Dynamic Spatial-Temporal Pattern-Enhanced Graph Network for Traffic Forecasting

Zilong Wu, Junwei Yang, Hongyan Mao*
Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, Shanghai, China
{51275902113, 51275902096}@stu.ecnu.edu.cn, hymao@sei.ecnu.edu.cn

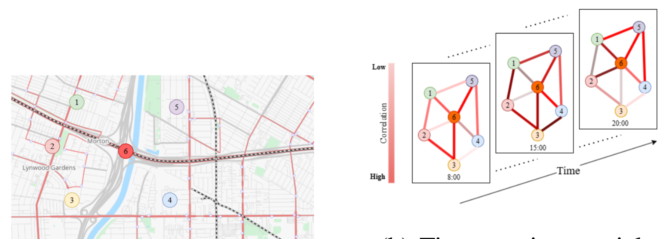
***Abstract*—Traffic prediction remains a critical challenge in urban transportation systems. While existing approaches leveraging Graph Neural Networks (GNNs) and attention mechanisms have demonstrated promising results, two key limitations persist: 1) The dynamic evolution of spatial dependencies with respect to both time and traffic context is often overlooked; 2) There is a lack of explicit modeling of complex traffic conditions and their interactions in forming diverse traffic patterns. To address these challenges, we propose a novel Dynamic Spatial-Temporal Pattern-Enhanced Graph Network (DSTPEGN). To tackle the first limitation, we propose a Time-Aware Graph Convolution Network (TAGCN) that captures dynamic spatial dependencies conditioned on traffic states. For the second limitation, we propose a novel Pattern-Enhanced Graph Convolution Network (PEGCN) to model fundamental spatiotemporal conditions and dynamically retrieve them to represent complex traffic patterns. We also simulate realistic traffic pattern propagation to capture real-world traffic nature. Additionally, we design a self-supervised task to regularize the learned pattern representations. A Multi-Graph Interactive Learning (MGIL) module is proposed to effectively fuse the output of multiple graph convolutions, enabling effective interaction between time-aware spatial correlations and traffic pattern representations. Extensive experiments on three real-world traffic datasets demonstrate that DSTPEGN achieves state-of-the-art performance.**

***Index Terms*—Traffic Prediction, Dynamic Embedding Learning, Traffic Pattern Modeling, Spatial-Temporal Graph, Self-Supervised learning**

I. Introduction

Urban traffic flow prediction is a fundamental task in intelligent transportation systems, aiming to forecast future traffic conditions based on historical observations [1], [2]. Accurate traffic prediction plays a crucial role in traffic management, route planning, and congestion mitigation. With the increasing availability of large-scale traffic data, data-driven approaches have become the dominant paradigm for traffic flow prediction.

In recent years, Spatiotemporal Graph Neural Networks (STGNNs) [3], [4]have emerged as an effective framework for traffic prediction by representing traffic networks as graphs and jointly modeling spatial and temporal dependencies. Existing STGNN-based methods typically integrate graph neural networks with sequential models such as recurrent neural networks or Transformers [5] to capture complex spatial–temporal correlations. These approaches have achieved notable performance improvements compared to traditional statistical and machine learning methods. However, we argue that existing STGNN-based methods still fail to adequately address two critical issues in traffic flow prediction. First, most existing approaches overlook the dynamic evolution of spatial dependencies [6]. As shown in Fig. 1, the spatial correlation between nodes is not static but evolves dynamically over time [7], [8].Existing methods usually rely on static or data-adaptive adjacency matrices that remain fixed across time, which limits their ability to capture time-varying spatial correlations under different traffic conditions. In real-world traffic systems, spatial correlations among nodes evolve dynamically and are closely related to current traffic states. Secondly, existing methods lack explicit modeling of regularizable traffic patterns [9]. As shown in Fig. 2, traffic patterns in urban area are often affected by various conditions [10]–[12] such as weather, accidents, holidays. As a result, it's difficult to model traffic conditions in an explicit manner. Most STGNN-based approaches implicitly encode traffic patterns within hidden representations learned by spatial–temporal encoders [13], [14], without explicitly representing, constraining, or organizing fundamental traffic conditions. As a result, the learned representations are often difficult to interpret and stabilize [15], and fail to capture how basic traffic conditions can be composed and propagated to form diverse traffic patterns under different contexts. In real-world traffic systems, recurring traffic patterns such as congestion or recovery exhibit localized spatial propagation, which are not explicitly modeled by existing methods [16],



(a) Road network of a certain zone.

(b) Time-varying spatial correlation between nodes.

Fig. 1. Visualizing time-varying spatial correlations in road networks. In (a), the real-world map displays road infrastructure and numbered circles indicating nodes in city. In (b), the temporal evolution of spatial correlations among nodes is shown at three representative time points. The color reflects the strength of correlation throughout the day, revealing the time-varying nature of traffic flow interactions.

* Corresponding author

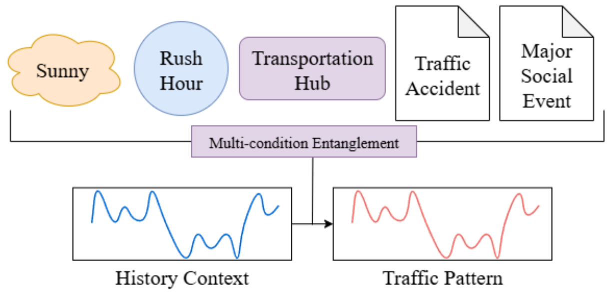


Fig. 2. An example of multiple basic traffic parameters combined with historical context indicate future traffic patterns

[17].

To address these limitations, we propose a novel Dynamic Spatial–Temporal Pattern-Enhanced Graph Network (DST-PEGN) for traffic flow prediction. First, to model dynamically evolving spatial dependencies, we propose a Time-Aware Graph Convolution Network (TAGCN) that adaptively captures spatial correlations conditioned on current traffic states by combining traffic-driven dynamic graphs with a learnable static graph. Second, to explicitly model regularizable traffic patterns, we propose a Pattern-Enhanced Graph Convolution Network (PEGCN) equipped with a Spatial-Temporal Dependency Memory Bank (STDMB). The memory bank stores a set of fundamental traffic pattern prototypes and enables dynamic retrieval of relevant patterns according to traffic context. Moreover, to simulate realistic traffic pattern propagation, the retrieved pattern representations are propagated over the adaptive graph structure, reflecting the locality of traffic pattern diffusion in real-world urban environments. A self-supervised learning task is further designed to regularize the learned pattern representations. A Multi-Graph Interactive Learning (MGIL) module is proposed to enable effective interaction between time-aware spatial correlations and pattern-enhanced representations.

In summary, the main contributions of this work are as follows:

- We propose a spatial-temporal traffic prediction framework that dynamically models spatial dependencies conditioned on traffic states, enabling adaptive capture of time-varying spatial correlations.
- We propose a Pattern-Enhanced Graph Convolution Network with a Spatial-Temporal Dependency Memory Bank to explicitly model regularizable traffic patterns and simulate their realistic spatial propagation, together with a self-supervised task to improve pattern representation quality.
- Extensive experiments on three real-world traffic datasets demonstrate that the proposed framework consistently outperforms state-of-the-art methods, while ablation and case studies further validate the effectiveness of each component.

II. Related Work

Early studies on traffic flow prediction mainly relied on classical statistical methods such as Historical Average (HA) and AutoRegressive Integrated Moving Average (ARIMA). These approaches assume linear relationships and stationary patterns, which limits their ability to model the complex and highly dynamic nature of real-world traffic systems.

With the development of machine learning, nonlinear models including Support Vector Regression (SVR) and Random Forest Regression (RFR) were introduced to improve prediction accuracy. Although these methods can capture nonlinear relationships to some extent, they heavily depend on manual feature engineering and lack the ability to automatically learn hierarchical spatial-temporal representations, which restricts their generalization capability.

In recent years, deep learning methods have significantly advanced traffic prediction. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been widely adopted to model spatial and temporal dependencies, respectively. However, CNN-based models are limited in capturing long-range temporal dependencies, while RNN-based models often struggle to effectively model complex spatial relationships in large-scale traffic networks.

To address these limitations, Spatiotemporal Graph Neural Networks (STGNNs) have emerged as a powerful framework for traffic prediction. Models such as STGCN integrate temporal convolution with graph convolution to jointly learn spatial and temporal correlations. Nevertheless, many STGNN-based methods rely on static or data-adaptive graphs that remain fixed over time, which hinders their ability to capture dynamically evolving spatial dependencies. AGCRN introduces a learnable adaptive adjacency matrix to model intrinsic spatial structures, but the learned graph is still static after training, limiting adaptability under changing traffic conditions.

III. Preliminaries

The road network is modeled as a spatial graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$, where $\mathcal{V}$ denotes the set of nodes with $|\mathcal{V}| = N$, $\mathcal{E}$ represents the set of edges describing spatial connections among nodes, and $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the adjacency matrix encoding the network topology. The spatial graph structure is assumed to remain fixed over time, and $\mathcal{G}$ can be either directed or undirected.

Traffic conditions on the road network are represented as spatiotemporal graph signals. At each time step t, the traffic state of all nodes is denoted by $\mathbf{X}_t \in \mathbb{R}^{N\times D}$, where D is the number of feature channels associated with each node, such as traffic speed and flow. Given T historical observations, the input traffic data are organized as a tensor $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_T\} \in \mathbb{R}^{T\times N\times D}$, where the three dimensions correspond to time, nodes, and features, respectively.

The objective of spatiotemporal traffic forecasting is to learn a mapping function $\mathcal{F}$ that maps a historical traffic sequence $(\mathbf{X}_{t-T+1}, \mathbf{X}_{t-T+2}, \ldots, \mathbf{X}_t)$ to future traffic states $(\hat{\mathbf{X}}_{t+1}, \hat{\mathbf{X}}_{t+2}, \ldots, \hat{\mathbf{X}}_{t+T_f})$, where T denotes the length of the historical input sequence and T_f represents the forecasting

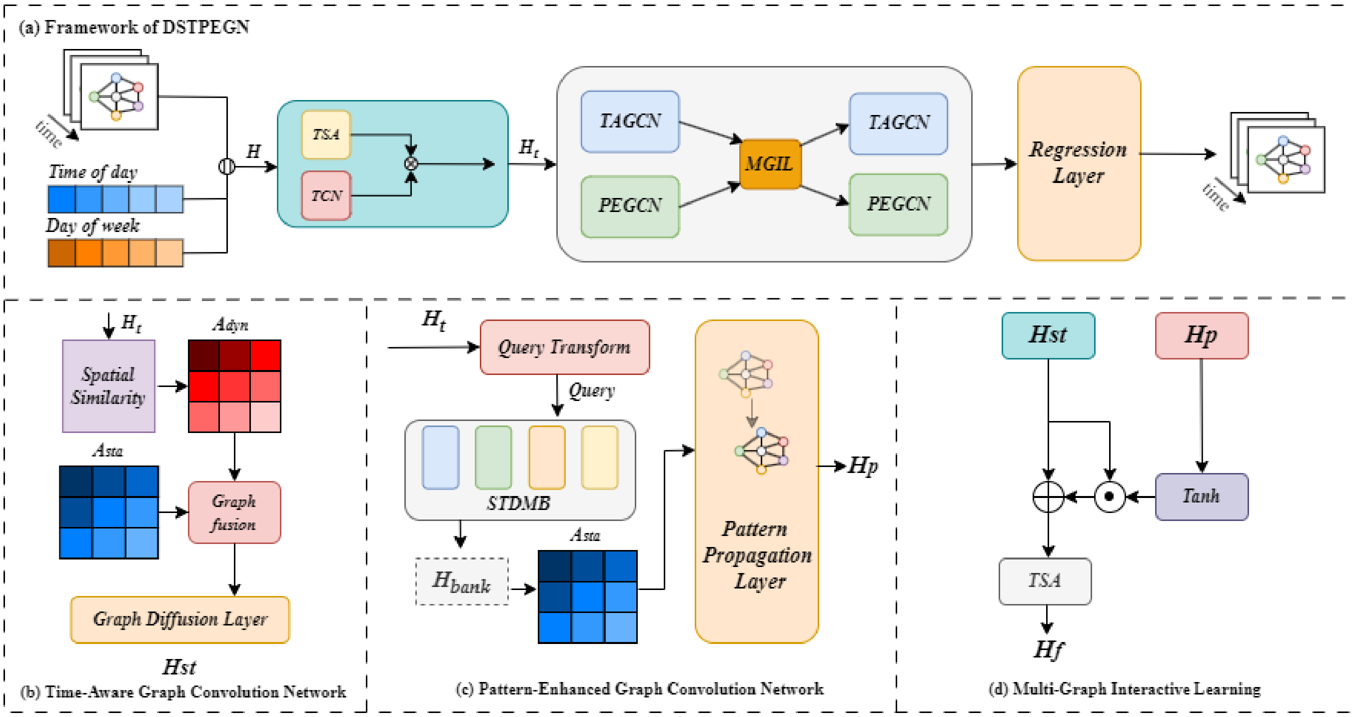


Fig. 3. Detailed framework of DSTPEGN.

horizon, with each predicted traffic state satisfying $\hat{\mathbf{X}}_{t+k} \in \mathbb{R}^{N \times D}$ for $k = 1, 2, \ldots, T_f$.

IV. Methodology

Fig. 3 depicts the detailed framework of the proposed model. The proposed model consists of four main modules. Temporal Dependency Extractor: jointly models long-term and short-term temporal dependencies; Time-Aware Graph Convolution Network (TAGCN): captures time-aware spatial correlations; Pattern-Enhanced Graph Convolution Network (PEGCN): We propose this novel module to explicitly models regularizable traffic patterns and simulate spatial propagation of traffic patterns in real-world urban conditions through a spatial-temporal dependency memory bank, a self-supervised task is also proposed to regularize the pattern representation retrieved from the memory bank; Multi-Graph Interactive Learning (MGIL): We propose this strategy to aggregate spatial representations learned from multiple graphs to generate refined spatial-temporal features.

A. Temporal Dependency Extractor

The Temporal Dependency Extractor employs temporal self-attention and temporal convolution, and models temporal dynamics from both global and local perspectives and fuses them by a gated mechanism.

We first concatenate the weekly embeddings, daily embeddings, and history data and processed by a FC layer to obtain the hidden representation $H \in \mathbb{R}^{T \times N \times C}$.

1) Temporal Self-Attention: The hidden representation $H \in \mathbb{R}^{T \times N \times C}$ is first processed by temporal self-attention to capture long-range temporal dependencies. The process is defined as:

$$H_{ta} = softmax(\frac{Q_t K_t^T}{\sqrt{d}})V_t \tag{1}$$

$$Q_t = W_{taq} H_t, K_t = W_{tak} H_t, V_t = W_{tav} H_t \tag{2}$$

where $W_{taq}, W_{tak}, W_{tav} \in \mathbb{R}^{C \times d}$ are learnable projection matrices.$H_t a$ represents the long-term temporal representation.

2) Temporal Convolution: To capture short-term temporal dynamics, H is processed by a Temporal Convolution Network (TCN) based on dilated causal convolutions. Dilated causal convolution expands the temporal receptive field while preserving temporal causality. The output of temporal convolution is given by:

$$H_{tcn} = H * \theta_d \tag{3}$$

where θ_d denotes the convolution kernel.

To effectively aggregate long-term and short-term temporal information, a temporal dependency fusion mechanism is applied to the outputs of temporal self-attention and temporal convolution. Specifically, the ReLU-activated attention output is modulated by the sigmoid-activated convolution output:

$$H_t = ReLU(H_{ta}) \odot sigmoid(H_{tcn}) \tag{4}$$

B. Time-Aware Graph Convolution Network

The Time-Aware Graph Convolution Network (TAGCN) aims to capture dynamic and long-range spatial dependencies conditioned on current traffic states.

Given the temporal representation $H_t \in \mathbb{R}^{T\times N\times d}$, TAGCN first utilizes a linear layer to obtain dynamic hidden representation H_t^{dyn} and then constructs a dynamic spatial correlation graph by measuring node-wise similarity at each time step. The time-aware dynamic adjacency matrix is computed as:

$$A_{\text{dyn}}[i][j] = \frac{exp(H_t^{dyn}[...,i]\cdot H_t^{dyn\top}[...,j])}{\sum_{j=1}^{N} exp(H_t^{dyn}[...,i]\cdot H_t^{dyn\top}[...,j])}, \quad (5)$$

$$A^{\text{dyn}} = \text{softmax}\left(\frac{\mathbf{H}_t^{\text{dyn}}\left(\mathbf{H}_t^{\text{dyn}}\right)^\top}{\sqrt{d}}\right), \quad (6)$$

where $A_{\text{dyn}} \in \mathbb{R}^{N\times N}$ reflects traffic-conditioned spatial correlations among nodes.

To jointly model static connectivity and dynamic spatial correlation, the dynamic graph A_{dyn} is combined with a learnable static graph A_{sta}. The two graphs are concatenated and transformed by a multi-layer perceptron followed by a softmax operation:

$$A_f = \text{Softmax}\big(\text{MLP}(A_{\text{dyn}}||A_{\text{sta}})\big), \quad (7)$$

where A_sta is a learnable graph, A_f denotes the fused adjacency matrix.

Based on the fused adjacency matrix A_f, TAGCN applies a graph diffusion convolution to aggregate spatial information across multiple hops. The spatial-temporal representation is computed as:

$$H_{st} = \sum_{k=0}^{K}(A_f)^k H_t W_s, \quad (8)$$

where K denotes the maximum diffusion step, W_s is a learnable weight matrix, and $(A_f)^k$ represents the k-th order graph diffusion operator.

C. Pattern-Enhanced Graph Convolution Network

The Pattern-Enhanced Graph Convolution Network (PEGCN) aims to explicitly model regularizable traffic patterns and their spatial propagation that are difficult to capture by conventional graph convolution methods. In this module, we propose a novel Spatial-Temporal Dependency Memory Bank to encode fundamental traffic conditions and provides constrained pattern representations for traffic prediction. And the spatial propagation of traffic patterns is simulated in this module by graph diffusion mechanism.

PEGCN maintains a Spatial-Temporal Dependency Memory Bank (STDMB) that stores a set of P spatial-temporal dependency prototypes:

$$\mathcal{M} = \{M_1, M_2, \ldots, M_P\}, \quad (9)$$

where each memory item represents a basic traffic condition, such as congestion, accident, rainy weather. Given the temporal representation $H_t \in \mathbb{R}^{T\times N\times d}$, PEGCN first applies a linear transformation to obtain the query representation:

$$Q = H_t W_q, \quad (10)$$

TABLE I
DATASET STATISTICS

Dataset	#Nodes	#Edges	#Days	#Data
PEMS03	358	340	59	16992
METR_LA	207	1722	119	34272
PEMS_BAY	325	2694	181	52116

where W_q is a learnable projection matrix.

To retrieve relevant traffic patterns from the memory bank, PEGCN employs a cosine similarity-based matching mechanism. Specifically, the cosine similarity between the query representation and each memory prototype is computed as:

$$s_p = \frac{Q\cdot M_p}{\|Q\|\,\|M_p\|}, \quad p = 1, 2, \ldots, P, \quad (11)$$

where s_p denotes the similarity score between the query and the k-th memory item. The similarity scores are normalized by a softmax function to obtain matching weights:

$$\alpha_p = \frac{\exp(s_p)}{\sum_{j=1}^{P}\exp(s_j)}. \quad (12)$$

Based on the matching weights, the pattern-enhanced representation is computed as a weighted combination of memory prototypes:

$$H_{\text{bank}} = \sum_{p=1}^{P}\alpha_p M_p, \quad (13)$$

where H_{bank} denotes the traffic pattern representation retrieved from the memory bank.

To simulate the propagation of traffic patterns under real-world urban conditions, PEGCN further applies a graph convolution operation to H_{bank} using the static adjacency matrix A_{sta}. The pattern-aware spatial representation is computed as:

$$H_p = \sum_{k=0}^{K}(A_{\text{sta}})^k H_{\text{bank}} W_p, \quad (14)$$

where W_p is a learnable weight matrix and $(A_{\text{sta}})^k$ denotes the k-th order graph diffusion operator based on the static graph structure. This design reflects the realistic assumption that traffic patterns such as congestion and incidents propagate across geographically adjacent regions rather than through dynamically correlated but distant nodes.

To further regularize the learned traffic pattern representations and improve their discriminative capability, we propose a self-supervised node classification task. Specifically, the propagated pattern representation is first processed by a temporal convolution layer:

$$E = \text{TCN}(H_p), \quad (15)$$

followed by a fully connected layer and a softmax function to produce the predicted labels:

$$\hat{y}' = \text{Softmax}(\text{FC}(E)). \quad (16)$$

TABLE II
MODEL PERFORMANCE COMPARISON ON PEMS_BAY AND METR_LA.

Dataset	Models	Horizon 3			Horizon 6			Horizon 12		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
METR_LA	GWNet	2.69	5.15	7.10%	3.12	6.27	8.47%	3.51	7.28	10.19%
	AGCRN	2.87	5.20	7.19%	3.23	6.52	8.51%	3.59	7.46	10.38%
	STNorm	2.81	5.57	7.40%	3.18	6.59	8.47%	3.57	7.51	10.24%
	STID	2.82	5.53	7.75%	3.19	6.57	9.39%	3.55	7.55	10.95%
	DSTAGNN	2.85	5.56	7.73%	3.25	6.59	9.11%	3.63	7.65	10.98%
	PDFormer	2.83	5.45	7.77%	3.20	6.46	9.19%	3.62	7.47	10.91%
	MHGNet	2.80	5.52	7.58%	3.18	6.62	9.26%	3.54	7.56	10.75%
	DSTPEGN	**2.61**	**5.09**	**6.84%**	**3.04**	**6.15**	**8.31%**	**3.35**	**7.19**	**9.88%**
PEMS_BAY	GWNet	1.30	2.73	2.74%	1.63	3.73	3.73%	1.99	4.60	4.71%
	AGCRN	1.35	2.88	2.91%	1.67	3.82	3.81%	1.94	4.50	4.55%
	STNorm	1.33	2.82	2.76%	1.65	3.77	3.66%	1.92	4.45	4.46%
	STID	1.31	2.79	2.78%	1.64	3.73	3.73%	1.91	4.42	4.55%
	DSTAGNN	1.41	2.94	3.01%	1.76	3.88	3.87%	2.09	4.67	4.74%
	PDFormer	1.32	2.83	2.78%	1.64	3.79	3.71%	1.91	4.43	4.51%
	MHGNet	1.31	2.74	2.75%	1.62	3.65	3.62%	1.90	4.36	4.43%
	DSTPEGN	**1.21**	**2.62**	**2.63%**	**1.49**	**3.57**	**3.60%**	**1.83**	**4.28**	**4.31%**

The self-supervised classification loss is defined as:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{N}\sum_{i=1}^{N} y_i' \log(\hat{y}_i'), \tag{17}$$

where y' denotes node-level pseudo labels that treat each node as a distinct entity, encouraging discriminative and well-regularized traffic pattern representations.

D. Multi-Graph Interactive Learning

The Multi-Graph Interactive Learning (MGIL) module aims to enable effective interaction between time-aware spatial correlations and traffic patterns. MGIL explicitly models the mutual influence between representations learned from different graphs and enhances spatial-temporal feature learning through interactive modulation.

Given the time-aware representation H_{st} obtained from TAGCN and the pattern-enhanced representation H_p learned by PEGCN, MGIL first performs interactive feature modulation. The pattern representation is nonlinearly transformed and used to modulate the time-aware features:

$$\tilde{H} = H_{st} + H_{st} \odot \tanh(H_p), \tag{18}$$

where $\odot$ denotes element-wise multiplication. This modulation mechanism allows regularizable traffic patterns to dynamically adjust time-aware representations learned from time-aware graphs.

The interactively modulated representation $\tilde{H}$ is further refined by a temporal self-attention module to model temporal interactions after spatial-pattern coupling:

$$H_f = \mathrm{TSA}(\tilde{H}), \tag{19}$$

where H_f denotes the refined spatial-temporal representation.

The refined representation H_f is fed back into both TAGCN and PEGCN for further processing. The resulting outputs are concatenated and projected through a linear transformation:

$$H_r = \mathrm{FC}\big(\text{TAGCN}(H_f)||\text{PEGCN}(H_f)\big), \tag{20}$$

where H_r represents the aggregated representation after multi-graph interaction.H_r is processed by as regression layer to produce predictions $\hat{Y}$.

E. Model Training

We jointly optimize the forecasting objective and the auxiliary self-supervised node-discrimination objective. The overall loss is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{pred}} + \lambda\,\mathcal{L}_{\text{cls}}, \tag{21}$$

where λ is a hyper-parameter balancing the contribution of the auxiliary task.Given the ground-truth traffic values $\mathbf{Y}$ and predictions $\hat{\mathbf{Y}}$, we adopt mean absolute error (MAE) as the forecasting loss.

V. EXPERIMENTS

A. Datasets

We evaluate our model on three real-world traffic datasets from the CalTrans Performance Measurement System (PeMS), including PEMS_BAY, METR_LA, and PEMS03. All datasets are aggregated at 5-minute intervals, and predefined spatial graphs are constructed based on the road network topology. Detailed dataset statistics are reported in Table I. Following common practice, we use the past 60 minutes (12 time steps) of traffic flow to predict traffic conditions for the next 60 minutes.

TABLE III
MODEL PERFORMANCE COMPARISON ON PEMS03.

Dataset	Models	MAE	RMSE	MAPE
PEMS03	GWNet	14.59	25.24	15.52
	AGCRN	15.27	26.65	15.89
	STNorm	15.32	25.93	14.37
	STID	15.33	27.40	16.40
	DSTAGNN	15.57	27.21	14.68
	PDFormer	14.94	25.39	15.82
	MHGNet	15.30	26.58	15.12
	DSTPEGN	**13.71**	**24.39**	**14.11**

B. Baseline Methods

DSTPEGN's performance is compared against the following baselines: GWNet [18], AGCRN [19], STNorm [20], STID [21], PDFormer [22], DSTAGNN [23], MHGNet [24].

C. Experiment Settings

The experiments are conducted on a system equipped with an NVIDIA GeForce RTX4090 GPU. The datasets PEMS_BAY, METR_LA, and PEMS03 are divided into training, validation, and test sets in a ratio of 6:2:2. The models are trained using the Adam optimization algorithm with a weight decay (wdecay) of 1.0×10^{-5} and an epsilon(eps) of 1.0×10^{-8}.We set the hidden dimension of the hidden representation as $d = 48$ and $C = 24$. The diffusion step is set to $K = 3$, and the memory bank size is set to $P = 64$. The whole model are implemented with 3 stacked layers, and each layer adopts a temporal dependency extractor, a TAGCN, a PEGCN and a MGIL module.

D. Experiment Results

As shown in Table II and Table III, our method achieves better performance across most metrics on all three datasets. We can draw the following conclusions: Compared to GNN based models (such as GWNet and AGCRN), DSTPEGN improves performance through learning dynamic long-range spatial dependency captured by attention mechanism. Compared to embedding-based models like STID, it learns spatiotemporal correlation that better reflect the intrinsic nature of traffic data. Additionally, when compared to attention-based models (like PDFormer), it exhibits better ability to explicitly learn more stable and constrained traffic pattern.

E. Ablation Study

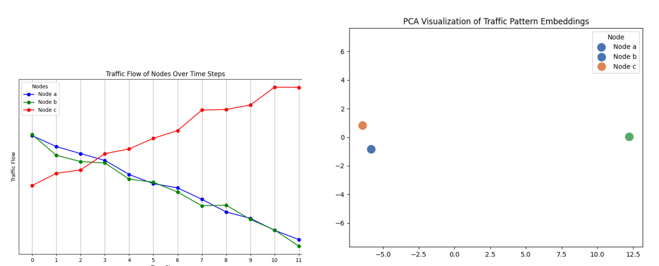
To evaluate the contribution of each component in the proposed DSTPEGN model, we conduct an ablation study on the PEMS_BAY dataset. The results are summarized in Table IV. The following variants are tested:

- **w/o tf**: This variant removes the temporal fusion mechanism and models temporal dependencies with a single branch. The performance drop shows the necessity of jointly modeling long- and short-term temporal dependencies.
- **w/o TAGCN**: This variant removes the Time-Aware Graph Convolution Network and relies only on a static graph. The degraded performance highlights the importance of time-aware spatial dependency modeling.

TABLE IV
ABLATION STUDY ON PEMS_BAY.

Dataset	Variant	MAE	RMSE	MAPE (%)
PEMS_BAY	w/o tf	1.92	3.92	3.86
	w/o TAGCN	1.79	3.91	3.79
	w/o PEGCN	1.67	3.78	3.66
	w/o MGIL	1.63	3.71	3.62
	w/ A_f	1.60	3.69	3.53
	w/o ssl	1.46	3.40	3.49
	DSTPEGN	**1.43**	**3.36**	**3.34**
PEMS03	w/o tf	20.02	31.15	21.90
	w/o TAGCN	18.12	29.13	20.16
	w/o PEGCN	14.94	25.42	15.93
	w/ MGIL	14.79	25.38	15.51
	w/ A_f	14.39	25.36	15.30
	w/o ssl	13.92	25.21	14.19
	DSTPEGN	**13.71**	**24.39**	**14.11**

- **w/o PEGCN**: This variant removes the Pattern-Enhanced Graph Convolution Network. The inferior results indicate that explicit modeling of regularizable traffic patterns is critical for accurate prediction.
- **w/ A_f**: This variant replaces the static graph A_{sta} with the fused dynamic graph A_f in PEGCN. The degraded performance suggests that traffic patterns tend to propagate locally across physically adjacent regions, validating the use of the static graph for pattern propagation.
- **w/o ssl**: This variant removes the self-supervised learning task. The performance decline indicates that the self-supervised objective provides effective regularization for learning stable pattern representations.
- **w/o MGIL**: This variant removes the Multi-Graph Interactive Learning module and directly concatenates H_{st} and H_p. The inferior performance demonstrates the importance of interactive coordination between spatial correlations and traffic patterns.



(a) Traffic flow (b) traffic pattern representation

Fig. 4. Visualization of traffic flows and retrieved traffic bank embeddings of three nodes in PEMS03.

F. Case Study: Visualization of Learned Traffic Patterns

To qualitatively evaluate whether the proposed Spatial-Temporal Dependency Memory Bank (STDMB) captures meaningful traffic patterns, we conduct a case study on the three traffic nodes selected from PEMS03 dataset. Fig. 4(a) illustrates the traffic flow evolution of three representative nodes over time, while Fig. 4(b) visualizes the corresponding pattern representations retrieved from STDMB using PCA.

As shown in Fig. 4(a), Node a and Node b exhibit similar decreasing traffic trends, whereas Node c demonstrates a distinct increasing pattern, indicating fundamentally different traffic dynamics. Fig. 4(b) shows that the pattern representations retrieved from STDMB (H_{bank}) preserve this structural similarity, where Node a and Node b are mapped to nearby positions in the embedding space, while Node c is clearly separated.

This observation suggests that STDMB is able to capture regularizable traffic patterns and map nodes with similar temporal evolution into consistent representations. In contrast, nodes with distinct traffic dynamics are assigned to different regions in the pattern embedding space, demonstrating the effectiveness of the proposed memory-based pattern modeling mechanism.

VI. Conclusion

In this paper, we propose DSTPEGN, a spatial–temporal traffic prediction framework that jointly models long-term and short-term temporal dependencies, time-aware spatial correlations, and regularizable traffic patterns. By proposing a time-aware graph convolution network, a pattern-enhanced graph convolution network that simulates the propagation of traffic patterns over physically adjacent regions, and a multi-graph interactive learning mechanism, DSTPEGN effectively captures dynamic spatial interactions and structured traffic behaviors. Extensive experiments demonstrate that DSTPEGN consistently outperforms state-of-the-art methods, while ablation and case studies verify the effectiveness of each component.

References

[1] Maryam Shaygan, Collin Meese, Wanxin Li, Xiaoliang George Zhao, and Mark Nejad, "Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities," *Transportation research part C: emerging technologies*, vol. 145, pp. 103921, 2022.

[2] David Alexander Tedjopurnomo, Zhifeng Bao, Baihua Zheng, Farhana Murtaza Choudhury, and A Kai Qin, "A survey on modern deep neural network for traffic prediction: Trends, methods and challenges," in *2023 IEEE 39th International Conference on Data Engineering (ICDE)*. IEEE, 2023, pp. 3795–3796.

[3] Chao Song, Youfang Lin, Shengnan Guo, and Huaiyu Wan, "Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, 2020, vol. 34, pp. 914–921.

[4] Chuanpan Zheng, Xiaoliang Fan, Cheng Wang, and Jianzhong Qi, "Gman: A graph multi-attention network for traffic prediction," in *Proceedings of the AAAI conference on artificial intelligence*, 2020, vol. 34, pp. 1234–1241.

[5] Di Jin, Jiayi Shi, Rui Wang, Yawen Li, Yuxiao Huang, and Yu-Bin Yang, "Trafformer: Unify time and space in traffic prediction," in *Proceedings of the AAAI conference on artificial intelligence*, 2023, vol. 37, pp. 8114–8122.

[6] Huaxiu Yao, Fei Wu, Jintao Ke, Xianfeng Tang, Yitian Jia, Siyu Lu, Pinghua Gong, Jieping Ye, and Zhenhui Li, "Deep multi-view spatial-temporal network for taxi demand prediction," in *Proceedings of the AAAI conference on artificial intelligence*, 2018, vol. 32.

[7] Kihwan Kim, Seungmin Jin, Sungahn Ko, and Jaegul Choo, "Stgrat: A spatio-temporal graph attention network for traffic forecasting," in *International conference on information and knowledge management*, 2020.

[8] Fuxian Li, Jie Feng, Huan Yan, Guangyin Jin, Fan Yang, Funing Sun, Depeng Jin, and Yong Li, "Dynamic graph convolutional recurrent network for traffic prediction: Benchmark and solution," *ACM Transactions on Knowledge Discovery from Data*, vol. 17, no. 1, pp. 1–21, 2023.

[9] Xiaoyang Wang, Yao Ma, Yiqi Wang, Wei Jin, Xin Wang, Jiliang Tang, Caiyan Jia, and Jian Yu, "Traffic flow prediction via spatial temporal graph neural network," in *Proceedings of the web conference 2020*, 2020, pp. 1082–1092.

[10] Zhihui Lin, Maomao Li, Zhuobin Zheng, Yangyang Cheng, and Chun Yuan, "Self-attention convlstm for spatiotemporal prediction," in *Proceedings of the AAAI conference on artificial intelligence*, 2020, vol. 34, pp. 11531–11538.

[11] Zhidan Liu, Zhenjiang Li, Kaishun Wu, and Mo Li, "Urban traffic prediction from mobility data using deep learning," *Ieee network*, vol. 32, no. 4, pp. 40–46, 2018.

[12] Huaxiu Yao, Xianfeng Tang, Hua Wei, Guanjie Zheng, Yanwei Yu, and Zhenhui Li, "Modeling spatial-temporal dynamics for traffic prediction," *arXiv preprint arXiv:1803.01254*, vol. 1, no. 9, 2018.

[13] Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, 2019, vol. 33, pp. 922–929.

[14] Lu Lin, Jianxin Li, Feng Chen, Jieping Ye, and Jinpeng Huai, "Road traffic speed prediction: A probabilistic model fusing multi-source data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 7, pp. 1310–1323, 2017.

[15] Huaxiu Yao, Xianfeng Tang, Hua Wei, Guanjie Zheng, Yanwei Yu, and Zhenhui Li, "Modeling spatial-temporal dynamics for traffic prediction," *arXiv preprint arXiv:1803.01254*, vol. 1, no. 9, 2018.

[16] Jingrui He, Wei Shen, Phani Divakaruni, Laura Wynter, and Rick Lawrence, "Improving traffic prediction with tweet semantics.," in *IJCAI*, 2013, pp. 1387–1393.

[17] Kan Guo, Yongli Hu, Zhen Qian, Hao Liu, Ke Zhang, Yanfeng Sun, Junbin Gao, and Baocai Yin, "Optimized graph convolution recurrent neural network for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 2, pp. 1138–1149, 2020.

[18] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang, "Graph wavenet for deep spatial-temporal graph modeling," *arXiv preprint arXiv:1906.00121*, 2019.

[19] Lei Bai, Lina Yao, Can Li, Xianzhi Wang, and Can Wang, "Adaptive graph convolutional recurrent network for traffic forecasting," *Advances in neural information processing systems*, vol. 33, pp. 17804–17815, 2020.

[20] Jinliang Deng, Xiusi Chen, Renhe Jiang, Xuan Song, and Ivor W Tsang, "St-norm: Spatial and temporal normalization for multi-variate time series forecasting," in *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 2021, pp. 269–278.

[21] Zezhi Shao, Zhao Zhang, Fei Wang, Wei Wei, and Yongjun Xu, "Spatial-temporal identity: A simple yet effective baseline for multivariate time series forecasting," in *Proceedings of the 31st ACM international conference on information & knowledge management*, 2022, pp. 4454–4458.

[22] Jiawei Jiang, Chengkai Han, Wayne Xin Zhao, and Jingyuan Wang, "Pdformer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction," in *Proceedings of the AAAI conference on artificial intelligence*, 2023, vol. 37, pp. 4365–4373.

[23] Shiyong Lan, Yitong Ma, Weikang Huang, Wenwu Wang, Hongyu Yang, and Pyang Li, "Dstagnn: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting," in *International conference on machine learning*. PMLR, 2022, pp. 11906–11917.

[24] Mei Wu, Yiqian Lin, Tianfan Jiang, and Wenchao Weng, "Mhgnet: Multi-heterogeneous graph neural network for traffic prediction," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5.