

DOI:10.19651/j.cnki.emt.2518749

基于因果干预的 BEV 车道线检测<sup>\*</sup>

李睿豪 于红绯

(辽宁石油化工大学人工智能与软件学院 抚顺 113005)

**摘 要:** 针对如光照突然变化、极端天气等环境干扰导致的鸟瞰图车道线检测中的特征模糊和误检问题,本文提出了一种基于因果干预的 BEV 车道检测框架。首先,为提升 BEV 空间转换过程中特征的表示效果,设计复合位置编码并融合至前视图特征,以保持空间连续性与一致性。其次,在获取 BEV 特征后构建因果干预模块,因果干预模块旨在通过生成反事实特征来显式地将车道线特征与环境干扰解耦,以提高模型在极端环境中的稳定性。最后,通过引入特征融合模块完成多尺度特征的动态校准与干扰抑制,并利用全局注意力机制实现 BEV 特征的增强。实验结果表明,在 Apollo 数据集的三个子集中,相比于性能第 2 的模型,F1 值提高了 0.8%、1%、3%;在 OpenLane 数据集内的包含极端天气、夜间及交叉路口等挑战性场景中,F1 值也达到了最佳。成功实现了车道线特征与环境干扰的显式解耦,为复杂环境下的自动驾驶感知提供了高鲁棒性解决方案。

**关键词:** 计算机视觉;车道线检测;因果干预;位置编码;鸟瞰图

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

## BEV lane detection based on causal intervention

Li Ruihao Yu Hongfei

(College of Artificial Intelligence and Software, Liaoning Petrochemical University, Fushun 113005, China)

**Abstract:** Aiming at the feature ambiguity and misdetection problems in bird's eye view lane line detection caused by environmental disturbances such as sudden changes in illumination and extreme weather, this paper proposes a causal intervention-based BEV lane detection framework. First, to enhance the representation of features during BEV spatial transformation, composite positional encoding is designed and fused to front view features to maintain spatial continuity and consistency. Second, the causal intervention module is constructed after acquiring the BEV features. The causal intervention module aims to explicitly decouple the lane line features from the environmental disturbances by generating counterfactual features to improve the stability of the model in extreme environments. Finally, the dynamic calibration of multi-scale features and interference suppression is accomplished by introducing the feature fusion module, and the global attention mechanism is utilized to achieve the enhancement of BEV features. The experimental results show that in the three subsets of the Apollo dataset, the F1 values are improved by 0.8%, 1%, and 3% compared to the model with the 2nd performance, and the F1 values are also optimal in the challenging scenarios within the OpenLane dataset that contain extreme weather, night, and intersections. The explicit decoupling of lane line features and environmental disturbances is successfully realized, providing a highly robust solution for autonomous driving perception in complex environments.

**Keywords:** computer vision; lane detection; causal intervention; positional encoding; bird's eye view

## 0 引 言

车道线检测是自动驾驶系统的重要组成部分,受到了学术界和工业界的广泛关注。实时准确的车道线检测是高

级自动驾驶的基础之一,在智能巡航控制,定位和地图构建中起着重要作用。目前 2D 车道线检测方法已经获得了显著的成果。但是通过前视图执行车道线检测的框架不适用于实际工业产品,下游任务通常要求车道位置以鸟瞰图

收稿日期:2025-05-06

<sup>\*</sup> 基金项目:国家自然科学基金(61702247)、辽宁省教育厅基本科研项目(LJKMZ20220723)、辽宁省教育厅基本科研项目(LJKMZ20220754)资助

(bird's eye view, BEV) 的形式呈现<sup>[1]</sup>。

为了适应下游任务,大部分 3D 车道线检测方法通过空间转换的方式将前视图特征转换成 BEV 特征,之后再投影到 BEV 空间。开创性工作 3D-LaneNet<sup>[2]</sup>首次将逆透视映射<sup>[3]</sup>(inverse perspective transformation, IPM)与基于锚点的车道表示相结合。其优势在于利用 IPM 显式建模透视变换,但其锚点设计对复杂拓扑车道的适应性有限。为了解决这一问题,Gen-LaneNet<sup>[4]</sup>提出了一个两阶段框架,通过几何引导的锚点生成提高了泛化能力,但代价是计算冗余。PersFormer<sup>[1]</sup>基于 Transformer 实现了前视图与 BEV 空间的特征动态对齐<sup>[5]</sup>,支持统一 2D/3D 锚点建模,但其计算复杂度较高,且视角畸变问题在长距离检测中仍存挑战。

针对效率与精度平衡,近期研究提出多种创新方法。PVALane<sup>[6]</sup>利用 2D 车道预测作为稀疏先验锚点,并结合轻量化的视图无关特征对齐模块,实现了实时推理。然而,其性能高度依赖 2D 检测的准确性,若 2D 检测失败,误差会传递至后续模块。Anchor3Dlane<sup>[7]</sup>和 CurveFormer<sup>[8]</sup>虽然绕开了传统视角转换带来的图像扭曲问题,但它们需要大量带有精确 3D 标记的数据来训练模型。然而,在极端天气下,这类方法的表现会大打折扣,因为训练数据中很少包含这类天气的真实场景,模型面对模糊的车道线时,很难准确预测出它们的 3D 形状和位置。

尽管现有方法在标准场景下已取得显著进展,但在真实驾驶环境中频繁出现的复杂干扰因素导致的图像特征模糊、几何畸变加剧现象<sup>[9]</sup>,严重削弱了现有车道线检测模型鲁棒性<sup>[10]</sup>。近年来,因果推断被视为解决环境干扰导致泛化瓶颈的关键方向。然而,现有相关方法仍存在显著局限:传统方法如多环境数据增强或对抗训练虽能缓解部分过拟合,但难以建模环境因子与车道语义的因果关联<sup>[11-12]</sup>。不变性学习方法(invariant risk minimization, IRM)<sup>[13]</sup>需多环境标注数据支撑,而实际车载场景中环境标签获取成本高昂;干预性生成模型<sup>[14]</sup>通过特征随机化打破虚假关联,但未显式解耦车道特征中的环境噪声。此外,反事实生成技术源于文本领域,难以适配车道线的高维空间特性。而结构因果模型<sup>[15]</sup>其复杂的图推理机制在高维视觉任务中计算效率较低。上述局限表明,现有因果推断方法尚不足以有效应对复杂环境干扰带来的鲁棒性难题。因此,如何构建一种对复杂环境干扰具备强免疫力的车道线检测框架,成为学术界与工业界亟待突破的核心挑战<sup>[10]</sup>。

本文聚焦于提升 BEV 车道线检测在复杂环境下的鲁棒性。本文的核心工作建立在 BEV-LaneDet<sup>[16]</sup>模型的基础之上,并针对其关键不足进行改进。BEV-LaneDet 利用基于多层感知机(multilayer perceptron, MLP)的空间转换模块将前视图特征映射为 BEV 特征,无需真实相机参数是其优势。然而,在转换过程中,MLP 将二维特征图压缩为一维向量,严重破坏了特征的空间结构信息,直接造成检测

精度下降。尽管后续改进模型 LaneCPP<sup>[17]</sup>利用物理先验驱动的三维参数化建模克服了传统 MLP 固定映射的缺陷,其方法仍未能有效解决复杂环境干扰与车道线语义深度耦合的影响。

为突破 BEV-LaneDet 的固有缺陷以及 LaneCPP 未能解决的问题,本文以 BEV-LaneDet 模型为基础,对其空间转换模块进行改造,并引入因果干预机制,提出以下针对性改进策略:

1)针对原空间转换模块破坏空间结构的问题,本文在空间转换模块中引入复合位置编码。通过该位置编码网络将绝对位置信息与相对空间关系深度融合,显著增强了输入特征空间位置信息的表达能力。

2)为提升特征质量并抑制噪声,本文对原模型结构进行优化,在空间转换模块后添加特征融合模块与全局注意力机制。特征融合模块通过多尺度特征交互抑制环境噪声干扰;全局注意力机制则利用长程依赖关系强化模型对弱纹理车道线的感知能力,避免关键细节丢失。

3)为从根本上解决环境干扰与车道语义耦合的问题,本文设计了因果干预模块。该模块通过生成反事实 BEV 特征,显式地剥离光照突变、动态遮挡等环境干扰因子与车道线语义的虚假关联,确保模型在极端条件下的输出稳定性与鲁棒性。

通过详尽的对比实验和消融实验,本文验证了所提整体模型及各改进模块的有效性。实验结果表明,本文方法在极端天气和夜间低照度等复杂场景下均取得了最优异的效果,显著提升了车道线检测的鲁棒性和精度。

## 1 模型设计

### 1.1 模型结构

如图 1 所示,本文网络由虚拟相机、特征提取、复合位置编码、空间转换、特征增强、因果干预模块和 3D 输出头 7 个部分构成,其中复合位置编码、特征增强、因果干预模块为文本所做改进部分,在图 1 中以黑实线框出。网络首先通过虚拟相机对输入的前视图进行预处理,将图像转换为内外参数一致的形式,这种预处理方式保证了空间关系的一致性。预处理后的图像作为输入,通过 EfficientNet<sup>[18]</sup>骨干网络进行特征提取,获取前视图特征。提取的前视图特征与复合位置编码拼接后,被送入 MLP 进行 BEV 转换。转换获得的 BEV 特征依次进入特征融合模块和全局注意力机制中进行特征增强,随后由因果干预模块去除混杂的环境因子,得到最终的 BEV 特征。将 BEV 平面划分为  $c_1 \times c_2$  个单元格并预测该平面上的车道位置。预测包括每个单元格的置信度,嵌入度,以及每个单元格中车道  $y$  方向的偏移量和高度信息  $z$ ,通过将每个分支的结果融合得到三维车道。

### 1.2 复合位置编码

传统基于 MLP 的 BEV 转换方法通过全连接网络直

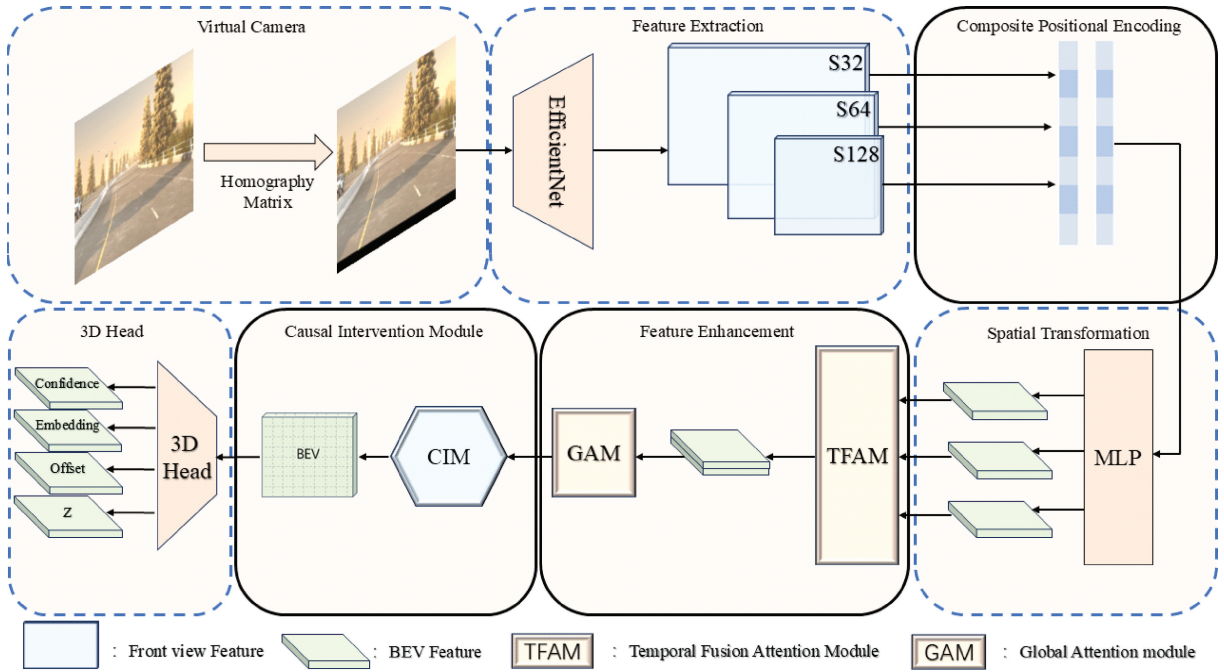


图 1 模型整体框架图  
Fig. 1 Overall framework of the model

接映射前视图特征至鸟瞰空间,其核心问题在于空间信息丢失与位置敏感性不足。具体表现为:1)空间结构破坏:MLP 需将 2D 特征图展平为 1D 向量,导致局部空间关联性被破坏。假设输入特征维度为  $C \times H \times W$ ,展平后为  $C \times (W \cdot H)$ ,此时相邻像素的几何关系无法有效保留;2)位置先验缺失:传统方法依赖隐式学习空间映射,缺乏显式位置编码,难以建模复杂透视变换。对于远距离或遮挡区域,特征位置歧义性显著增加。

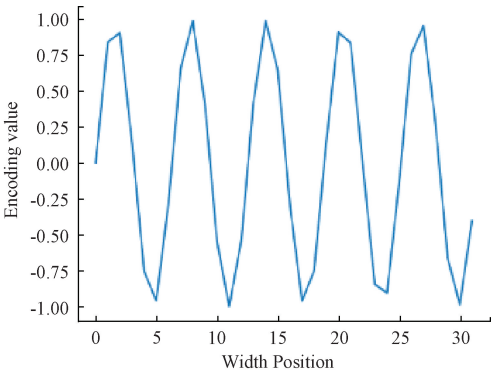
本研究提出一种复合位置编码器 (composite positional encoding,CPE),通过双通道正交编码分离横向与纵向位置特征,并通过增加通道维度,利用指数项实现多频带编码,由低频通道编码宏观布局,高频通道捕获局部细节,来解决上述问题。数学表达式如下:

$$CPE(c,h,w)=\begin{cases} \sin\left(\frac{w}{10\,000^{\frac{2c}{d}}}\right), & c\bmod 2=0 \\ \cos\left(\frac{h}{10\,000^{\frac{2(c-1)}{d}}}\right), & c\bmod 2=1 \end{cases} \quad (1)$$

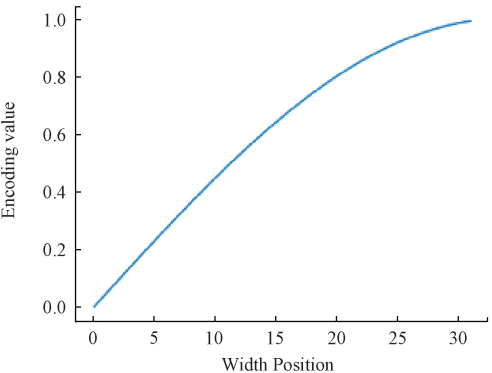
式中:  $d$  表示总通道数,  $c$  表示当前通道,  $h$  和  $w$  表示空间坐标,  $10\,000^{\frac{2c}{d}}$  为指数项。

如图 2 所示,对比了不同频率的编码模式,高频通道(如 Channel 0)的编码值在宽度方向快速震荡,适用于捕捉细粒度位置特征,低频通道(如 Channel 10)的编码值变化缓慢,用于建模全局位置关系。这种设计不仅保证了不同方向位置信息的正交性,还通过空间维度保持机制避免了

传统展平处理导致的空间信息丢失。



(a) 高频模式(Channel 0)  
(a) High frequency pattern (Channel 0)



(b) 低频模式(Channel 0)  
(b) Low frequency pattern (Channel 0)

图 2 不同编码模式展示

Fig. 2 Display of different encoding



之后,将原始前视图特征  $F_{img} \in \mathbf{R}^{C \times H \times W}$  与位置编码  $CPE \in \mathbf{R}^{C \times H \times W}$  沿通道维度拼接,得到增强特征  $F_{enhanced} = \text{Concat}(F_{img}, CPE) \in \mathbf{R}^{2C \times H \times W}$ ,随后通过两阶段全连接层实现 BEV 空间投影。

### 1.3 特征融合模块

在多尺度前视图特征经过 CPE 处理并通过基于 MLP 的 BEV 空间转换模块后,获得了多个含有大量位置信息的 BEV 特征。由大尺度前视图特征转换得到的 BEV 特征携带精确的局部信息,小尺度前视图特征转换后的特征会提供广泛的上下文信息。本文通过特征融合模块(temporal fusion attention module, TFAM)将携带不同信息的 BEV 特征进行融合,以此来获得更多的空间信息。

使用基于通道-空间双分支的特征融合模块 TFAM<sup>[19]</sup>实现了 BEV 特征的高效融合。TFAM 结构如图 3 所示,

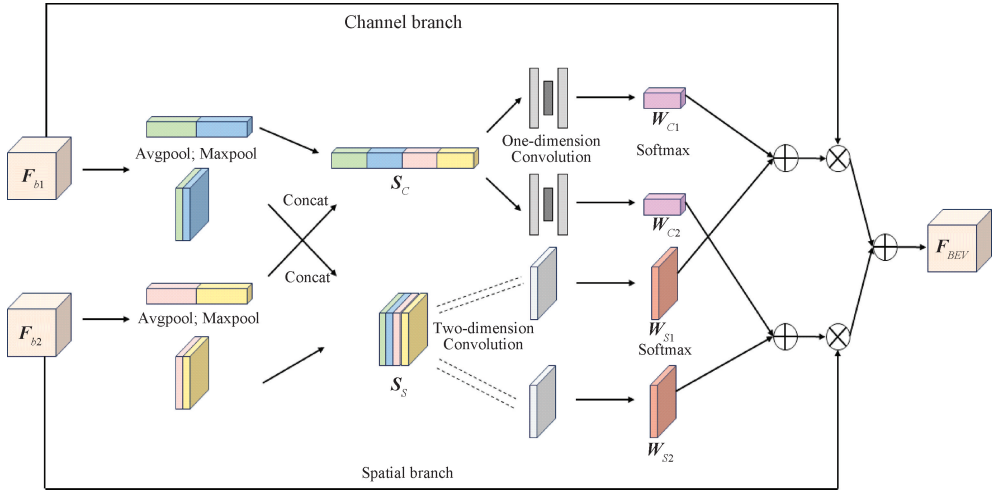


图 3 TFAM 结构图

Fig. 3 TFAM structure diagram

聚合后的通道特征和空间特征通过两个卷积得到不同 BEV 特征的权重,并通过 Softmax 使权重之和为 1。最后,将通道和空间权重与 BEV 特征相乘并进行汇总,以有效地融合 BEV 特征。数学表达式如下:

$$\begin{cases} W_{C1} = \text{Softmax}(\text{Conv}_1(S_c)) \\ W_{C2} = \text{Softmax}(\text{Conv}_2(S_c)) \end{cases} \quad (4)$$

$$\begin{cases} W_{S1} = \text{Softmax}(\text{Conv}_1(S_s)) \\ W_{S2} = \text{Softmax}(\text{Conv}_2(S_s)) \end{cases} \quad (5)$$

$$F_{BEV} = (W_{C1} + W_{S1}) \times F_{b1} + (W_{C2} + W_{S2}) \times F_{b2} \quad (6)$$

式中:  $F_{BEV}$  为输出的融合 BEV 特征,  $W_{C1}$ 、 $W_{C2}$  为通道分支得到的两个 BEV 特征的通道权重矩阵,  $W_{S1}$  和  $W_{S2}$  为空间权重矩阵。

### 1.4 全局注意力机制

本文对融合后获得的多个 BEV 特征进行了拼接操作。拼接后的特征包含了整个场景的上下文信息,增加了特征的丰富性。但与此同时,在拼接多个 BEV 特征后会导致特征表示的模糊性,模型可能难以区分哪些特征更为

该模块针对输入特征张量分别构建通道注意力分支和空间注意力分支并进行池化操作。通过池化操作,实现了对全局空间信息的压缩和提取,从而更好的捕捉输入特征图的重要信息<sup>[20]</sup>。之后通过聚合,分别生成通道特征和空间特征。聚合的特征保留了两个特征中更有价值的部分,同时抑制了不重要或误导性的信息。其数学解析式如式(2)和式(3)所示。

$$S_c = \text{Concat}(\text{Avg}(F_{b1}), \text{Max}(F_{b1}), \text{Avg}(F_{b2}), \text{Max}(F_{b2})) \quad (2)$$

$$S_s = \text{Concat}(\text{Avg}(F_{b1}), \text{Max}(F_{b1}), \text{Avg}(F_{b2}), \text{Max}(F_{b2})) \quad (3)$$

式中:  $F_{b1} \in \mathbf{R}^{C \times H \times W}$  和  $F_{b2} \in \mathbf{R}^{C \times H \times W}$  为输入特征,  $\text{Avg}(\cdot)$  和  $\text{Max}(\cdot)$  分别表示全局平均池化和全局最大池化,  $S_c$  和  $S_s$  分别表示生成的通道特征和空间特征。

重要。为了解决这个问题,本文考虑并使用全局注意力机制<sup>[21]</sup>(global attention mechanism, GAM),全局注意力结构如图 4 所示。全局注意力机制能够自动学习不同特征之间的依赖关系,并为每个特征分配一个权重,表示其在特定任务中的重要性。同时,全局注意力机制考虑到跨维交互的重要性,能够捕获通道、空间宽度和空间高度 3 个维度的显著特征,进一步增强空间信息。整个流程如式(7)和(8)中所示。

$$F_{mid} = M_C(F_{BEV}) \otimes F_{BEV} \quad (7)$$

$$F'_{BEV} = M_S(F_{mid}) \otimes F_{mid} \quad (8)$$

式中:  $F_{BEV} \in \mathbf{R}^{C \times H \times W}$  为输入特征,  $F_{mid}$  为中间特征、 $F'_{BEV}$  为输出特征,  $M_C$ 、 $M_S$  分别表示全局注意力机制内的通道注意力子模块和空间注意力子模块,  $\otimes$  为元素之间的乘法。

### 1.5 因果干预机制

在复杂动态交通场景中,车道线检测模型易受环境混杂因子(如光照变化、天气条件等)的干扰,导致特征表达



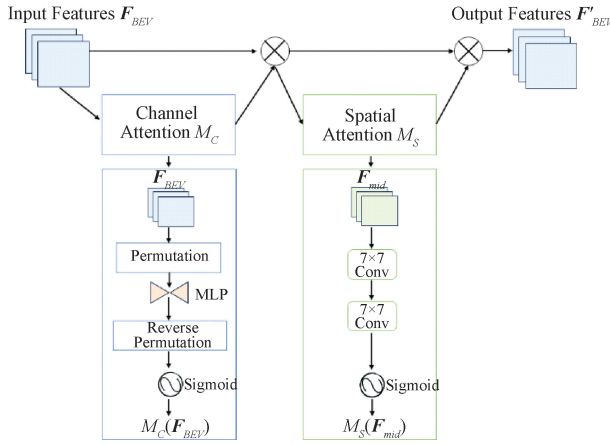


图 4 全局注意力结构图

Fig. 4 Global attention structure diagram

与车道线语义间存在虚假相关性。

本文提出一种轻量化因果干预模块（causal intervention module, CIM），通过环境因子估计与反事实特征生成，实现特征空间的环境因子解耦，流程如图 5 所示。该方法基于 Pearl<sup>[22]</sup> 的因果推断理论框架，建立结构化因果模型，定义观测变量三元组  $(\mathbf{F}'_{BEV}, \mathbf{Y}, \mathbf{E})$ ，其中  $\mathbf{F}'_{BEV} \in \mathbf{R}^{C \times H \times W}$  为输入特征图， $\mathbf{Y}$  为车道线预测结果， $\mathbf{E} \in \mathbf{R}^{16}$  为潜在环境因子。其因果结构可表示为  $\mathbf{E} \rightarrow \mathbf{F}'_{BEV} \leftarrow \mathbf{Y}$ ，表明环境因子  $\mathbf{E}$  作为混杂变量，同时影响特征  $\mathbf{F}'_{BEV}$  和预测结果  $\mathbf{Y}$ 。传统方法直接学习  $P(\mathbf{Y} | \mathbf{F}'_{BEV})$  会继承  $\mathbf{E}$  的偏差，而本方法通过干预  $\text{do}(\mathbf{F}'_{BEV} = \mathbf{F}^{cf}_{BEV})$ ，阻断  $\mathbf{E} \rightarrow \mathbf{F}'_{BEV}$  的后门路径。具体实现中，首先通过环境因子编码器预估环境因子：

$$\mathbf{E} = \psi_{\theta}(\mathbf{F}'_{BEV}) = \text{MLP}(\text{GAP}(\mathbf{F}'_{BEV})) \quad (9)$$

式中： $\psi_{\theta}$  为环境因子编码器， $\text{GAP}(\cdot)$  为全局平均池化， $\text{MLP}(\cdot)$  由两个全连接层组成，输出 16 维环境因子  $\mathbf{E}$ 。

随后，通过反事实生成器生成环境因子干预要素  $\mathbf{e}_0$ ，用于干预环境因子。最后构建反事实特征，模拟若环境因子被干预时的特征表达，迫使模型忽略混杂的环境因子，直接建模车道线本质属性与预测结果  $\mathbf{Y}$  的因果关联。数学表达式如下：

$$\mathbf{F}^{cf}_{BEV} = \mathbf{F}'_{BEV} + \Delta \mathbf{F}'_{BEV} = \mathbf{F}'_{BEV} + \phi_{\omega}(\mathbf{E}) \quad (10)$$

式中： $\phi_{\omega}$  为反事实生成器， $\mathbf{F}^{cf}_{BEV}$  为生成的反事实特征。

为防止环境因子编码器  $\psi_{\theta}$  泄露车道线语义信息，引入对抗训练约束，通过梯度反转层（如式(11)所示）迫使  $\mathbf{F}^{cf}_{BEV}$  的环境因子不可辨识。具体而言：环境编码器  $\psi_{\theta}$  的目标是尽可能准确地估计环境因子  $\mathbf{E}$ ，而 GRL 通过梯度反转，使得  $\mathbf{F}^{cf}_{BEV}$  的优化方向与  $\psi_{\theta}$  的目标相反，即  $\mathbf{F}^{cf}_{BEV}$  应尽可能不包含  $\mathbf{E}$  的信息。这种对抗机制迫使  $\mathbf{F}^{cf}_{BEV}$  与  $\mathbf{E}$  独立，即  $\mathbf{F}^{cf}_{BEV} \perp \mathbf{E}$ 。环境因子损失函数定义为式(12)。

$$\text{GRL}(\mathbf{F}^{cf}_{BEV}) = \mathbf{F}^{cf}_{BEV}, \quad \frac{\partial \text{GRL}}{\partial \mathbf{F}^{cf}_{BEV}} = -\lambda \mathbf{I} \quad (11)$$

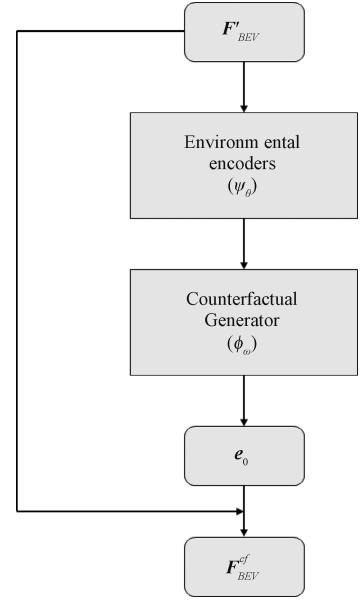


图 5 因果干预模块流程图

Fig. 5 Causal intervention module flowchart

$$L_E = E_x [\|\psi_{\theta}(\mathbf{F}^{cf}_{BEV})\|_2^2] \quad (12)$$

## 1.6 损失函数

本文在车道表示上沿用了 BEV-LaneDet 的车道表示方法，按照 BEV-LaneDet 的关键点表示法预测 BEV 网格的置信度、嵌入、偏移量以及高度，最后一起用作 3D 车道表示<sup>[16]</sup>。

车道置信度分支采用的是二元交叉熵损失函数，本文在 BEV 面上划分了  $c_1 \times c_2$  个单元格，通过预测每个单元格的置信度来判断此单元格内是否存在车道。如果置信度预测值为 1，认为此单元格内存在车道；如果置信度为 0，则认为车道不存在。损失函数表达如下：

$$L_{conf}^{3d} = \sum_i^{c_1 \times c_2} (\hat{p}_i \log p_i + (1 - \hat{p}_i) \log(1 - p_i)) \quad (13)$$

式中： $p_i$  表示预测的置信度的概率， $\hat{p}_i$  为通过 BEV 平面的分割图获得的真实值。

置信度分支只能判断是否存在车道，而无法知道每个单元格内的车道的身份，为此设计了嵌入度分支来区分车道。

$$L_{embed}^{3d} = \text{push\_weight} \times L_{push}^{3d} + \text{pull\_weight} \times L_{pull}^{3d} \quad (14)$$

式中：嵌入损失中包含  $\text{pull\_loss}$  和  $\text{push\_loss}$ ，即方差损失和距离损失。方差损失是不同车道的单元格之间嵌入度的方差平均值；距离损失是计算同一个车道的所有单元格之间嵌入度的平均值。

在有了置信度和嵌入度后，还需要去预测每个单元格内的车道的位。通过预测从单元格中心到车道所在位置的精确偏移量  $\Delta y$ ，可以得到  $y$  方向上的地面坐标。损失函数表达如下：

$$L_{offset}^{3d} = \sum_i^{c_1 \times c_2} 1_{obj} (\sigma(\Delta y_i) - \Delta \hat{y}_i) \quad (15)$$

式中:  $l_{obj}$  表示车道是否通过此单元格。 $\Delta y_i$  表示与预测的偏移量,  $\Delta \hat{y}_i$  表示与实际的偏移量。

通过置信度,嵌入度和偏移量只能获得在  $z=0$  情况下的坐标  $(x,y)$  想要获得三维车道线的位置  $(x,y,z)$ , 还需要车道的高度信息。直接预测每个单元格内的车道的高度, 并与真实的高度一起进行损失计算。

$$L_z^{3d} = \frac{1}{c_1 \times c_2} \sum_i^{c_1 \times c_2} (z_i - \hat{z}_i)^2 \tag{16}$$

式中:  $z_i$  为预测的车道高度,  $\hat{z}_i$  为真实的车道高度。

除了车道损失和对抗损失外, 还计算了前视图损失用来辅助监督。

$$L_{total} = w_E L_E + w_{conf}^{3d} L_{conf}^{3d} + w_{embed}^{3d} L_{embed}^{3d} + w_{offset}^{3d} L_{offset}^{3d} + w_z^{3d} L_z^{3d} + w_{seg}^{2d} L_{seg}^{2d} + w_{embed}^{2d} L_{embed}^{2d} \tag{17}$$

式中: 三维车道损失、环境因子损失与前视图损失之和即为总损失,  $w$  为各损失函数的权重。

## 2 实验与分析

### 2.1 数据集与评价指标

本文在 Apollo 3D 虚拟数据集和 OpenLane 数据集上对模型性能进行了评估。为了确保实验结果的可靠性, 采

用 Gen-LaneNet 的评估指标来验证上述的两个数据集, 包含不同场景下的 F1-Score 和 X/Z 误差, 实验中的最优结果以加粗格式标注, 次优结果以加下划线格式呈现。

Apollo 数据集是利用 Unity 游戏引擎渲染的虚拟数据集, 其中包含了 3 个场景: 路面平坦的场景, 很少被注意到的场景以及存在视觉变化的场景。数据集包括 10 500 张车道图片以及对应的车道真实数据。

OpenLane 数据集是目前最大的真实世界 3D 车道数据集, 包含 15 万个训练帧, 4 万个测试帧, 14 个车道类别以及场景标签和路线邻近目标注释。数据集包含 6 个场景, 分别为 Curve、Intersection、Merge\_split、Extreme\_weather、Night、Up\_down。

### 2.2 消融实验与分析

为了验证因果干预模块(CIM)、复合位置编码(CPE)、特征融合模块(TFAM)和全局注意力机制(GAM)的有效性, 本文在 Apollo 测试集上开展系统性消融实验, 结果如表 1 所示。实验以 EfficientNet-B0 为基线主干网络, 通过逐步叠加上述 4 个核心模块, 定量分析各组件对检测性能的贡献度。分析表 1 可以看出, 当单独使用各模块时 F1 分数均有所上升, 同时使用 4 个模块时, F1 分数达到了最好的指标, X error 也降至最低, 验证了各模块的有效性。

表 1 消融实验  
Table 1 Ablation experiment

EfficientNet-B0	CIM	CPE	TFAM	GAM	F1-Score	X error
✓	—	—	—	—	96.90	0.046/0.300
✓	✓	—	—	—	97.50(+0.6)	0.045/0.279
✓	—	✓	—	—	97.97(+1.07)	0.031/0.25
✓	—	—	✓	—	97.42(+0.52)	0.040/0.280
✓	—	—	—	✓	97.22(+0.32)	0.034/0.25
✓	✓	✓	—	—	98.02(+0.52)	0.030/0.248
✓	✓	—	✓	—	97.69(+0.79)	0.030/0.269
✓	✓	—	—	✓	97.96(+1.06)	0.034/0.281
✓	✓	✓	✓	✓	<b>98.20(+1.3)</b>	<b>0.028/0.249</b>

注: 加粗字体为每列最优值, ✓ 表示采用, — 表示未采用

同时, 为了验证多尺度前视图特征转换为多个 BEV 特征后, 不同 BEV 特征相互融合对模型精度的影响, 本文添加了表 2。其中  $b_{32}$  表示输入图像的 32 倍下采样, 通过空间转化获得的 BEV 特征, + 表示特征融合,  $b_{32} + b_{64}$ ,  $b_{64} + b_{128}$  表示两个融合后的特征进行拼接。

实验表明, 在多尺度前视图特征转换成 BEV 特征后, 对  $b_{32}$ ,  $b_{64}$  和  $b_{128}$  进行两两融合, 并将这两组融合特征进行拼接, 可以获得最优的模型性能。这种组合效果最佳的原因在于:  $b_{32}$  (全局语义强, 空间细节弱)、 $b_{64}$  (语义与细节平衡) 和  $b_{128}$  (局部细节强, 语义弱) 这 3 个关键尺度通过  $b_{64}$  形成连续的信息链。  $b_{32} + b_{64}$  融合提供了全局上下文与中等细节的互补,  $b_{64} + b_{128}$  融合则实现了中等语义对局部细节的有效支撑。最终的拼接操作整合了这条从全局到局部的完整且连贯的信息流 ( $b_{32} \rightarrow b_{64} \rightarrow b_{128}$ ), 覆盖了感知任

表 2 融合不同 BEV 特征带来的影响  
Table 2 The impact of fusing different BEV features

Different Combinations	F1-Score	X error	X error
		near	far
$b_{32} + b_{64}$	97.77	<u>0.035</u>	0.284
$b_{64} + b_{128}$	97.27	0.036	<u>0.250</u>
$b_{32} + b_{64}$ , $b_{32} + b_{64}$	95.37	0.044	0.399
$b_{32} + b_{64}$ , $b_{64} + b_{128}$	<b>98.2</b>	<b>0.028</b>	<b>0.249</b>
$b_{32} + b_{64}$ , $b_{128} + b_{256}$	97.95	0.028	0.297

务所需的最有效信息范围。相比之下, 其他组合要么信息覆盖不全 (如仅用单一组合  $b_{32} + b_{64}$  或  $b_{64} + b_{128}$ ), 要么存在尺度断裂或信息冗余 (如  $b_{32} + b_{64}$ ,  $b_{128} + b_{256}$  缺少  $b_{64} \rightarrow b_{128}$  的过渡,  $b_{32} + b_{64}$ ,  $b_{32} + b_{64}$  则是重复融合)。

本文的方法聚焦于为复杂环境(如雨雾、雪天)下的自动驾驶感知提供高鲁棒性解决方案,为了验证本文改进模块在复杂环境下的表现效果,本文在 OpenLane 数据集内的极端天气和夜间这两个具有挑战性的场景进行了消融实验,实验以 F1 分数为指标进行评估,Base 为基础模型未添加改进模块,详情如表 3 所示。实验表明,在面对复杂场景时,因果干预机制通过解耦环境干扰和车道线语义,有效抑制环境混淆因子影响,提升车道检测鲁棒性。

表 3 复杂场景下各模块效果对比

Table 3 Module efficacy in complex scenes		
Method	Extreme weather	Night
Base	53.4	53.4
Base with CIM	57.5	55.8
Base with CPE	56.1	54.9
Base with TFAM	54.0	53.7
Base with GAM	54.3	54.2

2.3 模型对比结果

为了验证模型的效果,本文在 Apollo 数据集和 OpenLane 数据集上进行了实验与前人的实验结果进行比较,包括 PersFormer<sup>[1]</sup>、CLGO<sup>[23]</sup>、CurveFormer<sup>[8]</sup>、BEV-LaneDet<sup>[16]</sup>、LaneCPP<sup>[17]</sup>等。与上述方法类似,本文

方法同样在三维空间进行车道线检测,并致力于解决前视图到 BEV 的视角转换问题。区别于上述方法,本文核心创新在于引入因果干预机制,显式建模并剥离环境干扰与车道语义的虚假关联,从而在极端复杂环境下实现更鲁棒、稳定的检测性能。在对比实验的性能指标方面,F1-Score 是精确率和召回率的调和平均数,综合反映模型整体性能<sup>[24]</sup>;X error near 和 X error far 分别表示车辆近处和远处车道线检测结果在水平方向上的误差,影响车辆的短期决策控制和远方道路状况预判规划;Z error near 和 Z error far 则分别代表车辆近处和远处车道线检测结果在高度方向上的误差,对车辆的高度控制、远方道路立体感知和提前决策至关重要。通过对比这些性能指标,可充分验证本文模型在准确性和鲁棒性等方面的优势和先进性。

实验结果表明,本文方法在 Apollo 数据集上取得了最好的效果,在 OpenLane 数据集的极端环境,夜晚等子集上也有着最好的发挥。表 4 对比了本文方法与其他模型在 Apollo 数据集不同场景下的性能表现。结果显示,本文方法在 F1 分数上全面优于基线方法,尤其在存在视觉误差的场景中提升显著。表 5 进一步表明,不管是在采用 EfficientNet-B0 主干网络的情况下,还是使用基础模型 BEV-LaneDet 的 ResNet-34 主干网络时,本文模型相较 BEV-LaneDet 在 F1 分数上均显著提升,验证架构改进有效性。

表 4 模型在 Apollo 数据集的不同场景下进行实验对比

Table 4 Experimental comparison of models in different scenarios on the Apollo dataset						
Scene	Method	F1-Score	X error near	X error far	Z error near	Z error far
Balanced Scenes	CLGO <sup>[23]</sup>	91.9	0.061	0.361	0.029	0.250
	PersFormer <sup>[1]</sup>	92.2	0.054	0.356	<b>0.010</b>	0.234
	CurveFormer <sup>[8]</sup>	95.8	0.078	0.326	0.018	0.219
	BEV-LaneDet <sup>[16]</sup>	96.9	<b>0.016</b>	<b>0.242</b>	0.02	0.216
	Anchor3Dlane <sup>[7]</sup>	95.4	0.045	0.300	0.016	0.223
	LaneCPP <sup>[17]</sup>	<u>97.4</u>	0.030	0.277	<u>0.011</u>	<b>0.206</b>
	Ours	<b>98.15</b>	<u>0.028</u>	<u>0.249</u>	0.018	<u>0.213</u>
Rarely Observed	CLGO <sup>[23]</sup>	86.1	0.147	0.735	0.071	0.609
	PersFormer <sup>[1]</sup>	87.5	0.107	0.782	<u>0.024</u>	0.602
	CurveFormer <sup>[8]</sup>	95.6	0.182	0.737	0.039	0.561
	BEV-LaneDet <sup>[16]</sup>	<u>97.6</u>	<b>0.031</b>	<b>0.594</b>	0.040	0.556
	Anchor3Dlane <sup>[7]</sup>	94.4	0.082	0.699	0.030	0.580
	LaneCPP <sup>[17]</sup>	96.2	0.073	0.651	<b>0.023</b>	<u>0.543</u>
	Ours	<b>98.59</b>	<u>0.054</u>	<u>0.605</u>	0.033	<b>0.542</b>
Visual Variations	CLGO <sup>[23]</sup>	87.3	0.084	0.464	0.045	0.312
	PersFormer <sup>[1]</sup>	89.6	0.074	0.430	<b>0.015</b>	0.266
	CurveFormer <sup>[8]</sup>	90.8	0.125	0.410	0.028	0.254
	BEV-LaneDet <sup>[16]</sup>	<u>95.0</u>	<u>0.027</u>	<u>0.320</u>	0.031	0.256
	Anchor3Dlane <sup>[7]</sup>	91.8	0.047	0.327	0.019	<b>0.219</b>
	LaneCPP <sup>[17]</sup>	90.4	0.054	0.327	0.020	<u>0.222</u>
	Ours	<b>96.26</b>	<b>0.026</b>	<b>0.263</b>	<u>0.017</u>	0.226



表 5 相同主干网络下的实验对比

Table 5 Experimental comparison under the same backbone

Backbone	Scene	Ours	BEV-LaneDet
ResNet-34	Balanced Scenes	97.95	96.9
	Rare Observed	98.47	97.6
	Visual Variations	95.57	95.0
EfficientNet-B0	Balanced Scenes	98.15	97.8
	Rare Observed	98.59	97.7
	Visual Variations	96.46	95.7

表 6 模型在 Openlane 数据集的不同场景下进行实验对比

Table 6 Experimental comparison of models in different scenarios on the Openlane dataset

方法	Up & Down	Curve	Extreme Weather	Night	Intersection	Merge & Split
3D-LaneNet <sup>[2]</sup>	40.8	46.5	47.5	41.5	32.1	41.7
Gen-LaneNet <sup>[4]</sup>	25.4	33.5	28.1	18.7	21.4	31.0
PersFormer <sup>[1]</sup>	42.4	55.6	48.6	46.6	40.0	50.7
CurveFormer <sup>[8]</sup>	45.2	56.6	49.7	49.1	42.9	45.4
BEV-LaneDet <sup>[16]</sup>	48.7	63.1	53.4	53.4	50.3	<u>53.7</u>
Anchor3Dlane <sup>[7]</sup>	46.7	57.2	52.2	47.8	45.4	51.2
LaneCPP <sup>[17]</sup>	<u>53.6</u>	<u>64.4</u>	56.7	<u>54.9</u>	52.0	<b>58.7</b>
PVALane <sup>[6]</sup>	52.6	<b>65.7</b>	<u>59.5</u>	<b>56.5</b>	<u>52.2</u>	<b>58.7</b>
Ours	<b>56.4</b>	59.1	<b>60.2</b>	<b>56.5</b>	<b>56.8</b>	51.0

表 7 模型在 Openlane 数据集不同场景下的详细评估

Table 7 Detailed evaluation of models in different scenarios on the Openlane dataset

场景	F1/% ↑ ↓	P/% ↑	R/% ↑	X-error/m ↓	Z-error/m ↓
				near/far	near/far
Up & Down	56.4	52.1	61.5	0.231/0.911	0.207/0.912
Curve	59.1	55.6	63.2	0.318/0.790	0.215/0.654
Extreme Weather	60.2	53.5	68.8	0.198/0.679	0.185/0.674
Night	56.5	46.8	71.4	0.246/0.591	0.254/0.596
Intersection	56.8	48.1	69.4	0.419/0.645	0.427/0.686
Merge & Split	51.0	45.9	57.5	0.395/0.664	0.324/0.589

2.4 可视化实验

车道线检测可视化实验在 V100 GPU + Pytorch 1.10 环境下进行,使用 Apollo 数据集和 Openlane 数据集进行测试。车道线拟合关键参数设定包含聚类分割图最小置信度为 0.9、不同集群嵌入边际为 6、集群最小点数为 15。在本文模型可视化结果中:如图 6 所示,Apollo 数据集展示图 6(a)白天,图 6(b)、(c)黄昏直线、曲线,图 6(d)、(e)、(f)夜间直线、坡道、曲线场景的检测结果;如图 7 所示,

在 Openlane 数据集的 6 个场景中,本文将模型与近几年的优化方法的 F1 分数进行对比,F1 分数越高,召回率和回归率越平衡,模型的性能越好,结果如表 6 所示。本文模型在 Up & Down、Extreme Weather、Night、Intersection 场景表现最优,在 Curve、Merge & Split 场景中表现还有待提升。为了进一步验证本文模型在不同场景下的性能,本文对 Openlane 数据集的不同场景进行了详细评估,包含 F1 分数、精确率(P)、召回率(R)和几何误差,结果如表 7 所示。实验结果表明,通过因果干预、复合位置编码等方法可以有效提升模型在极端天气、夜间等场景下的车道识别能力。

Openlane 数据集呈现图 7(a)强光、图 7(b)阴影、图 7(c)夜间、图 7(d)雨天场景的检测结果。所有可视化结果均来自本文模型且提供 Front View(原图)、BEV(鸟瞰图)、3D(三维空间)三视角对比。从视觉连续性评估指标来看,Apollo 数据集的可视化结果全程保持连续平滑;而 Openlane 数据集在强光与夜间场景下虽存在局部中断,但整体连续性仍得到有效维持。可视化结果表明,本文的方法在有复杂环境因素干扰的场景下可以准确检测出车道线。

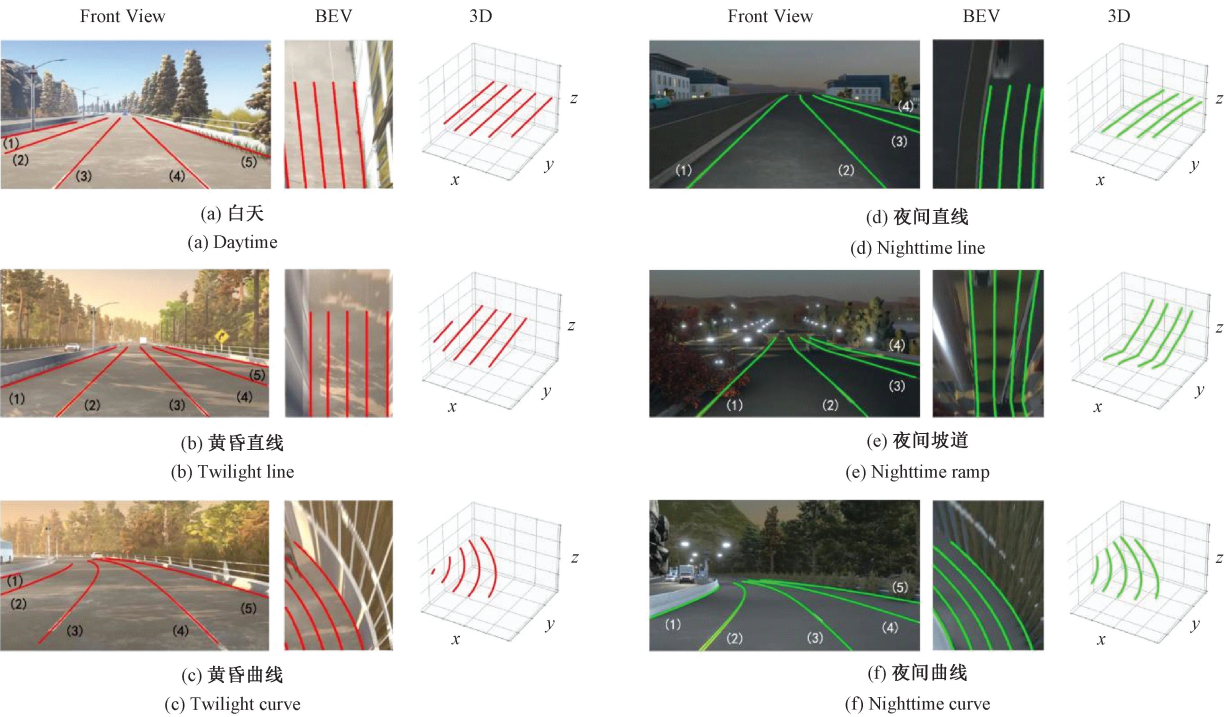


图 6 模型在 Apollo 数据集的检测结果

Fig. 6 Detection results by the model on the Apollo dataset

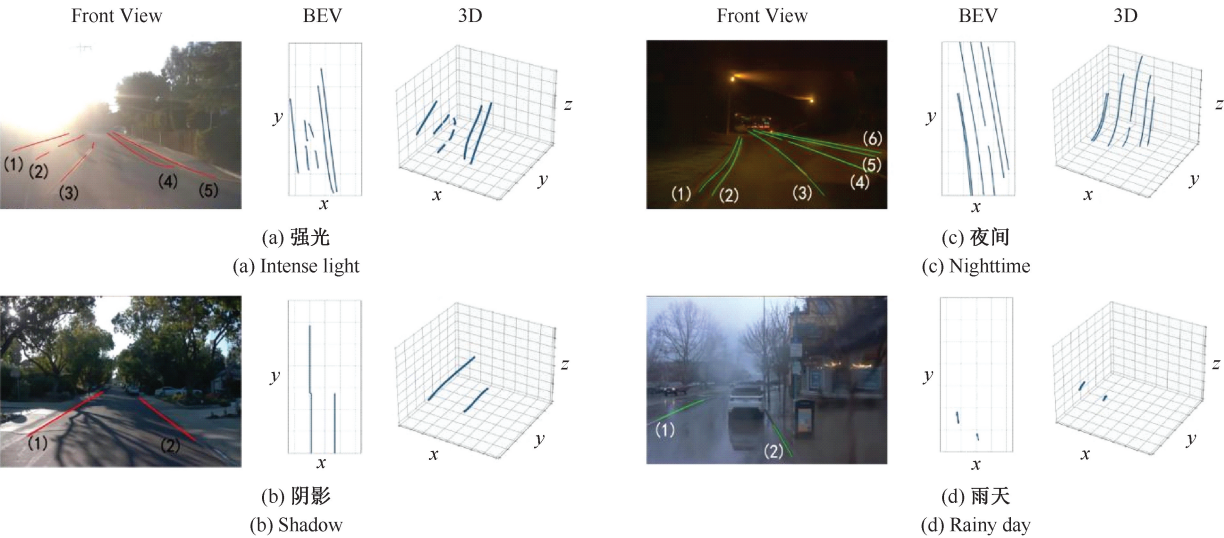


图 7 模型在 Openlane 数据集的检测结果

Fig. 7 Detection results by the model on the Openlane dataset

3 结 论

本文提出了一种鲁棒的 BEV 车道线检测方法。通过复杂位置编码解决了基于 MLP 的 BEV 空间转换方法中存在的空间信息丢失问题,并采用特征融合以及全局注意力机制对 BEV 特征进行增强。除此以外,设计的因果干预模块消除了混杂的环境因子对特征的影响。与近年来提出的车道线检测方法相比,本文模型在最具代表性的两

个 3D 车道线检测数据集中具有复杂环境的子集上取得了最好的效果。综上所述,本文方法在应对复杂环境的车道线检测任务中展现了卓越的性能。虽然在大曲率弯道等特定场景下的检测效果尚有改进空间,但其在复杂环境中的车道线检测能力已得到验证。未来工作将致力于算法的进一步优化,以期实现更加精准和鲁棒的车道线检测。

参考文献

[1] CHEN L, SIMA CH H, LI Y, et al. Persformer: 3D

- lane detection via perspective transformer and the openlane benchmark [C]. European Conference on Computer Vision, Cham: Springer Nature Switzerland, 2022: 550-567.
- [2] GARNETT N, COHEN R, PE'ER T, et al. 3D-lanenet: End-to-end 3D multiple lane detection[C]. IEEE/CVF International Conference on Computer Vision, 2019: 2921-2930.
  - [3] PHILION J, FIDLER S. Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3D[C]. Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XIV 16. Springer International Publishing, 2020: 194-210.
  - [4] GUO Y L, CHEN G, ZHAO P T, et al. Gen-LaneNet: A generalized and scalable approach for 3D lane detection [C]. Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XXI 16. Springer International Publishing, 2020: 666-681.
  - [5] LI ZH Q, WANG W H, LI H Y, et al. BEVFormer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers[J]. ArXiv preprint arXiv:2203.17270, 2022.
  - [6] ZHENG Z W, ZHANG X M, MOU Y Q, et al. PVALane: Prior-guided 3D lane detection with view-agnostic feature alignment[C]. AAAI Conference on Artificial Intelligence, 2024, 38(7): 7597-7604.
  - [7] HUANG SH F, SHEN ZH W, HUANG Z H, et al. Anchor3DLane: Learning to regress 3D anchors for monocular 3D lane detection [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 17451-17460.
  - [8] BAI Y F, CHEN ZH R, FU ZH J, et al. CurveFormer: 3D lane detection by curve propagation with curve queries and attention [C]. 2023 IEEE International Conference on Robotics and Automation (ICRA), 2023: 7062-7068.
  - [9] BIJELIC M, GRUBER T, MANNAN F, et al. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11682-11692.
  - [10] WANG Z, REN W Q, QIU Q. LaneNet: Real-time lane detection networks for autonomous driving[J]. ArXiv preprint arXiv:1807.01726, 2018.
  - [11] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks [J]. Communications of the ACM, 2020, 63 (11): 139-144.
  - [12] PRANJAL, VAISHNAVI, DIVYANSH, et al. Adversarial attacks on neural networks [C]. International Conference on Cyber Intelligence and Information Retrieval. Singapore: Springer Nature Singapore, 2023: 409-419.
  - [13] ROSENFELD E, RAKIKUMAR P, RISTESKI A. The risks of invariant risk minimization[J]. ArXiv preprint arXiv:2010.05761, 2020.
  - [14] KOCAOGLU M, SNYDER C, DIMAKIS A G, et al. CausalGAN: Learning causal implicit generative models with adversarial training[J]. ArXiv preprint arXiv:1709.02023, 2017.
  - [15] LOUZIOS C, SHALIT U, MOOIJ J M, et al. Causal effect inference with deep latent-variable models[J]. Advances in Neural Information Processing Systems, 2017, 30, DOI:10.48550/arXiv.1705.08821.
  - [16] WANG R H, QIN J, LI K Y, et al. BEV-LaneDet: A simple and effective 3D lane detection baseline[J]. ArXiv preprint arXiv:2210.06006, 2022.
  - [17] PITTLNER M, JANAI J, CONDURACHE A P. Lanecpp: Continuous 3D lane detection using physical priors [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 10639-10648.
  - [18] TAN M X, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C]. International Conference on Machine Learning. PMLR, 2019: 6105-6114.
  - [19] ZHAO S J, ZHANG X L, XIAO P F, et al. Exchanging dual-encoder-decoder: A new strategy for change detection with semantic guidance and spatial localization[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-16.
  - [20] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. European Conference on Computer Vision(ECCV), 2018: 3-19.
  - [21] LIU Y CH, SHAO Z R, HOFFMANN N. Global attention mechanism: Retain information to enhance channel-spatial interactions[J]. ArXiv preprint arXiv:2112.05561, 2021.



[22] PEARL J, GLYMOUR M, JEWELL N P. Causal inference in statistics; A primer[M]. Hoboken:John Wiley & Sons, 2016.

[23] LIU R J, CHEN D P, LIU T, et al. Learning to predict 3D lane shape and camera pose from a single image via geometry constraints[C]. AAAI Conference on Artificial Intelligence, 2022, 36(2): 1765-1772.

[24] 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 backbone [J]. Electronic Measurement Technology, 2024, 47(19):164-171.

作者简介

李睿豪,硕士研究生,主要研究方向为计算机视觉、目标检测。

E-mail:1007298391@qq.com

于红绯(通信作者),博士,副教授,主要研究方向为计算机视觉、图像处理。

E-mail:yuhongfei@lnpu.edu.cn