

基于自适应门控的双路激光雷达三维车道线检测

胡杰^{1,2,3}, 陈楠^{1,2,3}, 徐文才^{1,2,3*}, 昌敏杰^{1,2,3}, 徐博远^{1,2,3}, 王占彬^{1,2,3}, 郭启翔⁴¹武汉理工大学现代汽车零部件技术湖北省重点实验室, 湖北 武汉 430070;²武汉理工大学汽车零部件技术湖北省协同创新中心, 湖北 武汉 430070;³武汉理工大学湖北省新能源与智能网联车工程技术研究中心, 湖北 武汉 430070;⁴东风汽车股份有限公司商品研发院, 湖北 武汉 430100

摘要 局部和全局特征提取在车道线检测任务中扮演着重要的角色。针对现有的基于激光雷达的车道线检测算法局部特征提取层次单一和全局特征利用不充分的问题,提出一种基于自适应门控的双路激光雷达三维车道线检测算法。首先,通过全局特征金字塔结构,使网络提取具有全局相关性的多层次车道线特征;其次,引入双路结构,其中辅助通路将底层高分辨率纹理特征压缩到高级抽象语义特征从而降低计算复杂度,压缩的语义特征用作另一个构建的主通路的先验信息,帮助学习更加精细的底层纹理细节,并借助高效移动卷积模块,在缓解注意力窗口效应的同时产生更好的下采样特征;最后,设计一种自适应多阶门控模块来更好地利用上下文信息,使网络自适应捕捉车道线更具代表性的全局纹理及语义特征。在K-Lane数据集上的实验表明,所提算法 F_1 分数较主流模型提升2.6个百分点,在不良光照、严重遮挡条件下分别提升2.7个百分点和3.5个百分点。将算法部署在实车平台实现在线检测,验证了算法的工程实用价值。

关键词 激光雷达; 自动驾驶; 深度学习; 三维车道线检测; 自适应门控

中图分类号 TN958.98; TN249 **文献标志码** A

DOI: 10.3788/CJL230456

1 引言

随着人工智能的快速发展,自动驾驶成为了国内外交通行业发展的新趋势。自动驾驶汽车需要根据各种传感器获取的外部环境信息规划出安全有效的路线,其中车道线检测是环境感知系统的重要部分。车道线检测通过对车道特征的提取获得车道中每条车道线的准确位置和形状,得到车道线与车辆间的相对位置关系^[1]。在自动驾驶的实际应用场景中,大多基于视觉的方法^[2-5]对光照条件比较敏感,无法直接获得三维世界的深度信息。激光雷达获取的数据精度高、不易受光照变化影响,并且能产生更丰富的空间三维信息;同时激光雷达的点云不需要像图像一样投影到鸟瞰图(BEV)中进行运动规划,避免了投影带来的车道线失真问题^[6-8]。因此,基于激光雷达的三维车道线检测算法研究具有重要的意义。

目前,基于激光雷达的车道线检测方法主要分为两类:传统方法和深度学习方法。传统方法是通过对点云的强度或者反射率设定阈值的方法来检测车道点,如文献[9]使用固定的极坐标栅格地图来存储点的强度,并沿着方位角用阈值来过滤候选车道;文献[10]

提出一种结合亮度梯度和亮度统计直方图的边缘检测方法,可以从密集的点云数据中提取道路标记;文献[11]提出一种基于反射率的实时车道线检测方法,使用DBSCAN(density-based spatial clustering of applications with noise)聚类方法对过滤后的点云数据进行聚类。然而,这种依赖于预定义阈值参数的方法并不能适应不同的驾驶环境。深度学习方法是将激光点云输入到神经网络进行三维特征学习,直接预测出车道线的空间位置和分类置信度,如文献[12]提出一种激光雷达车道线检测网络,将前视摄像头图像与激光雷达点云融合产生BEV图像,以提高车道线检测性能;文献[13]提出一种基于卷积神经网络的车道线检测模型,使用来自激光雷达点云的BEV图像生成车道线建议;文献[14]提供了世界上第一个也是最大的激光雷达车道线检测数据集KAIST-Lane(K-Lane),并提出一种基于自注意力的全局特征提取网络用于在各种条件下检测车道线;后续,作者在此基础上,提出一种采用逐行检测的两阶段激光雷达车道线检测网络, F_1 分数较之前提升0.6%,参数量减少30%^[15]。

与传统方法相比,基于深度学习的方法更加关注的是如何有效提取车道线点云特征,但目前的方法大

收稿日期: 2023-01-12; 修回日期: 2023-02-20; 录用日期: 2023-03-15; 网络首发日期: 2023-03-22

基金项目: 湖北省科技重大专项(2020AAA001, 2022AAA001)

通信作者: *wencaixu_val@163.com

多是基于卷积神经网络提取车道线局部特征,忽略了全局特征的重要性,导致网络的鲁棒性不强且无法满足复杂场景下的工程需要。文献[14]首次将点云的全局特征提取应用于激光雷达车道线检测算法中,并且提供了基准模型 LLDN-GFC (lidar lane detection network-global feature correlator)。基准模型包括 BEV 编码网络、主干网络和检测头三部分。BEV 编码网络部分采用点投影编码器^[16],将点云数据投影到 xy 平面上并使用 ResNet^[17]生成二维特征图;主干网络采用 ViT (vision transformer) 模型^[18]进行特征提取,获取丰富的深层全局特征;最后将生成的特征图送入检测头,对车道进行定位和分类。然而,ResNet 模型提取的特征层次较为单一,且不能在网络的早期关注到车道线的全局特征,增加了主干网络提取全局特征的难度,使网络不能对距离较远的车道线目标进行精准的定位和分类;ViT 模型存在窗口效应,没有考虑窗口内信息的交互作用,因此不能充分地提取和利用全局特征,且存在参数量较大的问题,导致在不同光照、严重遮挡条件下模型检测性能出现下降。

针对以上问题,本文提出一种基于自适应多阶门控的双路激光雷达三维车道线检测网络 LLDN-AGDP (lidar lane detection network-adaptive gating and dual pathways)。首先在 BEV 编码网络部分提出全局特征金字塔网络 (GFPN) 结构,有效融合不同尺度多层次的特征,使网络能够提取具有全局相关的多层次语义和位置信息的车道线特征,增强对远距离车道线目标特征信息的捕获能力;然后在主干网络部分引入双路结构 (DPS)^[19],其中,辅助通路可以将底层高分辨率纹理特征压缩到高级抽象语义特征,降低计算复杂度,主通路可以有效利用辅助通路压缩的高级抽象语义特征作为学习更加精细的底层纹理细节时的先验信息,降低全局特征提取难度;针对因局部窗口注意力导致信息丢失的问题,引入改进高效移动卷积 (E-MBConv) 模块,有效交换像素之间的局部信息,增强表征能力;最后,设计了一种自适应多阶门控 (AMOG) 模块,根据网络结构层次,通过上下文信息之间的多阶空间交互,使网络能够自适应捕捉车道线更具代表性的全局纹理及语义特征。通过以上改进,使本文提出的车道线检测网络能够更有效地提取和利用车道线全局特征并聚合多层次特征,提高了模型的鲁棒性和检测精度,在不良光照、严重遮挡场景下,效果更加明显。

2 相关研究

2.1 特征金字塔

特征金字塔网络 (FPN) 由 Lin 等^[20]提出用于解决目标检测中多尺度变化问题。FPN 在增加极少计算量的情况下,利用卷积神经网络固有的多尺度金字塔结构,通过横向连接来构建自上而下不同尺度、具有高

度语义信息的特征金字塔网络结构。文献[21]提出一种基于轻量化扩充感受野模块的特征金字塔网络结构,增强网络对目标多尺度的适应性,提高小目标检测的准确率;文献[22]提出一种基于双注意力的特征金字塔网络,在 FPN 的横向连接中同时引入通道注意力和空间注意力,从而为特征金字塔带来更多语义信息。

2.2 Transformer

Transformer^[23]通过查询、键和值构建具有捕捉长距离依赖能力的自注意力模块,显示出优越的性能,被用于图像和点云数据处理。文献[24]在图像重建任务中提出一种基于上下文信息的 Transformer,通过卷积和 Transformer 并行的信息流动更好地提取光谱特征;文献[25]提出的 Swin Transformer 是一种基于移动窗口的层次化 Transformer,将特征图划分到不同固定大小的局部窗口中,移动窗口允许信息在相邻窗口中进行传递,并将自注意力计算限制在不重叠的局部窗口中降低计算成本;文献[26]提出的 CAT 采用一种交叉注意力机制,将局部窗口内的注意力和各窗口间的注意力结合起来,从而使局部信息和全局信息进行交互。

2.3 门控机制

门控作为一种注意力机制,能够忽略不相关信息而获取重要的特征信息^[27]。文献[28]在目标跟踪任务中提出一种二维门控模块,可以为所在层的每个通道、每个空间位置提供可学习的动态特征选择机制,从而自适应增强多尺度特征;VAN^[29]和 FocalNet^[30]分别采用大核扩张卷积和多个连续的 3×3 卷积来产生门控权重,以实现自注意力的自适应和远程相关;HorNet^[31]提出一种递归门控卷积,采用深度可分离卷积产生门控权重,并通过递归设计来执行高阶空间交互,基于卷积框架实现类似于 Transformer 的输入自适应、远程和高阶空间交互的特点。

3 基本原理

LLDN-AGDP 整体架构如图 1 所示,分为 BEV 编码网络、主干网络和检测头三部分^[14]。

BEV 编码网络利用点投影编码器将输入的三维点云投影至二维伪图像,通过带有 GFPN 结构的 ResNet34^[17]对伪图像提取特征生成二维 BEV 特征图。主干网络由 E-MBConv 模块、双路模块、融合模块和 AMOG 模块组成,用于提取深层次全局特征,称为双路多阶段全局特征提取网络。首先,E-MBConv 模块对 BEV 特征图划分的像素块建模捕获局部信息,产生更好的下采样特征;其次,利用双路模块中交叉注意力提取经过下采样和局部信息交互后特征图的全局特征,主通路的输出经过 AMOG 模块捕捉上下文信息,提取车道线更具代表性的全局纹理及语义特征;然后,将双路模块中两条通路的输出拼接后送入融合模块,通过并行信息传播增强网络深层全局特征提取,输出通过 AMOG 模块聚合车道线特征;最后,将融合模块

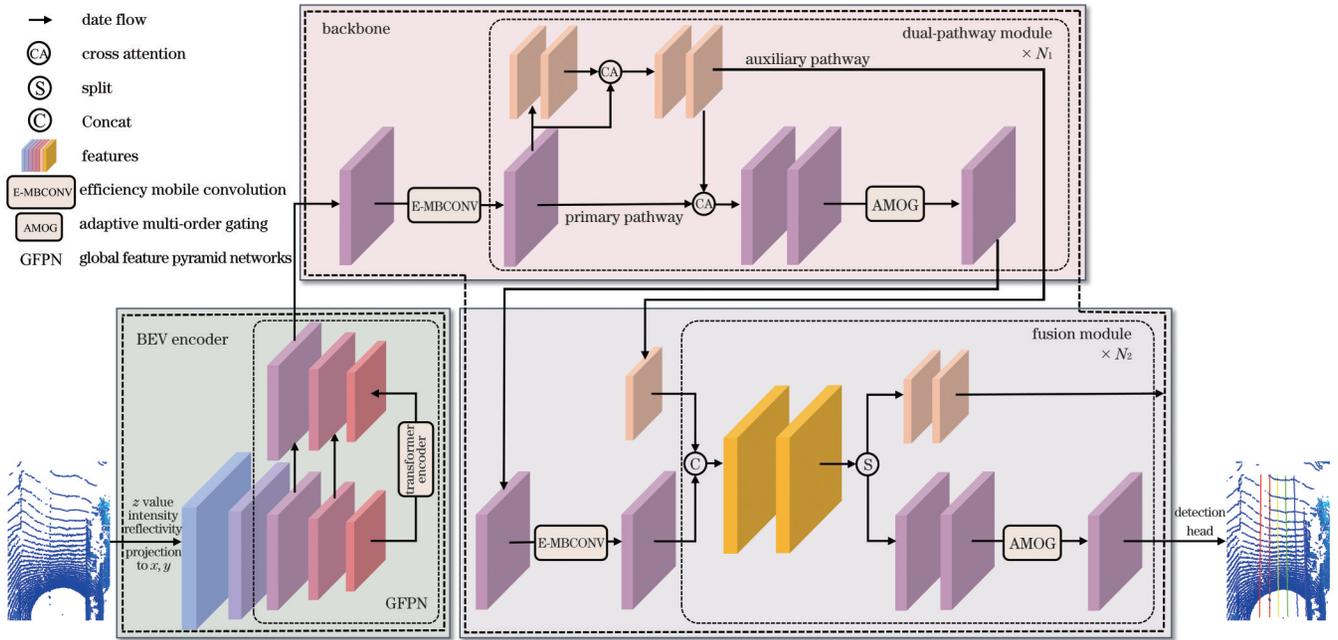


图 1 LLDN-AGDP 整体架构

Fig. 1 Overall architecture of LLDN-AGDP

的输出送入检测头进行车道线的定位和分类。LLDN-AGDP 各模块详细介绍如下。

3.1 BEV 编码网络

BEV 特征图中包含车道线丰富的特征信息,浅层特征具有强位置、弱语义信息,而深层特征具有强语义、弱位置信息,因此对 BEV 特征图进行多层次特征融合有利于车道线不同特征信息的提取和检测精度的提升。考虑到车道线形状细长,需要长距离建模,因此在最后一层特征图处理中通过 Transformer^[23]引入全局信息,再与上两层通过卷积操作生成的特征图进行特征融合,保留车道线更多全局语义和位置信息,提高对远距离车道线目标的识别精度,并使得送入主干网络中的特征图带有全局相关的多层次信息。

具体处理过程可视化如图 2(a)所示。BEV 编码网络首先将激光雷达产生的三维点云数据通过点投影机投影到 xy 平面,并利用点云额外的 z 轴坐标、反射率及强度信息生成伪图像的三个通道信息^[14]。为了在保证实时速度的同时不丢失车道信息,得到的伪图像在下采样之后使用一个自下而上的四阶段 ResNet34 网络进行特征提取,每个阶段都可以得到一层含有不同位置和语义信息的特征图,后三层特征图通过 GFPN 结构输出 BEV 编码网络最终的结果。GFPN 结构通过自上而下的路径和横向连接,将浅层的位置纹理特征和深层的抽象语义特征进一步融合。考虑到 Transformer 自注意力模块所需计算成本高,因此只在最后一层特征图的横向连接引入 Transformer 以得到

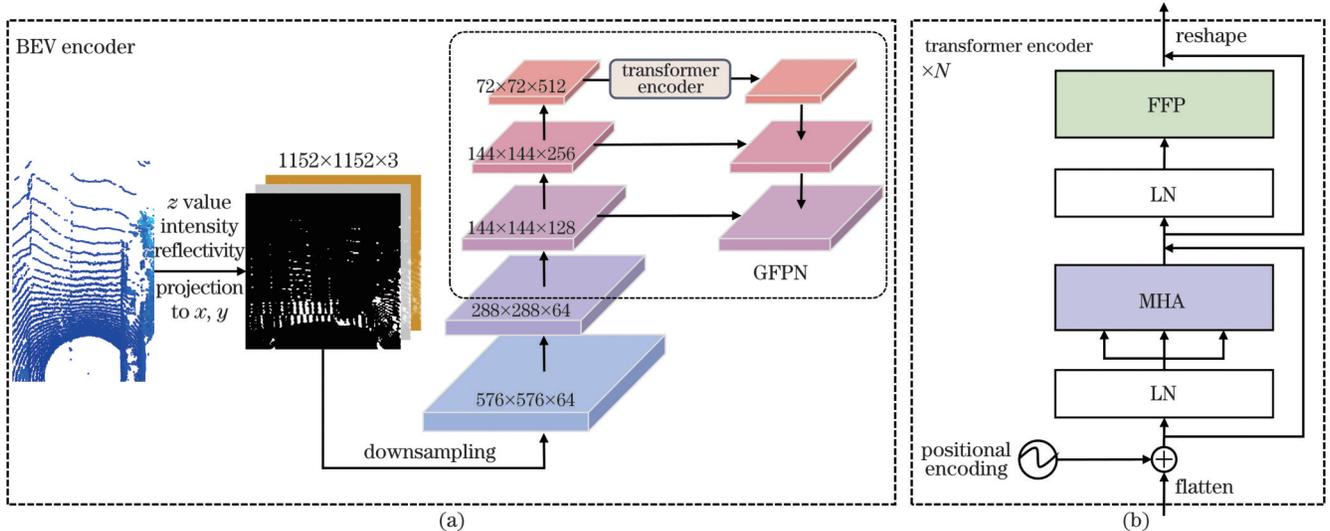


图 2 BEV 编码网络示意图。(a)GFPN 结构示意图;(b)Transformer 编码器

Fig. 2 Schematic diagram of BEV encoder. (a) Schematic diagram of GFPN structure; (b) transformer encoder

带有车道线全局信息的特征图,其他浅层横向连接采用 1×1 的卷积得到带有车道线局部信息的特征图,自上而下通过插值方式进行特征图的融合,在丰富车道线像素点的同时,也进一步融合了不同尺度多层次特征信息。最后将具有全局相关的多层次位置和语义信息的特征图通道数转为 64 后输入到主干网络中进行更高层次全局特征的提取。

Transformer 编码器结构如图 2(b) 所示,由一个多头注意力层(MHA)和一个前馈层(FFN)组成,在每层之前进行层归一化(LN)操作。计算细节如下:

$$z'_{in} = \text{MHA}[\text{LN}(z_{in}), \text{LN}(z_{in}), \text{LN}(z_{in})] + z_{in}, \quad (1)$$

$$z_{out} = \text{FFN}[\text{LN}(z'_{in})] + z'_{in}, \quad (2)$$

$$\text{MHA}(q, k, v) = \text{Concat}(h_1, \dots, h_i) W^o, \quad (3)$$

$$h_i = \text{Attention}(q W_i^q, k W_i^k, v W_i^v), \quad (4)$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_h}}\right)V, \quad (5)$$

式中: z_{in} 为输入; z'_{in} 为经过 MHA 层后的输出; z_{out} 为输出; LN 为层归一化操作; MHA 为多头注意力操作; FFN 为前馈操作; $Q, K, V \in \mathbb{R}^{(HW) \times D}$ 为生成的对应维度的矩阵; $\text{Concat}(\cdot)$ 为拼接操作; h_i 为第 i 个注意力头的结果; W^o, W_i^q, W_i^k, W_i^v 为权重矩阵; $\text{Attention}(\cdot)$ 为自注意力操作; softmax 为归一化指数函数操作; d_h

为每个注意力头的维度。

3.2 双路多阶段全局特征提取网络

BEV 编码网络中的 GFPN 结构只提取了最后一层特征图中的全局信息,并不能充分发挥 Transformer 中自注意力模块捕捉输入之间长距离依赖的能力,因此在主干网络部分引入改进了一种基于自适应门控的双路多阶段全局特征提取网络,可以有效提取丰富的深层次全局纹理特征和语义特征,并在保持车道线检测精度的前提下有效地降低网络计算复杂度。

双路多阶段全局特征提取网络的整体框架如图 3 所示,包括两个阶段,第一个阶段由一个 E-MBCONV 模块和 N_1 个双路模块组合而成,第二个阶段由一个 E-MBCONV 模块和 N_2 个融合模块组成。首先, E-MBCONV 模块将特征图大小从 $144 \times 144 \times 64$ 采样到 $144 \times 144 \times 8$,生成带有通道注意力的下采样特征图;然后,将特征图按 8×8 窗口划分为相同大小的特征补丁输入到双路模块,通过两条通路之间的交叉注意力减少自注意力层中标记数量,从而降低计算复杂度;输出的特征图通过 E-MBCONV 模块再输入到融合模块,进一步整合车道线全局纹理和语义特征,主通路和辅助通路输出被聚合到一起联合训练,并通过并行的信息传播增强网络对更深层全局特征的提取能力。两个阶段中每个双路模块和融合模块输出的特征图都会输入到 AMOG 模块,提取车道线更具代表性的全局纹理及语义特征。

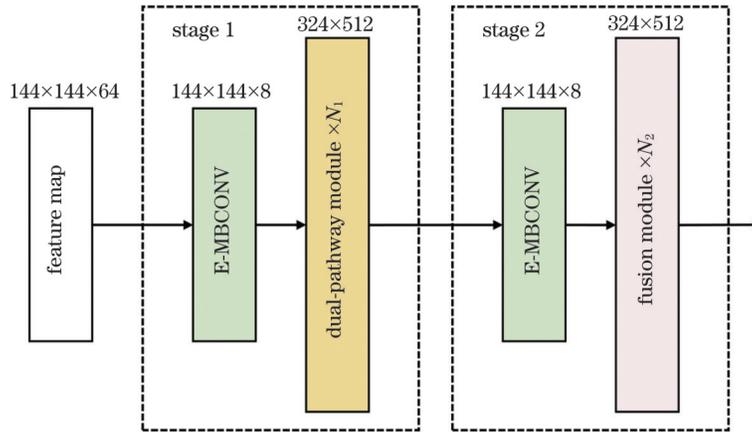


图 3 双路多阶段全局特征提取网络框架示意图

Fig. 3 Schematic diagram of dual-pathway multi-stage global feature extraction network framework

3.2.1 高效移动卷积模块

一般在特征提取过程中,特征图都会经过下采样层,以减少图像固有的冗余和计算过载,然而点云产生的特征不如图像稠密,下采样操作可能会损失过多信息。因此本文引入改进的 E-MBCONV 模块用来代替下采样操作,不仅可以产生更好的下采样特征,而且通过 E-MBCONV 中高效通道注意力(ECA)^[32]机制对输入特征图各通道重新分配权重,增强网络表征能力,并且因为卷积的局部感受野可以有效地交换相邻像素之间的局部信息,所以在双路模块及融合模块中不需要

增加额外的窗口移动机制。

E-MBCONV 模块结构如图 4 所示, E-MBCONV 模块是在移动卷积(MBCONV)^[33]模块的基础上采用了更加高效的通道注意力机制而形成的。 E-MBCONV 模块首先采用 1×1 卷积,将输入特征图通道 c 扩展至 4 倍,然后使用 3×3 的深度可分离卷积捕捉像素之间的局部空间相互作用,接着通过 ECA 模块使用全局信息来对特征图各通道重新分配权重,最后采用 1×1 卷积将特征映射回原始通道尺度,从而实现残差连接。 ECA 模块与压缩激励(SE)通道注意力模

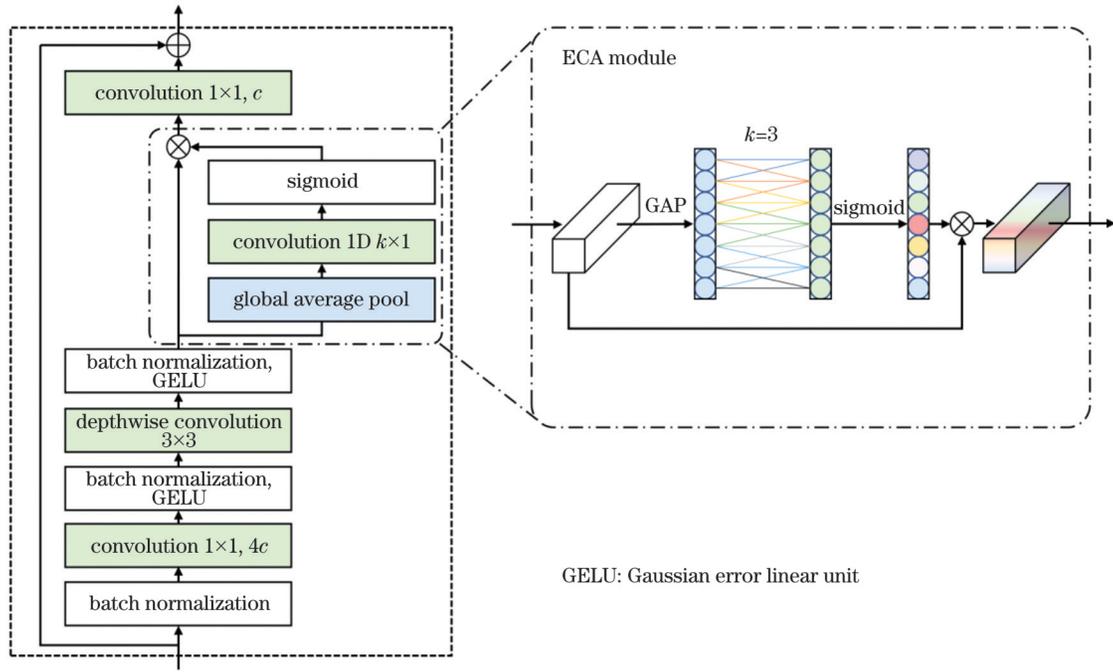


图 4 E-MBConv 模块示意图
Fig. 4 Schematic diagram of E-MBConv module

块^[34]相比,将SE模块中使用全连接层学习全部通道注意力信息的方式,改为使用自适应一维(1D)卷积来学习临近几个通道的注意力信息,大大降低了参数量。具体来说,首先使用全局平均池化(GAP)操作对输入特征图进行空间特征压缩,对压缩后的特征,使用自适应一维卷积进行邻近几个通道特征的学习,一维卷积的尺度与输入特征图的通道数自适应相关,图中尺度

k 为3,最后将学习到的通道权重矩阵与原始输入特征图进行逐通道相乘,最终输出具有高效通道注意力的特征图。与MBConv模块相比,E-MBConv模块具有更低的参数量,在双路多阶段全局特征提取网络中却可以达到与MBConv模块相当的性能。

3.2.2 双路模块和融合模块

如图5(a)所示,双路模块包括两条通路,分别是

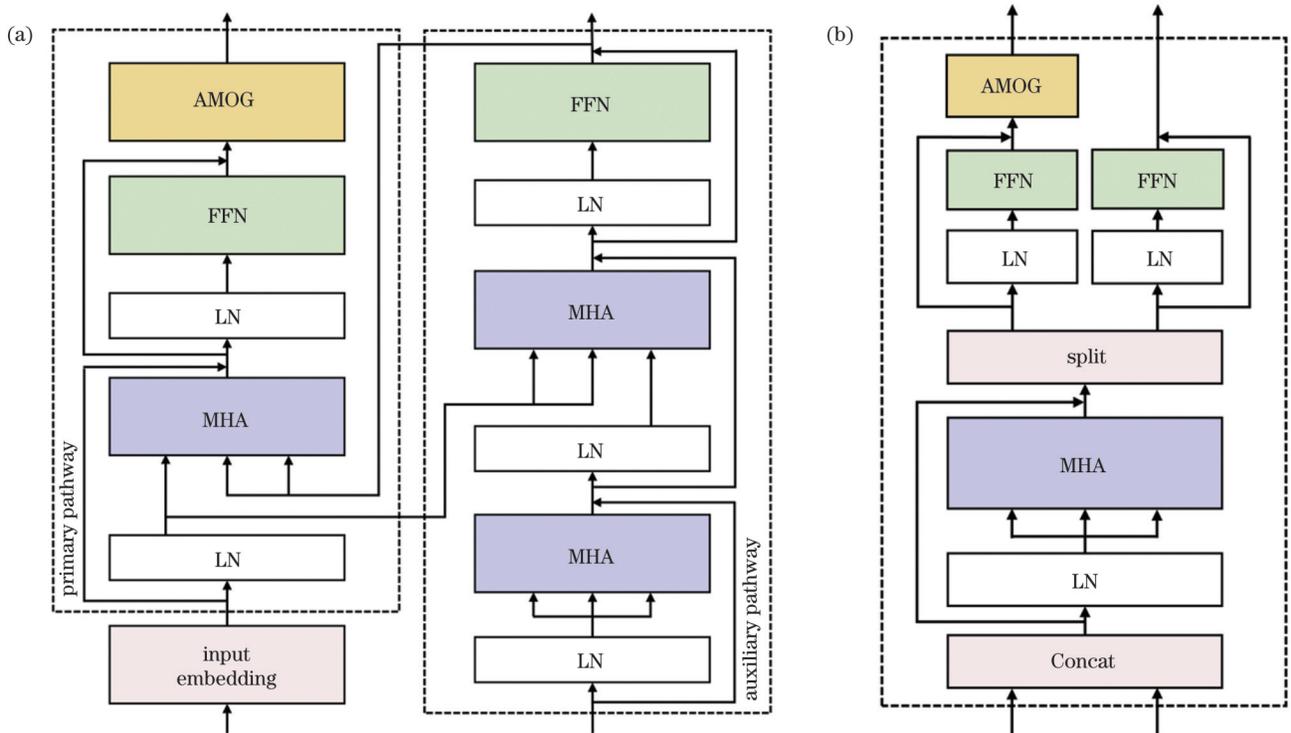


图 5 双路模块和融合模块。(a)双路模块;(b)融合模块
Fig. 5 Dual-pathway module and fusion module. (a) Dual-pathway module; (b) fusion module

主通路和辅助通路。辅助通路包括两个 MHA 层和一个 FFN 层,在每层前存在 LN 操作。主通路拥有和传统 Transformer 相似的结构,包括 MHA 层、FFN 层以及一个专门设计的 AMOG 模块。

首先,辅助通路将输入到双路模块的特征图压缩成抽象语义特征,减少标记数量;经过 LN 操作,保证数据分布的一致性,加快模型的收敛速度,减少网络整体的训练时间;再经过 MHA 层,通过自注意力机制对语义标记进行上下文编码,获取车道线全局特征,然后经过 LN 操作后利用由主通路传来的高分辨率纹理特征与抽象语义特征之间的交叉注意力来提取粗颗粒语义特征;最后经过 FFN 层得到带有深层全局信息的语义标记。辅助通路第一个 MHA 层和 FFN 层信息流动与传统 Transformer 相同,公式同式(1)和式(2),第二个 MHA 层输入的键和值来自主通路,信息流动如下所示:

$$z''_m = \text{MHA}[\text{LN}(z'_m), \text{LN}(x_m), \text{LN}(x_m)] + z'_m, \quad (6)$$

式中: x_m 为主通路输入; z'_m 为经过辅助通路第一个 MHA 层后的输出; z''_m 为经过第二个 MHA 层后的输出。

辅助通路输出的语义标记被输入到主通路,作为辅助学习车道线高分辨率纹理特征的先验信息,并且被当作一种增强的语义标记,提供给下一个双路模块的辅助通路。首先,主通路输入的特征图进行 LN 操作后,利用辅助通路输出的粗颗粒语义特征与主通路输入的高分辨率纹理特征的交叉注意力来精炼细颗粒纹理特征;然后,再次经过 LN 操作后通过 FFN 层增强自注意力的效果;最后,将特征图输入到 AMOG 模块,保留更多所需要的车道线全局特征。主通路信息流如下所示:

$$x'_m = \text{MHA}[\text{LN}(x_m), \text{LN}(z_{out}), \text{LN}(z_{out})] + x_m, \quad (7)$$

$$x_{out} = \text{AMOG}\{\text{FFN}[\text{LN}(x'_m)] + x'_m\}, \quad (8)$$

式中: z_{out} 为辅助通路输出; x'_m 为经过主通路 MHA 层后的输出; AMOG 为自适应多阶门控操作; x_{out} 为主通路输出。

在第一阶段双路模块中,由于车道线高分辨率纹理特征巨大的复杂性,使得主通路纹理标记之间的相互作用并未得到充分的利用。为了缓解这个问题,在第二阶段引入一个简单的融合模块,如图 5(b) 所示,对联合在一起的纹理和语义标记执行自注意力操作,从而充分利用标记间的相互作用进一步提取深层次的全局特征。由于两条通路的标记传达不同的信息,因此融合模块中采用两个不同的 FFN 层分别输出纹理和语义标记,最后将纹理标记送入 AMOG 模块得到最终的输出。

第一阶段由于双路模块相互作用、互为补充的信息流动方式,在进行反向传播时,既可以通过主通路到辅助通路的交叉注意力来补充由于抽象语义特征压缩

导致的信息丢失,又可以通过辅助通路到主通路的交叉注意力带来的先验信息降低高分辨率纹理特征提取的难度。第二阶段融合模块通过并行的信息传播,增强网络对更深层全局特征的提取能力,确保了复杂纹理标记间的深层次全局特征得到充分提取和利用,提高了车道线的检测精度。

3.2.3 自适应多阶门控模块

AMOG 模块利用跨层次特征图之间丰富的上下文信息来进行多阶空间交互,从而有效聚合更具代表性的车道线全局特征。AMOG 模块具有输入自适应、长距离建模及高效的特点。首先,AMOG 模块对输入特征图的数量没有限制,可自适应地根据输入特征图的数量进行相对应的空间交互;其次,AMOG 模块采用了大尺寸卷积核,增强了捕捉长期依赖的能力,也相对缓解了双路多阶段全局特征提取网络中窗口机制带来的局部信息损失;第三,AMOG 模块是基于深度可分离卷积实现的,相比于普通卷积及自注意力操作,避免了过高的计算复杂度。

AMOG 模块结构简单、即插即用,如图 6 所示,假设在双路多阶段全局特征提取网络中使用了 N_1 个双路模块,相应地会有 N_1 个特征图输入到 AMOG 模块。当 AMOG 模块储存了前两个特征图后,对第一个特征图用 7×7 大小、步长为 1(s1) 的深度可分离卷积进行局部特征权重的获取,得到一阶的门控权重矩阵,将得到的权重矩阵与第二个特征图进行逐位置相乘,经过一个线性层得到最终的特征图,称为一阶空间交互;一阶空间交互特征图输入到下一个双路模块得到第三个特征图,与之前操作相同,会得到二阶空间交互的结

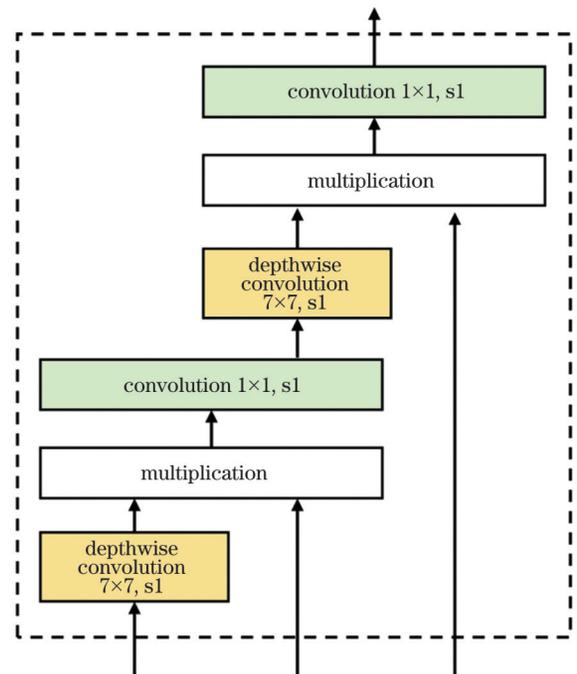


图 6 AMOG 模块二阶空间交互示意图

Fig. 6 Schematic diagram of second-order spatial interaction of AMOG module

果;以此类推,最终会得到 $N_1 - 1$ 阶空间交互的结果。

3.3 检测头和损失函数

LLDN-AGDP 的检测头和损失函数设置与基准模型 LLDN-GFC^[14] 相同,是将车道线检测问题看作实例分割问题,每条车道线被视为不同类别,同时每个像素都会被指定一个类别分数和一个置信度分数,用来表明每个像素是否为车道线点以及属于哪一条车道线。检测头由两个分割头组成,分别是分类头和置信度头。如图 7 所示,给定来自双路多阶段全局特征提取网络大小为 $H_{bev} \times W_{bev} \times C_{out}$ 特征图的输入,通过两

层连续的共享多层感知机(MLP)产生最终的预测输出。对于分类头,第一层共享 MLP 将特征图维度从 C_{out} 扩展到 $2C_{out}$ 以增强网络表征能力,第二层共享 MLP 将特征维度映射回 N_{cls} 。由于最多生成 6 条车道实例以及存在像素点为背景的可能,因此类别数目 N_{cls} 设置为 7。经过 softmax 函数获得 $H_{bev} \times W_{bev} \times N_{cls}$ 的分类特征图,显示每个网格每类的概率。对于置信度头,同样先将特征图维度扩展到 $2C_{out}$,第二层共享 MLP 将特征维度映射回 1,经过 sigmoid 函数获得 $H_{bev} \times W_{bev} \times 1$ 的置信度特征图,显示每个网格为车道线的概率。

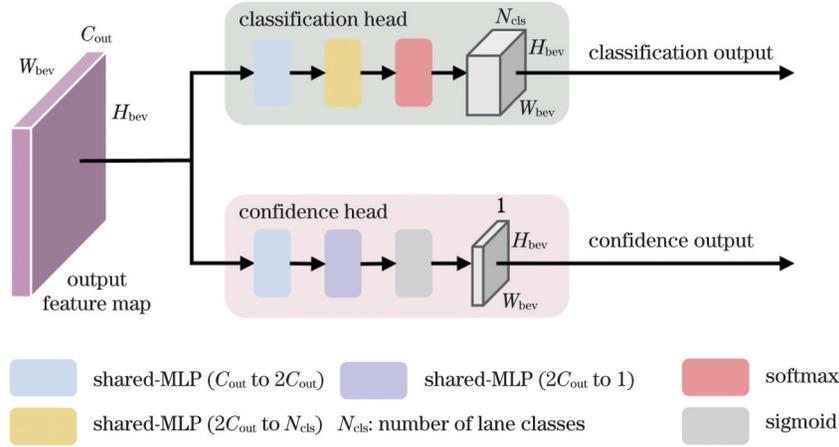


图 7 LLDN-AGDP 检测头

Fig. 7 Detection head of LLDN-AGDP

由于每一帧车道线样本数量明显小于背景样本的数量,因此采用处理正负样本极度不平衡问题的 softdice 损失^[35]作为置信度损失,定义为

$$L_{conf} = 1 - \frac{2 \sum_i^N \sum_j^M x_{conf_{i,j}} \hat{x}_{conf_{i,j}}}{\sum_i^N \sum_j^M x_{conf_{i,j}}^2 + \sum_i^N \sum_j^M \hat{x}_{conf_{i,j}}^2 + \epsilon}, \quad (9)$$

式中: L_{conf} 为置信度损失; $\hat{x}_{conf_{i,j}}$ 为网格 (i, j) 处置信度的预测值; $x_{conf_{i,j}}$ 为目标网格置信度的真实值; N, M 分别为网格的行数和列数; ϵ 设置为 10^{-12} 保证分母不为零。

对于分类损失,由于车道线共 6 个类别,选择广泛用于多分类问题的交叉熵损失,引导网络在训练阶段学习最大正确概率的车道线类别。总损失函数为置信度损失和分类损失之和,具体如下所示:

$$L_{cls} = \frac{1}{NM} \sum_i^N \sum_j^M \log [p(\hat{x}_{cls_{i,j}})], \quad (10)$$

$$p(\hat{x}_{cls_{i,j}}) = \frac{\exp(\hat{x}_{cls_{i,j}}^k)}{\sum_{k'=1}^C \exp(\hat{x}_{cls_{i,j}}^{k'})}, \quad (11)$$

$$L_{total} = L_{conf} + L_{cls}, \quad (12)$$

式中: $p(\hat{x}_{cls_{i,j}})$ 为网格 (i, j) 处正确类别 $k = x_{cls_{i,j}}$ 预测的 softmax 值; C 为车道线类别总数; $\hat{x}_{cls_{i,j}}^k$ 为网格 (i, j) 处类别为 $k = x_{cls_{i,j}}$ 的预测值; $\hat{x}_{cls_{i,j}}^{k'}$ 为网格 (i, j) 处类别为 k' 的预测值; L_{cls} 为分类损失; L_{total} 为总损失函数。

4 实验分析与讨论

4.1 数据集和评价指标

本文采用的是第一个公开的大型激光雷达车道线数据集 K-Lane, 该数据集包括 15382 frame 各种条件下最多带有 6 条车道线标注的激光雷达点云数据, 共分为 7687 frame 训练数据和 7695 frame 测试数据, 每组数据包含不同的道路条件和场景, 具体包括: (1) 光照条件, 如白天、夜晚; (2) 车道类型, 如城市道路、高速公路; (3) 车道遮挡等级, 根据被遮挡车道线数量共分为 6 个等级; (4) 匝道及车道弯曲等级, 其中弯曲等级分为无、轻微、急弯。各道路条件和场景所占比例如图 8 所示。

K-Lane 数据集采用 F_1 分数作为评价指标, 对预测结果上的每个网格分别评估是否存在车道点以及是否被正确分类。 F_1 分数是精确率和召回率的调和平均值, 定义为

$$F_1 = \frac{2}{\frac{1}{c_{Precision}} + \frac{1}{c_{Recall}}} = \frac{2 N_{TP}}{N_{TP} + 0.5(N_{FP} + N_{FN})}, \quad (13)$$

式中: $c_{Precision}$ 为精确率, 表示预测为真阳性样本中, 实际为真阳性样本的比例; c_{Recall} 为召回率, 表示实际为真阳性样本中, 预测为真阳性样本的比例; N_{TP} 为真阳性样本数量; N_{FP} 为假阳性样本数量; N_{FN} 为假阴性样本

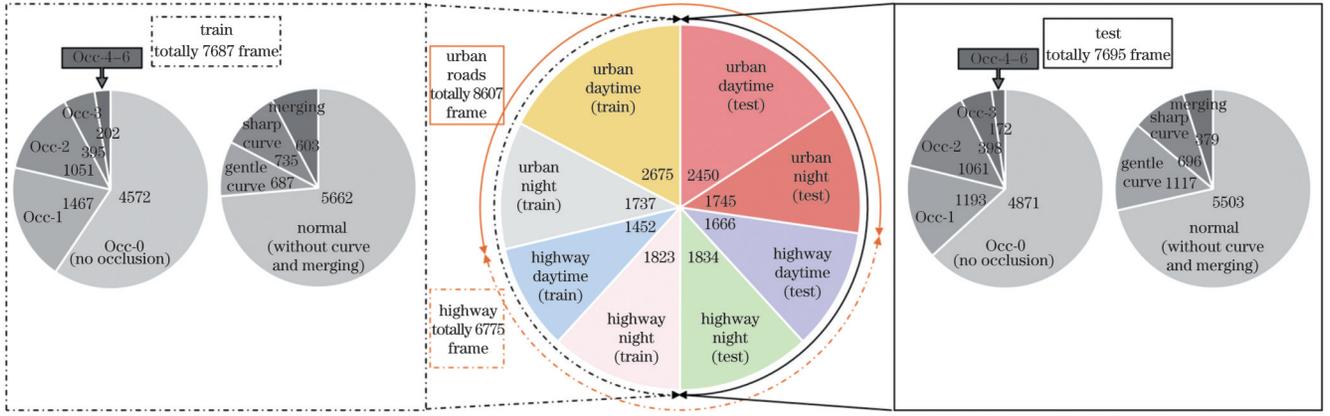


图 8 数据集分布

Fig. 8 Dataset distribution

数量。

4.2 训练细节和参数设置

本实验所有算法均在 Ubuntu 18.04 系统下, 基于框架 Pytorch 1.7.1 完成, 以端到端的方式在一张 Tesla V100 显卡中进行训练, 选用 Adam 优化器, batch size 设置为 4, 学习率设置为 0.0002, 置信度交并比 (IoU) 阈值设置为 0.5, 模型共迭代 60 个 epoch。在 BEV 编码网络部分, 网络参数与基准网络 LLDN-GFC^[14] 保持一致, 设置如下: 点投影器感兴趣点云范围 x 轴为 $[0, 46.08 \text{ m}]$, y 轴为 $[-11.52 \text{ m}, 11.52 \text{ m}]$, 划分的网格尺寸为 $[0.04 \text{ m}, 0.02 \text{ m}]$, GFPN 结构最终输出特征图通道数设置为 64。在主干网络部分, 双路模块和融合模块个数 N_1 、 N_2 通过消融实验都设置为 3, 注意力窗口大小设置为 8, 输入标记长度设置为 512, MHA 层注意力头个数设置为 16, 双路模块和融合模块 FFN 层扩展比分别设置为 8 和 3, 除 N_1 、 N_2 外均采用文献 [19] 中的默认值。在检测头部分, 根据文献 [14], C_{out} 设置为 1024, N_{cls} 设置为 7。

4.3 实验结果及分析

在本节中, 我们不仅将 LLDN-AGDP 与基准网络 LLDN-GFC 的性能进行对比, 还把 LLDN-GFC 中主干网络部分替换为其他先进的全局特征提取网络或卷积神经网络, 并与其进行对比。为了保证实验公平性, 测试方法与对比文献^[14] 完全一致。

替换的卷积神经网络与对比文献采用相同的设置, 分别是 RNF-S13、RNF-D23、RNF-C13, 其中, RNF 表示带有 FPN 结构的 ResNet, S13、D23、C13 分别代表使用步长卷积、扩张卷积^[36] 及 CBAM^[37] 实现的 13 层或者 23 层残差块。扩张卷积可以在不损失分辨率的情况下增加感受野; CBAM 执行类似于 LLDN-GFC 的自注意力机制, 却是沿着每个通道轴计算注意力权重, 意味着它不像 LLDN-AGDP 那样对所有输入的补丁执行全局相关操作。替换的全局特征提取网络采用 Swin Transformer^[25] 和 CAT^[26]。我们将上述网络在 K-Lane 测试集上进行测试, 测试结果如表 1 和表 2 所示。

表 1 不同算法在 K-Lane 测试集上的置信度 F_1 分数比较Table 1 Comparison of confidence F_1 scores of different algorithms on K-Lane test set

Method	Overall / %	Daytime / %	Night / %	Urban / %	Highway / %	Curve / %	No occlusion / %	Occ-4-6 / %	Frame rate / (frame·s ⁻¹)
RNF-S13	73.2	72.6	74.0	73.1	73.3	70.5	74.9	63.5	13.1
RNF-C13	78.0	77.6	78.5	77.7	78.3	76.0	79.6	69.3	13.0
RNF-D13	72.1	71.3	73.0	71.9	72.3	69.6	74.0	61.9	13.2
LLDN-GFC ^[14]	82.1	82.2	82.0	81.7	82.5	78.0	82.9	75.9	11.6
RLLDN-LC ^[15]	82.7	82.5	82.9	81.6	84.0	76.1	83.4	79.1	16.4
LLDN-Swin ^[25]	82.5	82.6	82.5	81.6	83.7	75.7	83.0	77.4	10.7
LLDN-CAT ^[26]	82.7	82.7	82.6	81.7	83.9	77.3	83.5	77.7	11.2
LLDN-AGDP	84.7	84.6	84.7	84.2	85.3	78.9	85.2	79.4	11.8

如表 1 所示, 所提出的 LLDN-AGDP 性能大幅度领先于主干网络采用卷积神经网络设计的激光雷达车道线检测算法, 原因是 LLDN-AGDP 对全局信息的利用, 大大降低了车道线检测的难度; 同时, LLDN-

AGDP 在各种道路及场景下性能全面优于 LLDN-GFC、RLLDN-LC 和其他主干网络替换成先进全局特征提取网络的激光雷达车道线检测算法, 且速度保持与基准网络 LLDN-GFC 相当的水平。尤其是在不良

表 2 不同算法在 K-Lane 测试集上的分类 F_1 分数比较Table 2 Comparison of classification F_1 scores of different algorithms on K-Lane test set

Method	Overall / %	Daytime / %	Night / %	Urban / %	Highway / %	Curve / %	No occlusion / %	Occ-4-6 / %	Frame rate / (frame·s ⁻¹)
RNF-S13	70.5	70.1	71.0	70.4	70.6	68.1	72.3	59.0	13.1
RNF-C13	75.3	75.1	75.5	74.8	76.0	73.1	77.0	65.3	13.0
RNF-D13	68.6	68.3	69.4	68.7	69.0	66.5	70.7	57.6	13.2
LLDN-GFC ^[14]	81.1	81.4	80.7	80.6	81.7	76.7	81.9	75.5	11.6
LLDN-Swin ^[25]	81.4	81.1	81.8	79.5	82.8	72.3	81.7	77.3	10.7
LLDN-CAT ^[26]	81.5	81.5	81.3	80.7	82.7	76.1	82.3	77.1	11.2
LLDN-AGDP	83.6	83.8	83.5	83.0	84.3	78.1	84.3	79.1	11.8

光照(夜晚)和严重遮挡(Occ-4~6)条件下,较基准网络 LLDN-GFC 平均置信度 F_1 分数分别提升 2.7 和 3.5 个百分点,原因是 LLDN-AGDP 通过 AMOG 模块利用丰富的上下文信息聚合深层次全局特征,同时通过 E-MBCONV 模块实现窗口内局部信息的交互,因此在不良光照和严重遮挡条件下,性能依然能够保持稳定。综上所述,LLDN-AGDP 算法表现出了优越的性能。

为了更直观地显示 LLDN-AGDP 的检测效果,展示部分 K-Lane 测试集上的检测结果,如图 9 所示,图中包括各种条件下的检测示例,其中每列表示一个条件,每行的上方显示图像数据,包括左上角的真实

BEV 标签和车道预测结果在前视图中的投影,下方显示带有车道预测结果的点云数据。从图 9(a)和图 9(b)的两列检测结果可知,白天和夜晚条件下的光照强度逐渐变化,但由于激光雷达相比于相机受光照影响较小,可以为网络提供高质量的点云数据输入,远处较暗部分的车道线仍可以被精准识别;从图 9(c)和图 9(d)的检测结果可以看出,在城市道路和高速公路场景下,LLDN-AGDP 在不同光照条件下都可以稳定输出预测结果;从图 9(e)弯道和图 9(g)严重遮挡条件的检测结果可知,被遮挡和弯道部分车道线识别存在一定的抖动,但仍可以较为准确地检测车道线的位置。

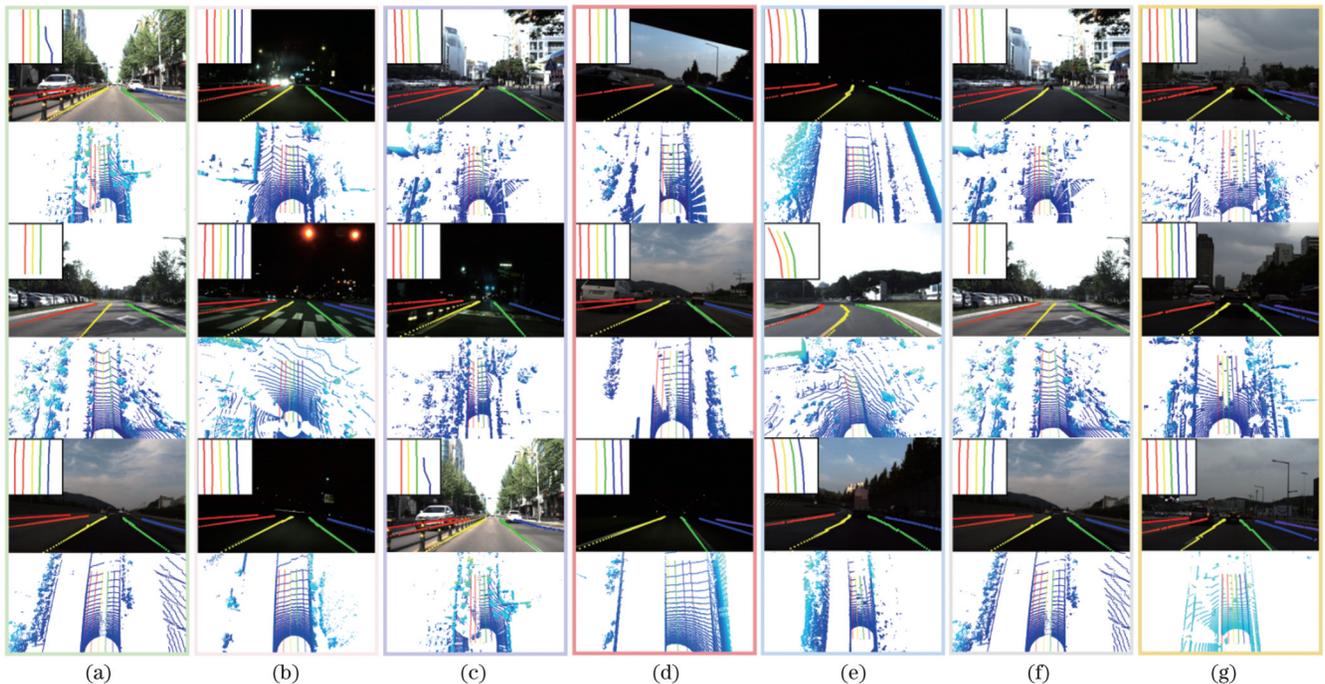


图 9 检测结果在 K-Lane 测试集上可视化。(a)白天;(b)夜晚;(c)城市道路;(d)高速公路;(e)弯道;(f)无遮挡;(g)严重遮挡

Fig. 9 Visualization of detection results on K-Lane test set. (a) Daytime; (b) night; (c) urban; (d) highway; (e) curve; (f) no occlusion; (g) occlusion Occ-4-6

4.4 注意力可视化

在本小节中,采用注意力分数可视化的方式,验证 LLDN-AGDP 在遮挡场景下的鲁棒性。LLDN-AGDP 及其他基于自注意力机制的网络利用全局长距

离建模的优势可以更多地关注特征图上有意义的区域。因此,可以通过可视化每个网络的注意力分布来对比各网络在遮挡条件下性能的优劣。可视化结果如图 10 所示。

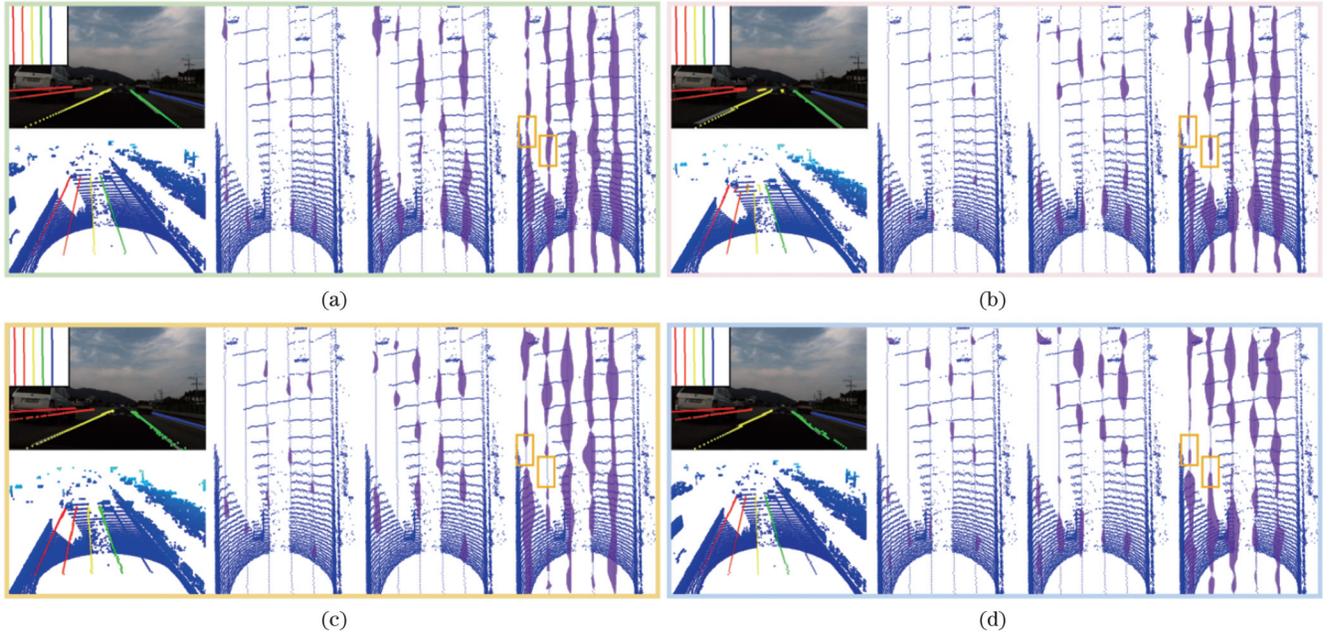


图 10 各网络注意力分数可视化结果。(a)LLDN-AGDP算法;(b)LLDN-GFC算法;(c)LLDN-Swin算法;(d)RLLDN-LC算法
Fig. 10 Visualization results of attention scores of each network. (a) LLDN-AGDP; (b) LLDN-GFC; (c) LLDN-Swin; (d) RLLDN-LC

从可视化结果可以看出,各网络通过削弱不相关特征的注意力权重来更加关注含有车道线特征的区域,从各阶段的注意力分布可知,随着网络层次的加深,各网络均继续扩大含车道线特征的感兴趣区域。从图 10(a)可以看出,为 LLDN-AGDP 使用 3 个融合模块足以确保感兴趣区域覆盖包含车道线的整个点云区域。此外,对于车道线遮挡区域(图中橙色方框位置),LLDN-AGDP 表现出比其他网络更强的关注,这种现象表明在车道线遮挡条件下,由于其独有的设计,LLDN-AGDP 通过增强车道线局部与全局信息、纹理与语义信息及上下文信息的交互补充,仍然使得被遮挡部分可以被精确地估计。

4.5 消融实验

为了验证本算法提出的每个模块结构在整个网络模型中的作用,进行了各模块有效性实验对比。测试实验均在 K-Lane 训练集上进行训练,在 K-Lane 测试集上进行验证,实验条件和参数均与之前保持一致,使用 F_1 置信度分数作为评价标准。

4.5.1 LLDN-AGDP 消融实验

LLDN-AGDP 消融实验结果如表 3 所示。表 3 中,(a) 是没有加任何模块的基准网络,即 LLDN-GFC 的检测结果;(b) 是在网络中引入 DPS 的检测结果,表明双路交叉注意力和联合训练可以充分利用不同层次特征之间的差异性,从而有利于更深层全局信息的挖掘;(c) 是在网络中加入 DPS 和 GFPN 结构后的检测结果,表明 GFPN 结构可以有效地将浅层特征中强位置信息与深层特征中带全局信息的强语义信息进行融合,提高车道线检测的精度;(d) 是在网络中加入 DPS、GFPN 结构和 E-

MBCONV 模块后的检测结果,表明 E-MBCONV 模块有利于缓解窗口自注意力的限制,实现窗口内局部信息的交互,充分发挥自注意力机制长距离建模的优势;(e) 是加入所有模块结构之后的检测结果,相比于基准网络 LLDN-GFC 平均置信度 F_1 分数提升 2.6 个百分点,网络对于车道线的检测能力得到了显著提升。通过以上对比,验证了本文所提出各模块的有效性。

表 3 LLDN-AGDP 消融实验结果

Table 3 Ablation experiment results of LLDN-AGDP

Method	DPS	GFPN	E-MBCONV	AMOG	Confidence of F_1 / %
(a)					82.1
(b)	✓				83.0
(c)	✓	✓			83.5
(d)	✓	✓	✓		84.1
(e)	✓	✓	✓	✓	84.7

4.5.2 各模块消融实验

为了消除各模块堆叠造成整个网络模型深度变化给性能带来的影响,设计了如下各模块的消融实验。

(1) 改变 DPS 中双路模块和融合模块的数量。考虑到本文提出的融合模块与基准对照模型 ViT 结构相似,参考文献[14],融合模块个数 N_2 设为 3,双路模块个数 N_1 分别设为 1、3、6 与之搭配进行实验;同时为验证融合模块在整个网络模型中的必要性,设置三组对照实验,将融合模块删除或替换为双路模块,分别单独使用 3、6、9 个双路模块进行全局特征的提取。实验结果如表 4 所示。

(2) 将 E-MBCONV 中高效通道注意力 ECA 模块

替换为 SE 模块或删除;将 GFPN 结构最后一层横向连接中的自注意力操作替换为 1×1 卷积操作;AMOG

模块取消重新分配权重的过程,只经过一个线性层直接输出结果。实验结果如表 5 所示。

表 4 DPS 消融实验结果

Table 4 Ablation experiment results of DPS

Method	$N_1=3, N_2=3$ (baseline)	$N_1=1, N_2=3$ (test 1)	$N_1=6, N_2=3$ (test 2)	$N_1=3, N_2=0$ (test 3)	$N_1=6, N_2=0$ (test 4)	$N_1=9, N_2=0$ (test 5)
Confidence of $F_1 / \%$	84.7	84.4	84.3	83.5	84.1	83.3

表 5 E-MBCONV、AMOG 和 GFPN 消融实验结果

Table 5 Ablation experiment results of E-MBCONV, AMOG and GFPN

Method	E-MBCONV			AMOG		GFPN	
	Baseline	MBCONV ^[33]	Removing ECA	Baseline	Removing depthwise convolution	Baseline	FPN ^[20]
Confidence of $F_1 / \%$	84.7	84.7	84.4	84.7	84.0	84.7	84.2

从 DPS 消融实验 1 和 2 结果可知,当融合模块个数 N_2 为 3 时,随着双路模块个数 N_1 的增加,LLDN-AGDP 的平均置信度 F_1 分数先上升再下降;从实验 3、4、5 结果可知,当只使用双路模块进行全局特征提取时,由于复杂纹理标记间的深层次全局特征未得到充分提取, F_1 分数均出现不同程度的下降,且随着 N_1 的增加, F_1 分数同样先上升再下降。从表 5 中 E-MBCONV 模块的消融实验结果可知,采用 ECA 模块可以达到与 SE 模块相同的检测性能,而去除通道注意力模块后网络的 F_1 分数下降 0.3 个百分点;AMOG 模块消融实验结果表明,舍弃模块中重新分配权重的过程会导致网络的性能大幅度下降, F_1 分数较基准下降 0.7 个百分点;GFPN 结构消融实验结果表明,由于标准 FPN 结构在最后一层中未引入全局信息,较采用

GFPN 结构网络平均置信度 F_1 分数下降 0.5 个百分点。

通过上述各模块消融实验,在消除网络深度对实验影响的条件下,进一步验证了各模块的有效性和必要性,同时通过实验得出较为理想的双路结构。

4.6 算法本地数据集部署

为了验证上述算法的工程性能,使用自动驾驶试验车 WHUT-E70 在校园场景下采集图像和激光雷达数据,将采集的数据按照 K-Lane 数据集格式进行标注,用标注好的本地数据集对算法进行训练。所用试验平台如图 11 所示,安装了摄像头、激光雷达、毫米波雷达、全球导航卫星系统 (GNSS)、惯性测量单元 (IMU) 等高性能传感器。



图 11 试验平台硬件部署

Fig. 11 Hardware deployment of test platform

算法在试验平台下的检测结果如图 12 所示。检测结果显示,LLDN-AGDP 网络可以对车道线进行准

确的检测和定位,即使被遮挡的车道线也可以被较准确地检测到。

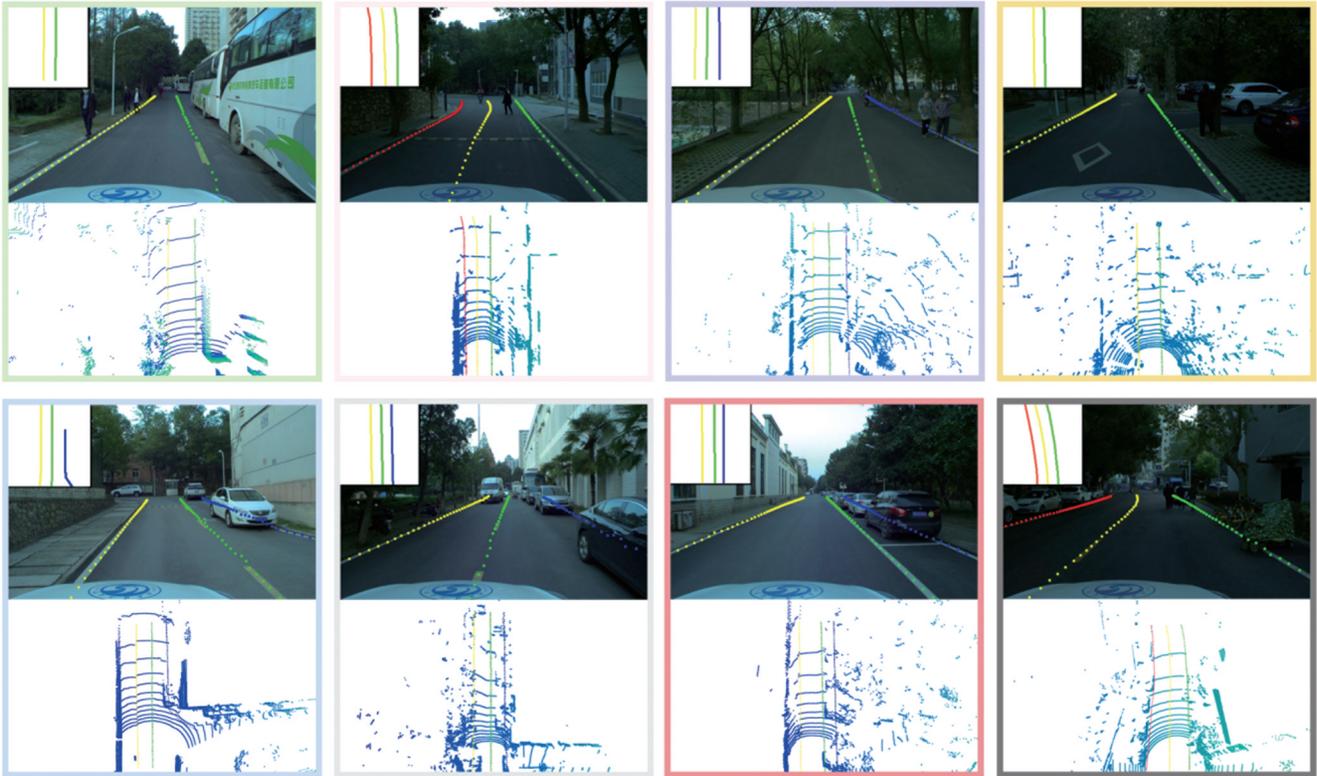


图 12 本地数据集检测结果

Fig. 12 Detection results of local data set

5 结 论

本文提出了一种基于自适应多阶门控的双路激光雷达三维车道线检测算法 LLDN-AGDP。在 BEV 编码网络部分提出 GFPN 结构,使网络能够有效融合不同层次的纹理特征和语义特征,提高对远距离车道线目标的识别能力,并在网络的早期就关注车道线的全局信息,为深层全局信息提取提供更加丰富的表征信息;在主干网络部分提出双路多阶段全局特征提取网络,通过两条通路相互作用、互为补充的信息流动方式,降低计算复杂度的同时减小深层全局特征提取的难度,引入 E-MBCONV 模块缓解自注意力窗口效应,产生更好的下采样特征,并构建 AMOG 模块从而利用上下文信息,自适应地聚合车道线更具代表性的全局特征,提高车道线检测精度。在 K-Lane 测试集上测试结果表明,本算法在不同的道路条件和场景下的平均 F_1 分数可达 84.7%,较主流模型提高 2.6 个百分点,在不良光照、严重遮挡条件下 F_1 分数分别提高了 2.7、3.5 个百分点。与其他方法相比,LLDN-AGDP 通过多层次的特征融合及对全局特征进一步的提取和利用表现出了更加优越的性能。最后,将 LLDN-AGDP 算法部署在实车上,验证了其工程应用价值。

参 考 文 献

- [1] Zheng T, Huang Y F, Liu Y, et al. CLRNNet: cross layer refinement network for lane detection[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 888-897.
- [2] Deng G L, Wu Y F. Double lane line edge detection method based on constraint conditions Hough transform[C]//2018 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), October 19-23, 2018, Wuxi, China. New York: IEEE Press, 2018: 107-110.
- [3] Liu L Z, Chen X H, Zhu S Y, et al. CondLaneNet: a top-to-down lane detection framework based on conditional convolution[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2022: 3753-3762.
- [4] Lee M, Lee J, Lee D, et al. Robust lane detection via expanded self attention[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), January 3-8, 2022, Waikoloa, HI, USA. New York: IEEE Press, 2022: 1949-1958.
- [5] Liu R J, Yuan Z J, Liu T, et al. End-to-end lane shape prediction with transformers[C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV), January 3-8, 2021, Waikoloa, HI, USA. New York: IEEE Press, 2021: 3693-3701.
- [6] Qian Q S, Hu Y H, Zhao N X, et al. Object tracking method based on joint global and local feature descriptor of 3D lidar point cloud[J]. Chinese Optics Letters, 2020, 18(6): 061001.
- [7] 胡杰, 安永鹏, 徐文才, 等. 基于激光点云的深度语义和位置信息融合的三维目标检测[J]. 中国激光, 2023, 50(10): 1010003. Hu J, An Y P, Xu W C, et al. 3D object detection based on deep semantic position information fusion of laser point cloud[J]. Chinese Journal of Lasers, 2023, 50(10): 1010003.
- [8] Wang C X, Zhang H R, Yang M, et al. Automatic parking based on a bird's eye view vision system[J]. Advances in Mechanical Engineering, 2014, 6: 847406.
- [9] Lindner P, Richter E, Wanielik G, et al. Multi-channel lidar processing for lane detection and estimation[C]//2009 12th

- International IEEE Conference on Intelligent Transportation Systems, October 4-7, 2009, St. Louis, MO, USA. New York: IEEE Press, 2009.
- [10] Yang R H, Li Q T, Tan J X, et al. Accurate road marking detection from noisy point clouds acquired by low-cost mobile lidar systems[J]. ISPRS International Journal of Geo-Information, 2020, 9(10): 608.
- [11] Ester M, Kriegel H P, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]// Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, August 2-4, 1996, Portland, Oregon, USA. New York: ACM Press, 1996: 226-231.
- [12] Bai M, Mattyus G, Homayounfar N, et al. Deep multi-sensor lane detection[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 1-5, 2018, Madrid, Spain. New York: IEEE Press, 2019: 3102-3109.
- [13] Martinek P, Pucea G, Rao Q, et al. Lidar-based deep neural network for reference lane generation[C]//2020 IEEE Intelligent Vehicles Symposium (IV), October 19-November 13, 2020, Las Vegas, NV, USA. New York: IEEE Press, 2021: 89-94.
- [14] Paek D H, Kong S H, Wijaya K T. K-Lane: lidar lane dataset and benchmark for urban roads and highways[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 19-20, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 4449-4458.
- [15] Paek D H, Wijaya K T, Kong S H. Row-wise lidar lane detection network with lane correlation refinement[EB/OL]. (2022-10-17) [2022-12-15]. <https://arxiv.org/abs/2210.08745>.
- [16] Ku J, Mozifian M, Lee J, et al. Joint 3D proposal generation and object detection from view aggregation[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 1-5, 2018, Madrid, Spain. New York: IEEE Press, 2019.
- [17] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [18] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale[EB/OL]. (2021-05-18) [2022-12-15]. <https://arxiv.org/abs/2010.11929>.
- [19] Yao T, Li Y H, Pan Y W, et al. Dual vision transformer[EB/OL]. (2022-07-11) [2022-12-15]. <https://arxiv.org/abs/2207.04976>.
- [20] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [21] 黄裕凯, 王青旺, 沈韬, 等. 基于 MobileNet 的多尺度感受野特征融合算法[J]. 激光与光电子学进展, 2023, 60(4): 0410024.
Huang Y K, Wang Q W, Shen T, et al. Multi-scale receptive field feature fusion algorithm based on MobileNet[J]. Laser & Optoelectronics Progress, 2023, 60(4): 0410024.
- [22] Xing H J, Wang S, Zheng D Z, et al. Dual attention based feature pyramid network[J]. China Communications, 2020, 17(8): 242-252.
- [23] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]//Advances in Neural Information Processing Systems, December 4-9, 2017, Long Beach, CA, USA. Cambridge: MIT Press, 2017: 5998-6008.
- [24] Wang L S, Wu Z L, Zhong Y, et al. Snapshot spectral compressive imaging reconstruction using convolution and contextual transformer[J]. Photonics Research, 2022, 10(8): 1848-1858.
- [25] Liu Z, Lin Y T, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2022: 9992-10002.
- [26] Lin H Z, Cheng X, Wu X Y, et al. CAT: cross attention in vision transformer[C]//2022 IEEE International Conference on Multimedia and Expo (ICME), July 18-22, 2022, Taipei, China. New York: IEEE Press, 2022.
- [27] 唐璐, 万良, 王婷婷, 等. DECANet: 基于改进 DeepLabv3+ 的图像语义分割方法[J]. 激光与光电子学进展, 2023, 60(4): 0410002.
Tang L, Wan L, Wang T T, et al. DECANet: image semantic segmentation method based on improved DeepLabv3+ [J]. Laser & Optoelectronics Progress, 2023, 60(4): 0410002.
- [28] Liang Z Q, Wang J S, Xiao G, et al. FAANet: feature-aligned attention network for real-time multiple object tracking in UAV videos[J]. Chinese Optics Letters, 2022, 20(8): 081101.
- [29] Guo M H, Lu C Z, Liu Z N, et al. Visual attention network[EB/OL]. (2022-02-20) [2022-12-15]. <https://arxiv.org/abs/2202.09741>.
- [30] Yang J W, Li C Y, Dai X Y, et al. Focal modulation networks [EB/OL]. (2022-03-22) [2022-12-15]. <https://arxiv.org/abs/2203.11926>.
- [31] Rao Y, Zhao W, Tang Y, et al. HorNet: efficient high-order spatial interactions with recursive gated convolutions[C]// Advances in Neural Information Processing Systems, November 28, 2022, New Orleans, LA, USA. Cambridge: MIT Press, 2022.
- [32] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.
- [33] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3 [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2020: 1314-1324.
- [34] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [35] Sudre C H, Li W Q, Vercauteren T, et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations[EB/OL]. (2017-07-11) [2022-12-25]. <https://arxiv.org/abs/1707.03237>.
- [36] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[EB/OL]. (2015-11-23) [2022-12-15]. <https://arxiv.org/abs/1511.07122>.
- [37] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.

Three-Dimensional Lane Detection Algorithm of Lidar Based on Adaptive Gating and Dual Pathways

Hu Jie^{1,2,3}, Chen Nan^{1,2,3}, Xu Wencai^{1,2,3*}, Chang Minjie^{1,2,3}, Xu Boyuan^{1,2,3},
Wang Zhanbin^{1,2,3}, Guo Qixiang⁴

¹*Hubei Key Laboratory of Advanced Technology for Automotive Components, Wuhan University of Technology, Wuhan 430070, Hubei, China;*

²*Hubei Collaborative Innovation Center for Automotive Components Technology, Wuhan University of Technology, Wuhan 430070, Hubei, China;*

³*Hubei Research Center for New Energy & Intelligent Connected Vehicle, Wuhan University of Technology, Wuhan 430070, Hubei, China;*

⁴*Commercial Product R&D Institute, Dongfeng Automobile Co., Ltd., Wuhan 430100, Hubei, China*

Abstract

Objective Lane detection plays an important role in automatic driving. It is the premise of lane keeping, lane departure warning and high-level automatic driving. Lidar has become a new direction in this field because it can generate more spatial three-dimensional (3D) information and is less affected by bad lighting, shading and other conditions. Currently, lane detection is mainly based on deep learning. Compared with traditional methods, it has higher detection accuracy and better robustness. The key of 3D lane detection with lidar based on deep learning is how to extract and utilize the feature information of lane point cloud completely and efficiently, otherwise it will not be able to cope with various lighting conditions and challenging scenes, which will have a great impact on the realization of automatic driving function. Therefore, how to fully extract and utilize the feature information of point cloud is the key to improve the accuracy of lane detection.

Methods The proposed lidar lane detection network, LLDN-AGDP, consists of three parts, i. e., bird's eye view (BEV) encoder network, backbone network, and detection head. In the BEV encoder part, the original 3D point cloud is projected into a two-dimensional (2D) pseudo-image by a point projector, and the feature is extracted by ResNet34 with global feature pyramid network (GFPN) (Fig. 2). By fusing multi-level features of different scales, a feature map with globally relevant multi-level semantic information is constructed. In the backbone network part, firstly, through the efficiency mobile convolution (E-MBCONV) module (Fig. 4), the local information between the window pixels is exchanged to generate better downsampling features and enhance the network representation ability. Then, the feature map is input into the dual-pathway module and the fusion module (Fig. 5), and the low-level high-resolution texture features are compressed into high-level abstract semantic features, thus reducing the computational complexity. When learning finer low-level texture details, the compressed high-level abstract semantic features can be used as prior information, thus reducing the difficulty of global feature extraction. Moreover, the adaptive multi-order gating (AMOG) module (Fig. 6) is embedded in the dual-pathway module and the fusion module, and the multi-order spatial interaction is carried out by using the rich context information between the cross-level feature maps, so that the network can adaptively extract lane lines. Finally, the lane lines are classified and located by the detection head.

Results and Discussion The proposed LLDN-AGDP network is tested and evaluated on the K-Lane test set (Tables 1 and 2). Compared with the comparison networks, the average confidence F_1 score and average classification F_1 score of LLDN-AGDP are 84.7% and 83.6%, respectively, and the performance of LLDN-AGDP is greatly ahead of the lidar lane detection network using convolutional neural network for the backbone. Meanwhile, LLDN-AGDP outperforms LLDN-GFC, RLLDN-LC and other lidar lane detection algorithms with advanced global feature extraction network in all kinds of roads and scenes. Under bad lighting and severe occlusion conditions, the average confidence F_1 score is increased by 2.7 and 3.5 percentage points, respectively, and the speed is at the same level as that of the benchmark network LLDN-GFC. Through the visualization of attention scores, the robustness of each model under occlusion conditions is compared and analyzed (Fig. 10). The comparison between LLDN-AGDP and other network attention visualization shows that LLDN-AGDP can pay more attention to the areas with lane characteristics and show stronger interest in the lane lines in the blocked areas. Then, the effectiveness of the proposed innovation module is further analyzed (Table 3). The results show that after adding GFPN structure to the network, it can effectively fuse the features of shallow strong position information with the features of deep strong semantic information with global correlation, which brings stronger representation ability to the network. After the introduction of the dual-pathway structure, the network can make full use of the differences between different levels of features to further dig deep-seated global information. After the E-MBCONV module is added, it is beneficial to alleviate the attention limitation of local windows in the dual-pathway structure and realize the interaction of information in the windows. After adding the AMOG module, the feature capture ability of the network is stronger by using multi-order spatial interaction of context information.

Conclusions A 3D lane detection algorithm LLDN-AGDP for lidar based on adaptive multi-order gating and dual pathways is proposed. GFPN structure is proposed in BEV encoder to enable the network to effectively fuse texture features and semantic features of different levels, and pay attention to the global information of lane lines at the early stage of the network. In the part of backbone, a dual-pathway global feature extraction network and AMOG module are proposed, which reduce the computational complexity and the difficulty of deep global feature extraction through the interactive and complementary information flow structure of the two pathways. The AMOG module can make use of rich context information and adaptively aggregate the more representative features of lane lines to improve the detection accuracy of lane lines. Moreover, the E-MBCONV module, which can effectively exchange the local information among the pixels in the window, is introduced. The test results on K-Lane test set show that the average F_1 score of the proposed algorithm can reach 84.7% under different road conditions and scenes, which is 2.6 percentage points higher than that of state-of-the-art model, and the F_1 scores under bad lighting and severe shading conditions are increased by 2.7 and 3.5 percentage points, respectively. Finally, LLDN-AGDP algorithm is deployed on a real vehicle to verify its engineering application value.

Key words lidar; automatic driving; deep learning; three-dimensional lane detection; adaptive gating