

一种轻量级多模态车辆轨迹预测算法

李贞妮, 孙晖, 郝梓彤, 肖冬

(东北大学信息科学与工程学院, 110819, 沈阳)

摘要: 针对自动驾驶汽车车载嵌入式计算平台存储和计算资源有限、车辆未来轨迹具有不确定性、周围环境信息复杂多变的问题,提出了一种基于 MobileNeXt 搭建的轻量级多模态车辆轨迹预测算法(CAM-MobileNeXt)。首先,利用 MobileNeXt 轻量级框架,构建了参数量和计算量均较少的车辆轨迹预测模型;其次,通过将单模态轨迹预测调整为多模态轨迹预测,以预测目标车辆可能存在的多条未来轨迹;最后,引入注意力机制,使其具备从众多输入信息中筛选出重要信息的能力,从而高效分配有限的存储和计算资源。在 L5 级别自动驾驶车辆轨迹数据集 Lyft 上开展轨迹预测实验,结果表明:所提算法具备较低的参数量和计算量,预测性能优于 Lyft 基线方法 ResNet50;与 MobileNeXt 相比,所提算法在 Lyft 数据集上的损失值降低了 11.9%,最终位移误差降低了 7.4%,平均位移误差降低了 11.4%。所提算法适合部署在自动驾驶汽车的车载嵌入式计算平台上,在对自动驾驶汽车的周围车辆进行准确多模态轨迹预测,以保证自动驾驶汽车安全行驶方面具有良好的应用前景。

关键词: 车辆轨迹预测;轻量级网络;多模态;注意力机制

中图分类号: TP391.4 **文献标志码:** A

DOI: 10.7652/xjtuxb202406002 **文章编号:** 0253-987X(2024)06-0014-10

A Lightweight Multi-Modal Vehicle Trajectory Prediction Algorithm

LI Zhenni, SUN Hui, HAO Zitong, XIAO Dong

(School of Information Science and Engineering, Northeastern University, Shenyang 110819, China)

Abstract: Aiming at the limited storage and computing resources of the embedded computing platform of self-driving vehicles, the uncertainty of future trajectory of the vehicle, and the complex and changeable surrounding environment information, a lightweight multimodal vehicle trajectory prediction algorithm (CAM-MobileNeXt) based on MobileNeXt is proposed. Firstly, a vehicle trajectory prediction model with fewer parameters and computations is constructed based on the MobileNeXt lightweight framework. Secondly, the trajectory prediction is adjusted from unimodal to multimodal to predict multiple potential future trajectories that may exist for the target vehicle. Finally, attention mechanism is introduced to enable the system to screen out important information and efficiently allocate limited storage and computing resources. In experiments conducted on the Lyft dataset for Level 5 autonomous vehicle trajectories, the results show that the proposed algorithm exhibits lower parameter and computational requirements, while outperforming the Lyft baseline method, ResNet50, in predictive performance. Compared

收稿日期: 2023-09-01。 作者简介: 李贞妮(1982—),女,讲师,硕士生导师。 基金项目: 国家自然科学基金资助项目(62273078);辽宁省博士启动基金计划资助项目(2021-BS-054);中央高校基本科研业务专项资金资助项目(N2204006)。

网络出版时间: 2024-01-23

网络出版地址: <https://link.cnki.net/urlid/61.1069.T.20240123.0913.002>

with MobileNeXt, the proposed algorithm has an 11.9% reduction in loss values on the Lyft dataset. It also exhibits a decrease of 7.4% in final displacement error and an 11.4% reduction in average displacement error. The proposed algorithm is suitable to be deployed on the embedded computing platform of self-driving vehicles, and performs accurate multi-modal trajectory prediction for the surrounding vehicles to ensure the safe driving of self-driving vehicles, indicating good application prospects.

Keywords: vehicle trajectory prediction; lightweight network; multi-modal; attention mechanism

自动驾驶系统由感知、决策、控制等多个模块组成^[1]。感知模块负责处理车载传感器采集的环境信息,并对车辆自身进行定位;决策模块负责规划车辆行驶路径,并对车辆行为进行实时决策;控制模块接收决策模块传来的控制指令,根据控制算法与执行机构的状态对车辆运动进行控制。

轨迹预测模块在自动驾驶系统中扮演着关键的角色,作为感知模块与决策模块之间的桥梁承担着重要任务。一旦感知模块获取并传递环境信息后,预测模块需要根据提供的环境信息对周围车辆的未来轨迹进行预测,决策模块再根据预测结果进行安全可行的路径规划和行为决策,提前规避潜在的碰撞或危险情况。正确的轨迹预测意味着感知模块检测到的环境信息能够对后续的决策模块和控制模块产生积极的影响,从而提前规避风险,降低事故发生的可能性,为人类带来更安全、更舒适的驾驶体验。

目前,车辆轨迹预测算法大体可分为3类,分别为基于波形处理的车辆轨迹预测算法^[2-3]、基于机器学习的车辆轨迹预测算法^[4]和基于深度学习的车辆轨迹预测算法^[5]。使用卡尔曼滤波算法的前提是系统状态模型已知,但由于目前还未能针对车辆行驶的轨迹信息建立出较好的数学模型,故卡尔曼滤波算法在车辆轨迹预测中往往不具有很好的效果。基于机器学习的车辆轨迹预测算法在车辆行驶轨迹数据已知的情况下能够具有较好的效果,但此类算法往往只能使用短期内的车辆历史轨迹,对具有长期依赖关系的轨迹难以做到精确建模,且处理非结构化数据时对模型的依赖性较大,在实际应用中可能会受到数据的限制。

深度学习近年来发展迅速,卷积神经网络(convolutional neural network, CNN)的广泛研究与应用使得人们可以提取大规模图像数据的特征,循环神经网络(recurrent neural network, RNN)的提出使得人们可以处理大规模的具有长期依赖关系的序列数据。赵健等^[6]提出了一种结合意图预测的

RNN模型,通过分析序列数据来预测车辆轨迹。Kawasaki等^[7]将卡尔曼滤波与长短时记忆网络(long short-term memory, LSTM)相结合,用于车辆的轨迹预测。此外,还有学者使用生成式对抗网络(generative adversarial network, GAN)^[8]学习历史数据的分布以生成多样化的轨迹,但其训练困难,很容易导致模式崩溃。考虑到实时系统限制的问题,并为了捕获空间的相关性,包括道路和交互等因素,研究人员引入包含丰富语义信息的高精地图,使用CNN模型进行轨迹预测^[9]。Nikhil等^[10]认为CNN模型在轨迹预测建模方面优于RNN模型,因为轨迹本质上是连续的,没有复杂的状态,而深层CNN模型作为一个高鲁棒性、高灵活性且高容量的函数逼近器,能够很好地模拟预测任务中输入与输出之间的复杂关系。此外,CNN模型还可以通过高效的卷积运算捕获数据中的空间和时间相关性。有研究者的实验结果表明:实时轨迹预测中,CNN模型能够比RNN模型具有更好的表现,这是由于该方法考虑了交通参与者和环境之间的相关性,将历史的感知信息处理成光栅化鸟瞰图(bird's-eye-view, BEV)图像训练CNN模型,使其可以从图像中提取场景上下文信息,捕捉不同交通场景中的模式和依赖关系,从而使预测轨迹更逼近真实轨迹^[11]。Gilles等^[12]也使用了类似的方法,将每个交通参与者的周围环境编码成栅格化图像作为CNN模型的输入,并输出热力图表示车辆的未来运动。Ye等^[13]将点云学习的方法引入轨迹预测以捕获时间和空间的相关性,用于预测未来时间步的车辆轨迹,但该方法在对动态交通场景的长期依赖建模和复杂特征表达方面仍存在一定不足。为此,研究人员将深度学习中的注意力机制(attention mechanism, AM)应用于轨迹预测,从大量信息中快速过滤出有价值的信息以缓解这一问题。Lin等^[14]提出了一种结合注意力的RNN模型,将其应用于预测车辆在高速公路上的运动轨迹。Cai等^[15]提出了环境注意力网络(environment-attention network, EA-Net),用于捕

获车辆与驾驶环境间的交互信息。然而,上述算法由于体积较大,因而导致推理效率不高。显然,想要在自动驾驶汽车中实现轨迹预测模块,仍存在一些实际的限制问题^[16]:①自动驾驶汽车车载嵌入式计算平台的存储和计算资源有限;②由于车辆的轨迹具有不确定性,自动驾驶汽车在行驶过程中需要预测其周围车辆不只一种可能的未来行驶轨迹,以保证车辆安全行驶;③由于相机运动、光照变化、物体遮挡等因素的影响,自动驾驶汽车的车载传感器采集到的周围环境信息复杂多变。

针对以上问题,本文提出了一种基于深度学习 MobileNeXt 算法的轻量级多模态车辆轨迹预测算法(CAM-MobileNeXt)。该算法的网络参数量和计算量较少,适合部署在存储和计算资源有限的自动驾驶汽车上,能对自动驾驶汽车周围车辆多种可能的未来轨迹进行预测。算法融合了坐标注意力(coordinate attention, CA)机制,具备从众多的输入信息中筛选出重要信息的能力,从而能够高效分配有限的计算资源。实验结果表明:提出的轻量级算法适合部署在存储和计算资源有限的自动驾驶汽车上,能以较高的准确度完成多模态轨迹预测,从而保证自动驾驶汽车的安全行驶。

1 CAM-MobileNeXt 算法描述

1.1 CAM-MobileNeXt 算法设计

基于 MobileNeXt 基础模型,本文提出了 CAM-MobileNeXt 轨迹预测算法融合 CA 注意力机制,对自动驾驶汽车周围车辆多种可能的未来轨迹进行预测。

CAM-MobileNeXt 网络的整体结构如图 1 所示,其中青绿色轨迹点为预测坐标轨迹点。车辆轨迹预测模型的第一层为尺寸为 3×3 的卷积层,用于接收包含目标车辆历史轨迹信息与周围环境信息的输入图像数据,其图像大小为 224×224 像素,通道维度则由模型选取的历史轨迹的帧数决定。该尺寸为 3×3 卷积层的卷积步长为 2,卷积核个数为 32。

在第一层卷积层之后,堆叠 19 个融合 CA 注意力的沙漏模块。在堆叠的沙漏模块之后,设置了一个全局平均池化层,用于将二维特征压缩为一维特征。车辆轨迹预测模型的最后一层是一个尺寸为 1×1 的卷积层,用于输出目标车辆未来一段时间的二维预测轨迹坐标点以及每条预测轨迹的概率。

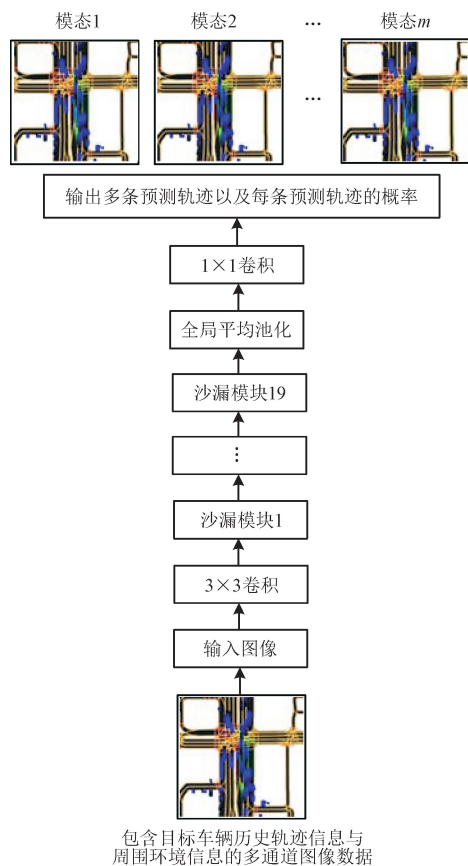


图 1 CAM-MobileNeXt 网络结构示意图

Fig. 1 The overall network architecture of CAM-MobileNeXt

1.2 沙漏模块设计

从 AlexNet^[17] 到 VGGNet^[18]、GoogLeNet^[19], 卷积神经网络的发展日趋成熟。随着网络性能的提高,网络的参数量和计算量也随之增加,很难部署和运行在存储空间和处理器性能有限的移动设备上。为解决该问题,轻量级神经网络应运而生。MobileNet^[20] 系列网络是专门为移动设备而设计的轻量级网络,旨在提高深度网络在有限硬件条件下的实时性能。MobileNetV1 使用深度可分离卷积替代标准卷积。MobileNetV2 继承了 MobileNetV1 的深度可分离卷积,同时借鉴了 ResNet^[21] 中的残差模块,提出引入跳跃连接的逆残差模块,如图 2(a)所示。MobileNetV3 在 MobileNetV2 的基础上改进了激活函数,使用了 h-swish 激活函数,还引入了压缩激励(squeeze and excitation, SE)模块,令网络通过学习自动获取到每个特征通道的重要性权重,并利用这些权重对特征进行加权处理,使网络能够更有效地利用特征信息,性能得到了提高^[22]。

本文提出的基于 MobileNeXt 的车辆轨迹预测算法的沙漏模块结构与逆残差模块不同,如图 2(b)

间位置注意力信息,卷积只能捕获局部关系,无法建立长程依赖关系。

CA 注意力机制不仅捕获通道注意力信息,还捕获方向感知和位置敏感的空间注意力信息^[27]。在 CA 模块中,为了减轻由二维全局池化引起的位置信息丢失,坐标注意力先将通道注意力的二维全局池化分解为两个一维池化,分别沿两个空间方向聚合特征,生成一对具有全局感受野且保留了精确位置信息的坐标描述矩阵;再将两个坐标描述矩阵分别编码,采用 Sigmoid 激活函数生成一对方向感知和位置敏感的坐标注意力特征,沿着一个空间方向捕获长程依赖关系,并沿着另一个空间方向保留精确的位置信息;最后将两个坐标注意力特征互补地作用于输入特征,以增强对关注对象的表示。输入模型的图像数据经过多个堆叠的融合 CA 注意力的沙漏模块,利用有限的计算资源自动学习图像特征的重要程度并关注重要特征,能够显著提升网络的特征表达能力,提高了网络性能。因此,相较于 SE 注意力机制和 CBAM 注意力机制,CA 注意力机制可以使网络获得更好的性能。

1.3 多模态损失函数设计

为了评估与优化多模态车辆轨迹预测模型的性能,本文设计了一种基于负对数似然(negative log-likelihood, NLL)的多模态损失函数。假设真实轨迹点 $s = (x, y)$, 其中 x 为真实轨迹横坐标, y 为真实轨迹纵坐标;预测轨迹点 $\hat{s} = (\hat{x}, \hat{y})$, 其中 $\hat{x}$ 为预测轨迹横坐标, $\hat{y}$ 为预测轨迹纵坐标。

未来一段时间内的真实轨迹 s_{real} 可定义为

$$s_{\text{real}} = \{s_i\}, i = 1, 2, \dots, t \quad (1)$$

式中: i 为时间步长索引; t 为未来时间总步长。

未来一段时间内预测的多模态轨迹 s_{pred} 可定义为

$$s_{\text{pred}} = \{\hat{s}_i^k\}, i = 1, 2, \dots, t, k = 1, 2, \dots, m \quad (2)$$

式中: k 为模态序数; m 为模态总数。

对应预测的多个模态置信度 c_{pred} 可定义为

$$c_{\text{pred}} = \{c^k\}, k = 1, 2, \dots, m \quad (3)$$

式中: c^k 为第 k 个模态的置信度。

假设真实轨迹遵循多维独立正态分布,则似然概率可写为

$$p(s_{\text{real}} | c_{\text{pred}}, s_{\text{pred}}) = \sum_{k=1}^m c^k \prod_{i=1}^t \mathcal{N}(s_i | \hat{s}_i^k) \quad (4)$$

式中: $\mathcal{N}(\cdot)$ 为正态分布。

多模态损失值 $L_{\text{oss}}^{\text{MM}}$ 可写为

$$\begin{aligned} L_{\text{oss}}^{\text{MM}} &= -\ln(p(s_{\text{real}} | c_{\text{pred}}, s_{\text{pred}})) = \\ &= -\ln\left(\sum_{k=1}^m \exp(\ln c^k + \sum_{i=1}^t \ln(\mathcal{N}(s_i | \hat{s}_i^k)))\right) = \\ &= -\ln\left(\sum_{k=1}^m \exp(\ln c^k - \frac{1}{2} \sum_{i=1}^t \|s_i - \hat{s}_i^k\|_2^2)\right) \end{aligned} \quad (5)$$

综上,本文提出的多模态车辆轨迹预测算法的损失函数可写为

$$\begin{aligned} L_{\text{oss}}^{\text{MM}} &= -\ln\left(\sum_{k=1}^m \sum \exp(\ln c^k - \right. \\ &\quad \left. \frac{1}{2} \sum_{i=1}^t \|s_i - \hat{s}_i^k\|_2^2)\right) \end{aligned} \quad (6)$$

2 实验与分析

2.1 数据集及评估指标

本文选用 Lyft 数据集验证车辆轨迹预测算法的性能。该数据集是迄今全球最大的 L5 级别自动驾驶车辆轨迹数据集,包含超过 1 000 h、26 000 km 的数据^[28]。该数据集由 17 万个场景组成,每个场景约 25 s,包含了自动驾驶车辆遇到的汽车、自行车、行人等交通参与者的轨迹信息,还包含了带有 15 242 个标记元素的高清语义鸟瞰图与卫星地图,如图 4 所示。图 4(a)为光栅化后的语义地图,可选择不同的图块和线条用以标记道路结构和交通参与者。图 4(b)为卫星的航拍地图,并标记了交通参与者。该数据集支持多种方式对原始数据进行有效采样,以实现交通场景的精确建模。因此,本文选用该数据集验证车辆轨迹预测算法的性能。



(a) 语义地图

(b) 航拍地图

图 4 Lyft 数据集图像数据^[28]

Fig. 4 Image data of Lyft dataset^[28]

在 Lyft 数据集中,原始数据可用图像形式表示。该图像是由多种传感器(激光雷达、GPS、惯性传感器等)采集的原始数据进行处理得到,其将环境信息光栅化为像素,以 0.5 m 的图像分辨率呈现,大小为 224×224 像素。语义鸟瞰图图像数据的可视化结果如图 5 所示,单帧的数据包含了 3 种图像,分别为单通道的目标车辆语义鸟瞰图、单通道的周

围障碍物语义鸟瞰图和3通道RGB格式的道路语义鸟瞰图。图5(a)和5(b)以黑色矩形框绘制车辆、行人等交通代理,显示其长、宽、位置及航向角等信息(通过平移和旋转整个场景以使目标车辆位于图像中央且与X轴对齐)。为了保持一致,所有代理的坐标都采用相同的局部坐标系。图5(c)显示了道路结构、车道线、交通信号等环境信息,其中黑色区域表示道路,彩色车道边界表示红绿灯状态,如绿色车道表示绿灯亮起。历史多帧的多通道语义鸟瞰图包含目标车辆及周围代理历史状态等动态信息和道路结构等静态信息,其可视化结果如图5(d)所示,其中绿色矩形为目标车辆,蓝色矩形为其他交通参与者。

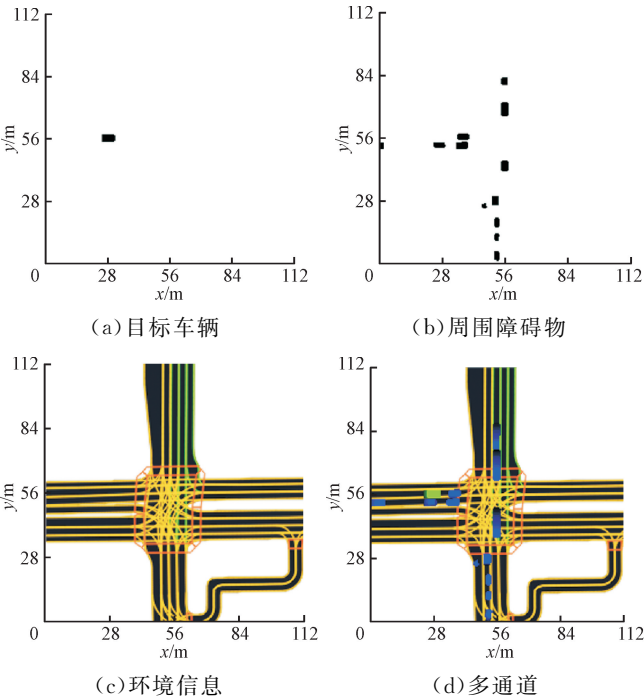


图5 语义鸟瞰图图像数据可视化结果
Fig. 5 Visualized results of BEV image data

本文将图5所示图像作为车辆轨迹预测模型的输入,通过计算,输出目标车辆未来一段时间的多模态轨迹。输入图像数据通道维度由模型选取的车辆历史轨迹的帧数决定,计算公式为

$$n_{in}^{channels} = 2n_{history}^{frames} + 3 \quad (7)$$

式中: $n_{in}^{channels}$ 为输入图像数据通道数; $n_{history}^{frames}$ 为历史轨迹帧数。

车辆轨迹预测模型的输出数据为目标车辆可能存在的多条预测未来轨迹的二维坐标点以及每条轨迹的概率,其特征维度的计算公式为

$$n_{out}^{frames} = 2mn_{future}^{frames} + m \quad (8)$$

式中: n_{out}^{frames} 为多模态输出轨迹数据特征数; $2mn_{future}^{frames}$ 为未来轨迹的帧数。

为了评价车辆预测算法的性能,本文选用3种车辆轨迹预测算法评价指标,分别为时间位移误差、平均位移误差和最终位置误差。时间位移误差(time displacement error, TDE)定义为预测时间范围内某一时间步的预测位置与真实位置之间的距离。平均位移误差(average displacement error, ADE)定义为预测时间范围内所有时间步的预测位置与真实位置之间距离的平均值。最终位移误差(final displacement error, FDE)定义为预测最终位置与真实最终位置之间的距离。不考虑预测时间范围内其他时间步长产生的误差,所定义的3种误差指标越小,算法的预测精度越高。3种误差的计算公式分别为

$$E_{TDE}^i = \|y_{pred}^i - y_{target}^i\|_2 \quad (9)$$

$$E_{ADE} = \frac{\sum_{i=1}^t \|y_{pred}^i - y_{target}^i\|_2}{t} \quad (10)$$

$$E_{FDE} = \|y_{pred}^{final} - y_{target}^{final}\|_2 \quad (11)$$

式中: E_{TDE}^i 为 i 时刻的时间位移误差; E_{ADE} 为平均位移误差; E_{FDE} 为最终位移误差; y_{pred}^i 为 i 时刻的预测位置; y_{target}^i 为 i 时刻的真实位置; y_{pred}^{final} 为预测最终位置; y_{target}^{final} 为真实最终位置。

在轻量化模型评估中,除了预测精度外,参数量和计算量同样至关重要。参数量是指模型的可学习参数数量,其数量的减少有助于节省存储空间和内存,提升模型加载速度。计算量则代表模型所需的计算资源,低计算量意味着对资源的需求减少,可更快实现推理任务,满足实时性需求。

综上,一个合格的轻量级轨迹预测模型需要在模型的参数量、计算量和准确度之间实现平衡,在设备的存储空间和处理器性能有限的前提下达到满足任务需求的准确性。

2.2 实验设置

本文算法基于 Pytorch1.7 深度学习框架进行开发,在 NVIDIA 2080Ti GPU 进行深度学习模型的训练。为了验证本文提出的轻量级多模态车辆轨迹预测算法各模块的有效性,在 Lyft 数据集上实验了多种不同的车辆轨迹预测算法,分别比较了 ResNet50、MobileNetV2、MobileNetV3、MobileNetXt、SE-MobileNetXt、CBAM-MobileNetXt 和 CAM-MobileNetXt 模型的性能。其中,ResNet50 模型为 Lyft 数据集官方提供的基准模型。

在数据参数方面,将历史轨迹帧数设置为 10,模型的输入为目标车辆 1.0 s (10 帧)的历史轨迹;将预测轨迹帧数设置为 50,模型的输出为目标车辆 5.0 s (50 帧)的预测轨迹;将预测轨迹的模态数设置为 3,模型输出目标车辆未来可能行驶的 3 条预测轨迹。由于 Lyft 数据集数据量巨大,为了减少训练时间和硬件消耗,仅使用数据集中 10% 的数据训练车辆轨迹预测模型。将训练样本数量设置为 23 355 429,批样本数量设置为 128,模型在训练数据上累计迭代约 18 万次。

在模型参数方面,为了使各轻量级模型的参数量和计算量大致相同,将 MobileNetV2 模型的宽度超参数设置为 1.0,MobileNetV3 模型的宽度超参数设置

为 0.75, MobileNeXt、SE-MobileNeXt、CBAM-MobileNeXt 和 CA-MobileNeXt 模型的通道超参数均设置为 1.0,SE-MobileNeXt、CBAM-MobileNeXt 和 CA-MobileNeXt 模型的通道降维率均设置为 16。

在训练参数方面,将优化器设置为随机梯度下降(stochastic gradient descent,SGD)优化器,其动量设置为 0.9,权重衰减设置为 4×10^{-5} ;将学习率设置为指数衰减学习率,其初始学习率设置为 0.1,衰减率设置为 0.95。

2.3 实验结果与分析

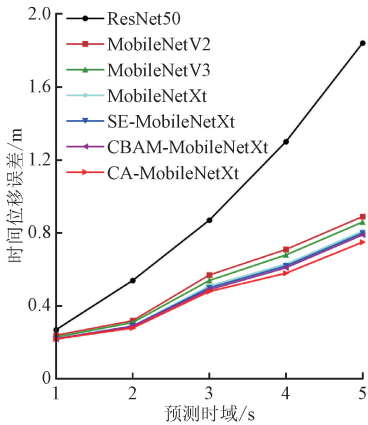
不同多模态车辆轨迹预测算法在 Lyft 数据集上的实验结果如表 1 所示,各模型的性能及模型参数量与计算量的对比结果如图 6 所示。

表 1 不同多模态车辆轨迹预测算法实验结果对比

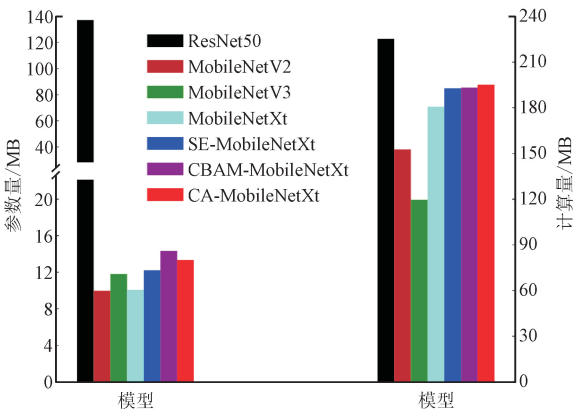
Table 1 Comparison of experimental results of different multimodal vehicle trajectory prediction algorithms

模型	参数量/ MB	计算量/ MB	损失值	时间位移误差/m					
				1 s	2 s	3 s	4 s	5 s	平均值
ResNet50(基准)	137.52	225.49	104.19	0.27	0.54	0.87	1.3	1.84	0.96
MobileNetV2	9.99	152.87	25.31	0.24	0.32	0.57	0.71	0.89	0.49
MobileNetV3	11.84	119.78	23.15	0.23	0.31	0.54	0.68	0.86	0.47
MobileNeXt	10.10	180.75	21.64	0.22	0.29	0.51	0.63	0.81	0.44
SE-MobileNeXt	12.23	192.87	20.38	0.22	0.29	0.50	0.62	0.80	0.43
CBAM-MobileNeXt	14.36	193.46	19.95	0.22	0.29	0.49	0.61	0.79	0.42
CAM-MobileNeXt	13.35	195.32	19.06	0.22	0.28	0.48	0.58	0.75	0.39

注:表中加粗数据为最优值。



(a) 时间位移误差



(b) 参数量和计算量

图 6 不同模型时间位移误差及参数量和计算量的对比结果

Fig. 6 Comparison results of displacement errors, parameter quantities and computational complexity among different models

由图 6 可见,MobileNeXt 模型的性能全面优于基准模型,前者的时间位移误差和参数量显著降低,计算量也下降了 19.8%。此外,经过计算可知,在参数量和计算量大致相同的情况下,MobileNeXt 模型在各项

数据指标中均优于轻量级网络 MobileNetV2 和 MobileNetV3 模型,其损失值为 21.64,最终位移误差为 0.81 m,平均位移误差为 0.44 m。

在 MobileNeXt 的基础上分别融合 SE、CBAM、

CA 3 种注意力机制,虽然会导致模型的参数量和计算量略大于原模型,但模型的预测精度均有一定提高。经过计算发现,本文设计的融合 CA 注意力的 CAM-MobileNeXt 模型预测准确度最高,其损失值为 19.06,平均位移误差为 0.39 m,最终位移误差为 0.75 m。虽然该模型的参数量相较于轻量网络 MobileNetV2、MobileNetV3、MobileNeXt 分别增长了 33.6%、12.7%、32.2%,但平均位移误差分别下降了 20.4%、17%、11.4%。

此外,在参数量和计算量相似的情况下,CAM-MobileNeXt 模型在预测性能上优于 SE-MobileNeXt 和 CBAM-MobileNeXt 模型,表现为各个时刻的时间位移误差均低于 SE-MobileNeXt 和 CBAM-MobileNeXt 模型,且预测时间越长降低得越明显,4 s 时的时间位移误差分别下降了 6.5%、4.9%,5 s 时的时间位移误差分别下降了 6.3%、5.1%。

综上,相较于 MobileNeXt 等参数量和计算量均较小的轻量级系列模型,CAM-MobileNeXt 在只稍微增加存储和计算成本的情况下,实现了更高的预测精度。相较于其他模型,CAM-MobileNeXt 的性能最优,在预测精度、参数量和计算量之间取得了良好平衡,能满足轻量级网络的要求。

选取多种场景下模型输出的多模态轨迹中置信度最高即概率最大的预测结果作为展示,CAM-MobileNeXt 轻量级多模态车辆轨迹预测算法的预测轨迹可视化结果如图 7 所示。图片以鸟瞰图的形式,从自动驾驶车辆(self-driving vehicle, SDV)的视角展示了多种交通场景下周围车辆的未来轨迹预测结果。场景中,使用矩形表示车辆,线条表示轨迹,彩色车道表示交通信号灯,并标注了自动驾驶车辆的运动轨迹。

图 7(a)为最常见的直行场景,此时 SDV 处于直行状态,预测同向车道的前后车辆以及邻近对向车道车辆将会直行。

图 7(b)为路口红灯场景,红色车道表示当前 SDV 行驶方向的交通信号为红灯,自动驾驶车辆处于等待状态,预测路口两侧车辆将直行通过路口,而 SDV(红色矩形)后方的车辆将会减速刹车,停在 SDV 后方与其一起等待交通信号灯变化。

图 7(c)为路口绿灯场景,绿色车道表示当前 SDV 行驶方向的交通信号为绿灯,自动驾驶车辆将直行通过路口,预测其同向车道的周围车辆将会直行或向左转弯进入另外车道,该路口两侧部分车辆的预测结果为保持停车趋势,等待其他车辆通过路口。

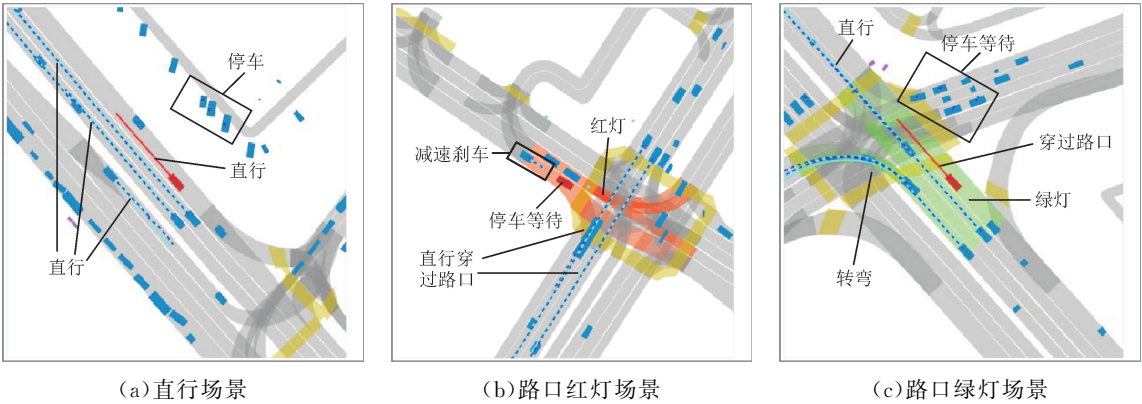


图 7 CAM-MobileNeXt 算法预测轨迹可视化结果

Fig. 7 Visualization of trajectory prediction results using the CAM-MobileNeXt algorithm

以上可视化结果表明,本文设计的 CAM-MobileNeXt 算法对于待预测的目标车辆在转弯、直行、停车、制动等常见的行驶场景,均能够以较高的准确度完成多模态轨迹预测。

3 结论与展望

本文针对自动驾驶中车辆轨迹预测提出了一种 CAM-MobileNeXt 轻量级多模态车辆轨迹预测算

法,其网络的参数量和计算量较少,适合于部署和运行在移动和嵌入式设备;通过将单模态轨迹预测算法改进为多模态轨迹预测算法,对自动驾驶汽车周围车辆多种可能的未来轨迹进行预测;融合 CA 注意力机制高效分配有限的计算资源,提高了模型的特征表达能力。仿真结果表明:本文设计的 CAM-MobileNeXt 轻量级多模态车辆轨迹预测算法在参数和计算成本较低的情况下,对于常见的交通场景

能够以较高的准确度完成多模态轨迹预测。

该算法适合部署在存储和计算资源有限的自动驾驶汽车上,能够辅助车辆安全行驶,但仍存在一定的不足和局限。由于算法尚未经过实际场景的部署和验证,因此车辆感知系统获取的数据可能会带有噪声,从而影响模型预测的准确性。在未来的工作中,可以探索预测算法与感知系统的整合与不确定性估计,提高算法的鲁棒性。此外,也可在部署过程中结合使用网络剪枝、量化和知识蒸馏等模型压缩方法,以期构建出更加高效的轻量网络应用于自动驾驶车辆。

参考文献:

- [1] 段续庭,周宇康,田大新,等. 深度学习在自动驾驶领域应用综述[J]. 无人系统技术, 2021, 4(6): 1-27.
DUAN Xuting, ZHOU Yukang, TIAN Daxin, et al. A review of deep learning applications for autonomous driving[J]. Unmanned Systems Technology, 2021, 4(6): 1-27.
- [2] 张亮修,张铁柱,吴光强. 考虑误差校正的智能车辆路径跟踪鲁棒预测控制[J]. 西安交通大学学报, 2020, 54(3): 20-27.
ZHANG Liangxiu, ZHANG Tiezhu, WU Guangqiang. Robust predictive control for intelligent vehicle path tracking considering error feedback correction[J]. Journal of Xi'an Jiaotong University, 2020, 54(3): 20-27.
- [3] 乔少杰,韩楠,朱新文,等. 基于卡尔曼滤波的动态轨迹预测算法[J]. 电子学报, 2018, 46(2): 418-423.
QIAO Shaojie, HAN Nan, ZHU Xinwen, et al. A dynamic trajectory prediction algorithm based on Kalman filter[J]. Acta Electronica Sinica, 2018, 46(2): 418-423.
- [4] 陈雪梅,李梦溪,王子嘉,等. 无人驾驶车辆城市交叉口周边车辆轨迹预测[J]. 汽车工程学报, 2021, 11(4): 235-242.
CHEN Xuemei, LI Mengxi, WANG Zijia, et al. Trajectory prediction of surrounding vehicles for unmanned vehicle at urban intersections[J]. Chinese Journal of Automotive Engineering, 2021, 11(4): 235-242.
- [5] 李雪松,张锬石,宋呈群,等. 自动驾驶场景下的轨迹预测技术综述[J]. 计算机工程, 2023, 49(5): 1-11.
LI Xuesong, ZHANG Qieshi, SONG Chengqun, et al. Review of trajectory prediction technology in autonomous driving scenes[J]. Computer Engineering, 2023, 49(5): 1-11.
- [6] 赵健,宋东鉴,朱冰,等. 数据机理混合驱动的交通车意图识别方法[J]. 汽车工程, 2022, 44(7): 997-1008.
ZHAO Jian, SONG Dongjian, ZHU Bing, et al. Traffic vehicles intention recognition method driven by data and mechanism hybrid[J]. Automotive Engineering, 2022, 44(7): 997-1008.
- [7] KAWASAKI A, SEKI A. Multimodal trajectory predictions for urban environments using geometric relationships between a vehicle and lanes[C]//2020 IEEE International Conference on Robotics and Automation (ICRA). Piscataway, NJ, USA; IEEE, 2020: 9203-9209.
- [8] GUO Hongyan, MENG Qingyu, ZHAO Xiaoming, et al. Map-enhanced generative adversarial trajectory prediction method for automated vehicles[J]. Information Sciences, 2023, 622: 1033-1049.
- [9] CHOU F C, LIN T H, CUI Henggang, et al. Predicting motion of vulnerable road users using high-definition maps and efficient ConvNets[C]//2020 IEEE Intelligent Vehicles Symposium (IV). Piscataway, NJ, USA; IEEE, 2020: 1655-1662.
- [10] Nikhil N, MORRIS B T. Convolutional neural network for trajectory prediction[C]//Computer Vision-ECCV 2018 Workshops. Cham, Germany; Springer International Publishing, 2019: 186-196.
- [11] CASAS S, GULINO C, SUO S, et al. The importance of prior knowledge in precise multimodal prediction[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway, NJ, USA; IEEE, 2020: 2295-2302.
- [12] GILLES T, SABATINI S, TSISHKOU D, et al. HOME: heatmap output for future motion estimation[C]//2021 IEEE International Intelligent Transportation Systems Conference (ITSC). Piscataway, NJ, USA; IEEE, 2021: 500-507.
- [13] YE Maosheng, CAO Tongyi, CHEN Qifeng. TPCN: temporal point cloud networks for motion forecasting[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA; IEEE, 2021: 11313-11322.
- [14] LIN Lei, LI Weizi, BI Huikun, et al. Vehicle trajectory prediction using LSTMs with spatial-temporal attention mechanisms[J]. IEEE Intelligent Transportation Systems Magazine, 2022, 14(2): 197-208.
- [15] CAI Yingfeng, WANG Zihao, WANG Hai, et al. Environment-attention network for vehicle trajectory prediction[J]. IEEE Transactions on Vehicular Technology, 2021, 70(11): 11216-11227.

- [16] MOZAFFARI S, AL-JARRAH O Y, DIANATI M, et al. Deep learning-based vehicle behavior prediction for autonomous driving applications: a review [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(1): 33-47.
- [17] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [18] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2015-04-10)[2023-05-26]. <https://arxiv.org/abs/1409.1556>.
- [19] SZEGEDY C, LIU Wei, JIA Yangqing, et al. Going deeper with convolutions [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2015: 1-9.
- [20] HOWARD A G, ZHU Menglong, CHEN Bo, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. (2017-04-17)[2023-06-10]. <https://arxiv.org/abs/1704.04861>.
- [21] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.
- [22] HOWARD A, SANDLER M, CHEN Bo, et al. Searching for MobileNetV3 [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2019: 1314-1324.
- [23] SANDLER M, HOWARD A, ZHU Menglong, et al. MobileNetV2: inverted residuals and linear bottlenecks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 4510-4520.
- [24] 张宸嘉, 朱磊, 俞璐. 卷积神经网络中的注意力机制综述 [J]. 计算机工程与应用, 2021, 57(20): 64-72. ZHANG Chenjia, ZHU Lei, YU Lu. Review of attention mechanism in convolutional neural networks [J]. Computer Engineering and Applications, 2021, 57(20): 64-72.
- [25] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7132-7141.
- [26] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
- [27] HOU Qibin, ZHOU Daquan, FENG Jiashi. Coordinate attention for efficient mobile network design [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2021: 13708-13717.
- [28] HOUSTON J, ZUIDHOF G, BERGAMINI L, et al. One thousand and one hours: self-driving motion prediction dataset [C]//Proceedings of the 2020 Conference on Robot Learning. Chia Laguna Resort, Sardinia, Italy: PMLR, 2021: 409-418.

(编辑 李慧敏 刘杨)