

APDP-Net: 基于区域注意力的轻量级全景驾驶感知网络

郎愚迟¹, 郭 迟^{1†}, 詹 骄²

1. 武汉大学 电子信息学院, 湖北 武汉 430072;
2. 武汉大学 卫星导航定位技术研究中心, 湖北 武汉 430072

收稿日期: 2025-04-16 † 通信联系人 E-mail: guochi@whu.edu.cn

基金项目: 湖北省重大科技专项(2022AAA009)

第一作者: 郎愚迟, 男, 硕士生, 主要从事计算机视觉相关研究。E-mail: 2023206550003@whu.edu.cn

摘 要: 全景驾驶感知旨在同时完成交通目标检测、可行驶区域分割与车道线检测三种感知任务, 以满足自动驾驶对于可通行性预测的需求。现有方法通常采用多任务学习网络同时处理多个任务, 虽在性能上取得了一定进展, 但普遍存在计算复杂度高、预测精度与实时性不平衡等问题。针对这些问题, 提出一种基于区域注意力的轻量级多任务全景驾驶感知网络: APDP-Net。首先, 引入残差高层次聚合网络 R-ELAN 作为主干网络, 该网络中的区域注意力模块以局部、连续注意力机制为特点, 在保持感受野的同时降低了计算成本; 其次, 设计特征金字塔结构的 GhostFPN-ELAN 颈部网络, 引入幽灵卷积, 通过合理压缩特征空间减少了计算量; 最后, 针对车道线检测任务, 引入极化自注意力机制, 建立了像素级特征间的长距离联系, 增强了网络的空间特征表达能力。实验结果表明, APDP-Net 通过轻量化网络设计, 在显著降低计算开销的同时保持了出色的预测性能, 为端侧自动驾驶感知提供了切实可行的解决方案。

关 键 词: 多任务学习; 全景驾驶感知; 区域注意力; 自动驾驶

中图分类号: TP391.4

文献标志码: A

文章编号: 1671-8836(2026)04-0507-12

APDP-Net: Area Attention-Based Lightweight Panoptic Driving Perception Network

LANG Yuchi¹, GUO Chi^{1†}, ZHAN Jiao²

1. Electronic Information School, Wuhan University, Wuhan 430072, Hubei, China;
2. GNSS Research Center, Wuhan University, Wuhan 430072, Hubei, China

Abstract: Panoptic driving perception aims to simultaneously accomplish three perception tasks—traffic object detection, drivable area segmentation, and lane line detection—to meet the requirements of drivability prediction in autonomous driving. Existing methods usually adopt multi-task learning networks to process multiple tasks simultaneously, which have achieved certain progress in performance but generally suffer from high computational complexity as well as an imbalance between prediction accuracy and real-time performance. To address these problems, a lightweight multi-task panoramic driving perception network based on area attention, named APDP-Net, is proposed. First, the Residual-Efficient Layer Aggregation Network (R-ELAN) is introduced as the backbone network, in which the area attention module, characterized by a local and continuous attention mechanism, maintains the receptive field while reducing computational cost. Second, a GhostFPN-ELAN neck network with a feature pyramid structure is designed, introducing Ghost Convolution to reduce computation by reasonably compressing the feature space. Finally, for the lane line detection task, a Polarized Self-Attention mechanism is introduced to establish long-range connections between pixel-level features, enhancing the spatial feature representation capability of the network. Experimental results show that APDP-Net, through lightweight network design, significantly

引用格式: 郎愚迟, 郭迟, 詹骄. APDP-Net: 基于区域注意力的轻量级全景驾驶感知网络[J]. 武汉大学学报(理学版), 2026, 72(4): 507-518. DOI: 10.14188/j.1671-8836.2025.0098.

LANG Yuchi, GUO Chi, ZHAN Jiao. APDP-Net: Area Attention-Based Lightweight Panoptic Driving Perception Network [J]. *J. Wuhan Univ. (Nat. Sci. Ed.)*, 2026, 72(4): 507-518. DOI: 10.14188/j.1671-8836.2025.0098(Ch).

reduces computational overhead while maintaining excellent prediction performance, providing a feasible solution for edge-side autonomous driving perception.

Key words: multi-task learning; panoptic driving perception; area attention; autonomous driving

0 引言

自动驾驶对周围环境的全景驾驶感知(Panoptic Driving Perception)依赖于三大核心视觉任务:交通目标检测、可行驶区域分割以及车道线检测。交通目标检测旨在当前视觉帧中识别出机动车、行人、交通标识等对象,并精确定位其空间坐标与轮廓尺寸,以规避潜在碰撞风险并遵守交通规则。可行驶区域分割(亦称路面检测)主要负责识别车辆可安全通行的区域,排除障碍物及非道路区域的干扰,确保车辆路径规划的有效性 with 安全性。车道线检测则聚焦于对道路标线的识别与追踪,从而为路径规划和车道保持等功能提供关键语义支撑。这三项视觉任务的协同,为自动驾驶系统提供了对周边环境的全面理解,是车辆实现智能决策与安全驾驶的技术基石。

在深度学习技术的持续推动下,诸多针对这三种单一视觉任务的高效算法不断涌现。在交通目标检测领域,以 Fast R-CNN^[1]为代表的两阶段目标检测器在精度表现上尤为突出;YOLO 系列作为单阶段目标检测器的典型代表,因其在预测精度与实时性之间的良好平衡而被广泛应用于工业部署。例如,YOLOv5^[2]通过引入 CSPDarknet53 主干网络与 C3 模块显著提升了特征提取效率,成为工业级场景下的常用方案;YOLOv8^[3]在结构上进一步优化,采用无锚点检测机制与 C2f 模块,在预测速度与预测精度方面实现突破;最新的 YOLOv12^[4]则聚焦于基于注意力机制的网络设计,并通过引入多任务学习,使 YOLO 系列在预测精度与泛化能力方面取得更好平衡。在可行驶区域分割领域, PSPNet^[5]与 DeepLabv3+^[6]等经典方法通过多尺度特征融合策略,在复杂道路场景中实现了准确的可行区域分割。随后,基于 Transformer 架构的 TransFuser^[7]通过融合多模态感知数据(如图像与 LiDAR 信息),进一步增强了模型在动态与多变场景下的可行区域分割能力。在车道线检测领域,SCNN^[8]引入基于行传播的空间消息传递机制,有效提升了对弯曲与模糊车道线的检测性能。ENet-SAD^[9]在保持轻量化的同时提升了特征表达能力。UFAST^[10]通过行分类与全局特征压缩策略,实现了高达 250 FPS 的推理速度,显著提升了车道线检测在实时系统中的

应用价值。LaneATT^[11]引入基于锚点驱动的注意力机制与多尺度融合模块,在远距离车道线检测中 F1 达到 81.6%。CondLaneNet^[12]通过条件卷积生成动态参数,有效解决了车道交叉与遮挡场景下的检测难题。CLRNet^[13]则通过跨层特征聚合与曲线参数建模,在复杂道路曲线识别任务中展现出更强的鲁棒性。

上述方法在各自的视觉任务中均展现出优异的性能,但在实际自动驾驶系统中,由于车载计算资源的限制,多个独立模型的并行部署容易导致系统响应延迟与数据同步异常等问题,从而影响车辆的行驶安全。为了解决这一问题,研究者们构建了不同的多任务学习网络,例如 MultiNet^[14]、DLT-Net^[15]、YOLOP^[16]、HybridNets^[17]、YOLOPv2^[18]以及 YOLOPX^[19]。它们通过统一的多任务学习框架实现了交通目标检测、可行驶区域分割和车道线检测等任务的同时处理,并通过任务间信息共享提升了模型的整体感知性能与数据处理效率。然而,这些基于共享架构的多任务学习网络仍面临两个方面的挑战。一方面,为了追求更优的多任务预测性能,现有方法(如 HybridNets、AYOLOM(s)^[20]等)通常采用密集卷积堆叠结构,导致模型参数规模庞大(例如 HybridNets 达 13.7 MB)、计算负载过高,难以满足资源有限的车载嵌入式设备对实时性的要求。另一方面,常用的注意力机制,如 SE(压缩-激励注意力)^[21]、CBAM(卷积块注意力)^[22],在全局特征加权过程中引入了显著的额外计算开销,进一步影响了模型的实时性。

针对自动驾驶多任务感知网络中计算开销与性能平衡的关键难题,本文提出了一种基于区域注意力的轻量级全景驾驶感知网络——APDP-Net(Area Attention-Based Lightweight Panoptic Driving Perception Network)。实验结果表明,APDP-Net 在显著降低计算开销的同时保持了出色的预测性能。

本文主要贡献包括:1) 引入残差高层次聚合网络(Residual-Efficient Layer Aggregation Network, R-ELAN)作为主干网络,通过区域注意力机制,有效聚焦关键区域特征,提高多尺度下的特征表达能力,同时相较于以往自注意力机制降低了计算成本;2) 设计了基于特征金字塔(Feature Pyramid Network, FPN)结构的颈部网络 GhostFPN-

ELAN,通过采用幽灵卷积(GhostConv)^[23]替代传统卷积,进一步减少模型计算量;3) 在车道线检测分支引入极化自注意力(Polarized Self-Attention, PSA)^[24]模块,以低计算成本增强像素级特征间的长距离上下文依赖,提高网络对细节信息的感知能力,从而提升车道线检测的精度与连续性。

1 本文网络

1.1 网络概述

APDP-Net的整体结构如图1所示,由共享编码器(包括主干网络和颈部网络)与三个针对不同任

务而设计的解码器组成。APDP-Net端到端的预测过程如下所述:1) 为了与现有多任务学习网络(如YOLOP、YOLOPX等)保持一致,首先对原始图像进行预处理,得到尺寸为 384×640 的输入图像。2) 输入图像经过主干网络处理后,得到分辨率从 92×160 到 12×20 不等的特征图(即 C_2, C_3, C_4, C_5)。3) 这些特征图被送入颈部网络进行特征融合,得到特征图 $P_3, P_4, P_5, N_3, N_4, N_5$ 。其中, N_3, N_4, N_5 被送入交通目标检测分支, P_4 被送入可行驶区域分割分支, C_2 和 P_3 被送入车道线检测分支。4) 三个不同分支接收相应特征输入,经过处理后生成全景驾驶感知输出。APDP-Net的网络结构细节如图2所示。

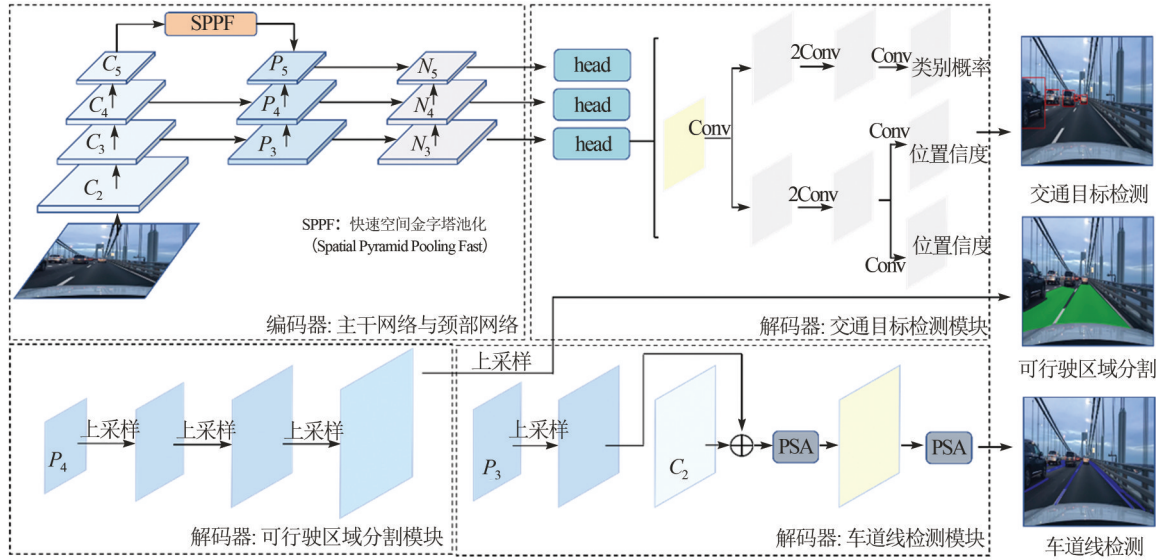


图1 APDP-Net结构示意图

Fig. 1 APDP-Net structure schematic diagram

1.2 主干网络R-ELAN

APDP-Net的主干网络采用YOLOv12中的R-ELAN,该主干网络通过引入区域注意力机制,不仅能够高效提取特征,还能在多个尺度和层级上捕捉特征,以提高网络的鲁棒性。R-ELAN首先通过C3K2模块和卷积(Conv)模块进行特征提取;其中,YOLOv12使用的C3K2模块借鉴了Bottleneck结构的设计思路,利用瓶颈层有效压缩计算量并增强特征提取能力,同时通过额外增加的特征传递路径进一步提升网络的表达能力。然后,R-ELAN通过区域注意力(A2C2f)模块以较小的计算成本引入注意力机制,进一步增强了网络的表征能力。区域注意力与传统自注意力的计算过程以及计算成本比较如下:

传统自注意力机制的计算过程如下:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d^k}}\right)\mathbf{V} \quad (1)$$

式中,Attention表示自注意力操作; $\mathbf{Q}$ 表示查询(Query); $\mathbf{K}$ 表示键(Key); $\mathbf{V}$ 表示值(Value)。

对于输入 $X \in \mathbb{R}^{H \times W \times C}$,其注意力计算成本 cost_1 为:

$$\text{cost}_1 = O(2(H \times W)^2 \frac{C}{h} + (H \times W)^2) \quad (2)$$

式中, O 表示计算复杂度; $H \times W$ 表示输入特征图尺寸(H 为特征图高度, W 为特征图宽度); C 表示特征通道数; h 表示自注意力头数。

为了降低传统自注意力机制的计算成本,一些方法将自注意力操作限制在局部范围内。例如,对于输入尺寸为 $H \times W$ 的特征图,Swin-Transformer^[25]将自注意力操作限制在尺寸为 $\frac{H}{I} \times \frac{W}{I}$ 的窗口内,其中 I 为自定义的特征图分段数。CSwin-Transformer^[26]则采用十字形的窗口自注意力机制,将注意力计算限制在宽度为 sw (CSwin-Transformer的

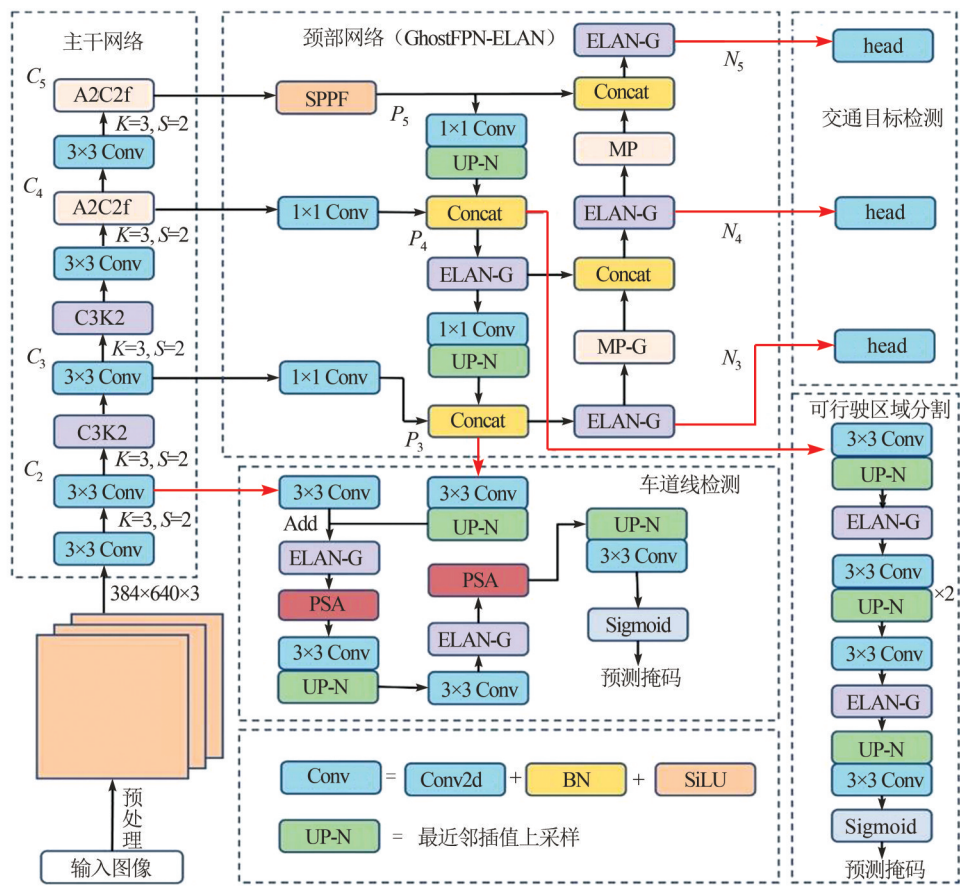


图2 APDP-Net结构细节图
Fig. 2 Details of APDP-Net architecture

预设超参数,表示感受野的宽度)的十字形水平或垂直条纹内。其中,sw随着层深从1逐步增大到8,而非固定的 $\frac{H}{I} \times \frac{W}{I}$ 窗口。这些设计能够有效地降低注意力机制计算成本,但其复杂的重塑(Reshape)操作会带来额外的内存访问开销。为此,A2C2f将特征图分为 $H \times \frac{W}{I}$ 或 $\frac{H}{I} \times W$ 的长条形

注意力窗口,在引入注意力机制的同时,最大程度地降低了重塑操作的额外开销,并保持了有效的计算效率。如图3所示,图中不透明红色方块表示当前计算位置,半透明红色区域则代表注意力机制的感受野。例如,当I取4时,区域注意力操作的计算成本cost₂为:

$$\text{cost}_2 = O\left(\frac{1}{2}(H \times W)^2 \frac{C}{h} + (H \times W)^2\right) \quad (3)$$

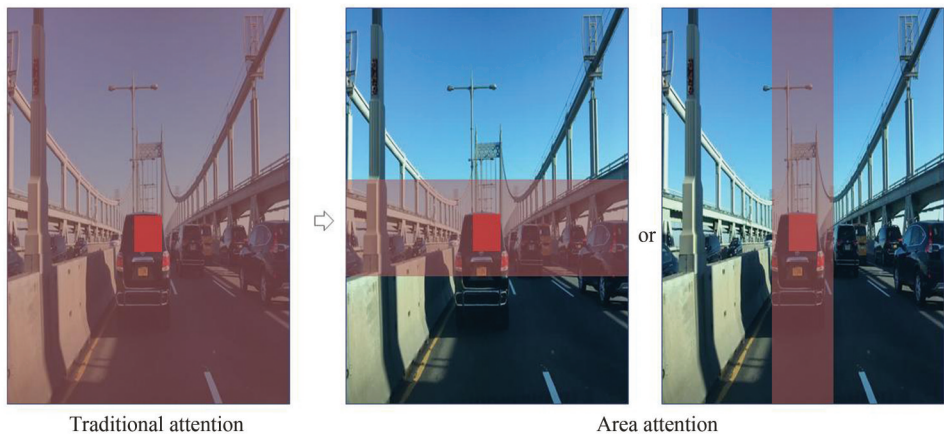


图3 传统注意力与区域注意力对比
Fig. 3 Comparison between traditional attention and area attention

将(3)式与(2)式对比可知,相比于传统自注意力机制,使用区域注意力机制后计算成本显著降低,从而有助于提高网络的预测速度。

1.3 颈部网络 GhostFPN-ELAN

本文构建了一种基于特征金字塔结构的颈部网络——GhostFPN-ELAN(见图2)。其中,ELAN-G是该网络的核心组成模块,能够引入幽灵卷积(GhostConv)有效降低网络计算量和参数量,同时通过优化特征通道设计引入更多梯度信息,提升训练稳定性与特征表达能力。

GhostConv的结构如图4所示,其核心思想是将标准卷积分为两步:首先使用较少的通道进行普通卷积(默认输出通道为输入通道的一半);随后对该输出再进行一次尺寸为 5×5 的深度可分离卷积,从而生成“伪特征”;最后,将两次卷积的输出进行通道拼接,形成最终的输出特征。通过这种方式,GhostConv能够以更低的计算代价保持特征表达能力。

此外,ELAN-G还对原始的ELANBlock结构进行了改进,ELANBlock和ELAN-G的结构示意图如图5所示。原始的ELANBlock结构在特征输入后,首先进入两条由 1×1 卷积构成的分支,其中一条分支继续与两个由两组 3×3 卷积组成的特征提取模块串联,最终将这四条从浅到深的特征分支进行通道拼接,以增强特征多样性和层级表达。ELAN-G则对原始ELANBlock结构的第三、第四梯度分支进行了结构细化,将其拆分为四条连续的梯度流路径。通过引入更多并行的细粒度梯度通道,ELAN-G能够捕获更为连续和丰富的梯度信息,有效增强模型的表达能力和训练稳定性。

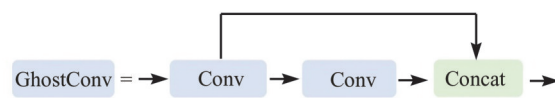


图4 GhostConv结构图

Fig. 4 Structure of GhostConv

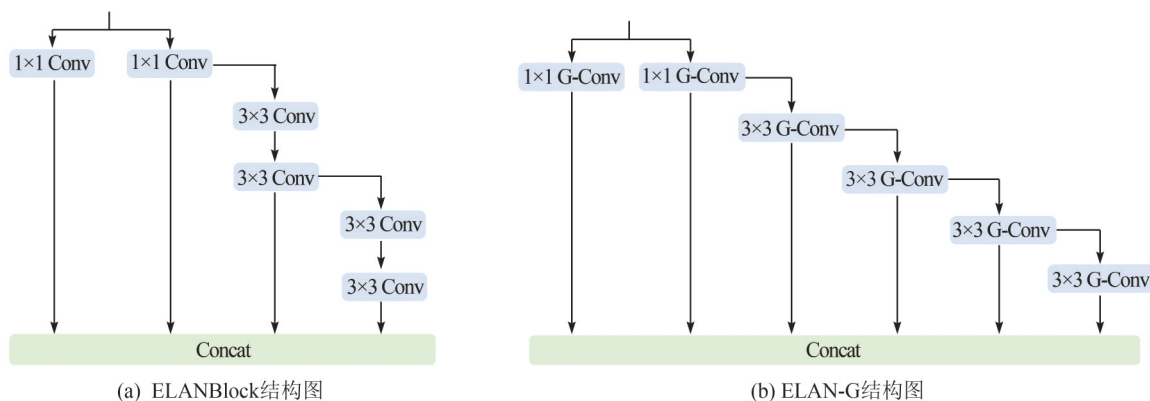


图5 ELANBlock与ELAN-G结构对比图

Fig. 5 ELANBlock and ELAN-G structure comparison chart

GhostFPN-ELAN的执行过程为:首先将 C_5 输入SPPF模块进行多尺度特征池化与融合,然后通过 1×1 的卷积模块对融合特征、 C_4 、 C_3 进行通道调整。接着沿着自顶向下的特征传递路径,将融合特征(P_5)通过1个 1×1 卷积模块与上采样模块后,与调整后的 C_4 特征图进行融合得到 P_4 ,并以同样的方式将 P_4 分别通过ELAN-G、 1×1 卷积、上采样三个模块后,与经调整的 C_3 特征图融合,输出 P_3 ,实现高层语义信息和底层空间细节的融合。最后,通过自底向上的特征传递路径输出最终的 N_3 、 N_4 、 N_5 。

1.4 交通目标检测分支

已有研究如YOLOP、HybridNets等对于交通目标检测通常采用基于锚框的方法,这种方法能够有效利用不同尺度的特征图上预设的先验锚框,从而精准地定位目标。然而,这种方法在训练时依赖超参数的设置,训练稳定性较差,同时也限制了网

络对目标形状变化的适应性。为了增强交通目标检测的鲁棒性,保障行驶安全,本文采用无锚框的解耦目标检测分支(如图1所示)。该分支首先通过一个 1×1 卷积模块对输入特征图进行通道调整,然后利用两个并行分支,每个分支包含两个 3×3 卷积模块进行特征提取。最后,分类分支、目标分支和回归分支分别采用一个卷积层输出预测结果。其中,分类分支和目标分支输出目标的类别概率和目标置信度,而回归分支直接输出边界框的位置信息。

1.5 可行驶区域分割与车道线检测分支

可行驶区域分割和车道线检测之间存在许多不同特性。如果像YOLOP那样采用相同的预测分支与输入特征完成两种不同的分割任务,可能导致两个任务之间相互干扰,进而影响网络的预测性能。受YOLOP^[19]的启发,本文采用一种特征解耦

策略,为各任务提供符合其任务特性的输入特征。

由于可行驶区域的像素占比远大于车道线,因此可行驶区域分割应更侧重于大区域分割,在低分辨率特征图基础上即可实现准确的预测。如图 1 所示,本文方法仅将低分辨率特征图 P_4 作为可行驶区域分割分支的输入,并通过 4 个上采样操作,即可做到准确分割。

对于车道线检测,由于车道线在输入图像中的占比较少,因此本文引入高分辨率特征图 C_2 输入以提供更多细节信息(如图 1 所示)。由于 C_2 获取自特

征提取的早期阶段,为了减弱前期噪声的影响,本文在 C_2 经过 1 个 3×3 卷积模块初步处理噪声后,衔接了一个 ELAN-G 模块(如图 2 所示),同时衔接第一个 PSA 模块以充分利用高分辨率的特征信息。在经过 PSA 模块后,通过两个卷积模块进一步提取特征。在通过一个上采样模块将特征尺寸恢复到原图大小后,嵌入第二个 PSA 模块以在原图尺寸恢复特征间的联系。其中,PSA 模块基于极化自注意力机制(Polarized Self-Attention Mechanisms)构建(如图 6 所示)。

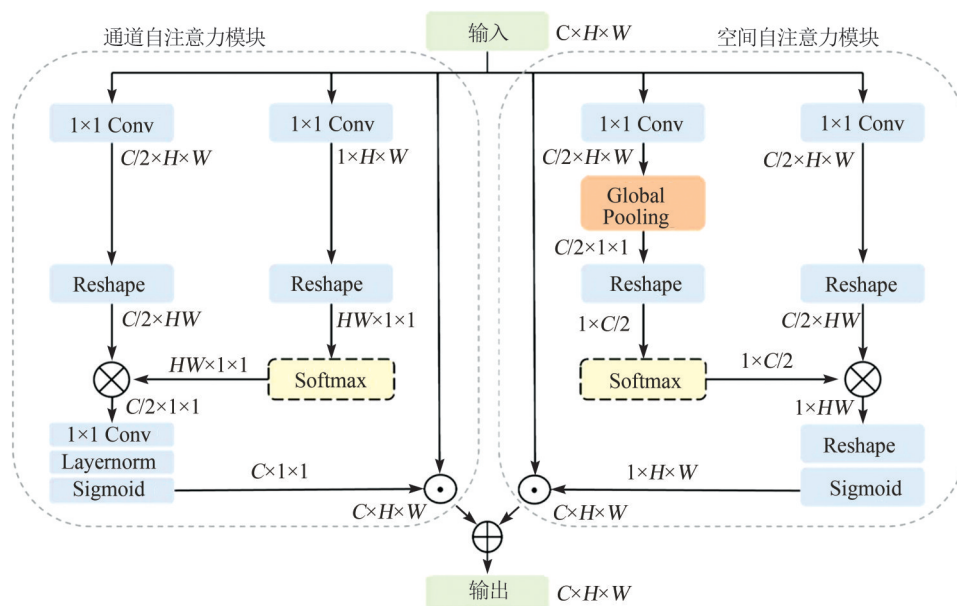


图 6 PSA 模块原理图

Fig. 6 Schematic diagram of the PSA module

极化自注意力机制的执行过程如下:首先将输入特征图分别送入通道自注意力模块与空间自注意力模块进行处理。在通道注意力模块中,输入特征通过两个 1×1 卷积,分别生成查询与键值矩阵,通道数被压缩为原始的一半($C/2$),空间尺寸保持不变。随后将特征在空间维度上展平,并计算查询与键值之间的矩阵乘法以获得通道注意力图,经过 softmax 归一化后,再通过一个 1×1 卷积、层归一化(LayerNorm)与 Sigmoid 激活函数,生成最终的通道注意力权重。该权重与原始特征图按通道维度进行逐元素乘法,实现通道方向上的特征增强。

在空间注意力模块中,输入同样经过两个 1×1 卷积生成空间维度的查询与键值矩阵,随后对键值矩阵进行全局池化以提取全局上下文信息。接着,将两个多维矩阵调整形状并进行矩阵乘法以获得空间注意力图,通过 softmax 和 Sigmoid 函数归一化处理后,生成空间注意力权重。最终,将其与原始特征图按空间维度进行逐元素相乘,实现空间维度

上的特征增强。

通道注意力模块与空间注意力模块的输出通过逐元素相加的方式融合,联合建模通道与空间维度的相关性,显著提升本文方法对关键信息的提取能力与目标区域的感知效果。通过这种轻量级的极化自注意力机制,本文方法能够在保持低计算成本的同时,将特征图的方向与尺度解耦,增强起像素级特征之间的长距离联系,从而提升整体特征表达能力。

1.6 多任务损失函数

本文提出的多任务损失函数 L_{ml} 由目标检测损失函数 L_{det} 、可行驶区域分割损失函数 L_{da} 以及车道线检测损失函数 L_{ll} 组成。即:

$$L_{ml} = \lambda_{det} L_{det} + \lambda_{da} L_{da} + \lambda_{ll} L_{ll} \quad (4)$$

式中, λ_{det} 、 λ_{da} 、 λ_{ll} 为平衡三个损失函数的超参数,设置与 YOLOP 保持相同, λ_{det} 、 λ_{da} 设置为 0.2, λ_{ll} 则设置为 0.02。

对于交通目标检测损失 L_{det} , 其计算公式为:

$$L_{\text{det}} = \lambda_{\text{cls}} L_{\text{cls}} + \lambda_{\text{obj}} L_{\text{obj}} + \lambda_{\text{reg}} L_{\text{reg}} \quad (5)$$

式中, L_{cls} 、 L_{obj} 是常用的二元交叉熵(Binary Cross Entropy, BCE)损失函数, L_{reg} 采用了交并比(Intersection Over Union, IOU)损失函数。 λ_{cls} 、 λ_{obj} 、 λ_{reg} 分别设置为 1.0、1.0、5.0。

对于可行驶区域分割损失 L_{da} , 同样采用 BCE 损失函数。

对于车道线检测损失 L_{ll} , 采用焦点(Focal)损失和特沃斯基(Tversky)损失的组合, 以有效应对车道线检测中存在的类不平衡情况:

$$L_{\text{ll}} = L_{\text{focal}} + L_{\text{tv}} \quad (6)$$

2 实验与分析

2.1 实验设置

2.1.1 数据集

为验证所提出的 APDP-Net 在全景驾驶感知任务中的实际效果和泛化能力, 本文选用 BDD100K (Berkeley DeepDrive 100K) 数据集作为基准数据集。BDD100K 是面向自动驾驶的大规模基准数据集, 分为训练集、验证集、测试集三部分, 分别包含 70 000、20 000 和 10 000 张图片。该数据集覆盖了多样化的天气条件(如晴天、雨天、雪天、雾天)、光照环境(白天/夜间)以及复杂的交通场景(如城市道路、高速公路、乡村道路等)。利用 BDD100K 数据集进行训练有效提升了网络在实际驾驶场景下的表现。

2.1.2 实验细节

APDP-Net 建立在 Pytorch 框架上, 使用 AdamW 优化器进行参数更新, 并将初始学习率设置为 0.000 15。学习率通过余弦退火法进行调整, 动量设置为 0.937。使用小批量方法遍历数据集, 批量大小设置为 32, 共在 RTX 4090 上训练 200 轮。在验证阶段, 将交通目标检测的 NMS 阈值设置如下: 置信度阈值为 0.3, IOU 阈值为 0.45。

2.1.3 评价指标

按照惯例^[16, 18-20]对不同的视觉任务采取不同的评价指标。在交通目标检测中, 采用召回率(Recall), 以及平均精度(mean Average Precision, mAP, 本文采用 mAP50%, 即 IOU 阈值为 0.5 时的平均精度)作为评价指标; 可行驶区域分割中, 采用平均交并比(mIOU)作为评价指标; 车道线检测中, 采用像素精度(Pixel accuracy)以及交并比(IOU)作为评价指标。在推理速度测试中, 以每秒帧率

(Frames Per Second, FPS)作为评价指标。以上不同方面的评价指标, 其数值越大, 均代表模型对应性能越优。

2.2 定量比较

为全面评估 APDP-Net 的性能和有效性, 将 APDP-Net 与近年来提出的经典模型及最新同类多任务网络进行对比。比较方法包括: Faster R-CNN^[1]、Enet^[27]、MultiNet^[14]、YOLOP^[16]、DLT-Net^[15]、HybridNet^[17]、SCNN^[8]、ENet-SAD^[6]、PSPNet^[5]、YOLOv8n^[3](seg)、YOLOv8n^[3](det)、YOLOv12(det)^[4]、YOLOv12(seg)^[4] 以及 A-YOLOM(n)^[20]、A-YOLOM(s)^[20]。

2.2.1 计算成本对比

首先将经典方法 YOLOP、轻量级多任务方法 A-YOLOM(n)、A-YOLOM(s)以及 APDP-Net 进行计算成本对比。所有实验均在 RTX 4060 上进行, 分别评估了各方法在模型参数量、推理速度(取批量大小为 1 时的 FPS)等方面的表现, 实验结果如表 1 所示。

表 1 不同方法的参数量对比

Table 1 Comparison of parameter counts among different methods

方法	骨干网络	输入大小	参数量/MB	FPS
YOLOP	CSPDarknet	384×640	7.90	31
AYOLOM(s)	CSPDarknet	384×672	4.43	34
AYOLOM(n)	CSPDarknet	384×672	13.61	39
APDP-Net	R-ELAN	384×640	10.71	48

由表 1 可知, 本文提出的 APDP-Net 在模型参数量方面仅略大于早期经典方法 YOLOP。得益于高效主干网络 R-ELAN 的引入, APDP-Net 在推理速度上表现更为优越, 相比 YOLOP 和 A-YOLOM(n) 分别提升了 17 FPS 和 9 FPS, 从而体现了 APDP-Net 在实时性方面的显著优势。

2.2.2 交通目标检测结果对比

表 2 展示了各方法在 BDD100K 数据集上的交通目标检测结果。从表 2 中可以看出, APDP-Net 相较于 MultiNet、DLT-Net、Faster R-CNN 等方法在两个指标上均有显著提升。与 YOLOP、YOLOv8n(det)、YOLOv12n(det)、A-YOLOM(n)、A-YOLOM(s)相比, APDP-Net 在召回率上分别提高了 2.8、9.2、6.3、6.1、4.5 个百分点。同时, 其 mAP50% 达到 77.1%, 略高于 YOLOv12n(det), 接近 A-YOLOM(n), 在精度与召回的平衡上展现出良好竞争力。

表 2 不同方法的交通目标检测结果对比

Table 2 Comparison of traffic object detection results among different methods

方法	Recall	mAP50
MultiNet ^[14]	81.3	60.2
DLT-Net ^[15]	89.4	68.4
Faster R-CNN ^[1]	81.2	64.9
YOLOv8n(det) ^[3]	82.2	75.1
YOLOv12n(det) ^[4]	85.1	75.6
YOLOP ^[16]	88.6	76.5
A-YOLOM(n) ^[20]	85.3	78.0
A-YOLOM(s) ^[20]	86.9	81.1
APDP-Net	91.4	77.1

2.2.3 可行驶区域分割结果对比

表3对比了不同方法在BDD100K数据集上的可行驶区域分割结果。结果表明,APDP-Net在mIOU上显著优于PSPNet、YOLOP、YOLOv12n(seg)、HybridNet及A-YOLOM(s),分别提高了3.2、3.6、5.1、2.3和1.8个百分点。这一结果表明APDP-Net在像素级场景理解中具备出色的预测精度和鲁棒性。

2.2.4 车道线检测结果对比

表4展示了各方法在BDD100K数据集上的车道线检测结果。APDP-Net为了简化模型训练,将BDD100K数据集中两条边缘线表示的车道线合并为一条中心线,同时训练集中车道线宽度设为8 px,而验证集则为2 px,导致IOU指标较低。相比之下,像素精度更能真实反映车道线检测的连续性与准确性。从表4中可以看出,APDP-Net在像素精度方面显著优于早期方法Enet、SCNN、ENET-SAD。相比近期的多任务学习方法YOLOP、YOLOv8n(seg)、YOLOv12n(seg)、A-YOLOM(s)分别提高了17.9、7.9、5.8、3.5个百分点。优势明显,这说明其在实际

表 3 不同方法的可行驶区域分割结果对比

Table 3 Comparison of drivable area segmentation results among different methods

方法	mIOU
MuitiNet ^[14]	71.6
DLT-Net ^[15]	72.1
PSPNet ^[5]	89.6
YOLOP ^[16]	89.2
YOLOv12n(seg) ^[4]	87.7
HybridNet ^[17]	90.5
A-YOLOM(n) ^[20]	90.5
A-YOLOM(s) ^[20]	91.0
APDP-Net	92.8

表 4 不同方法的车道线检测结果对比

Table 4 Comparison of lane detection results among different methods

方法	Pixel accuracy	IOU
Enet ^[27]	34.12	14.64
SCNN ^[8]	35.79	15.84
ENet-SAD ^[6]	36.56	16.02
YOLOP ^[16]	70.5	26.2
YOLOv8n(seg) ^[3]	80.5	22.9
YOLOv12n(seg) ^[4]	82.6	22.8
A-YOLOM(n) ^[20]	81.3	28.2
A-YOLOM(s) ^[20]	84.9	28.8
APDP-Net	88.4	22.6

应用中具有更强的稳定性与实用价值。

2.2.5 消融实验

本节对APDP-Net进行了全面的消融实验,以验证各模块对网络性能表现的具体贡献。其中,基准线设置为YOLOP,主干网络为CSPDarknet。所有实验均保持与APDP-Net相同的训练设置。实验结果如表5所示。

表 5 APDP-Net消融实验结果对比表

Table 5 Ablation study results of APDP-Net

基准线	Recall/%	mAP50/%	mIOU/%	Pixel accuracy/%	IOU/%	参数量/MB
YOLOP ^[16]	88.6	76.5	89.3	80.6	19.2	7.90
+R-ELAN	89.8	76.6	92.0	85.4	19.7	9.46
+FPN-ELAN	91.3	77.5	92.5	85.3	20.1	11.20
+ELAN-G	90.1	76.5	91.9	84.9	19.6	8.76
+Anchor free	92.1	77.9	91.7	84.6	19.6	10.37
+PSA	91.4	77.1	92.8	88.4	22.6	10.37

注:YOLOP、+R-ELAN、+FPN-ELAN、+ELAN-G、+Anchor free、+PSA 分别表示以YOLOP作为网络结构,在其基础上改进主干网络为R-ELAN,改进颈部网络为FPN-ELAN,改进ELANBlock模块为ELAN-G模块,改进交通目标检测分支为Anchor free、在车道线检测分支中加入PSA模块。

实验结果表明:1) 引入YOLOv12的主干网络R-ELAN后,网络在所有任务的表现上都获得了提升,同时网络的参数量也有所增大;2) 加入嵌入ELANBlock的颈部网络FPN-ELAN后,进一步提升了网络的交通目标检测性能,但也导致网络规模进一步增大;3) 将解码器中的ELANBlock模块替换为轻量级的ELAN-G模块大幅减少了网络的参数,虽准确率略有下降,但实现了精度和效率的平衡;4) 将交通目标检测分支替换为Anchor-free结构后,进一步提升了交通目标检测的性能,同时保持了其他两个任务的稳定精度;5) 在车道线检测分支嵌入PSA模块后,显著提升了车道线检测精度,验证了该模块在增强像素级特征间联系和捕捉空间细节方面的有效性。消融实验结果表明:APDP-Net通过系统化的网络结构设计,在三个感知任务中均实现了更优异的性能,同时保持了轻量级的模型规模,并具有更快的推理速度。

2.3 定性比较

为了验证APDP-Net在不同驾驶场景下的泛化能力,本文将BDD100K验证集划分为白天、夜晚两种驾驶场景的子集,并在每个子集上对APDP-Net与YOLOP和AYOLOM(s)进行了对比。实验结果如图7、8所示。

图7从上到下分别为输入图像,真实标注结果(Ground Truth)、YOLOP、A-YOLOM(s)以及APDP-Net的可视化结果,从左到右为四种不同驾驶场景。由场景一可知,APDP-Net的车道线检测表现明显优于其他方法。由场景二可知,APDP-Net相比YOLOP、AYOLOM(s)在弯道处的可行驶区域分割和车道线检测结果均更好。由场景三可知,APDP-Net在检测横向、倾斜车道线时同样更加准确。由场景四可知,APDP-Net相较于YOLOP,小目标检测的能力更强,相比于AYOLOM(s),车道线检测结果更为完整、连续。



图7 不同方法白天驾驶场景结果对比图

Fig. 7 Comparison of different methods under daytime conditions

图8同样包含四种不同驾驶场景。由场景一可知,APDP-Net在夜晚低光照条件下表现出色,其可行驶区域分割和车道线检测的结果更为精准。由场景二可知,面对十字路口等复杂交通场景APDP-Net依然能够表现出更优的车道线检测和可行驶区域分

割性能。由场景三可知,在夜晚弯道的情景下,APDP-Net不仅对于车道线的检测结果依然准确,同时对于远处小目标检测的性能更好,为自动驾驶提供了更可靠的安全保障。由场景四可知,APDP-Net相较于其他方法,车道线的检测效果更加连续、完整。



图 8 不同方法夜晚驾驶场景结果对比图

Fig. 8 Comparison of different methods under night conditions

总的来说,APDP-Net在多种复杂驾驶场景中表现出色,尤其在车道线检测、可行驶区域分割以及小目标检测方面优于其他方法。无论是在白天还是夜晚低光照条件下,其检测结果的准确性和连续性都为自动驾驶提供了更强的安全保障。

2.4 Nvidia Jetson NX 部署

为了验证 APDP-Net在实际场景中的应用效果,本文将 APDP-Net 实际部署于 NVIDIA Jetson NX 嵌入式平台。在部署过程中,使用了 TensorRT C++ API 将 APDP-Net 序列化为 TensorRT engine,并通过权值量化与层间融合等技术,在不损失精度的前提下,实现了 35 FPS 的实时推理速度。此外,为了引入真实图像并进行实时处理,本文设计了一个基于 C++ 的高性能部署库,并连接了 Stereo ZED 2 相机,完成了从图像捕捉、网络处理到显示输出的全流程。图 9 中左侧为显示终端生成的模拟驾驶场景视频流,通过 ZED2 深度相机对其进行图像数据采集,然后通过部署于 NVIDIA Jetson NX 平台的 APDP-Net 进行实时处理;图 9 中右侧屏幕展示了 APDP-Net 实时处理后的多任务驾驶场景感知结果,包括交通目标检测、车道线检测及可行驶区域分割的可视化结果。该端到端部署系统完整复现了自动驾驶感知系统的实际工作流程,直



图 9 APDP-Net 部署效果图

Fig. 9 Deployment results of APDP-Net

观验证了本方法在真实场景下的实时推理能力与部署可行性。

2.5 国产边缘计算平台 Lynxi HS110 部署

Lynxi HS110是由北京灵汐科技在第二代天机芯、新一代类脑计算芯片基础上发展来的,以 Lynxi KA200(-S)类脑计算芯片为核心的国产异构边缘计算平台。其配备国产瑞芯微 RK3568 处理器以及 ZYNQ ZU9EV FPGA 图像采集卡,能够高效、实时地处理图像,AI 算力达 48 TOPS (Tera Operations Per Second, 每秒万亿次运算)。联合 Lynxi 提供的 C++ 高性能软件栈 LynOS,配备海康威视摄像头,本文同样在 HS110 上设计完成 C++ 高性能部署,

经过实测接入摄像头后仍然可达 19 FPS, 从而满足了实时性和准确性需求。

3 结 语

本文针对自动驾驶多任务感知网络的计算效率与性能平衡问题, 提出基于区域注意力的轻量级网络 APDP-Net。通过引入 R-ELAN 主干网络, 设计基于特征金字塔结构的 GhostFPN-ELAN 颈部网络, 结合车道线极化自注意力模块在 BDD100K 数据集上以 10.71 MB 参数量实现 48 FPS 的推理速度, 可行驶区域分割的 mIOU 达 92.8%, 车道线检测像素精度达到 88.4%, 交通目标检测召回率达 91.4%, mAP 达 77.1%。此外, 该方法在 NVIDIA Jetson NX 和国产 Lynxi HS110 平台上分别实现了 35 FPS 和 19 FPS 的推理速度。总的来说, APDP-Net 为边缘端自动驾驶系统提供了高效的全景感知解决方案。

尽管本研究聚焦于 2D 图像空间下的轻量级多任务感知, 并在满足实时性与边缘部署需求方面取得了一定成果, 但在复杂场景中的空间建模能力仍存在局限性。为进一步提升自动驾驶系统的环境理解能力与运行安全性, 未来工作将逐步向基于鸟瞰图 (Bird's Eye View, BEV) 表示的感知框架过渡。通过融合多视角信息并构建统一的空间语义表达, 期望实现更稳定的多任务协同感知能力, 同时增强系统对复杂道路结构和全局场景的理解。

参考文献:

- [1] REN S Q, HE K M, GIRSHICK R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(6): 1137-1149. DOI:10.1109/TPAMI.2016.2577031.
- [2] GLENN J, NISHIMURA K, MINEEVA T, *et al.* YOLOv5: You Only Look Once, Version 5 [EB/OL]. [2025-05-18]. <https://github.com/ultralytics/yolov5>.
- [3] GLENN J, CHAURASIA A, QIU J. YOLOv8: You Only Look Once, Version 8 [EB/OL]. [2023-01-10]. <https://github.com/ultralytics/yolov8>.
- [4] TIAN Y J, YE Q X, DOERMANN D. YOLOv12: Attention-centric real-time object detectors [EB/OL]. 2025; arXiv: 2502.12524. <https://arxiv.org/abs/2502.12524>.
- [5] ZHAO H, SHI J, QI X, *et al.* Pyramid Scene Parsing Network [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(2): 352-363. DOI: 10.1109/TPAMI.2016.2522835.
- [6] CHEN L C, ZHU Y K, PAPANDREOU G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation [C]//Computer Vision — ECCV 2018. Cham: Springer, 2018: 833-851. DOI:10.1007/978-3-030-01234-2_49.
- [7] CHITTA K, PRAKASH A, JAEGER B, *et al.* TransFuser: Imitation with transformer-based sensor fusion for autonomous driving [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, **45**(11): 12878-12895. DOI:10.1109/TPAMI.2022.3200245.
- [8] PAN X G, SHI J P, LUO P, *et al.* Spatial as deep: Spatial CNN for traffic scene understanding [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018, **32**(1): 7276-7283. DOI:10.1609/aaai.v32i1.12301.
- [9] HOU Y N, MA Z, LIU C X, *et al.* Learning lightweight lane detection CNNs by self attention distillation [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE Press, 2020: 1013-1021. DOI:10.1109/ICCV.2019.00110.
- [10] QIN Z Q, WANG H Y, LI X. Ultra fast structure-aware deep lane detection [C]//Computer Vision — ECCV 2020. Cham: Springer, 2020: 276-291. DOI: 10.1007/978-3-030-58586-0_17.
- [11] TABELINI L, BERRIEL R, PAIXÃO T M, *et al.* Keep your eyes on the lane: Real-time attention-guided lane detection [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2021: 294-302. DOI: 10.1109/CVPR46437.2021.00036.
- [12] 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). New York: IEEE Press, 2022: 3753-3762. DOI: 10.1109/ICCV48922.2021.00375.
- [13] ZHENG T, HUANG Y F, LIU Y, *et al.* CLNet: Cross layer refinement network for lane detection [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2022: 888-897. DOI:10.1109/CVPR52688.2022.00097.
- [14] TEICHMANN M, WEBER M, ZÖLLNER M, *et al.* MultiNet: Real-time joint semantic reasoning for autonomous driving [C]//2018 IEEE Intelligent Vehicles Symposium (IV). New York: IEEE Press, 2018: 1013-1020. DOI:10.1109/IVS.2018.8500504.
- [15] QIAN Y Q, DOLAN J M, YANG M. DLT-net: Joint detection of drivable areas, lane lines, and traffic objects [J]. *IEEE Transactions on Intelligent Transportation Systems*,

- 2020, **21** (11): 4670-4679. DOI: 10.1109/TITS.2019.2943777.
- [16] WU D, LIAO M W, ZHANG W T, *et al.* YOLOP: You only look once for panoptic driving perception [J]. *Machine Intelligence Research*, 2022, **19** (6): 550-562. DOI:10.1007/s11633-022-1339-y.
- [17] VU D, NGO B, PHAN H. HybridNets: End-to-end perception network [EB/OL]. 2022; arXiv: 2203.09035. <https://arxiv.org/abs/2203.09035>.
- [18] HAN C, ZHAO Q C, ZHANG S Y, *et al.* YOLOPv2: Better, faster, stronger for panoptic driving perception [EB/OL]. 2022; arXiv: 2208.11434. <https://arxiv.org/abs/2208.11434>.
- [19] ZHAN J, LUO Y R, GUO C, *et al.* YOLOPX: Anchor-free multi-task learning network for panoptic driving perception [J]. *Pattern Recognition*, 2024, **148**: 110152. DOI:10.1016/j.patcog.2023.110152.
- [20] WANG J Y, JONATHAN WU Q M, ZHANG N. You only look at once for real-time and generic multi-task [EB/OL]. 2023; arXiv: 2310.01641. <https://arxiv.org/abs/2310.01641>.
- [21] HU J, SHEN L, ALBANIE S, *et al.* Squeeze-and-excitation networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, **42** (8): 2011-2023. DOI:10.1109/TPAMI.2019.2913372.
- [22] WOO S, PARK J, LEE J Y, *et al.* CBAM: Convolutional block attention module [C]//Computer Vision - ECCV 2018.2018: 3-19. DOI: 10.1007/978-3-030-01234-2_1.
- [23] HAN K, WANG Y H, TIAN Q, *et al.* GhostNet: More features from cheap operations [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2020: 1577-1586. DOI:10.1109/CVPR42600.2020.00165.
- [24] LIU H J, LIU F Q, FAN X Y, *et al.* Polarized self-attention: Towards high-quality pixel-wise regression [EB/OL]. 2021; arXiv: 2107.00782. <https://arxiv.org/abs/2107.00782>.
- [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). New York: IEEE Press, 2022: 9992-10002. DOI:10.1109/ICCV48922.2021.00986.
- [26] DONG X Y, BAO J M, CHEN D D, *et al.* CSWin transformer: A general vision transformer backbone with cross-shaped windows [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2022: 12114-12124. DOI:10.1109/CVPR52688.2022.01181. [LinkOut]
- [27] PASZKE A, CHAURASIA A, KIM S, *et al.* ENet: A deep neural network architecture for real-time semantic segmentation [EB/OL]. 2016; arXiv: 1606.02147. <https://arxiv.org/abs/1606.02147>.

□