**相关工作：**

**交通流量预测**

交通流量预测是智慧交通系统中的一个重要研究方向，旨在基于历史数据对未来一段时间的交通信息进行预测。基于图学习的相关方法近年来被广泛应用于交通预测领域。DCRNN[25] 通过扩散卷积网络建模空间关系，并利用循环神经网络捕捉时间依赖性，从而实现对交通数据的时空特性进行有效表征。一些研究[30-32] 使用节点间的欧氏距离来度量空间相关性，这种方法虽然能够直观地反映交通节点在地理空间中的分布关系，但仍存在局限性。因此，研究[33-38] 利用节点特征去构建动态交通网络，并通过图神经网络 (GNNs) 在动态交通图上进行建模，从而更加灵活地捕获时空关系。在深层 GNN中，多次迭代的消息传递会导致节点特征趋同，导致过平滑现象[39-40] 。为解决这些问题，基于注意力机制的交通预测模型[41-46] 受到研究者的广泛关注，显著提升了模型对远距离时空依赖关系的建模能力。

同时，研究者们探索了多层感知机（MLP）在交通预测领域的应用[47-49] ，旨在降低模型复杂度并减少训练开销。类似地，基于神经架构搜索的交通预测模型[50-53] 通过自动化搜索最优模型架构，进一步节省了人工设计成本，同时保证了预测结果的准确性。近年来，面向复杂交通场景的预测问题（如少样本交通预测和跨城市交通预测）也受到广泛关注，研究者致力于提升模型在复杂交通场景的预测鲁棒性和准确性。其中，少样本交通预测问题聚焦于在数据稀缺的条件下实现准确的流量预测，常通过迁移学习[54-56] 和元学习[57] 等技术来提升模型的泛化能力。跨城市预测模型[58-62] 则关注如何将已训练的模型迁移到新的城市场景，其核心挑战在于如何有效应对不同城市间交通模式的差异性。由此可见，城市交通预测作为智慧城市建设的核心环节，已吸引了众多研究者的广泛关注。随着交通数据的多样性和复杂性不断增加，构建鲁棒性强、泛化能力优异的交通预测模型已成为当前的重点研究方法。

研究方法：

**基于自适应时空上下文学习的交通流量预测模型**

1. **引言**

随着现代城市的快速发展和城市化进程的推进，交通预测已成为智能交通系统和智慧城市建设中的关键要素。精准的交通预测对于拥堵管理、路径规划及应急响应等应用场景至关重要。通过分析历史数据，城市交通系统旨在预测路网未来状态（如流量、速度与密度等指标）[1]。有效的交通预测能显著提升城市交通系统运行效率、缩短出行时间，并改善居民整体生活质量。

近年来，随着城市交通系统复杂度的持续提升，交通预测研究呈现爆发式增长。这类研究通常将交通网络建模为时空图结构：交通传感器对应图中的节点，节点间的相互依赖关系则构成边[2]。图神经网络（GNN）因其能通过消息传递机制聚合邻域节点信息的特点[1][3][4]，已成为交通预测的主流方法。但基于GNN的模型存在过度平滑问题，可能导致节点表征趋同[5][6]；同时由于处理复杂图结构时的高计算与内存开销，其效率瓶颈也备受诟病[7]。此外，受注意力机制在时空建模中成功应用的启发，学界相继提出多种基于注意力的模型[8][9]。这些方法通过加权时空特征的重要性，利用注意力机制来提升模型性能。

现有交通预测方法通常基于大规模数据集中存在隐含不变时空模式的假设。然而，由于时空数据固有的动态特性，历史训练数据与未来测试数据之间可能出现不可预见的分布偏移[10]。交通模式极易受多种外部因素影响，包括不可预测的社会事件、气候变化导致的天气异常、区域建设规划以及城市经济发展等。例如在某些意外期间，大规模封锁政策导致交通流量骤减，这类突发情况可能引发交通模式的剧烈变化，从而显著降低预测模型的有效性。

为应对这一挑战，我们从一个新颖的视角探索交通预测问题，强调以快速自适应方式学习有效时空表征的重要性。我们提出基于日粒度数据的自适应交通预测方法，仅利用单日数据即可训练出适应动态交通场景的模型。这种每日自适应机制使模型能够快速响应由外部因素引起的交通模式变化。通过每日增量训练，模型既能捕捉最新交通趋势，又能及时应对特殊事件或极端天气等突发状况。这种实时响应能力对于维持动态环境中交通预测的准确性与可靠性至关重要。如图1所示，该方法不仅达到了理想的预测精度，还能灵活适应跨城市、跨领域等多种应用场景。

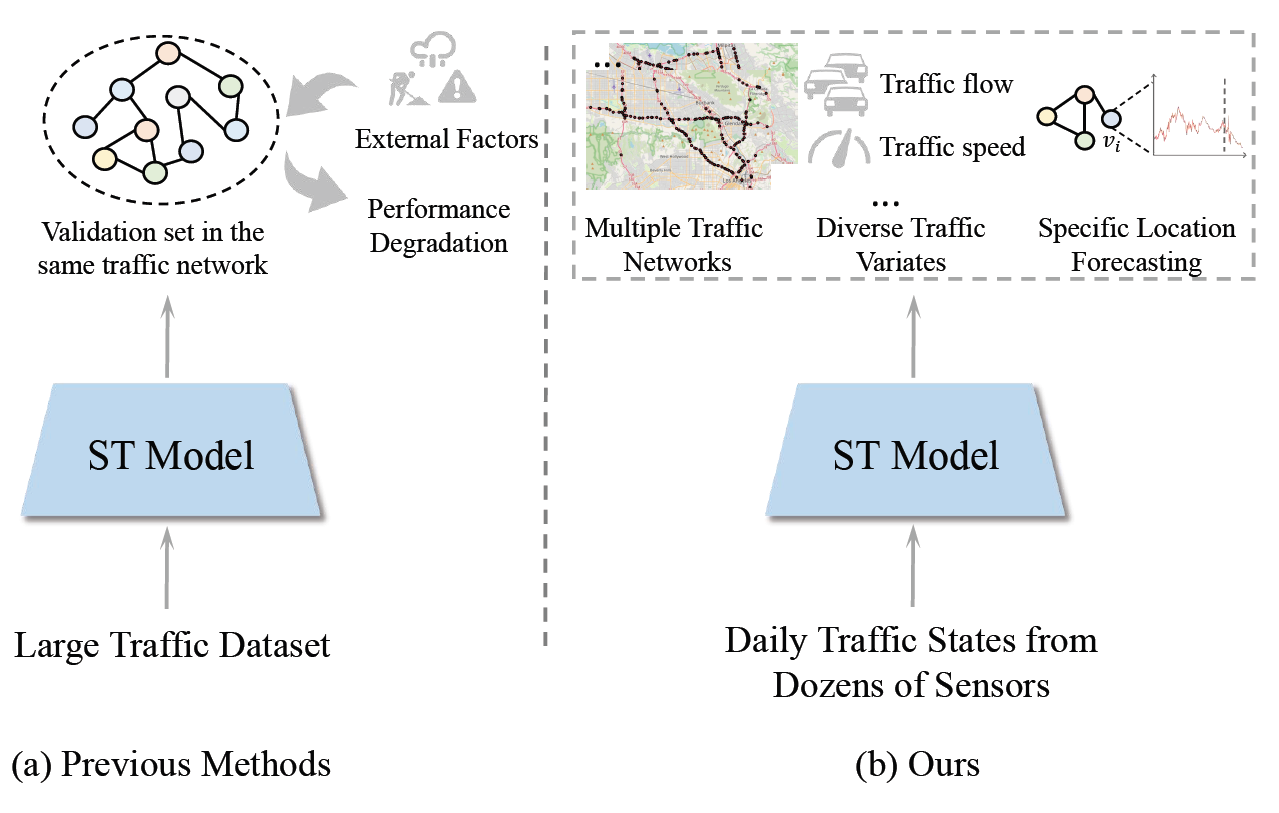


图1、交通预测框架

与传统方法依赖长期历史数据挖掘不变时空模式不同，本研究聚焦实时交通信息与应用场景特性，提出了一种新型自适应时空上下文学习方法（ASTCL），可实现高精度的日粒度交通预测。本方法创新性地设计了节点级预测任务——以目标节点（如单个交通传感器）的信息聚合作为模型输入，而非全路网数据。这种设计具有三重优势：精准定位​：针对特定关注区域进行定向预测，相比城市级路网预测更具成本效益；​强迁移性​：训练完成的模型可突破特定路网结构限制，适用于城市内部其他区域或跨城市应用；数据增效​：基于图结构的节点级预测使训练样本量与交通节点数量成正比，有效缓解日粒度预测中的数据稀缺问题。

针对单一节点数据存在的单调性与稀疏性缺陷，本研究通过构建自适应时空上下文进行全局信息补偿：基于语义相似度，为每个目标节点动态整合相同时段内其他传感器的历史数据。这种上下文增强机制通过聚合具有时空语义关联的节点信息，显著提升了数据多样性与时序覆盖度。为建模动态上下文中的节点关系，本工作提出动态序列图结构，其边权重随时间序列动态演化。在此基础上，创新性地设计了时空联合模块，通过联合注意力机制与联合卷积操作，同步捕捉空间、时间及时空交叉维度的复杂关联。ASTCL仅包含数千参数，模型复杂度与计算需求显著低于主流方法，四个真实交通数据集上的实验表明，ASTCL均稳定超越现有最优基线模型，各项指标提升幅度达1.2%-5.7%

**方法描述**：

本研究将由交通传感器收集的历史交通数据表示为 ，其中，代表可用传感器的数量，表示时间范围，代表所研究的交通特征。在此基础上，采用滑动窗口机制将数据序列划分为多个时间区间，每个窗口大小为 ，从而得到 。通常情况下，直接作为模型输入，并依据的大小进行多轮训练。然而，在基于单日数据的交通预测场景中，由于每天的数据量有限，的取值通常仅为数百，导致生成的训练样本数量显著受限。

为应对这一挑战，本研究提出一种基于节点级预测的训练任务。具体而言，对于交通网络中的目标节点（即交通传感器），模型以其自适应时空上下文作为输入，预测其未来 个时间步的交通状态，而非整个交通图的状态。通过这一设计，训练样本数量将随着交通传感器数量的增加而成倍增长，从而有效缓解数据稀缺问题。该预测任务的定义如下：



其中，为预测结果。

尽管节点级预测能够扩充训练样本数量，但单个节点的交通数据往往缺乏多样性，难以提供足够的信息支持模型的稳定训练。近期研究表明，节点的交通状态主要受到其局部邻域的影响，因此通过融合相邻节点的信息构建目标节点的局部或区域上下文。然而，这些方法未能充分考虑全局空间的影响，限制了模型的预测能力。为此，本研究提出自适应时空上下文的概念，旨在从全局视角为目标节点引入更加多样化的数据，使模型能够捕捉更广泛的交通环境及其对目标节点的影响模式。在时间窗口 内，目标节点 的特征可表示为：。本文的目标是自适应地识别对目标节点相关性强的节点集合，并将这些节点聚合为自适应时空上下文。

为了量化不同节点间交通状态的相关性，本研究采用快速动态时间规整（fastDTW）算法 ，该方法基于经典的动态时间规整（DTW）技术，通过fastDTW计算得到的数值可用于衡量节点间交通模式的相似程度，从而有效评估两个节点在相同时段内的序列相似性。节点与之间的相关性权重通过高斯核函数定义如下：



其中，表示与之间的相关性权重，权重值越高表明二者关系越密切；表示fastDTW计算；为所有值的标准差。基于的值，本研究选择与目标节点交通模式最相似的个节点，并将其特征拼接以构建自适应时空上下文。目标节点本身也被包含在该自适应上下文中，以确保其自身的时空模式得到保留。这种方法使模型能够动态地从全局交通网络中选择相关信息，从而增强预测的鲁棒性和准确性。

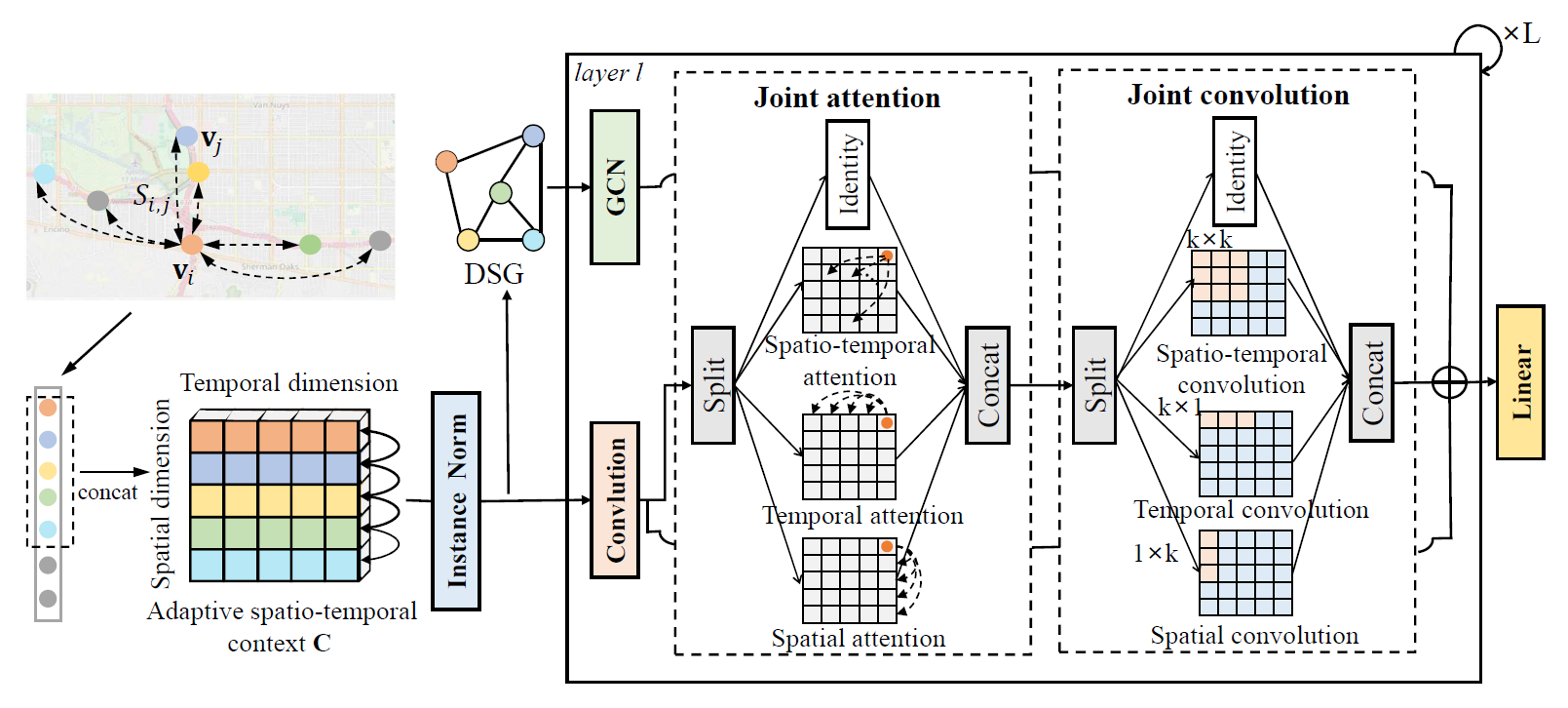
ASTCL的模型设计与实现，主要包括时空联合学习、动态序列建模和模型预测三个方面的内容，ASTCL模型的整体架构如图2所示。ASTCL的核心任务是从自适应时空上下文中捕捉复杂的时空相关性，并预测交通状态的未来趋势。在时间维度上，历史交通数据揭示了随时间推移不断重复的模式和趋势；在空间维度上，目标节点的交通状况通常受到自适应上下文中其他节点的影响，这种影响源于交通节点间的语义相似性。以往的研究通常采用独立的空间、时间和时空编码器，并将它们按序堆叠以提取各维度的特征。然而，这种堆叠策略不仅增加了模型的深度和复杂性，还可能导致不同维度之间交互信息的丢失。本研究提出了两种联合时空学习方法：联合注意力机制和联合卷积。这两种操作能够在单层编码器中同时学习空间、时间和时空表示，同时保持较低的计算复杂度。

给定自适应时空上下文作为输入，ASTCL首先对输入数据应用实例归一化，归一化后的数据将在模型的最终预测阶段进行反归一化。如图2所示，归一化后的时空上下文经过一层卷积神经网络得到自适应时空上下文的初始特征嵌入，特征维度从转换为，得到特征张量 。为了学习更高层次的时空表示，本文将特征提取模块进行了次堆叠，对于第 层的输出，第一步的卷积计算过程如下：



其中，是学习到的特征嵌入； 是第 层的输出；\ 表示卷积操作；是激活函数；是可学习参数。

为了充分利用自注意力机制在捕捉长距离依赖关系方面的优势，本研究设计了一种联合时空注意力机制，旨在单阶段内高效建模空间、时间和时空相关性。具体而言，首先将 根据特征通道均匀地划分为四部分，得到 、、 和 ，每部分都是具有相同特征维度的张量。然后，针对每个特征划分利用不同的注意力机制来捕捉各种依赖关系。对于，本项目利用多头空间自注意力机制来捕捉同一时间步内节点之间的空间



图、ASTCL模型框架图

依赖关系。具体地，对于时间戳 ，首先通过特征变换得到查询（Q）、键（K）和值（V）矩阵，公式如下：



其中，、 和 是可学习参数。接着，通过自注意力操作获得时间步  上所有节点的注意力图。空间注意力图通过计算同一时间步内不同节点之间的成对关系，揭示了沿时间维度不变的空间属性。最终，空间自注意力的输出通过以下公式计算：



对于，本方法应用时间自注意力机制来捕捉每个节点在时间维度上的相关性。对于则应用时空注意力机制来捕捉动态的时空相关性，全面理解交通模式在时间和空间上的演变规律。为了保持特征一致性并避免模型过度复杂化，对最后一个部分 应用恒等映射，直接保留其原始特征。随后将所有输出进行拼接，并通过通道重排操作重新排列学习到的特征，最终得到联合注意力模块的输出。

在联合注意力模块之后，为了建模局部的时空相关性，本研究提出一种联合卷积操作。首先采用与联合注意力模块类似的分割策略将划分为多个部分，并对每部分应用不同的卷积模块。这里使用三种不同的卷积核大小：、和 ，分别从空间、时间和时空视角提取特征。最后，联合卷积模块的输出通过残差连接与输入特征融合，其公式如下：



其中，、 和 是不同卷积核的可学习参数； 表示卷积操作；是通道重排列；学习到的时空表示。

图神经网络在交通预测中被广泛应用于捕捉空间依赖性。构建交通图的常见方法是通过相似度或距离度量预定义一个连接图，以这种方式生成的图通常是时间不变的，并应用于所有时间戳，如图3所示。为了更好地捕捉交通动态，近期研究提出在每个时间戳构建动态图，如图 3所示。然而，这种方式通常忽略了不同时间间隔之间的相互作用，从而限制了模型的预测性能。

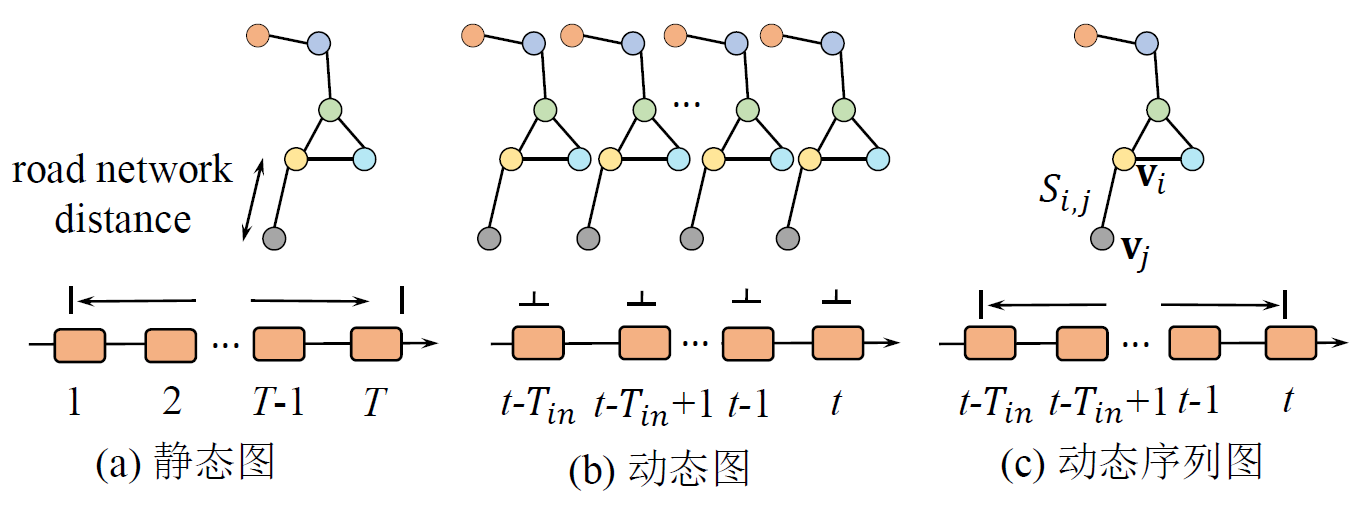


图3、不同图的构建方式

针对上述问题，本研究提出动态序列图（Dynamic Sequence Graph，DSG），旨在建模节点在整个输入序列中的复杂关系，动态捕捉空间和时间依赖性。DSG通过计算节点间的成对边权重，并设置阈值来限制图的密度，以确保边权重能够准确反映节点在空间和时间维度上的动态关系。DSG的构建在数据准备阶段完成，不会在训练过程中引入额外的计算开销。

基于DSG，本文为自适应时空上下文生成相应的邻接矩阵 并采用图卷积网络从图中提取特征。GNN通过聚合和转化邻域信息来平滑节点信息，给定跳数，第层的计算如下：



其中，是经过图学习得到的特征，  是可学习的参数， 表示转移矩阵的幂级数，通过 来生成。

ASTCL的最后阶段是将来自动态序列图建模、联合时空学习和残差连接的特征结合起来，其公式如下：



其中，是第 层的输出，综合了空间、时间和时空维度的特征信息。在通过堆叠的 个学习阶段后，可以得到高层时空特征 。为了生成多步交通预测，本研究应用一层MLP将时空特征映射为预测结果：



其中，表示未来步预测结果。该模型的训练目标是最小化预测结果和真实交通信息 之间的平均绝对误差，其损失函数定义为：



其中， 是ASTCL中的可学习参数。

**实验与分析：**

**实验设置**

本研究利用 PyTorch 框架实现了所提出的模型，并在 PEMS04 和 PEMS08 数据 集来进行交通流量预测任务的评估，METR-LA 和 PEMS-BAY 数据来进行交通 速度预测任务的评估。为了模拟快速适应的交通 场景，本文将每个数据集中的第一天的交通数据作为所有基准模型的训练集 。具体而言， PEMS04 和 PEMS08 数据集的最后 40% 数据按 1:1 的比例划分为验证集和测试集；而 METR-LA 和 PEMS-BAY 数据集的最后 30% 数据按 1:2 的比例划分。本文利用过去一小时的 交通数据（Tin = 12）预测未来三个时段的交通信息，分别为：15 分钟（Tout = 3）， 30 分钟（Tout = 6）， 和 60 分钟（Tout = 12）。通过参数网格搜索，构建自适应 上下文时使用的节点数为 15，模型的层数和特征维度在最佳表现的设置分别为 2 和 32。

**方法对比**

表1和2分别展示了 ASTCL 和其他基线模型在交通流量预测和交通速度预测上的表现。可以看出，ASTCL 在四个数据集上均优于其他先进的交通预 测模型，在所有评估指标上均取得了最佳性能。ASTCL 在单日训练数据情况下，能够快速适应不同交通的状况，在数据稀疏情况下表现出优秀的预测性能，展现出较强的鲁棒性和泛化能力。特别是在 PEMS04 和 PEMS08 这两个误差较大的数据集上，ASTCL 在大多数情况下实现了超过 30% 的性能提升。对于 METR-LA 和 PEMS-BAY 数据集，ASTCL 同样表现最佳，与次佳的模型相比，性能提升最高达 11.7%。从预测准确度上看， ASTCL 在捕捉和预测复杂交通模式方面的能力较为突出，同时保证了预测结果的鲁棒性和稳定性。

其次，在需要模型快速适应的单日交通预测任务当中，传统统计方法和深 度学习模型均面临挑战。统计方法如 HA 和 VAR 虽能利用单日交通数据做出预 测，但对异常值高度敏感，且难以适应交通模式的动态性和复杂性。另一方面， 数据驱动的模型在需要快速适应的单日交通预测中可能会因为过拟合问题而导致性能下降。如表 4-1 所示，AGCRN 通过可学习的节点嵌入来生成自适应图， 但在仅有一天数据的情况下，难以学习到有效的交通节点表征，而 GWNET 和STAEformer 也面临类似的问题。值得注意的是，在交通速度预测任务中

表1、交通流量预测实验结果

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 数据集 | PEMS04 | | | PEMS08 | | |
| 方法 | 15 分钟 | 30 分钟 | 60 分钟 | 15 分钟 | 30 分钟 | 60 分钟 |
|  | MAE RMSE MAPE | MAE RMSE MAPE | MAE RMSE MAPE | MAE RMSE MAPE | MAE RMSE MAPE | MAE RMSE MAPE |
| HA | 121.31 154.05 102.74 | 121.31 154.05 102.74 | 121.31 154.05 102.74 | 119.73 149.72 155.53 | 119.73 149.72 155.53 | 119.73 149.72 155.53 |
| VAR | 46.72 62.01 23.18 | 80.62 108.92 44.47 | 51.42 65.35 36.71 | 44.63 62.02 14.99 | 71.83 91.71 26.77 | 108.34 144.77 38.83 |
| ASTGCN | 41.08 58.85 31.78 | 42.34 60.67 32.00 | 46.90 67.30 38.42 | 23.40 35.27 17.95 | 25.08 38.05 18.94 | 31.16 45.91 22.88 |
| AGCRN | 160.25 213.83 100.67 | 160.74 214.41 100.22 | 160.87 214.58 100.12 | 160.33 207.24 108.17 | 159.48 206.33 108.76 | 159.24 206.13 108.83 |
| STSGCN | 41.96 64.12 30.23 | 44.04 66.43 31.65 | 50.22 73.91 34.93 | 35.13 53.83 24.60 | 36.56 55.74 25.23 | 42.05 62.93 27.57 |
| Bi-STAT | 33.28 51.20 21.19 | 37.30 58.01 23.63 | 47.15 72.65 30.25 | 22.86 33.10 16.34 | 24.76 36.45 17.92 | 32.13 46.79 23.09 |
| STID | 83.29 103.01 62.34 | 97.49 123.25 67.01 | 122.46 156.90 76.15 | 61.91 80.01 39.41 | 64.04 84.99 38.04 | 79.10 106.81 47.71 |
| FourierGNN | 29.37 44.12 24.50 | 37.65 55.26 31.85 | 48.47 70.89 39.57 | 24.01 35.60 16.45 | 30.27 44.82 20.29 | 43.56 62.56 26.83 |
| PDFormer | 32.00 44.65 33.26 | 79.41 115.50 121.85 | 96.94 131.52 155.95 | 65.35 91.46 108.86 | 93.03 121.78 176.17 | 101.42 129.15 197.72 |
| STAEformer | 82.15 107.78 58.66 | 92.23 119.36 71.45 | 96.54 125.44 74.51 | 38.40 51.44 30.76 | 44.24 58.84 35.11 | 52.38 69.45 35.88 |
| **ASTCL20** | 22.20 34.88 14.54 | 26.25 40.96 16.77 | 34.41 52.79 22.09 | 18.01 27.60 11.53 | 20.86 32.54 13.08 | 28.54 43.43 17.59 |
| **ASTCL50 ASTCL** | 21.67 34.19 14.20  **21.36 33.78 14.05** | 25.28 39.58 16.28  **24.90 39.01 16.17** | 33.55 51.41 21.99  **32.83 50.01 21.36** | 16.84 26.10 10.80  **16.77 25.95 11.15** | 19.74 30.87 12.61  **19.55 30.43 12.76** | 26.44 40.78 16.70  25.77 39.46 16.86 |

，部分深度学习模型的表现甚至不如统计方法 VAR，这进一步突显了深度学习模型在单日交通预测任务上的局限性。相比之下，ASTCL 通过引入自适应时空上下文学习而丰富了时空语义，在预测性能和鲁棒性上取得了显著提升。

为了进一步分析 ASTCL 在复杂交通场景下的预测能力，本文在交通图传感 器不足的情景下对模型性能进行了评估。这种情况一般发生在智慧城市早期发展 阶段，由于交通传感器部署不足，许多城市地区会面临数据稀缺的问题。少样本 交通预测可以有效评估模型在有限数据条件下的适应能力。为了模拟传感器稀 疏的交通网络，本方法将在四个数据集中随机移除大部分的交通节点，模型需要 在仅保留一小部分交通传感器数据的条件下进行训练，并预测整个网络的交通 模式。为确保公平性，实验中的验证集和测试集以及其他设置与基准模型保持一 致，这意味着模型不仅在未来的时间序列上进行评估，还需要在未见过的传感器上进行预测。这种实验设计显著增加了预测难度，并且在交通预测领域尚未得到 充分研究。在表 4-1和4-2 中，分别用 ASTCL50和 ASTCL20表示仅保留 50%和20%可用交通传感器训练的模型。例如，在PEMS08数据集上，ASTCL20 的训练集仅包含来自34个（占PEMS08的20%）交通传感器的单日交通数据。在实验 结果表明，在少样本设置下，ASTCL20和ASTCL20相比基准模型仍表现出优越 的预测性

表2、交通流量预测实验结果

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 数据集 | METR-LA | | | | | | | | | PEMS-BAY | | | | | | | | |
| 方法 | 15 分钟 | | | 30 分钟 | | | 60 分钟 | | | 15 分钟 | | | 30 分钟 | | | 60 分钟 | | |
|  | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE |
| HA | 11.30 | 15.86 | 30.99 | 11.30 | 15.86 | 30.99 | 11.30 | 15.86 | 30.99 | 6.42 | 9.12 | 14.71 | 6.42 | 9.12 | 14.71 | 6.42 | 9.12 | 14.71 |
| VAR | 7.74 | 12.50 | 23.02 | 8.10 | 11.25 | 19.15 | 7.15 | 10.79 | 18.18 | 4.24 | 6.86 | 7.55 | 4.33 | 6.39 | 7.39 | 5.32 | 8.10 | 8.61 |
| STGCN | 8.75 | 16.13 | 8.84 | 10.24 | 18.42 | 9.89 | 12.21 | 21.08 | 13.78 | 4.64 | 9.67 | 6.29 | 4.84 | 9.63 | 6.52 | 5.21 | 10.38 | 6.95 |
| DCRNN | 3.71 | 7.92 | 9.18 | 4.78 | 10.17 | 11.28 | 6.48 | 13.01 | 17.38 | 2.22 | 5.08 | 6.57 | 2.64 | 5.81 | 7.44 | 3.29 | 7.01 | 8.97 |
| GWNET | 3.24 | 6.09 | 8.44 | 3.83 | 7.58 | 11.01 | 4.90 | 9.51 | 15.49 | 1.76 | 3.73 | 4.52 | 2.24 | 4.92 | 5.71 | 2.97 | 6.78 | 7.53 |
| GMAN | 10.57 | 16.79 | 21.06 | 11.01 | 17.52 | 21.91 | 11.73 | 18.61 | 23.44 | 3.82 | 6.48 | 8.98 | 4.05 | 6.92 | 9.50 | 4.49 | 7.77 | 10.54 |
| MTGNN | 11.54 | 14.61 | 32.38 | 10.96 | 13.80 | 30.74 | 11.22 | 13.98 | 30.51 | 5.17 | 10.14 | 14.32 | 5.26 | 10.34 | 14.59 | 5.53 | 10.78 | 31.09 |
| STID | 5.48 | 10.07 | 19.08 | 5.84 | 10.51 | 20.30 | 6.26 | 11.38 | 22.21 | 2.94 | 6.05 | 8.18 | 4.07 | 9.03 | 11.96 | 4.80 | 10.28 | 13.84 |
| FourierGNN | 4.17 | 7.11 | 10.57 | 4.81 | 8.92 | 12.77 | 5.35 | 10.01 | 14.13 | 2.46 | 5.42 | 5.47 | 3.55 | 6.50 | 7.61 | 6.45 | 9.40 | 13.20 |
| STAEformer | 5.77 | 9.68 | 14.55 | 6.16 | 10.59 | 16.26 | 6.89 | 12.13 | 19.20 | 3.68 | 6.61 | 8.71 | 3.93 | 7.12 | 9.35 | 4.39 | 8.06 | 10.32 |
| **ASTCL20** | 2.90 | 5.84 | 7.67 | 3.52 | 7.26 | 9.84 | 4.65 | 9.36 | 13.60 | 1.62 | 3.57 | 3.42 | 2.16 | 4.98 | 4.85 | 2.99 | 6.93 | 7.03 |
| **ASTCL50** | 2.89 | 5.74 | 7.72 | 3.50 | 7.18 | 9.76 | 4.61 | 9.21 | 13.69 | 1.57 | 3.44 | 3.37 | 2.10 | 4.87 | 4.65 | 2.94 | 6.84 | 6.77 |
| **ASTCL** | **2.81** | **5.71** | **7.50** | **3.43** | **7.16** | **9.65** | **4.46** | **9.10** | **13.17** | **1.56** | **3.42** | **3.33** | **2.07** | **4.83** | **4.62** | **2.89** | **6.71** | **6.74** |

能。而相比于 ASTCL，ASTCL50和ASTCL20的平均性能仅分别下降了 2% 和 5%，这一结果进一步证明了 ASTCL 模型在少样本交通预测中的适应性和鲁棒性。

**消融实验**

为了验证 ASTCL 中各个组件的有效性，本文设计了如下消融实验：w/o STA： 去除联合时空注意力机制；w/o STC：去除联合时空卷积；w/oDSG：去除与动态序 列图相关的模块； w/o IN：去除实例归一化。表3展示了在 PEMS04 和 PEMS08 数据集上预测时长 60 分钟的实验结果。结果表明所有变体的性能均低于完整的 ASTCL 模型，证明了每个组件对预测重要性。其中，去除联合时空注意力机制 会导致模型性能显著下降， 这说明长距离的时空依赖对预测起关键性作用。本文 还探讨了联合注意力和联合卷积种各个结构的作用，联合时空学习模块利用三 种操作来从不同角度建模交通网络：空间建模（S）、时间建模（T）和时空建模（ST）。如图 3所示，整体的联合注意力和联合卷积模块在所有变体中表现出了最佳性能。通过结合空间模块和时间模块， 模型能够比单独使用任一模块获得更 好的预测精度。综上所述， 本方法所提出的联合时空学习能够同时建模复杂的时 空相关性，并有效提升了预测准确率。

表3、消融实验

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 方法 | PEMS04 | | | PEMS08 | | |
|  | MAE | RMSE | MAPE | MAE | RMSE | MAPE |
| w/o STA | 35.54 | 53.57 | 23.50 | 32.88 | 47.45 | 24.11 |
| w/o STC | 33.94 | 51.53 | 22.15 | 28.60 | 42.40 | 21.77 |
| w/o DSG | 33.71 | 51.31 | 21.72 | 26.54 | 39.86 | 17.27 |
| w/o IN | 33.92 | 51.89 | 22.30 | 27.54 | 40.94 | 21.20 |
| ASTCL | **32.83** | **50.01** | **21.36** | **25.77** | **39.46** | **16.86** |

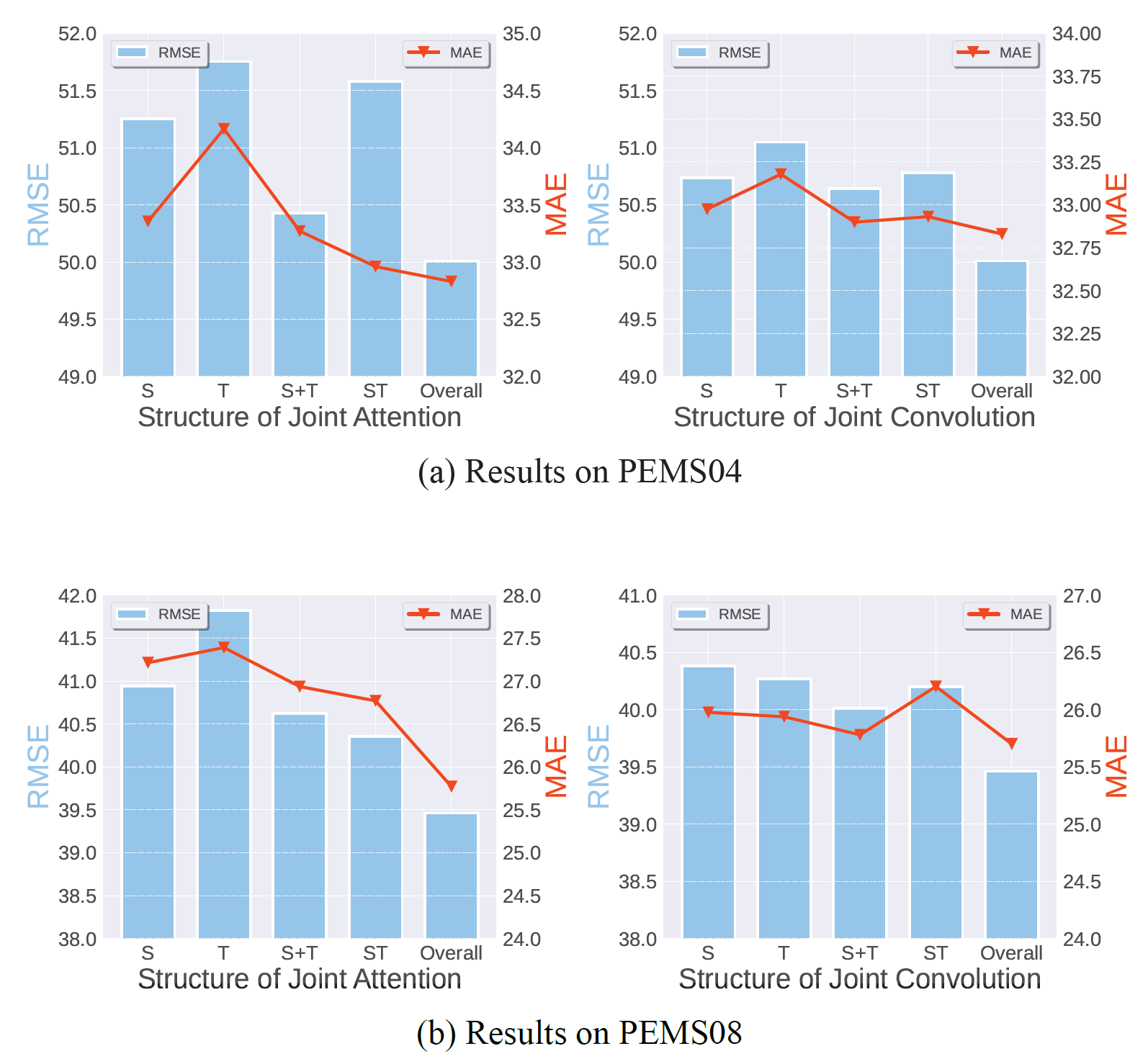


图3、时空联合学习分析

**参数分析**

关于参数分析中模型层数和特征维度的实验结果如图 4所示。在 PEMS04 和 METR-LA 数据集上，采用2 层架构能够在保证高性能的同时，有效缓解过拟 合并降低了计算成本。相比之下，对于 PEMS08 和 PEMS-BAY 数据集，较深的 5 层模型可以进一步降低 RMSE，表明这些数据集可能需要更复杂的模型结构来 捕捉其内在特征。此外，关于特征维度的选择，较大的维度通常能够增强模型 的表示能力，但同时会增加参数量，导致更高的计算成本和潜在的过拟合风险。 基于实验结果，本文在 METR-LA 数据集中选择特征维度为 16，而在 PEMS04、 PEMS08 和 PEMS-BAY 数据集中选择 32。这种差异化的参数配置不仅反映了不同数据集的特性，也在性能与计算效率之间实现了平衡。

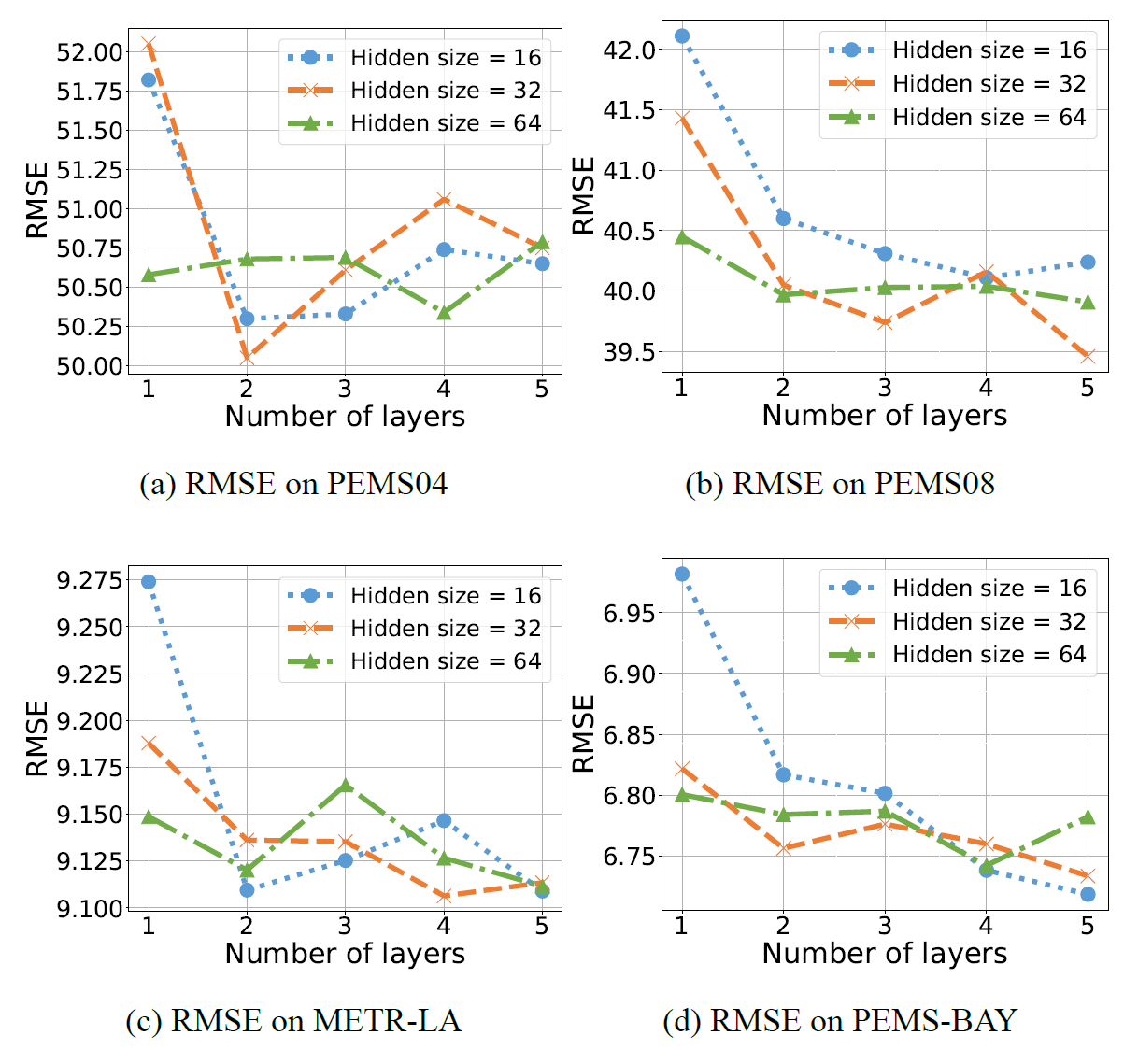


图4、参数分析

**鲁棒性分析**

本节从跨城市和跨域预测两个实验角度来分析 ASTCL 在复杂交通预测场景 中的鲁棒性。

表 4中展示了 ASTCL 在跨城市交通预测任务中的实验结果，从跨城市迁 移的角度来论证 ASTCL 的鲁棒性。迁移实验的设计包括包括在源数据集上预训 练模型，并将其直接迁移到目标数据集进行预测。ASTCL 与两种先进的迁移基 模型 TransGTR 和 pFedCTP 进行了对比。根据其原始训练策略， 首先在完整的源 数据集上训练 TransGTR 和 pFedCTP，然后使用目标数据集中的一天数据进行微 调。相比之下， ASTCL 仅使用源数据集中的一天数据进行预训练，并直接在目标 数据集上进行评估，无需微调或额外操作。这种训练模型的差异凸显了 ASTCL 的高效性，因为它避免了大量的预训练或微调过程。实验结果表明，ASTCL 在 跨城市迁移预测任务优于 TransGTR 和 pFedCTP，并实现了最多 10% 的平均性能 提升, 这一结果充分证明了 ASTCL 在预测交通流量和交通速度方面的可迁移性 和鲁棒性。特别值得注意的是，ASTCL 可以在任何交通数据集上进行训练，并直接应用于其他城市的类似预测任务，无需额外的训练过程，从而显著降低了计 算成本。

表4、跨城市预测实验结果

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 源数据 | 目标数据 | 方法 | 15 分钟 30 分钟 60 分钟 | | | | | | | | |
| MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE |
| METR-LA | PEMS-BAY | TransGTR  pFedGTP  ASTCL | 2.25 1.78  **1.52** | 5.44 3.45  **3.31** | 6.87 3.89  **3.25** | 2.81 2.32  **2.05** | 6.50 4.77  **4.68** | 8.35 5.23  **4.59** | 3.46 3.21  **2.86** | 7.71 6.58  **6.51** | 10.12 7.38  **6.74** |
| PEMS-BAY | METR-LA | TransGTR  pFedGTP  ASTCL | 3.52 3.15  **2.88** | 7.34 5.78  **5.75** | 9.69 8.42  **7.62** | 4.35 3.93  **3.51** | 8.96 7.23  **7.22** | 12.65 11.23  **9.60** | 5.21 5.06  **4.62** | 10.10 9.41  **9.34** | 15.95 15.30  **13.16** |
| PEMS04 | PEMS08 | TransGTR  pFedGTP  ASTCL | 18.03 18.48  **17.15** | 26.83 27.81  **26.66** | \ 12.02  **11.97** | 20.77 21.19  **20.11** | 33.96 33.40  **33.28** | \ 14.29  **14.10** | 27.68 27.59  **27.50** | 41.93 42.29  **41.74** | \ 22.46  **22.30** |
| PEMS08 | PEMS04 | TransGTR  pFedGTP  ASTCL | 21.69 22.30  **21.67** | 34.51 35.03  **34.25** | 14.40  \  **13.77** | 25.74 26.25  **25.30** | 40.38 40.76  **39.56** | 17.50  \  **17.12** | **32.94**  34.57 32.98 | 50.49 51.97  **50.24** | 22.82  \  **22.44** |

表 5、跨域预测实验结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 源数据 | 目标数据 | 方法 | MAE | RMSE | MAPE |
| METR-LA | PEMS04 | ASTCL  pFedGTP FT-ASTCL | 47.92 35.22  **33.05** | 66.48 52.39  **51.69** | 40.24  \  **21.17** |
| PEMS08 | ASTCL  pFedGTP FT-ASTCL | 41.34 28.00  **26.66** | 57.51 41.45  **41.01** | 32.03 23.28  **16.43** |
| PEMS-BAY | PEMS04 | ASTCL  pFedGTP FT-ASTCL | 45.42 35.23  **32.87** | 62.99 52.12  **51.92** | 37.67  \  **21.24** |
| PEMS08 | ASTCL  pFedGTP FT-ASTCL | 38.31 27.68  **26.41** | 53.17 41.52  **40.93** | 30.93 23.58  **16.65** |

第二类实验旨在从跨域预测的角度来评估 ASTCL 在不同预测领域中的鲁棒 性和泛化能力。对于跨城市预测而言，其主要考察模型在不同城市间迁移的能力，仍针对相同的预测变量；而跨领域评估则侧重于模型适应不同类型预测变量的能力。具体而言，本方法将在一种交通预测任务（例如，交通速度预测）上训练的模型迁移到另一种预测任务中（例如，交通流量预测）。由于不同预测变量之间固有的异质性，直接将预训练模型迁移到新的预测任务上通常会导致性能下降。为此，本文引入了 FT-ASTCL 变体，通过在目标变量数据集上进行额外的 20 轮微调，以提升预训练模型的性能。实验结果如表 5所示，通过简单的微调，FT-ASTCL 的性能提升至与在目标数据集上全量训练模型相当的水平。同时，ASTCL 在跨域场景中的预测表现要优于 pFedCTP，证明了 ASTCL 在不同交 通预测任务中的广泛适用性和良好表现，为其在实际应用中的灵活部署提供了有力支持。通过简单的微调，ASTCL 能够有效克服跨领域迁移中的性能下降问题，为多任务交通预测提供了可能的解决方案。

**本章小结**

本研究中针对交通预测的外部适应性提出 ASTCL 模型，旨在用单日的的交 通数据学习到高效的时空表征以适应不同的交通预测环境。具体而言，ASTCL 通过构建目标节点的自适应时空上下文，并结合联合注意力机制与联合卷积操 作，有效捕捉了交通数据中复杂的时空相关性。此外， 为了建模交通模式的动态 性，本研究还提出了一种动态序列图结构，进一步增强了模型对交通流变化的表 征能力。在四个真实交通数据集上的综合实验表明，ASTCL 在预测性能上要优 于现有方法，并在不同数据集上均表现出较强的鲁棒性和适应性。同时， ASTCL 不仅能够直接应用于多种交通数据集，还可以灵活预测多种交通属性，展现了较 为广泛的适用性。

**参考文献：**

[1] Programme U N U N H S. World cities report 2022: Envisaging the future of cities[M]. United Nations Research Institute for Social Development, 2022.

[2]Moorthy C, Ratcliffe B. Short term traffic forecasting using time series methods [J]. Transportation Planning and Technology, 1988, 12(1): 45-56.

[3]Moreira-Matias L, Gama J, Ferreira M, et al. Predicting taxi–passenger demand using streaming data[J]. IEEE Transactions on Intelligent Transportation Sys- tems, 2013, 14(3): 1393-1402.

[4]Shekhar S, Williams B M. Adaptive seasonal time series models for forecasting short-term traffic flow[J]. Transportation Research Record, 2007, 2024(1): 116- 125.

[5]Sun S, Zhang C, Yu G. A bayesian network approach to traffic flow forecasting [J]. IEEE Transactions on Intelligent Transportation Systems, 2006, 7(1): 124-

132.

[6]Zhang Y, Liu Y. Traffic forecasting using least squares support vector machines [J]. Transportmetrica, 2009, 5(3): 193-213.

[7] Xia Y, Chen J. Traffic flow forecasting method based on gradient boosting de- cision tree[C]//International Conference on Frontiers of Manufacturing Science and Measuring Technology. 2017: 413-416.

[8]Zhang J, Zheng Y, Qi D. Deep spatio-temporal residual networks for citywide crowd flows prediction[C]//AAAI Conference on Artificial Intelligence. 2017:

1655-1661.

[9] Yao H, Wu F, Ke J, et al. Deep multi-view spatial-temporal network for taxi demand prediction[C]//AAAI Conference on Artificial Intelligence. 2018: 2588- 2595.

[10] Pan Z, Wang Z, Wang W, et al. Matrix factorization for spatio-temporal neural networks with applications to urban flow prediction[C]//International Conference on Information and Knowledge Management. 2019: 2683-2691.

[30]Yu B, Yin H, Zhu Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting[C]//International Joint Conference on Artificial Intelligence. 2018: 3634-3640.

[31]LiM, Zhu Z. Spatial-temporal fusion graph neural networks for traffic flow fore- casting[C]//AAAI Conference on Artificial Intelligence. 2021: 4189-4196.

[32]Song C, Lin Y, Guo S, et al. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting[C]// AAAI Conference on Artificial Intelligence. 2020: 914-921.

[33]Wu Z, Pan S, Long G, et al. Graph wavenet for deep spatial-temporal graph mod- eling[C]//International Joint Conference on Artificial Intelligence. 2019: 1907-1913.

[34] Wu Z, Pan S, Long G, et al. Connecting the dots: Multivariate time series fore- casting with graph neural networks[C]//SIGKDD International Conference on Knowledge Discovery and Data Mining. 2020: 753-763.

[35] Bai L, Yao L, Li C, et al. Adaptive graph convolutional recurrent network for traffic forecasting[J]. International Conference on Neural Information Processing Systems, 2020, 33: 17804-17815.

[36] Guo S, Lin Y, Wan H, et al. Learning dynamics and heterogeneity of spatial- temporal graph data for traffic forecasting[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 34(11): 5415-5428.

[37]Lan S, Ma Y, Huang W, et al. Dstagnn: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting[C]//International Conference on Ma- chine Learning. 2022: 11906-11917.

[38]Duan W, He X, Zhou Z, et al. Localised adaptive spatial-temporal graph neu- ral network[C]//SIGKDD International Conference on Knowledge Discovery and Data Mining. 2023: 448-458.

[39]Liu M, Gao H, Ji S. Towards deeper graph neural networks[C]//SIGKDD Inter- national Conference on Knowledge Discovery and Data Mining. 2020: 338-348.

[40]Xhonneux LP, Qu M, Tang J.Continuous graph neural networks[C]// International Conference on Machine Learning. 2020: 10432-10441.

[41] Zheng C, Fan X, Wang C, et al. GMAN: A graph multi-attention network for

traffic prediction[C]//AAAI Conference on Artificial Intelligence. 2020: 1234- 1241.

[42]Guo S, Lin Y, Feng N, et al. Attention based spatial-temporal graph convolu- tional networks for traffic flow forecasting[C]//AAAI Conference on Artificial Intelligence. 2019: 922-929.

[43]Chen C, Liu Y, Chen L, et al. Bidirectional spatial-temporal adaptive transformer for urban traffic flow forecasting[J]. IEEE Transactions on Neural Networks and Learning System, 2023, 34(10): 6913-6925.

[44]杜圣东, 李天瑞, 杨燕, 等. 一种基于序列到序列时空注意力学习的交通流预 测模型[J]. 计算机研究与发展, 2020, 57(8): 1715-1728.

[45]魏荣. 基于注意力机制的短时交通流预测模型研究[J]. 交通科技与管理, 2024, 5(20): 42-44.

[46]李慧慧, 曹倩霞, 单振宇, 等. 基于注意力时空图神经网络的多任务交通流预 测[J]. 交通科学与工程: 1-12.

[47] Liu X, Liang Y, Huang C, et al. Do we really need graph neural networks for traffic forecasting?[J]. CoRR, 2023, abs/2301.12603.

[48]Wang Z, Nie Y, Sun P, et al. ST-MLP: A cascaded spatio-temporal linear frame- work with channel-independence strategy for traffic forecasting[J]. CoRR, 2023, abs/2308.07496.

[49]Zhang Z, Huang Z, Hu Z, et al. Mlpst: Mlp is all you need for spatio-temporal prediction[C]//International Conference on Information and Knowledge Manage-ment. 2023: 3381-3390.

[50]Pan Z, Ke S, Yang X, et al. Autostg: Neural architecture search for predictions of spatio-temporal graph[C]//International Conference of World Wide Web. 2021: 1846-1855.

[51]Ke S, Pan Z, He T, et al. Autostg+: An automatic framework to discover the optimal network for spatio-temporal graph prediction[J]. Artificial Intelligence, 2023, 318: 103899.

[52]Wu X, Zhang D, Guo C, et al. Autocts: Automated correlated time series fore- casting[J]. Proceedings of the VLDB Endowment, 2021, 15(4): 971-983.

[53]Wu X, Zhang D, Zhang M, et al. Autocts+: Joint neural architecture and hyperpa- rameter search for correlated time series forecasting[J]. Proceedings of the ACM on Management of Data, 2023, 1(1): 1-26.

[54] Lin B Y, Xu F F, Liao E Q, et al. Transfer learning for traffic speed prediction: Apreliminary study[C]//AAAI Conference on Artificial Intelligence Workshops:

WS-18. 2018: 174-177.

[55]LuB, Gan X, Zhang W, et al. Spatio-temporal graph few-shot learning with cross- city knowledge transfer[C]//SIGKDD International Conference on Knowledge Discovery and Data Mining. 2022: 1162-1172.

[56] Zhang C, Zhang H, Qiao J, et al. Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data[J]. IEEE Journal on Selected Areas in Communications, 2019, 37(6): 1389-1401.

[57]Yao H, Liu Y, Wei Y, et al. Learning from multiple cities: A meta-learning approach for spatial-temporal prediction[C]//International Conference of World Wide Web. 2019: 2181-2191.

[58]Zhang Y, Lu H, Liu N, et al. Personalized federated learning for cross-city traffic prediction[C]//International Joint Conference on Artificial Intelligence. 2024:

5526-5534.

[59]Jin Y, Chen K, Yang Q. Transferable graph structure learning for graph-based traffic forecasting across cities[C]//SIGKDD International Conference on Knowl- edge Discovery and Data Mining. 2023: 1032-1043.

[60]Fang Z, Wu D, Pan L, et al. When transfer learning meets cross-city urban flowprediction: Spatio-temporal adaptation matters[C]//International Joint Confer- ence on Artificial Intelligence. 2022: 2030-2036.

[61]Wei Y, Zheng Y, Yang Q. Transfer knowledge between cities[C]//SIGKDD In-

ternational Conference on Knowledge Discovery and Data Mining. 2016: 1905- 1914.

[62]Wang L, Geng X, Ma X, et al. Cross-city transfer learning for deep spatio- temporal prediction[C]//International Joint Conference on Artificial Intelligence.2019: 1893-1899.