

# 面向时空交通栅格流量预测的3D通道注意力网络\*

童凯南<sup>1</sup>,林友芳<sup>1</sup>,刘军<sup>2</sup>,郭晟楠<sup>1</sup>,万怀宇<sup>1</sup>

(1. 北京交通大学计算机与信息技术学院,北京 100044; 2. 中国民航信息网络股份有限公司,北京 101318)

**摘要:**城市交通流量预测对交通管理和公共安全具有重要意义。然而,交通栅格流量数据中的规律在时刻变化,在城市中存在全局范围的时空关系,并且不同特征通道在每个城市区域上有不同的重要性。为解决这些挑战并做出更准确的预测,设计了一种新颖的时空神经网络模型——3D通道注意力网络(three-dimensional channel-wise attention networks, 3D-CANet)。提出一个3D通道内注意力(three-dimensional inner channel attention, 3D-InnerCA)单元来动态捕获各个通道中不同的全局时空相关性,同时设计通道间注意力(inter channel attention, InterCA)单元来自适应地重校准每个区域上不同特征通道的贡献。在3个真实交通栅格流量数据集上的实验结果表明,3D-CANet模型的预测能力优于其他对比方法,证明了模型的有效性。

**关键词:**时空数据;交通栅格流量;3D通道注意力;通道内注意力;通道间注意力

**中图分类号:**TP391 **文献标志码:**A **开放科学(资源服务)标识码(OSID):**

**文章编号:**1001-2486(2022)03-041-09



听语音  
与作者互动  
聊科研

## 3D channel-wise attention network for spatio-temporal traffic raster flow prediction

TONG Kainan<sup>1</sup>, LIN Youfang<sup>1</sup>, LIU Jun<sup>2</sup>, GUO Shengnan<sup>1</sup>, WAN Huaiyu<sup>1</sup>

(1. School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China;

2. TravelSky Technology Limited, Beijing 101318, China)

**Abstract:** Urban traffic flow forecasting is of great significance for traffic management and public safety. However, the correlations of traffic raster flow change with time. There are global spatio-temporal correlations in the city, and the contributions of channel-wise features vary on each city region. To tackle these challenges and make more accurate prediction, a novel spatio-temporal neural network model, named 3D-CANet (three-dimensional channel-wise attention network), was designed. A 3D-InnerCA (three-dimensional inner-channel attention) unit was proposed to dynamically capture the global spatio-temporal correlations for different channel-wise features. Meanwhile, an InterCA (inter-channel attention) unit was designed to adaptively recalibrate the contributions of different channel-wise features on each region. The experimental results on three real-world traffic raster flow datasets demonstrate that the predictive performance of the 3D-CANet model was better than the others, which proved the validity of the model proposed.

**Keywords:** spatio-temporal data; traffic raster flow; 3D channel-wise attention; inner channel attention; inter channel attention

在现代智能交通系统<sup>[1]</sup> (intelligent transportation systems, ITS) 和高级旅行者信息系统 (advanced traveler information systems, ATIS) 中,交通流量预测任务能够为政府、出租车公司和游客提供准确可靠的信息,起到至关重要的作用<sup>[2-3]</sup>。例如,政府的交通管理部门可以通过预测交通流量高峰出现的时间和地点来提前采取相应的交通管理措施<sup>[4]</sup>。出租车公司可以通过预测未来的交通流量来优化车辆调度任务。此外,游客也

可以根据预测的交通状况制订合理的出行计划<sup>[5]</sup>。因此,准确预测交通流量对保障社会安全,促进经济发展和提升游客出行体验有重大意义。

现有交通数据预测模型大致可以分为两类:传统方法和深度学习方法。传统的统计或机器学习方法如 XGBoost<sup>[6]</sup> 等缺乏处理高维的空间数据能力,无法学习交通数据中复杂的非线性信息。不过随着深度学习的发展,使得这些问题可以逐渐被解决<sup>[7]</sup>。

\* 收稿日期:2021-06-16

基金项目:中国博士后科学基金资助项目(2021M700365)

作者简介:童凯南(1996—),男,北京人,硕士研究生,E-mail:kntong@bjtu.edu.cn;

万怀宇(通信作者),男,副教授,博士,博士生导师,E-mail:hywan@bjtu.edu.cn

递归神经网络<sup>[8]</sup> (recurrent neural network, RNN)能够建模数据中的时间信息。其变体长短期记忆神经网络<sup>[9]</sup> (long short-term memory, LSTM)和门控循环单元<sup>[10]</sup> (gated recurrent unit, GRU)进一步增强了这种能力。最近,文献[2, 11]使用基于 LSTM 的网络来预测短期交通速度和交通流量,基于长短期记忆网络的关联时间序列预测<sup>[12]</sup> (correlated time series oriented neural network based on LSTM, CTS-LSTM)模型考虑了多个序列之间的相关性。此外,空间相关性对交通数据预测至关重要,却很少有基于 LSTM 的方法能对其进行建模。

幸运的是,研究人员找到了使用卷积神经网络(convolutional neural networks, CNN)来捕获空间信息的解决方案,例如卷积长短期记忆网络<sup>[13]</sup> (convolutional LSTM, ConvLSTM)模型和深度时空残差网络<sup>[14]</sup> (spatio-temporal residual networks, ST-ResNet)模型。它们不仅能捕获空间相关性,还能捕获时间相关性。此外,基于注意力机制的时空图卷积<sup>[15]</sup> (attention based spatial-temporal graph convolutional networks, ASTGCN)模型和时空同步图卷积<sup>[16]</sup> (spatial-temporal synchronous graph convolutional networks, STSGCN)模型等基于图卷积(graph convolutional networks, GCN)的模型性能也很强大,不过并不适用于本研究聚焦的栅格数据预测问题。

近年来,ST-3DNet<sup>[17]</sup>和DeepSTNPlus<sup>[18]</sup>模型都在交通栅格流量预测问题上取得了进展。然而,与其他基于 CNN 的模型一样受到卷积核大小的限制,它们都无法有效地捕获远程空间相关性。如图 1 所示,在办公区上班的人们可能住在远近不同的住宅区 A 和 B,这些区域之间就会因为通

勤而存在很强的相关性,这就需要模型能够捕获大范围甚至全局的时空联系。一些研究试图通过堆叠多层 CNN 来解决此问题,但这样会使得模型训练变得非常困难。在文献[19-23]中,自注意力机制可以用来捕获全局范围的空间相关性,但并不能对不同通道捕获不同的相关性,于是本文提出 3D 通道内注意力单元(three-dimensional inner channel attention, 3D-InnerCA)来学习每个通道内的全局时空相关性。

此外,受 SENet<sup>[24]</sup>计算 CNN 特征通道之间相互关系的启发,ST-3DNet 设计了一个重新校准模块,以探索每个区域上各个通道的贡献,不过这些贡献是静态学习得到的,并不能很好地适应数据的动态性。假设有一个特征通道关注的是某些复杂的交通情景,比如立交桥。那么在有立交桥的区域,这个特征就比其他特征有更大贡献。相应的,这个特征在有立交桥的区域也比在其他区域更具重要性。需要强调的是,上述关系也是动态变化的。例如,人们可能在周末外出郊游,那么某些关注通勤状况的特征在周末就会变得不那么重要。而现有交通栅格流量预测模型无法应对上述情况。因此本文在此基础上设计 InterCA,从而动态计算各个特征通道在每个区域上的贡献。

### 1 问题定义

在本节中给出本文中的一些定义。

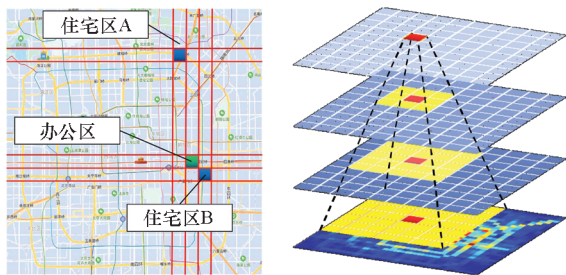
**定义 1** 交通栅格流量。通过将城市划分为  $I \times J$  的栅格图,每个栅格就代表城市的一个区域,  $C_m$  个通道代表  $C_m$  种类型的交通数据,例如流入流量、流出流量。此时的交通栅格流量的形式就像一张图片,可以得到第  $t$  时间段内的数据  $X_t \in \mathbf{R}^{C_m \times I \times J}$ 。

**定义 2** 交通栅格流量预测问题。在本研究中,旨在根据历史数据预测未来的交通流量。给定第  $t-T$  到第  $t$  时间段的历史交通栅格流量序列  $X = [X_{t-T}, X_{t-T+1}, \dots, X_t] \in \mathbf{R}^{C_m \times T \times I \times J}$ ,模型应该输出第  $t+1$  时间段的预测结果  $\hat{X}_{t+1}$ 。

### 2 模型结构

#### 2.1 总体架构

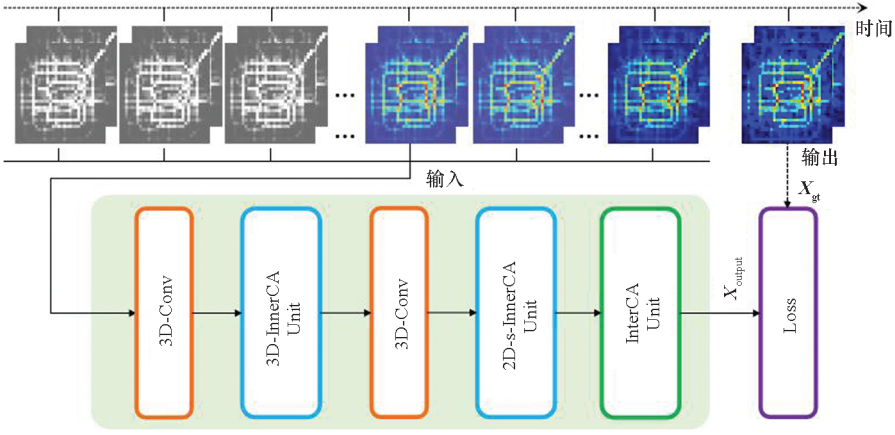
为更有效地捕获时空相关性,提出一种基于深度学习的端到端模型 3D-CANet。模型总体架构如图 2 所示。首先使用一层卷积核大小为  $1 \times 3 \times 3$  的 3D 卷积来捕获数据的局部特征。之后堆叠 3 个 3D-InnerCA 单元以捕获全局时空动态相



(b) 堆叠卷积层来捕获全局空间信息

(a) 城市栅格图  
(a) Citywide grid map (b) Stacking CNNs to explore global spatial correlation

图 1 CNN 在城市交通流量预测分析中的应用  
Fig. 1 Applications of CNNs for citywide traffic flow analytics



注:3D-Conv为3D卷积;3D-InnerCA Unit为3D通道内注意力单元;2D-s-InnerCA Unit为2D空间通道内注意力单元;InterCA Unit为通道间注意力单元。

图2 3D-CANet模型的整体架构

Fig. 2 Overall architecture of 3D-CANet

关性。接下来,使用一层有  $C$  个大小为  $3 \times 3 \times 3$  卷积核的3D卷积层,用于聚合时间维度上的信息。此时,时间信息已被充分捕获,为进一步探索全局空间动态相关性<sup>[17]</sup>,堆叠2个2D空间通道内注意力单元(2D spatial inner channel attention, 2D-s-InnerCA)。最后,使用InterCA单元来动态地计算每个区域上各特征通道的贡献。

## 2.2 3D-InnerCA

由于交通栅格流量具有动态的全局时空相关性,基于CNN的模型受卷积核大小的限制,并不能捕获大范围的信息。并且由于参数固定,模型不具有自适应输入数据的动态能力,因此它们不能充分学到数据中蕴含的信息。如图3所示,提出一个能动态探索各个特征通道内的全局时空相

关性的模块,称为3D-InnerCA,其中BN+ReLU表示批归一化和激活函数ReLU。文献[19-23]提出的模型大多使用自注意力机制来捕获栅格流量中的空间信息,可以动态计算每个区域和其他所有区域之间的相关性,但这种相关性对于所有通道是共享的。考虑到不同的特征通道中时空相关性可能有所区别,因此在3D-InnerCA单元中对通道进行分组,以学习分组后各个通道内的动态全局相关性。具体来说,3D-InnerCA单元由3D-s-InnerCA和3D时间通道内注意力单元(3D temporal inner channel attention, 3D-t-InnerCA)组成,分别用来动态捕获全局空间相关性和时间相关性。将两个部分的结果拼接成 $2C$ 个通道,并使用大小为 $3 \times 3 \times 3$ 的 $C$ 个卷积核进行3D卷积,从而把 $2C$ 个通道聚合为 $C$ 个通道。为保证训练效率,结构中使用残差学习<sup>[25]</sup>。

### 2.2.1 3D-s-InnerCA

3D-s-InnerCA的结构如图4所示。它用来捕获分组后各个通道内的动态全局空间相关性。具体来说,设第 $k$ 层3D-s-InnerCA的输入为 $\mathbf{X}_s^{(k-1)} \in \mathbf{R}^{C \times T \times I \times J}$ ,其中 $T$ 是时间段的数量,首先将形状转化为 $\mathbf{R}^{C \times T \times N}$ ,这里 $N = I \times J$ 。然后分别使用3个全连接层从 $N$ 维输入特征映射到 $N$ 维隐含特征,并将通道分为 $G$ 组,根据式(1)~(3),转化成 $\mathbf{Q}_s \in \mathbf{R}^{G \times T \times N \times (C/G)}$ , $\mathbf{K}_s \in \mathbf{R}^{G \times T \times N \times (C/G)}$ 和 $\mathbf{V}_s \in \mathbf{R}^{G \times T \times N \times (C/G)}$ 。接下来,根据式(4),计算注意力结果 $\mathbf{X}_{\text{spatial}} \in \mathbf{R}^{G \times T \times N \times (C/G)}$ ,并将其形状转为 $\mathbf{R}^{C \times T \times I \times J}$ ,即

$$\mathbf{Q}_s = \text{Group}(\mathbf{W}_{\text{qs}} \mathbf{X}_s^{(k-1)}) \quad (1)$$

$$\mathbf{K}_s = \text{Group}(\mathbf{W}_{\text{ks}} \mathbf{X}_s^{(k-1)}) \quad (2)$$

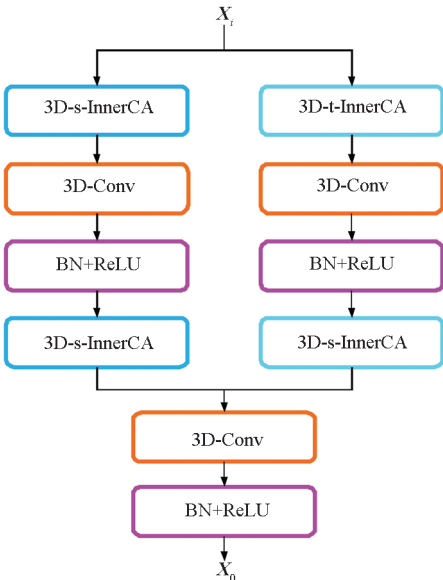


图3 3D通道内注意力单元

Fig. 3 3D-InnerCA unit

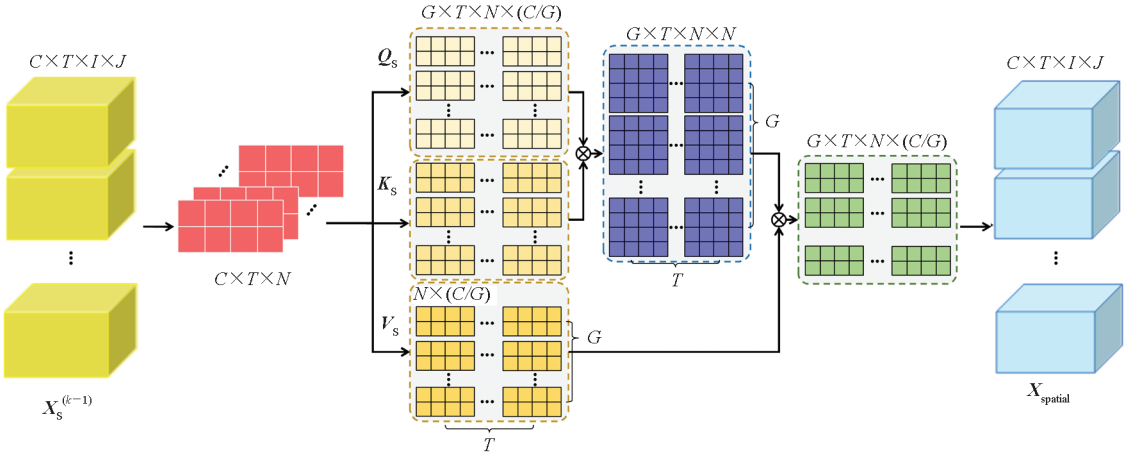


图 4 3D 空间通道内注意力

Fig. 4 3D-s-InnerCA

$$V_s = \text{Group}(W_{vs}X_s^{(k-1)}) \quad (3)$$

$$X_{\text{spatial}} = Q_s K_s^T V_s \quad (4)$$

其中,  $W_{qs}$ 、 $W_{ks}$  和  $W_{vs}$  是全连接层的参数。这里不像传统的注意力那样使用激活函数 Softmax, 因为它会限制模型的表达能力并增加训练成本<sup>[26]</sup>。需要注意的是, 2D-s-InnerCA 的结构与 3D-s-InnerCA 类似, 只是没有时间维度上的操作。

### 2.2.2 3D-t-InnerCA

为探究动态全局时间相关性, 设计如图 5 所示的 3D-t-InnerCA。与 3D-s-InnerCA 相似, 给定第 1 层的输入  $X_T^{(l-1)} \in \mathbf{R}^{C \times T \times I \times J}$ , 首先将其转为  $\mathbf{R}^{C \times T \times N}$ 。同样在经过 3 个全连接层之后, 分别将

其通道划分为  $G$  组。与 3D-s-InnerCA 不同的是, 根据式 (5) ~ (7), 将变量转换为  $Q_T \in \mathbf{R}^{G \times T \times (N \times C/G)}$ ,  $K_T \in \mathbf{R}^{G \times T \times (N \times C/G)}$  和  $V_T \in \mathbf{R}^{G \times T \times (N \times C/G)}$ , 并根据式 (8) 计算注意力结果  $X_{\text{temporal}} \in \mathbf{R}^{C \times T \times (N \times C/G)}$ , 从而捕获动态全局时间相关性。最后, 将结果转换为  $\mathbf{R}^{C \times T \times I \times J}$ , 即

$$Q_T = \text{Group}(W_{qt}X_T^{(k-1)}) \quad (5)$$

$$K_T = \text{Group}(W_{kt}X_T^{(k-1)}) \quad (6)$$

$$V_T = \text{Group}(W_{vt}X_T^{(k-1)}) \quad (7)$$

$$X_{\text{temporal}} = Q_T K_T^T V_T \quad (8)$$

其中,  $W_{qt}$ 、 $W_{kt}$  和  $W_{vt}$  是全连接层的参数。

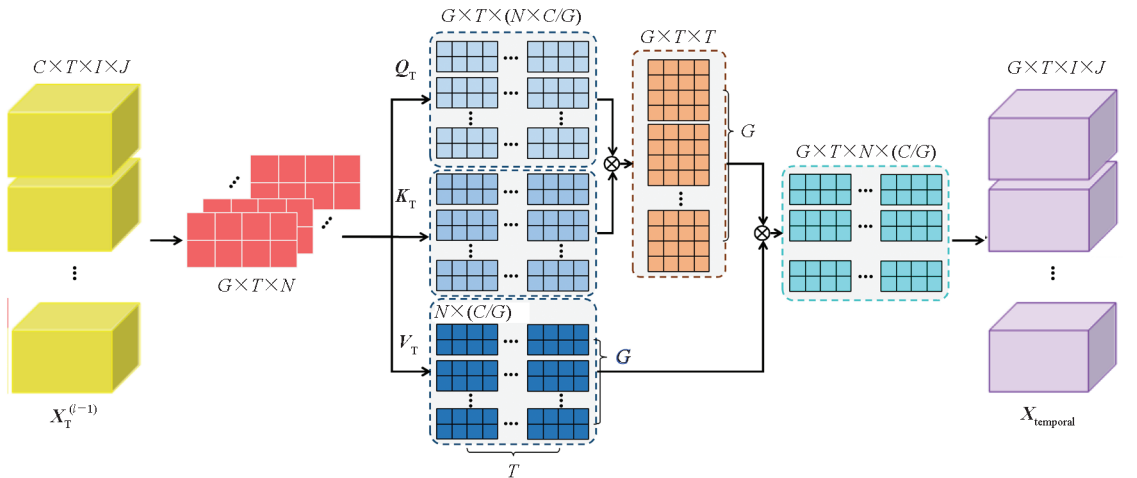


图 5 3D 时间通道内注意力

Fig. 5 3D-t-InnerCA

## 2.3 InterCA

由于交通流量中复杂的相关性, 不同特征通道对同一个区域的贡献不同, 而同一个特征对不同区域的重要性也不同。不仅如此, 特征通道的贡献还会随着时间而动态变化。因此, 提出通道

间注意力单元, 并使用 2D 卷积神经网络来动态学习每个区域上特征通道之间的关系。

将 InterCA 单元放置在 2D-s-InnerCA 单元之后, 以自适应地计算不同特征通道对每个区域的贡献。用  $X_{\text{inter}} \in \mathbf{R}^{C \times I \times J}$  来表示 InterCA 单元的输

入,并将输出表示为  $\mathbf{X}_{\text{out}} \in \mathbf{R}^{C_{\text{out}} \times I \times J}$ ,这里设置输出通道数  $C_{\text{out}} = C_{\text{in}}$ 。对于每个输出通道  $\mathbf{X}_{\text{out}}^{(p)} \in \mathbf{R}^{I \times J}$ ,首先计算所有输入通道的贡献  $\mathbf{W}_{\text{out}}^{(p)} \in \mathbf{R}^{C \times I \times J}$ ,也就是给  $C_{\text{out}}$ 个输出通道计算  $C_{\text{out}}$ 个重要性矩阵。每个重要性矩阵表示每个区域上所有输入通道对对应输出通道的重要性,即

$$\mathbf{W}_{\text{out}}^{(p)} = f(\mathbf{W}^{(p)} * \mathbf{X}_{\text{inter}} + b^{(p)}), p = 1, \dots, C_{\text{out}} \quad (9)$$

其中,“\*”表示使用  $C$ 个大小为  $3 \times 3$ 的卷积核进行卷积运算。如图6所示,输出通道  $\mathbf{X}_{\text{out}}^{(p)}$ 可表示为

$$\mathbf{X}_{\text{out}}^{(p)} = \sum_{i=1}^C \mathbf{W}_{\text{out}}^{(p),(i)} \circ \mathbf{X}_{\text{inter}}^{(i)} \quad (10)$$

其中,“ $\circ$ ”是逐元素乘法。将全部  $C_{\text{out}}$ 个通道拼在一起,则3D-CANet最终输出预测结果  $\mathbf{X}_{\text{out}} \in \mathbf{R}^{C_{\text{out}} \times I \times J}$ 。

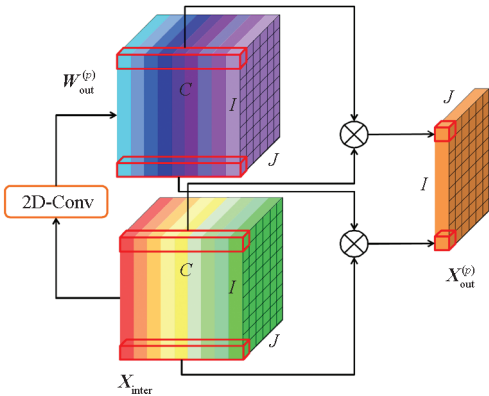


图6 通道间注意力单元

Fig. 6 Inter-channel attention unit

## 2.4 损失函数

模型通过最小化损失函数来进行训练,损失函数为预测值与真实值之间的均方误差,即

$$\text{Loss} = \|\mathbf{X}_{\text{out}} - \mathbf{X}_{t+1}\|_2^2 \quad (11)$$

## 3 实验

在3个不同的交通流量数据集上进行实验,与基准模型相比,实验结果证明了3D-CANet模

型的有效性。

### 3.1 数据集

本研究使用TaxiBJ、BikeNYC和TaxiNYC3个交通流量数据集,它们的统计信息如表1所示。

1) TaxiBJ: TaxiBJ数据集由轨迹数据和外部因素数据组成。其中轨迹数据是在四段时间内的北京出租车GPS数据,包括:2013年7月1日—2013年10月30日,2014年3月1日—2014年6月30日,2015年3月1日—2015年6月30日和2015年11月1日—2016年4月10日。每个时间段的长度设定为30 min。将轨迹数据转化为栅格数据,并选取后4周数据为测试集,其余数据为训练集。

2) BikeNYC: 像TaxiBJ一样,BikeNYC数据集由轨迹数据和外部因素数据组成。其轨迹数据来源于纽约自行车系统,时间跨度为2014年4月1日—2014年9月30日。每个时间段的长度设定为1 h。将轨迹数据转化为栅格数据,并选取最后10 d的数据为测试集,其余数据为训练集。

3) TaxiNYC: TaxiNYC数据集由22 349 490条纽约出租车行驶数据组成,时间跨度为2015年1月1日—2015年3月1日。每个时间段的长度设定为30 min。将轨迹数据转化为栅格数据,并选取最后20 d为测试集,其余数据为训练集。

### 3.2 基准模型

将3D-CANet与以下8个基准模型进行比较:

1) HA: 如同文献[14,17,27]的做法,将每个区域历史数据中对应时间段的均值作为该区域流入和流出流量的预测值。

2) XGBoost<sup>[6]</sup>: 一种提升树的可扩展机器学习系统,属于机器学习方法,能够根据各个区域的历史数据进行预测。

表1 交通流量数据集的统计信息

Tab. 1 Statistics of the traffic flow datasets

数据集	TaxiBJ	BikeNYC	TaxiNYC
城市	北京	纽约	纽约
时间跨度	2013.07.01—2013.10.30 2014.03.01—2014.06.30 2015.03.01—2015.06.30 2015.11.01—2016.04.10	2014.04.01—2014.09.30	2015.01.01—2015.03.01
时段长度	30 min	1 h	30 min
网格划分	(32, 32)	(16, 8)	(10, 20)
时段数量	22 459	4 392	2 880

3) LSTM<sup>[9]</sup>: 长期短期记忆网络是一种 RNN 网络,用于时间序列预测,能够根据各个区域的历史数据进行预测。

4) GRU<sup>[10]</sup>: 门控循环网络也是一种 RNN 网络,用于时间序列预测,能够根据各个区域的历史数据进行预测。

5) ConvLSTM<sup>[13]</sup>: ConvLSTM 结合 CNN 和 LSTM,可以像 CNN 一样捕获空间信息,也能像 LSTM 一样挖掘时间信息,且能同时对全部区域的历史数据进行时空信息提取并预测。

6) ST-ResNet<sup>[14]</sup>: 它是一种基于深度神经网络的时空数据预测模型。使用 3 个残差网络分别建模近邻性、周期性和趋势性的空间相关性,然后融合 3 个部分的结果最终输出。

7) ST-3DNet<sup>[17]</sup>: 基于深度神经网络的时空数据预测模型。它首先利用 3D 卷积在时间和空间维度上捕获信息。与 ST-ResNet 不同,它仅考虑近邻性和趋势性,然后进行融合输出。

8) DeepSTNPlus<sup>[18]</sup>: 基于深度神经网络的时空数据预测模型。它在整个模型最开始就进行近邻性、周期性和趋势性的融合,然后在模型末尾使用多尺度融合网络来融合多级别特征,展现出城市流量预测任务的最佳性能。

### 3.3 参数设置

本文基于 PyTorch<sup>[27]</sup> 实现 3D-CANet 模型。在本研究中,预测两种交通数据:流入和流出流量。因此, $C_m$  设定为 2。设置输入数据的长度为 6,也就是模型根据最近 6 个时间段内的数据以进行预测。使用最大值最小值归一化方法将输入数据缩放到  $[-1, 1]$  范围内。在 TaxiBJ、BikeNYC 和 TaxiNYC 上,分别选择最近 4 周、10 d 和 20 d 的数据作为测试集,其他数据作为训练集。在 3 个数据集上训练的批大小分别为 4、8、4,学习率为 0.01。计算预测值和真实值的均方根误差 (root mean square error, RMSE) 和绝对平均误差 (mean absolute error, MAE) 作为评价指标:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \|X_{out}^{(i)} - X_{t+1}^{(i)}\|^2} \quad (12)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |X_{out}^{(i)} - X_{t+1}^{(i)}| \quad (13)$$

其中, $N$  为所有预测值的数量。

在 TaxiBJ、BikeNYC 和 TaxiNYC 上,3D-CANet 的第一层 3D 卷积分别使用 16、64 和 64 个卷积

核。在探索不同数量并考虑训练开销的情况下,设定在模型中堆叠 3 个 3D-InnerCA 单元和 2 个 2D-s-InnerCA 单元,每个单元分别有 2 层注意力。其他的 3D 卷积都应用  $C = 64$  个大小为  $3 \times 3 \times 3$  的卷积核,2D 卷积都使用  $C = 64$  个大小为  $3 \times 3$  的卷积核。

### 3.4 实验结果比较分析

首先给出 3D-CANet 在 3 个数据集上与其他 8 个基准模型的比较,结果如表 2 所示。模型根据最近 6 个时间段来预测下一个时间段的数据,其中 ST-ResNet 和 DeepSTNPlus 分别另有 3 个周期性和趋势性数据,ST-3DNet 有 3 个趋势性数据。运行每个模型 5 次,并计算出评价指标的平均值和标准差。结果表明,本文提出的模型性能比其他基准模型更优。对于 ST-ResNet、ST-3DNet 和 DeepSTNPlus 模型,它们除了要有近邻性数据还需要获取周期性和趋势性信息;而 3D-CANet 模型仅需要近邻性数据就能获得更好的性能。这表明 3D-CANet 模型比基准模型能更有效地捕获时空特征。实际上每个基准模型都有自身的特点。HA、XGBoost、LSTM 和 GRU 只能预测单个区域的序列数据,并且都不能建模空间相关性。ConvLSTM、ST-ResNet、ST-3DNet 和 DeepSTNPlus 等模型确实可以捕获空间和时间信息。然而受卷积核大小的限制,它们并不能很好地捕获全局信息。因此在城市范围更大、栅格数量更多的 TaxiBJ 数据集上,它们的性能与 3D-CANet 差距更大。相对的,3D-CANet 使用 3D-InnerCA 单元,能够利用注意力机制动态捕获各个特征通道内的全局时空信息。此外,模型的 Inter-CA 可以动态学习每个区域上各特征通道的贡献,更增强了其表征能力。

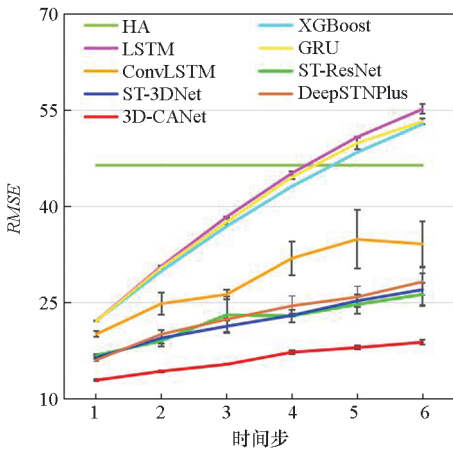
### 3.5 模型分析

#### 3.5.1 多步预测比较

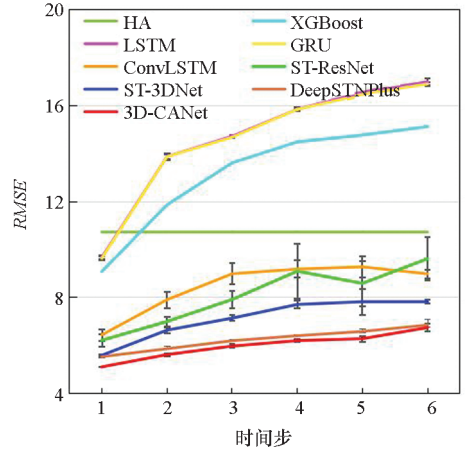
图 7 显示了 3D-CANet 模型与基准模型在三个数据集上的多步预测结果。模型要预测未来 1 至 6 个时间段的数据,每个时间段为 1 h。可以看出,预测时间越长,两种评价指标就越大。结果表明,本文提出的模型在所有模型中表现最好,这证明了 3D-InnerCA 单元捕获全局信息的能力。此外,InterCA 可以动态计算每个区域上不同特征通道的贡献,这使得本文提出的 3D-CANet 模型更加强大。

表 2 三个数据集上的预测结果  
Tab.2 Prediction results on three datasets

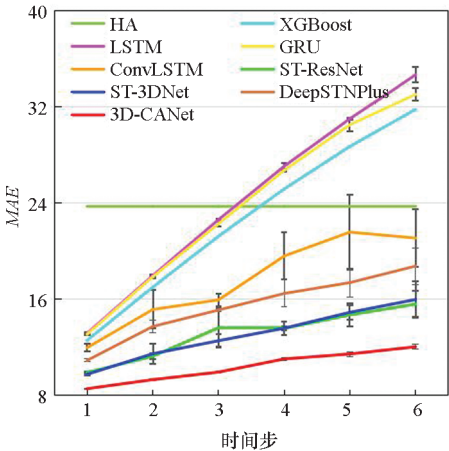
模型	TaxiBJ		BikeNYC		TaxiNYC	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
HA	46.21	23.70	10.73	5.84	26.52	8.02
XGBoost	22.07	12.53	9.08	5.33	17.20	5.98
LSTM	22.10 ± 0.05	13.13 ± 0.06	9.67 ± 0.09	6.75 ± 0.17	23.26 ± 5.13	8.28 ± 1.72
GRU	22.04 ± 0.07	13.04 ± 0.05	9.63 ± 0.09	6.65 ± 0.14	22.29 ± 4.24	7.88 ± 1.30
ConvLSTM	20.03 ± 0.44	11.92 ± 0.32	6.44 ± 0.24	4.25 ± 0.20	17.74 ± 0.77	7.61 ± 0.43
ST-ResNet	16.73 ± 0.19	9.86 ± 0.09	6.21 ± 0.24	3.80 ± 0.13	19.89 ± 1.53	7.74 ± 0.58
ST-3DNet	16.38 ± 0.19	9.70 ± 0.09	5.59 ± 0.05	3.48 ± 0.02	16.36 ± 1.40	6.21 ± 0.55
DeepSTNPlus	16.02 ± 0.25	10.86 ± 0.11	5.53 ± 0.12	2.80 ± 0.05	13.67 ± 0.26	5.29 ± 0.06
3D-CANet	12.86 ± 0.11	8.51 ± 0.05	5.11 ± 0.01	2.64 ± 0.00	12.85 ± 0.23	5.14 ± 0.11



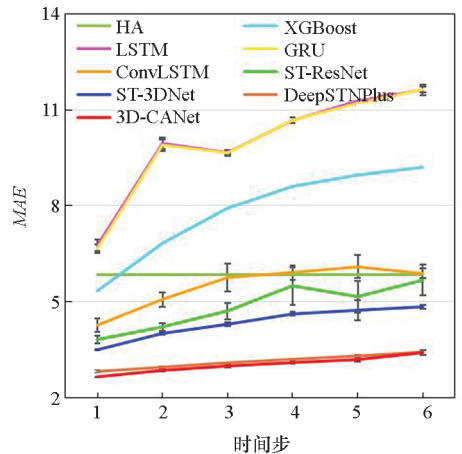
(a) TaxiBJ 的 RMSE  
(a) RMSE on TaxiBJ



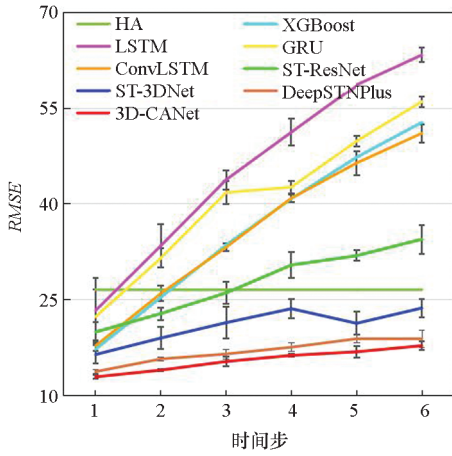
(c) BikeNYC 的 RMSE  
(c) RMSE on BikeNYC



(b) TaxiBJ 的 MAE  
(b) MAE on TaxiBJ

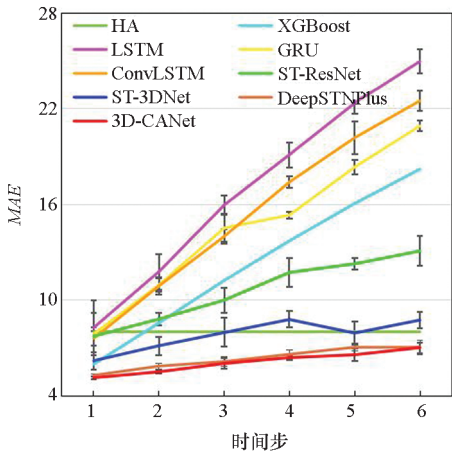


(d) BikeNYC 的 MAE  
(d) MAE on BikeNYC



(e) TaxiNYC 的 RMSE

(e) RMSE on TaxiNYC



(f) TaxiNYC 的 MAE

(f) MAE on TaxiNYC

图 7 多步预测结果

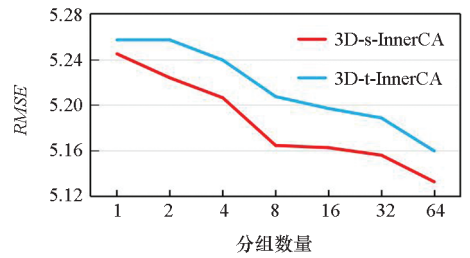
Fig. 7 Multi-step prediction result

### 3.5.2 通道分组的数量

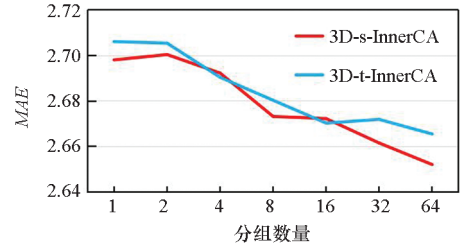
在 3D-InnerCA 单元中对通道进行分组,从而捕获每组通道内的全局相关性。如图 8 所示,实验结果表明,组数越多,模型预测的效果越好。因此,将组数设置为最大值,即通道数。这正如同本文的观点,每个通道内的全局相关性都是不同的。

### 3.6 消融实验

这里通过消融实验来验证模型各组件的效果。在表 3 中显示 3D-CANet 模型的每个组件以及它们组合后的效果。3D-Conv 模型将 3D-CANet 模型中 3D-InnerCA 单元的注意力结构替换为 3D 卷积,并且用有  $C_{out} = 2$  个大小为  $3 \times 3$  的卷积核的 2D 卷积来替代 InterCA 单元。3D-Conv\_InterCA 模型在此基础上添加了 InterCA 单元。通过比较这两个模型,可以发现 InterCA 单元使模型性能更好。此外,3D-CANet\_s 和 3D-



(a) RMSE



(b) MAE

图 8 在 BikeNYC 数据集上不同通道分组数量对 3D-s-InnerCA 和 3D-t-InnerCA 的效果

Fig. 8 Effect of different numbers of groups in 3D-s-InnerCA and 3D-t-InnerCA on BikeNYC

CANet\_t 模型分别为在 3D-InnerCA 单元中保留时间和空间分支的 3D-CANet 模型,结果表明仅用 3D-s-InnerCA 或 3D-t-InnerCA 都能使预测结果更为准确。当然,同时使用它们时,3D-InnerCA 单元的性能达到最佳。总结来说,每个组件都有助于提升 3D-CANet 模型的预测能力。

表 3 在 BikeNYC 上的消融实验

Tab. 3 Ablation experiment on BikeNYC

模型	RMSE	MAE
3D-Conv	$5.46 \pm 0.10$	$2.79 \pm 0.05$
InterCA	$5.31 \pm 0.03$	$2.73 \pm 0.03$
3D-CANet_s	$5.13 \pm 0.01$	$2.65 \pm 0.01$
3D-CANet_t	$5.16 \pm 0.04$	$2.66 \pm 0.01$
3D-CANet	$5.11 \pm 0.01$	$2.64 \pm 0.00$

## 4 结论

本文设计用于预测交通栅格流量的时空模型 3D-CANet。通过对特征通道进行分组,提出的 3D 通道内注意力单元从时间和空间两个方面进行注意力计算,动态学习不同特征通道中的全局时空相关性,有效提升了模型的时空建模能力。此外,3D 通道间注意力单元能够针对不同特征通道在各个区域上的重要性进行动态学习,提升了模型的代表能力。经过 3 个交通流量数据集的实验评



估,结果表明本文提出的 3D-CANet 模型性能优于其他 8 个基准模型。

## 参考文献 (References)

- [1] ZHANG J P, WANG F Y, WANG K F, et al. Data-driven intelligent transportation systems: a survey [J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(4): 1624 – 1639.
- [2] MA X L, TAO Z M, WANG Y H, et al. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data [J]. Transportation Research Part C: Emerging Technologies, 2015, 54: 187 – 197.
- [3] FOULADGAR M, PARHAM M, ELMASRI R, et al. Scalable deep traffic flow neural networks for urban traffic congestion prediction [C]//Proceedings of International Joint Conference on Neural Networks, 2017: 2251 – 2258.
- [4] WANG Y B, GEROLIMINIS N, LECLERCQ L. Recent advances in ITS, traffic flow theory, and network operations [J]. Transportation Research Part C: Emerging Technologies, 2016, 68: 507 – 508.
- [5] ROTA B C R, SIMIC M. Traffic flow optimization on freeways [J]. Procedia Computer Science, 2016, 96: 1637 – 1646.
- [6] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system [C]//Proceedings of the 22nd ACM SIGKDD International Conference, 2016: 785 – 794.
- [7] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553): 436 – 444.
- [8] ELMAN J L. Finding structure in time [J]. Cognitive Science, 1990, 14(2): 179 – 211.
- [9] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735 – 1780.
- [10] CHUNG J Y, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [EB/OL]. (2014 – 10 – 11) [2021 – 05 – 05]. <https://arxiv.org/abs/1412.3555v1>.
- [11] ZHAO Z, CHEN W H, WU X M, et al. LSTM network: a deep learning approach for short-term traffic forecast [J]. IET Intelligent Transport Systems, 2017, 11(2): 68 – 75.
- [12] WAN H Y, GUO S N, YIN K, et al. CTS-LSTM: LSTM-based neural networks for correlated time series prediction [J]. Knowledge-Based Systems, 2020, 191: 105239.
- [13] SHI X, CHEN Z, WANG H, et al. Convolutional LSTM network: a machine learning approach for precipitation nowcasting [C]//Proceedings of Advances in Neural Information Processing Systems, 2015, 28: 802 – 810.
- [14] ZHANG J, ZHENG Y, QI D. Deep spatio-temporal residual networks for citywide crowd flows prediction [C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence, 2017: 1655 – 1661.
- [15] GUO S N, LIN Y F, FENG N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 922 – 929.
- [16] SONG C, LIN Y F, GUO S N, et al. Spatial-temporal synchronous graph convolutional networks: a new framework for spatial-temporal network data forecasting [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1): 914 – 921.
- [17] GUO S N, LIN Y F, LI S J, et al. Deep spatial-temporal 3D convolutional neural networks for traffic data forecasting [C]//IEEE Transactions on Intelligent Transportation Systems, 2019, 20(10): 3913 – 3926.
- [18] LIN Z Q, FENG J, LU Z Y, et al. DeepSTN+: context-aware spatial-temporal neural network for crowd flow prediction in metropolis [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 1020 – 1027.
- [19] ZHANG H, GOODFELLOW L, METAXAS D, et al. Self-attention generative adversarial networks [EB/OL]. (2018 – 05 – 21) [2021 – 05 – 05]. <https://arxiv.org/abs/1805.08318>.
- [20] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local neural networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7794 – 7803.
- [21] RAMACHANDRAN P, PARMAR N, VASWANI A, et al. Stand-alone self-attention in vision models [C]//Proceedings of Advances in Neural Information Processing Systems 32, 2019.
- [22] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16 × 16 words: transformers for image recognition at scale [EB/OL]. (2020 – 10 – 22) [2021 – 05 – 05]. [https://arxiv.org/abs/2010.11929v1?utm\\_campaign=AI%2520Scholar%2520Weekly%2520&utm\\_medium=email&utm\\_source=Revue%2520newsletter](https://arxiv.org/abs/2010.11929v1?utm_campaign=AI%2520Scholar%2520Weekly%2520&utm_medium=email&utm_source=Revue%2520newsletter).
- [23] CORDONNIER J B, LOUKAS A, JAGGI M. On the relationship between self-attention and convolution [C]//Proceedings of International Conference on Learning Representations, 2020.
- [24] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 7132 – 7141.
- [25] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770 – 778.
- [26] KATHAROPOULOS A, VYAS A, PAPPAS N, et al. Transformers are RNNs: fast autoregressive transformers with linear attention [EB/OL]. (2020 – 06 – 29) [2021 – 05 – 05]. <http://arxiv.org/abs/2006.16236>.
- [27] PASZKE A, GROSS S, MASSA F, et al. Pytorch: an imperative style, high-performance deep learning library [EB/OL]. (2019 – 12 – 03) [2021 – 05 – 05]. <http://arxiv.org/abs/1912.01703>.