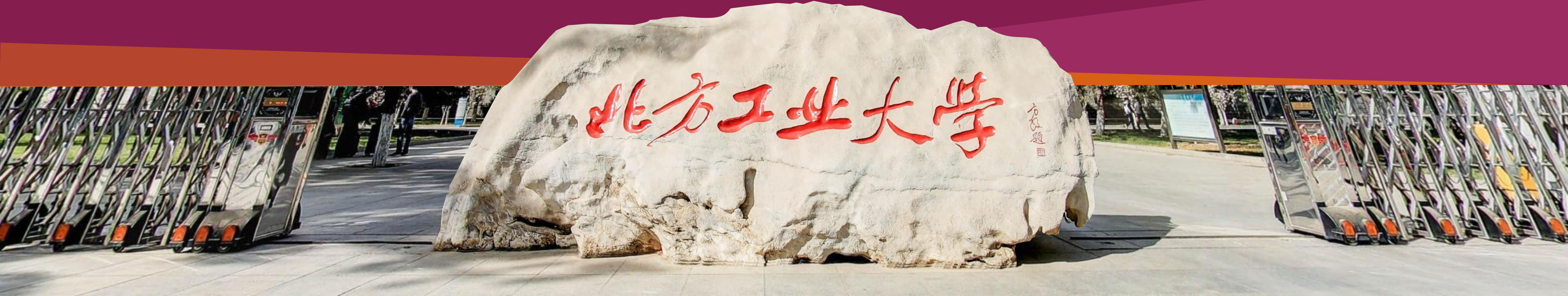
North China University of Technology
Beijing Key Laboratory on Integration and Analysis of Large-Scale Stream Data

An Evolving Transformer Network based on Hybrid Dilated Convolution for Traffic Flow Prediction

Qi Yu，Weilong Ding*，Maoxiang Sun，Hongmin Cai



Introduction

Background

- Traffic congestion is a pain point on many big city highways
- Road sensors capture massive and complex traffic data
- Relying on traffic data for traffic flow prediction can help alleviate congestion

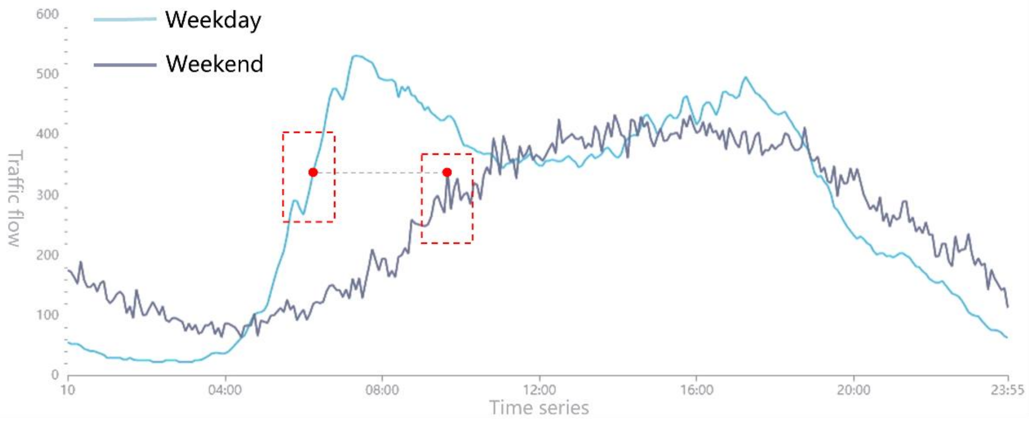
Value

- Advanced traffic management
- Optimized route planning
- Road construction and project design

Challenges

Dynamic Temporal Relationships

- **Efficiently Extracting Relevance**
  - the evolving attention mechanism
- **Influence of Surrounding Context**
  - a novel convolutional embedding layer
- **Periodicity**
  - periodic positional embedding



Related Work

Traffic Flow Prediction

- **statistical methods**
  - HA, ARIMA
- **traditional machine learning methods**
  - KNN, SVR
- **deep learning methods**
  - LSTM [1], LSTM-BILSTM [2], G-CNN [3]

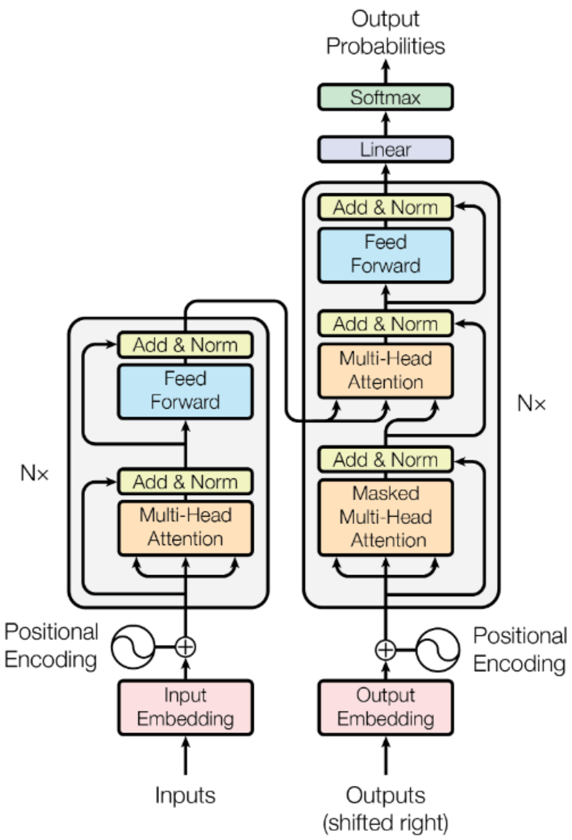
[1] Williams, R., Hochreiter, S., Schmidhuber, J.: Long short-term memory.

[2] Ma, C., Dai, G., Zhou, J.: Short-term traffic flow prediction for urban road sections based on time series analysis and lstm_bilstm method. IEEE Transactions on Intelligent Transportation Systems p. 5615–5624.

[3] Yi, S., Ju, J., Yoon, M.K., Choi, J.: Grouped convolutional neural networks for multivariate time series.

Related Work

Transformer Networks



- **for traffic flow prediction**
  - Traffic transformer
- **improve important modules**
  - Informer [4], Pyraformer
- **design a new architecture**
  - Autoformer [5], Scaleformer

[4] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond efficient transformer for long sequence time-series forecasting. Proceedings of the AAAI Conference on Artificial Intelligence p. 11106–11115.

[5] Wu, H., Xu, J., Wang, J., Long, M.: Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting.

Preliminary

Traffic Flow Tensor

- **all nodes over the total T time slices**
  - $\{X_1, \cdots, X_n, \cdots, X_N\} \in \mathbb{R}^{N \times T}$
- **node n at the last T time steps**
  - $X_n = \{x_1, \cdots, x_t, \cdots, x_T\} \in \mathbb{R}^{T}$

Preliminary

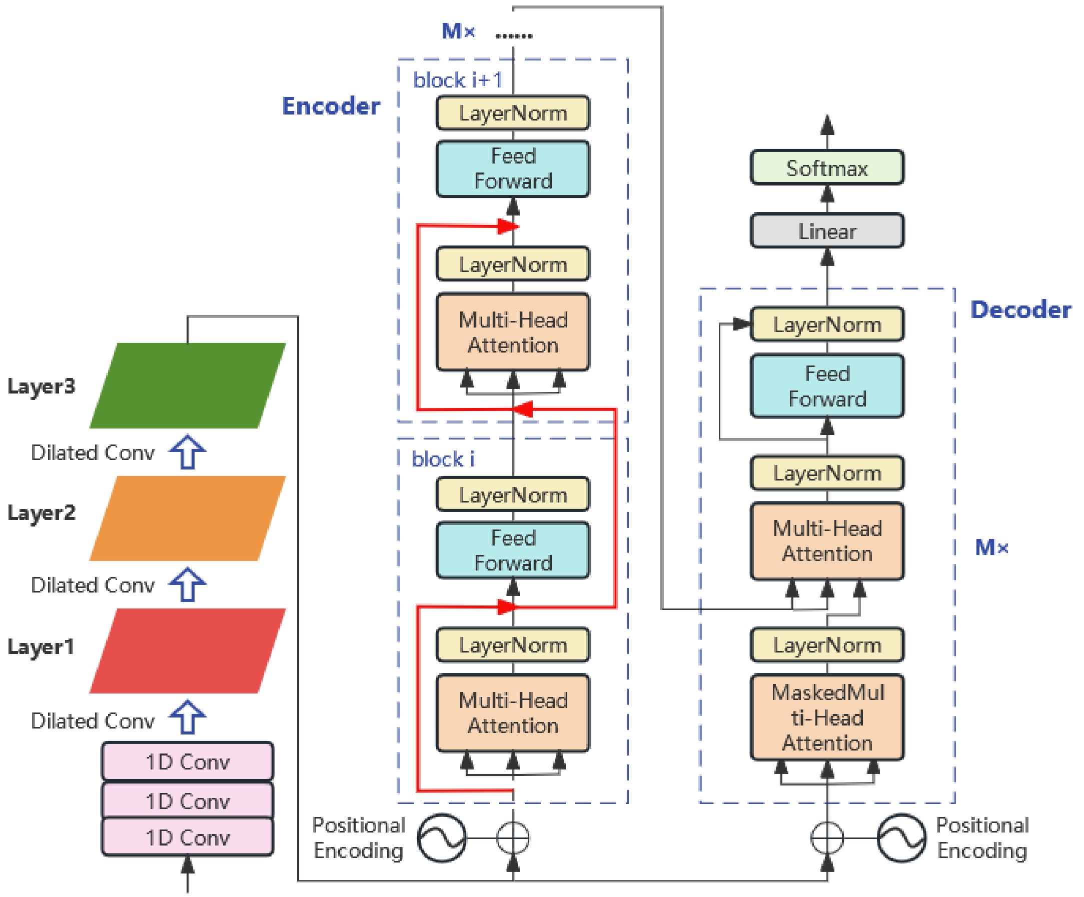
Traffic Flow Prediction Problem

- By fitting a complex function $\tilde{f}$, traffic values for the coming P time steps can be forecasted based on traffic data from N nodes over the previous T time steps.

$$[x_{T+1}, \cdots, x_{T+p}, \cdots, x_{T+P}] = \tilde{f}([x_1, \cdots, x_t, \cdots, x_T; \theta]).$$

Methodology

An Evolving Transformer Network based on Hybrid Dilation Convolution



Methodology

Hybrid Dilated Convolution For Data Embedding

Local relevance of time series

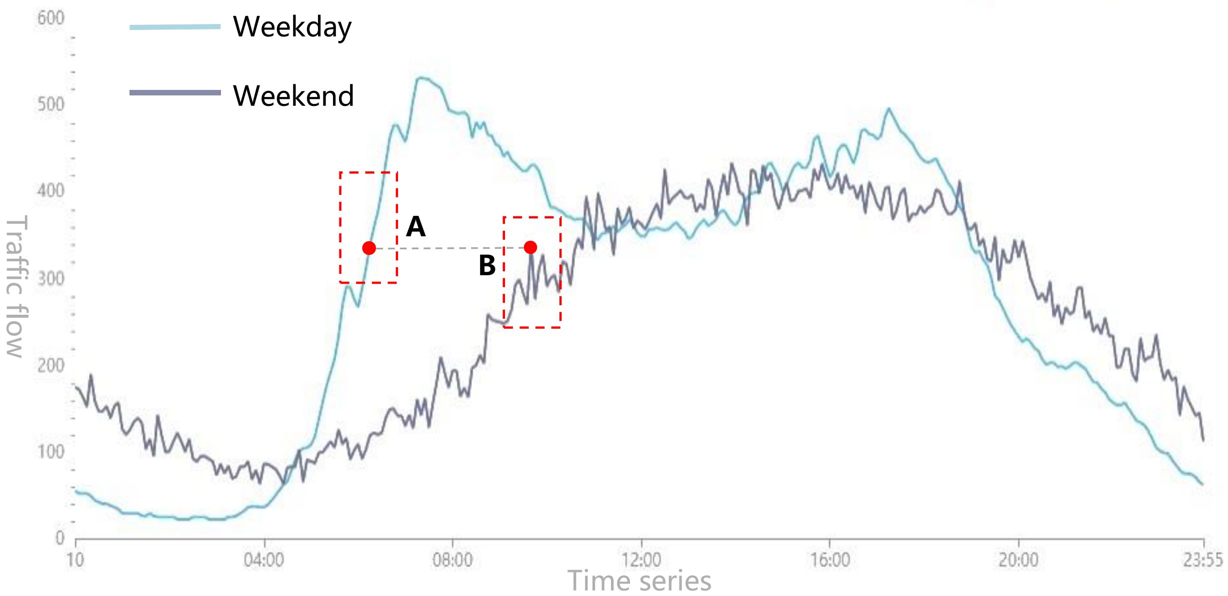
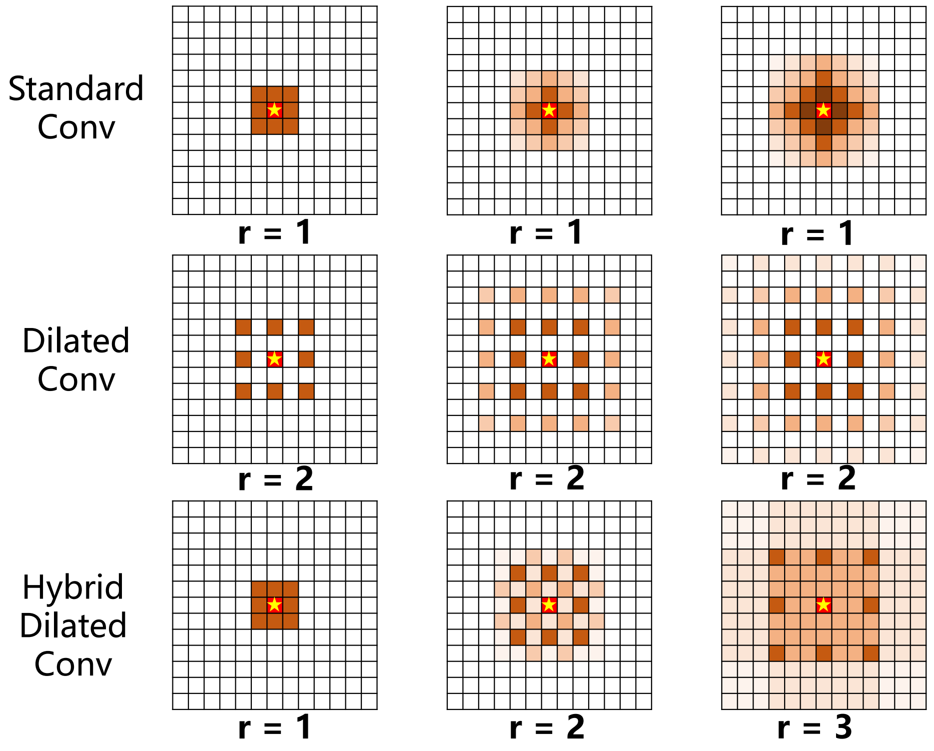


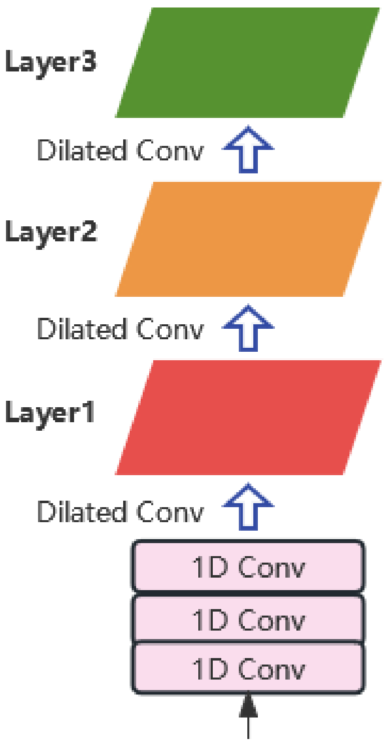
Fig. 1. shows time series on PeMSD4 dataset, where two curves represent the traffic flow of a node on a weekday and a weekend respectively

- Points A and B have the same value which is just the median within a one hour, but they have completely different fluctuation trends in subsequent time steps.

Hybrid Dilated Convolution For Data Embedding

Hybrid Dilated Convolution





$$X_{emb} \in \mathbb{R}^{B \times T_{in} \times D_{emb}}$$

$$\Big\uparrow \; HDC(\cdot)$$

$$X_{input} \in \mathbb{R}^{B \times T_{in} \times D_{in}}$$

Methodology

Positional Encoding

◆ **Periodic positional embedding**

- $$\begin{cases} X_{pos}(2t) & = sin(\frac{2\pi t}{\boxed{period}}) \\ X_{pos}(2t+1) & = cos(\frac{2\pi t}{\boxed{period}}) \end{cases}$$

- $$\mathcal{X} = Concat(X_{pos}, X_{emb})$$

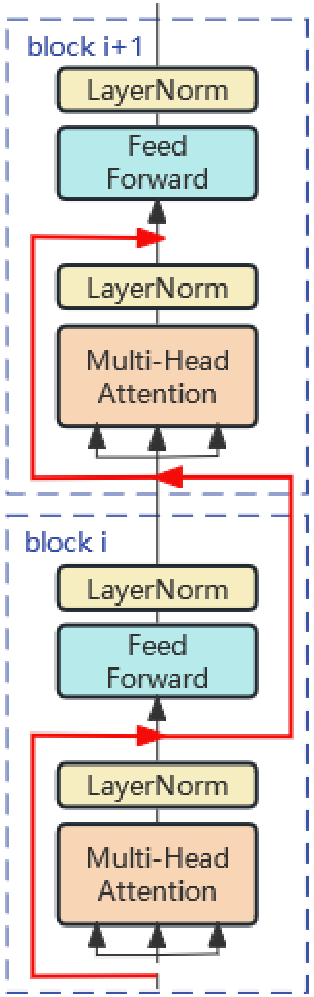
Methodology

Evolving Transformer

The Evolving Mechanism

- Add residual connections between adjacent encoder blocks

- $$\mathcal{X}_{in}^{i} = \alpha \cdot \boxed{\mathcal{X}_{res}^{i-1}} + (1-\alpha) \cdot \mathcal{X}_{out}^{i-1},$$
  $$\mathcal{X}_{res}^{i} = \beta \cdot \boxed{\mathcal{X}_{in}^{i}} + (1-\beta) \cdot Attention(\mathcal{X}_{in}^{i}),$$
  $$\mathcal{X}_{out}^{i} = LayerNorm(FeedForward(\mathcal{X}_{res}^{i})).$$



Evolving Transformer

Multi-Head Attention

- The Scaled Dot-Product Attention model
- $Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{D_{pe} + D_{emb}}})V.$

$$head_h = Attention(Q_h, K_h, V_h),$$

$$Q_h = QW_q^h, K_h = KW_k^h, V_h = VW_v^h, h = 1, \cdots, H.$$

Feed Forward Networks

- $FeedForward(\mathcal{X}_{res}) = max(a_r\mathcal{X}_{res}, \mathcal{X}_{res})W_r + b_r.$

Evaluation

Dataset

- **PeMSD4**
- **PeMSD8**
  - Data aggregated at 5-minute intervals, i.e., 12 sample points per hour.

Datasets	Nodes	Time Interval	Timesteps	Time Range
PeMSD4	307	5min	16992	1/1/2018-2/28/2018
PeMSD8	170	5min	17856	7/1/2016-8/31/2016

Evaluation

Baseline

- HA: History Average Model
- ARIMA: Autoregressive Integrated Moving Average Model
- KNN: K-Nearest Neighbor Model
- SVR: Support Vector Regression Model
- LSTM [1]: Long Short-Term Memory Model
- GRU [6]: Gate Recurrent Unit Model
- RPConvformer [7]: A novel Transformer-based deep neural network for traffic flow prediction

Evaluation

Setting

- **Training : Validation : Test** = 6 : 2 : 2
- **Goal:** Predicting the next hour's data using the past day's data.
- **Evaluation Metrics**
  - $$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|,$$
  - $$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2},$$
  - $$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{n}\frac{\left|y_i - \hat{y}_i\right|}{y_i} \times 100\%$$

Result Analysis

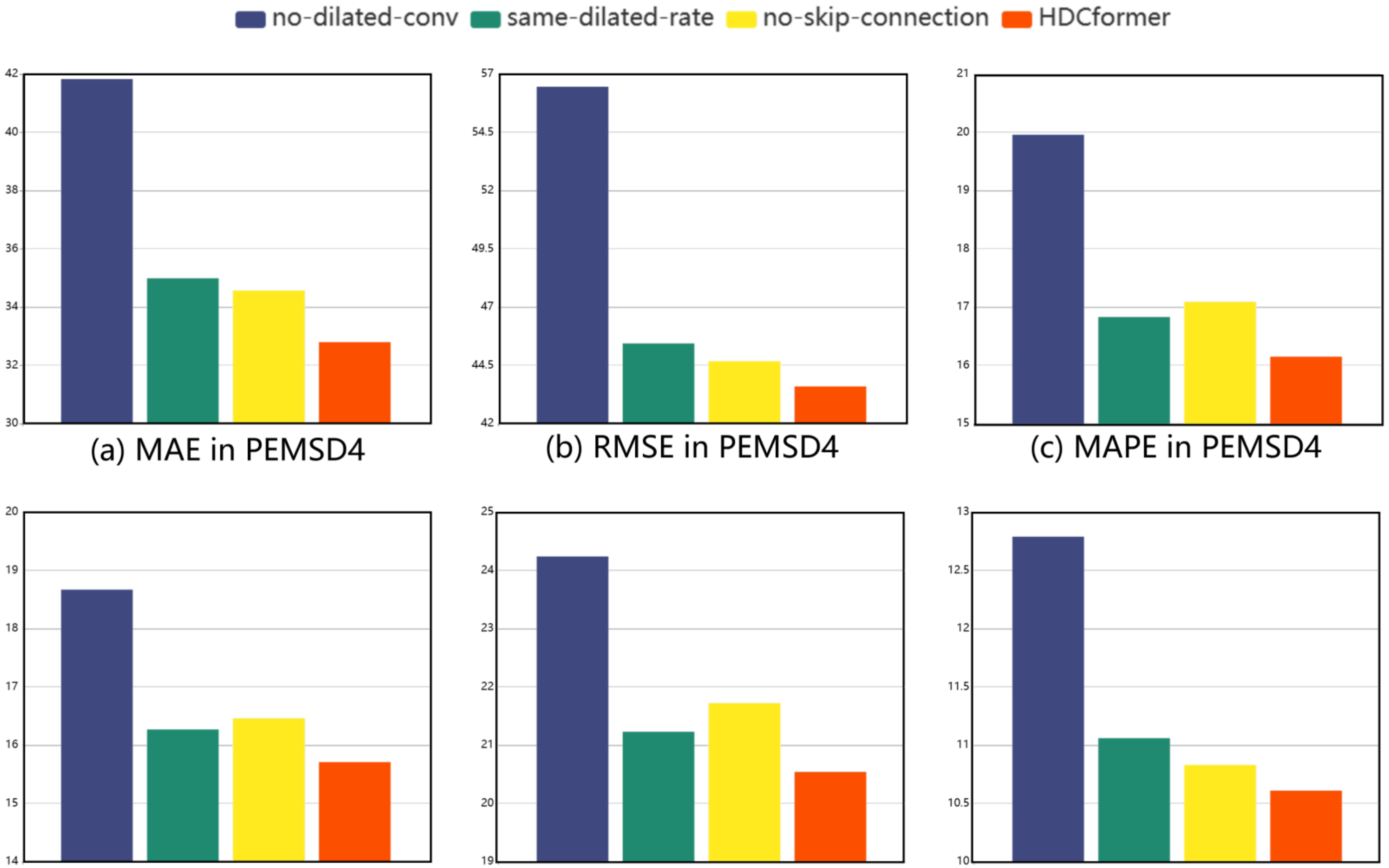
Performance Comparison

Model	PeMSD4			PeMSD8		
	MAE	RMSE	MAPE(%)	MAE	RMSE	MAPE(%)
HA	47.17	70.14	22.98	28.46	36.3	25.92
ARIMA	64.34	84.20	36.93	30.00	38.22	27.76
KNN	52.86	72.25	26.10	22.49	29.85	18.65
SVR	53.81	71.48	29.02	21.54	27.55	19.50
LSTM	38.50	52.06	19.23	19.75	25.96	16.96
GRU	39.78	52.25	22.52	20.19	26.68	17.01
RPConvformer	35.8	47.5	17.51	16.15	21.08	11.02
HDCformer(ours)	**32.80**	**43.60**	**16.15**	**15.71**	**20.54**	**10.61**

- PeMSD4 ↓8.12%
- PeMSD8 ↓3.00%

Result Analysis

Ablation Experiments



(a) MAE in PEMSD4 (b) RMSE in PEMSD4 (c) MAPE in PEMSD4

(d) MAE in PEMSD8 (e) RMSE in PEMSD8 (f) MAPE in PEMSD8

Reference

[1] Williams, R., Hochreiter, S., Schmidhuber, J.: Long short-term memory.

[2] Ma, C., Dai, G., Zhou, J.: Short-term traffic flow prediction for urban road sections based on time series analysis and lstm_bilstm method. IEEE Transactions on Intelligent Transportation Systems p. 5615–5624.

[3] Yi, S., Ju, J., Yoon, M.K., Choi, J.: Grouped convolutional neural networks for multivariate time series.

[4] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond efficient transformer for long sequence time-series forecasting. Proceedings of the AAAI Conference on Artificial Intelligence p. 11106–11115.

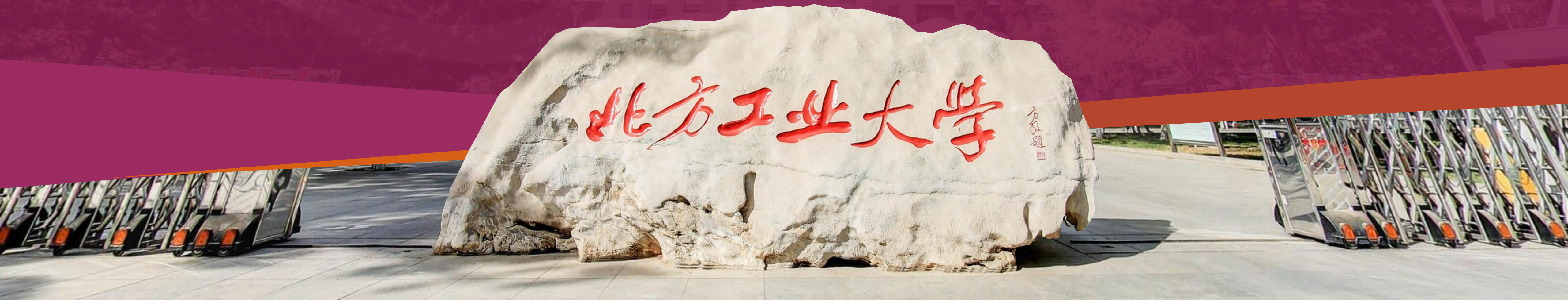
[5] Wu, H., Xu, J., Wang, J., Long, M.: Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting.

[6] Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling (Dec 2014).

[7] Wen, Y., Xu, P., Li, Z., Xu, W., Wang, X.: Rpconvformer: A novel transformerbased deep neural networks for traffic flow prediction (Jan 2023).

North China University of Technology
Beijing Key Laboratory on Integration and Analysis of Large-Scale Stream Data

Thanks!



EAI CollaborateCom

An Evolving Transformer Network based on Hybrid Dilated Convolution for Traffic Flow Prediction

Qi Yu, Weilong Ding*,
Maoxiang Sun, Hongmin Cai