长期Transformer和自适应傅里叶变换的动态图卷积交通流预测研究

张 红* 伊 敏 张玺君 李 扬 张鹏程
(兰州理工大学计算机与通信学院 兰州 730050)

摘 要：针对交通流长期趋势性与非平稳性不易有效建模，以及交通流的隐藏动态时空特征难以捕捉的问题，该文提出一种基于长期Transformer和自适应傅里叶变换的动态图卷积交通流预测模型(ADGformer)。其中，长期门控卷积层通过掩码子序列Transformer从长历史序列中学习压缩的、上下文信息丰富的子序列表示，并利用膨胀门控卷积从子序列的时间表示中有效捕捉交通流的长期趋势特征。并设计一种动态图构造器生成动态可学习图，并利用可学习动态图卷积对节点间潜在的和时变的空间依赖关系进行建模以有效捕获交通流的动态隐藏空间特征。其次，自适应频谱块利用傅里叶变换来增强特征表示并捕获长短期的交互作用，同时通过自适应阈值处理来降低交通流的非平稳性。实验结果表明，所提ADGformer模型具有较好的预测性能。

关键词：交通流预测；动态图卷积；长期Transformer；傅里叶变换；非平稳性

中图分类号：TN911.7 **文献标识码**：A **文章编号**：1009-5896(2025)07-2249-14

DOI: 10.11999/JEIT241076 **CSTR**: 32379.14.JEIT241076

1 引言

近年来，智能交通系统(Intelligent Transportation System, ITS)的快速发展使得人口数量和机动车保有量不断增加，交通系统的道路容量不足、交通堵塞等问题刻不容缓。因此，针对交通路网路况复杂动态变化、流量分布失衡等问题，交通路网系统的交通流预测已然成为现阶段研究热点。准确实时的交通流预测为交通管理提供有效的决策支持，以有效缓解交通拥堵并实现绿色低碳出行[1]。

交通流预测具有复杂的空间依赖性和非线性的时间关系，传统的交通流预测方法自回归积分移动平均法(Auto Regressive Integrated Moving Average, ARIMA)[2]和向量自回归(Vector Autoregressive, VAR)[3]等需对交通数据进行特定的假设，难以捕捉复杂的非线性时间相关性。随着深度学习的发展，研究人员通常使用图神经网络(Graph Neural Networks, GNNs)[4]来提取道路节点之间的空间相关性，并使用递归神经网络(Recurrent Neural Network, RNN)[5]或时间卷积网络(Temporal Convolutional Network, TCNs)[6]来捕获时间相关性。然而，这类方法难以表达具有复杂拓扑结构的交通路网，不能准确地挖掘交通流量的时间演化特性。针对这一问题，研究人员引入图卷积网络(Graph Convolutional Network, GCN)，聚合相邻节点的信息，以进一步结合有关交通网络的先验知识并捕获复杂的时空特征。例如，Liu等人[7]提出了时空双自适应图卷积网络模型(SpatioTemporal Dual Adaptive Graph Convolutional Network model, ST-DAGCN)，利用双自适应邻接矩阵来捕获空间相关性。现有大多数基于GCN模型利用预定义静态图来描述交通节点之间的空间相关性，忽略了路网中的节点之间相似的功能特性以及动态交通模式的影响，无法有效捕捉其动态变化特性。为解决该问题，动态图卷积网络通过构造动态图充分考虑交通流数据的周期性和动态性以更好地捕捉时空特征。然而，现有大多数基于动态图结构的模型无法有效地捕获交通路段间的动态隐藏空间关系，那么深入、充分地挖掘交通系统潜在的模式和动力学以动态地对路段的空间依赖关系进行有效建模是一关键问题。

随着时间步长的增加，交通序列间的长期依赖关系难以有效捕捉，绝大多数模型仅限于使用短历史序列作为模型输入，但短期数据中的信息通常不足以充分表示历史交通数据中存在的丰富的时间特征，交通流长历史序列的趋势性和周期性特征有助于模型对交通流复杂趋势的判断。Transformer是一种基于自注意机制的神经网络体系结构, 相比于传统的深度学习模型，具有更好的并行性能和更短的训练时间，在交通流预测问题上得到显著效果。

收稿日期：2024-12-05；改回日期：2025-05-25；网络出版：2025-06-07
*通信作者： 张红 zhanghong@lut.edu.cn
基金项目：甘肃省重点研发项目(23YFGA0063)，国家自然科学基金(62363022, 61663021)，甘肃省自然科学基金(22JR5RA226, 23JRRA886)，甘肃省教育厅产业支撑计划(2023CYZC-35)
Foundation Items: The Key R&D Program of Gansu Province (23YFGA0063), The National Natural Science Foundation of China (62363022,61663021),The Natural Science Foundation of Gansu Province(22JR5RA226, 23JRRA886), Gansu Provincial Department of Education: Industrial Support Plan Project (2023CYZC-35)

由于Transformer[8]通过其编码器将输入交通序列编码成一个高维特征向量表示，自注意力机制使得模型能够同时考虑交通序列中所有位置的信息，使得其在长期交通流预测任务中表现突出，不仅能够解决RNN及其变体存在的长距离依赖问题，还支持并行化计算，进而提高模型的预测性能。然而，Transformer模型在处理输入序列时，通常会受限于其可处理的序列长度，无法有效处理关键的上下文信息。因此，如何利用Transformer有效地学习长历史序列中的关键时间信息和长期趋势特征是一重要问题。由于Transformer架构都是复杂多变的，无法有效地处理复杂的、有噪声的交通流序列，傅里叶变换方法[9]可用于获取经过消噪的历史交通流量测量值，以减少局部噪声对构造的同化模型的影响，并提高同化结果的准确性。故而，如何利用傅里叶变换在降低交通数据的非平稳性的同时捕获数据中的长期和短期相互作用是一关键问题。

针对上述问题，本文提出一种基于长期Transformer和自适应傅里叶变换的动态图卷积交通流预测模型(long-term transformer and Adaptive fourier transform for Dynamic Graph convolutional traffic flow prediction model, ADGformer)，ADGformer不仅对隐藏的空间关系进行了动态建模，还在有效捕获了交通流的长短期依赖关系的同时降低了交通数据的非平稳性。主要贡献如下：

(1)构建了长期门控卷积层，利用掩码子序列(Masked Sub-series) Transformer以预训练的方式从一小部分未掩码子序列及其时间上下文中推断出掩码子序列的内容，使得模型从长历史序列中有效地学习与上下文相关的子序列时间表示；通过结合门控机制和1D膨胀卷积的膨胀门控卷积从子序列的时间表示中提取交通流的长期趋势特征，以有效提高模型的长期预测能力。

(2)建立了空间卷积层，根据动态空间关系传播节点的隐藏状态设计了一种动态图构造器以生成动态可学习图，并提出一种可学习动态图卷积通过动态邻接矩阵上的消息传递，将邻居节点的隐藏状态聚合到焦点节点以有效地捕获交通流的动态隐藏空间特征。

(3)设计了一种基于傅里叶变换的自适应频谱块，对交通流的全局频域信息进行有效地学习，并且利用自适应全局和局部滤波器减轻噪声以降低交通流数据序列的非平稳性，有效地捕获交通数据中的长期和短期相互作用。

(4)在公开的3个交通数据集上做了大量的对比试验，实验结果表明所建立的模型比现有流行的模型具有更好的预测效果，证明了ADGformer模型的有效性和准确性。

2 相关工作

2.1 基于图卷积网络的交通流预测

早期深度学习模型将交通路网划分为规则网络以构建不同区域的欧式空间结构，并使用卷积神经网络捕获空间相关性。如Li等人[10]提出了深度时空自适应3D卷积神经网络(deep Spatio-Temporal Adaptive 3D convolutional neural Networks, ST-A3DNet)，构建自适应3D卷积模块同时捕捉时空关系。由于现实中的交通路网结构是复杂的非欧氏拓扑结构，CNN无法有效提取交通流的空间依赖特性。图卷积网络的出现将非欧氏结构的交通图数据应用于卷积操作中，以有效地捕捉路网拓扑结构特征。Chen等人[11]设计了基于节点连接强度矩阵的图卷积网络来预测节点的流量；Zhang等人[12]主要利用时空注意力(Spatio-Temporal Attention Block, STA-Block)、GCN和标准卷积网络，分别对交通流的周期性、空间相关性和时间依赖性进行建模；Ling等人[13]采用具有残差图注意力的耦合图卷积网络来动态学习交通站点之间和内部的不同空间特征。然而，上述方法主要依赖于预定义的静态图结构，无法有效学习节点间潜在的动态相关性信息。故而，动态图卷积网络成功地将动态图结构应用于GCN中。Wang等人[14]引入自适应动态拓扑图模块和曼-肯德尔(Mann-Kendall, MK)时间预测模块，有效解决了交通流的自相似性和空间异质性特征；Li等人[15]通过特征增强的下采样动态图卷积网络更好地模拟节点之间的动态关系。邓涵优等人[16]利用卷积网络和动态图卷积网络交互学习及交通流奇偶子序列特征交互融合，进而同时学习交通流序列的局部时空相关性和全局时空相关性。由于大多数基于动态图卷积的方法主要利用路段距离、兴趣点(Point Of Interesting, POI)相似度等先验知识构建相关的交通图结构以捕获道路之间的空间依赖关系，无法有效地动态捕捉交通路网间的空间相关性。因此，如何在不依赖先验知识的情况下动态地捕捉道路之间的空间依赖性成为一大研究热点。Peng等人[17]通过构建交通流概率图，并利用强化学习策略生成动态序列图，以提高模型长期预测能力。唐晨嘉等人[18]提出了一种自适应动态图时空预测模型，将节点嵌入与时间编码结合为时空编码构建动态邻接图以表示节点关系的时间演化。蒋挺等人[19]构建了一种新的基于输入交通信号分解的动态时空融合的交通预测模型(Decomposed Dynamic

Spatial-Temporal Graph Convolutional Network, DDSTGCN)，并提出了动态图学习模块，考虑了空间依赖的动态性质。然而，以上模型未充分考虑交通路网中节点间隐藏的动态时空相关性。

2.2 基于Transformer的交通流预测

Transformer模型是序列到序列任务的高效解决方法，是最早完全依赖于自注意力机制来计算其输入和输出的表示的转换模型，不仅实现时间序列之间信息的交互，还实现了并行化训练。因此，Transformer逐渐被应用于交通流预测中并取得优异的效果。Wei等人[20]利用图扩散和Transformer来捕获交通流时空模式的复杂性和动态性。时空融合嵌入Transformer(Spatial–Temporal Fusion Embedding Transformer, STFEformer)[21]借鉴了Transformer模型的思想，通过关注相似节点之间的交互来增强动态和远程空间相关性的检测。研究者表明Transformer在相比于循环神经网络和卷积神经网络，Transformer更适合对长距离依赖进行建模，且模型更具可解释性，在性能上实现进一步的提高。Zhu等人[22]结合图卷积注意网络(Graph Convolutional ATtention network, GCAT)和Transformer有效预测未来的单步、短期和长期多步交通流。Transformer增强型周期性时间卷积网络(Transformer-Enhanced periodic Temporal Convolution Network, TE-TCN)[23]集成 Transformer多头注意力机制和门控循环单元(Gated Recurrent Unit, GRU)有效学习长期周期性模式，并应用两个并行的时间卷积网络捕获短期周期依赖性。为了有效提取交通数据的长期时间相关性，Zhang等人[24]构建了一种基于Transformer和鲸鱼优化算法的自适应图卷积网络。在保证模型参数可以并行训练的同时，以上模型能够实现输入序列的各个时间步长之间的直接连接，而不依赖于RNN的线性链结构或CNNs中的感受野以更好地捕获交通数据中的长期依赖性。然而，上述方法仅使用较短的历史时间序列来进行预测，未充分考虑长历史序列特有的长期趋势特征和非平稳性。

2.3 基于非平稳性的交通流预测

交通流序列数据具有非平稳性和时间依赖性的显著特性，Transformer模型虽在交通流预测方面取得了显著成绩，但由于其复杂多变的网络架构，容易出现过拟合和计算效率低的问题，忽略了交通流序列数据具有非平稳性的特性，导致模型预测性能降低。因此，有效处理复杂的、有噪声的历史序列数据已然成为交通流预测的一个研究热点。Zang等人[25]提出了一种基于迁移学习的方法实现目标领域数据稀缺性的精确非平稳预测。Yin等人[26]利用变分模态分解(Variational Mode Decomposition, VMD)算法将交通流序列分解为一系列相对平稳的模态分量，以降低其短期非平稳性。Huang等人[27]构建了一种由 VMD, CNN和GRU相结合的混合预测模型，将原始时间序列分解为K个分量进行VMD处理，去除噪声部分，提高建模精度。以上预测方法虽有效地降低了交通流的非平稳性，却没有从局部全局的角度充分考虑交通流的空间非平稳性。Sun等人[28]设计了一种全局时空动态捕获网络(Global Spatio-Temporal Dynamic Capturing Network, GSTDCN)，有效地模拟交通流的全局空间依赖性、局部空间非平稳性和时间周期性。然而，尽管这些方法可以很好地处理短期交通序列，不能处理复杂的、有噪声的长期时间序列。

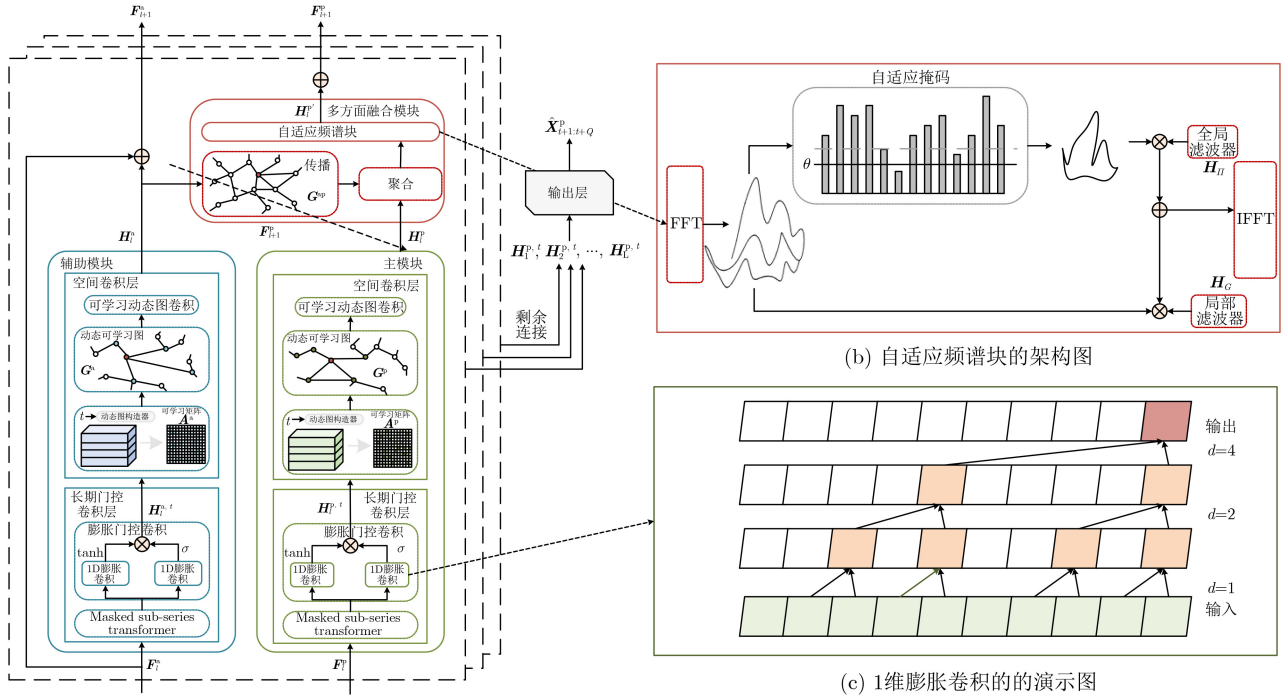
3 模型结构

所提ADGformer模型总体框架如图1(a)所示，主要由长期门控卷积层、空间卷积层和多方面融合模块堆叠而成。其中，长期门控卷积层利用由子序列时间表征学习器(Subsequent Temporal Representation Learner, STRL)和自监督任务头组成的Masked Sub-series Transformer有效地从长时间序列中学习压缩的、上下文信息丰富的子序列表示，通过巧妙结合1D膨胀卷积和门控机制的膨胀门控卷积模块从子序列的时间表示中提取交通流的长期趋势特征。空间卷积层依据不同的节点之间的动态隐藏关系构建一个动态图构造器生成动态可学习图，并运用于图卷积构成可学习动态图以有效地捕获交通流的动态隐藏空间特征。多方面融合模块根据“传播和聚合”的思想运用可学习动态图卷积将交通流的辅助特征和主特征进行有效融合，并利用自适应频谱块有效学习交通流全局信息且通过自适应全局和局部滤波器降低交通流数据序列的非平稳性。

3.1 问题定义

定义1(交通路网) 交通路网定义为图$\boldsymbol{G}=(V,E,\boldsymbol{A})$，节点集$V(|V|=N)$表示为交通网络中的传感器或道路交叉口，根据传感器的空间距离或道路的拓扑关系可以得到边集E。在t时刻时，交通路网为图$\boldsymbol{G}_t=(V,E_t,\boldsymbol{A}_t)$。

定义2(交通特征) 由于交通特征包括交通速度特征和交通流量特征等，且这些交通特征间存在潜在的相关性。故而，将历史交通速度$\boldsymbol{X}_{\mathrm{p}}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_{N_{\mathrm{p}}})$作为主要特征，将历史交通流量$\boldsymbol{X}_{\mathrm{a}}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_{N_{\mathrm{a}}})$作为辅助特征。其中，$N_{\mathrm{p}}$，$N_{\mathrm{a}}$表示为路段



(a) ADGformer模型总体框架

(b) 自适应频谱块的架构图

(c) 1维膨胀卷积的的演示图

图 1 模型框架图

数。$\boldsymbol{X}_t^{\mathrm{p}} \in R^{N_{\mathrm{p}}}$, $\boldsymbol{X}_t^{\mathrm{a}} \in R^{N_{\mathrm{a}}}$分别表示所有路段在$t$时刻的交通流观测值和辅助特征。

定义3(交通流预测) 交通流预测的目的是给定历史交通流数据从而预测未来一段时间内所有路段的交通流。通过给定历史步长为P的交通流特征，建立模型f来预测未来Q个时间步长的交通流特征，即$\hat{\boldsymbol{X}}_{t+1:t+Q}^{\mathrm{p}} = f(\boldsymbol{X}_{t-P+1:t}^{\mathrm{p}}, \boldsymbol{X}_{t-P+1:t}^{\mathrm{a}})$。

3.2 长期门控卷积层

长期门控卷积层(Long-term Gated Convolutional layers, LGConv)主要由Masked Sub-series Transformer和膨胀门控卷积(Dilated Gated Convolution, DGConv)组成。其中，Transformer模型通过掩码重建任务进行预训练，以获得压缩的和上下文相关的子序列时间表示。DGConv利用1维膨胀卷积和门控机制从所得子序列表示中有效地提取长期趋势特征。

3.2.1 Masked Sub-series Transformer

Masked Sub-series Transformer从少量子序列及其时间上下文推断掩码子序列的内容，使得模型可以有效地从长时间序列学习压缩的、上下文信息丰富的子序列表示。

(1)掩码策略：Masked Sub-series Transformer的掩码策略主要考虑两个重要因素，即掩码的基本单位和掩码比率。现有的交通流预测方法通常以5 min作为输入交通数据的基本单位，无法有效地捕获长时间序列的趋势特征。故而，所提模型将长时间历史序列划分为包含多个时间步长的等长子序列，并将子序列作为模型输入的基本单位。

在自然语言处理中，词的语义信息更加丰富，冗余部分较少且信息密度高，故而BERT模型所设置的掩码比率较低，即15%。图像数据的信息密度相对较低，像素点具有空间连续性，因此MAE使用75%的掩码率。由于长期交通流数据类似于具有时间连续性和低信息密度的图像，需要较高的掩蔽比，故而采用75%的随机掩码。

(2)表征学习模型：如图2所示，Masked Sub-series Transformer主要由子序列时间表征学习器(STRL)和自监督任务头两部分组成。STRL利用Transformer编码器学习子序列的时间表示，自监督任务头根据未掩码子序列和掩码标记的时间表示重构完整的长序列。首先将长历史交通流序列$\boldsymbol{F}_{\mathrm{l}} = [\boldsymbol{X}_{t-\Gamma}, \boldsymbol{X}_{t-\Gamma+1}, \cdots, \boldsymbol{X}_{t-1}]$划分为包含$I$个时间步长的不重叠子序列，且子序列数量为$N = \Gamma/I$，随机屏蔽75%的子序列，所屏蔽子序列集表示为$\boldsymbol{X}_{\mathrm{masked}}$，未屏蔽子序列$\boldsymbol{X}_{\mathrm{unmasked}}$为STRL的输入数据

$$\boldsymbol{S}_{\mathrm{unmasked}} = \mathrm{STRL}(\boldsymbol{X}_{\mathrm{unmasked}}) \tag{1}$$

其中，$\boldsymbol{S}_{\mathrm{unmasked}}$为STRL的输出。自监督任务头由Transformer解码层和线性输出层组成，并且在给定$\boldsymbol{S}_{\mathrm{unmasked}}$和可学习掩码标记$\boldsymbol{S}_{[\mathrm{MASK}]}$的情况下重建完整的长序列，如式(2)所示

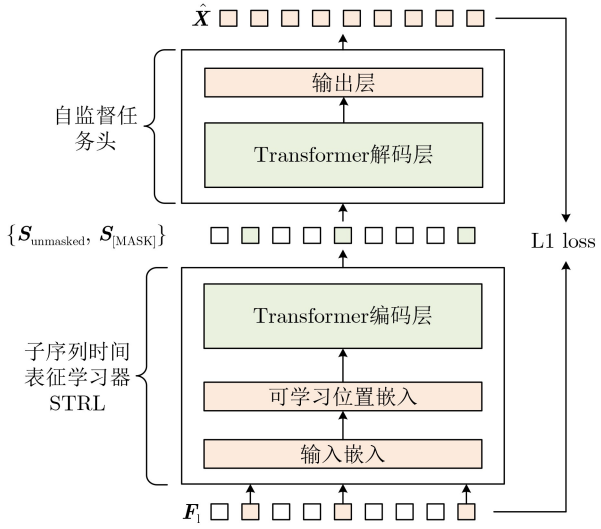


图 2 Masked Sub-series Transformer的结构组成

$$[\boldsymbol{X}_{\text{masked}}, \hat{\boldsymbol{X}}_{\text{unmasked}}] = \text{TaskHead}([\boldsymbol{S}_{\text{unmasked}}, \boldsymbol{S}_{[\text{MASK}]}]) \tag{2}$$

为使重构的$\hat{\boldsymbol{X}}_{\text{masked}}$和掩码真值之间的误差尽可能小，故预训练阶段在计算损失时只考虑掩码子序列

$$L_{\text{pretrain}}(\boldsymbol{X}_{\text{masked}}, \boldsymbol{X}_{\text{masked}}; \boldsymbol{\Theta}_{\text{T}}) = \left\| \hat{\boldsymbol{X}}_{\text{masked}} - \boldsymbol{X}_{\text{masked}} \right\| R \tag{3}$$

其中$\boldsymbol{\Theta}_{\text{T}}$为整个Transformer的可学习参数。

3.2.2 膨胀门控卷积

短期历史序列信息量相对较少，无法有效推断未来交通流的复杂变化，而长期历史序列可以帮助模型确定未来时刻交通流的波动。为此，提出DG-Conv在时间维度上处理数据以有效捕获交通流的长期依赖性。如图1(c)所示，采用1D膨胀卷积层从子序列时态表示中提取交通流的长期趋势特征。该模块的感受野随着1D扩张卷积层数量的增加而呈指数增长，以有效地捕获趋势特征，同时避免梯度消失等问题。扩张卷积运算表示为

$$\boldsymbol{x}_d^* \boldsymbol{C}(m) = \sum_{j=0}^{k-1} \boldsymbol{C}(m) \boldsymbol{x}(m - d \times j) \tag{4}$$

其中，m表示序列$\boldsymbol{x}$中的第m个元素，$\boldsymbol{C} \in R^k$为卷积核，d表示膨胀率。因此，膨胀卷积层表示为

$$\boldsymbol{D}_i = \text{Conv}(\boldsymbol{D}_{i-1}) = \text{MaxPool}(\text{gelu}(\boldsymbol{D}_{(i-1)_d}^* \boldsymbol{C}_i)) \tag{5}$$

其中采用最大池操作降低维数，第i层的膨胀率d设置为2^i。当$i=1$时，模块的输入是子序列时间表示S的集合，即$\boldsymbol{D}_1 = \boldsymbol{S} = [\boldsymbol{S}_1, \boldsymbol{S}_2, \cdots, \boldsymbol{S}_N]$。长期趋势特征$\boldsymbol{F}_l^{\text{long}}$为最后一个卷积层的输出。

根据门控机制在处理序列数据方面表现出优势，因此将其用于膨胀门控卷积层以提高模型容量，表示为

$$\boldsymbol{H}_l^t = \tanh(\boldsymbol{W}_{f,l} \blacklozenge \boldsymbol{F}_l^{\text{long}}) \odot \sigma(\boldsymbol{W}_{g,l} \blacklozenge \boldsymbol{F}_l^{\text{long}}) \tag{6}$$

其中，$\blacklozenge$菱形符号为扩张卷积运算，$\odot$为逐元素乘法运算符，$\sigma(\cdot)$为sigmoid函数，$\boldsymbol{W}(\cdot)$为卷积滤波器的可学习参数。

3.3 空间卷积层

由于交通流的动态特性是不断变化的且路段间具有隐藏的空间依赖关系，因此设计了空间卷积层，通过动态图构造器生成动态可学习图并构建可学习动态图卷积，以有效捕获交通路网间的空间隐藏关系及交通流的动态空间特征。

3.3.1 动态图构造

为有效地对不同的节点之间的关系进行自适应建模，首先考虑交通状态的周期性，假设一天中同一时刻的交通路网共享一个图。因此构建邻接矩阵$\boldsymbol{A} \in R^{N_{\text{t}} \times N \times N}$，$N_{\text{t}}$为一天的时段数。在时刻$t$的交通图为$\boldsymbol{A}_{\phi(t)}$，然而此方法的复杂度是$O(N_{\text{t}}NN)$，且随着图的增大，复杂度呈2次方增长。在现实生活中，动态图的某些结构可跨时间和空间共享。例如，两个相邻的时间段可以在一天中相关联，并且人们会在早上从不同的居住区离开到不同的工作区域。因此，如图3所示，受Tucker分解方法启发形成邻接张量，即构建一个具有空间依赖性的可学习参数的未知张量，并在训练期间通过反向传播更新动态图。

具体地，利用3个可学习矩阵和1个可学习核心张量，即时隙嵌入$\boldsymbol{E}^{\text{t}} \in R^{N_{\text{t}} \times d}$，源节点嵌入$\boldsymbol{E}^{\text{c}} \in R^{N_{\text{c}} \times d}$，目标节点嵌入$\boldsymbol{E}^{\text{e}} \in R^{N_{\text{e}} \times d}$，核心张量$\boldsymbol{E}^{\text{b}} \in R^{d \times d \times d}$。邻接张量计算为

$$\boldsymbol{A}'_{t,i,j} = \sum_{o=1}^{d} \sum_{q=1}^{d} \sum_{r=1}^{d} \boldsymbol{E}^{\text{b}}_{o,q,r} \boldsymbol{E}^{\text{t}}_{t,o} \boldsymbol{E}^{\text{e}}_{i,q} \boldsymbol{E}^{\text{c}}_{j,r} \tag{7}$$

$$\boldsymbol{A}''_{t,i,j} = \max(0, \boldsymbol{A}'_{t,i,j}) \tag{8}$$

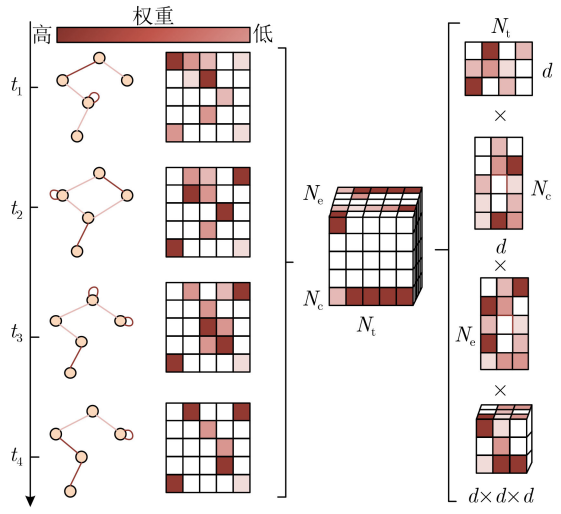


图 3 动态图的构造

$$\boldsymbol{A}_{t,i,j}=\frac{\mathrm{e}^{\boldsymbol{A}''_{t,i,j}}}{\sum_{n=1}^{N_{\mathrm{a}}}\mathrm{e}^{\boldsymbol{A}''_{t,i,n}}} \tag{9}$$

其中，$\boldsymbol{E}^{\mathrm{t}}$, $\boldsymbol{E}^{\mathrm{c}}$, $\boldsymbol{E}^{\mathrm{e}}$分别表示为时隙数、原始节点数、目标节点数和嵌入维数。且共生成3个邻接张量$\boldsymbol{A}^{\mathrm{p}}\in R^{N_t\times N_{\mathrm{p}}\times N_{\mathrm{p}}}$，$\boldsymbol{A}^{\mathrm{a}}\in R^{N_{\mathrm{t}}\times N_{\mathrm{a}}\times N_{\mathrm{a}}}$和$\boldsymbol{A}^{\mathrm{ap}}\in R^{N_{\mathrm{t}}\times N_{\mathrm{p}}\times N_{\mathrm{a}}}$表示动态图$\boldsymbol{G}^{\mathrm{p}}$, $\boldsymbol{G}^{\mathrm{a}}$和$\boldsymbol{G}^{\mathrm{ap}}$。

3.3.2 可学习动态图卷积

有效捕获不同路段交通流的空间依赖性可用于提高模型预测性能。为对节点之间潜在关系和时变的空间依赖关系进行有效建模，提出一种可学习动态图卷积，在不同的时间对不同的动态图进行卷积。

给定交通流输入$\boldsymbol{X}_{t-p+1}$，$\boldsymbol{A}_{\phi(t)}$反映了t时刻节点间的空间关系。如图4所示，首先将每个界定的输入隐藏状态进行隔离，并从邻接张量中有效提取$\boldsymbol{A}_{\phi(t)}$后根据其动态空间关系将隐藏状态组织为图信号。依据DCRNN方法，将交通流视为通过邻接矩阵、节点隐藏状态和可学习参数的矩阵乘法来实现的扩散过程。

通过动态图构造器生成可学习动态图运用于图卷积运算。每一步从动态图张量中获取带有时间的可学习邻接矩阵，通过加权链接聚合邻居的隐藏状态来更新焦点节点的隐藏状态，可学习动态图卷积定义为

$$\boldsymbol{H}_l=\sum_{k=0}^{K}(\boldsymbol{A}_{\phi(t)})^k\boldsymbol{H}_l^t\boldsymbol{W}^k \tag{10}$$

其中，$\boldsymbol{H}_l^t$表示第l块中长期门控卷积层的输出隐藏状态和第l块中可学习动态图卷积的输入，$\boldsymbol{W}^k$为可学习参数。

3.4 多方面融合模块

由于主特征与辅助特征之间存在潜在关系并相互影响，故而提出多方面融合模块，基于交通动态图视角对交通流主特征和辅助特征进行有效融合，以实现更加精准的交通流预测。并且构建一个自适应频谱块对交通流的全局特征信息进行有效捕获，并利用局部滤波器降低交通流的非平稳性以提高模型性能。

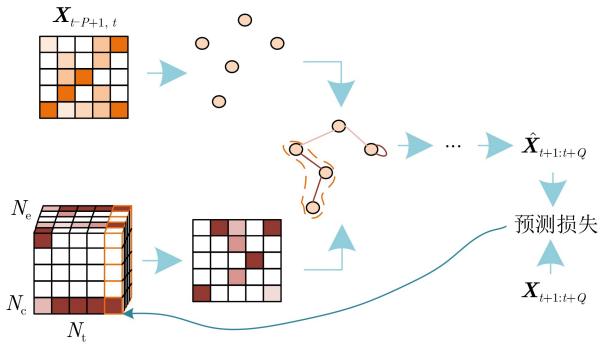


图 4 可学习动态图卷积

3.4.1 可主特征和辅助特征框架

交通路网中，上游流量的增加会影响未来的下游流量，从而对下游交通速度的下降产生影响，因此，根据具有辅助特征的节点之间所存在的时空依赖关系，为主特征和辅助特征建立独立的时空块，且对l层结构进行分离，获得主特征$\boldsymbol{H}_l^{\mathrm{p}}$和辅助特征$\boldsymbol{H}_l^{\mathrm{a}}$的隐藏状态，并遵循“传播和聚合”范式对其进行有效融合

$$\boldsymbol{H}_l^{\mathrm{p}'}=\mathrm{Aggregate}(\boldsymbol{H}_l^{p},\mathrm{Propagate}(\boldsymbol{H}_l^{\mathrm{a}})) \tag{11}$$

具体地，通过可学习矩阵$\boldsymbol{A}^{\mathrm{ap}}\in R^{N_{\mathrm{t}}\times N_{\mathrm{p}}\times N_{\mathrm{a}}}$表示动态图$\boldsymbol{G}^{\mathrm{ap}}$，以捕获主特征及辅助特征间的空间依赖性，传播模块将最相关的辅助隐藏状态传播到所需的主节点，计算为

$$\boldsymbol{H}_l^{\mathrm{ap}}=\mathrm{Propagate}(\boldsymbol{H}_l^{\mathrm{a}})=\boldsymbol{A}_{\phi(t)}^{\mathrm{ap}}\boldsymbol{H}_l^{\mathrm{a}}\boldsymbol{W} \tag{12}$$

其中，$\boldsymbol{H}_l^{\mathrm{ap}}\in R^{N_{\mathrm{p}}\times P\times d_{\mathrm{p}}}$表示辅助特征对主节点的影响，$d_{\mathrm{p}}$为主隐藏状态维数。为有效聚合两者之间的信息，采用聚合函数SUM(·)以获得较高的表征能力，表示为

$$\boldsymbol{H}_l^{\mathrm{p}'}=\mathrm{Aggregate}(\boldsymbol{H}_l^{p},\boldsymbol{H}_l^{\mathrm{ap}})=\boldsymbol{H}_l^{p}+\boldsymbol{H}_l^{\mathrm{ap}} \tag{13}$$

若存在多个辅助特征$\{\boldsymbol{H}_l^{\mathrm{a}_1},\boldsymbol{H}_l^{\mathrm{a}_2},\cdots,\boldsymbol{H}_l^{\mathrm{a}_M}\}$且辅助节点数量$N_{\mathrm{a}_i}(i\in[1,M])$，将以上过程扩展为

$$\boldsymbol{H}_l^{\mathrm{p}'}=\boldsymbol{H}_l^{p}+\sum_{\varepsilon=1}^{M}\boldsymbol{A}_{\phi(t)}^{\mathrm{ap}_\varepsilon}\boldsymbol{H}_l^{\mathrm{a}_\varepsilon}\boldsymbol{W} \tag{14}$$

其中$\boldsymbol{A}^{\mathrm{ap}_\varepsilon}\in R^{N_{\mathrm{t}}\times N_{\mathrm{p}}\times N_{\mathrm{a}_\varepsilon}}$为第$\varepsilon$个辅助特征与主特征间的动态图。

3.4.2 自适应频谱块

如图1(b)所示，为在有效学习交通流全局-局部空间信息的同时降低交通流的非平稳性，提出了基于离散傅立叶变换(Discrete Fourier Transform, DFT)的自适应频谱块，主要由全局循环卷积和自适应局部滤波器两部分组成，以有效隔离了交通流数据序列的噪声高频分量。

(1)快速傅里叶变换：DFT通过考虑交通流序列$\boldsymbol{H}_l^{\mathrm{p}'}(n)$，其中$0\le n\le N-1$。利用1维DFT将交通流序列变换为频域表示

$$\boldsymbol{X}[\eta]=\sum_{n=0}^{N-1}\boldsymbol{H}_l^{\mathrm{p}'}(n)\mathrm{e}^{-\mathrm{j}(2\pi/N)\eta n}:=\sum_{n=0}^{N-1}\boldsymbol{H}_l^{\mathrm{p}'}(n)\Sigma_N^{\eta n} \tag{15}$$

其中，$\Sigma_N=\mathrm{e}^{-\mathrm{j}(2\pi/N)}$由连续傅里叶变换在时域和频域离散所得。由于DFT的双射性质，原始序列

$\boldsymbol{H}_l^{\mathrm{p}'}(n)$通过逆DFT(Inverse Discrete Fourier Transform, IDFT)得到

$$\boldsymbol{H}_l^{\mathrm{p}'}(n)=\frac{1}{N}\sum_{\eta=0}^{N-1}\boldsymbol{X}[\eta]\mathrm{e}^{\mathrm{j}(2\pi/N)\eta n} \tag{16}$$

给定$\boldsymbol{\Omega}_{\mathrm{PE}}$实现快速傅里叶变换(Fast Fourier Transform, FFT)，FFT利用$\Sigma_N^{\eta n}$的对称性和周期性，将DFT计算复杂度从$O(N^2)$优化$O(N\log_2 N)$，上述过程表示为

$$\boldsymbol{H}=\boldsymbol{\Psi}[\boldsymbol{\Omega}_{\mathrm{PE}}]\in\mathbb{C}^{C\times\Upsilon'} \tag{17}$$

其中，$\boldsymbol{\Psi}[\cdot]$表示1维FFT运算，Υ'为频域中变换序列长度。交通流序列的每个通道被独立地变换，从而产生综合频域表示$\boldsymbol{H}$，其封装了跨所有通道的原始时间序列的频谱特性。

(2)高频噪声的自适应滤除：高频成分通常代表偏离潜在趋势或感兴趣信号的快速波动，使其显得更加随机且难以解释。给定从FFT运算获得的频域表示$\boldsymbol{H}$，通过计算$\boldsymbol{H}$的功率谱$P=\|\boldsymbol{H}\|^2$，有助于有效识别主频分量。

自适应局部滤波器通过可训练阈值θ为自适应地从功率谱P_S中滤除高频分量，且该阈值根据交通数据的频谱特性进行调整，为在训练期间通过反向传播进行优化的可学习参数以区分基本信号频率和噪声。自适应阈值表示为

$$\boldsymbol{H}_{\mathrm{filtered}}=\boldsymbol{H}\odot(P_S>\theta) \tag{18}$$

其中，$\odot$为逐元素乘法，且$(P_S>\theta)$为二进制掩码，其中保留功率高于阈值θ的频率，并且滤除其他频率。

(3)可学习滤波器：采用两组可学习滤波器：从原始交通频率数据$\boldsymbol{H}$学习的全局滤波器和从自适应滤波数据$\boldsymbol{H}_{\mathrm{filtered}}$学习的局部滤波器，表示为

$$\left.\begin{aligned}\boldsymbol{H}_G&=\boldsymbol{W}_G\odot\boldsymbol{H}\\ \boldsymbol{H}_{\Pi}&=\boldsymbol{W}_{\Pi}\odot\boldsymbol{H}_{\mathrm{filtered}}\end{aligned}\right\} \tag{19}$$

其中$\boldsymbol{W}_G$, $\boldsymbol{W}_{\Pi}$为滤波器的可学习参数。通过整合滤波后的交通特征以捕获全面的光谱细节，即，$\boldsymbol{H}_{\mathrm{integrated}}=\boldsymbol{H}_G+\boldsymbol{H}_{\Pi}$。且滤波器运算等效于循环卷积更加有效地捕捉交通流序列中的周期性。

(4)逆傅里叶变换：应用快速傅里叶逆变换(Inverse Fast Fourier Transform, IFFT)将频域交通数据转换回时域交通数据。故而，所得时域交通数据$\boldsymbol{\Omega}'$为

$$\boldsymbol{\Omega}'=\boldsymbol{\Psi}^{-1}[\boldsymbol{H}_{\mathrm{integrated}}] \tag{20}$$

IFFT确保增强的交通流特征与输入交通流序列的原始数据结构保持一致。

3.5 其他组件

ADGformer模型中，$\boldsymbol{F}_l^{(\cdot)}$是第$l$块的输入和第$l-1$块的输出，且在第1块之前，输入交通数据由全连接层进行转换

$$\left.\begin{aligned}\boldsymbol{F}_1^{\mathrm{p}}&=\boldsymbol{W}_{\mathrm{in}}^{\mathrm{p}}\boldsymbol{X}^{\mathrm{p}}+\boldsymbol{g}_{\mathrm{in}}^{\mathrm{p}}\\ \boldsymbol{F}_1^{\mathrm{a}}&=\boldsymbol{W}_{\mathrm{in}}^{\mathrm{a}}\boldsymbol{X}^{\mathrm{a}}+\boldsymbol{g}_{\mathrm{in}}^{\mathrm{a}}\end{aligned}\right\} \tag{21}$$

并为每块进行剩余连接操作

$$\left.\begin{aligned}\boldsymbol{F}_{l+1}^{\mathrm{p}}&=\boldsymbol{H}_l^{\mathrm{p}'}+\boldsymbol{F}_l^{\mathrm{p}}\\ \boldsymbol{F}_{l+1}^{\mathrm{a}}&=\boldsymbol{H}_l^{\mathrm{a}}+\boldsymbol{F}_l^{\mathrm{a}}\end{aligned}\right\} \tag{22}$$

此外，在每个长期门控卷积层之后添加跳跃连接到输出层。通过连接不同深度层的隐藏状态并传递到两个完全连接层中，以预测未来Q步骤的最终主要交通流特征

$$\boldsymbol{H}=\|_{l=1}^{L}\mathrm{reshape}(\boldsymbol{H}_l^{\mathrm{p},t}) \tag{23}$$

$$\begin{aligned}\hat{\boldsymbol{X}}_{t+1:t+Q}^{\mathrm{p}}=\boldsymbol{W}_{\mathrm{fc2}}\cdot\mathrm{ReLU}(\boldsymbol{W}_{\mathrm{fc1}}\cdot\mathrm{ReLU}(\boldsymbol{H})\\ +\varepsilon_{\mathrm{fc1}})+\varepsilon_{\mathrm{fc2}}\end{aligned} \tag{24}$$

其中，$\|$为串联操作，$\mathrm{reshape}(\cdot)$为重塑隐藏状态$\boldsymbol{H}_l^{\mathrm{p},t}$的函数，$\boldsymbol{W}_{(\cdot)}$和$\varepsilon_{(\cdot)}$是可学习参数。将MAE作为损失函数以端到端训练模型

$$\mathrm{Loss}=\frac{1}{Q\times N_{\mathrm{p}}}\sum_{i=t+1}^{t+Q}\sum_{j=1}^{N_{\mathrm{p}}}\left|\boldsymbol{X}_{i,j}^{\mathrm{p}}-\hat{\boldsymbol{X}}_{i,j}^{\mathrm{p}}\right| \tag{25}$$

4 实验

4.1 数据集

所有实验均在旧金山弗朗西斯科湾区数据集PEMSD4、圣贝纳迪诺数据集PEMSD8和英国高速公路交通数据集England上进行。其中，PEMSD4和PEMSD8是由PeMS(Caltrans Performance Measurement System)收集的美国加利福尼亚州2个地区的真实高速公路2个月流量数据，道路传感器每30 s收集1次数据，以5 min为间隔进行聚合。England是一个包含了英国高速公路交通信息的综合性数据集，以15 min为间隔进行聚合，共收集了6个月的交通数据。数据集的详细信息如表1所示。

4.2 实验设置

所有实验在配备14 vCPU Intel(R) Xeon(R) Platinum 8362 CPU @ 2.80GHz, GPU:RTX

表 1　数据集详细信息

数据集	节点	边	时间步长(min)	时间跨度(月)
PEMSD4	307	16 969	5	2
PEMSD8	170	17 833	5	2
England	314	17 353	15	6

3090(24 GB) * 1的计算机环境下执行。使用Pytorch1.11.0, Python3.8和 ubuntu20.04完成ADGformer模型的搭建和训练，并将数据集按照6:2:2的比例分为训练集、验证集和测试集，并采用Adam优化器对模型进行优化，对于PEMSD4和PEMSD8，通过线性插值法填充缺失值。经多次实验，模型中每个模块的块数为8，每层膨胀率为[1,2,1,2,1,2,1,2]，空间卷积层的最大深度为2，通道大小为16，动态图构造器中的隐藏维数设置为16，bach-size为16，学习率设为0.001。为验证提出 ADGformer的有效性，采用平均绝对误差(Mean Absolute Error, MAE)、平均绝对百分误差(Average Absolute Percentage Error, MAPE)和误差均方根(Root Mean Square Error, RMSE)3个指标对模型的性能进行了评估

$$\mathrm{MAE}=\frac{1}{N}\sum_{i=1}^{N}|\boldsymbol{Y}_i-\hat{\boldsymbol{Y}}_i| \tag{26}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{Y}_i-\hat{\boldsymbol{Y}}_i\right)^2} \tag{27}$$

$$\mathrm{MAPE}=\frac{1}{N}\sum_{i=1}^{N}\left|\frac{\boldsymbol{Y}_i-\hat{\boldsymbol{Y}}_i}{\boldsymbol{Y}_i}\right| \tag{28}$$

其中，N为路网节点数，$\boldsymbol{Y}_i$代表真实值，$\hat{\boldsymbol{Y}}_i$为预测值。

4.3 基线模型

为验证所提模型的有效性，将ADGformer与以下基线模型进行比较：

(1)HA(Historical Average)[29]：历史平均方法，历史平均交通流量信息作为预测依据。

(2)VAR(Vector AutoregRessive)[3]：用于捕捉多个量之间的关系。

(3)XGBoost(eXtreme Gradient Boosting)[30]：一种基于梯度增强树的方法。

(4)LR(Logistic Regression)[31]：一种利用投入产出线性相关性的回归模型。

(5)ASTGCN(Attention based Spatial-Temporal Graph Convolutional Networks)[32]：将时空注意力机制集成到时空图卷积网络中以捕获交通流的动态时空特征。

(6)GMAN(Graph Multi-Attention Network)[33]：基于输入特征、空间嵌入和时间嵌入的考虑时空相关性的图多注意模型。

(7)DCRNN(Diffusion Convolutional Recurrent Neural Network)[34]：一种结合扩散图卷积和递归神经网络的时空网络。

(8)MTGNN(Multivariate Time series prediction model based on Graph Neural Network)[35]：一种利用外部特征生成单向自适应图的时空网络。

(9)GWNet(Graph WaveNet)[36]：一种将图卷积网络和扩张因果卷积网络进行结合的方法。

4.4 实验结果

4.4.1 对比实验

在PEMSD4,PEMSD8和England 3个数据集上，所提ADGformer模型与基线模型分别在Horizon 3, Horizon 6和Horizon 12 3个时间段上进行了预测性能比较，ADGformer模型在3个数据集上均实现了最佳精度。

从表2可以看出，所提出的ADGformer模型无论在短期还是长期均已达到最佳预测效果，且基于深度学习的模型(ASTGCN, GMAN, DCRNN, MTGNN和GWNet)的预测精度高于传统统计方法。在PEMSD4数据集上，与VAR和LR等传统预测模型相比，ADGformer模型在Horizon 3下的MAE分别改进了约21.57%, 16.72%，其原因是所提模型能够拟合序列数据集的非线性和复杂性。在Horizon 6下，与依赖于预定义图结构的ASTGCN和DCRNN等模型相比，ADGformer模型在PEMSD8数据集下的MAPE分别提高了20.39%, 12.31%，表明所提ADGformer模型通过动态图构造器生成动态可学习图并将其运用至图卷积运算中，有效地对不同的节点之间的关系进行自适应建模，进一步捕获交通路网间的空间隐藏关系及交通流的动态空间特征。

交通流的长期预测比短期预测具有更多的不确定性，导致预测长期的交通流更加困难，使得大多数现有模型的预测性能不佳。GWNet和MTGNN模型利用自适应图结构有效地利用了历史交通数据中潜在的空间依赖性从而取得较好的预测性能。然而，在Horizon 12下，与GWNet和MTGNN相比，所提出的模型ADGformer在England数据集下的RMSE分别改进了3.35%, 2.44%，表明ADGformer模型通过Masked Sub-series Transformer有效地从长时间序列中学习压缩的、上下文信息丰富的子序列表示，并利用膨胀门控卷积从子序列的时间表示中提取交通流的长期趋势特征，且通过自适应频谱块降低了交通流数据的非平稳性，使其随着预测时间步长的增加，ADGformer模型的预测性能仍然表现良好。通过对比试验证明，所提模型ADGformer的短期预测性能优于基线模型，长期预测性能同样优于基线模型，表明了ADGformer在有效提取路网间动态复杂时空潜在关系的同时也对长期趋势特

表 2　各种模型在3个数据集上的预测结果

数据集	模型	Horizon 3			Horizon 6			Horizon 12		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
PEMSD4	HA	3.567 9	6.778 7	0.080 3	3.568 9	6.779 7	0.080 4	3.570 7	6.781 5	0.080 4
	VAR	1.662 4	3.088 2	0.032 7	2.125 0	4.015 6	0.043 3	2.568 7	4.834 7	0.053 6
	XGBoost	1.447 1	2.973 2	0.028 8	1.915 9	4.095 5	0.041 3	2.532 7	5.323 9	0.058 5
	LR	1.565 5	3.138 4	0.030 9	2.075 8	4.291 8	0.043 9	2.795 1	5.597 8	0.062 7
	ASTGCN	1.542 8	3.142 8	0.032 3	1.937 4	4.106 2	0.042 5	2.442 4	5.139 2	0.055 2
	GMAN	1.376 0	2.980 3	0.027 9	1.621 3	3.790 6	0.034 8	1.865 0	4.479 1	0.041 6
	DCRNN	1.429 5	2.880 5	0.028 1	1.853 1	3.873 1	0.039 1	2.366 6	4.899 1	0.052 7
	MTGNN	1.348 5	2.851 7	0.026 8	1.645 4	3.752 3	0.034 8	1.930 6	4.497 2	0.042 6
	GWNet	1.326 6	2.815 9	0.026 6	1.645 6	3.764 0	0.035 6	1.955 0	4.556 0	0.044 6
	ADGformer	**1.303 8**	**2.728 6**	**0.025 6**	**1.606 0**	**3.684 3**	**0.032 7**	**1.862 9**	**4.367 6**	**0.041 1**
PEMSD8	HA	2.811 9	5.676 3	0.063 1	2.809 9	5.673 0	0.063 0	2.804 7	5.664 7	0.062 7
	VAR	1.132 7	2.071 2	0.022 4	1.716 6	3.305 3	0.035 6	2.166 1	4.250 0	0.046 7
	XGBoost	1.199 7	2.569 8	0.024 1	1.560 4	3.503 0	0.033 8	2.000 8	4.484 0	0.045 8
	LR	1.271 1	2.633 5	0.024 4	1.661 4	3.560 6	0.033 5	2.168 1	4.579 4	0.045 9
	ASTGCN	1.379 4	2.993 4	0.029 7	1.644 6	3.619 4	0.035 8	1.994 5	4.288 0	0.043 6
	GMAN	1.137 4	2.675 2	0.023 4	1.323 7	3.395 0	0.029 2	1.512 0	4.052 4	0.035 5
	DCRNN	1.195 7	2.467 9	0.023 6	1.534 7	3.303 4	0.032 5	1.905 1	4.145 1	0.042 7
	MTGNN	1.140 2	2.573 8	0.023 4	1.397 7	3.491 3	0.031 4	1.632 3	4.249 0	0.039 4
	GWNet	1.118 4	0.023 3	2.553 3	1.384 9	3.532 7	0.031 9	1.604 3	4.242 4	0.039 1
	ADGformer	**1.046 9**	**2.451 6**	**0.019 1**	**1.318 5**	**3.282 0**	**0.028 5**	**1.505 8**	**4.051 4**	**0.035 3**
England	HA	7.047 3	12.213 3	0.099 3	7.044 1	12.210 6	0.099 3	7.034 1	12.200 5	0.099 2
	VAR	3.213 5	5.725 4	0.041 9	4.101 9	7.576 4	0.056 3	4.837 2	8.948 7	0.069 3
	XGBoost	3.192 4	6.717 7	0.044 0	4.298 8	8.667 3	0.063 1	5.523 7	10.475 6	0.082 9
	LR	3.773 2	7.433 1	0.050 5	5.281 6	9.704 5	0.073 7	6.566 4	11.599 1	0.094 3
	ASTGCN	3.214 1	6.540 4	0.043 9	3.926 2	7.917 7	0.055 6	4.601 0	9.079 3	0.066 4
	GMAN	2.611 7	6.287 4	0.036 2	2.955 8	7.333 9	0.043 1	3.306 6	8.140 4	0.049 3
	DCRNN	2.836 3	6.106 2	0.038 1	3.491 3	7.447 3	0.048 7	4.281 3	8.721 7	0.061 0
	MTGNN	2.535 2	5.978 9	0.035 4	2.921 6	7.088 9	0.043 0	3.335 9	7.993 6	0.050 7
	GWNet	2.521 6	5.961 0	**0.034 8**	2.961 0	7.128 4	0.043 6	3.446 1	8.068 7	0.051 3
	ADGformer	**2.519 0**	**5.943 5**	**0.034 8**	**2.907 9**	**7.045 0**	**0.042 8**	**3.243 4**	**7.798 8**	**0.048 8**

征进行了有效捕获并降低了交通流的非平稳性，增强了模型的长期预测能力。

为了更好地解释ADGformer模型，将ADGformer模型与XGBoost, ASTGCN和DCRNN在PEMSD4和England数据集上的实验结果进行可视化，如图5所示，ADGformer模型一直优于XGBoost, ASTGCN和DCRNN，随着预测时长的增加，预测误差增长幅度较小，在Horizon 12时，ADGformer的预测误差均明显低于其他对比模型，表明该模型在长期预测中的预测性能更加优越。

4.4.2　消融实验

为了验证ADGformer模型不同模块的性能，在PEMSD4, PEMSD8和England数据集上进行了消融实验并分析实验结果，如表3所示，消融实验分别在Horizon 3, Horizon 6和Horizon 12上进行预测，变体模型如下所示：

NMS-Trans：不含Masked Sub-series Transformer模块；

NLGConv：不含长期门控卷积层(Masked Sub-series Transformer和膨胀门控卷积)；

NASB：不含自适应频谱块；

NDGConv：不含动态图结构。

由结果可知，随着预测步长的增加，ADGformer模型的性能优于其他变体模型, 证实了所提模型中每个模块的有效性。在Horizon 12上，ADGformer相较于变体模型NMS-Trans, NLGConv, NASB和

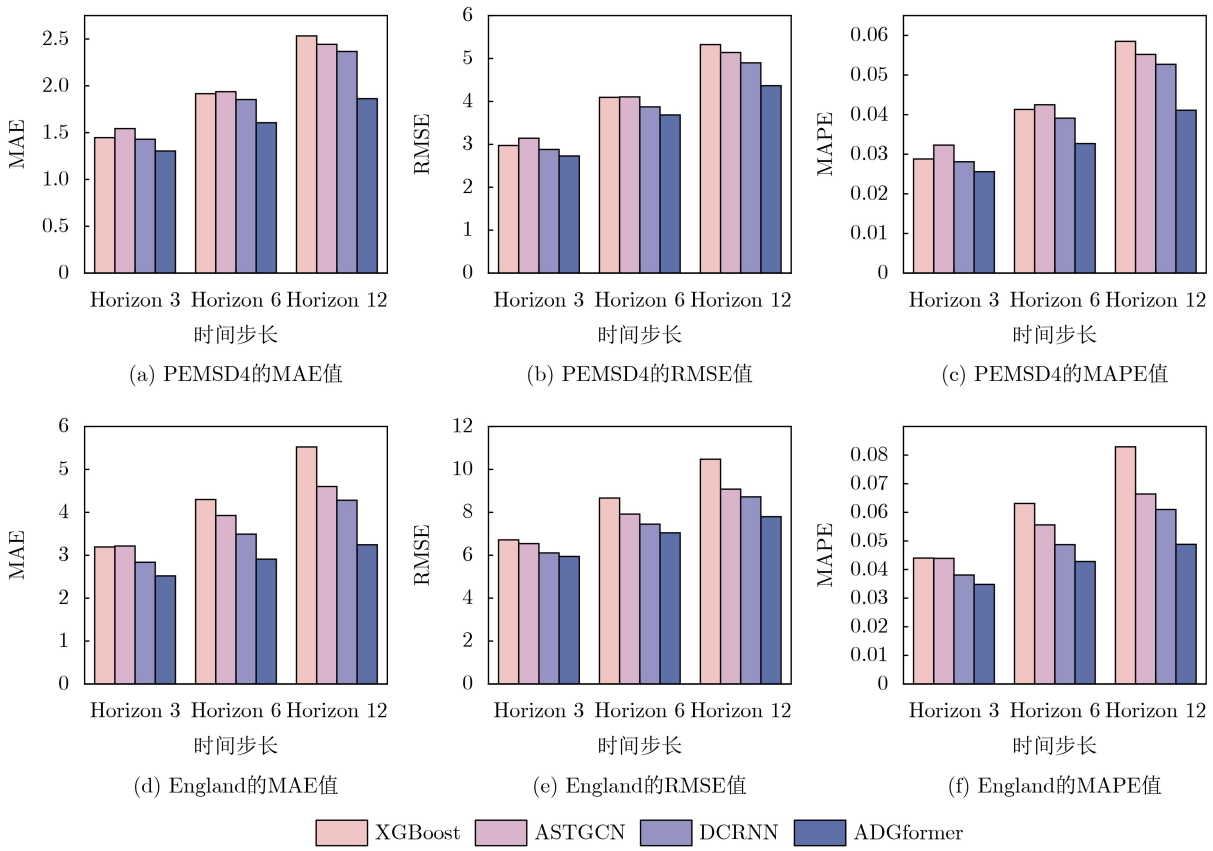


图 5 PEMSD4和England数据集上不同模型性能比较

表 3 ADGformer模型与变体模型在不同时间步长的预测性能

数据集	模型	Horizon 3			Horizon 6			Horizon 12		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
PEMSD4	NMS-Trans	1.330 3	2.810 3	0.026 2	1.643 3	3.727 1	0.035 1	1.934 3	4.493 5	0.043 2
	NLGConv	1.353 8	2.828 6	0.027 6	1.656 0	3.784 3	0.036 7	1.962 9	4.567 6	0.044 1
	NASB	**1.303 2**	2.772 3	0.025 8	1.611 1	3.694 7	0.033 9	1.880 2	4.399 0	0.041 9
	NDGConv	1.345 2	2.828 4	0.026 4	1.684 3	3.782 7	0.035 7	2.016 0	4.625 8	0.045 1
	ADGformer	1.303 8	**2.728 6**	**0.025 6**	**1.606 0**	**3.684 3**	**0.032 7**	**1.862 9**	**4.367 6**	**0.041 1**
PEMSD8	NMS-Trans	1.103 9	2.483 2	0.021 9	1.360 0	3.332 8	0.029 1	1.607 1	4.076 5	0.036 3
	NLGConv	1.111 4	2.500 5	0.022 2	1.389 6	3.447 0	0.030 4	1.666 7	4.263 7	0.038 5
	NASB	1.121 8	2.566 9	0.022 9	1.398 2	3.533 4	0.031 0	1.631 3	4.232 8	0.037 9
	NDGConv	1.149 5	2.563 2	0.023 9	1.442 5	3.497 0	0.031 9	1.724 6	4.263 1	0.039 7
	ADGformer	**1.046 9**	**2.451 6**	**0.019 1**	**1.318 5**	**3.282 0**	**0.028 5**	**1.505 8**	**4.051 4**	**0.035 3**
England	NMS-Trans	2.528 2	5.994 9	0.035 3	**2.889 0**	7.109 5	0.043 3	3.288 5	7.900 6	0.050 6
	NLGConv	2.513 2	6.271 1	0.037 3	2.988 9	7.178 7	0.044 2	3.297 7	7.993 7	0.059 9
	NASB	2.524 7	5.984 7	0.035 3	2.911 9	7.092 5	0.044 3	3.267 8	7.917 2	0.050 3
	NDGConv	2.555 7	6.088 4	0.036 2	2.957 3	7.258 3	0.044 2	3.339 8	7.997 9	0.051 3
	ADGformer	**2.519 0**	**5.943 5**	**0.034 8**	2.907 9	**7.045 0**	**0.042 8**	**3.243 4**	**7.798 8**	**0.048 8**

NDGConv在数据集PEMSD4上的MAE分别降低了约3.69%, 5.09%, 0.92%和7.59%，在数据集England上的MAPE分别降低了约3.56%, 18.53%, 2.98%和4.87%，证明了ADGformer不仅有效地提取交通流隐藏的动态时空特征，且在高效捕捉交通流长历史序列间的长趋势特征的同时降低交通数据间的非平稳性，提高了模型长期预测能力。为更好地表明ADGformer模型随预测步长的性能变化，将PEMSD8和England数据集上的实验结果进行可视化，见图6。NMS-Trans模型性能优于NLGConv，

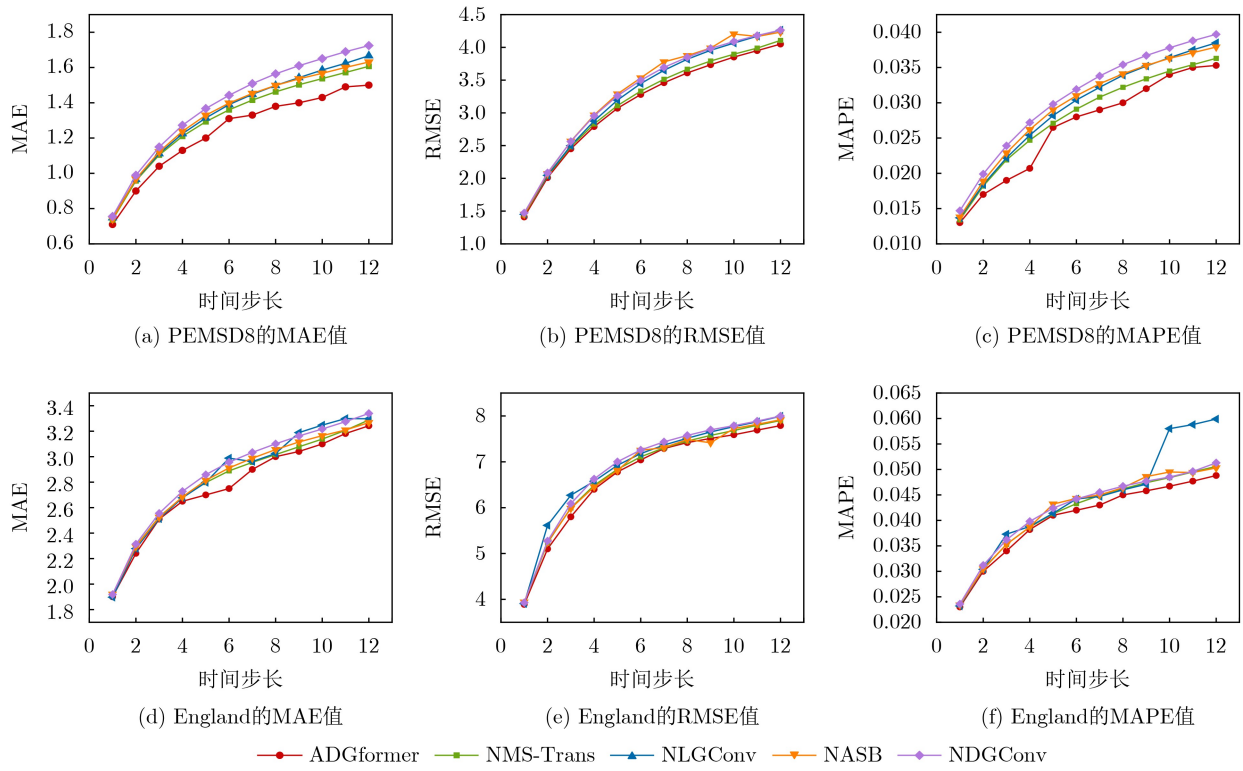


图 6 PEMSD8和England数据集上与变体模型性能对比

证明了Masked Sub-series Transformer和膨胀门控卷积的巧妙结合使得ADGformer能够有效捕获交通流的长趋势特征和长期依赖性。与NASB和NDGConv相比，表明了自适应频谱块在有效降低交通流非平稳性的同时有效捕捉了长短期相互作用，且动态图构造器所生成的动态图结构充分考虑了交通路网节点间隐藏的动态时空相关性。

5　结束语

为了解决现有交通流预测模型所存在的问题：(1)无法有效捕捉路网节点间隐藏的动态时空相关性；(2)无法有效地学习长历史序列中的关键时间信息和长期趋势特征；(3)无法有效降低交通数据的非平稳性，提出了一种交通流组合预测模型ADGformer，实现了精准的长期交通流预测。ADGformer首先通过Masked Sub-series Transformer对长历史交通流序列进行预训练以有效学习信息丰富的子序列表示，并巧妙结合门控机制和1D膨胀卷积以充分捕捉交通流的长期趋势特征，提高模型的长期预测能力。其次，ADGformer从动态交通模式出发建立空间卷积层，运用于可学习动态图卷积以有效捕获路网间隐藏的动态时空相关性。另外，ADGformer设计了基于傅里叶变换的自适应频谱块来增强特征表示，通过自适应阈值来减轻噪声，降低交通流数据的非平稳性的同时考虑了长短期的相互作用。交通流的特性复杂多变，在未来的工作中将考虑周期性等特性，提高模型预测性能。

参 考 文 献

[1] CINI N and AYDIN Z. A deep ensemble approach for long-term traffic flow prediction[J]. *Arabian Journal for Science and Engineering*, 2024, 49(9): 12377–12392. doi: 10.1007/s13369-023-08672-1.

[2] SATTARZADEH A R, KUTADINATA R J, PATHIRANA P N, *et al*. A novel hybrid deep learning model with ARIMA Conv-LSTM networks and shuffle attention layer for short-term traffic flow prediction[J]. *Transportmetrica A: Transport Science*, 2025, 21(1): 2236724. doi: 10.1080/23249935.2023.2236724.

[3] CHENG Zeyang, LU Jian, ZHOU Huajian, *et al*. Short-term traffic flow prediction: An integrated method of econometrics and hybrid deep learning[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(6): 5231–5244. doi: 10.1109/TITS.2021.3052796.

[4] HE Silu, LUO Qinyao, DU Ronghua, *et al*. STGC-GNNs: A GNN-based traffic prediction framework with a spatial–temporal Granger causality graph[J]. *Physica A: Statistical Mechanics and its Applications*, 2023, 623: 128913. doi: 10.1016/j.physa.2023.128913.

[5] ZHU Hailong, XIE Yawen, HE Wei, *et al*. A novel traffic flow forecasting method based on RNN - GCN and BRB[J]. *Journal of Advanced Transportation*, 2020, 2020: 1–11. doi:

10.1155/2020/7586154.

[6] SUN Lijun, LIU Mingzhi, LIU Guanfeng, *et al*. FD-TGCN: Fast and dynamic temporal graph convolution network for traffic flow prediction[J]. *Information Fusion*, 2024, 106: 102291. doi: 10.1016/j.inffus.2024.102291.

[7] LIU Yutian, FENG Tao, RASOULI S, *et al*. ST-DAGCN: A spatiotemporal dual adaptive graph convolutional network model for traffic prediction[J]. *Neurocomputing*, 2024, 601: 128175. doi: 10.1016/j.neucom.2024.128175.

[8] BING Qichun, ZHAO Panpan, REN Canzheng, *et al*. Short-Term traffic flow forecasting method based on secondary decomposition and conventional neural network–transformer[J]. *Sustainability*, 2024, 16(11): 4567. doi: 10.3390/su16114567.

[9] WANG Runjie, SHI Wenzhong, LIU Xianglei, *et al*. An adaptive cutoff frequency selection approach for fast fourier transform method and its application into short-term traffic flow forecasting[J]. *ISPRS International Journal of Geo-Information*, 2020, 9(12): 731. doi: 10.3390/ijgi9120731.

[10] LI He, LI Xuejiao, SU Liangcai, *et al*. Deep spatio-temporal adaptive 3D convolutional neural networks for traffic flow prediction[J]. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2022, 13(2): 1–21. doi: 10.1145/3510829.

[11] CHEN Jian, WANG Wei, YU Keping, *et al*. Node connection strength matrix-based graph convolution network for traffic flow prediction[J]. *IEEE Transactions on Vehicular Technology*, 2023, 72(9): 12063–12074. doi: 10.1109/TVT.2023.3265300.

[12] ZHANG Hong, CHEN Linlong, CAO Jie, *et al*. Traffic flow forecasting of graph convolutional network based on spatio-temporal attention mechanism[J]. *International Journal of Automotive Technology*, 2023, 24(4): 1013–1023. doi: 10.1007/s12239-023-0083-9.

[13] LING Jiahao, LAN Yuanchun, HUANG Xiaohui, *et al*. A multi-scale residual graph convolution network with hierarchical attention for predicting traffic flow in urban mobility[J]. *Complex & Intelligent Systems*, 2024, 10(3): 3305–3317. doi: 10.1007/s40747-023-01324-9.

[14] WANG Xinqiang, SHANG Yihui, and LI Guoyan. DTM-GCN: A traffic flow prediction model based on dynamic graph convolutional network[J]. *Multimedia Tools and Applications*, 2024, 83(41): 89545–89561. doi: 10.1007/s11042-024-18348-z.

[15] LI Zuhua, WEI Siwei, WANG Haibo, *et al*. ADDGCN: A novel approach with down-sampling dynamic graph convolution and multi-head attention for traffic flow forecasting[J]. *Applied Sciences*, 2024, 14(10): 4130. doi: 10.3390/app14104130.

[16] 邓涵优, 陈红梅, 肖清, 等. 基于多头注意力动态图卷积网络的交通流预测[J]. 太原理工大学学报, 2024, 55(1): 172–183. doi: 10.16355/j.tyut.1007-9432.2023BD004.

DENG Hanyou, CHEN Hongmei, XIAO Qing, *et al*. Dynamic graph convolution network with multi-head attention for traffic flow prediction[J]. *Journal of Taiyuan University of Technology*, 2024, 55(1): 172–183. doi: 10.16355/j.tyut.1007-9432.2023BD004.

[17] PENG Hao, DU Bowen, and LIU Mingsheng, *et al*. Dynamic graph convolutional network for long-term traffic flow prediction with reinforcement learning[J]. *Information Sciences*, 2021, 578: 401–416. doi: 1016/j.ins.2021.07.007. doi: 10.1016/j.ins.2021.07.007.

[18] 唐晨嘉, 曾伟, 赵振兴. 基于自适应动态图卷积循环网络的交通流预测[J]. 武汉理工大学学报(交通科学与工程版), 2024, 48(3): 415–420. doi: 10.3963/j.issn.2095-3844.2024.03.002.

TANG Chenjia, ZENG Wei, and ZHAO Zhenxing. Traffic flow prediction based on adaptive dynamic graph convolutional recurrent network[J]. *Journal of Wuhan University of Technology (Transportation Science & Engineering)*, 2024, 48(3): 415–420. doi: 10.3963/j.issn.2095-3844.2024.03.002.

[19] 蒋挺, 杨柳, 刘亚林, 等. 基于分解动态时空分解框架预测交通流量[J]. 科学技术与工程, 2025, 25(7): 3007–3017. doi: 10.12404/j.issn.1671-1815.2307832.

JIANG Ting, YANG Liu, LIU Yalin, *et al*. Traffic flow prediction based on the dynamic spatial-temporal decomposition framework[J]. *Science Technology and Engineering*, 2025, 25(7): 3007–3017. doi: 10.12404/j.issn.1671-1815.2307832.

[20] WEI Siwei, YANG Yang, LIU Donghua, *et al*. Transformer-Based spatiotemporal graph diffusion convolution network for traffic flow forecasting[J]. *Electronics*, 2024, 13(16): 3151. doi: 10.3390/electronics13163151.

[21] YANG Hanqing, WEI Sen, and WANG Yuanqing. STFEformer: Spatial–Temporal fusion embedding transformer for traffic flow prediction[J]. *Applied Sciences*, 2024, 14(10): 4325. doi: 10.3390/app14104325.

[22] ZHU Changfeng, YU Chunxiao, and HUO Jiuyuan. Research on spatio-temporal network prediction model of parallel–series traffic flow based on transformer and GCAT[J]. *Physica A: Statistical Mechanics and its Applications*, 2023, 610: 128414. doi: 10.1016/j.physa.2022.128414.

[23] REN Qianqian, LI Yang, and LIU Yong. Transformer-enhanced periodic temporal convolution network for long short-term traffic flow forecasting[J]. *Expert Systems with Applications*, 2023, 227: 120203. doi: 10.1016/j.eswa.2023.120203.

[24] ZHANG Chen, WU Yue, SHEN Ya, *et al*. Adaptive graph convolutional recurrent network with transformer and whale optimization algorithm for traffic flow prediction[J].

Mathematics, 2024, 12(10): 1493. doi: 10.3390/math12101493.

[25] ZANG Linlin, WANG Tao, ZHANG Bo, *et al*. Transfer learning-based nonstationary traffic flow prediction using AdaRNN and DCORAL[J]. *Expert Systems with Applications*, 2024, 258: 125143. doi: 10.1016/j.eswa.2024.125143.

[26] YIN Lisheng, LIU Pan, WU Yangyang, *et al*. ST-VGBiGRU: A hybrid model for traffic flow prediction with spatio-temporal multimodality[J]. *IEEE Access*, 2023, 11: 54968–54985. doi: 10.1109/ACCESS.2023.3282323.

[27] HUANG Xiaoting, MA Changxi, ZHAO Yongpeng, *et al*. A hybrid model of neural network with VMD–CNN–GRU for traffic flow prediction[J]. *International Journal of Modern Physics C*, 2023, 34(12): 2350159. doi: 10.1142/S0129183123501590.

[28] SUN Haoran, WEI Yanling, HUANG Xueliang, *et al*. Global spatio - temporal dynamic capturing network - based traffic flow prediction[J]. *IET Intelligent Transport Systems*, 2023, 17(6): 1220–1228. doi: 10.1049/itr2.12371.

[29] LI Yaguang and SHAHABI C. A brief overview of machine learning methods for short-term traffic forecasting and future directions[J]. *SIGSPATIAL Special*, 2018, 10(1): 3–9. doi: 10.1145/3231541.3231544.

[30] CHEN Tianqi and GUESTRIN C. XGBoost: A scalable tree boosting system[C]. The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, USA, 2016: 785–794. doi: 10.1145/2939672.2939785.

[31] LAI Qiang and CHEN Peng. LEISN: A long explicit–implicit spatio-temporal network for traffic flow forecasting[J]. *Expert Systems with Applications*, 2024, 245: 123139. doi: 10.1016/j.eswa.2024.123139.

[32] GUO Shengnan, LIN Youfang, FENG Ning, *et al*. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[C]. Proceedings of the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, Honolulu, USA, 2019: 922–929. doi: 10.1609/aaai.v33i01.3301922.

[33] ZHENG Chuanpan, FAN Xiaoliang, WANG Cheng, *et al*. GMAN: A graph multi-attention network for traffic prediction[C]. The 34th AAAI Conference on Artificial Intelligence, New York, USA, 2020: 1234–1241. doi: 10.1609/aaai.v34i01.5477.

[34] LI Yaguang, YU R, SHAHABI C, *et al*. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting[C]. The 6th International Conference on Learning Representations, Vancouver, Canada, 2018: 1–16. doi: 10.48550/arXiv.1707.01926.

[35] WU Zonghan, PAN Shirui, LONG Guodong, *et al*. Connecting the dots: Multivariate time series forecasting with graph neural networks[C]. The 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, California, USA, 2020: 753–763. doi: 10.1145/3394486.3403118.

[36] TIAN Chenyu and CHAN W K. Spatial-temporal attention wavenet: A deep learning framework for traffic prediction considering spatial-temporal dependencies[J]. *IET Intelligent Transport Systems*, 2021, 15(4): 549–561. doi: 10.1049/itr2.12044.

张　红：女，博士，副教授，研究方向为机器学习、大数据分析、智能交通.

伊　敏：女，硕士生，研究方向为智能交通.

张玺君：男，博士，副教授，研究方向为大数据处理、数据挖掘分析、嵌入式系统、机器学习、智能交通.

李　扬：男，硕士生，研究方向为智能交通.

张鹏程：男，硕士生，研究方向为智能交通.

责任编辑：余　蓉

Long-term Transformer and Adaptive Fourier Transform for Dynamic Graph Convolutional Traffic Flow Prediction Study

ZHANG Hong　YI Min　ZHANG Xijun　LI Yang　ZHANG Pengcheng

(*School of Computer & Communication, Lanzhou University of Technology, LanZhou* 730050, *China*)

Abstract:

Objective The rapid development of Intelligent Transportation Systems (ITS), combined with population growth and increasing vehicle ownership, has intensified traffic congestion and road capacity limitations. Accurate and timely traffic flow prediction supports traffic control centers in road management and assists travelers in avoiding congestion and improving travel efficiency, thereby relieving traffic pressure and enhancing road network utilization. However, current models struggle to capture the complex, non-stationary, and long-

term dependent characteristics of traffic systems, limiting prediction accuracy. To address these limitations, deep learning approaches—such as dynamic graph convolution and Transformer architectures—have gained prominence in traffic prediction. This study proposes a dynamic graph convolutional traffic flow prediction model (ADGformer) that integrates a long-term Transformer with an adaptive Fourier transform. The model is designed to more effectively capture long-term temporal dependencies, handle non-stationary patterns, and extract latent dynamic spatiotemporal features from traffic flow data.

Methods The ADGformer model, based on a Long-Term Transformer and Adaptive Fourier Transform, comprises three primary components:a stacked Long-term Gated Convolutional (LGConv) layer, a spatial convolutional layer, and a Multifaceted Fusion Module. The LGConv layer integrates a Masked Sub-series Transformer and dilated gated convolution to extract long-term trend features from temporally segmented traffic flow series, thereby enhancing long-term prediction performance. The Masked Sub-series Transformer—comprising the Subsequent Temporal Representation Learner (STRL) and a self-supervised task head—learns compact and context-rich representations from extended time series sub-segments. The spatial convolutional layer incorporates a dynamic graph constructor that generates learnable graphs based on the hidden states of spatially related nodes. These dynamic graphs are then processed using learnable dynamic graph convolution to extract latent spatial features that evolve over time. To address the non-stationarity of traffic flow sequences, an adaptive spectral block based on Fourier transform is introduced.

Results and Discussions The ADGformer model adaptively models inter-node relationships through a dynamic graph constructor, effectively capturing spatial dependencies within traffic networks and the evolving spatial characteristics of traffic flows. It also learns compressed, context-rich subsequence representations and long-term temporal patterns from extended input sequences. The adaptive spectral block, based on the Fourier transform, reduces the non-stationarity of traffic flow data. To evaluate model performance, this study conducts comparison and ablation experiments on three benchmark datasets. On the PEMSD4 dataset, the ADGformer model reduces MAE under Horizon 3 by approximately 21.57% and 16.72% compared with traditional models such as VAR and LR, respectively. Under Horizon 12 on the England dataset, ADGformer reduces RMSE by 3.35% and 2.44% compared with GWNet and MTGNN, respectively. ADGformer achieves the highest prediction accuracy across all three datasets (Table 2). Visualization comparisons with XGBoost, ASTGCN, and DCRNN on PEMSD4 and England datasets further confirm its superior long-term predictive performance (Fig. 5). The model maintains robust performance as the prediction horizon increases. To assess the contributions of individual components, ablation experiments are performed for Horizons 3, 6, and 12. On Horizon 12 of the PEMSD4 dataset, ADGformer improves MAE by approximately 3.69%, 5.09%, 0.92%, and 7.59% relative to the variant models NMS-Trans, NLGConv, NASB, and NDGConv, respectively. On the England dataset, MAPE is reduced by 3.56%, 18.53%, 2.98%, and 4.87%, respectively (Table 3). Visual results of the ablation study (Fig. 6) show that ADGformer consistently outperforms its variants as the forecast step increases, confirming the effectiveness of each module in the proposed model.

Conclusions To address the limitations of existing traffic flow prediction models—namely, (1) failure to capture hidden dynamic spatio-temporal correlations among road network nodes, (2) insufficient learning of key temporal information and long-term trend features from extended historical sequences, and (3) limited ability to reduce the non-stationarity of traffic data—this study proposes a combinatorial traffic flow prediction model, ADGformer, which enables accurate long-term traffic flow prediction. ADGformer employs a Masked Sub-series Transformer to pre-train long historical sequences and extract compact, information-rich subsequence representations. A gated mechanism combined with one-dimensional dilated convolution is then applied to capture long-term trend features and enhance temporal modeling capacity. The model also integrates a spatial convolutional layer constructed from dynamic traffic patterns. Through learnable dynamic graph convolution, it effectively captures evolving spatial dependencies across the network. Furthermore, an adaptive spectral block based on Fourier transform is designed to enhance feature representation and reduce non-stationarity by applying adaptive thresholding to suppress noise, while preserving both long- and short-term interactions. Given the complex and variable nature of traffic flow, future work will consider integrating additional features such as periodicity to further improve prediction performance.

Key words: Traffic flow prediction; Dynamic graph convolution; Long-term transformer; Fourier transform; Non-stationarity