

基于 LSTR 和 Vit-CoMer 骨干的车道线检测方法^{*}郑 剑^{1,2} 郭亦畅^{1,2}

(1. 江西理工大学信息工程学院 赣州 341000; 2. 江西理工大学宜春新能源锂电研究院 宜春 336023)

摘 要: 针对 LSTR 算法在实际应用中存在的提取特征尺度单一及缺乏对车道局部特征有效捕捉的问题。本文首次将 Vit-CoMer 骨干网络用于车道线检测任务中,提出 LSCoMer 车道线检测模型。首先,在特征提取网络后使用 MRFP 丰富多尺度特征,提高检测精度;其次,在 Transformer 结构的开始和结束位置集成 CTI 模块,以促进 CNN 的局部特征与 Transformer 的全局特征之间有效融合,强化后者在局部细节上的敏感性。实验结果表明,本文方法在 TuSimple 数据集上准确率为 96.68%,较原 LSTR 方法提升 0.5% 且显著优于 PolyLaneNet 等同类方法,在 CULane 数据集中,本文方法 F_1 分数比 LSTR 方法提升 3.02%。

关键词: 车道线检测;LSTR 算法;Transformer;多尺度特征

中图分类号: TN209 **文献标识码:** A **国家标准学科分类代码:** 510.99

Lane detection method based on LSTR and ViT-CoMer

Zheng Jian^{1,2} Guo Yichang^{1,2}

(1. School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, China;

2. Yichun Lithium New Energy Industry Research Institute, Jiangxi University of Science and Technology, Yichun 336023, China)

Abstract: Addressing the limitations of the LSTR algorithm in practical applications, particularly its single-scale feature extraction and lack of effective capture of local lane features, this paper introduces the Vit-CoMer backbone network for the first time in lane detection tasks, proposing the LSCoMer lane detection model. Initially, the model employs a MRFP module after the feature extraction network to enrich multi-scale features, thereby enhancing detection accuracy. Additionally, a CTI module is integrated at both the beginning and the end of the Transformer structure to promote effective fusion between CNN's local features and Transformer's global features, enhancing the latter's sensitivity to local details. Experimental results indicate that this method achieves an accuracy of 96.68% on the TuSimple dataset, which is a 0.5% improvement over the original LSTR method and significantly outperforms similar methods like PolyLaneNet. On the CULane dataset, our method improves the F_1 score by 3.02% compared to the LSTR method.

Keywords: lane detection; LSTR algorithm; Transformer; multi-scale features

0 引 言

智能驾控系统^[1-2]是传统车企与新能源车企不断攻坚的技术领域,在自动驾驶技术蓬勃发展的今天,它能够为驾驶员提供全场景路面信息,并结合激光雷达和视觉传感器保障车辆行车安全。

在过去的研究中,车道线检测方法主要分为两大类:传统手工提取特征方法和深度学习方法。传统方法主要依靠手工提取车道特征,这些方法又可分为基于特征方法^[3-5]和基于模型的方法^[6-7]。前者将输入图像经过灰度变换后,利

用边缘检测算法,如 Canny 边缘检测算法等,提取特征信息,再使用 Hough 变换检测车道位置,后者则利用数学模型建立与结构化路面相吻合的车道线参数模型。传统方法实现简单,在光照条件充足的结构化路面表现良好,但其在复杂环境下抗噪能力不足,拟合车道线时不能很好利用全局上下文信息,检测效率不高。

随着深度学习在图像处理领域的快速发展,深度学习技术受到研究者青睐。基于深度学习的车道线检测方法分为基于检测的方法^[8-10]和基于分割的方法^[11-14]。Qin 等^[9]首次将车道线检测问题视为居于全局特征的行锚分类定位

问题,显著降低了计算成本,同时提出一种结构性损失函数拟合车道线位置,实时性达到同期最优,但基于行锚分类方法对位置信息感知较弱,检测精度较低。Tabelini 等^[10]提出一种新颖的基于锚点和注意力机制的车道线检测模型 LaneATT,该模型使用锚点进行特征池化步骤,提出基于锚点的注意力机制,提升了模型处理车道线存在遮挡和缺失场景的能力。Pan 等^[11]充分利用卷积的空间特性,提出 SCNN 空间卷积神经网络,该方法从多个方向对特征图切片并沿切片方向进行卷积计算,此策略增加了特征间信息交互。然而,该方法增加了特征信息在相邻切片之间的传递成本,且推理速度较慢。Liu 等^[12]基于条件卷积提出 CondLaneNet,一种自顶向下的车道实例检测方法。此方法根据输入场景中车道分布动态调整卷积核的权重,模型实现了高精度和鲁棒性的统一。Neven 等^[13]采用实例分割思想提出 LaneNet 车道检测网络,并加入透视变换结构(H-Net),在后处理阶段,将语义分割和像素聚类结合,车道检测任务被转换为像素级的分类预测任务。基于分割的方法能准确探测图像中车道线的位置,但此类方法通常有复杂的后处理,实时性难以满足要求。Liu 等^[14]将 Transformer^[15]结构引入车道线检测任务中,提出 LSTR (lane shape prediction with Transformers)检测模型,该方法创新地用 Transformer 结构完全替代特征提取及后处理步骤,通过端到端的方式直接输出车道线形状参数,极大程度简化了处理流程,增加了模型的可解释性,检测速度达到了惊人的 420 fps。张莹等^[16]提出 LSTR 的改进方法,引入 CBAM 模块和掩码单头注意力机制,利用通道和空间上的信息捕捉更多细节,虽然 CBAM 对局部特征有增强作用,但 Transformer 架构的本质更擅长处理全局特征,单靠 CBAM 模块的提升仍显有限,且多头注意力剪枝虽降低复杂度,但可能导致信息损失,从而降低检测精度。尽管 LSTR 模型在同期实现 SOTA 效果,但其在建模过程中也存在某些不足:一是 Transformer 结构通常在相同分辨率下处理输入特征,导致其在处理多尺度特征方面存在局限;二是对车道局部特征有效捕捉能力不足,当车道线出现局部磨损、断裂或部分遮挡时,LSTR 模型可能无法响应某些局部变化。

针对 LSTR 模型在实际应用中存在的上述两点不足,本文基于 LSTR 模型,结合 Vit-CoMer 骨干网络在密集检测中的算法思想,提出 LSCoMer 车道检测模型,在保证实时性的前提下,提升检测精度。相较于 LSTR 的改进方法^[16],本文不局限于简单的模块插入,而是深度融合局部与全局信息,突破了传统 Transformer 对局部细节捕捉能力的限制。本文的主要贡献如下:

1) 首次将 Vit-CoMer 密集检测骨干应用于车道线检测任务中,创新性地提出了 LSCoMer 车道检测模型,通过对不同尺度特征的有效提取和深度融合,使模型在复杂交通环境中良好鲁棒性和精确度。

2) 为解决 LSTR 模型提取特征尺度单一的问题,本文使用多感受野特征金字塔模块(multi-receptive field pyramid module, MRFP)对骨干网络提取的特征进行增强,特征金字塔结构和不同大小卷积核的深度可分离卷积丰富了多层次语义信息,进一步提升了模型的检测能力。

3) 为提高 Transformer 结构对局部信息有效捕捉的能力,本文对 LSTR 模型结构进行优化,在 Transformer 结构的开头和结尾处增加 CNN-Transformer 双向交互融合模块(CNN-Transformer bidirectional Interaction fusion module, CTI),促进 CNN 与 Transformer 的之间的信息交互,缓解编码器处理序列时缺乏局部特征信息的问题。

4) 本文模型集成了两阶段 CoMerBlock,实验证明了本文模型在实际应用中的有效性。

1 LSTR 算法

1.1 算法原理

LSTR 算法^[14]是结合深度学习与 Transformer 架构的高效车道线检测算法,如图 1 所示,算法结构由特征提取骨干网络、Transformer 结构、前馈神经网络和匈牙利损失函数组成。首先输入图像 I 经过特征提取骨干提取得到丰富语义信息的底层视觉特征,这些特征随后转换成一维序列 S 输入 Transformer 编码器中,同时引入位置编码 E_p 帮助模型理解输入数据的顺序关系。编码器输出序列 S_e 同学习位置嵌入序列 E_{LL} 流入入 Transformer 解码器中,解码器首先处理初始化查询序列 S_q 作为预测每条车道线的起始位置,然后通过注意力机制与 S_e 进行交互,这种交互允许解码器层在预测车道线时利用编码器捕获的全局特征,从而输出包含车道线形状编码的序列 S_d 。一组前馈神经网络基于序列 S_d 预测车道线具体形状参数。LSTR 使用匈牙利算法计算损失函数,在模型的预测输出和真实标签之间进行最优匹配,确保预测结果与真实情况的一致性。

1.2 车道形状参数化表示

LSTR 算法的主要特点是将车道线拟合过程描述为曲线参数的输出,参数主要由道路结构和摄像机摆放位姿决定,这种拟合策略省略了后处理步骤同时增加了计算机输出的物理解释性。模型将车道线定义为道路上的多项式,通常在宽敞无遮挡的路面上,使用三次曲线来拟合车道形状:

$$X = kZ^3 + mZ^2 + nZ + b \quad (1)$$

其中, k, m, n, b 为实数系数,且 $k \neq 0$ 。路面上的点用 (X, Z) 表示,当摄像机光轴平行路面位置时,车道线被映射到平面上曲线:

$$u = \frac{k'}{v^2} + \frac{m'}{v} + n' + b' \times v \quad (2)$$

其中, k', m', n', b' 为摄像机内外参数与实数系数的复合参数。 (u, v) 表示平面内的像素点。当摄像机固定位置与路面呈 ϕ 角度时,拟合车道线的曲线变换为:

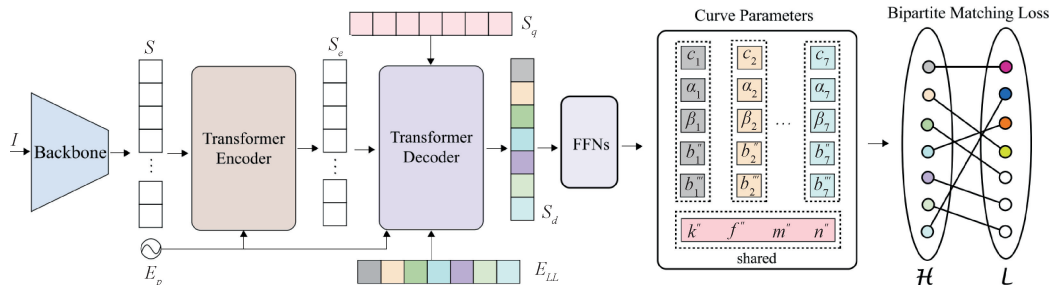


图 1 LSTR 算法结构图

Fig. 1 Structure of LSTR algorithm

$$u' = \frac{k' \times \cos^2 \phi}{(v' - f \sin \phi)^2} + \frac{m' \cos \phi}{(v' - f \sin \phi)} + n' + \frac{b' \times v'}{\cos \phi} - b' \times f \tan \phi \quad (3)$$

其中, f 表示以像素为单位的焦距, (u', v') 表示对应螺旋距转换后的点位。当曲线参数与相机倾斜角 ϕ 相结合, 平面内对应曲线被表示为:

$$u' = \frac{k''}{(v' - f'')^2} + \frac{m''}{(v' - f'')} + n' + b'' \times v' - b'' \quad (4)$$

其中, n' 和 b'' 为常数, 其他参数意义与上述多项式相同。在真实路面场景中, 如图 1 所示, 车道线曲率较大处往往出现在画面最前端, 且车道线之间通常以平行方式排列, 基于以上观测, 所有车道线可共享参数 k'' 、 f'' 、 m'' 、 n'' , 则画面中第 t 条车道线 g_t 可被表示为:

$$g_t = (k'', f'', m'', n', b_i'', b_i'', \alpha_i, \beta_i) \quad (5)$$

其中, α, β 指垂直车道线起始点和结束点的偏移量, 每条车道线只在偏置项和上下边界处存在差别。

2 LSCoMer 算法

2.1 模型介绍

LSCoMer 模型整体架构如图 2 所示, 本文模型运用了 Vit-CoMer^[17] 有效解决特征尺度单一和 Transformer 结构对局部信息捕捉不敏感两类问题的算法思想, 在保留 LSTR 整体框架的同时, 新增了 Vit-CoMer 骨干中的 Patch Embedding 模块、MRFP 模块、CTI 模块, 以解决车道线检测任务对不同尺度特征的精确提取和有效融合的技术难点。首先, Patch Embedding 分支将图像压缩至原始分辨率的 1/16, 提升了特征表达能力且便于后续模块的处理。同时, 原始图像进入特征提取骨干分支得到具有不同分辨率的特征金字塔, 各分支产生的特征将经过两段 CoMerBlock 进行特征交互。CoMerBlock 作为特征信息处理阶段, 集成了 VitCoMer 中的 MRFP 和 CTI 模块, 以及 LSTR 中的 Transformer 架构。在每个阶段, MRFP 模

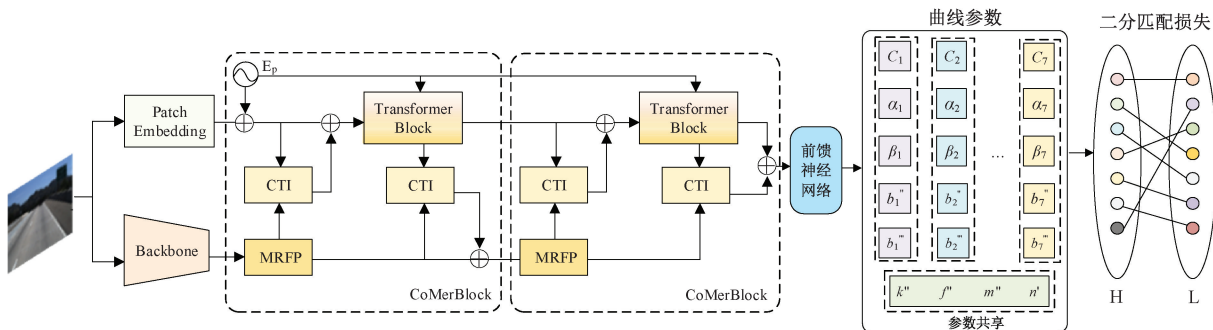


图 2 LSCoMer 模型架构图

Fig. 2 Architecture diagram of the LSCoMer model

块先对特征金字塔进行增强, 然后通过首尾 CTI 模块与 Transformer 特征进行双向交互, 从而获得具有丰富语义信息的多尺度特征。最终, 展平后特征进入参数预测模块拟合车道线。

2.2 多感受野特征金字塔

MRFP 模块由特征金字塔、线性层和多感受野深度可分离卷积层(MRC)组成。特征金字塔能够提供多尺度的视觉特征帮助模型更好地适应不同宽度和不同远近的车道线, 而 MRC 由卷积核尺寸多样的深度可分离卷积层构

成, 这种设计允许网络通过不同尺寸的卷积核捕获不同尺度的空间信息, 从而感知到小范围的细节和大范围的上下文信息。

如图 3 所示, MRFP 的输入是特征提取骨干网络生成的 3 组不同尺度的特征金字塔 C_1 、 C_2 、 C_3 , 这些特征首先被展平并顺序拼接成一个 tokens 序列。

序列经过线性层降维处理后, 根据通道数被分成两组, 分别进入具有 3×3 和 5×5 卷积核大小的深度可分离卷积层(DWConv)进行处理。处理得到的特征序列经过线

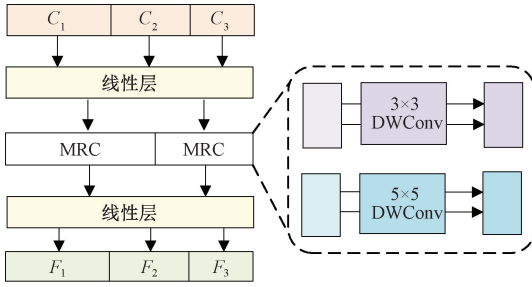


图 3 MRFP 模块结构图

Fig. 3 MRFP module structure

性层恢复到特征原始维度,最终输出特征序列 F_1 、 F_2 、 F_3 。这一过程有效解决了 Transformer 结构提取特征尺度单一的问题,极大地提升了模型对复杂场景的感知能力。整个过程如式(6)所示。

$$F = FC(DWConv(FC(C))) \quad (6)$$

其中, C 为输入特征, FC 为线性层, $DWConv$ 为深度可分离卷积, F 为 MRFP 模块输出特征。

2.3 CNN-Transformer 双向交互模块

为解决 Transformer 结构在捕获局部细节特征上的不足,在 Transformer 的首部和尾部增加 CTI 模块,这一策略保留了 Transformer 的整体结构并增加了 CNN 和 Transformer 之间的信息交互,进一步增强了模型对车道线细节特征的关注。

如图 4 所示,在 CTI 模块中,Transformer 生成的特征 $X \in R^{\frac{H}{16} \times \frac{W}{16} \times D}$ 与 MRFP 模块输出的特征序列 F 进行融合,其中, $F = \{F_1, F_2, F_3\}$, $F \in R^{(\frac{HW}{8^2} + \frac{HW}{16^2} + \frac{HW}{32^2}) \times D}$ 。考虑到特征信息尺度之间的差异,将具有相同尺度的特征 $F_2 \in R^{\frac{H}{16} \times \frac{W}{16} \times D}$ 与 $X \in R^{\frac{H}{16} \times \frac{W}{16} \times D}$ 进行融合得到 F'_2 ,则融合后特征序列表示为 $F' = \{F_1, F'_2, F_3\}$ 。

由于 CNN 和 Transformer 两种架构在设计原理和核心机制上有显著的不同,导致它们在处理数据时的表现形

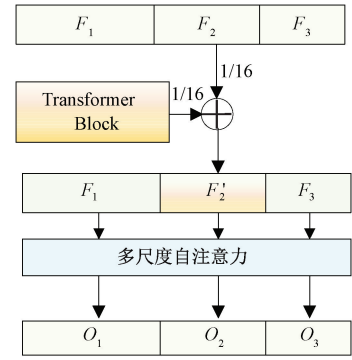


图 4 CTI 模块结构图

Fig. 4 CTI module structure

态也存在差异,如 CNN 通过其层级结构逐步实现从低频到高频信息的提取,而 Transformer 则更有效地处理需要强全局语义理解的任务。为统一数据形态,采用多尺度自注意力机制处理融合后的特征集合 F' ,具体而言,多尺度可变形自注意力^[18]对不同层级的特征图进行分析和整合,接着通过前馈神经网络和规范化层进一步处理,以确保数据的稳定性和模型的泛化能力。整个过程如式(7)所示。

$$O = FFN(Attention(norm(F'))) \quad (7)$$

其中, F' 为分辨率为 $1/8$ 、 $1/16$ 、 $1/32$ 的多尺度特征, $norm(\cdot)$ 为归一化层, $Attention(\cdot)$ 为可变形注意力机制, $FFN(\cdot)$ 为前馈神经网络。

3 实 验

3.1 数据集及评价指标

本文在 TuSimple 和 CULane^[14]数据集上评估 LSCoMer 模型在实际场景中的表现。TuSimple 数据集包含 6 408 张 $1\,280 \times 720$ 的高分辨率公路驾驶图像,CULane 数据集则包含了 133 235 张分辨率为 $1\,640 \times 590$ 且带有车道线标注的图像,涵盖了多种复杂交通环境,检测难度更大。有关数据集详细信息如表 1 所示。

表 1 TuSimple 和 CULane 数据集详细信息

Table 1 TuSimple and CULane dataset details

数据集	帧数	训练集	验证集	测试集	分辨率	场景种类	最大车道数
TuSimple	6 408	3 626	358	2 782	$1\,280 \times 720$	1	5
CULane	133 235	88 880	9 675	34 680	$1\,640 \times 590$	9	4

TuSimple 数据集提供准确率(accuracy, Acc)、假阳性率(false positive, FP)、假阴性率(false negative, FN)作为模型性能的评价指标。计算公式如下:

$$\begin{cases} Acc = \frac{\sum_{clip} C_{clip}}{\sum_{clip} S_{clip}} \\ FP = \frac{F_{pred}}{N_{pred}} \\ FN = \frac{M_{pred}}{N_{gt}} \end{cases} \quad (8)$$

其中, C_{clip} 表示预测正确的车道点的数量, S_{clip} 表示

标注的车道真实点数量。 F_{pred} 表示负样本被预测为正样本的数量, N_{pred} 表示所有预测为正样本的数量。 M_{pred} 表示正样本被预测为负样本的数量, N_{gt} 表示全部正样本的数量。

在 CULane 数据集中,采用 F_1 分数作为评价指标,它指的是召回率(recall)和精确率(precision)的调和平均值。本文中,当车道线真实像素值与预测值的重合部分(IoU)大于 0.5 时,则预测值为真阳性(TP),否则为假阳性(FP)。此外,未被标注的真实车道线像素点预测值为假阴性(FN)。 F_1 分数计算公式如下:

$$\begin{cases} Precision = \frac{TP}{TP + FP} \\ Recall = \frac{TP}{TP + FN} \\ F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \end{cases} \quad (9)$$

为提高模型泛化能力,本文采用数据增强方法提升训练数据的多样性,数据增强方法包括:水平翻转、随机旋转、水平平移、随机裁剪以及颜色抖动。

除上述评价指标外,增加 FPS 及模型运算操作次数(MACs)为评价指标评估模型性能。

3.2 实验环境及参数设置

本文实验采用处理器为: Intel(R) Xeon(R) Platinum 8358P CPU@2.60 GHz, GPU 为: NVIDIA GeForce RTX 3090, 操作系统为: Ubuntu18.04, CUDA11.0。超参数设置时,采用 Adam 自适应算法优化模型,学习率设置为 0.0001, 批次大小为 16, 总迭代次数为 600 000 次, 每迭代 100 次时输出损失值。图 5 展示了训练过程中二分匹配损失函数变化过程, 当模型迭代次数不断增加, 损失值呈现下降趋势直至模型达到收敛效果。

表 2 不同方法在 TuSimple 数据集上性能对比

方法	Acc/%	FP/%	FN/%	FPS	MACs/G	无后处理过程
PolyLaneNet ^[19]	93.36	2.91	3.38	115	1.784	✓
PINet ^[20]	96.70	2.94	2.63	30	—	×
LaneATT(ResNet-34) ^[10]	95.57	3.76	3.10	171	—	×
LaneATT(ResNet-122) ^[10]	96.12	5.47	2.26	26	—	×
ENet-SAD ^[21]	96.64	6.02	2.05	75	—	×
LSTR ^[13]	96.18	2.91	3.38	420	0.574	✓
LSTR 改进方法 ^[16]	96.31	2.76	3.22	420	—	✓
LSCoMer(本文)	96.68	2.76	3.20	180	0.582	✓

图 6 展示了 LSCoMer 与 PolyLaneNet 的可视化对比, 相比 PolyLaneNet, 本文提出的 LSCoMer 模型表现出更高的准确性、连续性和稳定性。在具有直路、弯道和有阴影的检测场景下, LSCoMer 检测到的车道线更加平滑且无中断, 对细节的处理也更加精细, 如在阴影和复杂路况下仍能准确检测车道线。而 PolyLaneNet 在复杂场景中容易

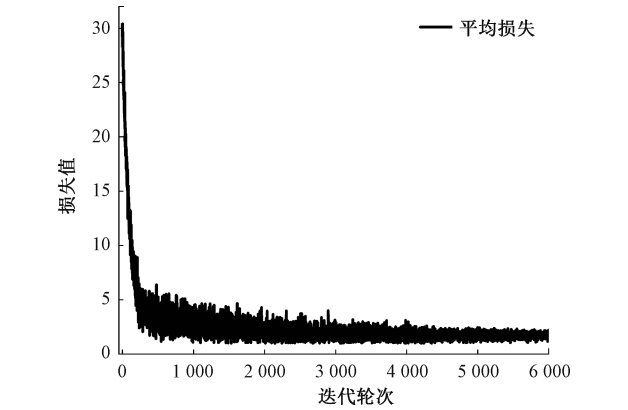


图 5 平均损失函数曲线图
Fig. 5 Average loss function graph

3.3 模型对比试验

为评估 LSCoMer 模型性能, 本文在 TuSimple 数据集上与车道线检测领域中具有代表性的先进方法进行了对比实验, 实验结果如表 2 所示。本文方法相比使用分割的 SCNN 方法, 不仅精度得到提升, 且减少了复杂的后处理步骤, 在模型推理速度上表现更好。对于基于锚框的代表性方法 UFLD 模型, 尽管预定义锚框的策略能提高模型检测速度, 但检测精度受制于锚框的固定形状, 反而有所下降。本文方法弥补了 LSTR 模型中 Transformer 结构对局部特征捕捉不佳的不足, 检测精度相比 LSTR 提升了 0.5%, 相比同类以参数回归车道线位置的 PolyLaneNet 算法相比, 准确度提升了 3.32%, 模型运算操作次数仅为其 1/3, 误检率和漏检率均有所提升, 相比 SCNN 等其他类型车道检测算法, 本文模型的各项指标表现优异。

出现检测错误、车道线偏离或中断的现象, 细节处理不够精细。因此, LSCoMer 比 PolyLaneNet 在实际应用中提供了更可靠和精确的车道线检测结果。

为评估本文模型在复杂场景下的检测能力, 将模型在 CULane 数据集上与其他方法进行对比, 结果如表 3 所示。



图 6 LSCoMer 与 PolyLaneNet 可视化对比

Fig. 6 Comparison of LSCoMer and PolyLaneNet visualization

表 3 不同方法在 CULane 数据集上对比

Table 3 Comparison of different methods on CULane dataset

方法	F_1 分数/%
SALMNet ^[22]	63.20
LSTR ^[13]	65.19
LSTR 改进方法 ^[16]	67.30
LSCoMer(本文)	68.21

为改善 LSTR 模型在提取特征尺度单一和缺乏内部信息交互的问题,本文模型集成两阶段 CoMerBlock,提升了车

道线检测模型在复杂城市道路上的检测能力, F_1 得分较 LSTR 提升了 3.02%。相比使用注意力模块缓解 Transformer 内部缺乏局部信息的 LSTR 改进方法^[16],本文采用了多尺度特征金字塔增强特征并结合 CNN-Transformer 双向融合交互特征信息的策略,效果更为显著。

3.4 消融实验

为验证 LSCoMer 模型中各组件在车道线检测实际应用中的有效性,本文在多车道场景的 TuSimple 数据集及复杂场景的 CULane 数据集上分别进行消融实验,基准模型选择 LSTR 模型,实验数据如表 4 所示。

表 4 消融实验结果

Table 4 Results of ablation experiments

Baseline	MRFP	CTI	CoMerBlock×2	CoMerBlock×4	Acc/%	F_1 /%
					TuSimple	CULane
✓	—	—	—	—	96.13	65.19
✓	✓	—	—	—	96.26	66.37
✓	—	✓	—	—	96.30	65.89
✓	✓	✓	—	—	96.47	67.65
✓	—	—	✓	—	96.68	68.21
✓	—	—	—	✓	96.68	68.23

为更清晰对比各模块效果,如图 7 所示,本文将消融实验结果可视化,包括每条车道的拟合效果及对应热力图。此外,在消融实验中还探究了不同数量的 CoMerBlock 对模型效果的影响。实验结果表明 MRFP 模块对特征信息的增强效果明显,其主要依赖于不同尺度的特征组及不同卷积核大小的深度可分离卷积,有效缓解了 LSTR 模型特征提取尺度单一的局限性。由于 Transformer 结构在捕捉局部信息能力上的不足,CTI 模块的增加促进了 CNN 和 Transformer 结构间多尺度信息的交互,检测精度和 F_1 分数分别提升了 0.17% 和 0.70%。此外,Transformer 结构仅在单尺度特征上使用自注意力,CoMerBlock(MRFP+

CTI)通过增强的多尺度特征信息在 CNN 和 Transformer 间流动,提升了模型的远程建模能力。从图 7 展示的热力图来看,MRFP 和 CTI 模块的增加有效提升了检测模型对车道线的关注。

实验对比了集成单个 CoMerBlock、CoMerBlock×2 和 CoMerBlock×4 的模型表现。结果表明,CoMerBlock×2 效果明显好于单一 CoMerBlock 模块,尽管 CoMerBlock×4 集成到模型中相比 CoMerBlock×2 有微弱效果提升,但模型参数量剧增。从车道线检测任务对场景实时性和模型轻量化的需求考虑,本文选择 CoMerBlock×2 是最优解。

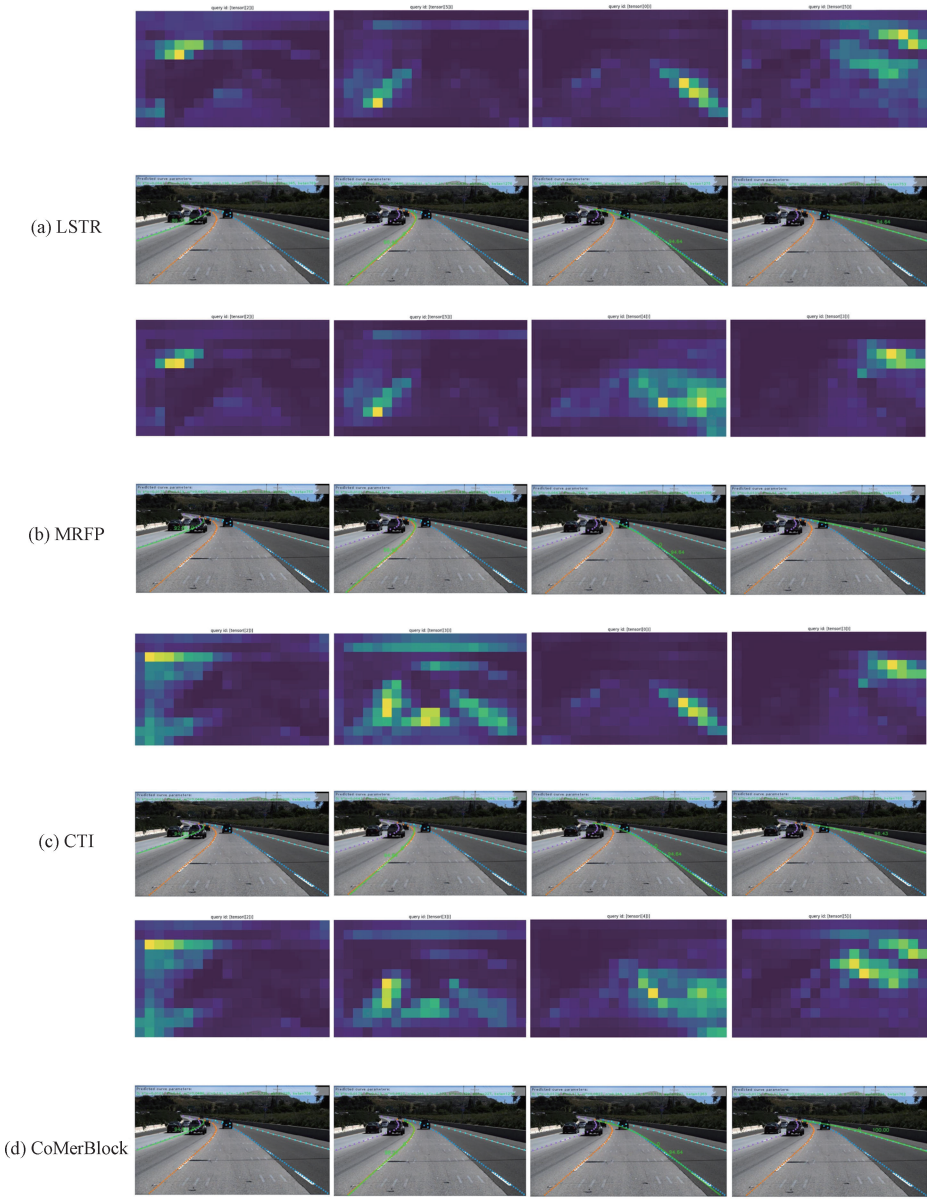


图 7 消融实验效果图

Fig. 7 Ablation experiment effect diagram

4 结 论

为解决 LSTR 车道检测模型中 Transformer 结构存在的特征提取尺度单一、局部信息提取能力不足的问题,本文提出一种具有 MRFP 模块和 CTI 模块的高效车道线检测模型 LSCoMer。MRFP 模块利用不同卷积核大小的深度可分离卷积层扩展感受野,捕捉多尺度特征,而 CTI 模块通过逐元素相加和多尺度自注意力,融合 CNN 和 Transformer 的特征,增强了模型的长程建模能力和局部信息捕捉能力。实验验证了本文模型的有效性,与使用参数回归的同类检测模型 PolyLaneNet 相比,准确度提升了 3.32%,模型运算次数减少 68%,与 LSTR 及其他类型车

道检测方法相比,本文模型各项指标均表现较好。未来的工作将继续提升模型在不同场景下的检测精度。

参考文献

[1] 张新钰, 高洪波, 赵建辉, 等. 基于深度学习的自动驾驶技术综述[J]. 清华大学学报(自然科学版), 2018, 58(4): 438-444.
ZHANG X Y, GAO H B, ZHAO J H, et al. Overview of deep learning intelligent driving methods[J]. Journal of Tsinghua University(Science and Technology), 2018, 58(4): 438-444.

[2] 宋绍京, 陆婷婷, 孙翔, 等. 面向自动驾驶的多任务环境感知算法[J]. 电子测量技术, 2023, 46(24): 157-163.

- SONG SH J, LU T T, SUN X, et al. Multi-task environment sensing algorithms for autonomous driving [J]. *Electronic Measurement Technology*, 2023, 46(24): 157-163.
- [3] KE H M. Research on lane line detection and tracking under complex road conditions [J]. *Frontiers in Science and Engineering*, 2021, 1(3): 162-165.
- [4] LI Y D, CHEN L G, HUANG H B, et al. Nighttime lane markings recognition based on Canny detection and Hough transform [C]. 2016 IEEE International Conference on Real-time Computing and Robotics (RCAR), 2016.
- [5] HUANG Y P, LI Y W, HU X, et al. Lane detection based on inverse perspective transformation and Kalman filter [J]. *KSII Transactions on Internet and Information Systems (TIIS)*, 2018, 12(2): 643-661.
- [6] SHIN B S, TAO J L, KLETTE R. A superparticle filter for lane detection [J]. *Pattern Recognition*, 2015, 48(11): 3333-3345.
- [7] NIU J W, LU J, XU M L, et al. Robust lane detection using two-stage feature extraction with curve fitting [J]. *Pattern Recognition*, 2016, 59: 225-233.
- [8] WANG J SH, MA Y CH, HUANG SH F, et al. A keypoint-based global association network for lane detection [C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [9] QIN Z Q, WANG H Y, LI X. Ultra fast structure-aware deep lane detection [C]. *Computer Vision-ECCV 2020: 16th European Conference*, 2020.
- [10] TABELINI L, BERRIEL R, PAIXAO T M, et al. Keep your eyes on the lane: Real-time attention-guided lane detection [C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [11] PAN X G, SHI J P, LUO P, et al. Spatial as deep: Spatial CNN for traffic scene understanding [C]. *AAAI Conference on Artificial Intelligence*, 2018.
- [12] LIU L ZH, CHEN X H, ZHU S Y, et al. CondLaneNet: A top-to-down lane detection framework based on conditional convolution [C]. *IEEE/CVF International Conference on Computer Vision (CVPR)*, 2021.
- [13] NEVEN D, BRABANDERE B D, GEORGOULIS S, et al. Towards end-to-end lane detection: An instance segmentation approach [C]. 2018 IEEE Intelligent Vehicles Symposium (IV), 2018.
- [14] LIU R J, YUAN Z J, LIU T, et al. End-to-end lane shape prediction with transformers [C]. *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2021.
- [15] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. *ArXiv preprint arXiv: 1706.03762*, 2017.
- [16] 张莹, 张露露, 孙月, 等. LSTR 算法的改进及在车道线检测中的应用 [J]. *小型微型计算机系统*, 2024, 45(8): 1863-1868.
- ZHANG Y, ZHANG L L, SUN Y, et al. Improvement of LSTR algorithm and its application in lane detection [J]. *Journal of Chinese Computer Systems*, 2024, 45(8): 1863-1868.
- [17] XIA CH L, WANG X L, LYU F, et al. Vit-CoMer: Vision transformer with convolutional multi-scale feature interaction for dense predictions [J]. *ArXiv preprint arXiv: 2403.07392*, 2024.
- [18] XIA ZH F, PAN X R, SONG SH J, et al. Vision transformer with deformable attention [C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [19] TABELINI L, BERRIEL R, PAIXÃO T M, et al. PolyLaneNet: Lane estimation via deep polynomial regression [C]. 2020 25th International Conference on Pattern Recognition (ICPR), 2020.
- [20] KO Y, LEE Y, AZAM S, et al. Key points estimation and point instance segmentation approach for lane detection [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 23(7): 8949-8958.
- [21] HOU Y N, MA ZH, LIU CH X, et al. Learning lightweight lane detection cnns by self attention distillation [C]. *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.
- [22] XU X M, YU T F, HU X W, et al. SALMNet: A structure-aware lane marking detection network [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 22(8): 4986-4997.

作者简介

郑剑, 博士, 副教授, 硕士生导师, 主要研究方向为机器视觉。

E-mail: zhengjian@jxust.edu.cn

郭亦畅(通信作者), 硕士研究生, 主要研究方向为计算机视觉、车道线识别。

E-mail: guoyichang12@126.com