Active Arbitration: Decoupling Spatio-Temporal Duality for Efficient Traffic Forecasting

JiaJun Yu[1] , **Fang Yuan**[1] , **Guang-Yong Chen**[2] , **Min Gan**[1*]
[1]College of Computer Science and Technology, Qingdao University, Qingdao, China
[2]College of Computer and Data Science, Fuzhou University, Fuzhou, China
{yujiajun, yuanfang2}@qdu.edu.cn, cgykeda@mail.ustc.edu.cn, aganmin@aliyun.com

Abstract

Spatio-temporal forecasting is inherently bottlenecked by a phenomenon we term Spatio-Temporal Duality: the paradoxical coexistence of time-varying physical lags and instantaneous semantic synchrony. Constrained by passive coupling, existing architectures often employ indiscriminate spatial aggregation that fails to dynamically arbitrate interaction intensity based on traffic states, forcing models to rely on redundant deep stacking to approximate complex dynamics. We address this issue by introducing Time Arbitrated Spatial GNN (TAS-GNN), an efficient framework that arbitrates spatial interactions through time. Our model leverages deep temporal semantics to dynamically and proactively manage the aggregation of spatial domains, effectively decoupling physical connectivity from semantic relevance. In addition, to ensure model simplicity, we also introduce spectral decoupling via Discrete Wavelet Transform (DWT) to capture multiscale dependencies with minimal overhead. Experiments on three real-world datasets (PEMS04, 07, 08) show that TAS-GNN not only outperforms the current baseline model in accuracy, but also significantly improves the inference speed while reducing the number of parameters.

1 Introduction

Accurate traffic prediction is the key to an Intelligent Transportation Systems [Jiang and Luo, 2022]. The Spatio-Temporal Graph Neural Networks (STGNNs) [Li *et al.*, 2018b; Yu *et al.*, 2018] relies on the inherent homogeneity assumption in graph neural networks-spatial proximity determines the correlation of features, and this has led to significant progress in this field. However, we believe that this paradigm has significant theoretical deviations when applied to complex transportation systems, due to the dual nature of time and space in traffic evolution. That is, although the physical propagation follows a clear time-delay path through local neighbors, the relevant semantic synchronicity is often not limited by spatial distance.

Figure 1 illustrates this duality, which causes a fundamental conflict between physical propagation delay and semantic synchronization. The figure shows that while the semantic neighbor (node 35) exhibits immediate waveform synchronization, the upstream neighbor (node 77) shows a significant lag (60 minutes). This phenomenon creates a serious predicament that is difficult to solve using the existing paradigm. Traditional spatial-temporal neural networks (STGNNs [Yu *et al.*, 2018]) are limited by the reliance on static spatial priors and thus cannot capture dynamic propagation delays. Meanwhile, recent Transformer-based models(STAEformer [Liu *et al.*, 2023]) have begun to address this by adopting redundant deep stacking and complex dynamic changes. This not only makes them susceptible to false correlations due to heterogenous noise (node 92, 134) but also significantly increases computational costs.

In this work, we propose **TAS-GNN (Time-Arbitrated Spatial GNN)**, a streamlined framework that shifts the paradigm from Spatial-for-Temporal aggregation to **Time-Arbitrated Spatial Interaction**. Unlike methods that treat spatial dependencies as static or uniform, TAS-GNN leverages deep temporal contexts to explicitly arbitrate spatial aggregation. This mechanism dynamically engages topological neighbors during flow propagation while pivoting towards temporal inertia during saturation (congestion), effectively decoupling physical delays from semantic synchronization without requiring deep stacking. In summary, our key contributions are as follows:

- **Time-Arbitrated Spatial Interaction:** We propose a novel time arbitration mechanism that dynamically adjusts spatial aggregation by utilizing deep temporal context as a gating signal. This model adaptively filters noise and captures phase dependencies, effectively decoupling physical delays from semantic synchrony.
- **Resource-Efficient Architecture:** We designed an efficient framework that performs frequency disentanglement through the integrated Discrete Wavelet transform (DWT) and replaces computationally intensive global attention with linearly complex arbitration units. TAS-GNN significantly reduces parameter count and computational overhead, addressing the bottlenecks in deploying complex Transformers.
- **Comprehensive Evaluation:** Extensive experiments

*Corresponding author.

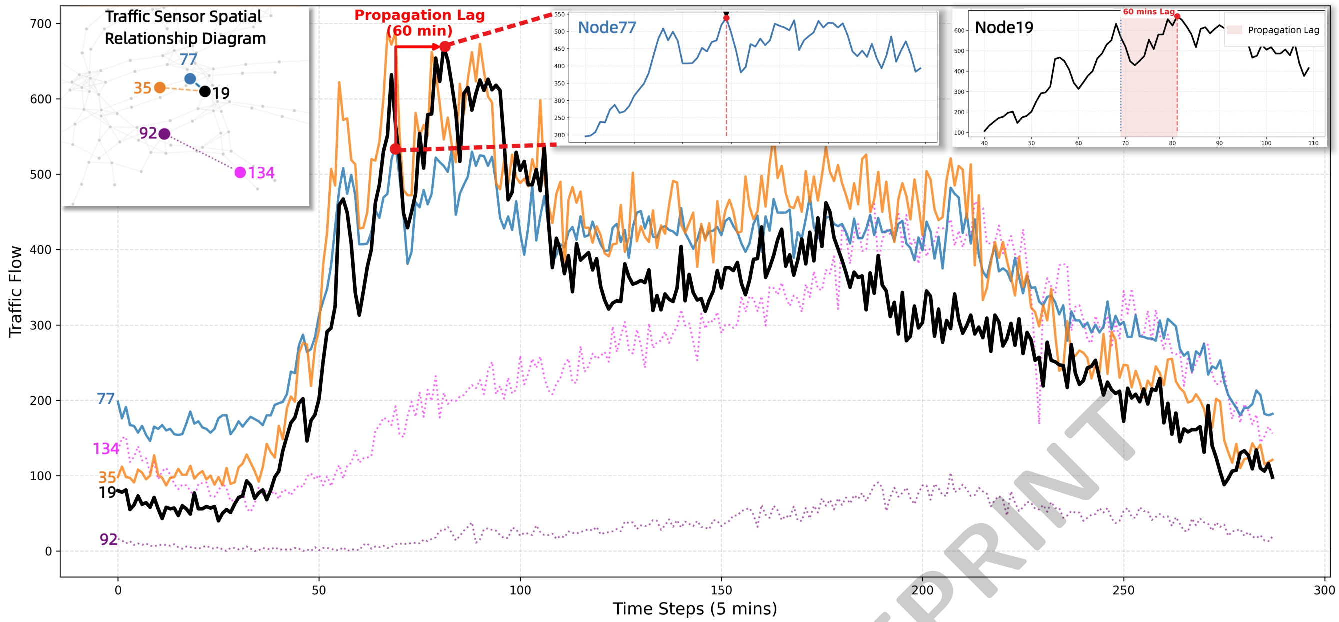

Figure 1: Empirical observation of the duality of time and space in the transportation network. The main figure compares the traffic flow time series of the target node 19 and its three different types of neighbors: (1) Upstream physical neighbors (node 77): shows a significant physical propagation lag of approximately 12 time steps (60 minutes); (2) Semantic neighbors (node 35): although there is no direct connection, it shows immediate semantic synchronization; (3) Heterogeneous nodes (nodes 92, 134): shows unrelated heterogeneous flow patterns. The upper left inset shows the geographical spatial distribution of these nodes and their topological connection relationships.

on three real-world benchmarks (PEMS04, PEMS07, PEMS08) demonstrate that TAS-GNN outperforms competitive baselines, it offers superior inference speed and low latency, providing a practical and scalable solution for real-time edge deployment.

2 Related Work

2.1 Graph-based Approaches

Modeling spatial dependencies in traffic forecasting has transitioned from rigid physical constraints to data-driven adaptation. Pioneering works, such as DCRNN [Li *et al.*, 2018b] and STGCN [Yu *et al.*, 2018], anchored diffusion processes to predefined adjacency matrices. While effective for extracting local correlations, these methods are limited by the rigid assumption of stationarity, restricting interactions to fixed physical connections. To mitigate reliance on incomplete physical priors, adaptive graph learning emerged. Although adaptive adjacency matrices (GraphWaveNet [Wu *et al.*, 2019] and AGCRN [Bai *et al.*, 2020]) have been introduced to learn the underlying structure, these methods inherently have the problem of passive coupling. They typically treat spatial dependence as a static prior or auxiliary feature and apply a fixed aggregation mechanism at all time steps. Therefore, they are unable to dynamically distinguish between physical propagation lags and semantic synchrony, and often require redundant depth to approximate these constantly changing patterns.

2.2 Attention-based Approaches

Recent works like GMAN [Zheng *et al.*, 2020], PDFormer [Jiang *et al.*, 2023] and STAEformer [Liu *et al.*, 2023] leverage global self-attention to capture long-range dependencies.These models utilize the global self-attention mechanism to identify long-range dependencies, effectively capturing the semantically synchronous part of bivalence. However, this approach also leads to the problem of indiscriminate aggregation. Since there are no constraints of physical topology, global attention is prone to be affected by false associations from different nodes. Furthermore, to compensate for the lack of topological induction, these models rely on deep stacking of attention layers, resulting in heavy computational overhead that hinders edge deployment.

2.3 Spatio-Temporal Interaction: From Passive Coupling to Active Arbitration

For spatio-temporal interactions, existing research has primarily explored different coupling methods between spatial and temporal modules [Wang *et al.*, 2020], but most follow the paradigm of spatial serving temporal. Whether employing alternating convolutional layers (STGCN [Yu *et al.*, 2018]) or integrating graph convolutions into recurrent unit gates (DCRNN [Li *et al.*, 2018b]), these approaches treat spatial aggregation as a preprocessing step to enhance temporal features [Guo *et al.*, 2019; Song *et al.*, 2020]. While effective for simple dependencies, this paradigm performs Indiscriminate Aggregation, failing to reconcile the inherent Spatio-Temporal Duality: the conflict between time-delayed physical propagation and instantaneous semantic synchrony. By passively mixing these conflicting signals, prior methods struggle to adapt to rapid traffic state transitions.

TAS-GNN fundamentally inverts this hierarchy by implementing the Time-Arbitrated Spatial Interaction (TASI) paradigm. Rather than treating spatial features as static inputs, we leverage deep temporal contexts as a governing signal to explicitly arbitrate spatial weight allocation. Unlike indiscriminate aggregation, this design ensures that spatial message passing is strictly synchronized with evolving traffic trends, effectively decoupling physical connectivity from semantic relevance.

2.4 Efficient Modeling: Beyond Deep Stacking via Multi-Scale Reconstruction

Achieving SOTA accuracy often relies on stacking deep attention layers [Zheng *et al.*, 2020; Liu *et al.*, 2023], yet this incurs quadratic complexity ($O(L^2)$) that hinders edge deployment. Multi-resolution modeling offers a viable path to efficiency. Notably, MedGNN [Fan *et al.*, 2025] validated hierarchical graph learning in medical classification, achieving SOTA results by capturing physiological correlations at varying scales. Nevertheless, transposing this to traffic forecasting encounters a fundamental granularity discrepancy. Unlike medical tasks identifying global discrete states, traffic forecasting demands capturing continuous, fine-grained propagation delays. Static hierarchical graphs remain insufficient to model such time-varying physical lags.

To bridge this gap, TAS-GNN reformulates the multi-resolution paradigm for ITS. Rather than relying on static hierarchies, we integrate Discrete Wavelet Transform (DWT) within our Time-Arbitrated framework. This design retains the spectral efficiency of multi-scale feature extraction—inspired by MedGNN—while dynamically decoupling physical propagation from semantic synchrony. The result is a Pareto-optimal architecture that surpasses heavyweight baselines without the burden of deep recursive stacking.

3 Problem Definition

We represent the traffic network as a graph $G = (V, E, A_{\text{static}})$, where V denotes the set of sensor nodes N. A_{static} is constructed using the k-nearest neighbors (k-NN) algorithm with Gaussian kernel weights. The edge weight formula is $A_{ij} = \exp(-\frac{\text{dist}(v_i,v_j)^2}{\sigma^2})$, where node j is among the top K nearest neighbors of node i (with k set to 8). Finally, row normalization is applied to the matrix to ensure propagation stability. Let $X_t \in \mathbb{R}^{N\times C}$ denote the traffic features at time t (flow), where C is the number of feature channels. Given a historical observation sequence of length T, $\mathcal{X} = \{X_{t-T+1}, \ldots, X_t\}$, our goal is to predict the future traffic states for the next τ time steps, $\mathcal{Y} = \{X_{t+1}, \ldots, X_{t+\tau}\}$. This task aims to simultaneously capture stable physical dependencies induced by A_{static} and instantaneous semantic correlations driven by dynamic attention. The forecasting problem can be formalized as learning a θ-parameterized mapping function F_θ:

$$\hat{\mathcal{Y}} = F_\theta(\mathcal{X}, G, \mathcal{T}), \qquad \hat{\mathcal{Y}} \in \mathbb{R}^{\tau\times N\times C} \tag{1}$$

where $\hat{\mathcal{Y}}$ is the predicted sequence and $\mathcal{T}$ denotes auxiliary temporal information (time-of-day).

4 Methodology

4.1 Overall Framework

We model traffic flow as a non-stationary spatio-temporal signal consisting of low-frequency trends, high-frequency fluctuations, and topology-constrained propagation (Spectral–Temporal–Spatial triangulation). Motivated by this decomposition, we adopt a wavelet-enhanced temporal encoder to provide multi-scale temporal representations without relying on deep stacking. To capture diffusion-like propagation over the road network while maintaining structural stability, we combine a multi-order spatial propagation module with a static-graph-biased attention mechanism. Building on these designs, we propose TAS-GNN, a time-arbitrated spatial model (Figure 2, which consists of four components: (1) **Domain-Differentiated Embedding**, (2) **Shallow Temporal Context Encoding**, (3) **Time-Arbitrated Spatial Generation**, and (4) **Residual Decoding**.

4.2 Domain-Specific Embedding and Feature Extraction

Unlike traditional methods, we adopt the domain-differentiated embedding approach based on the fundamental differences between time series and spatial topology. The temporal flow achieves the reduction of the encoder's learning complexity by integrating **Wavelet Feature Extraction**, **Position Embedding**, **Adaptive Embedding**, and **Shared Temporal Context** (TOD & DOW). At the same time, it establishes causal logic based on the sequence order and captures the implicit non-periodic dynamics to complement the relatively rigid clock-like pattern of the latter.

The spatial flow utilizes **Node Embedding** and **Shared Temporal Context** (TOD & DOW) to index the geographical spatial identity. The shared time context plays a unique role in achieving mode decoupling: In the temporal flow, it stipulates the periodic patterns, while in the spatial flow, it gives dynamic spatial correlation to the static signals.

Overall, domain embedding provides disentangled priors by modality and strengthens the multi-scale expression of temporal modeling and the time-arbitrated modeling of spatial dependence, thereby laying the foundation for decoupling physical constraints from semantic relevance.

4.3 Shallow Temporal Context Encoding

To capture global temporal context, we employ a 4-layer Transformer encoder[Vaswani *et al.*, 2017] to contextually encode the time series, aggregating information into node-specific deep temporal representations $H_{temp} \in \mathbb{R}^{N\times T\times D}$ via multi-head attention mechanisms. Concurrently, a pre-processed Discrete Wavelet Transform (DWT) decomposes the input into low-frequency trends and high-frequency details, achieving information orthogonality. The wavelet transform decomposes the input into frequency subbands, facilitating the separation of high-frequency components from low-frequency trends. This allows the shallow Transformer to focus solely on synchronizing semantic waveforms within traf-

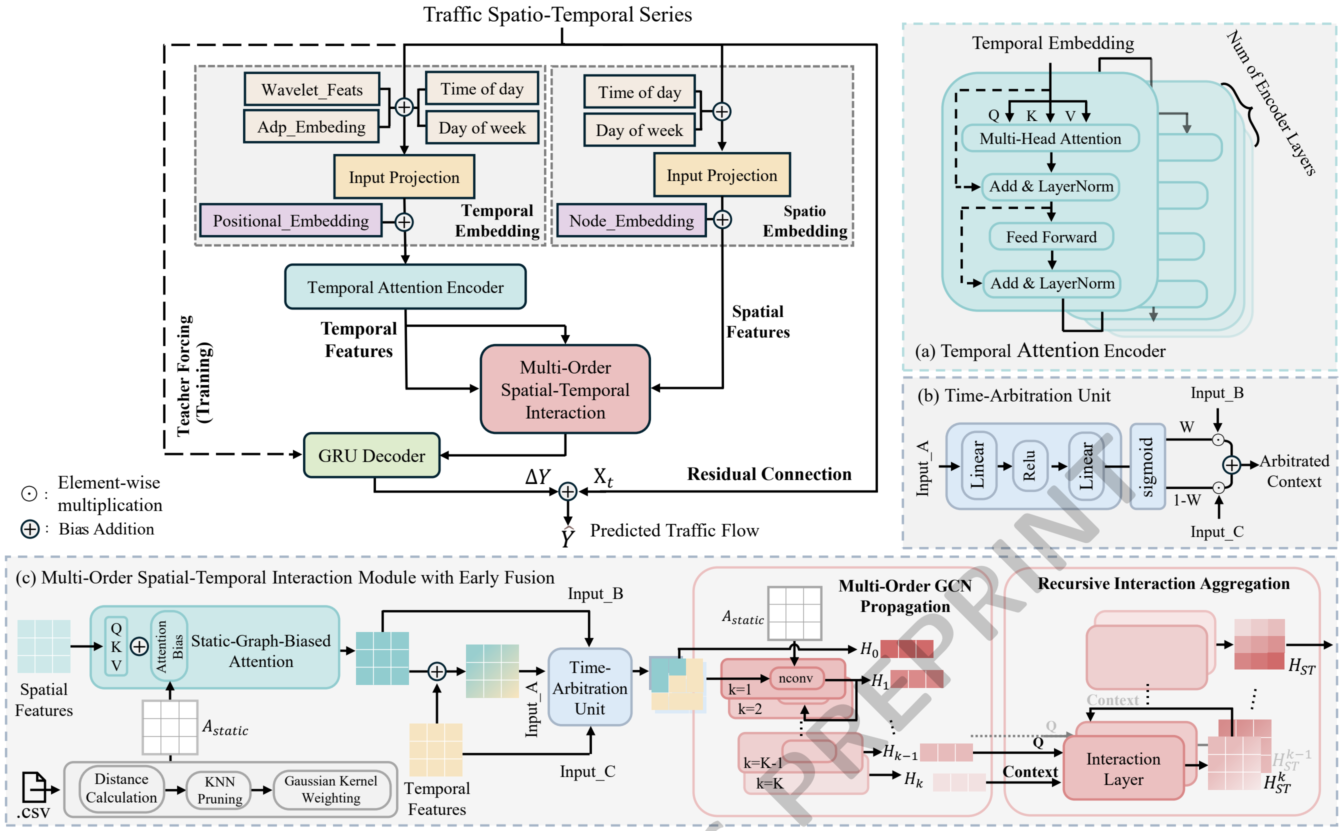

Figure 2: Time-Arbitrated Spatial Graph Neural Network **TAS-GNN**

fic flow data, eliminating the need for deeper layers to fit noise. The detailed analysis related to this is provided in Supplementary Material 1. (Supplementary materials are available at https://github.com/highhhhh/TAS-GNN).

Furthermore, the deep context encoding is strictly confined to the time domain ($T \ll N$), combined with an efficient spatial attention mechanism in the subsequent generation stage that differs from the $\mathcal{O}(N^2)$ complexity of all layers in standard spatio-temporal Transformers. It effectively avoids computational redundancy from spatio-temporal stacking, thereby reducing the model's parameter count.

4.4 Time-Arbitrated Spatial Generation

The deep temporal feature H_{temp} conditionally guides the dynamic spatial graph. Through three coupled steps, H_{temp} modulates spatial generation as a control signal, capturing macro-scale traffic patterns [Bai *et al.*, 2020; Wu *et al.*, 2019].

Static-Graph-Biased Spatial Attention. We treat the static network structure as a structural bias, directly injecting it into the dynamic attention computation process. For the dynamic spatial features of a given input, we first generate the query matrix, key matrix, and value matrix through linear projection [Vaswani *et al.*, 2017]:

$$\mathbf{Q} = X_{\text{spatial}}\mathbf{W}_Q, \mathbf{K} = X_{\text{spatial}}\mathbf{W}_K, \mathbf{V} = X_{\text{spatial}}\mathbf{W}_V \quad (2)$$

Although purely data-driven attention mechanisms can capture global semantic correlations, they are prone to overfitting on coefficient signals [Wu *et al.*, 2020]. Therefore, we introduce Astatic as a structural prior to anchor the attention mechanism to physical topological structures. The attention scores are computed as follows:

$$\text{Score}_{ij} = \frac{\mathbf{Q}_i \mathbf{K}_j^\top}{\sqrt{d}} + \lambda \cdot A_{\text{static},ij} \quad (3)$$

By making $\lambda \cdot A_{static}$ learnable, the model automatically balances dependencies between rigid physical topology and data-driven semantic relevance, adaptively optimizing structural biases across different flow patterns. This mechanism enables the model to focus on physically adjacent nodes, effectively preventing attention mechanisms from diverging on coefficient data while retaining the ability to capture latent long-range correlations. After computing the attention scores, we perform a weighted sum over the Value matrix, then add residual connections and layer normalization (LayerNorm) to obtain the dynamic spatial feature $H_{spatial} \in \mathbb{R}^{N \times D}$ for each time step given the input.

Time-Arbitrated Spatial Interaction Unit. After acquiring spatial features H_{spatial} that fuse static and dynamic information, we employ an arbitration fusion layer to achieve on-demand spatial aggregation. Traffic flow exhibits state-dependent interactions: free-flow is governed by temporal continuity; congestion formation actively recruits spatial neighbors for physical propagation; whereas saturated congestion relies on temporal inertia, utilizing spatial information merely as a corrective signal. Therefore, we process the spatio-temporal state—generated by concatenating local spatial features $H_{spatial}$ with the aligned node-wise temporal feature H_{temp} —through a multilayer perceptron to produce arbitration coefficients $z \in [0,1]^{N\times D}$ that participate in spatio-temporal information fusion. During the interaction phase, we slice the temporal feature H_{temp} at the current time step t to obtain $H_{temp}^{(t)} \in \mathbb{R}^{N\times D}$, ensuring strict point-wise alignment with the spatial feature $H_{spatial}$.

$$z = \sigma\big(\mathbf{W}_{g2} \cdot \mathrm{ReLU}\big(\mathbf{W}_{g1} \cdot \big[H_{\text{spatial}} \oplus H_{\text{temp}}\big]\big) + \mathbf{b}_g\big) \quad (4)$$

$$H_{\text{fused}} = z \odot H_{\text{spatial}} + (1 - z) \odot H_{\text{temp}} \quad (5)$$

This arbitration unit functions as a traffic condition classifier. By taking into account the current traffic flow conditions and using time states to distinguish between physical propagation delay and semantic synchronization, the model is able to dynamically decide whether to focus on time dependence or capture the characteristics of physical traffic propagation.

Multi-Order Propagation. To simulate the physical diffusion process of traffic flow across road networks, we feed the time-corrected feature H_{fused} into a Multi-Order Graph Convolutional Network(GCN) [Li *et al.*, 2018b], employing a multi-hop diffusion strategy to expand the receptive field of nodes [Wu *et al.*, 2019]. To rigorously simulate the physical diffusion process, define $H^{(0)} = H_{\text{fused}}$, and define the row-normalized transfer matrix as $\tilde{A} = D^{-1}A_{static}$. The k-th order propagation is then formulated as:

$$H^{(k)} = \text{nconv}(H^{(k-1)}, \tilde{A}) = \tilde{A}H^{(k-1)} \quad (6)$$

This normalization ensures that feature magnitudes remain stable across multi-hop diffusion steps.

By stacking K layers of graph convolutions, the model captures spatial influences ranging from immediate neighbors to Khop distant neighbors. To avoid the over-smoothing issue in deep GCNs [Li *et al.*, 2018a], we employ a Recursive Interaction Aggregation layer instead of using the last layer's output. This layer sequentially fuses features of different orders $\{H_0, H_1, H_K\}$ through a series of interaction layers, thereby integrating global context (higher-order features) while preserving local details (lower-order features). The resulting spatio-temporal context H_{ST} incorporates temporally validated, multi-scale spatial dependencies, providing discriminative feature representations for subsequent decoding.

4.5 Decoding and Optimization

To address the nonstationarity of traffic data, we designed a residual decoder based on GRU [Chung *et al.*, 2014]. Unlike directly predicting absolute values, this decoder uses the spatiotemporal hidden state H_{sr} generated by the encoder as its initial state, focusing on predicting the increment at the current time step relative to the previous moment. The final prediction result is:

$$\Delta\hat{Y} = \text{GRU}_{\text{dec}}(H_{ST}), \qquad \hat{Y} = X_{\text{last}} + \Delta\hat{Y} \quad (7)$$

This residual architecture reduces fitting difficulty and accelerates convergence. To further enhance stability, we employ two strategies:

Curriculum Learning via Teacher Forcing. During the initial training phase, the model uses ground truth labels as the next input for the decoder with a high probability to accelerate convergence. As training progresses, this probability linearly decays, forcing the model to gradually adapt to using its own predictions as input. This ensures an effective balance between rapid convergence during training and long-term prediction stability [Bengio *et al.*, 2015].

Differential Learning Rate Mechanism. In the spatial modeling module of the model, in order to prevent the sensitive Sigmoid activation function in the early fusion gate from experiencing polarization of weights due to a uniform and high global learning rate. To address this, we assign a lower learning rate multiplier to the Arbitration parameters [Howard and Ruder, 2018]. This strategy ensures a smooth evolution of the gate weights, enabling the model to more robustly search for the optimal spatio-temporal fusion ratio.

5 Experiments

5.1 Experiment Setup

Dataset. We evaluate TAS-GNN on three standard real-world public traffic benchmarks collected by the Caltrans Performance Measurement System (PeMS)[Chen *et al.*, 2001]: PEMS04 ($N = 307, T = 16,992$), PEMS07 ($N = 883, T = 28,224$), and PEMS08 ($N = 170, T = 17,856$). All data are aggregated at 5-minute intervals, focusing on traffic flow prediction.

Baselines. We compare TAS-GNN against a wide range of baselines, ranging from statistical methods (HA, ARIMA) and deep time-series models (LSTM) to advanced spatio-temporal neural networks. Specifically, we include classic GNN-based models such as STGCN, DCRNN, ASTGCN, GWNet, STSGCN, and AGCRN. Furthermore, we benchmark against recent Transformer-based SOTA models, including PDFormer, STAEformer, as well as advanced GNN variants such as MedGNN, and LightST.

Settings. Implemented in PyTorch on an NVIDIA A6000 GPU, our model is trained using AdamW and Huber Loss [Huber, 1964] ($\delta = 2.0$). Data is split 6:2:2 with horizon $T = \tau = 12$. The encoder uses 4 layers ($D = 96, H = 6$). We set the batch size to 32 and max epochs to 200 (early stopping patience=30). Notably, a 0.25 learning rate multiplier is applied to the arbitration parameters for stability.

5.2 Overall Performance

The overall performance of TAS-GNN and all baselines on PEMS04, PEMS07, and PEMS08 is reported in Table 1. The results show that deep spatio-temporal models such as AGCRN and GWNet generally outperform traditional methods by better modeling spatial dependencies. TAS-GNN achieves the best or highly competitive performance across the three datasets, with stable multi-run results on PEMS08, e.g., MAE 3.42 ± 0.04, RMSE 22.83 ± 0.05, and MAPE $8.45\% \pm 0.07\%$. These results indicate that TAS-GNN effectively captures dynamic traffic patterns and reduces forecasting errors through temporal-context-guided spatial modeling.

Datasets	PEMS04			PEMS07			PEMS08		
Method	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
HA [Ahmed and Cook, 1979]	26.35	42.16	16.74	30.32	56.66	12.80	23.30	40.58	14.49
ARIMA [Box *et al.*, 2015]	33.71	48.8	24.18	38.17	59.27	19.46	31.09	44.32	22.73
LSTM [Hochreiter and Schmidhuber, 1997]	32.07	44.29	16.11	31.59	54.33	14.03	27.58	38.09	10.39
DCRNN [Li *et al.*, 2018b]	24.41	37.78	17.10	26.8	41.44	11.67	19.91	30.94	12.39
STGCN [Yu *et al.*, 2018]	20.95	32.51	14.40	22.54	33.82	11.20	16.72	24.81	11.58
ASTGCN [Guo *et al.*, 2019]	25.89	40.24	17.00	29.42	45.05	13.00	21.83	33.4	13.00
GWNet [Wu *et al.*, 2019]	22.64	35.70	16.03	23.77	37.99	10.09	16.83	27.08	11.18
STSGCN [Song *et al.*, 2020]	21.68	34.02	14.50	24.87	39.77	10.65	17.67	27.44	11.42
AGCRN [Bai *et al.*, 2020]	19.79	32.49	13.10	22	36.85	9.09	16.77	26.43	10.51
PDFormer [Jiang *et al.*, 2023]	18.33	29.99	12.25	19.73	32.76	8.39	13.57	23.29	9.19
STAEformer [Liu *et al.*, 2023]	18.29	30.43	12.11	19.14	32.58	8.01	13.46	23.25	8.88
LightST [Zhang *et al.*, 2023]	18.12	**29.72**	11.87	19.29	**32.4**	8.06	13.46	23.06	8.87
MedGNN [Fan *et al.*, 2025]	22.21	37.14	14.71	29.34	52.59	12.53	20.70	32.38	12.26
STD-PLM [Huang *et al.*, 2025]	18.16	30.21	11.89	19.25	32.84	8.06	**13.31**	23.19	8.84
TAS-GNN(ours)	**17.94**	29.79	**11.57**	**19.01**	32.47	**7.78**	13.38	**22.79**	**8.38**

Table 1: Performance comparison on the PEMS04, PEMS07, and PEMS08 datasets, evaluated by MAE, RMSE, and MAPE (%).

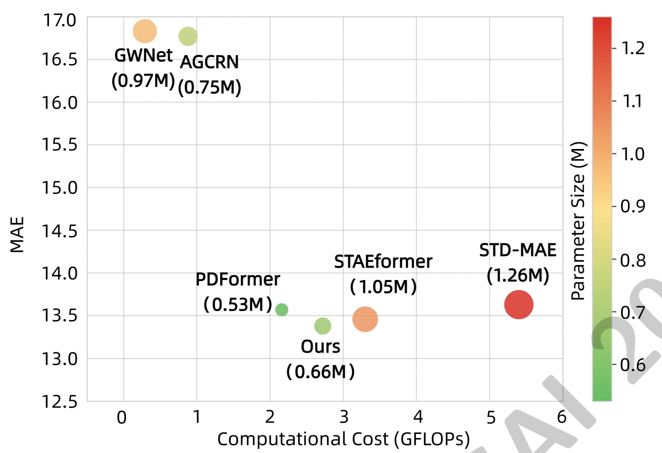

Figure 3: Efficiency–accuracy trade-off on PEMS08. The x-axis reports computational cost in GFLOPs and the y-axis reports MAE (lower is better). Bubble size and color indicate the number of parameters (in millions).

5.3 Efficiency and Computational Complexity Analysis

As shown in Figure 3, we evaluate overall efficiency by jointly considering parameter counts and theoretical computational costs (GFLOPs). The results show that TAS-GNN lies on the Pareto-efficient frontier between predictive performance and resource consumption.

Specifically, compared with the heavyweight SOTA STAEformer, TAS-GNN achieves comparable or superior accuracy while reducing the parameter count by 37% and significantly lowering GFLOPs, validating that the proposed time-arbitrated interaction is a more efficient alternative to deep stacking. While the lightweight PDFormer exhibits slightly lower computational costs, it suffers from a performance bottleneck. TAS-GNN surpasses PDFormer in accuracy with only a marginal increase in computation, effectively trading negligible cost for robustness grounded in physical topology. Detailed complexity derivations are provided in Supplementary Material 2.

To further validate practical deployment capabilities on resource-constrained edge devices, we conducted latency testing on the PEMS08 dataset: the model achieves an inference speed of just 6.79 ms/sample under high-throughput scenarios with Batch=32 (10.57 ms/sample at Batch=1). This result fully demonstrates that TAS-GNN not only possesses low theoretical complexity but also exhibits outstanding efficiency in practical large-scale data stream processing and real-time prediction scenarios.

5.4 Ablation Study

To verify the effectiveness of each key component in the proposed framework, we conducted the following ablation tests and comparisons on PEMS04 and PEMS08: (1)**Temporal-only**: Removes the spatial module, retaining only the Transformer-based temporal modeling and decoding process. (2)**Serial Stack**: Adopts a "time → space" serial architecture (Transformer → MRGNN → Decoder). (3)**Late Fusion (Gated)**: Fuses temporal and spatial features via a gating mechanism before decoding. (4)**w/o Arbitration**: Removes the time arbitration unit. Replaces the time arbitration unit with simple element-wise summation for feature fusion. (5)**w/o Spectral**: Removes the Discrete Wavelet Transform (DWT) module. (6)**w/o Static Graph Prior**: Removes the predefined static adjacency matrix, relying entirely on data-driven graph learning. All the ablation variants were evaluated under the same experimental settings as the complete model, using MAE, RMSE, and MAPE as evaluation metrics. The results are summarized in Table 2.

Table 2 quantitatively validates the contribution of each component, with TAS-GNN consistently achieving the lowest error rates. Specifically, the performance drop in Temporal-only confirms the necessity of spatial modeling, while the poor results of Serial Stack highlight the superiority of our

Setting	PEMS04			**PEMS08**		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
Full model	**17.94**	**29.79**	**11.57**	**13.38**	**22.79**	**8.38**
Temporal-only	18.18	30.31	11.66	13.58	22.98	8.55
Serial Stack	18.55	30.70	12.01	14.20	23.56	9.35
Late Fusion	18.07	30.26	11.66	13.60	23.04	8.55
w/o Arbitration	18.06	30.23	11.76	13.60	22.93	8.47
w/o Spectral	18.09	30.25	11.79	13.54	24.13	9.01
w/o Static Graph Prior	18.03	30.35	11.68	13.63	22.95	8.59

Table 2: Ablation study on PEMS04 and PEMS08.

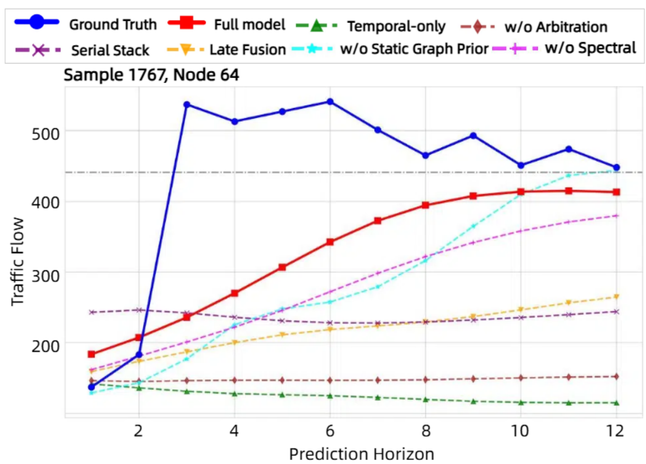

Figure 4: Performance comparison of traffic flow prediction between TAS-GNN and ablation variants under typical peak congestion scenarios.

parallel interaction paradigm. Crucially, the w/o Spectral variant exhibits a notable rise in RMSE, confirming that a priori frequency disentanglement is essential for the lightweight encoder to resist noise.

Although global metrics show modest gains, Figure 4 reveals the robustness of the model under critical conditions. In the dynamic trend scenario (Sample 1767), the Temporal-only variant (green) fails to capture the rapid upward momentum due to a lack of spatial context. In contrast, TAS-GNN leverages its parallel architecture to track surges with minimal lag. Crucially, the w/o Arbitration variant (brown) also suffers from severe mean reversion, failing to track the peak despite having access to spatial features. This indicates that static fusion alone is insufficient for regime shifts. In contrast, TAS-GNN (red) leverages the time-arbitrated mechanism to actively detect the state change and engage spatial neighbors, successfully tracking the surge with minimal lag. We have provided more diverse visualizations of transportation modes in Supplementary Material 3.

5.5 Arbitration Weight Visualization

To intuitively understand the efficacy of the time-arbitrated mechanism, we visualize the dynamic arbitration weights (α) of the early fusion module for Node 297 in PEMS04 (Figure 5), where the red dashed curve represents the learned importance of spatial neighbors. Initially, during free-flow periods (Steps 0–50), the spatial weight remains low (< 0.2), indicating that traffic evolution is dominated by temporal inertia (semantic synchrony) rather than physical interactions. As traffic volume increases (Steps 50–100), the weight rises dramatically, reflecting the active recruitment of upstream/downstream information to capture physical propagation lag. Crucially, during the peak congestion phase (Steps 100–250), the spatial weight stabilizes at approximately 0.35. This dynamic adaptation serves as compelling empirical evidence for the Spatio-Temporal Duality: it confirms that traffic states are not static but oscillate between propagation-dominant and synchrony-dominant modes. By stabilizing at this specific coupling coefficient, TAS-GNN effectively arbitrates this duality, balancing dominant temporal continuity (65%) with necessary spatial correction (35%) to filter out high-frequency noise without overreacting to transient oscillations.

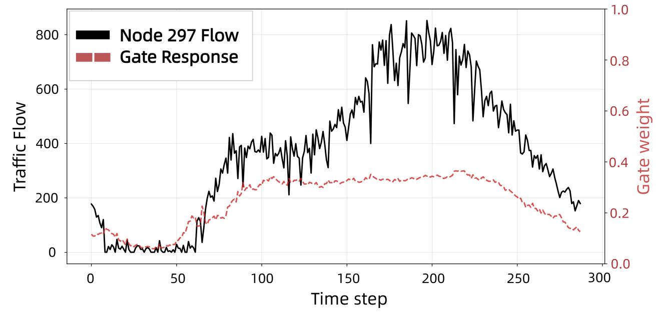

Figure 5: Gate behavior visualization on PEMS04 (Node 297). The black curve shows the traffic flow, while the red dashed curve shows the arbitration weight (spatial importance, normalized to $[0, 1]$). The x-axis is indexed by 5-minute time steps.

6 Conclusion

This paper re-examines the fundamental challenges of traffic prediction from the perspective of temporal-spatial duality theory. By proposing TAS-GNN, we have demonstrated that the inefficiency of existing models is not due to insufficient depth, but rather to the passive coupling of physical and semantic dependencies. Our research results show that replacing indiscriminate stacking with an active arbitration mechanism (TASI) enables the model to switch between capturing physical lags and semantic synchronizations in real time, thereby achieving the optimal balance between accuracy and efficiency. Moreover, it validates that precise mechanism modeling not only reduces the number of parameters and accelerates the training convergence speed, but also can serve as a powerful alternative to the brute force of parameter redundancy. This provides a solution for deploying high-performance, context-aware prediction models on resource-constrained edge devices.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 62576183 and 62073082.

References

[Ahmed and Cook, 1979] M. S. Ahmed and A. R. Cook. Analysis of freeway traffic time-series data by using box-jenkins techniques. *Transportation Research Record*, 722:1–9, 1979.

[Bai *et al.*, 2020] Lei Bai, Lina Yao, Can Li, Xianzhi Wang, and Can Wang. Adaptive graph convolutional recurrent network for traffic forecasting. In *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS)*, pages 17804–17815, Vancouver, Canada, December 2020. Curran Associates, Inc.

[Bengio *et al.*, 2015] Samy Bengio, Oriol Vinyals, Navdeep Jaitly, and Noam Shazeer. Scheduled sampling for sequence prediction with recurrent neural networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems (NeurIPS)*, pages 1171–1179. MIT Press, 2015.

[Box *et al.*, 2015] George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, Hoboken, New Jersey, 2015.

[Chen *et al.*, 2001] Chao Chen, Karl Petty, Alexander Skabardonis, Pravin Varaiya, and Zhanfeng Jia. Freeway Performance Measurement System: Mining loop detector data. *Transportation Research Record*, 1748(1):96–102, 2001.

[Chung *et al.*, 2014] Junyoung Chung, Çağlar Gülçehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of Gated Recurrent Neural Networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

[Fan *et al.*, 2025] Wei Fan, Jingru Fei, Dingyu Guo, Kun Yi, Xiaozhuang Song, Haolong Xiang, Hangting Ye, and Min Li. Towards multi-resolution spatiotemporal graph learning for medical time series classification. In *Proceedings of the ACM Web Conference 2025 (WWW)*, pages 5054–5064, Sydney, Australia, April 2025. ACM.

[Guo *et al.*, 2019] 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 33rd AAAI Conference on Artificial Intelligence (AAAI)*, pages 922–929, Honolulu, Hawaii, January 2019. AAAI Press.

[Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[Howard and Ruder, 2018] Jeremy Howard and Sebastian Ruder. Universal language model fine-tuning for text classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 328–339. Association for Computational Linguistics, 2018.

[Huang *et al.*, 2025] Yiheng Huang, Xiaowei Mao, Shengnan Guo, Yubin Chen, Junfeng Shen, Tiankuo Li, Youfang Lin, and Huaiyu Wan. STD-PLM: Understanding both spatial and temporal properties of spatial-temporal data with PLM. In *Proceedings of the 39th AAAI Conference on Artificial Intelligence*, pages 11817–11825, Philadelphia, Pennsylvania, USA, February 2025. AAAI Press.

[Huber, 1964] Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

[Jiang and Luo, 2022] Weiwei Jiang and Jiayun Luo. Graph neural network for traffic forecasting: A survey. *Expert Systems with Applications*, 207:117921, 2022.

[Jiang *et al.*, 2023] 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 37th AAAI Conference on Artificial Intelligence (AAAI)*, pages 4365–4373, Washington, DC, USA, February 2023. AAAI Press.

[Li *et al.*, 2018a] Qimai Li, Zhige Li, and Lingpei Han. Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, pages 3538–3545, New Orleans, USA, 2018. AAAI Press.

[Li *et al.*, 2018b] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *Proceedings of the 6th International Conference on Learning Representations (ICLR)*, Vancouver, Canada, April 2018. OpenReview.net.

[Liu *et al.*, 2023] Hang Liu, Jiabo Wang, and Meng Chen. Spatio-temporal adaptive embedding makes vanilla transformer SOTA for traffic forecasting. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM)*, pages 4125–4129, Birmingham, UK, October 2023. ACM.

[Song *et al.*, 2020] 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 34th AAAI Conference on Artificial Intelligence (AAAI)*, pages 914–921, New York, USA, February 2020. AAAI Press.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS)*, pages 6000–6010, Long Beach, CA, USA, 2017. Curran Associates, Inc.

[Wang *et al.*, 2020] Senzhang Wang, Jiannong Cao, and Philip S. Yu. Deep learning for spatio-temporal data mining: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 34(8):3681–3700, 2020.

[Wu *et al.*, 2019] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. Graph wavenet for deep spatial-temporal graph modeling. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1907–1913, Macao, China, August 2019. ijcai.org.

[Wu *et al.*, 2020] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, pages 753–763, Virtual Event, 2020. ACM.

[Yu *et al.*, 2018] Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3634–3640, Stockholm, Sweden, July 2018. ijcai.org.

[Zhang *et al.*, 2023] Hongjun Zhang, Zonghan Wu, Ye Liu, and Shirui Pan. LightST: Lightweight spatio-temporal graph neural network for traffic flow forecasting. In *Proceedings of the 2023 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, Gold Coast, Australia, 2023. IEEE.

[Zheng *et al.*, 2020] Chuanpan Zheng, Xiaoliang Fan, Cheng Wang, and Jianzhong Qi. GMAN: A graph multi-attention network for traffic prediction. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, pages 1234–1241, New York, USA, 2020. AAAI Press.