# WTI Futures Price Forecasting Based on Multi-Graph Fusion Spatiotemporal Attention Network

Junke Huang[a*], Hui Qu[b]

School of Management and Engineering, Nanjing University, Nanjing, China

[a*] 1287171434@qq.com, [b] linda59qu@nju.edu.cn

**Abstract.** This study develops a Multi-Graph Fusion Spatiotemporal Attention Network (MG-STAN) to better capture the evolving interactions between crude oil markets and related financial systems. The proposed framework incorporates temporal embeddings, spatial attention modules, and a multi-graph structure to reflect diverse inter-market relationships. Using a dataset covering 2011–2024 that includes commodity futures, supply-demand factors, and financial indicators, our proposed MG-STAN models consistently and significantly outperform conventional deep learning models. Notably, a three-graph fusion strategy—combining correlation, K-nearest neighbor and dynamic time warping graphs—achieves the best results, suggesting that selectively integrating heterogeneous graphs can enhance forecasting accuracy. The findings underscore the value of multi-graph designs and attention mechanisms in modeling market complexity, and offer new perspectives for price forecasting and energy finance research.

**Keywords:** Crude oil futures, Price forecast, Spatiotemporal graph neural network, Multi-graph fusion

## 1 Introduction

Crude oil is pivotal to global energy systems, with price fluctuations significantly affecting economic stability. The futures market serves as a key venue for price discovery and risk management[1]. This study focuses on enhancing crude oil futures price forecasting precision through advanced spatiotemporal modeling.

Forecasting methodologies have evolved from linear econometric models (e.g., ARIMA, GARCH)[2,3], limited by non-stationarity and nonlinearity, to deep learning approaches. ANNs, RNN variants (LSTM, GRU), and CNNs capture temporal patterns and multi-scale features[4,5]. GNNs model inter-market relationships via graph representations citep{GNN1}, with spatiotemporal variants (e.g., STGCNs) jointly encoding spatial and temporal dependencies[6]. However, financial markets lack intrinsic graph structures. While static graphs rely on predefined relationships, adaptive methods (e.g., GAT) learn dynamic connections[7,8]. Multi-graph frameworks integrate diverse topologies to better represent complex spatial dependencies, showing promise in financial forecasting[9,10].



D. Magni et al. (eds.), *Proceedings of the 2026 3rd International Conference on Applied Economics, Management Science and Social Development (AEMSS 2026)*, Advances in Economics, Business and Management Research 389,
https://doi.org/10.2991/978-94-6239-672-2_35

To address these challenges, we propose MG-STAN (Multi-Graph Fusion Gated Recurrent Spatial Attention Network), comprising: (1) a temporal embedding module for short- and long-term dependencies; (2) a spatial attention module using GAT and Transformers; (3) a multi-graph fusion module integrating heterogeneous graph structures; and (4) a convolution-based prediction module. This work advances GNN-based spatiotemporal modeling for crude oil markets.

Main contributions:

(1) Proposes MG-STAN integrating attention mechanisms and multi-graph fusion for precise crude oil price forecasting.

(2) Constructs multiple graph structures to capture diverse market relationships from complementary perspectives.

(3) Designs an effective fusion layer that integrates multi-graph information, significantly improving prediction accuracy over single-graph baselines.

The paper is organized as follows: Section 2 details the MG-STAN framework; Section 3 describes the dataset and experimental setup; Section 4 presents results and analysis; Section 5 concludes with future directions.

## 2 Proposed Method

In Section 2.1, we will first outline the problem we aim to address along with its symbolic definitions. Section 2.2 will provide a detailed description of the four graph construction methods we employ and the rationale behind their selection. Section 2.3 will give a brief overview of our model architecture, followed by an in-depth explanation of the computational processes involved in each component of the model.

### 2.1 Problem description

We model the crude oil market and $N-1$ related series as nodes in an undirected graph. Each node has associated short-term and long-term time series, denoted$X_{i,\text{short}}$and$X_{i,\text{long}}$. The inter-market dependencies are represented by four adjacency matrices $A_1, A_2, A_3, A_4$, where $A_i \in \mathbb{R}^{N\times N}$. The goal is to predict the crude oil futures price $\hat{y}$:

$$\hat{y} = F\left(\left[\left(X_{1,\text{short}}, X_{1,\text{long}}\right) \ldots \left(X_{N,\text{short}}, X_{N,\text{long}}\right)\right]; A_1; A_2; A_3; A_4\right) \tag{1}$$

Where $F$ is our predictive model, trained to minimize the error between $\hat{y}$ and the true value.

### 2.2 Graph construction

We construct four adjacency matrices $A_1, A_2, A_3, A_4$ based on different relational measures between the node time series.

**2.2.1 Correlation Coefficient Matrix.**

The adjacency matrix $A_1$ uses absolute Pearson correlation to measure linear dependence between nodes. For nodes$V_i = \{v_{i,1}, v_{i,2} \dots v_{i,T}\}$ and $V_j = \{v_{j,1}, v_{j,2} \dots v_{j,T}\}$ with time series of length $T$, the weight $a_{i,j}^1$ is:

$$a_{i,j}^1 = \frac{|\Sigma_{k=1}^T (v_{i,k} - \bar{v}_i)(v_{j,k} - \bar{v}_j)|}{\sqrt{\Sigma_{k=1}^T (v_{i,k} - \bar{v}_i)^2} \sqrt{\Sigma_{k=1}^T (v_{j,k} - \bar{v}_j)^2}} \tag{2}$$

Where $v_{\mathrm{i}}$, and $v_{\mathrm{i}}$ are the mean values of the respective sequences.

**2.2.2 Granger Causality Matrix.**

Matrix $A_2$ is built using Granger causality tests. A binary connection is established if a significant causal relationship $p_{\mathrm{value}} = 0.1$ exists in either direction between two nodes:

$$a_{i,j}^2 = \begin{cases} 1, & \text{if the causality test between } V_i, V_j \text{ is significant in either direction} \\ 0, & \text{else} \end{cases} \tag{3}$$

Here, $a_{i,j}^2 = 1$ denotes a significant causal link.

**2.2.3 K-Nearest Neighbor Matrix.**

Matrix $A_3$ is constructed via the KNN algorithm. The Euclidean distance is first computed. Each node is then connected to its $K$ nearest neighbors:

$$a_{i,j}^3 = \begin{cases} 1 & \text{if } V_j \text{ is among the } K \text{ nearest neighbors of } V_i \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

The value $K$ is determined through experimental tuning.

**2.2.4 Dynamic Time Warping Matrix.**

Matrix $A_4$ uses Dynamic Time Warping (DTW) distance. The normalized DTW cost between sequences $V_i$ and $V_j$ is calculated. A symmetric connection is formed if each node is among the other's $K$ smallest-DTW neighbors:

$$a_{i,j}^4 = \begin{cases} 1 & \text{if } DTW(V_i, V_j) \text{ is among the } K \text{ smallest DTW distances} \\ 0 & \text{otherwise} \end{cases}$$

## 2.3 The Parameter $K$controls the Sparsity of $A_4$.Model Architecture

To capture spatiotemporal dependencies for crude oil price prediction, we propose the Multi-Graph Fusion Spatial-Temporal Attention Network (MG-STAN). As shown in Fig. 1, the model consists of four synergistic modules: 1) a Temporal Embedding Module that encodes multi-scale lag information; 2) a Spatial Attention Module that models dynamic inter-variable correlations; 3) a Multi-Graph Fusion Module that integrates

features from diverse graph structures; and 4) a Prediction Module that outputs the final forecast.

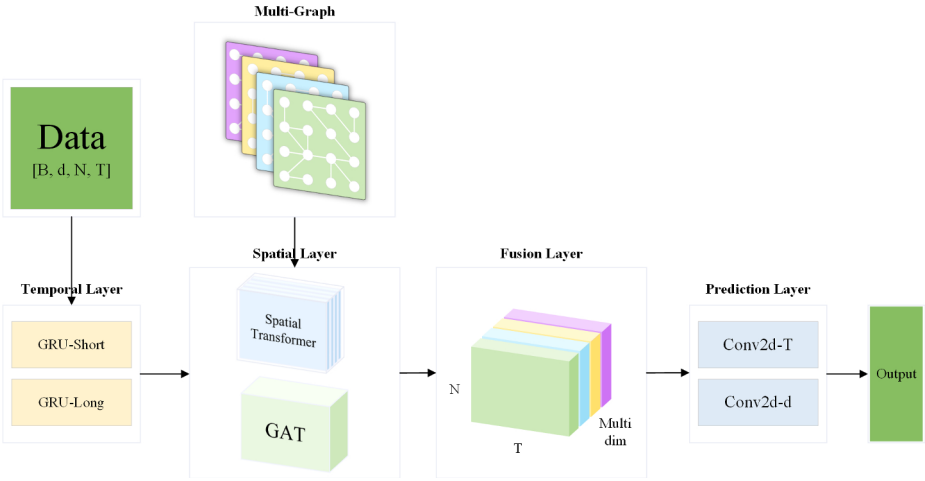


**Fig. 1.** The architecture of the proposed MG-STAN

### 2.3.1 Temporal Module.

To model multi-scale temporal patterns, we employ two parallel Gated Recurrent Unit (GRU) sub-modules. The first processes high-frequency, short-term sequences (e.g., daily data), while the second handles low-frequency, long-term sequences (e.g., monthly data). Each GRU captures temporal dependencies within its respective scale through its gating mechanisms. The final outputs from both sub-modules are combined via a broadcasting sum, fusing immediate patterns with long-term trends into a unified temporal representation.

### 2.3.2 Spatial Module.

This module dynamically models the correlations between different market variables (nodes). It integrates a Graph Attention Network (GAT) and a Spatial Transformer. The GAT assigns adaptive attention weights to a node's neighbors, focusing on the most influential connections. The Spatial Transformer provides a complementary mechanism for learning broader spatial relationships. A gating mechanism dynamically blends the outputs from these two components, allowing the model to balance and utilize both local attention and global transformation features effectively.

### 2.3.3 Fusion Module.

The fusion module integrates spatial features extracted from the different graph structures defined in Section 2.2. Instead of simple averaging, we concatenate the feature tensors from each graph along the feature dimension. This strategy preserves the

unique relational information captured by each graph type (e.g., causality, similarity) and provides a richer, multi-perspective feature set for the final prediction stage.

#### 2.3.4 Prediction Module.

The final module takes the fused multi-graph features and applies two consecutive one-dimensional convolutional layers. These layers progressively compress the combined temporal and feature dimensions, ultimately generating the predicted price value for the target crude oil futures node. This design efficiently distills the complex integrated information into a precise forecast.

## 3 Experimental Results and Analysis

### 3.1 Comparison with Baseline Deep Learning Models

We compare the out-of-sample performance of our proposed Multi-Graph Fusion Spatial-Temporal Attention Network (MG-STAN) with five baseline models: FNN, CNN, GCN, RNN, and GRU. Fig. 2 visualizes the predictions, showing that the four MG-STAN variants closely track actual prices, whereas baseline models exhibit larger deviations during volatile periods.

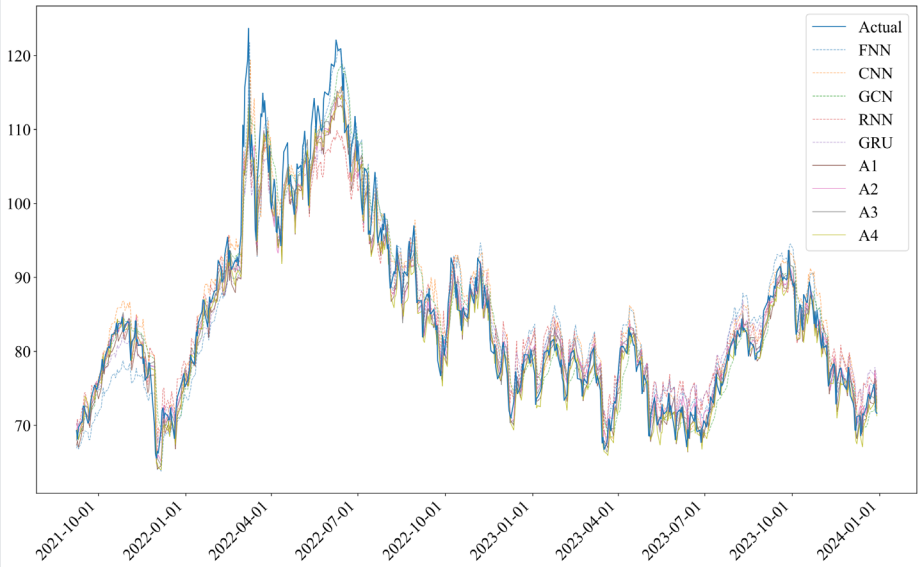


**Fig. 2.** WTI price prediction results.

Table 1 reports the average prediction errors. The MG-STAN models achieve significantly lower errors across all metrics than the baselines. Diebold-Mariano tests against the best benchmark (GCN) confirm this superiority is statistically significant at the 1% level.

We further apply the Model Confidence Set (MCS) test; results are in Table 2. Only the four MG-STAN models are consistently retained in the confidence set (p-values > 0.05) for all error metrics, demonstrating robust superiority. Among them, MG-STAN-A2 (using the Granger causality graph) shows the best performance, with p-values of 1.0000.

**Table 1.** Out-of-sample average losses of different models.

| Model | MSE | RMSE | MAE | MAPE (%) |
|---|---|---|---|---|
| FNN | 12.5550 | 3.5433 | 2.8998 | 3.4771 |
| CNN | 13.1329 | 3.6239 | 2.7604 | 3.2393 |
| GCN | 11.3928 | 3.3753 | 2.5269 | 2.9574 |
| RNN | 12.0227 | 3.4674 | 2.5900 | 2.9838 |
| GRU | 11.4928 | 3.3901 | 2.6264 | 3.0736 |
| MG-STAN-A1 | 8.7728*** | 2.9619*** | 2.2036*** | 2.5088*** |
| MG-STAN-A2 | 8.2111*** | 2.8655*** | 2.0677*** | 2.3362*** |
| MG-STAN-A3 | 8.2512*** | 2.8725*** | 2.1303*** | 2.4227*** |
| MG-STAN-A4 | 8.4740*** | 2.9110*** | 2.1732*** | 2.4859*** |

Note: ***, **, * indicate that the two-sided DM test against the GCN model is statistically significant at the 1%, 5%, and 10% levels, respectively.


**Fig. 3.** Cumulative error plots of WTI price forecasts.

Fig. 3 plots the cumulative error difference between each MG-STAN model and the GCN benchmark. Model A2 (Granger) maintains the most stable and largest error reduction. Models A3 (KNN) and A4 (DTW) also show consistent gains. Model A1 (Correlation) exhibits less stable advantages, particularly during high-volatility periods (e.g., the 2022 energy crisis), likely because correlation graphs capture only symmetric linear relationships, which are less effective for modeling directional dependencies in turbulent markets.

**Table 2.** Out-of-sample MCS test p-values of different models.

| Model | MSE | RMSE | MAE | MAPE (%) |
|---|---|---|---|---|
| FNN | 0.0087 | 0.0102 | 0.0010 | 0.0013 |
| CNN | 0.0032 | 0.0047 | 0.0004 | 0.0013 |
| GCN | 0.0565 | 0.0571 | 0.0063 | 0.0032 |
| RNN | 0.0143 | 0.0139 | 0.0056 | 0.0032 |
| GRU | 0.0370 | 0.0380 | 0.0056 | 0.0032 |
| MG-STAN-A1 | 0.2181 | 0.2204 | 0.0729 | 0.0967 |
| MG-STAN-A2 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| MG-STAN-A3 | 0.9343 | 0.9345 | 0.3772 | 0.2889 |
| MG-STAN-A4 | 0.2181 | 0.2204 | 0.0729 | 0.0967 |

## 3.2 Comparison of Multi-Graph Models

This section evaluates the impact of different graph structures on model performance by comparing four single-graph variants of MG-STAN (A1: similarity, A2: Granger causality, A3: KNN, A4: DTW) and a baseline with a fully connected graph (A0). As shown in Fig. 4, all specifically constructed graphs (A1-A4) outperform the fully connected baseline (A0) across all error metrics (MSE, RMSE, MAE, MAPE), with average reductions of 11.29%, 11.51%, 8.50%, and 11.87%, respectively. This confirms the importance of deliberate graph design.

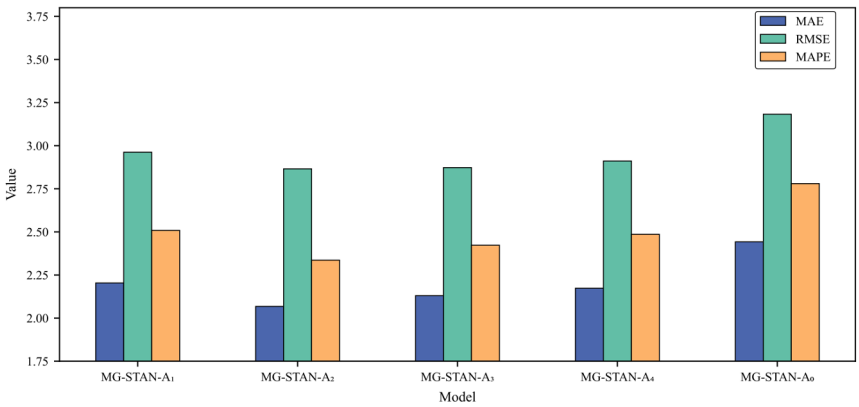


**Fig. 4.** Comparison results of MAE, RMSE and MAPE from different single-graph models.

Among single-graph models, A2 (Granger causality) performs best (e.g., RMSE: 2.8655, MAE: 2.0677), validating the importance of causal relationships. Models A3 (KNN) and A4 (DTW) show robust performance, while A1 (similarity) is weaker, likely due to noise from dense correlation-based connections.

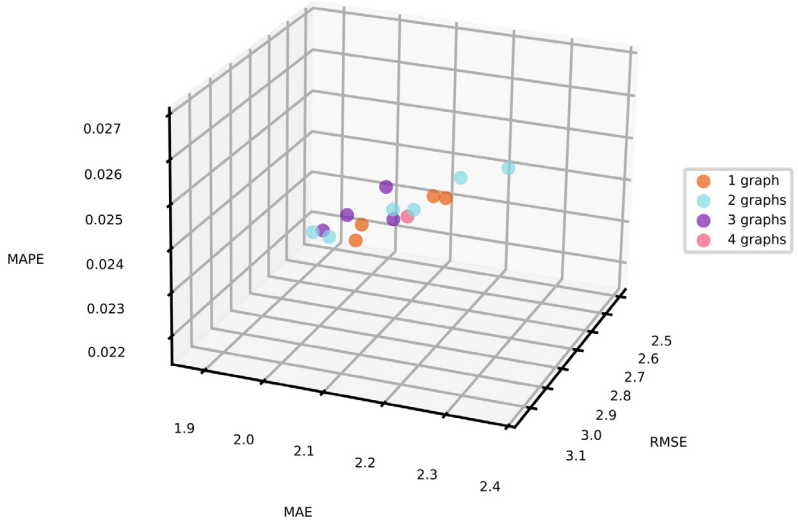


**Fig. 5.** Comparison results from using different numbers of graphs and the combination methods.

We further examine multi-graph fusion by combining different graphs. Table 3 shows that a ternary-graph configuration significantly outperforms unary and binary models. Fig. 5 illustrates that model performance generally improves with more graphs, peaking with the ternary combination A134 (MSE: 6.20, RMSE: 2.49). The MCS test in Table 4 confirms A134's statistical superiority (p-values: 1.0000).

**Table 3.** Out-of-sample average losses of different STAN models.

| Model | MSE | RMSE | MAE | MAPE (%) |
|---|---|---|---|---|
| MG-STAN-A12 | 7.6008 | 2.7570 | 2.0818 | 2.3727 |
| MG-STAN-A13 | 6.5056 | 2.5506 | 1.8641 | 2.1945 |
| MG-STAN-A14 | 7.4883 | 2.7365 | 2.0592 | 2.3636 |
| MG-STAN-A23 | 8.7651 | 2.9606 | 2.2505 | 2.5596 |
| MG-STAN-A24 | 8.1693 | 2.8582 | 2.1246 | 2.4247 |
| MG-STAN-A34 | 6.2524 | 2.5005 | 1.8771 | 2.1642 |
| MG-STAN-A123 | 7.3853 | 2.7176 | 2.0451 | 2.4099 |
| MG-STAN-A124 | 8.3601 | 2.8914 | 2.1555 | 2.4400 |
| MG-STAN-A134 | 6.2000 | 2.4900 | 1.8613 | 2.1716 |
| MG-STAN-A234 | 6.7512 | 2.5983 | 1.9399 | 2.2706 |
| MG-STAN-A1234 | 8.0369 | 2.8349 | 2.1099 | 2.3975 |

However, performance scaling is non-monotonic. Some binary/ternary combinations (e.g., A23, A124) underperform, and the quaternary model (A1234) yields no significant gain. This suggests: (1) information saturation occurs beyond three graphs, (2) some graph combinations introduce feature interference, and (3) optimal performance depends on selecting complementary graph types, not merely increasing their number.

In summary, multi-graph fusion effectively enhances prediction, but requires careful selection of graph types and quantities to balance diversity and complementarity.

**Table 4.** Model Confidence Set (MCS) Test Results

| Model | MSE | RMSE | MAE | MAPE (%) |
|---|---|---|---|---|
| MG-STAN-A12 | 0.2051 | 0.2049 | 0.3645 | 0.4839 |
| MG-STAN-A13 | 0.4343 | 0.4236 | 0.9435 | 0.7844 |
| MG-STAN-A14 | 0.2135 | 0.2195 | 0.3309 | 0.5012 |
| MG-STAN-A23 | 0.0024 | 0.0026 | 0.0005 | 0.0008 |
| MG-STAN-A24 | 0.0854 | 0.0803 | 0.0275 | 0.0830 |
| MG-STAN-A34 | 0.7963 | 0.7977 | 0.9254 | 1.0000 |
| MG-STAN-A123 | 0.2142 | 0.2229 | 0.3645 | 0.4839 |
| MG-STAN-A124 | 0.0073 | 0.0072 | 0.0010 | 0.0008 |
| MG-STAN-A134 | 1.0000 | 1.0000 | 1.0000 | 0.8946 |
| MG-STAN-A234 | 0.2142 | 0.2229 | 0.3645 | 0.4839 |
| MG-STAN-A1234 | 0.2010 | 0.2049 | 0.1521 | 0.4839 |

### 3.3 Conclusion

This study proposes a Multi-Graph Fusion Gated Recurrent Spatial Attention Network (MG-STAN) for forecasting crude oil futures. The framework integrates three key components: (1) an enhanced spatiotemporal graph neural network that models markets as dynamic nodes using static graphs and learned attention; (2) a multi-graph architecture capturing cross-market interactions from diverse perspectives; and (3) a novel fusion layer that combines heterogeneous graph representations.

Inspired by multi-head attention, the multi-graph design processes interactions through parallel pathways in independent feature subspaces. Empirical results show the framework outperforms conventional deep learning models and single-graph variants in prediction accuracy and robustness, particularly in characterizing cross-market correlations and nonlinear dynamics.

Future research will explore: (1) extending the model to other energy commodities; (2) incorporating alternative data (e.g., news sentiment) to enrich graph representations; and (3) developing long-cycle forecasting by integrating macroeconomic indicators. These directions aim to build a more comprehensive framework for analyzing energy market dynamics.

## References


1. Deng S, Zhu Y, Duan S, et al. High-frequency forecasting of the crude oil futures price with multiple timeframe predictions fusion[J]. Expert Systems with Applications, 2023, 217: 119580.
2. Xiang Y, Zhuang X H. Application of ARIMA model in short-term prediction of international crude oil price[J]. Advanced materials research, 2013, 798: 979-982.
3. Morana C. A semiparametric approach to short-term oil price forecasting[J]. Energy economics, 2001, 23(3): 325-338.
4. Shambora W E, Rossiter R. Are there exploitable inefficiencies in the futures market for oil?[J]. Energy Economics, 2007, 29(1): 18-27.
5. Jin B, Xu X. Price forecasting through neural networks for crude oil, heating oil, and natural gas[J]. Measurement: Energy, 2024, 1: 100001.
6. Foroutan P, Lahmiri S. Deep learning-based spatial-temporal graph neural networks for price movement classification in crude oil and precious metal markets[J]. Machine learning with applications, 2024, 16: 100552.
7. Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[J]. arXiv preprint arXiv:1710.10903, 2017.
8. Liu Y, Xiao W, Chu T. Bi-graph attention network for energy price forecasting via multiple time scale learning[J]. Neural Computing and Applications, 2023, 35(21): 15943-15959.
9. Chai D, Wang L, Yang Q. Bike flow prediction with multi-graph convolutional networks[C]//Proceedings of the 26th ACM SIGSPATIAL international conference on advances in geographic information systems. 2018: 397-400.
10. He Y, Li L, Zhu X, et al. Multi-graph convolutional-recurrent neural network (MGC-RNN) for short-term forecasting of transit passenger flow[J]. IEEE transactions on intelligent transportation systems, 2022, 23(10): 18155-18174.