

时空动态图卷积架构赋能交通速度预测问题

朱雨婷*, 孙德山

辽宁师范大学数学学院, 辽宁 大连

收稿日期: 2025年12月27日; 录用日期: 2026年1月21日; 发布日期: 2026年1月29日

摘要

精准的交通速度预测是智能交通系统造福人类出行的前提, 但现有研究对于交通速度的预测较少, 且难以捕捉与融合路网空间特征和交通速度的时序特征, 导致预测的精度以及泛化能力不足。为充分捕捉交通速度的时序特征和路网的空间特征, 本研究提出一种基于Transformer与图神经网络的STiGHT架构进行时空特征的捕捉, 结合预定义静态图与融合时序特征的动态图进行建模, 利用门控机制融合时空特征进行建模, 对路网交通速度进行预测。基于在真实路网数据METR-LA上的预测任务验证了STiGHT模型预测精度显著优于基线模型, 误差结果对比表明了该模型在实际应用上的可行性。STiGHT模型通过结合动态图、静态图、Transformer架构以及门控机制更好地融合捕捉了时空特征, 提高了路网交通速度的预测精度以及稳定性。

关键词

Transformer, 图神经网络, 门控机制, 交通速度预测

Empowering Traffic Speed Prediction via Spatio-Temporal Dynamic Graph Convolutional Architecture

Yuting Zhu*, Deshan Sun

School of Mathematics, Liaoning Normal University, Dalian Liaoning

Received: December 27, 2025; accepted: January 21, 2026; published: January 29, 2026

Abstract

Accurate traffic speed prediction is a prerequisite for intelligent transportation systems to benefit

*通讯作者。

human travel. However, existing research on traffic speed prediction is limited and struggles to capture and integrate the spatial characteristics of the road network and the temporal characteristics of traffic speed, resulting in insufficient prediction accuracy and generalization ability. To fully capture the temporal characteristics of traffic speed and the spatial characteristics of the road network, this study proposes a STiGHT architecture based on Transformer and graph neural networks to capture spatiotemporal features. It combines predefined static graphs with dynamic graphs that integrate temporal features for modeling and uses a gating mechanism to fuse spatiotemporal features for modeling, thereby predicting the traffic speed of the road network. The prediction task on the real road network data METR-LA verified that the prediction accuracy of the STiGHT model is significantly better than that of the baseline model. The comparison of error results indicates the feasibility of this model in practical applications. The STiGHT model better integrates and captures spatiotemporal features by combining dynamic graphs, static graphs, the transformer architecture, and the gating mechanism, thereby improving the prediction accuracy and stability of traffic speed on the road network.

Keywords

Transformer, Graph Neural Network, Gating Mechanism, Traffic Speed Prediction

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着时代的发展和科技的进步,近年来汽车保有量持续增加,更多人选择私家车出行。然而私家车出行带给人们出行便利的同时,交通拥堵现象也接踵而至,智能交通系统已然成为热门话题。交通预测作为智能交通系统的重要组成部分更是应用广泛[1]。近年来,有关学者对交通预测的研究源源不断,其主要方法可大致分为三类:传统统计学方法、机器学习方法以及深度学习方法[2]。统计学方法如历史平均值法(HA)[3]、差分自回归(ARIMA)[4]等过于依赖平稳性假设,然而交通流数据具有突发性与周期性,这使得传统方法存在局限;此外,传统方法中的参数不具备可学习性,无法准确反映真实交通状况。近年来,随着机器学习的发展,支持向量机(SVM)[5]、随机森林(RF)[6]以及K-近邻模型(KNN)[7]等方法应运而生,机器学习模型通过数据驱动的参数空间映射实现特征表征,其架构具有自适应性特征。紧接着由于数据规模的指数型增长以及反向传播的应用,深度学习理念得到广泛应用,卷积架构被更加频繁地用于交通预测问题。虽然传统卷积神经网络(CNN)[8]能够用于捕获时间依赖性,但其步长往往不可见,缺乏灵活性;循环神经网络(RNN)[9]虽然可以捕捉序列之间的相关性,但耗时较高;长短期记忆网络(LSTM)[10]虽然能够捕捉长期依赖,但对于交通预测的空间依赖性特征捕捉能力欠缺;同时,图神经网络(GNN)[11]可以对空间特征进行捕捉,却忽略了时间依赖。

针对上述问题,本研究旨在设计出一种新的时空特征提取架构,利用双头Transformer架构提取时间特征,利用路网物理拓扑结构以及动态滤波器分别生成静态图与动态图,最后利用门控机制将时空特征解耦融合以提取时空特征,整合形成时空特征提取框架,充分提取交通流数据中的时空依赖,结合Adam优化器以及余弦退火学习率调度优化进行交通速度的预测,提升预测精度与泛化能力。

2. 预备知识

2.1. 交通速度预测问题

交通预测任务可以被表述为一项带有辅助先验知识的多元时间序列预测任务。其先验知识通常为根

据实际道路拓扑结构预定义的邻接矩阵，具体表现为：将道路网络看作一个加权有向图 $G = (v, \varepsilon, A)$ ，其中 v 为包含 $N = |v|$ 个顶点的集合， ε 为连接顶点的边的集合， $A \in R^{N \times N}$ 则表示根据点与边生成的加权邻接矩阵。

交通速度指的是在给定的时间段内通过某个特定路段的车辆的平均速度，交通速度预测任务的目的可概括为在给定的某个路段历史观测数据基础上来预测未来这个路段的交通速度[12]。假定在图 G 上观察到的图信号 $X \in R^{N \times P}$ ，其中 P 为每个节点的特征数量， $X^{(t)}$ 表示 t 时刻观测到的图信号，交通速度预测问题就是期望学习一个函数 f ，使得给定图 G 以及其上 T' 个历史图信号，该函数能够映射得到未来 T 个图信号，即

$$\left[X^{(t-T'+1)}, \dots, X^t, G \right] \xrightarrow{f} \left[X^{(t+1)}, \dots, X^{t+T} \right] \quad (1)$$

2.2. Transformer 模型

Transformer 是 Vaswani [13] 等人于 2017 年所提出的一种基于自注意力机制的深度学习模型，是编码器 - 解码器架构的一个重要实例，没有任何卷积层或循环神经网络层，最初应用于在文本数据上的序列到序列学习，但现在已经推广到各种现代的深度学习例如语言、视觉、语音和强化学习领域。

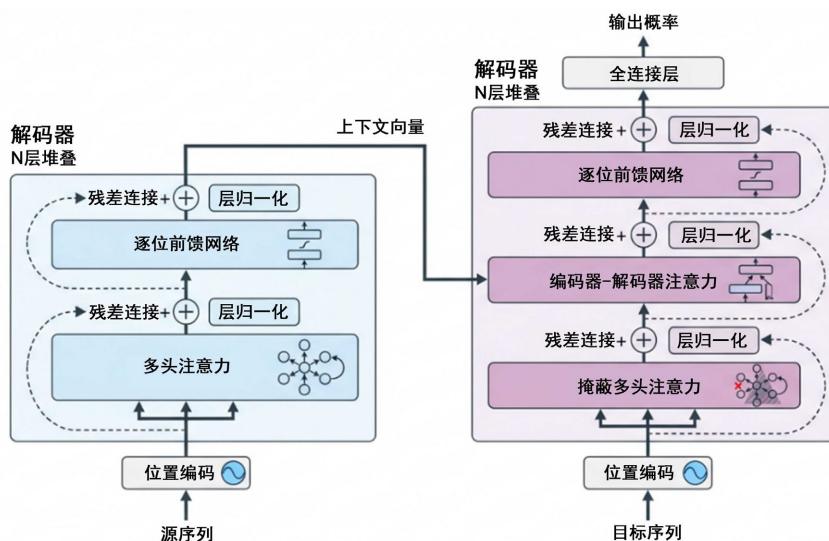


Figure 1. Transformer encoder-decoder architecture

图 1. Transformer 编码器解码器架构

如图 1 所示，Transformer 的编码器由多个包含多头注意力汇聚子层和前馈神经网络子层的相同层叠加而成，且每个子层都采用了残差连接。

Transformer 解码器也是由多个相同的层叠加而成的，并且层中使用了残差连接和层规范化。除了编码器中描述的两个子层之外，解码器还在这两个子层之间插入了第三个子层，称为编码器 - 解码器注意力(encoder-decoder attention)层[14]。

在时间序列预测中，该架构可以捕捉数据中的长期依赖关系和周期性模式，从而提高预测精度。

2.3. 图神经网络

图神经网络在处理非欧空间特征建模方面具有优势，因此在交通预测问题的空间依赖建模中得到广

泛使用。通常被分为谱域图卷积、空域图卷积以及图注意网络三类[15]。

谱域图卷积就是将信号通过图傅里叶变换的形式转换到谱域中进行特征提取, 但过于依赖拉普拉斯矩阵, 因此适合进行静态图建模。而通过空域图卷积进行节点消息传递, 在此基础上提出了扩散图卷积以及常微分方程建模等方式。

近些年由于自注意力机制的出世以及广泛应用, 图注意力网络被提出, 其主要架构为应用图上注意力来加权聚合邻居节点的信息以进行动态图建模。

3. 模型架构

3.1. 时间特征建模

交通数据可以看作时间序列, 下一时刻的交通速度依赖于上一时刻的数据, 且具有周期性与季节性。合理捕获数据中的时间特征是交通速度预测的重中之重[16]。

本文通过设计 Transformer 架构来提取交通数据中的时间依赖关系, 包括早晚高峰的周期性、与交通速度的传播延迟和突发事件等, 结合位置编码来提供时序位置关系。具体建模过程如图 2:

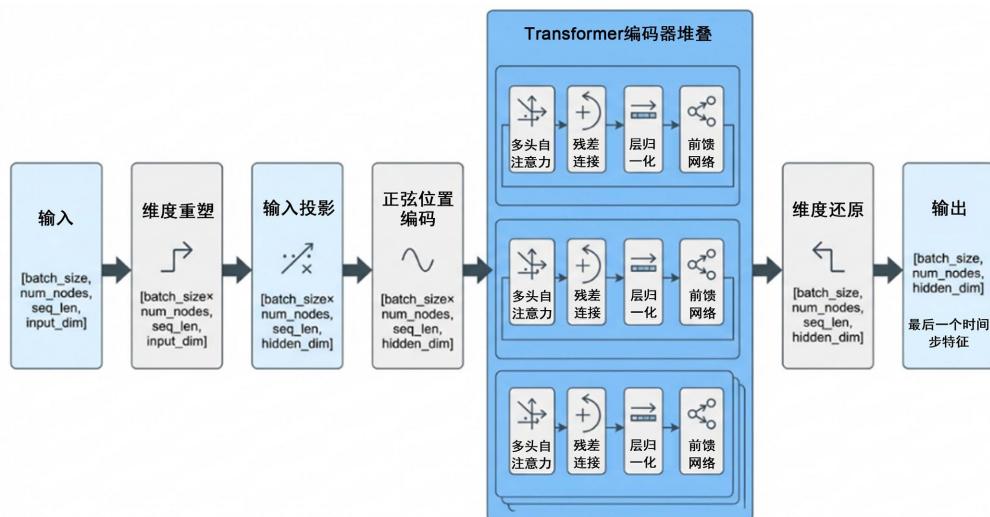


Figure 2. Modeling of time characteristics

图 2. 时间特征建模

首先将原始交通数据进行重塑操作, 合并节点维度与批次维度, 使得每个节点的时间序列都能被独立处理, 随后将特征维度通过线性投影方式进行映射,

$$x_p = W_p \cdot x' + b_p \quad (2)$$

其中 W_p 与 b_p 为可学习参数。

接着进行位置编码的注入。由于 Transformer 架构本身不具备时序位置感知能力, 因此引入正弦余弦位置编码(Sinusoidal Position Encoding)为模型提供时间步的绝对位置信息[17]。对于时间步 pos 与特征维度 i , 定义奇数维度位置编码为:

$$P_E(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/\text{hidden_dim}}}\right) \quad (3)$$

偶数维度位置编码为

$$P_E(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/\text{hidden_dim}}}\right) \quad (4)$$

该编码模式通过周期性函数来建模时间的周期性，且能很好地将该性能泛化到更长的未被学习到的序列。位置编码与投影后的特征通过残差连接进行融合：

$$x'' = x_p + P_E \quad (5)$$

其中 x'' 为融合位置编码的特征信息， P_E 为位置编码矩阵，融合过程中采用了广播机制。

随后将 x'' 输入多层 Transformer 编码器，每一层都由多头自注意力(Multi-Head Attention)和前馈网络(Feed Forward Network)组成，并通过残差连接与层归一化稳定训练。多头注意力中将输入分别映射为查询、键、值矩阵，分割为 h 个并行的注意力头后计算缩放点积注意力，将多头结果品结构整合并通过前馈网络，即两层线性变换进行运算，每个子层通过残差连接将输入与输出进行融合并采用层归一化处理。

$$FFN(x) = \max(0, x \cdot W_1 + b_1) \cdot W_2 + b_2 \quad (6)$$

该时间编码器的核心优势在于其能通过自注意力机制动态地学习时间步之间的关联权重，捕捉非周期性的长期依赖；结合正弦余弦位置编码能够增强对时间周期的建模能力；多层堆叠与残差连接结构能够很好地提高特征表达能力，适配交通数据的复杂动态性特征。

3.2. 空间特征建模

对于空间信息建模，本文不仅仅考虑到静态路网拓扑结构，更要考虑到交通状态的实时性与多变性，所以采取动态静态图建模并应用门控机制进行融合。相对于普通的加权和形式，门控机制能够更好地捕捉动态图与静态图之间的重要性权重，提高模型的泛化能力[18]。

3.2.1. 静态图卷积

静态图卷积是基于传感器分布特点生成预定义的邻接矩阵来捕获路网固定的拓扑结构。将表征静态图拓扑结构的邻接矩阵定义为 $A_s \in R^{N \times N}$ ，其中 $A_{s,ij}$ 表示节点 i 与节点 j 的连接权重：

$$A_{s,ij} = \begin{cases} -\exp(-d_{ij}/\sigma), & i, j \text{ 存在物理连接} \\ 0, & i, j \text{ 不存在物理连接} \end{cases} \quad (7)$$

其中 d_{ij} 为节点间的实际距离， σ 为距离衰减系数。

首先对预定义的邻接矩阵进行标准化处理：

$$A'_s = D^{-\frac{1}{2}} (A_s + I) D^{-\frac{1}{2}} \quad (8)$$

其中 A 为原始邻接矩阵， I 为添加了自环的单位矩阵， D 是度矩阵，其中 $D_{ii} = \sum_j A_{ij}$ ，将标准化的邻接矩阵与通过权重矩阵 W 映射完的特征维度进行图卷积操作来输出静态空间特征 H' ：

$$H_s = A'_s \odot H \otimes W \quad (9)$$

该静态图卷积层通过标准化邻接矩阵定义的卷积核来捕捉交通路网的固定空间关联，同时还能避免梯度消失或梯度爆炸问题，提高模型的泛化能力。

3.2.2. 动态图卷积

由于交通速度具有变化性，仅仅依赖静态图难以捕捉到其变化。为有效捕捉交通网络中节点间的动态演化的空间依赖，设计了一种时间索引动态滤波器，根据时间特征生成动态变化的邻接矩阵来捕捉变化的空间依赖，其核心假设为：节点的空间关联强度由节点特征与时间上下文共同决定[19]。

定义 $\tilde{A}(t) \in R^{N \times N}$ 为时间步 t 的函数, $\tilde{A}_{ij}(t)$ 表示时间步 t 节点 i 与节点 j 的动态关联权重。首先将 Transformer 时间编码器输出的时间特征映射为全局时间嵌入以捕捉当前时间上下文:

$$t = \text{Linear}(f) = f \cdot W_t + b_t \quad (10)$$

其中 f 为时间特征, W_t 、 b_t 为可学习的参数, 分别代表投影权重与偏置。投影过程中采取广播操作使得时间嵌入能够作用于所有可用节点。

接着将节点的时间特征与全局时间嵌入进行拼接, 通过非线性变换生成每个节点的动态权重向量, 表征节点在当前时间上下文下的特征模式。

$$z_i = \text{ReLU}(\text{Linear}([h_i; T_i])) = \text{ReLU}([h_i; T_i] \cdot W_1 + b_1) \quad (11)$$

$$w_i = \sigma(\text{Linear}(z_i)) = \sigma(z_i \cdot W_2 + b_2) \quad (12)$$

其中 h_i 为节点 i 的输入特征, $[h_i; T_i]$ 表示特征拼接, W_1 、 b_1 为中间层权重与偏置, W_2 、 b_2 为输出层权重与偏置, $\sigma(\cdot)$ 为 Sigmoid 激活函数, 将权重约束至 $[0,1]$ 区间, 所有节点生成的动态权重矩阵为 W_d 。

考虑到在真实交通系统中节点间的相互影响往往是非对称性的, 采用双投影法来生成两个独立的节点表征。所有节点生成的动态源权重矩阵为 W_t^s , 动态目标权重矩阵为 W_t^t , 节点间的非对称动态关联权重 $\tilde{A}_{t,ij}$ 由源结点 i 的动态权重向量与目标节点 j 的动态权重向量点积计算而得:

$$\tilde{A}_{t,ij} = W_{t,i}^s \cdot (W_{t,j}^t)^T \quad (13)$$

\tilde{A} 显然为非对称矩阵, 可反映出节点之间的相似度。在计算过程中为保持稳定性需对该矩阵进行和 softmax 标准化操作:

$$\tilde{A}' = \frac{\exp(\tilde{A}_{t,ij})}{\sum_{k=1}^N \exp(\tilde{A}_{t,ik})} \quad (14)$$

最后动态图卷积通过聚合每个节点在动态图中所有目标邻居的信息来更新节点特征:

$$H_d = \sigma(\tilde{A}' \cdot H \cdot W) \quad (15)$$

动态图卷积通过时间特征驱动生成动态的邻接矩阵, 可以做到自适应地捕捉空间依赖, 具有很强的泛化能力。其中 T_t 动态滤波器的结构使得时空信息能够有效耦合时空信息, 增强鲁棒性。

3.3. 时空依赖整合

在动态图与静态图分别建立邻接矩阵的基础上, 本文采取门控融合机制而不是简单加权和形式将二者进行有机结合。融合的核心思想为: 根据当前提取的时间与空间特征整合得到当前交通速度, 自动学习静态拓扑与动态关联的重要性占比。对于每个节点, 门控系数取值范围为 $[0,1]$, 系数越接近 1, 说明模型越依赖静态路网拓扑, 反之系数越接近 0, 模型越侧重于动态的空间关联[20]。

首先对静态图卷积输出的特征 H_s 和动态图卷积输出的特征 H_d 进行线性投影, 进行统一特征分布的预处理方式增强表达能力, 接着将二者进行拼接, 通过非线性变换生成逐个节点, 逐个维度的门控系数 G 表征静态特征权重占比, 最后基于门控系数完成对二者加权融合, 输出整合后的空间特征 H_f

$$H'_s = H_s \cdot W_s + b_s \quad (16)$$

$$H'_d = H_d \cdot W_d + b_d \quad (17)$$

$$G = \sigma(\text{Linear}[H'_s; H'_d]) = \sigma([H'_s; H'_d] \cdot W_g + b_g) \quad (18)$$

$$H_f = G \odot H'_s + (1 - G) \odot H'_d \quad (19)$$

其中 $[; \cdot]$ 表示拼接, 拼接后维度为 $R^{B \times N \times 2H}$, W_g 、 b_g 为门控层可学习参数, \odot 为 Hadamard 积。

融合过程中可实现逐维度自适应, 端到端的学习与残差特性兼容, 通过加入残差连接来提升训练稳定性:

$$H_f^{res} = \text{LayerNorm}(G \odot H'_s + (1 - G) \odot H'_d + H_{input}) \quad (20)$$

H_{input} 为融合模块输入特征, 即 Transformer 时间编码输出, LayerNorm 为层归一化操作来稳定特征分布。该融合模块输出为整合后的时空特征, 该特征包含交通网络的固定拓扑信息、随时间演化的空间关联信息以及时间序列的依赖信息, 做到时空依赖的融合, 更加适用于交通预测。

4. 实验

4.1. 实验设置

4.1.1. 数据集

为了验证 STiGHT 模型的效果, 本研究在真实数据集 META-LA 上进行试验, META-LA 数据为洛杉矶高速公路环形检测器收集的公共交通速度数据集, 选取时间范围为 2012 年 3 月 1 日至 2012 年 6 月 30 日, 包含 207 个传感器, 采样频率为 5 分钟, 适配交通速度预测问题。

4.1.2. 参数设置

本研究中所涉及的参数可分为三类: 模型结构参数、训练超参数和特征维度参数。针对可训练参数选取不同初始化方式, 如表 1 所示:

Table 1. Parameter categories and initialization methods

表 1. 参数类别与初始化方法

参数类别	线性层权重	偏置项	图卷积权重	Transformer 位置编码
初始化方法	Xavier 正态初始化	0	正交初始化	固定正余弦编码

为防止模型出现过拟合, 增强模型泛化能力, 选取正则化参数如表 2:

Table 2. Values of regularization parameters

表 2. 正则化参数取值

参数名称	Dropout	L2 正则化系数	LeakyReLU 负斜率
数值	0.1/0.2	1e-4	0.2

4.1.3. 训练策略

本文所提出的 STiGHT 架构采用端到端的多步预测训练策略, 首先进行数据预处理时采用 Z-Score 标准化方法进行, 应用滑动窗口构造序列, 划分数据集, 其中训练集: 测试集 = 7:3。

该模型采用 Adam 优化器进行训练, 同时采用余弦退火调度(CosineAnnealingLR)与预热(Warmup)策略调度学习率。在训练过程中采用梯度裁剪策略防止出现梯度爆炸, 结合早停机制与混合精度训练, 实现适配入门级 GPU 的轻量化模型架构, 同时保证预测精度。

4.1.4. 评估标准

本研究采用多损失融合策略, 融合 3 类评估标准如表 3 所示, 采用均方误差损失作为损失函数, 兼顾鲁棒性与误差约束, 其中所有指标均基于反归一化的数据计算而来。

Table 3. Calculation formula of evaluation function

表 3. 评估函数计算公式

指标名称	符号	公式
平均绝对误差	MAE	$MAE = \frac{1}{M} \sum_{i=1}^M (\hat{y}_i - y_i)^2$
均方根误差	RMSE	$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M (\hat{y}_i - y_i)^2}$
平均绝对百分比误差	MAPE	$MAPE = \frac{1}{M} \sum_{i=1}^M \left \frac{\hat{y}_i - y_i}{y_i} \right $

4.2. 实验结果(图 3, 图 4)

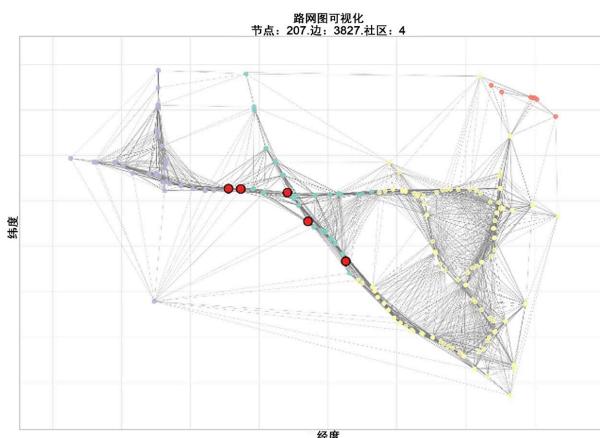


Figure 3. Road network topology structure

图 3. 路网拓扑结构

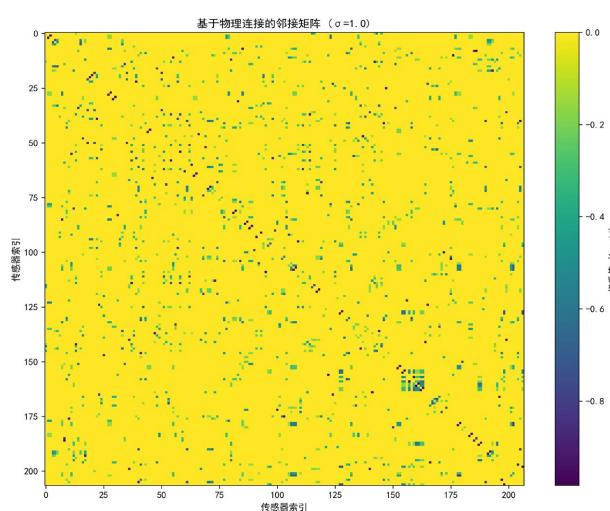


Figure 4. Sensor distribution heat map

图 4. 传感器分布热力图

基于上述模型以及实验设置, 本研究在 METR-LA 数据集上进行预测。秉持着让初学者也能够更好地学习交通预测问题的初心, 本研究所应用的架构相对轻量化, 适配于入门级 GPU。

基于该数据 207 个传感器节点构建邻接矩阵与物理连接距离分布图(如图 5 所示), 为后续空间特征提取做准备。

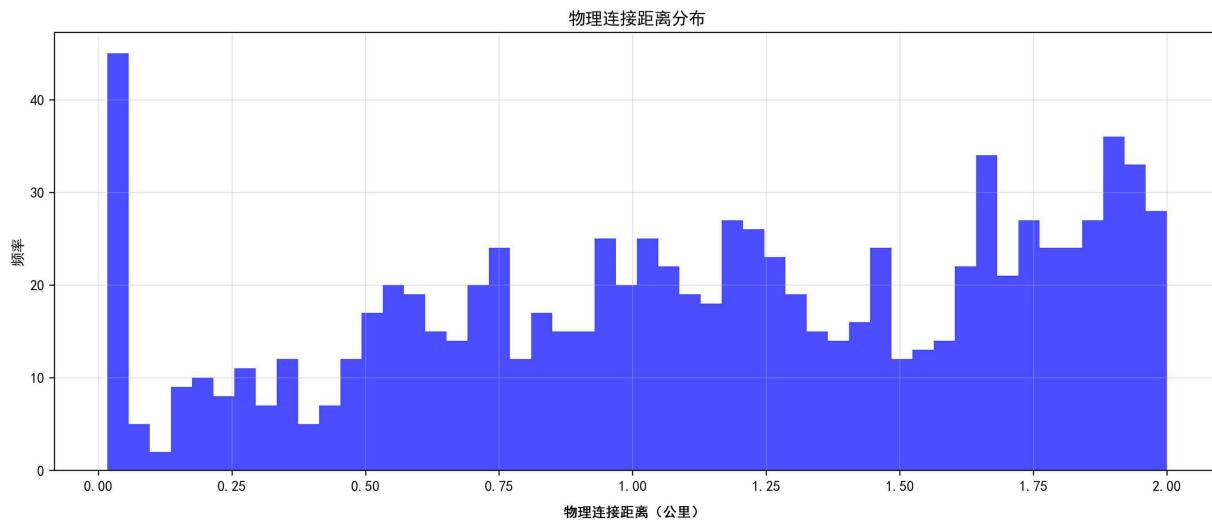


Figure 5. Distribution map of sensor physical distance

图 5. 传感器物理距离分布图

实验对 15 分钟、30 分钟、45 分钟以及 60 分钟的交通信息进行预测, 训练历史如图 6 所示:

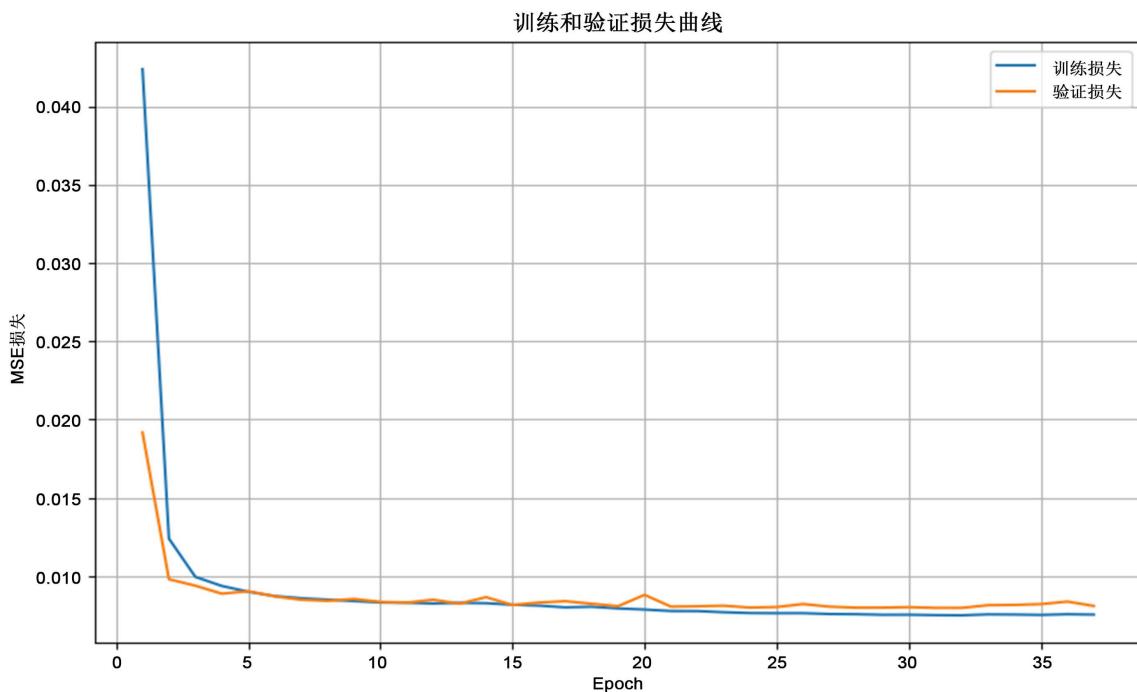


Figure 6. Training loss curve of the STiGHT model

图 6. STiGHT 模型训练损失曲线

由图像可以看出, 模型的训练损失随着迭代轮次的进行而不断下降, 最终达到收敛。测试损失也有相同趋势且在正常范围内波动, 二者具有良好的泛化能力。

4.3. 基线模型对比

接着为了进一步验证 STiGHT 模型的有效性, 本文将其与传统基线模型进行了对比, 基线模型介绍如下:

ARIMA: 差分移动平均模型是基于统计学方法的模型, 广泛应用于时间序列预测问题。

SVM: 支持向量机模型, 核心思想为寻找一个回归超平面使得绝大多数样本的预测误差落在预设 ε -不敏感带内, 同时最大化超平面间隔。

RF: 随机森林模型, 基于集成学习的 Bagging 算法, 核心思想为通过随机采样构建多棵决策树。

Graph WaveNet: 基于 GNN 与 RNN 架构, 同时整合扩散图卷积与门控一维膨胀卷积来进行预测。

对于所有基线模型, 本文采取与 STiGHT 模型相同的数据预处理方式与相同的损失函数进行评估。

所有模型在不同时间步的 MAE、RMSE 以及 MAPE 值如表 4 所示: 可以看出随着预测时间步的增加, 所有模型的损失都有所增加, 这表明短时预测效果比长时间预测效果要好。长期预测可能出现更多的变数与不确定性, 因此损失会有所升高。STiGHT 在交通速度预测问题上相较于其他基线模型损失较低, 15 分钟预测相对准确, 且损失随着时间步的增加没有过于波动, 说明 STiGHT 模型具有良好的预测性能与泛化能力。

Table 4. Comparison results of baseline models

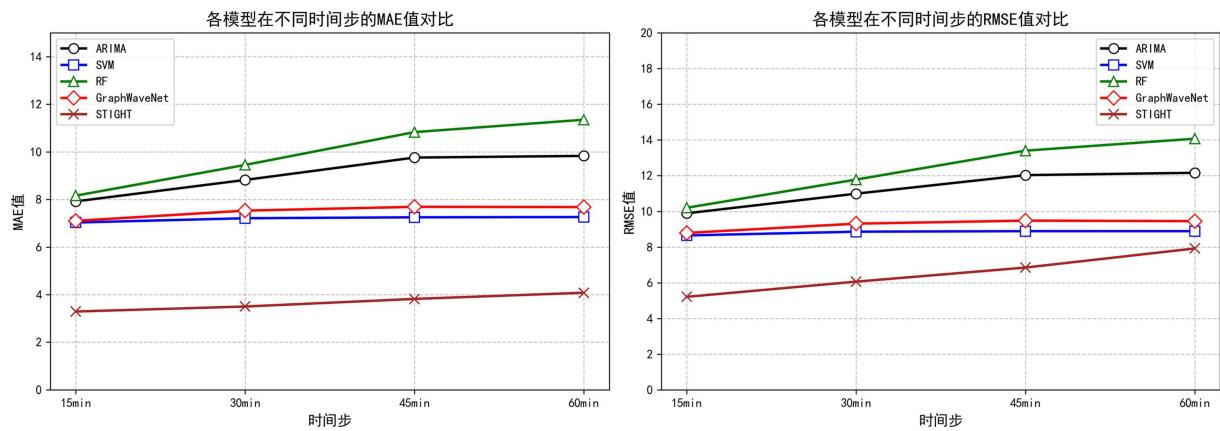
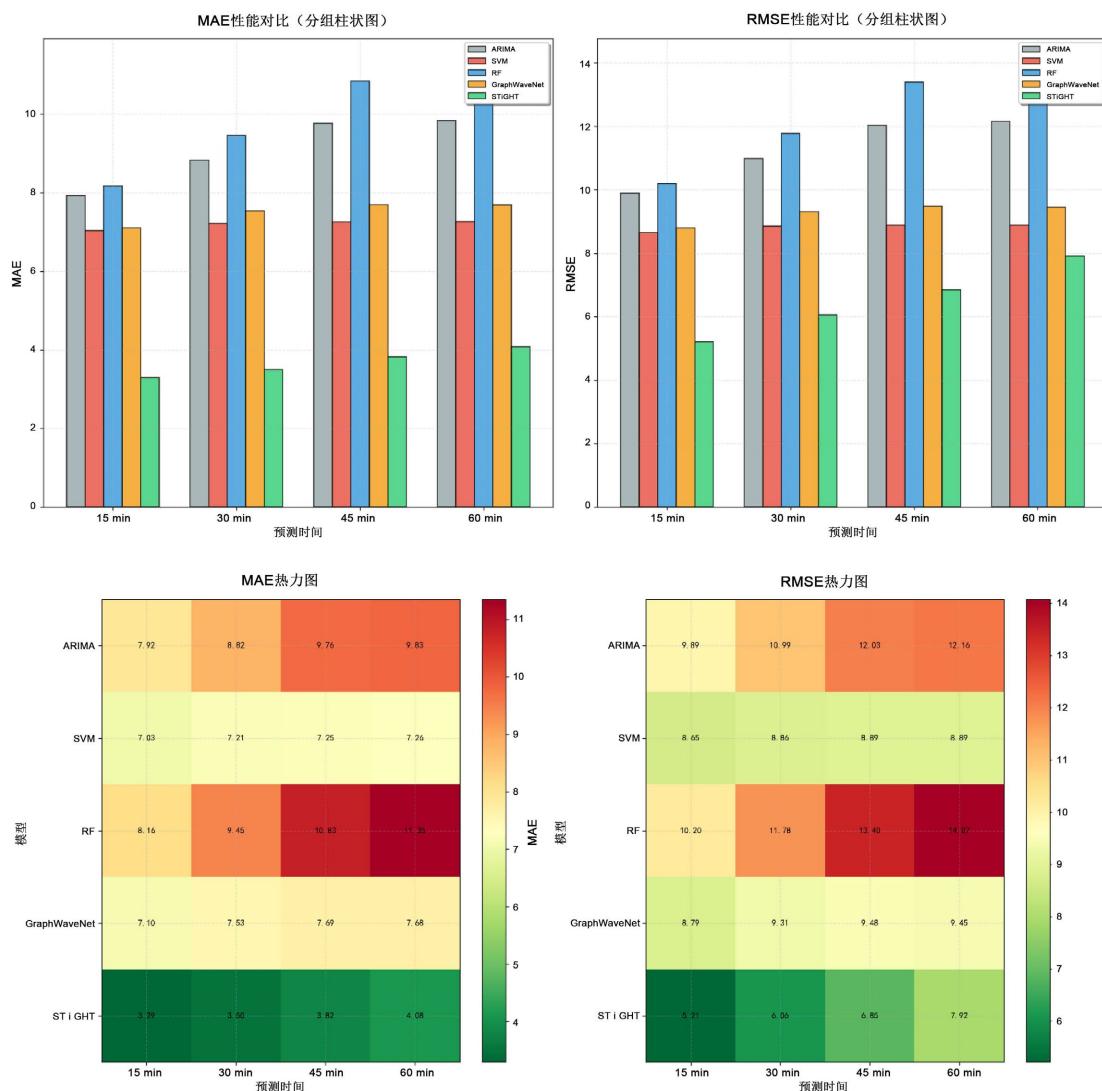
表 4. 基线模型对比结果

模型	15 min			30 min			45 min			60 min		
	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
ARIMA	7.92	9.89	11.78	8.82	10.99	13.19	9.76	12.03	14.67	9.83	12.16	14.78
SVM	7.03	8.65	10.56	7.21	8.86	10.83	7.25	8.89	10.88	7.26	8.89	10.88
RF	8.16	10.20	12.14	9.45	11.78	14.12	10.83	13.40	16.29	11.35	14.07	17.06
GraphWaveNet	7.10	8.79	10.66	7.53	9.31	11.34	7.69	9.48	11.58	7.68	9.45	11.55
STiGHT	3.29	5.21	7.54	3.50	6.06	8.84	3.82	6.85	10.41	4.08	7.92	11.61

各个模型不同时间步下的 MAE 与 RMSE 变化曲线(图 8)更能够直观地反映出各个模型的预测性能。基线对比柱状图与热力图能够更加清晰地看出 STiGHT 模型的优势所在。

由图 7 可以看出, STiGHT 模型在所有时间步和所有评估指标上都显著优于其他基线模型, 特别是在短期预测中优势最为明显。虽然随着预测时间步的增加, MAE 与 RMSE 值均有所升高, 但波动范围不大。SVM 在基线模型中表现最佳, 且计算复杂度低, 在资源受限场景下仍有应用价值。ARIMA 与 RF 模型在长时间预测时的 MAE 与 RMSE 会明显升高, 说明在长时间预测上存在局限性。其他方法损失值虽然没有随着时间步的增加而显著升高, 但损失值较大。GraphWaveNet 作为图神经网络, 能够有效捕捉交通网络的空间依赖关系, 在中长期预测中表现优异。STiGHT 模型不仅具有较低的损失值, 而且在时间步增加过程中近似呈现出直线状态, 说明其不仅预测准确率高, 稳定性强, 而且具备长时预测的能力。

综上所述, STiGHT 模型不仅误差小, 而且在不同预测时长下都保持稳定表现, 这对于实际交通管理决策至关重要。

**Figure 7.** Comparison of MAE and RMSE for each model at different time steps**图 7.** 各个模型不同时间步 MAE 与 RMSE 对比**Figure 8.** Performance comparison analysis**图 8.** 性能对比分析

4.4. 消融实验

为验证 STiGHT 模型中所采用的动静态图门控结合方式的有效性, 设置如下表 5 消融实验变体名称
表 5 消融实验:

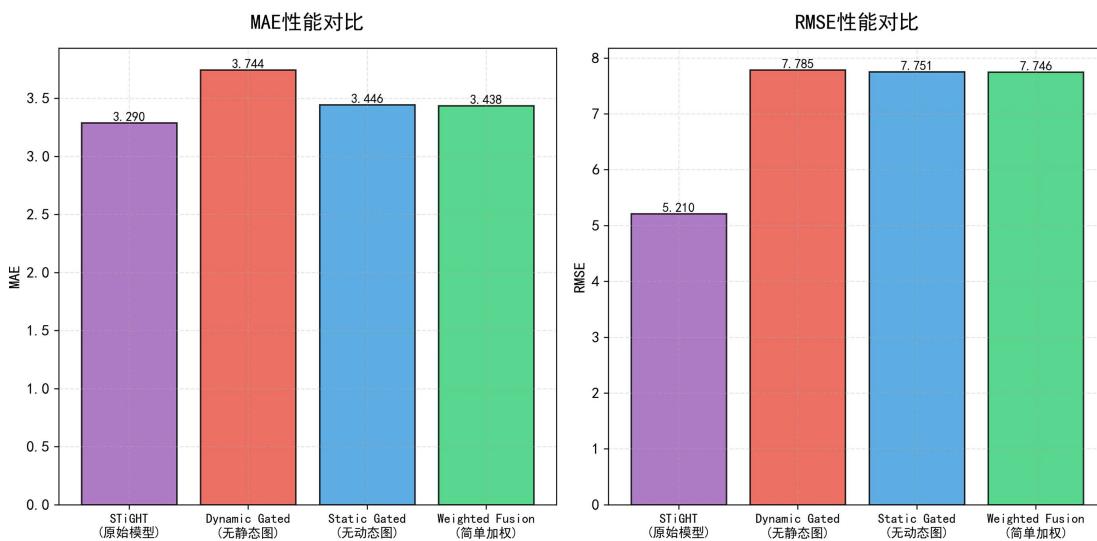
Table 5. Variant name of ablation experiment**表 5. 消融实验变体名称**

实验名称	符号表示
无预定义静态图的模型	Dynamic Gated
无动态图的模型	Static Gated
无门控融合机制应用简单加权方法的模型	Weighted Fusion
原始模型	STiGHT

消融实验采用去掉单个模块形式来验证此模块的重要性, 分别验证了预定义静态图、动态图以及门控融合机制的重要性。消融实验在真实数据 METR-LA 上进行了 15 分钟预测, 消融实验进行 5 次, 取平均值得出表 6 结果。根据实验结果绘制柱状图如图 9 所示来直观对比模型变体和原模型的性能。

Table 6. Comparison of ablation experiment results**表 6. 消融实验结果对比**

模型变体	MAE	RMSE
Dynamic Gated	3.74	7.79
Static Gated	3.45	7.75
Weighted Fusion	3.44	7.75
STiGHT	3.29	5.21

**Figure 9.** Comparison of ablation experiment performance**图 9. 消融实验性能对比**

无论是实验结果还是性能对比图都表明 STiGHT 模型表现得最好，无论是去掉预定义静态路网拓扑结构静态图，还是动态路网结构图，或是将动静态图进行简单加权融合而非门控机制整合，都会使模型有效性下降，静态图与动态图模块缺一不可，且简单加权方法远不如门控机制。故得出结论，本文所提出的 STiGHT 模型效果最佳。

参考文献

- [1] 郭璐. 智能交通系统在智慧城市建设中的作用与优化路径[J]. 城市建设理论研究(电子版), 2024(32): 226-228.
- [2] Tedjopurnomo, D.A., Bao, Z., Zheng, B., Choudhury, F. and Qin, A.K. (2020) A Survey on Modern Deep Neural Network for Traffic Prediction: Trends, Methods and Challenges. *IEEE Transactions on Knowledge and Data Engineering*, **34**, 1544-1561. <https://doi.org/10.1109/tkde.2020.3001195>
- [3] Xu, X., Zhang, L., Kong, Q., Gui, C. and Zhang, X. (2022) Enhanced-Historical Average for Long-Term Prediction. *2022 2nd IEEE International Conference on Computer, Control and Robotics (ICCCR)*, Shanghai, 18-20 March 2022, 115-119. <https://doi.org/10.1109/iccr54399.2022.9790148>
- [4] Shahriari, S., Ghasri, M., Sisson, S.A. and Rashidi, T. (2020) Ensemble of ARIMA: Combining Parametric and Bootstrapping Technique for Traffic Flow Prediction. *Transportmetrica A: Transport Science*, **16**, 1552-1573. <https://doi.org/10.1080/23249935.2020.1764662>
- [5] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. <https://doi.org/10.1023/a:1022627411411>
- [6] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/a:1010933404324>
- [7] Cover, T. and Hart, P. (1967) Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, **13**, 21-27. <https://doi.org/10.1109/tit.1967.1053964>
- [8] Zhang, W., Yu, Y., Qi, Y., Shu, F. and Wang, Y. (2019) Short-Term Traffic Flow Prediction Based on Spatio-Temporal Analysis and CNN Deep Learning. *Transportmetrica A: Transport Science*, **15**, 1688-1711. <https://doi.org/10.1080/23249935.2019.1637966>
- [9] Williams, R.J. and Zipser, D. (1989) A Learning Algorithm for Continually Running Fully Recurrent Neural Networks. *Neural Computation*, **1**, 270-280. <https://doi.org/10.1162/neco.1989.1.2.270>
- [10] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [11] Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M. and Monfardini, G. (2009) The Graph Neural Network Model. *IEEE Transactions on Neural Networks*, **20**, 61-80. <https://doi.org/10.1109/tnn.2008.2005605>
- [12] 黄佳妮. 基于时空特征学习的交通流量预测研究[D]: [硕士学位论文]. 南京: 南京林业大学, 2025.
- [13] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.
- [14] 阿斯顿·张, 扎卡里·C·立顿, 李沐, 等. 动手学深度学习[M]. 北京: 人民邮电出版社, 2023: 601.
- [15] 贾兴利, 曲远海, 朱浩然, 等. 交通预测中的时空图神经网络研究综述: 从模型解构到发展路径[J/OL]. 交通运输工程学报, 2025: 1-35. <https://doi.org/10.19818/j.cnki.1671-1637.2026.01.003>, 2025-12-23.
- [16] Yu, B., Yin, H. and Zhu, Z. (2018) Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, Stockholm, 13-19 July 2018, 3634-3640. <https://doi.org/10.24963/ijcai.2018/505>
- [17] 唐光旭. 基于 Transformer 和 Bi-LSTM 模型的船舶轨迹预测[J]. 珠江水运, 2024(17): 109-111.
- [18] Li, F., Feng, J., Yan, H., Jin, G., Yang, F., Sun, F., et al. (2023) Dynamic Graph Convolutional Recurrent Network for Traffic Prediction: Benchmark and Solution. *ACM Transactions on Knowledge Discovery from Data*, **17**, 1-21. <https://doi.org/10.1145/3532611>
- [19] Li, Y., Yu, R., Shahabi, C., et al. (2017) Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. *6th International Conference on Learning Representations, ICLR 2018*, Vancouver, 30 April-3 May 2018.
- [20] 唐进君, 段一鑫, 商淑杰, 等. 多特征融合时空图卷积的高速公路交通流量预测[J]. 公路交通科技, 2025, 42(6): 22-31.