2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Traffic Flow Prediction Using Multi-Scale Convolution and Attention Mechanisms

Pengfei Qi[1], Jinlai Zhang[2](✉), Chulin Li[1], Linlong Lei[2]
Wei Hao[2], and Xiong Jiang[3]

[1] School of Physics and Electronic Science, Changsha University of Science and Technology, Changsha, 410114, Hunan, China
[2] College of Mechanical and Vehicle Engineering, Changsha University of Science and Technology, Changsha, 410114, Hunan, China
[3] Changsha Planning and Design Institute Co., Ltd., Changsha, 410114, Hunan, China

Abstract. Traffic flow prediction is a critical task in intelligent transportation systems, significantly improving the efficiency of traffic management and scheduling. However, the complexity and diversity of traffic flow data pose substantial challenges to existing prediction methods. In particular, the frequent temporal variations and spatial characteristics in complex spatiotemporal data are difficult to handle effectively. To address this issue, this paper proposes MCANet, a novel prediction model designed to capture the intricate spatiotemporal features in traffic flow prediction through the integration of multi-scale convolution and attention mechanisms. Specifically, we introduce the Large Kernel Decomposition and Spatio-Temporal Selection (LKD-STS) module to enhance the model's ability to extract multi-scale features in traffic flow prediction, enabling it to better capture traffic patterns at different temporal scales. Additionally, we propose the Global Channel Spatial Attention (GCSA) module to improve the model's capability in capturing multi-scale traffic features and preserving spatial-channel information. Furthermore, we introduce the Partial Convolution Batch-normalization GELU (PCBG) module, which reduces redundant computations and memory access through partial convolution techniques, thereby enhancing the model's efficiency. Compared to the baseline model and other state-of-the-art (SOTA) traffic flow prediction models, MCANet demonstrates superior performance on the Flight and Traffic datasets. Notably, MCANet efficiently captures complex spatiotemporal features, maintaining stable performance in high-frequency traffic flow prediction tasks. Experimental results show that MCANet excels in traffic flow prediction tasks with varying prediction horizons. Particularly, for a prediction length of T=96, MCANet outperforms SOTA models such as MSGNet and TimesNet, with Mean Squared Error (MSE) reductions of 2.7% and 5.7%, respectively.

Keywords: Multi-Scale Convolution, Attention Mechanism, Traffic Flow Prediction, Time Series.

1 Introduction

Traffic flow prediction plays a pivotal role in the optimization and management of intelligent transportation systems (ITS). The growing complexity and volume of traffic data generated by urban environments, as well as the increasing need for real-time decision-making, have led to an increasing demand for robust traffic flow prediction models. Effective traffic prediction is crucial for various applications, including traffic congestion control, route planning, and predictive maintenance, all of which contribute to enhancing traffic safety, reducing travel time, and improving overall traffic efficiency.

Traffic flow prediction is inherently a spatiotemporal task, where the goal is to predict the future traffic states at different locations over a specified time horizon. Unlike traditional forecasting tasks, traffic flow prediction faces unique challenges due to the dynamic, non-linear, and complex nature of traffic systems. Traditional time-series forecasting methods, such as ARIMA [1], LSTM [6], and GRU [5], while capable of handling time-series data, still perform inadequately when confronted with complex tasks involving long-range dependencies. The traffic flow data is influenced by a myriad of factors, such as weather conditions, road incidents, and human-driven behaviors, which vary not only across time but also across different spatial regions. These fluctuations, combined with the interdependence between traffic patterns at various locations, make accurate prediction a difficult and computationally expensive task. Traditional methods, such as time-series analysis and linear regression, often fail to capture the complexity inherent in these spatiotemporal dependencies, leading to suboptimal predictive performance. Recently, deep learning-based models such as Informer [19] and Autoformer [12] have introduced self-attention mechanisms to enhance the modeling of long-term dependencies; however, they still face challenges in addressing multi-scale features and complex spatial dependencies. To overcome this limitation, convolutional neural networks (CNNs) [20-24] have been introduced to capture spatial features by applying convolutional operations across the spatial domain of traffic data. Despite these advancements, most existing methods still struggle with the trade-off between efficiently modeling both spatial and temporal dependencies in a unified framework, leading to less accurate predictions, especially in high-frequency and large-scale traffic flow prediction tasks.

To address these challenges, this paper proposes MCANet, a novel traffic flow prediction model that leverages the strengths of both convolution and attention mechanisms to simultaneously capture intricate spatial and temporal dependencies in traffic data. At the heart of MCANet is the Large Kernel Decomposition and Spatio-Temporal Selection (LKD-STS) module, which enhances the model's ability to extract multi-scale features from traffic flow data. The LKD-STS module is specifically designed to address the issue of capturing traffic patterns at different temporal scales. In addition, MCANet incorporates the Global Channel Spatial Attention (GCSA) module to further enhance its ability to capture spatial features. The GCSA module operates by applying a global attention mechanism that learns the relationships between different spatial locations and channels in the traffic flow data. By focusing on the most relevant features at each spatial location and time step, the GCSA module improves the model's ability to preserve essential spatial and temporal information, leading to more accurate and robust predictions. Moreover, MCANet introduces the Partial Convolution Batch-normalization GELU (PCBG) module, which enhances the model's computational efficiency. The PCBG module reduces redundant computations and memory access by employing partial convolution techniques, enabling faster training and inference.

The experimental evaluation of MCANet on benchmark datasets, such as the Flight and Traffic datasets, demonstrates its superior performance compared to SOTA models, including MSGNet [2] and TimesNet [11]. Notably, MCANet shows a significant reduction in Mean Squared Error (MSE) compared to these baseline models, with reductions of 2.7% and 5.7% for a prediction length of T=96, respectively. These results highlight the effectiveness of MCANet in handling complex spatiotemporal features and maintaining stable performance, even in high-frequency traffic flow prediction tasks. The results also reveal that MCANet significantly outperforms existing models, such as MSGNet, Informer, and Autoformer, across key evaluation metrics, including MSE and MAE. Notably, MCANet excels in capturing the spatiotemporal features of traffic flow, thereby enhancing its ability to predict complex traffic patterns. Our contributions can be summarized as follows:

- We propose a novel deep learning architecture for traffic flow prediction that integrates multi-scale convolution and attention mechanisms to capture both spatial and temporal dependencies effectively.
- We introduce the LKD-STS module, which improves the model's ability to capture multi-scale temporal features by leveraging large kernel convolutions and spatio-temporal selection.
- Through extensive experiments on multiple datasets, we demonstrate that MCANet outperforms existing SOTA models.

2 Methodology

2.1 Overview of the proposed MCANet

In this paper, we propose an innovative method based on Multi-Scale Convolution and Attention Mechanisms, for simplicity, we refer to this method as MCANet. To overcome the inherent limitations of traditional methods in handling multi-scale spatiotemporal correlations, we propose the integration of multi-scale convolution with attention mechanisms. This is especially important in complex spatiotemporal data, where frequent changes across different time periods and spatial features pose challenges for effective processing. Our proposed MCANet is illustrated on the left side of Fig. 1, where each Block entails a four-step sequence: 1) Extracting periodic frequency information from the input time series via "FFT for Periods"; 2) Modeling multi-scale correlations with the Scale and Mid Layer modules, followed by reshaping the data's dimensions; 3) Refining features through the LKD-STS, GCSA,

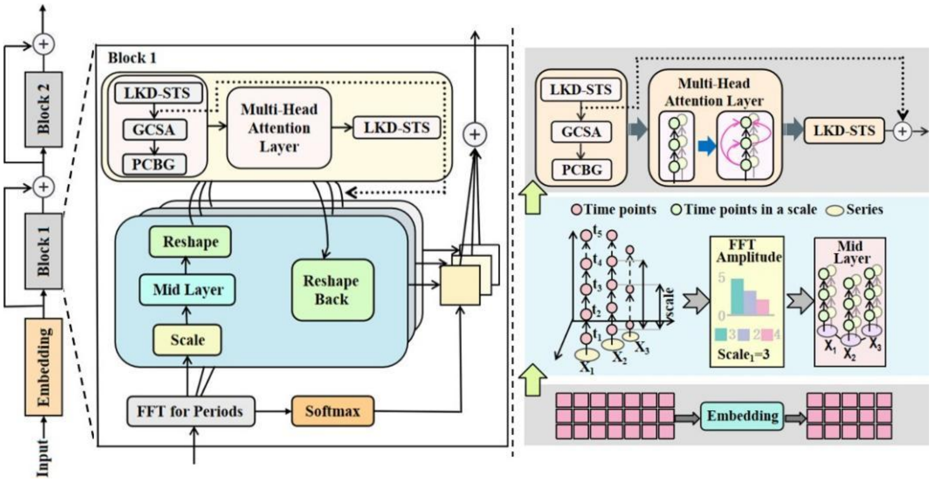


Fig. 1. The overall architecture of MCANet.

and PCBG modules, then fusing them with Multi-Head Attention Layer; 4) Integrating the final representation into the backbone network via an addition operation. When it comes to the innovations of this paper, firstly, to capture traffic patterns across different time scales, we introduce the LKD-STS module. Secondly, to effectively identify and reinforce key traffic spatial features, we incorporate the GCSA module. Finally, we propose the PCBG module, which alleviates redundant computation through partial convolution, thereby improving the model's efficiency. Furthermore, to mitigate the potential overfitting issue arising from the introduction of new modules, we incorporate residual connections between the two LKD-STS modules, ensuring the model remains robust in handling the diverse features and dynamic dependencies in complex traffic data [8]. In the following subsections, we provide a detailed introduction to each of these three modules.

2.2 LKD-STS Module

To address the need for multi-scale feature extraction [25-28] in traffic flow prediction and capture traffic patterns across different time scales, we propose the Large Kernel Decomposition and Spatio-Temporal Selection (LKD-STS) module, as illustrated in Fig. 2. The LKD-STS utilizes its unique Large Selective Kernel mechanism to dynamically adjust the receptive field size based on the input data requirements [15]. This feature is particularly effective in capturing patterns across different time scales in traffic data, making it especially suitable for complex traffic scenarios where both short-term fluctuations and long-term trends coexist. Compared to traditional selective convolution modules, such as SKNet [13], ResNeSt [14], and SCNet [9], the time-selective mechanism of LKD-STS is more flexible and adaptive, enabling it to effectively handle the variable patterns encountered in traffic flow prediction tasks.

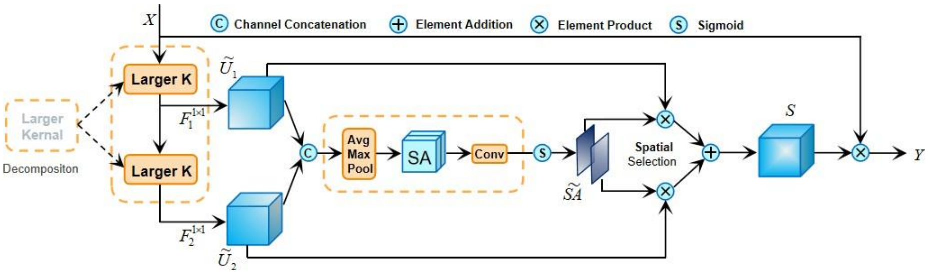


Fig. 2. The key components of the LKD-STS.

Moreover, the varying requirements across different time scales, including short-term and long-term fluctuations, necessitate the recognition of distinct temporal characteristics. The LKD-STS model leverages its Large Kernel Selection module to dynamically capture comprehensive temporal context information. This capability enables it to detect changes in traffic flow with greater precision across diverse time scales. The core part of LKD-STS includes:

Decomposition of Large Convolutions: LKD-STS decomposes large convolution kernels into sequences of multiple depth-wise convolution kernels, ensuring a large receptive field while effectively reducing the number of parameters. This decomposition is described by the following formula:

$$RF_1 = k_1, RF_i = d_i(k_i - 1) + RF_{i-1} \tag{1}$$

where k_i and d_i represent the size and dilation rate of the i-th convolution kernel, respectively. By decomposing large convolution kernels, LKD-STS can dynamically adjust to capture both temporal and spatial features in complex traffic scenarios, ensuring the extraction of rich contextual information.

Spatio-Temporal Selection Mechanism: After extracting multi-scale spatiotemporal features, LKD-STS employs a spatio-temporal selection mechanism to dynamically weight features at different spatial and temporal positions, further enhancing the flexibility and effectiveness of feature extraction. This mechanism generates selection masks through max pooling and average pooling, and uses the following formula to combine the features at each position:

$$S = F(\sum_{i=1}^{N} (S_i \cdot U_i)) \tag{2}$$

where $S_i$represents the spatio-temporal selection mask corresponding to each convolution kernel.

Through the selective use of large convolution kernels and a dynamic spatiotemporal selection mechanism, LKD-STS is able to effectively capture large-scale contextual information at lower layers while optimizing and refining fine-grained features at higher layers. As a result, LKD-STS demonstrates excellent performance in traffic flow prediction, particularly in recognizing traffic patterns across different time scales, as well as in complex traffic spatial scenarios, such as intersections and congested areas. This design significantly enhances the model's predictive accuracy and robustness.

2.3 GCSA Module

Although traditional attention mechanisms (such as SENet [10], CBAM [18], etc.) improve feature extraction performance to some extent, they often lead to the loss of global features when processing the channel and spatial dimensions separately. This occurs due to dimensionality reduction and independent operations, making it difficult to fully capture multi-dimensional interaction information. To highlight key spatial features in complex traffic scenarios, we propose the Global Channel Spatial Attention (GCSA) module, which effectively enhances the model's ability to capture multi-scale traffic features and preserve spatial-channel information.

The GCSA module employs a global attention mechanism, introducing 3D displacement operations and a Multi-Layer Perceptron (MLP), allowing the channel attention submodule to retain global information across space, time, and channels, and amplify cross-dimensional dependencies.

Meanwhile, the spatial attention module replaces pooling with convolution operations, effectively preventing feature loss. It also leverages group convolutions and channel shuffling to reduce computational complexity.

In the GCSA module, channel attention first preserves the integrity of the three-dimensional information through 3D displacement and then reprocesses the information using a two-layer MLP to enhance feature representation capability. In the spatial attention component, the 7×7 convolution kernel not only increases the model's sensitivity to spatial dependencies but also effectively controls the growth of parameters through group convolutions. The overall workflow is illustrated in Fig. 3 and is specifically described by the following two formulas:

$$F_2 = M_c(F_1) \otimes F_1 \tag{3}$$

$$F_3 = M_s(F_2) \otimes F_2 \tag{4}$$

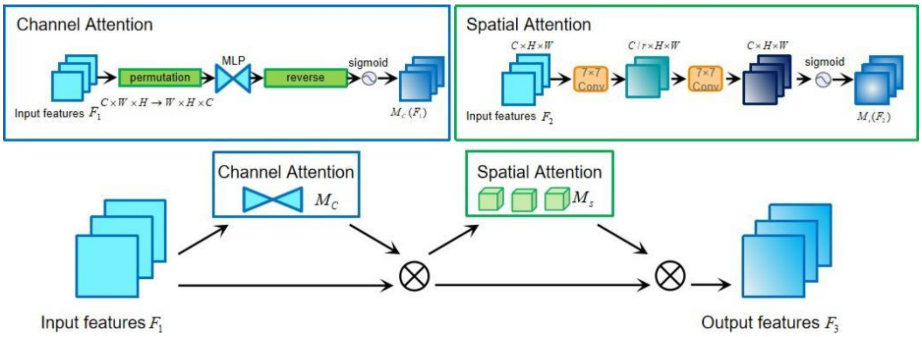


Fig. 3. The GCSA module.

where M_c and M_s represent the channel and spatial attention maps, respectively, and $\otimes$ denotes element-wise multiplication.

The GCSA module can preserves global features and enhance spatial-channel dependencies, particularly in peak and off-peak period predictions, thereby eliminating the information loss caused by pooling. This allows the model to more accurately perceive dynamic interactions between sensors. At the same time, GCSA optimizes the fusion of spatial and channel features, enabling MCANet to better capture the dynamic changes in traffic flow across different time scales and spatial dimensions, significantly improving the model's prediction accuracy and robustness in complex traffic environments, and helping traffic management systems more effectively address peak traffic and unexpected events.

2.4 PCBG Module

In order to extract spatial features more efficiently by minimizing redundant computations and memory accesses, we propose the Partial Convolution Batch-normalization (PCBG) module. Specifically, the PCBG module combines partial convolution (PConv) [3], Batch Normalization, and GELU activation functions, with the detailed structure shown in Fig. 4.

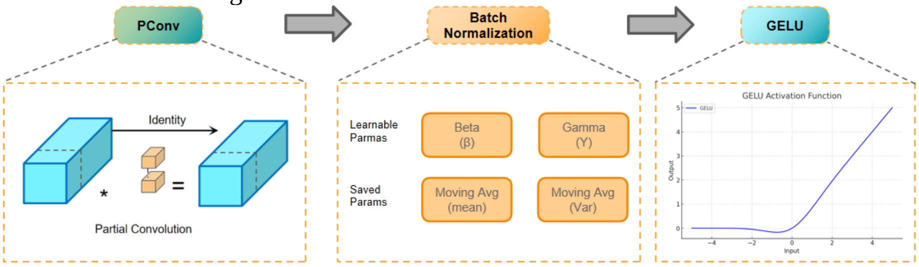


Fig. 4. Details of the PCBG module

In traffic prediction tasks, traditional convolution operations have significant limitations, particularly when handling spatiotemporal dynamic features. Traditional convolutions perform computations uniformly across all channels, leading to a large number of redundant operations. For example, when processing the spatiotemporal features of traffic flow, the input data often contains redundant spatial information. Traditional convolutions apply the same computations to each input channel, which results in low computational efficiency and increased dependency on computational resources and memory. This leads to computational bottlenecks and excessive memory access requirements during the inference of large-scale traffic data, making it challenging to meet the performance demands of real-time applications.

The proposed PCBG module significantly reduces the computational load by introducing PConv, optimizing the model’ s efficiency. The PConv module applies convolution filters only to a subset of the input channels while leaving the remaining channels unchanged, thereby optimizing the process of spatial feature extraction. The design of PConv is shown on the left side of Fig. 4. FLOPs can be expressed as:

$$h \times w \times k^2 \times c_p^2. \tag{5}$$

where $\mathrm{c_p}$ represents the number of partial channels. Typically, the partial ratio of PConv is $\mathrm{r} = \frac{\mathrm{c_p}}{\mathrm{c}} = \frac{1}{4}$, and the FLOPs is 1/16 of the standard convolution. This provides MCANet with a distinct advantage in applications requiring efficient inference.

In addition, the PCBG module integrates Batch Normalization and the GELU activation function, which provide the model with greater stability and stronger nonlinear expressiveness when handling complex traffic data. Batch Normalization standardizes the output features, improving the model’ s training stability on traffic data, accelerating convergence, and reducing issues such as vanishing and exploding gradients. The GELU activation function, compared to traditional ReLU, is smoother when processing dynamic traffic data and can effectively preserve subtle differences in the input features, thus improving the model’ s prediction accuracy.

Therefore, compared to traditional convolution, the PCBG module not only reduces redundant computations and improves efficiency, but also enhances the model’ s stability and accuracy in traffic flow prediction. Specifically, in scenarios involving long time series and complex dynamic features, the PCBG module significantly boosts the model’ s overall performance in traffic prediction [8].

3 Experimental Results

In this section, we perform extensive experiments to validate the effectiveness of MCANet. In Section 3.1, we introduce the traffic flow dataset used for training. Section 3.2 details the five baseline models used for comparison and the loss function used for training. In Section 3.3, we describe the experimental setup for training. In Section 3.4, we validate the capability of the enhanced model, MCANet, which consists of the LKD-STS, GCSA, and PCBG modules, as proposed in this paper. We also analyze the contribution of the PCBG module in reducing computational cost. In Section 3.5, we analyze and compare the impact of different modules on the overall model performance through ablation studies. Finally, in Section 3.6, we present visualized prediction examples on the Flight and Traffic datasets to provide intuitive comparisons of different models in capturing trend variations and periodic patterns.

3.1 Datasets

Table 1. Description of datasets.

Datasets	Dim	Input Length	Output Length	{Train / Test / Valid Size}	Frequency
Flight	7	96	{96, 192, 336, 720}	(18317, 2633, 5261)	Hourly
Traffic	862	96	{96, 192, 336, 720}	(12185,1757,3509)	Hourly

Flight dataset [2]: The Flight dataset, obtained from the OpenSky Network platform, contains flight operation records related to the COVID-19 pandemic. As a widely used benchmark dataset in the original MSGNet paper, it is utilized to assess model robustness under out-of-distribution (OOD) conditions. The dataset spans from January 2019 to December 2021 and includes flight information from seven major European airports: EDDF, EHAM, LEMD, LFPG, LIRF, LSZH, and UUEE.

It comprises a diverse set of attributes, including flight number, airline, departure and arrival airports, scheduled and actual departure/arrival times, flight delay durations, and operational status (e.g., on-time, delayed, or canceled) [4]. Due to the significant disruptions in air traffic caused by the COVID-19 outbreak, a considerable portion of the dataset consists of OOD samples, making it a valuable resource for evaluating model generalization in non-stationary environments.

Traffic dataset [19]: The Traffic dataset, provided by the California Department of Transportation, contains occupancy rate data from highway segments in the San Francisco Bay Area. As one of the widely adopted benchmark datasets in the MSGNet study, it is commonly used in time series forecasting tasks. Thedata were collected by 862 sensors deployed along major highways between 2015 and 2016, with a temporal resolution of one measurement per hour.

Each record in the dataset reports the occupancy rate of a specific highway segment as a percentage, representing the proportion of time that segment was occupied by vehicles. This enables a comprehensive characterization of traffic patterns across different times and locations.

To ensure consistency with the experimental configuration of MSGNet, we adopt an identical data partitioning strategy for both datasets, dividing them into training, validation, and test sets with a ratio of 7:1:2. All subsets are standardized using the mean and standard deviation computed from the training set to ensure fair comparisons across models. Detailed partitioning information used in the experiments is summarized in Table 1.

3.2 Baseline Models and Evaluation Metrics

We compared our approach against five time series prediction methods, including the MSGNet [2] framework with an adaptive graph convolution module and Transformer-based models such as Informer [19] and Autoformer [12]. Addi- tionally, we included the linear model DLinear [16] and considered TimesNet [11], which employs period decomposition and currently attains SOTA performance. All models are trained using Mean Squared Error (MSE) as the loss function.

For model evaluation, we employ two metrics: MSE and Mean Absolute Error (MAE), defined as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{6}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{7}$$

where y_i and $\hat{y}_i$ represent the i-th true value and predicted value, respectively, and n denotes the size of the test set [17].

3.3 Experimental Setup

All experiments were conducted on a Tesla P100 GPU. We adopted Mean Squared Error (MSE) as the primary loss function for model training. To ensure reproducibility, the input sequence length was fixed at L = 96, and the prediction lengths were selected from T = {96, 192, 336, 720}. The initial learning rate was set to 1×10^{-4}, and the batch size was set to 32. Each model was trained for a maximum of 10 epochs, with early stopping applied when necessary to prevent overfitting.

To maintain fairness and consistency, we applied the same set of hyperparameters across both the Flight and Traffic datasets, as summarized in Table 2. In particular, the number of temporal scales was set to k = 5, and the embedding dimension of the node features was fixed at 100. The model depth d_{model} was set to 16 for Flight and 1024 for Traffic, reflecting the difference in input feature dimensionality. We used 2 ScaleGraph blocks and a MixHop order of 2 to capture both local and multi-hop dependencies. The number of attention heads was set to 8, and optimization was performed using the Adam optimizer [7].

Table 2. Hyper-parameters on Flight and Traffic.

Datasets	**Flight/Traffic**
Epochs	10
Batch size	32
Loss	MSE
Learning rate	1e-4
k	5
Dim of E	100
d_{model}	{16,1024}
ScaleGraph block	2
Mixhop order	2
Heads	8
Optimizer	Adam [7]

3.4 Results and Analysis

As shown in Table 3, MCANet demonstrates outstanding overall performance on the Flight dataset, achieving either the best or second-best results across all prediction horizons. Its superiority is particularly evident in short-term forecasting tasks. For instance, at a prediction length of T = 96, MCANet achieves an MSE of 0.178 and an MAE of 0.296, highlighting its strong capability in modeling complex temporal dependencies.

Table 3. The forecast results.

Metric	Models	MCANet		MSGNet		TimesNet		DLinear		Informer		Autoformer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Flight	96	**0.178**	**0.296**	0.183	0.301	0.237	0.350	0.221	0.337	0.333	0.405	0.204	0.319
	192	**0.187**	**0.303**	0.189	0.306	0.224	0.337	0.220	0.336	0.358	0.421	0.200	0.314
	336	0.203	**0.317**	0.206	0.320	0.289	0.394	0.229	0.342	0.398	0.446	**0.201**	0.318
	720	**0.250**	**0.356**	0.253	0.358	0.310	0.408	0.263	0.366	0.476	0.484	0.345	0.426
Traffic	96	**0.575**	0.330	0.610	0.349	0.593	**0.321**	0.650	0.396	0.719	0.391	0.613	0.388
	192	0.623	0.360	0.634	0.370	0.617	**0.337**	**0.598**	0.370	0.696	0.379	0.616	0.382
	336	0.657	0.383	0.670	0.392	0.629	0.339	**0.605**	0.373	0.777	0.420	0.622	**0.337**
	720	0.709	0.412	0.727	0.422	**0.640**	**0.350**	0.645	0.394	0.864	0.472	0.660	0.408

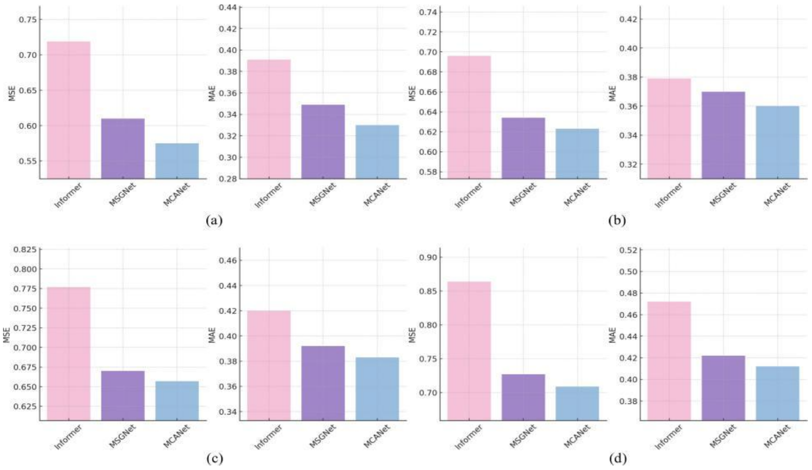


Fig. 5. Performance comparison

On the more complex and volatile Traffic dataset, MCANet also exhibits robust and highly accurate forecasting performance. Although certain baseline models yield slightly lower MSE values at specific prediction steps, MCANet consistently achieves the lowest MAE across all horizons, indicating a significant advantage in error control and overall stability. For example, in the short-term prediction task at T = 96, MCANet yields a markedly lower MAE compared to models such as DLinear and Informer, further validating its effectiveness in capturing short-term traffic fluctuations.

In addition to predictive performance, we further conducted a comparative analysis of training efficiency and resource consumption across different convolutional structures within MCANet. Table 4 presents the GPU memory usage and per-epoch training time under various prediction lengths for two model variants—one employing standard 3×3 convolutions and the other incorporating the proposed PCBG module—on the Flight and Traffic datasets.

The results clearly demonstrate that the PCBG module significantly reduces both GPU memory consumption and training time across all settings. For instance, on the Flight dataset with a prediction length of 96, the standard convolutional model consumed 0.879 GB of GPU memory, while the PCBG-based model reduced this to 0.828 GB, yielding a reduction of approximately 5.8%. Similarly, the training time decreased from 363.863 seconds to 345.826 seconds, representing an improvement of about 5.0This optimization becomes even more pronounced on the Traffic dataset. At a prediction length of 96, the PCBG-based model reduced GPU memory consumption by 18.1%, with usage dropping from

30.078 GB to 24.626 GB. In addition, it shortened the training time from 835.382 seconds to 795.037 seconds, resulting in an approximate 4.8% gain in training efficiency.

Moreover, as the prediction horizon increases, the standard convolutional model exhibits steadily rising memory and computational overheads, whereas the PCBG-based architecture maintains a more stable resource consumption profile. For example, at a prediction length of 720 on the Traffic dataset, the PCBG model utilized only 30.171 GB of GPU memory, compared to 31.019 GB for the standard convolutional model—a reduction of approximately 2.7%—while also decreasing the training time by nearly 29 seconds.

Table 4. Comparison of GPU Memory Usage and Training Time

Dataset	Model	Pred Length	GPU Memory (GB)	Running Time (s/epoch)
Flight	Ours (Conv 3×3)	96	0.879	363.863
		192	0.883	370.765
		336	0.838	358.048
		720	0.887	357.274
	Ours (PCBG)	96	0.828	345.826
		192	0.878	346.951
		336	0.834	340.260
		720	0.885	334.998
Traffic	Ours (Conv 3×3)	96	30.078	835.382
		192	30.921	818.558
		336	30.597	817.671
		720	31.019	832.445
	Ours (PCBG)	96	24.626	795.037
		192	30.537	802.731
		336	30.265	809.565
		720	30.171	803.175

Table 5. Ablation study on the Traffic dataset.

Baseline	LKD-STS	GCSA	PCBG	96		192		336		720	
				MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
✓				0.610	0.349	0.634	0.370	0.670	0.392	0.727	0.422
✓	✓			0.588	0.343	0.632	0.365	0.668	0.388	0.724	0.418
✓		✓		0.584	0.336	0.625	0.360	0.657	0.386	0.710	0.412
✓			✓	0.599	0.350	0.640	0.368	0.670	0.391	0.723	0.419
✓		✓	✓	0.576	0.335	0.627	0.364	0.659	0.387	0.714	0.415
✓	✓		✓	0.589	0.348	0.633	0.367	0.666	0.390	0.718	0.416
✓	✓	✓		**0.574**	0.332	0.625	0.362	0.658	0.386	**0.708**	0.414
✓	✓	✓	✓	0.575	**0.330**	**0.623**	**0.360**	**0.657**	**0.383**	0.709	**0.412**

3.5 Ablation Studies

We conducted ablation studies to examine the contribution of multi-scale convolution and attention mechanisms to the effectiveness of MCANet. In this analysis, we tested the performance of model by removing individual modules and assessing their impact on the public Traffic dataset, as illustrated in Table 5.

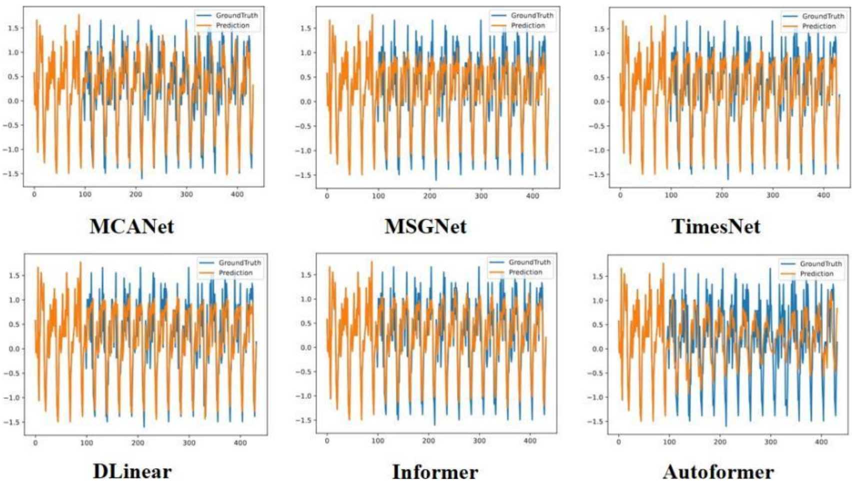
The results demonstrate that each module positively contributes to model performance across different prediction horizons. The incorporation of the LKD-STS module consistently reduced prediction errors for all forecast lengths, particularly in short-term forecasting. For instance, at T=96, the MSE decreased from 0.610 to 0.588, and the MAE decreased from 0.349 to 0.343, indicating that this module effectively enhances the model’s ability to capture temporal dependencies by dynamically adjusting the receptive field. In comparison, the GCSA module yielded even more substantial improvements. At T=96, the MSE and MAE dropped to 0.584 and 0.336, respectively, and further declined to 0.710 and 0.412 at T=720. These results highlight the GCSA module’s notable advan- tage in modeling long-term spatiotemporal channel interactions.

Although the performance gains introduced by the PCBG module were relatively modest, its contribution to computational efficiency was particularly significant. As shown in the resource consumption comparison in Table 4, on the Traffic dataset with a prediction length of 96, the use of the PCBG module reduced GPU memory usage from 30.078 GB to 24.626 GB and shortened training time from 835.382 seconds to 795.037 seconds, representing reductions of 18.1% and 4.8%, respectively. Therefore, the PCBG module demonstrates strong practical value by substantially decreasing computational overhead while maintaining stable performance.

Furthermore, combining multiple modules led to synergistic improvements in prediction accuracy. When both the LKD-STS and GCSA modules were incorporated, the model achieved the lowest MSE at T=96 and T=720, with values of 0.574 and 0.708, respectively, suggesting that these modules offer complementary strengths in temporal modeling and spatial feature integration. However, some fluctuations in the MAE were observed with this combination. Upon further inclusion of the PCBG module, the model maintained competitive performance in terms of MSE and achieved the lowest MAE across all prediction lengths—0.330,0.360, 0.383, and 0.412 at T=96, 192, 336, and 720, respectively—further enhancing the model’s overall stability and prediction accuracy.

3.6 More Showcases

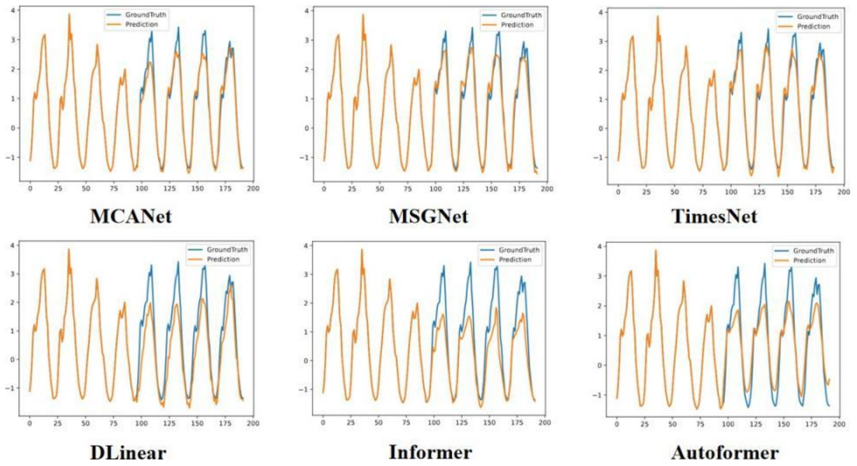
We conduct a series of visualized case studies in Fig. 6 and Fig. 7. Compared to SOTA models, our MSGNet exhibits a superior ability to capture trend variations and periodic patterns, yielding more accurate representations of both the global structure and local fluctuations in the time series.



MCANet MSGNet TimesNet

DLinear Informer Autoformer

Fig. 6. Visualization of the prediction with input length 96 and output length 336. The sequence id is 40.



MCANet MSGNet TimesNet

DLinear Informer Autoformer

Fig. 7. Visualization of the prediction with input length 96 and output length 96. The sequence id is 100.

4 Conclusion

In this paper, we propose a novel traffic flow prediction model, denoted as MCANet. The model enhances its multi-scale feature extraction capability by introducing the LKD-STS module, and further improves its capacity to capture multi-scale traffic characteristics and retain spatial–channel information through the GCSA module. Additionally, we integrate the PCBG module to reduce redundant computations and memory access, thus improving the model's efficiency. Experimental results show that, compared with the baseline and other SOTA prediction models, MCANet achieves lower MSE across various prediction horizons. These findings highlight the pivotal role of multi-scale convolution and attention mechanisms in advancing performance for traffic flow prediction tasks. In the future, we aim to explore more complex traffic scenarios and further optimize and expand our model to handle larger-scale traffic data and more diverse traffic patterns.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (No. 62403076 and 52472399), the Humanities and Social Science Fund of Ministry of Education (No. 24YJCZH416) and Science and Technology Innovative Research Team in Higher Educational Institutions of Hunan Province (New energy intelligent vehicle technology, 2024RC1029).

References

1. Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.: Time series analysis: forecasting and control. John Wiley & Sons (2015)
2. Cai, W., Liang, Y., Liu, X., Feng, J., Wu, Y.: Msgnet: Learning multi-scale interseries correlations for multivariate time series forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 11141–11149 (2024)
3. Chen, J., Kao, S.h., He, H., Zhuo, W., Wen, S., Lee, C.H., Chan, S.H.G.: Run, don't walk: chasing higher flops for faster neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12021 – 12031 (2023)
4. Chen, X., Chen, L., Xie, W., Mueller, N.D., Davis, S.J.: Flight delays due to air pollution in china. Journal of Environmental Economics and Management 119, 102810 (2023)
5. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)
6. Hochreiter, S.: Long short-term memory. Neural Computation MIT-Press (1997)
7. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
8. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
9. Li, J., Wang, Z., Gong, D., Wang, C.: Scnet3d: Rethinking the feature extraction process of pillar-based 3d object detection. IEEE Transactions on Intelligent Transportation Systems (2024)
10. Li, K., Geng, Q., Zhou, Z.: Exploring scale-aware features for real-time semantic segmentation of street scenes. IEEE Transactions on Intelligent Transportation Systems (2023)
11. Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., Long, M.: Timesnet: Temporal 2d-variation modeling for general time series analysis. arXiv preprint arXiv:2210.02186 (2022)

12. Wu, H., Xu, J., Wang, J., Long, M.: Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. Advances in neural information processing systems 34, 22419–22430 (2021)
13. Wu, W., Zhang, Y., Wang, D., Lei, Y.: Sk-net: Deep learning on point cloud via end-to-end discovery of spatial keypoints. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 6422–6429 (2020)
14. Yang, L., Zhong, J., Zhang, Y., Bai, S., Li, G., Yang, Y., Zhang, J.: An improving faster-rcnn with multi-attention resnet for small target detection in intelligent autonomous transport with 6g. IEEE Transactions on Intelligent Transportation Systems 24(7), 7717–7725 (2022)
15. Ying, L., Miao, D., Zhang, Z.: 3wm-augnet: A feature augmentation network for remote sensing ship detection based on three-way decisions and multi-granularity. IEEE Transactions on Geoscience and Remote Sensing (2023)
16. Zeng, A., Chen, M., Zhang, L., Xu, Q.: Are transformers effective for time series forecasting? In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 11121–11128 (2023)
17. Zhan, Y., Liu, J., Ou-Yang, L.: scmic: a deep multi-level information fusion framework for clustering single-cell multi-omics data. IEEE Journal of Biomedical and Health Informatics (2023)
18. Zhang, X., Cao, X., Zhang, H., Shen, Y., Yuan, X., Cui, Z., Lu, Z.: An intelligent obstacle detection for autonomous mining transportation with electric locomotive via cellular vehicle-to-everything and vehicular edge computing. IEEE Transactions on Intelligent Transportation Systems (2023)
19. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond efficient transformer for long sequence time-series forecasting. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 11106–11115 (2021)
20. Zhang, J., Meng, Y., Wei, J., Chen, J., & Qin, J.: A novel hybrid deep learning model for sugar price forecasting based on time series decomposition. Mathematical Problems in Engineering, 6507688.(2021).
21. Zhang, J., Meng, Y., Wu, J., Qin, J., Yao, T., & Yu, S.: Monitoring sugar crystallization with deep neural networks. Journal of Food Engineering, vol. 280, 109965.(2020).
22. Wu, X., Meng, Y., Zhang, J., Wei, J., & Zhai, X.: Amodal segmentation of cane sugar crystal via deep neural networks. Journal of Food Engineering, vol 348, 111435.(2023).
23. Lu, G., He, D., & Zhang, J. Energy-saving optimization method of urban rail transit based on improved differential evolution algorithm. Sensors, vol 23, 378. (2022).
24. Wu, J., Zhang, J., Zhu, J., Wang, F., Si, B., Huang, Y., ... & Meng, Y.: Lightweight peach detection using partial convolution and improved Non-maximum suppression. Journal of Visual Communication and Image Representation, 104495. (2025).
25. Du, R., Feng, R., Gao, K., Zhang, J., Liu, L.: Self-supervised point cloud prediction for autonomous driving. IEEE Transactions on Intelligent Transportation Systems. (2024).
26. Duan, Y., Meng, L., Meng, Y., Zhu, J., Zhang, J., Zhang, J., & Liu, X. MFSA-Net: Semantic Segmentation With Camera-LiDAR Cross-Attention Fusion Based on Fast Neighbor Feature Aggregation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. (2024).
27. Gao, K., Li, X., Hu, L., Liu, X., Zhang, J., Du, R., & Li, Y. STMF-IE: A Spatial-Temporal Multi-Feature Fusion and Intention-Enlightened Decoding Model for Vehicle Trajectory Prediction. IEEE Transactions on Vehicular Technology. (2024).
28. Wu, J., Zhang, J., Zhu, J., Duan, Y., Fang, Y., Zhu, J., ... & Meng, Y. Multi-scale convolution and dynamic task interaction detection head for efficient lightweight plum detection. Food and Bioproducts Processing, 149, 353-367. (2025).