

DOI:10.19651/j.cnki.emt.2519485

基于 Transformer 和 XCA 注意力的车道线检测方法^{*}

李睿 敖银辉 陈新盛 李桂伸

(广东工业大学机电工程学院 广州 510000)

摘要: 针对 LSTR 算法在复杂道路场景下存在的局部细节捕捉能力不足与计算复杂度高的问题,本研究提出动态多路径协方差 Transformer 检测模型 DMCTR。首先,构建特征增强与抑制模块,通过增强与抑制特征操作,缓解传统卷积对弯曲车道及虚线段等弱特征的漏检问题;其次,构建动态增强双通道注意力模块,利用可变形卷积适应车道几何形变,并结合双注意力机制增强局部几何特征;最后在 Transformer 架构中引入交叉协方差注意力机制,进而替换 Transformer 编码器中的多头自注意力。实验结果表明,在 TuSimple 数据集上,DMCTR 准确率达到 96.74%,较基线 LSTR 模型提升 0.56%;在 CULane 数据集复杂场景下,F1 值提升 4.48%,车道线模糊、夜晚强光等特殊场景检测精度提升显著,使得模型在保持实时性(353 fps)的同时,有效解决了传统方法在复杂场景下的特征建模瓶颈。

关键词: 车道线检测;LSTR 算法;Transformer;交叉协方差注意力机制

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 line detection method based on Transformer and XCA attention

Li Rui Ao Yinhui Chen Xinsheng Li Guishen

(School of Mechanical and Electrical Engineering, Guangdong University of Technology, Guangzhou 510000, China)

Abstract: Aiming at the problems of insufficient local detail capture ability and high computational complexity of the LSTR algorithm in complex road scenarios, this paper proposes a dynamic multi-path covariance Transformer detection model DMCTR. Firstly, a feature boosting and suppression module is constructed. Through feature enhancement and suppression operations, the problem of missed detection of weak features such as curved lanes and dashed line segments in traditional convolution is alleviated. Secondly, construct a dynamic enhancement dual-path aggregation block, utilize deformable convolution to adapt to lane geometric deformation, and combine the dual-attention mechanism to enhance local geometric features. Finally, the cross-covariance attention is introduced into the Transformer architecture to replace the multi-head self-attention in the Transformer encoder. The experimental results show that on the TuSimple dataset, the accuracy rate of the method proposed reaches 96.74%, which is 0.56% higher than that of the baseline LSTR model. In the complex scenarios of the CULane dataset, the F1 has increased by 4.48%, and the detection accuracy in special scenarios such as blurred lane lines and strong light at night has been significantly improved. This enables the model to maintain real-time performance (353 fps) while effectively solving the feature modeling bottleneck of traditional methods in complex scenarios.

Keywords: lane line detection;LSTR algorithm;Transformer;cross-covariance attention mechanism

0 引言

智能驾驶系统^[1]作为汽车智能化转型的核心技术,近年来在环境感知与决策规划领域取得显著突破。其中,车道线检测作为自动驾驶视觉感知的基础环节,直接影响着车道保持、自适应巡航等关键功能的可靠性。近年来,基于深度学习的检测方法逐渐成为主流,其技术路线主要可分

为基于检测的范式和基于分割的范式两类。Pan 等^[2]提出 SCNN 网络,其采用自上而下的方法,将每个车道视为独立类别,执行多类别语义分割,并通过并行分类分支来预测车道的存在,最终结合分类与分割结果得到车道实例。这虽利于实例区分,但需预定义最大车道数,且车辆换道时易出现标注歧义;Neven 等^[3]提出 LaneNet,通过实例嵌入对分割结果聚类,实现车道实例分割,聚类精度高但存在耗时严

收稿日期:2025-07-29

* 基金项目:广东省自然科学基金(2022A1515012080)项目资助

重的问题,不利于实时处理。Tabelini 等^[4]提出线路锚点特征池化方法 LaneATT,使用轻量级主干网络,结合注意力聚合机制,提升模糊车道检测性能。UFLD V2^[5]将逐行分类扩展为逐行/列分类,解决水平车道检测问题,但其多分类策略在行和列分类选择上过于简单,限制了模型的泛化能力。随着 Transformer 在目标检测中的广泛应用,该领域的研究逐渐从 YOLO、Faster R-CNN^[6]等密集预测范式转向 DETR 等设置预测范式。Liu 等^[7]将 Transformer 架构引入车道线检测任务中,从而提出了 LSTR (lane shape prediction with transformers)检测模型。该网络采用 Transformer 架构以端到端的方式直接输出车道线形状参数,极大地简化了处理流程,增强了模型的可解释性,并实现了 420 fps 的高检测速度。张莹等^[8]在 LSTR 中引入了 CBAM 模块和掩码单头注意力机制,试图通过通道和空间信息的融合来捕捉更多细节。然而,Transformer 架构本质上更擅长处理全局特征,仅依靠 CBAM 模块的增强作用难以充分提升局部特征的表现,而且多头注意力机制的剪枝可能引发信息丢失,从而降低检测精度。郑剑等^[9]通过在 LSTR 基础上引入多感受野特征金字塔模块和 CNN-Transformer 双向交互模块,解决了特征尺度单一和局部信息捕捉不足的问题。但因其仍采用传统卷积神经网络 (convolutional neural networks, CNN),故仍对于弯曲、虚线段等弱特征的提取能力不足,且模块引入导致计算开销增大。综上如文献[8-9]所提到,传统 CNN 存在感受野的形状单一固定及感受范围较小的局限性,使用单一 CNN 形式会导致捕捉车道线的连续性特征与不规则特征的能力不足;且全局 Transformer^[10]因计算复杂度高导致实时性下降,且对局部细节响应不足,使其在复杂道路场景(如弯道、模糊等)下仍面临显著挑战。

本研究在 LSTR 模型的基础上,提出动态多路径协方差 Transformer 检测模型 DMCTR (dynamic multipath covariance transformer, DMCTR),旨在解决现有方法受限于传统 CNN 感受野的固定形状与较小感受范围,致使车

道线连续性特征与不规则特征难以被有效捕捉的问题。主要贡献包括:

1) 构建特征增强与抑制模块 (feature boosting and suppression module, FBSM),通过抑制显著区域并迫使网络挖掘潜在在细节特征,缓解了传统卷积对弯曲车道、虚线段等弱特征及长距离上下文关联性特征的漏检问题。

2) 设计动态增强双通道注意力模块 (dynamic enhancement dual-path aggregation block, DEDA-Block),该模块创新性地可将可变形卷积 (deformable convolutional networks, DCN) 与双注意力机制 (dual attention block, DA-block) 相结合。DCN 通过自适应调整感受野以贴合车道几何形变,而 DA-block 则协同增强局部几何特征的表达能,从而有效克服了单一卷积形式在复杂场景下的特征提取局限。

3) 引入交叉协方差注意力机制 (cross-covariance attention, XCA) 替换标准多头自注意力 (multi-head attention, MSA),在显著降低计算复杂度的同时,增强全局空间结构的建模能力。

1 LSTR 算法

1.1 算法原理结构与改进策略对比

LSTR 是一种基于 Transformer 的端到端车道线检测模型,其核心在于直接输出参数化的车道形状方程,而非传统方法中基于逐像素分割或锚框检测的离散表征。如图 1 所示, LSTR 模型的整体架构包括 Resnet 网络^[11]、Transformer 架构、前馈神经网络、匈牙利损失函数。ResNet18 作为特征提取网络从输入图像提取出底层视觉特征,将其展平为一维序列 S 。输入 Transformer 模块中,位置编码 E_p 被嵌入输入特征以编码空位置信息。编码器输出的特征序列 S_e 与隐式学习位置差异的学习位置嵌入 E_{LL} 进行耦合后输入至 Transformer 解码器。解码器以可学习车道查询向量 S_q 作为初始化条件,通过多头注意力机制与 S_e 进行层次化特征交互,交互过程生成的车道表征序列 S_d 经前馈神经网络解码为参数化车道方程。

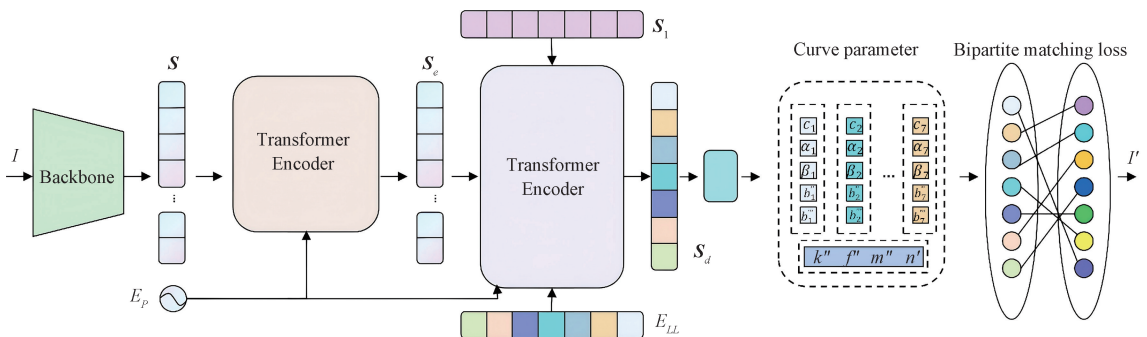


图 1 LSTR 算法原理

Fig. 1 Schematic diagram of the LSTR algorithm

然而, LSTR 模型仍存在复杂场景下的局部细节捕捉能力不足与计算复杂度高的问题。为此一系列基于 LSTR

的改进方法被提出,如表 1 所示。文献[8]中所提到的 LSTR 改进方法,侧重于局部增强与结构裁剪,但仅依赖

CBAM 模块难以充分捕捉局部特征,且对注意力头的剪枝会导致信息丢失,从而限制检测精度。文献[9]中所提到的 LSCoMer 则转向多尺度特征融合,但其核心仍基于传统 CNN,对弱特征的提取能力有限,且新增模块带来了额

外的计算开销。本研究提出的 DMCTR 则采用弱特征挖掘、形变自适应注意力、优化全局建模的策略,系统性地解决了弱特征漏检、形变适应性差和计算成本高等核心挑战。

表 1 改进 LSTR 算法的策略对比

Table 1 Comparison of strategies for improved LSTR algorithms

方法	改进策略	解决的问题
LSTR 改进 ^[8]	局部增强:在编码器末端引入注意力模块; 复杂度裁剪:减少解码器注意力头数	局部特征捕捉能力弱;部分计算冗余
LSCoMer ^[9]	多尺度融合:构建特征金字塔; 双向架构交互:引入双向交互模块	特征尺度单一; 全局模型缺乏局部敏感性
DMCTR	弱特征挖掘:抑制显著特征以关注弱目标; 形变自适应注意力:结合可变形卷积与双注意力; 优化全局建模:引入 XCAAttention 替换标准自注意力	弯曲/虚线等弱特征漏检; 车道几何形变适应性差; MSA 计算复杂度高

1.2 LSTR 参数化表达

LSTR 采用一种基于多项式的车道形状模型,将车道线表示为道路上的三次多项式曲线来近似单车车道线:

$$X = kZ^3 + mZ^2 + nZ + b \tag{1}$$

其中, k, m, n, b 是模型参数, (X, Z) 表示地面的点。

在相机成像平面上,该曲线投影为:

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

其中, k', m', n', b' 是参数与相机内参和外参的组合, (u, v) 是图像平面上的像素点,对于倾斜相机,曲线进一步转换为:

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

其中, ϕ 是相机的俯仰角, f 是焦距。通过结合参数与俯仰角,最终得到倾斜相机平面上的曲线形式。

2 DMCTR

2.1 DMCTR 网络架构介绍

如图 2 所示为 DMCTR 网络架构。在特征提取阶段,对 ResNet18 骨干网络进行多层级特征增强。在 MaxPool 层及 Layer1/2 输出端嵌入 FBSM, FBSM 可实现显著区域抑制,迫使网络关注弯曲、虚线段等易忽略细节。Layer2 输出的 32 通道中级特征经 FBSM 二次处理,进一步挖掘车道线多尺度纹理特征。在深层特征处理阶段,于 Layer4 输出端部署 DEDA-Block, DEDA-Block 为后续 Transformer 模块的特征建模任务挖掘出更全面的特征信息。

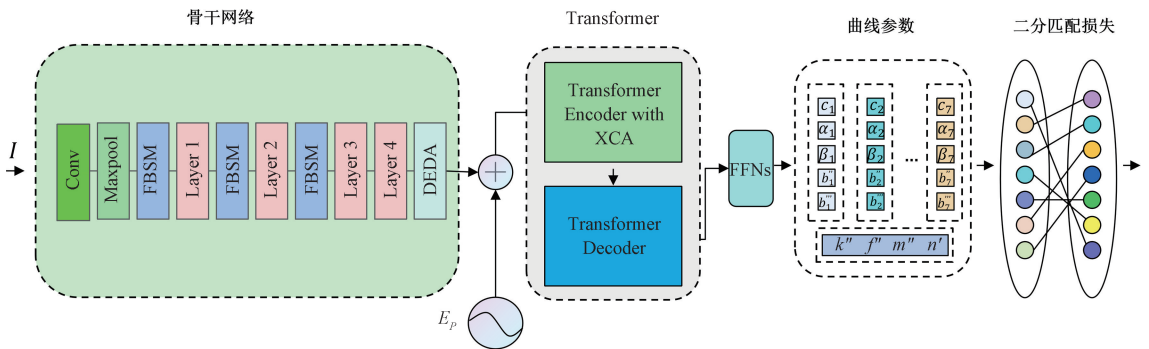


图 2 DMCTR 网络架构

Fig. 2 DMCTR network architecture

随后,特征将被展平为大小为 $HW \times C$ 的序列 S ,其中 HW 表示序列的长度, C 是通道的数量。在全局特征建模阶段,针对 Transformer 编码器中多头自注意力计算复杂度高、局部特征捕捉不足的问题,将 Transformer 中的 MSA 替换为 XCA 模块,在降低计算复杂度的同时,增强

对车道线的空间结构建模能力。最终,展平后特征进入参数预测模块拟合车道线。

2.2 FBSM 模块

传统卷积对显著特征能够较好的捕捉,而会忽略一些细节特征。具体而言,CNN 的卷积层通过滑动窗口的方式

对输入图像进行特征提取,其权重共享机制使得网络能够快速识别出图像中的显著特征。这些显著特征通常具有较高的信息量和较强的视觉辨识度,因此在训练过程中,网络容易优先学习这些显著特征。然而,这种学习方式也使得网络对一些细节特征的关注度不足。细节特征往往包含了丰富的语义信息和局部结构信息,如实际行驶路面中会存在弯曲、虚线、模糊车道,而这些车道线可能会对预测车道线接下来的走向起到至关重要的作用。为弥补传统卷积对潜在特征提取不足的缺点,故首次将应用于细粒度视觉分类(FGVC)领域的FBSM模块^[12]引入车道线检测任务。该模块能够显式地迫使网络关注多个不同的车道线特征部分。如图3所示,FBSM模块的输入是backbone输出的特征 $X \in \mathbb{R}^{C \times W \times H}$,其中 C 、 W 和 H 分别表示通道数、宽度和高度。将特征映射 X 沿着宽度方向均匀划分为 k 个部分。经过划分之后,输入一个 1×1 卷积操作来评估每一个部分的重要程度:

$$A^i = \text{Relu}(\text{Conv}(\mathbf{X}^{(i)})) \in \mathbb{R}^{1 \times \frac{W}{k} \times H} \quad (4)$$

其中,Relu函数是非线性激活函数,应用非线性函数Relu去除负激活。通过全局平均池化(GAP)计算每个部分的重要性因子 b'_i 。

$$b_i = \text{GAP}(A^{(i)}) \quad (5)$$

将重要性因子 b_i 进行归一化操作,可得 b'_i :

$$b'_i = \frac{\exp(b_i)}{\sum_{j=1}^k \exp(b_j)} \quad (6)$$

基于归一化后的重要性因子 $\mathbf{B} = (b'_1, \dots, b'_k)^T$,经计算可以得到显著特征部分。通过抑制操作,可以获得抑制特征 \mathbf{X}_s ,其中 $\mathbf{S} = (S_1, \dots, S_k)$, β 为超参数,用来控制特征抑制的程度。

$$S_i = \begin{cases} 1 - \beta, & b'_i = \max(\mathbf{B}) \\ 1, & \text{其他} \end{cases} \quad (7)$$

$$\mathbf{X}_s = \mathbf{S} \otimes \mathbf{X} \quad (8)$$

通过抑制操作,可以得到潜在的特征图 \mathbf{X}_s 。

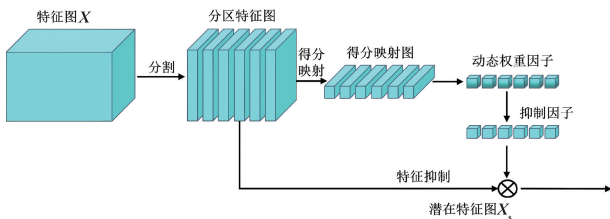


图3 FBSM结构

Fig. 3 Structure of FBSM

通过这种方式,FBSM能够在增强最显著特征的同时,抑制其在后续阶段的影响力,从而迫使网络挖掘其他潜在的特征。

2.3 DEDA-Block

Transformer虽然擅长捕捉全局上下文,但它通常缺

乏对通道维度关系的显式建模。针对这一问题,本研究受到DA-TransUNet^[13]的启发,故特将DA-block引入检测网络。DA-block通过引入通道注意力模块,补足这一短板,实现了空间与通道双重注意力,从而提取更加丰富和判别性更强的特征。同时,将DCN与DA-block两者相结合,使用深度可分离卷积(depthwise separable convolution, DSConv)^[14]替换DA-block前端的用于实现通道数调整的普通卷积,从而提出了DEDA-Block。

文献[7-8]所采用的传统CNN存在感受野的形状固定与感受范围较小的局限,而车道线具有快速变化特点且形状非线性,导致传统CNN对车道线所具有的连续性特征与不规则特征捕捉能力不足。而DCN^[15]在卷积和池化操作中引入动态、可学习的偏移量,可实现自适应空间采样,它可以自适应的改变感受野,能更好地获取车道线特征。因此将DCN与DA-block相结合,如图4所示。DCN的偏移量由一个单独的卷积层生成,这些偏移量从目标任务中学习,无需额外的监督信息,增强了卷积核的灵活性,使其能够适应不规则的几何变换。具体而言,DCN可以表示为:

$$\text{Output}(p_n) = \sum_{m=1}^M W(p_m) \cdot I(p_n + p_m + \Delta p_m) \quad (9)$$

$$\Delta p_m = \text{Offset}(\text{Input}) \quad (10)$$

其中, p_n 是输出特征图中的位置, p_m 是卷积核中的位置, Δp_m 是偏移量,由偏移量生成网络Offset生成。 $W(p_m)$ 是卷积核的权重, $I(p_n + p_m + \Delta p_m)$ 是输入特征图在偏移后的位置的值得。DCN模块添加的参数和计算量相对较小,因此不会显著增加模型的复杂度,Offset是一个卷积层,用于生成偏移量。这种偏移量使得DCN模块在处理具有复杂几何变换的图像特征时表现出色,使其在车道线检测中能够更好地捕捉弯曲车道线的特征。

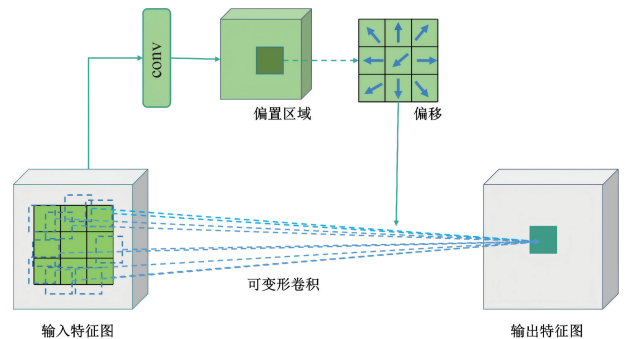


图4 DCN结构

Fig. 4 Structure of DCN

如图5所示,特征图首先经过DSConv模块,DSConv模块由深度卷积和逐点卷积构成。DSConv通过解耦空间滤波与通道融合,将输入通道数减少到原通道数的1/8。

$$\mathbf{X}_{\text{reduced}} = \text{DSConv}(\mathbf{X}_{\text{input}}) \quad (11)$$

其中, $\mathbf{X}_{\text{input}}$ 是输入特征图, $\mathbf{X}_{\text{reduced}}$ 是经过深度可分离

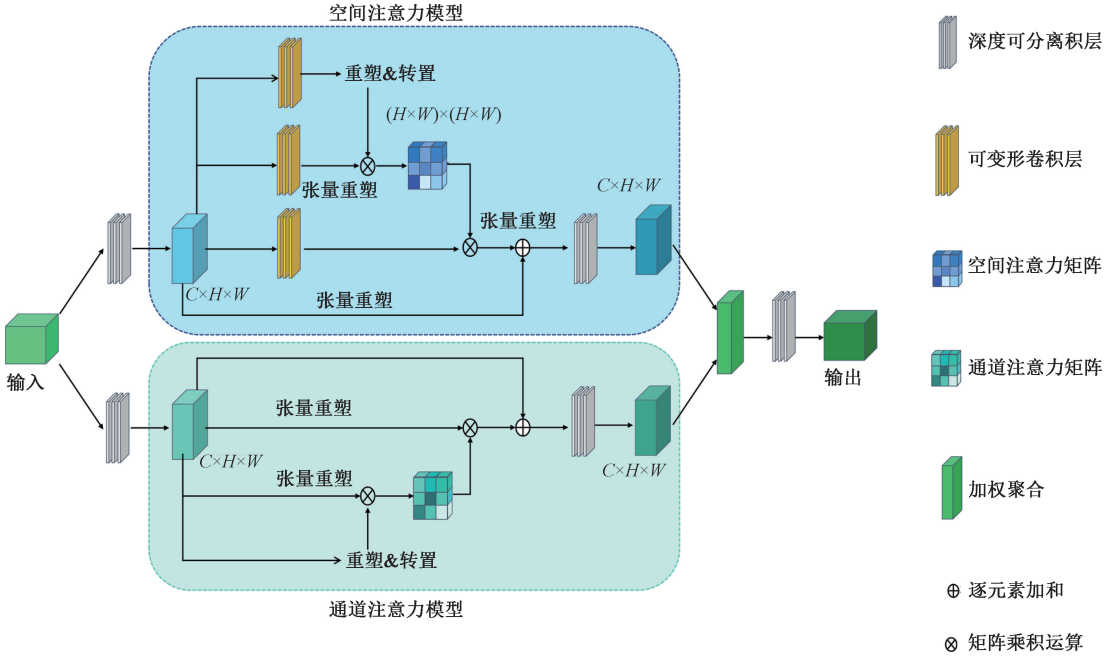


图 5 DEDA-Block 结构
Fig. 5 DEDA-Block structure

卷积处理后的特征图。经通道处理后的特征图分别输入到位置注意力模块 (positional attention module, PAM) 和通道注意力模块 (channel attention module, CAM)。

PAM 捕获特征图中不同位置之间的空间依赖关系, 通过计算位置特征的加权和来更新特定位置的特征。具体而言, 大小为 $C \times H \times W$ 原始特征图进入 PAM。随后, 将特征图输入进 DCN 处理, 获得 3 个特征向量 $PAM_Feature$ 。其中两个特征图重塑为 $C \times N$ 的尺寸, 矩阵相乘后使用 $softmax$ 计算空间注意力权重图 S_{PAM} 。然后将矩阵 $PAM_Feature$ 重塑后和 S_{PAM} 的转置之间执行矩阵乘法, 将结果重塑与原始特征图维度一致。将其与权重值 α 相乘, 并与原始特征图进行矩阵求和运算, 得到最终输出 E_{PAM} 。PAM 的计算过程如式(12)~(15)所示。

$$Offset = Conv_{offset}(X_{reduced}) \quad (12)$$

$$PAM_Feature = DCN(X_{reduced}, Offset) \quad (13)$$

$$S_{PAM} = Softmax(PAM_Feature \cdot PAM_Feature) \quad (14)$$

$$E_{PAM} = \alpha \sum(S_{PAM} \cdot PAM_Feature) + X_{reduced} \quad (15)$$

CAM 则专注于提取通道间的相关性, 通过通道特征的加权和来增强通道特征表达。CAM 与 PAM 的不同之处在于, 其将原始特征直接重塑后在原始特征和它的转置之间进行矩阵乘法。随后, 应用 $softmax$ 层获得通道注意力权重向量 S_{CAM} 。随后 S_{CAM} 的转置和 $CAM_Feature$ 之间矩阵相乘后将结果重塑回初始的输入尺寸。然后将结果乘以权重 β , 并对其以通道为元素的矩阵求和运算, 得到最终输出 E_{CAM} 。CAM 的计算过程如式(16)~(20)所示。

$$CAM_Feature = Reshape(X_{reduced}) \quad (16)$$

$$S_{CAM} = Softmax(CAM_Feature \cdot CAM_Feature) \quad (17)$$

$$E_{CAM} = \beta \sum(S_{CAM}^T \cdot CAM_Feature) + X_{reduced} \quad (18)$$

在 PAM 和 CAM 提取的特征经过进一步处理后, 两个分支的特征被融合并进行通道数恢复, $Output$ 是最终的输出特征图。融合过程如式(19)、(20)所示。

$$Fused_Feature = E_{PAM} + E_{CAM} \quad (19)$$

$$Output = Conv_{restore}(Fused_Feature) \quad (20)$$

2.4 XCA 注意力模块

在标准的 Transformer 模块中, 使用传统自注意力往往带来更高的浮点运算数和延迟, 这在实时性检测任务中将会提出更高的算力要求和更长的响应时间。XCA 模块^[16]在全局特征表达中使用 key 和 query, 消除了传统自注意力在时间和内存方面的二次复杂性^[17]。其线性复杂度为 $O(Nd^2/h)$ 大幅降低了计算量和内存占用, 使其能够高效处理高分辨率图像。另一方面, XCA 模块的交互矩阵大小固定, 与输入图像分辨率无关, 增强了模型对图像分辨率变化的鲁棒性, 使其在训练和测试时对分辨率变化更加灵活。

如图 6 所示, XCA 模块将输入特征形状 $H \times W \times C$ 重塑为 $HW \times C$ 。通过 3 个线性层分别生成 Q (query)、 K (key) 和 V (value)。随后对 Q 和 K 应用 L2 范数进行平滑处理, 以优化权重参数。与传统方法不同, XCA 模块不在空间维度上通过 K^T 和 Q 的运算得到形状为 $HW \times HW$ 的矩阵, 而是在通道维度上计算注意力。具体而言,

对 Q 进行转置得到 Q^T , 进而通过计算可得到形状为 $C \times C$ 的矩阵。经过 $Softmax$ 函数处理后, 得到注意力得分矩阵, 将其与对应的 V 相乘, 即:

$$XC_{Attention}(Q, K, V) = V \times Softmax\left(\frac{K^T Q}{\tau}\right) \quad (21)$$

其中, 每个输出 $token$ 的输入维度是其在 V 中对应的 $token$ 输入的 dv 特征的凸组合; 注意力权重 τ 是基于交叉协方差矩阵计算的。特征通过计算得到注意力图后, 依次通过线性层、归一化层和激活函数, 生成非线性特征。

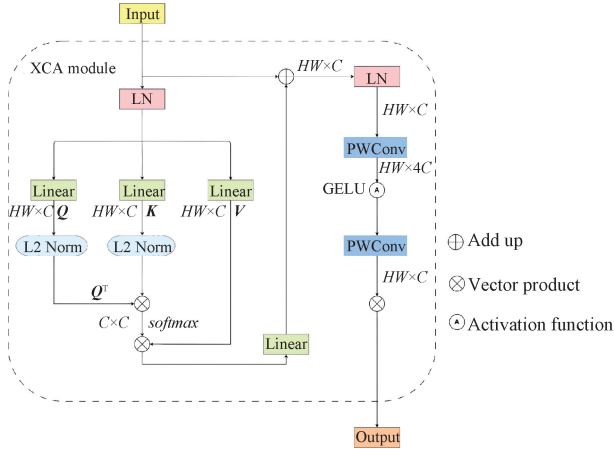


图6 XCA 模块结构

Fig. 6 XCA structure

3 实验设置

3.1 数据集与评价指标

本研究使用 TuSimple 和 CULane 数据集评估模型性能。TuSimple 数据集^[18] 包含 6 408 张图像, 涵盖高速公路场景中的车道线遮挡、残缺或缺失等情况, 其中涉及良好和中等天气。CULane 数据集^[19] 包含 133 235 张分辨率为 $1\ 640 \times 590$ 的图像, 涵盖北京地区的郊区、市区和高速公路场景, 其中分为 9 个场景 (如拥挤、黑夜、缺少可见线、曲线、阴影等), 该数据集适用于复杂路况测试。

针对两个数据集的场景特点与标注方式, 本研究采用了不同的性能评价指标。TuSimple 数据集侧重于车道线点的精确匹配, 因此选用准确率 (accuracy, Acc)、假阳性率 (false positive, FP)、假阴性率 (false negative, FN) 3 个指标作为性能的评价指标。计算公式如式(22)所示。

$$\left\{ \begin{aligned} Acc &= \frac{\sum_{clip} C_{clip}}{\sum_{clip} S_{clip}} \\ FP &= \frac{F_{pred}}{N_{pred}} \\ FN &= \frac{M_{pred}}{N_{gt}} \end{aligned} \right. \quad (22)$$

其中, C_{clip} 和 S_{clip} 分别表示预测正确的车道点数量和

标注真实的车道点数量; F_{pred} 表示预测错误的车道数量; N_{pred} 表示预测到的所有车道数量; M_{pred} 表示没有预测到的真实车道数量; N_{gt} 是标注已有的车道数量。

CULane 数据集注重车道实例的完整性检测, 因此采用计算预测车道线和真实车道线的交并比 (intersection over union, IoU) 得到 F_1 评分作为评价指标。其中真阳性 (TP) 表示 $IoU > 0.5$ 的情况, 而假阳性 (FP) 表示 $IoU \leq 0.5$ 的情况; 假阴性 (FN) 表示预测的车道线不存在的情况。公式如式(23)所示。

$$\left\{ \begin{aligned} Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{TP + FN} \\ F_1 &= \frac{2 \times Precision \times Recall}{Precision + Recall} \end{aligned} \right. \quad (23)$$

本实验中, 采用的硬件配置包括: 处理器为 13thGenIntel(R) Core(TM) i5-13400FCPU@2.50 GHz, GPU 为: NVIDIA GeForce RTX3060, 操作系统选用 Ubuntu20.04, CUDA12.0。实验使用 Pytorch 深度学习框架, 采用匈牙利损失作为损失函数, 采用 Adam 算法优化模型, 设定学习率为 0.000 1, 样本批量大小为 16, 迭代次数为 500 000 次。

3.2 对比实验

因 CULane 和 TuSimple 数据集所采用的数据格式不同, 故本研究在上述两个数据集上分别选取了在该数据集上具有广泛公开结果或表现突出的典型方法进行对比, 如表 2 所示。在 CULane 复杂场景下的对比实验中, DMCTR 的 F_1 分数达到 69.67%, 较原始 LSTR (65.19%) 提升 4.48%, 也优于其他 LSTR 改进版本。该结果验证了 DMCTR 在特征增强与计算效率方面的有效平衡。尤其在交叉路口、夜间、遮挡等极端场景中, 其 DEDA-Block 与 XCA 模块能够更准确地捕捉车道线局部细节与全局结构, 显著提升了模型的鲁棒性与泛化能力。

表 2 不同方法在 CULane 数据集上对比

Table 2 Comparison of different methods on the CULane dataset

方法	F1 分数
SCNN ^[2]	68.53
LaneNet ^[3]	63.51
UFLD V2 ^[5]	68.34
LSTR ^[7]	65.19
LSTR 的改进方法 ^[8]	67.30
LSCoMer ^[9]	68.21
DMCTR	69.67

如表 3 所示, 在 TuSimple 数据集上, DMCTR 模型与包括 LaneATT、PolyLaneNet 在内的 8 种主流车道线检测

方法以及多种 LSTR 改进模型进行了全面对比。实验结果显示,DMCTR 在准确率上达到 96.74%,相较于基线模型 LSTR 的 96.18%,提升了 0.56%。同时,DMCTR 在 FP 和 FN 上也表现优异,分别为 2.71%和 2.93%,相较于 LSTR 的 2.91% FP 与 3.38% FN,DMCTR 在误检与漏检控制方面相比较其他算法均有明显改善;可以看到,DMCTR 不仅在准确率上优于多数主流方法与两种 LSTR 改进方法,而且在误检与漏检控制方面也表现均衡。

表 3 不同方法在 TuSimple 数据集上对比

Table 3 Comparison of different methods on the TuSimple dataset

方法	Acc/%	FP/%	FN/%	FPS/fps
LaneATT ^[4]	95.57	3.76	3.10	171
UFLD V2 ^[5]	96.22	3.18	4.37	—
PolyLaneNet ^[20]	93.36	2.91	3.38	115
ADNet ^[21]	96.23	2.91	3.29	—
GANet-L ^[22]	96.44	2.64	2.47	63
ENet-SAD ^[23]	96.64	6.02	2.05	75
LSTR ^[7]	96.18	2.91	3.38	420
LSCoMer ^[8]	96.68	2.76	3.20	180
LSTR 改进方法 ^[9]	96.31	2.76	3.22	420
DMCTR	96.74	2.71	2.93	353

3.3 消融实验

为验证 DMCTR 各模块的有效性,本节基于 TUSimple(简单场景)和 CULane(复杂场景)数据集,对各模块进行实验分析。如表 4 所示,以原始 LSTR 模型作为基线,引入各模块进行性能分析。单独加入 DEDA 模块后,两项指标分别提升至 96.45%和 67.06%,证明了该模块通过 DCN 和 DA-block 有效增强了车道线形变的识别能力。单独引入 FBSM 模块使 CULane 的 F1 分数显著提升至 67.52%,验证了其特征抑制机制在挖掘弱特征方面的有效性。而单独使用 XCA 模块在将 F1 分数提升至 66.34%的同时,推理速度达到 453 fps,模型计算复杂度为 0.365,体现了其在保持性能的同时提升计算效率的优势。

模块组合实验进一步揭示了其协同作用,DEDA-Block 与 FBSM 的组合将模型性能提升至 96.69%和 68.43%;XCA 与 FBSM 的结合使得模型在保持 386 fps 的高速度的同时,将 F1 分数提至 68.95%;而 DEDA-Block 与 XCA 的组合则在性能与效率间取得了良好平衡。实验结果表明所有组合均有效推动了模型性能的持续提升。

最终,完整 DMCTR 模型在 TuSimple 和 CULane 数据集上分别达到 96.74%和 69.67%的最高性能,同时模型计算复杂度为 0.541,推理速度保持在 353 fps,验证了各模块在车道线检测任务上的有效性。

表 4 消融实验结果

Table 4 Results of ablation study

Baseline	DEDA-Block	FBSM	XCA	Accuracy/%		MACs/G	FPS/fps	
				TuSimple	CULane		TuSimple	
✓				96.18	65.19	0.574		420
✓	✓			96.45	67.06	0.616		372
✓		✓		96.34	67.52	0.592		386
✓			✓	96.22	66.34	0.365		453
✓	✓	✓		96.69	68.43	0.663		338
✓	✓		✓	96.56	68.47	0.462		392
✓		✓	✓	96.50	68.95	0.398		386
✓	✓	✓	✓	96.74	69.67	0.541		353

3.4 可视化实验对比

为直观展示 DMCTR 模型的改进效果,本研究在多种典型场景下将其与基线模型 LSTR 的检测结果进行了可视化对比。如图 7 (a)所示,面对模糊、断续的虚线车道线,LSTR 的预测结果出现了明显的断裂,且对中间模糊的车道漏检,而 DMCTR 凭借 FBSM 模块对弱特征的增强能力,能够生成连续、准确的车道线。在图 7 (b)的夜晚强光反射和车道线模糊场景中,LSTR 的预测因强光干扰和夜晚光照不足而严重失真,将对向车道灯光误检为车道,且同样出现中间车道线漏检。DMCTR 仍有效检出中间的

模糊车道线。在 TuSimple 数据集下,DMCTR 同样表现优异。如图 7 (c)的长直道场景,两种模型均能较好处理,但 DMCTR 的预测曲线更为平滑,且可检测距离较 LSTR 更远。而在图 7 (d)长弯道场景中,LSTR 的预测曲线在弯道远端开始偏离真实车道,DMCTR 则可将远端的弯道清晰完整检出。

综上所述,DMCTR 模型在虚线、模糊、强光、弯道等多种挑战性场景下,均能实现更可靠、更精确的车道线检测,其对于局部细节的捕捉能力和全局结构的理解能力均优于基线模型。

原图
Original image

LSTR

DMCTR

(a) CULane 数据集中的模糊、虚线车道线
(a) Blurred and dotted lane lines in the CULane dataset原图
Original image

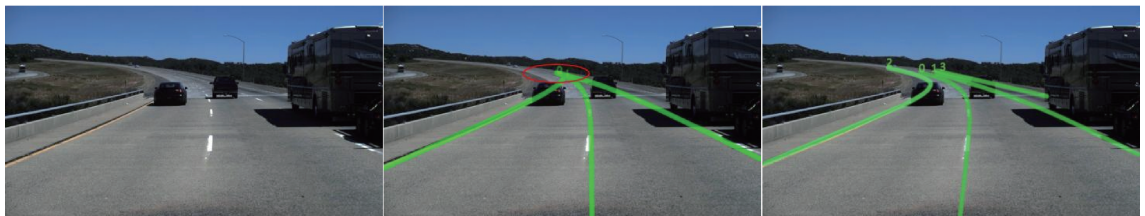
LSTR

DMCTR

(b) CULane 数据集中的夜晚、强光反射车道线
(b) Night and strong light reflection lane lines in the CULane dataset原图
Original image

LSTR

DMCTR

(c) TuSimple 数据集中的长直车道线
(c) Long straight lane lines in the TuSimple dataset原图
Original image

LSTR

DMCTR

(d) TuSimple 数据集中的长弯道车道线
(d) Long curve lane lines in the TuSimple dataset

图 7 各类场景下图像检测效果对比

Fig. 7 Comparison of image detection effects in various scenarios

4 结 论

本研究提出了一种动态多路径协方差 Transformer 车道线检测模型 DMCTR。针对骨干网络中特征挖掘不均衡的问题,引入 FBSM 模块,通过多层级嵌入 FBSM 模块,利用特征抑制机制迫使网络关注易忽略的细节,解决了传统卷积在骨干网络中对弯曲、虚线段等弱特征响应不足的难题。在局部特征提取方面,创新性地构建了 DEDA-Block 模块,将可变形卷积的几何形变适应能力与双注意力机制的协同增强效应进行高效融合,解决了传统单一卷积形式在弯曲车道场景下感受野固定、以及复杂环境中局部特征

捕捉不足的瓶颈问题,显著提升了模型对复杂车道几何形状的建模能力。在全局建模方面,通过引入 XCA 注意力机制,解决了 Transformer 编码器中多头自注意力计算复杂度高、难以兼顾实时性与全局依赖建模的矛盾。保持较低的线性复杂度的同时,获得了进行高效且强大的空间结构建模能力。实验结果表明,通过一系列创新改进使得 DMCTR 在 TuSimple 和 CULane 数据集上准确率 Acc 与 F1 分数分别提升至 96.74% 与 69.67%,同时保持 353 fps 的实时性。

未来的研究工作将进一步改进网络结构,优化模型结构以降低计算复杂度,探索多模态数据融合以增强特征表

达能力,并在更多复杂场景下验证模型的泛化能力,以推动车道线检测技术的发展。

参考文献

- [1] 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 perception algorithm for autonomous driving [J]. Electronic Measurement Technology, 2023,46(24):157-163.
- [2] 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, DOI: 10.48550/arXiv.1712.06080.
- [3] NEVEN D, DE BRABANDERE, GEORGOULIS S, et al. Towards end-to-end lane detection: an instance segmentation approach [C]. 2018 IEEE Intelligent Vehicles Symposium(IV). IEEE, 2018: 286-291.
- [4] 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, 2021: 294-302.
- [5] QIN Z Q, WANG H Y, LI X. Ultra fast structure-aware deep lane detection[C]. European Conference on Computer Vision, Cham: Springer International Publishing, 2020: 276-291.
- [6] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28 (4): 562-574.
- [7] 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, 2021: 3694-3702.
- [8] 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 line detection[J]. Journal of Chinese Computer Systems, 2024, 45(8):1863-1868.
- [9] 郑剑,郭亦畅.基于 LSTR 和 Vit-CoMer 骨干的车道线检测方法[J].电子测量技术,2024,47(19):164-171.
ZHENG J, GUO Y CH. Lane detection method based on LSTR and ViT-CoMer[J]. Electronic Measurement Technology, 2024,47(19):164-171.
- [10] 邵凯,吴广,梁燕,等.基于局部特征编解码的自动驾驶 3D 目标检测[J].系统工程与电子技术,2025,47(10): 3168-3178.
SHAO K, WU G, LIANG Y, et al. Local feature encode-decoding based 3D target detection of autonomous driving [J]. Systems Engineering and Electronics, 2025,47(10):3168-3178.
- [11] HE K M, ZHANG X Y, REN SH Q, et al. Deep residual learning for image recognition [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [12] SONG J Q, YANG R Y. Feature boosting suppression and diversification for fine-grained visual classification[C]. 2021 International Joint Conference on Neural Networks(IJCNN), 2021: 1-8.
- [13] SUN G Q, PAN Y ZH, KONG W K, et al. DA-TransUNet: Integrating spatial and channel dual attention with transformer U-net for medical image segmentation [J]. Frontiers in Bioengineering and Biotechnology, 2024, 12(5): 1398237.
- [14] CHOLLET F. Xception: Deep learning with depthwise separable convolutions [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017(7): 1251-1258.
- [15] DAI J F, QI H ZH, XIONG Y W, et al. Deformable convolutional networks [C]. IEEE International Conference on Computer Vision, 2017: 764-773.
- [16] EI-NOUBY A, TOUVRON H, CARON M, et al. XcIT: Cross-covariance image transformers [J]. Advances in Neural Information Processing Systems, 2021, 34(5): 20014-20027.
- [17] 左银华,赵明岩,吴顺海,等.基于改进 MobileViT 模型的大球盖菇品质分级实验研究[J].实验技术与管理, 2023, 40(9):81-90.
ZUO Y H, ZHAO M Y, WU SH H, et al. Experimental study on quality grading of stropharia rugoso-annulata based on improved MobileViT model[J]. Experimental Technology and Management, 2023, 40(9):81-90.
- [18] 李沐原,张兰春,张博源.基于多尺度空洞融合注意力的车道线检测算法[J].电子测量技术,2024,47(23): 84-92.
LI M Y, ZHANG L CH, ZHANG B Y. Lane detection method based on multi-scale dilated fusion attention [J]. Electronic Measurement Technology, 2024,47(23):84-92.
- [19] 蒯晓,丁飞,张登银.多分支融合注意力机制的车道线检测模型[J].电子测量与仪器学报, 2023, 37(4): 35-43.

- KUAI X, DING F, ZHANG D Y. Lane detection model of multi-branch fusion attention mechanism[J]. Journal of Electronic Measurement and Instrumentation, 2023,37(4):35-43.
- [20] TABELINI L, BERRIEL R, PAIXAO T M, et al. Polylanenet: Lane estimation via deep polynomial regression[C]. 2020 25th International Conference on Pattern Recognition(ICPR), 2021: 6150-6156.
- [21] LIU ZH, LIN W J, LI X P, et al. ADNet: Attention-guided deformable convolutional network for high dynamic range imaging[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 463-470.
- [22] 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, 2022: 1392-1401.
- [23] 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.

作者简介

李睿, 硕士, 主要研究方向为图像处理、自动驾驶。

E-mail: lirui970704@163.com

敖银辉(通信作者), 博士, 教授, 主要研究方向为自动驾驶。

E-mail: aoyinhui@gdut.edu.cn