the mutual information $I(c_t; z_t)$ and $I(c_t; \tilde{z}_t)$ are maximized to learn temporal heterogeneity:

$$\begin{aligned} \max I(c_t; z_t) &= \iint p(z_t \mid c_t)\, p(c_t) \log\frac{p(z_t \mid c_t)\, p(c_t)}{p(z_t)} \\ &= KL\left(p(z_t \mid c_t)\, p(c_t) \,\|\, p(z_t)\, p(c_t)\right) \end{aligned} \tag{15}$$

where $KL(\cdot)$ is the Kullback-Leibler divergence. Next, Jensen-Shannon (JS) divergence is used to measure the mutual information due to unbounded property of KL divergence, which further is redefined as:

$$\begin{aligned} \max I(c_t; z_t) = &-\left[\log\rho\left(\log\frac{2p(c_t)}{p(c_t)+q(c_t)}\right)\right] \\ &-\left[\log\left(1-\rho\left(\log\frac{2p(c_t)}{p(c_t)+q(c_t)}\right)\right)\right] \end{aligned} \tag{16}$$

where $\rho(\cdot)$ is a discriminator. In experiments, negative sample estimation is used to optimize the information maximization loss. Specifically, we consider node level and city level representations at the same time step as positive, while representations at different time steps are considered negative. Through this design, positive supervision will encourage consistency in time specific urban traffic trends (such as peak hours, weather factors), while
negative supervision will help capture time heterogeneity

between different time steps. the discriminator $\rho(\cdot)$ is then used to distinguish between negative and positive sample pairs to capture time heterogeneity. Then, the loss of the information-driven temporal heterogeneity learning is as follows:

$$L_{itl} = -\max\left(I\left(c_t; z_t\right) + I\left(c_t; \tilde{z}_t\right)\right) \tag{17}$$

2.5. The network optimization

The overall loss L that guides spatio-temporal heterogeneous learning for the traffic flow prediction is as follows:

$$L = L_m + L_a + \alpha L_e + \beta L_{csh} + \gamma L_{itl} \tag{18}$$

where α, β, and γ are trade-off parameters. L_m denotes MSE loss:

$$L_m = \mathrm{MSE}\left(y_{\text{true}}, y_{\text{pred}}\right) \tag{19}$$

where y_{pred} denotes the traffic flow prediction from Z by a linear predictor. y_{true} denotes true traffic flow values.

3. Result and discussion

3.1. Dataset and Setup

Datasets and Metrics [16, 17]: Four public real-world traffic flow datasets are utilized to evaluate the predictive performance of the proposed method. NYCBike1 records from April 1, 2014 to September 30, 2014, measured at 30 -minute intervals. NYCBike2 records from July 1, 2016 to August 29, 2016, measured at 30-minute intervals. NYCTaxi records from January 1, 2015 to March 1, 2015, measured at 30-minute intervals. BJTaxi records from March 1, 2015 to June 30, 2015, measured at 60 -minute intervals. Each dataset is partitioned into training, validation, and test sets following a 70% : 10% : 20% ratio respectively. MAE and MPAE are used as metrics to evaluate the predictive performance of the proposed method.

3.2. Implementation detail

We use PyTorch as the deep learning framework for model implementation. All models are trained on a single NVIDIA Tesla V100 GPU with 32 GB memory. For model training, we adopt the Adam optimizer with a learning rate of 0.001 , which is adjusted using a cosine annealing scheduler. The batch size is set to 64 for all experiments.

3.3. Performance Comparison

Comparison methods. Six traffic flow prediction methods are used to verify the performance of the proposed method, containing ARIMA [1], GMLP [2], LSTTN [7], Dstagnn [13], BST [14], and DSHL [15]. Comparison analysis: As shown in Table 1, the proposed method achieves the best performance compared with other methods. Specifically, in terms of MAE and MAPE metrics across all four datasets, EHCL consistently outperforms the existing state-of-the-art methods. For instance, on the NYC Bikel dataset, the inflow and outflow MAE of the proposed method are reduced by 11.5 and 13.3 respectively compared to the second best method DSHL. These results demonstrate that the proposed method effectively captures the spatio-temporal heterogeneity in traffic flow data and provides more accurate predictions. The superior performance of the proposed method can be attributed to its novel spatio-temporal heterogeneous learning framework. The dual spatio-temporal feature extraction module leverages semantic and structural augmentations to enhance the representation learning of traffic flow data, capturing both spatial and temporal dependencies more comprehensively. The cluster-invariant spatial heterogeneity learning approach effectively distinguishes different traffic patterns across urban regions, while the information-driven temporal heterogeneity learning injects time-aware heterogeneity into node representations, enhancing the model's ability to handle dynamic traffic conditions. Overall, these components work synergistically to enable the proposed method to achieve superior traffic flow forecasting performance.

3.4. Ablation Analysis

This section conducts loss ablation experiments of the proposed method. There are four variants. Our w/o L_a denotes the removal of the regularization loss that avoids partitioning all nodes into a cluster. Our w/o L_e denotes the removal of the inter-view correlation loss. Our w/o L_{csh} denotes the removal of the loss of the cluster-invariant spatial heterogeneity learning. Our w/o L_{ith} denotes the removal of the loss of the information-driven temporal heterogeneity learning.

The ablation results shown in the Table 2 lead to two key observations. First, ablation of any loss function diminishes the model's prediction performance, demonstrating the usefulness of each individual loss. Second, the combination of all losses yields the optimal performance, which validates the rationality of the overall loss design. Specifically, each loss function plays a distinct and indispensable role in the model's performance. For example, L_a ensures diverse traffic pattern representation across regions, L_e facilitates effective multi-view information integration, L_{csh} helps distinguish regional traffic patterns, and L_{ith} endows the model with dynamic temporal adaptability. When all losses work synergistically, they enable the model to fully capture the complexity of traffic data, including spatial and

Table 1. Comparison results on four datasets about MAE and MAPE

Dataset		Type	ARIMA	BST	Dstagnn	DSHL	LSTTN	GMLP	Ours
NYCBike1	MAE	Inflow	0.107	0.073	0.055	0.068	0.058	0.065	0.052
		Outflow	0.113	0.080	0.057	0.072	0.061	0.068	0.055
	MAPE	Inflow	0.331	0.254	0.255	0.317	0.265	0.321	0.244
		Outflow	0.350	0.274	0.264	0.343	. 276	0.329	0.256
NYCBike 2	MAE	Inflow	0.089	0.128	0.056	0.052	0.053	0.058	0.050
		Outflow	0.087	0.115	0.053	0.050	0.049	0.055	0.048
		Inflow	0.289	0.465	0.322	0.274	0.293	0.307	0.241
		Outflow	0.282	0.419	0.305	0.268	0.280	0.300	0.239
NYCTaxi	MAE	Inflow	0.209	0.522	0.151	0.137	0.137	0.163	0.125
		Outflow	0.168	0.417	0.121	0.108	0.108	0.125	0.105
	MAPE	Inflow	0.215	0.651	0.227	0.227	0.229	0.240	0.187
		Outflow	0.212	0.641	0.220	0.220	0.224	0.233	0.187
BJTaxi	MAE	Inflow	0.215	0.528	0.121	0.131	0.127	0.138	0.115
		Outflow	0.216	0.527	0.122	0.132	0.128	0.139	0.116
	MAPE	Inflow	0.231	0.655	0.155	0.187	0.172	0.193	0.152
		Outflow	0.207	0.655	0.156	0.188	0.174	0.194	0.153

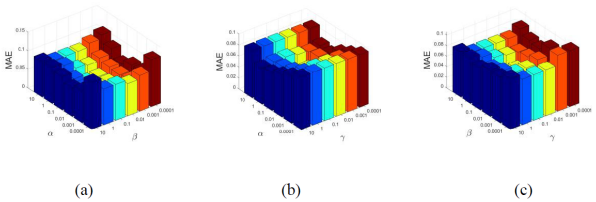
Table 2. Ablation Analysis of losses on four datasets in terms of MAE

Variant	NYCBike1		NYCBike2		NYCTaxi		BJTaxi	
	Inflow	Outflow	Inflow	Outflow	Inflow	Outflow	Inflow	Outflow
Our w/o L_a	0.055	0.055	0.055	0.055	0.055	0.055	0.055	0.055
Our w/o L_e	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.060
Our w/o L_{csh}	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.063
Our w/o L_{ith}	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062
Ours	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052

temporal heterogeneity, thereby maximizing prediction accuracy.

3.5. Parameter Analysis

Following [18, 19], this section conducts parameter analysis experiments of α, β, and γ on the NYCBike1 dataset in terms of MAE of inflow. These parameters are limited to $\{10, 1, 0.1, 0.01, 0.001, 0.0001\}$. During the experiments, one parameter is varied while the other two are kept fixed, and the changes in MAE are recorded in Fig. 2. The results show that encouraging performance is achieved when all three parameters are within the range of [1,0.1]. Therefore, in the subsequent parameter search on all datasets, the search range for these three parameters is set to this interval.



(a) (b) (c)

Fig. 2. Parameter analysis of α, β, and γ on the NYCBike1 dataset in terms of MAE of inflow

3.6. Cluster Analysis

This section conducts cluster number analysis K on the four datasets. In the experiments, the cluster number K is set as $\{3, 4, 5, 6, 7\}$. As shown in Fig. 3, for the NYCBike1 dataset, the metric value decreases from 0.064 at K = 3 to 0.052 at K = 6, then slightly increases to 0.055 at K = 7. Similarly, the NYCBike2 dataset's value drops from 0.058 at K = 3 to 0.05 at K = 6, before rising to 0.053 at K = 7. For NYCTaxi, the value decreases from 0.142 at K = 3 to 0.125 at K = 6, then goes up to 0.13 at K = 7. The BJTaxi dataset sees a slight increase from K = 4 to 5 . Overall, the trend indicates that K = 6 is the optimal choice for these four datasets.

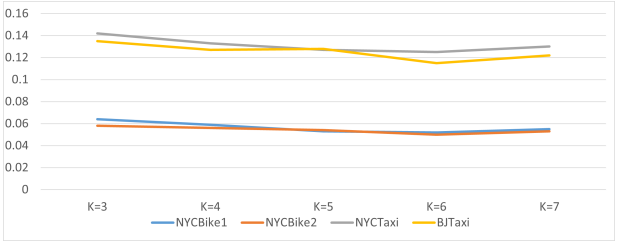


Fig. 3. Cluster number analysis K on the four datasets

4. Conclusions

This paper proposes a novel spatio-temporal heterogeneous learning framework for traffic flow forecasting. The framework effectively captures spatial and temporal heterogeneity in traffic flow data through three key components: dual spatio-temporal feature extraction that leverages semantic and structural augmentations to enhance representation learning, cluster-invariant spatial heterogeneity learning that distinguishes traffic patterns across urban regions, and information-driven temporal heterogeneity learning that injects time-aware heterogeneity into node representations. Experimental results on multiple real-world datasets show superior performance over existing methods in terms of MAE and MAPE metrics. In future research directions, we will also explore the integration of multi-source data fusion techniques to combine traffic flow data with other relevant data types, such as weather conditions, special events, and public transportation schedules. This fusion aims to provide a more comprehensive understanding of the factors influencing traffic flow and enhance the model's adaptability to various real-world scenarios. Additionally, we intend to investigate the application of reinforcement learning strategies to dynamically adjust model parameters based on real-time traffic conditions, thereby improving the model's responsiveness and prediction accuracy in rapidly changing environments.

5. Acknowledgments

This work was supported in part by the Gansu Province Natural Foundation under Grant No.23JRRE0740, and the Doctoral Foundation of Fuyang Normal University under Grant No.2025KYQD0031.

References

[1] W. Zhao, G. Yuan, Y. Zhang, X. Liu, S. Liu, and L. Zhang, (2025) *"An Interpretable and Efficient Multi-scale Spatio-Temporal Neural Network for Traffic Flow Forecasting"* **Expert Systems with Applications**: 128961. DOI: 10.1016/j.eswa.2025.128961.

[2] Y. Luo, J. Zheng, X. Wang, X. Jiang, Z. Zhu, et al., (2025) *"ST-GMLP: A concise spatial-temporal framework based on gated multi-layer perceptron for traffic flow forecasting"* **Neural Networks *184***: 107074. DOI: 10.1016/j.neunet.2024.107074.

[3] J. Gao, M. Liu, P. Li, A. A. Laghari, A. R. Javed, N. Victor, and T. R. Gadekallu, (2023) *"Deep Incomplete Multiview Clustering via Information Bottleneck for Pattern Mining of Data in Extreme-Environment IoT"* **IEEE Internet of Things Journal *11***(16): 26700–26712. DOI: 10.1109/JIOT.2023.3325272.

[4] W. Hu, Y. Wu, and Z. Yang, (2024) *"An analysis of credit risk prediction for small and micro enterprises"* **Journal of Artificial Intelligence Research *1***(2): 1–14. DOI: 10.70891/JAIR.2024.110004.

[5] J. Gao, M. Liu, P. Li, J. Zhang, and Z. Chen, (2024) *"Deep Multiview Adaptive Clustering With Semantic Invariance"* **IEEE Transactions on Neural Networks and Learning Systems *35***(9): 12965–12978. DOI: 10.1109/TNNLS.2023.3265699.

[6] S. Mandia, R. Mitharwal, and K. Singh, (2024) *"Automatic student engagement measurement using machine learning techniques: A literature study of data and methods"* **Multimedia Tools and Applications *83***(16): 49641–49672. DOI: doi.org/10.1007/s11042-023-17534-9.

[7] Q. Luo, S. He, X. Han, Y. Wang, and H. Li, (2024) *"LSTTN: A long-short term transformer-based spatiotemporal neural network for traffic flow forecasting"* **Knowledge-Based Systems *293***: 111637. DOI: 10.1016/j.knosys.2024.111637.

[8] W. Kong, Z. Guo, and Y. Liu. "Spatio-temporal pivotal graph neural networks for traffic flow forecasting". In: *Proceedings of the AAAI conference on artificial intelligence*. ***38***. 8. 2024, 8627–8635. DOI: 10.1609/aaai.v38i8.28707.

[9] Z. Li, J. Zhou, Z. Lin, and T. Zhou, (2024) *"Dynamic spatial aware graph transformer for spatiotemporal traffic flow forecasting"* **Knowledge-based systems *297***: 111946. DOI: 10.1016/j.knosys.2024.111946.

[10] Q. Bing, P. Zhao, C. Ren, X. Wang, and Y. Zhao, (2024) *"Short-term traffic flow forecasting method based on secondary decomposition and conventional neural network–transformer"* **Sustainability *16***(11): 4567. DOI: 10.3390/su16114567.

[11] Z. Fang, Q. Long, G. Song, and K. Xie. "Spatial-temporal graph ode networks for traffic flow forecasting". In: *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*. 2021, 364–373.

[12] G. Huo, Y. Zhang, B. Wang, J. Gao, Y. Hu, and B. Yin, (2023) *"Hierarchical spatio–temporal graph convolutional networks and transformer network for traffic flow forecasting"* **IEEE Transactions on Intelligent Transportation Systems *24***(4): 3855–3867. DOI: 10.1109/TITS.2023.3234512.

[13] S. Lan, Y. Ma, W. Huang, W. Wang, H. Yang, and P. Li. "Dstagnn: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting". In: *International conference on machine learning*. 2022, 11906–11917.

[14] C. Chen, Y. Liu, L. Chen, and C. Zhang, (2022) *"Bidirectional spatial-temporal adaptive transformer for urban traffic flow forecasting"* **IEEE Transactions on Neural Networks and Learning Systems *34***(10): 6913–6925. DOI: 10.1109/TNNLS.2022.3183903.

[15] Y. Zhao, X. Luo, W. Ju, C. Chen, X.-S. Hua, and M. Zhang. "Dynamic hypergraph structure learning for traffic flow forecasting". In: *2023 IEEE 39th International Conference on Data Engineering (ICDE)*. 2023, 2303–2316.

[16] J. Gao, C. Guo, Y. Liu, P. Li, J. Zhang, and M. Liu. "Dynamic-static Feature Fusion with Multi-scale Attention for Continuous Blood Glucose Prediction". In: *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2025, 1–5. DOI: 10.1109/ICASSP49660.2025.10888380.

[17] J. Ji, J. Wang, C. Huang, J. Wu, B. Xu, Z. Wu, J. Zhang, and Y. Zheng. "Spatio-temporal self-supervised learning for traffic flow prediction". In: *Proceedings of the AAAI conference on artificial intelligence*. **37**. 4. 2023, 4356–4364. DOI: 10.1609/aaai.v37i4.25555.

[18] Y. Jiang and S. Yin, (2023) *"Heterogenous-view occluded expression data recognition based on cycle-consistent adversarial network and K-SVD dictionary learning under intelligent cooperative robot environment"* **Computer Science and Information Systems *20***(4): 1869–1883. DOI: 10.2298/CSIS221228034J.

[19] Y. Qi, Z. Men, and S. Xie, (2025) *"Research on Wind and Photovoltaic Power Generation Forecasting Method Based on Digital Twin and Deep Learning"* **Software Engineering *28***(03): 57–63. DOI: 10.19644/j.cnki.issn2096-1472.2025.003.011.