

基于边缘特征引导的智能汽车语义分割方法

霍震¹, 金立生^{1,2}, 华强³, 贺阳¹

(1. 燕山大学车辆与能源学院, 河北秦皇岛 066004; 2. 燕山大学河北省特种运载装备重点实验室, 河北秦皇岛 066004; 3. 比亚迪汽车工业有限公司, 广东深圳 518118)

摘要: 针对现有语义分割算法在自动驾驶场景中未能充分挖掘目标边缘特征的问题, 提出了一种基于边缘特征引导的智能汽车语义分割方法。首先, 采用边缘特征增强模块挖掘原始图像的边缘信息作为编码器的并行输入, 增强神经网络的边缘特征表达。其次, 提出边缘注意力模块融合原始图像特征和边缘特征, 平衡图像不同区域原始特征与边缘特征在语义分割任务的关注度。最后, 在解码器中设计了高效通道金字塔池化模块, 提高不同尺度池化后上下文语义特征的利用效率。基于 Cityscapes 数据集的实验结果表明: 本文方法的语义分割性能达到 76.1% 的 MIoU 和 132.6 f/s 的算法运行速度, 与基线算法 PP-LiteSeg 相比, MIoU 提升了 2.2%, 实现了语义分割精度与实时性方面的平衡。

关键词: 智能汽车; 语义分割; 边缘特征增强; 边缘注意力机制; 高效通道金字塔池化

中图分类号: U491.2 **文献标志码:** A **文章编号:** 1671-5497(2025)09-3032-10

DOI: 10.13229/j.cnki.jdxbgxb.20250454

Edge feature-guided semantic segmentation method for intelligent vehicle

HUO Zhen¹, JIN Li-sheng^{1,2}, HUA Qiang³, HE Yang¹

(1. School of Vehicle and Energy, Yanshan University, Qinhuangdao 066004, China; 2. Hebei Key Laboratory of Special Carrier Equipment, Yanshan University, Qinhuangdao 066004, China; 3. BYD Auto Industry Company Limited, Shenzhen 518118, China)

Abstract: Regarding the issue that existing semantic segmentation algorithms fail to fully exploit target edge features in autonomous driving scenarios, an edge feature-guided semantic segmentation method for intelligent vehicles was proposed. First, an edge feature enhancement module was introduced to extract edge information from the original image as a parallel input to the encoder, strengthening the neural network's edge feature representation. Second, an edge attention module was proposed to fuse original image features and edge features, adaptively balancing their contributions in different regions for semantic segmentation. Finally, an efficient channel pyramid pooling module was designed in the decoder to improve the utilization efficiency of contextual semantic features after pooling at different scales. Experimental

收稿日期: 2025-05-24.

基金项目: 国家重点研发计划项目(2023YFB2504400); 国家自然科学基金项目(52472440); 河北省自然科学基金(F2024203112).

作者简介: 霍震(1995-), 男, 博士研究生. 研究方向: 智能汽车感知与定位. E-mail: vehicle_huo@stumail.ysu.edu.cn

通信作者: 金立生(1975-), 男, 教授, 博士. 研究方向: 智能汽车, 主动安全. E-mail: jinls@ysu.edu.cn

results on the Cityscapes dataset demonstrate that the proposed method achieves 76.1% MIoU with a processing speed of 132.6 f/s, outperforming the baseline PP-LiteSeg by 2.2% MIoU while maintaining real-time performance, balancing segmentation accuracy and computational efficiency.

Key words: intelligent vehicle; semantic segmentation; edge feature enhancement; edge attention mechanism; efficient channel pyramid pooling

0 引言

语义分割作为计算机视觉领域的一项关键任务,其在车辆自动驾驶^[1]、虚拟现实^[2]、智能交通^[3]、医学影像分析^[4]等领域发挥着至关重要的作用。语义分割任务将图像中的每个像素点预测为相应的类别标签,形成一张像素级的标签地图^[5]。这一过程帮助计算机深入理解图像内容,并为后续的分析或决策奠定基础。在智能汽车面对的自动驾驶场景中,通过语义分割可以识别并区分道路、行人、车辆等物体,从而为自动驾驶决策提供重要的信息支持^[6]。

近年来,图像语义分割任务得到了广泛的研究和应用。随着深度学习技术的快速发展,一系列基于卷积神经网络(Convolutional neural network, CNN)的语义分割模型被提出^[7-9]。其中,全卷积网络(Fully convolutional networks, FCN)^[7]作为首个端到端的语义分割模型,能够直接从不同尺寸输入图像中生成像素级的分割结果。为了进一步提升语义分割精度,U-Net^[8]提出对称编码器-解码器的网络结构设计,将浅层特征信息与深层语义信息相结合,显著提升了小目标的分割精度。DeepLab^[9]引入空洞卷积代替池化层,在保持特征分辨率的同时扩大感受野,并结合条件随机场(Conditional random field, CRF)进一步提升了分割精度。PSPNet^[10]提出金字塔池化模块(Pyramid pooling module, PPM),其通过多尺度池化聚合上下文信息,解决复杂环境下上下文信息缺失的问题。虽然这些方法在分割精度上取得了显著进展,但由于其复杂的网络结构,通常需要消耗更多计算资源,导致无法满足需要在线估计和实时决策等领域的实时性需求。

智能汽车语义分割任务的关键是在保证精度的前提下,满足智能汽车的实时性需求。Paszke等^[11]提出一种轻量级语义分割网络(ENet),其通过快速下采样和瓶颈结构设计等操作减少模型参数量,有效降低了模型计算复杂度。BiseNet^[12]针

对现有轻量级语义分割模型普遍存在空间细节特征缺失的问题,提出了一种融合空间路径和语义路径的双向分割模型,实现计算速度与精度的平衡。PP-LiteSeg^[13]设计了一种轻量级实时语义分割模型,利用灵活轻量级解码器平衡了编码器和解码器的计算成本,并采用轻量级金字塔池化模块提取全局上下文特征,在保证模型分割精度的同时,提升了算法实时性。PIDNet^[14]作为一种轻量级三支网络结构,将分割任务与PID控制器之间建立关联,提出利用边缘注意力机制融合空间信息和上下文信息。尽管上述方法在实时性方面取得了一定的进展,但未能充分挖掘图像边缘特征带来的潜在优势,由于神经网络的不可解释性,模型难以准确提取目标边缘特征,从而导致边缘区域分割精度较低。近年来,许多方法^[15,16]通过引入边缘特征提高语义分割任务中边缘区域的分割精度,GSCNN^[15]创新性地提出采用双流CNN网络架构,其通过引入门控机制将语义特征与边缘信息相结合,以增强对目标边界和小目标的分割能力。Yu等^[16]在边缘分支网络结构中引入非边缘抑制层,有效强化边缘特征表达,同时抑制非边缘噪声干扰。然而,这些方法通常依赖复杂的双流网络结构,通过额外设计边缘分支并结合边缘监督损失以引导语义分割,虽然在一定程度上提升了边缘分割精度,但也显著提高了计算复杂度,限制了其在自动驾驶等实时性要求较高场景中的应用。

针对上述问题,本文提出了一种基于边缘特征引导的智能汽车语义分割方法。由于图像边缘特征作为语义分割任务的一种重要潜在特征,本文采用边缘特征增强模块(Edge feature enhancement module, EFEM)人工提取原始图像的边缘信息,显性地为语义分割网络提供目标边缘特征。语义分割网络编码器由原始图像和边缘图像两个并行分支组成,本文设计边缘注意力模块(Edge attention module, EAM)以融合原始图像特征和边缘增强图像特征,提高原始图像特征中的特定

边缘区域关注度。为了实现语义分割精度与实时性间的平衡,本文在编码器中提出采用高效通道金字塔池化模块(Efficient channel pyramid pooling module, ECPPM),通过聚合不同尺度上下文信息提升模型分割精度。与现有边缘引导方法不同,本文方法避免了冗余的双流网络结构设计,采用高效的 Canny 算子^[17]提取边缘特征,并通过边缘注意力模块实现边缘信息与语义特征有效融合。

1 基于图像边缘特征增强的语义分割网络模型构建

1.1 总体框架

基于图像边缘特征引导的智能汽车语义分割

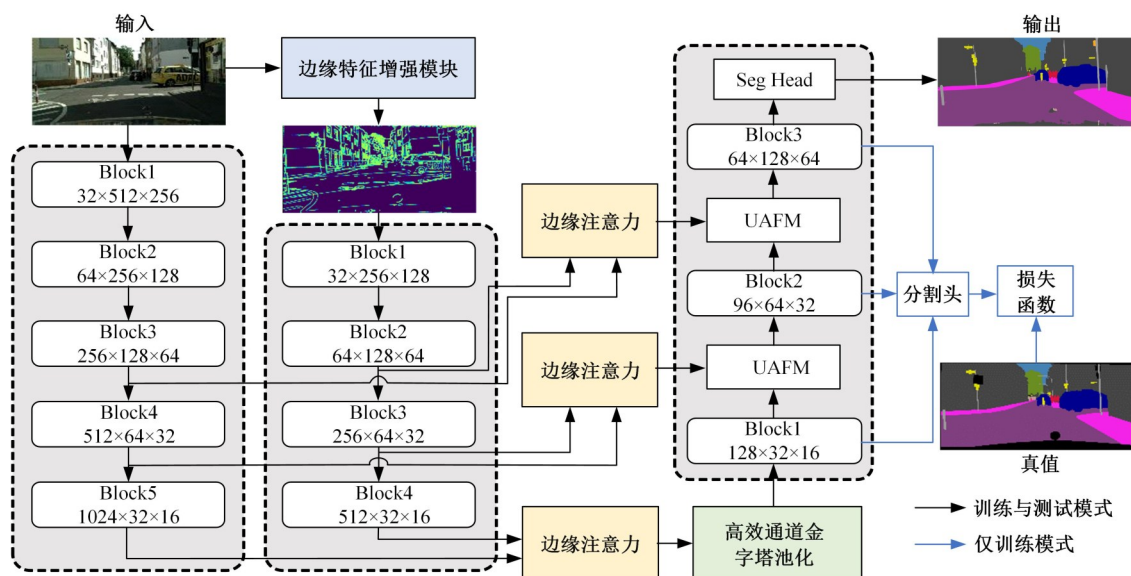


图1 基于边缘特征引导的语义分割网络框架

Fig. 1 Edge feature-guided semantic segmentation network framework

进一步地,本文提出通过采用边缘注意力模块计算边缘特征和原始图像特征的不同区域注意力权重,融合边缘特征和原始图像特征。不同尺度的上下文信息聚合可有效提高语义分割网络精度,为了在保证算法实时性的前提下提升解码器的特征利用效率,本文提出了基于高效通道注意力的金字塔池化模块,通过不同大小的池化层聚合上下文信息,并采用高效通道注意力分配各聚合通道权重,进一步提升模型的分割精度。

1.2 边缘特征增强模块

在语义分割任务中,网络模型通常能够有效识别目标内部区域特征,但对目标边缘的分割精

度较低。这一局限性主要是由于神经网络的可解释性较差,导致模型难以准确提取和利用目标边缘特征,从而影响目标轮廓的分割效果。为了提升语义分割算法在目标边缘处的分割精度,本文提出了边缘特征增强模块。通过增强目标边缘特征,使得网络在分割时能够更加关注目标的轮廓信息,从而提高边缘像素点的分割精度。边缘特征可作为语义分割模型中划分像素点类别的重要潜在特征。目标边缘是原始图像中像素值突变的区域,一般通过计算像素点在不同方向的梯度确定目标边缘。本文首先利用 Canny 算子^[17]获得原始图像 f 的边缘信息 $\text{grad}(f)$, 其计算方法

如下所示:

$$\text{grad}(f) = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (1)$$

式中: g_x 和 g_y 分别为图像在 x 方向和 y 方向的梯度值。

为了加强边缘特征,填补中断和不完整的边缘信息,进一步采用膨胀操作对图像边缘信息进行扩张。图像膨胀计算方法如下所示。

$$\text{Edge_dil}(i,j) = \text{Edge} \oplus \text{Kernal} = \{i,j | (\text{Edge})_{ij} \cap \text{Kernal} \neq \emptyset\} \quad (2)$$

式中: $\text{Edge_dil}(i,j)$ 为膨胀操作后的图像; Edge

为边缘图像; Kernal 为膨胀结构元素,本文取 5×5 大小的矩形作为膨胀结构元素。

扩张后图像的边缘特征更加显著,为进一步突出原始边缘特征在膨胀图上的影响,最终,通过计算原始边缘图像和边缘膨胀图像的二范数获取边缘特征增强图像 $\text{Edge_en}(i,j)$:

$$\text{Edge_en}(i,j) = \|\text{Edge}(i,j), \text{Edge_dil}(i,j)\|_2 \quad (3)$$

式中: $\|\cdot\|_2$ 为 L_2 范数计算。

边缘特征增强示例如图2所示,其中图2(a)为原始图像示例,由Canny算子计算获得的原始图像边缘信息示例如图2(b)所示,进一步对边缘特征增强后的图像示例如图2(c)所示。

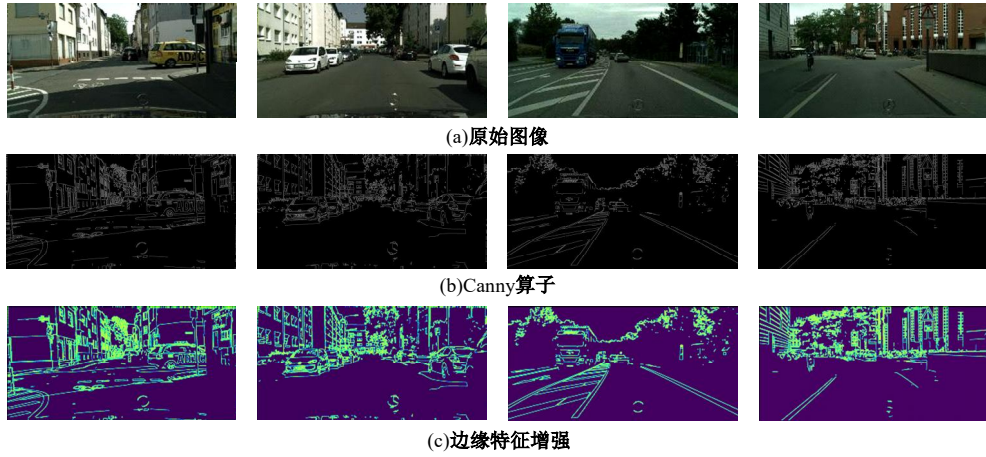


图2 边缘特征增强示例

Fig. 2 Example of edge feature enhancement

1.3 边缘注意力模块

在基于边缘引导的语义分割网络中,边缘特征和原始图像特征通常被单独处理,缺乏有效融合方式,导致网络无法充分利用边缘特征和原始图像特征之间的关系。本文提出构建边缘注意力模块融合两部分特征,强调原始图像特征中的边缘信息,动态调整原始图像特征和边缘特征的关注度,以提升语义分割任务中目标边缘的分割精度。本文提出的边缘注意力模块如图3所示。

注意力机制通过对特征的不同位置给予不同的权重来关注意重点区域特征^[19]。传统自注意力模块的Key张量、Query张量和Value张量(Q 、 K 、 V)均来源于相同特征图。与传统自注意力机制不同,本文边缘注意力机制的Key张量来源于边缘特征,Query张量和Value张量来源于原始图像特征。边缘注意力模块通过借助边缘特征,强调

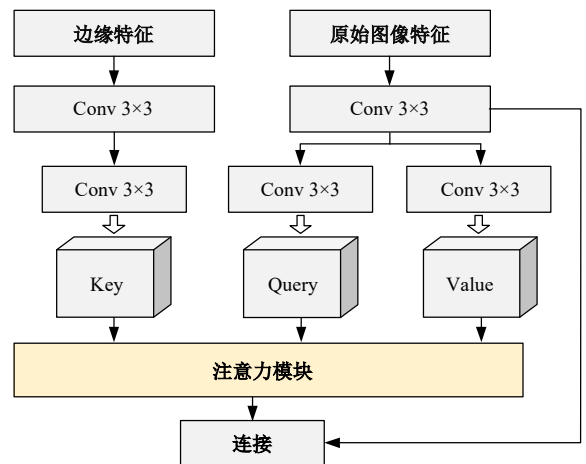


图3 边缘注意力模块

Fig. 3 Edge attention module

原始图像特征中的特定区域,其利用点积计算边缘特征张量Key与原始图像特征张量Query之间的相似度,并利用Softmax函数获取归一化权值,其计算公式如下所示。

$$\alpha_{ij} = \text{Attention}(q_i, k_j) = \text{Softmax}\left(\frac{q_i \cdot k_j^T}{\sqrt{d_k}}\right) \quad (4)$$

式中: α_{ij} 为归一化权值, 表示通道 i 对通道 j 的注意力权重; $\text{Attention}(q_i, k_j)$ 为 q_i 对 k_j 的注意力分数, 其中 q_i 为第 i 个通道的 Query 向量, k_j 为第 j 个通道的 Key 向量; $\text{Softmax}(\cdot)$ 为归一化激活函数; $\sqrt{d_k}$ 为比例因子, 表示 Key 张量的通道数。

之后, 采用归一化权值对张量 Value 中各通道的数据进行加权, 得到各通道的输出 z_i 为:

$$z_i = \sum_j \alpha_{ij} v_j \quad (5)$$

1.4 高效通道金字塔池化模块

在语义分割任务中, 目标的语义信息通常存在于不同尺度的上下文区域。为了在不同尺度下充分利用上下文信息, 传统的池化方法通常存在信息丢失或分辨率不足的问题。金字塔池化模块 (Pyramid pooling module, PPM)^[10] 通过对输入特征图进行不同尺度的池化操作, 并将这些池化结果进行融合, 从而有效捕捉图像中不同尺寸的上下文信息, 增强模型对不同尺度信息的表达能力。传统 PPM 模块对不同尺度池化结果融合将导致

通道数倍增, 由于不同尺度池化结果融合后各通道的特征表达能力存在差异, 本文提出 PPM 与高效通道注意力 (Efficient channel attention, ECA)^[20] 相结合的高效通道金字塔池化模块 (ECPPM)。ECPPM 模块通过引入多尺度池化操作来聚合丰富的上下文信息, 并结合高效通道注意力机制, 对不同通道的特征进行加权, 以优化各尺度信息的表达, 在平衡精度与实时性的前提下, 进一步提升 PPM 对上下文信息的表达能力。

本文提出的高效通道金字塔池化模块结构如图 4 所示。高效通道金字塔池化模块由 3 个全局平均池化操作组成, 其池化尺寸分别为 1×1 、 2×2 和 4×4 。随后采用连接 (Concat) 操作将经过 3 个池化操作后上采样的特征图与原始特征图融合, 输出通道数为原始特征通道的 4 倍, 其包含了原始特征信息和不同尺度上下文信息。由于聚合后通道数倍增, 本引入 ECA 模块动态调整不同通道的特征权重。ECA 模块首先采用全局平均池化方法生成大小为 $C \times 1 \times 1$ 的特征向量, 对特征向量进行卷积核为 5 的 1D 卷积和激活函数映射后, 将多尺度池化后获得的特征通道进行逐元素相乘。

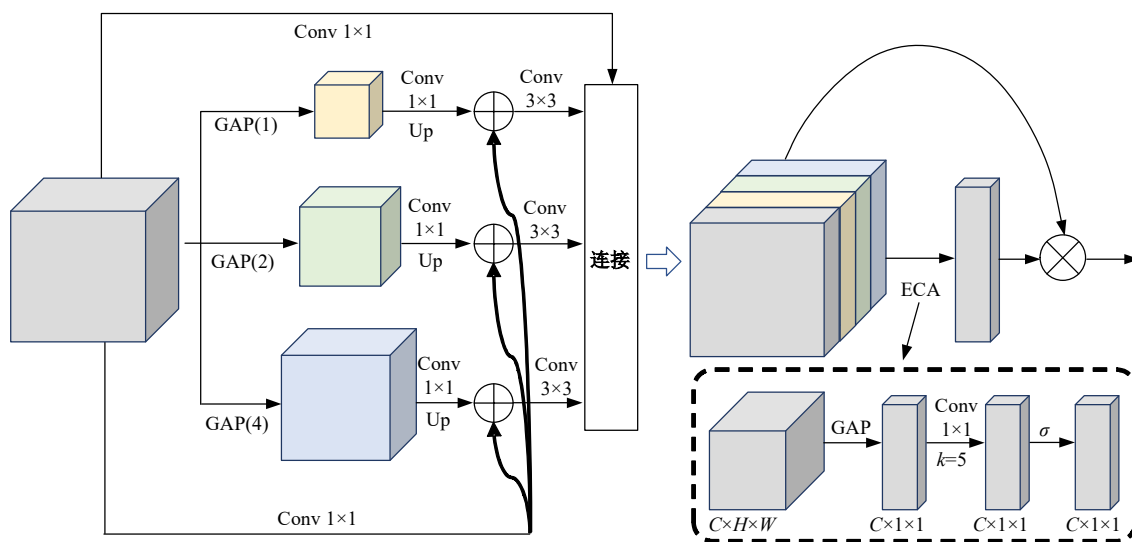


图 4 高效金字塔池化模块

Fig. 4 Efficient channel pyramid pooling module

2 实验设计及结果分析

为了证明本文基于图像边缘特征增强的实时语义分割方法的优越性, 在大型开源语义分割 Cityscapes 数据集^[21]上进行算法性能评估, 同时在真实道路场景下开展实车测试以验证其有

效性。

2.1 实验准备

2.1.1 数据集

Cityscapes 数据集被广泛应用于城市场景理解和自动驾驶领域。数据集由 5 000 张高质量的城市街景图像组成, 其中训练集包含 2 975 张图

像,验证集包含500张图像,测试集包含1 525张图像,每张图像的分辨率为 $2\,048 \times 1\,024$ 。数据集中标注了19个类别,包括道路(Road)、建筑物(Building)、行人(Person)和交通标志(Traffic Sign)等类别,每个类别都具有像素级的语义标签信息,为本文在自动驾驶场景中的语义分割任务提供数据支撑。

2.1.2 评价指标

为了评估本文算法的性能,参考当前主流的语义分割评估方法^[22,23],本文采用平均交并比(Mean intersection over union, MIoU)评估语义分割算法精度,采用算法推理速度(Frames per second, FPS)评估语义分割算法实时性。其中, FPS为单位时间(1 s)内处理的图像帧数; MIoU为像素级别下的预测与真实标注的平均重叠程度, MIoU的计算公式可表示为:

$$\text{MIoU} = \frac{1}{N} \sum_{i=1}^N \frac{|\text{Pre}_i \cap \text{Truth}_i|}{|\text{Pre}_i \cup \text{Truth}_i|} \quad (6)$$

式中: N 为语义分割类别总数; Pre_i 为第 i 类的预测区域; Truth_i 为第 i 类的真实区域; $|\text{Pre}_i \cap \text{Truth}_i|$ 为第 i 类预测区域与真实区域的交集面积; $|\text{Pre}_i \cup \text{Truth}_i|$ 为第 i 类预测区域与真实区域的并集面积。

由于本文方法旨在提高目标边缘分割精度,因此,本文进一步引入边缘分割精度的评价指标。具体而言,本文通过计算一定阈值范围内预测掩码的边界 F -score 评估算法边缘分割质量^[24]。其中,本文选用阈值分别为3、5、9和12个像素。

2.1.3 实施细节

本文提出的基于边缘特征引导的智能汽车语义分割方法训练与测试的硬件平台由 NVIDIA GeForce RTX 4070 GPU 和 16 GB RAM 组成。软件代码采用基于 PaddlePaddle 的 PaddleSeg 开源图像分割套件开发,软件环境由 Python 3.9.21 版本、PaddlePaddle 2.6.2 版本、CUDA 12.6 版本和 cudnn 8.9.7 版本构成。

在训练过程中,为了增强模型的泛化能力,对训练集图像进行随机水平翻转、随机裁剪和缩放,裁剪后图像尺寸为 $1\,024 \times 512$ 。本文采用随机梯度下降(Stochastic gradient descent, SGD)作为优化器对模型进行训练,动量值为0.9,权重衰减为

0.000 04。本文共训练500个 epoch, batch_size 大小设置为16,训练过程初始化学率为0.01,学习率变化策略为多项式学习率衰减,衰减速率为0.9,学习率衰减过程如下:

$$lr = lr_{\text{init}} \times \left(1 - \frac{\text{iter}}{\text{max_iter}}\right)^{0.9} \quad (7)$$

式中: lr 为当前学习率; max_iter 为训练阶段最大迭代数; iter 为当前迭代数; lr_{init} 为初始学习率。

2.2 对比实验结果及分析

为了验证本文方法的有效性和先进性,将本文算法与现有主流语义分割模型在 Cityscapes 数据集上进行性能对比分析。不同方法在 Cityscapes 数据集的测试结果如表1所示,本文方法分别在采用 STDC1 和 STDC2 作为骨干网络(Backbone)时达到了74.6%、76.1%的 MIoU 和177.8、132.6 FPS 的算法运行速度,实现了分割精度和实时性方面的平衡。同样采用边缘引导的 GSCNN 算法虽然表现出良好的分割性能,但由于其复杂的模型结构和高分辨率输入导致计算量增加,使其无法满足实时分割应用需求。与基线算法 PP-LiteSeg 相比,在采用相同的 Backbone 基础上,本文方法的分割精度分别提高了2.6%和2.2%的 MIoU。虽然本文提出的边缘特征引导网络结构增加了算法耗时,但本文算法运行速度依然能够满足智能汽车在自动驾驶场景语义分割的速度需求。

本文选取一部分 Cityscapes 验证集上的分割结果,进行定性实验结果分析。为了验证本文算法的先进性,额外选取了 STDC1^[18]、PP-LiteSeg-T^[13]和 BiSeNetV1^[12]模型的分割结果进行比较。图5展示了不同算法的可视化分割结果,其中每个示例均包含了原始图像、真值图像和各方法的分割结果图。与其他对比方法的可视化分割结果相比,本文方法具有更准确的分割结果,在第1个示例中,仅有本文算法有效分割出远距离的“Person”实例,体现出本文算法在远距离小目标实例的分割能力。在第2、第3和第4个示例中,本文算法分别对“truck”“fence”和“wall”实例进行有效分割,精准分割了目标边缘和内部像素。由此表明,本文算法通过增强模型对边缘特征和多尺度上下文特征的表达能力,有效提升了模型对远距离小目标和边缘区域的分割性能。

表 1 不同方法语义分割性能对比结果

方法	年份	输入图像尺寸	骨干网络	MIOU/%	速度/(f·s ⁻¹)
Enet ^[11]	2016	512×1 024	—	58.3	76.9
BiSeNetV1 ^[12]	2018	768×1 536	Xception39	68.4	105.8
BiSeNetV1 ^[12]	2018	768×1 536	ResNet18	74.7	65.5
GSCNN ^[15]	2019	1 024×2 048	—	80.8	—
STDC1 ^[18]	2021	512×1 024	STDC1	71.9	250.4
STDC2 ^[18]	2021	512×1 024	STDC2	73.4	188.6
BiSeNetV2 ^[25]	2021	512×1 024	—	72.6	156
BiSeNetV2-L ^[25]	2021	512×1 024	—	75.3	47.3
Edgenet ^[26]	2020	512×1 024	—	71.0	31.4
PP-LiteSeg-T ^[13]	2022	512×1 024	STDC1	72.0	273.6
PP-LiteSeg-B ^[13]	2022	512×1 024	STDC2	73.9	195.3
BiSeNetV3 ^[27]	2023	512×1 024	STDC1	73.5	244.3
BiSeNetV3 ^[27]	2023	512×1 024	STDC2	74.5	180.3
ALANet ^[28]	2024	512×1 024	—	74.4	115.6
本文(STDC1)	—	512×1 024	STDC1	74.6	177.8
本文(STDC2)	—	512×1 024	STDC2	76.1	132.6

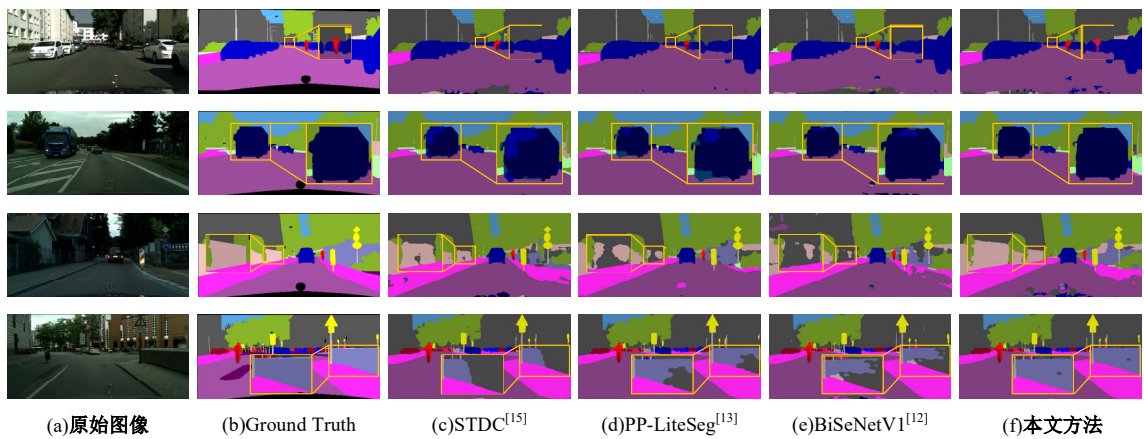


图 5 不同方法语义分割结果可视化

Fig. 5 Visualization of semantic segmentation results for different methods

2.3 消融实验结果与分析

为了验证本文方法中各模块的有效性,进一步开展基于Cityscapes数据集的消融实验。本文采用STDC2作为消融实验的backbone,以PP-LiteSeg为基线(baseline)算法,随后依次添加本文提出的边缘特征增强模块(EFEM)、边缘注意力模块(EAM)和高效通道金字塔池化模块(ECPPM),消融实验结果如表2所示。

实验结果表明,在依次添加各模块后,模型的分割精度不断提升,且模型在不同容错阈值下的边缘分割质量均呈现稳定提升趋势,证明了各模块在提升语义分割精度性能上的有效性。采用边缘特征增强模块后,与基线算法相比,模型的MIoU提升了0.5%,在3、5、9和12像素(pixel,表

2中简称为px)阈值下,F-score分别提升了2.5%、3.3%、2.5%和2.4%,表明该模块的引入能够引导模型关注目标边缘信息,进而提升了模型的语义分割精度。提出使用边缘注意力模块对边缘特征与原始图像特征融合后,模型的MIoU提升了0.6%,F-score分别提升了3.3%、2.7%、2.8%和4.0%,表明边缘注意力模块能够在原始特征图像不同区域上分配合适的边缘特征权重,有效平衡了神经网络对原始图像特征与边缘特征的关注度。进一步引入高效通道金字塔池化模块后,模型的MIoU提升了1.1%,F-score分别提升了0.9%、1.0%、0.8%和0.4%,验证了该模块能够提升模型多尺度上下文信息的表达能力。

为了更加直观地展现各模块的有效性,本文

选择 4 个示例进行可视化分析,消融实验可视化分割结果如图 6 所示,同样包含原始图像、真值图像和各消融方法的分割结果图像。与表 2 结果一致,在逐步引入各模块后,可视化分割结果逐渐接

近真值。由此表明,本文提出的边缘特征增强和边缘注意力模块能够有效捕获并利用原始图像的边缘特征,高效通道金字塔池化模块提高了不同尺度池化后上下文语义特征的利用效率。

表 2 消融实验语义分割性能对比结果

Table 2 Ablation study comparison of semantic segmentation performance									
基线算法	EFEM	EAM	ECPPM	MIOU/%	速度/(f·s ⁻¹)	F-score(3 px)	F-score(5 px)	F-score(9 px)	F-score(12 px)
✓	—	—	—	73.9	195.3	67.8	71.5	74.4	75.2
✓	✓	—	—	74.4	152.6	69.5	73.9	76.3	77.0
✓	✓	✓	—	75.0	139.3	71.8	75.9	78.5	80.1
✓	✓	✓	✓	76.1	132.6	72.5	76.7	79.2	80.5

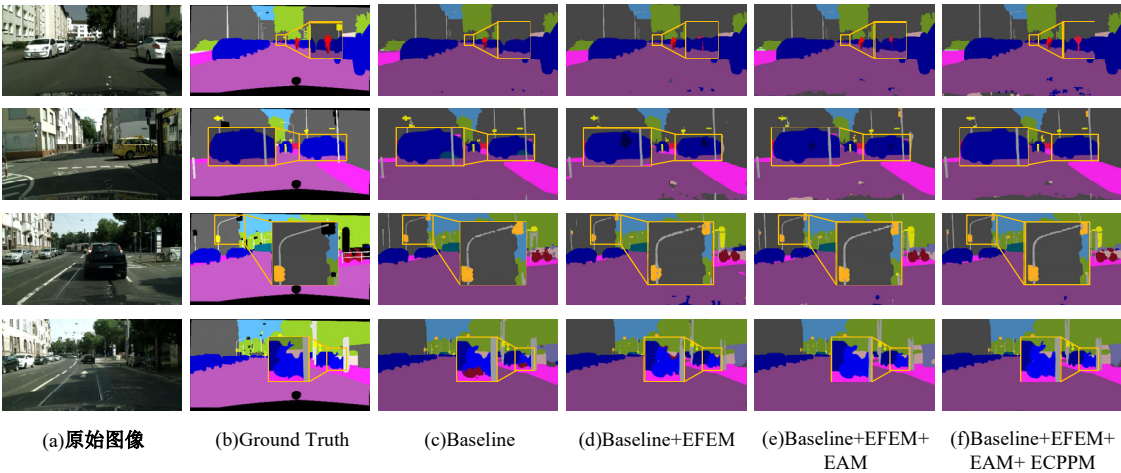


图 6 消融实验语义分割结果可视化

Fig. 6 Visualization of semantic segmentation results for ablation study

2.4 实车试验

为了验证基于图像边缘特征引导的智能汽车语义分割方法在真实道路场景中的适用性,将本文算法部署在 WEY Coffee 02 智能汽车上,其搭载了森云 SG2 摄像头和移动计算平台。

本文方法在真实道路场景的语义分割结果如图 7 所示,由图可以看出,真实道路场景中的车辆、行人等目标被精准分割,实车试验结果表明本文方法可以满足自动驾驶汽车在真实道路上的应用需求。

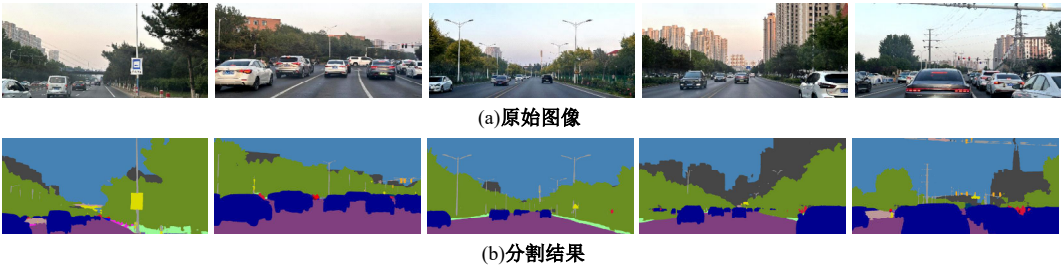


图 7 真实道路场景语义分割结果可视化

Fig. 7 Visualization of semantic segmentation results of real road scenes

3 结 论

(1)编码器利用边缘特征增强模块提取的边缘图像数据作为并行分支输入,提高了模型对目标轮廓的关注度。进一步提出采用边缘注意力模块融合编码器的两个并行分支特征,包括原始图

像特征和边缘特征,提高了目标边缘的分割精度。
(2)解码器引入了高效通道注意力,提出了高效通道金字塔池化模块,通过不同尺度池化聚合上下文信息并利用高效通道注意力分配各聚合通道权重,提升了解码器的特征表达能力。

参考文献:

- [1] 闵莉, 董冰洁, 安冬. 基于多注意力机制与跨特征融合的语义分割算法[J]. 计算机工程, 2024, 50(8): 282-289.
Min Li, Dong Bing-jie, An Dong. Semantic segmentation algorithm based on multi-attention mechanism and cross-feature fusion[J]. Computer Engineering, 2024, 50(8): 282-289.
- [2] Liang Y, Xu Z, Rasti S, et al. On the use of a semantic segmentation micro-service in AR devices for UI placement[C] //2022 IEEE Games, Entertainment, Media Conference, Bridgetown, Barbados, 2022: 1-6.
- [3] Yang L, Bai Y, Ren F, et al. Lcfnets: compensation strategy for real-time semantic segmentation of autonomous driving[J]. IEEE Transactions on Intelligent Vehicles, 2024, 9(4): 4715-4729.
- [4] Zhou Z, Rahman Siddiquee M M, Tajbakhsh N, et al. Unet++: a nested U-net architecture for medical image segmentation[C] //Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Granada, Spain, 2018: 3-11.
- [5] Mo Y, Wu Y, Yang X, et al. Review the state-of-the-art technologies of semantic segmentation based on deep learning[J]. Neurocomputing, 2022, 493: 626-646.
- [6] Muhammad K, Hussain T, Ullah H, et al. Vision-based semantic segmentation in scene understanding for autonomous driving: recent achievements, challenges, and outlooks[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(12): 22694-22715.
- [7] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(4): 640-651.
- [8] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation [C] //Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 2015: 234-241.
- [9] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.
- [10] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 6230-6239,.
- [11] Paszke A, Chaurasia A, Kim S, et al. Enet: a deep neural network architecture for real-time semantic segmentation[J/OL]. [2025-05-10]. <https://arxiv.org/abs/1606.02147>
- [12] Yu C, Wang J, Peng C, et al. BiSeNet: bilateral segmentation network for real-time semantic segmentation[C] //Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 2018: 334-349.
- [13] Peng J, Liu Y, Tang S, et al. Pp-liteseg: a superior real-time semantic segmentation model[J/OL]. [2025-05-10]. <https://arxiv.org/abs/2204.02681>
- [14] Xu J, Xiong Z, Bhattacharyya S P. PIDNet: a real-time semantic segmentation network inspired by PID controllers[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, 2023: 19529-19539.
- [15] Takikawa T, Acuna D, Jampani V, et al. Gated-SCNN: Gated shape cnns for semantic segmentation [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019: 5228-5237.
- [16] Yu L, Yao A, Duan J. Improving semantic segmentation via decoupled body and edge information[J]. Entropy, 2023, 25(6): No. 891.
- [17] Ding L, Goshtasby A. On the Canny edge detector [J]. Pattern Recognition, 2001, 34(3): 721-725.
- [18] Fan M, Lai S, Huang J, et al. Rethinking bisenet for real-time semantic segmentation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021: 9711-9720.
- [19] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48-62.
- [20] Wang Q, Wu B, Zhu P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 11531-11539.
- [21] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 3213-3223.

- [22] Zhou Q, Wang L, Gao G, et al. Boundary-guided lightweight semantic segmentation with multi-scale semantic context[J]. IEEE Transactions on Multimedia, 2024, 26: 7887-7900.
- [23] 江晟,路启,夏森磊,等. 基于特征融合注意力的RGB-D语义分割算法[J/OL]. [2025-05-10]. <https://doi.org/10.13229/j.cnki.jdxbgxb.20250159>
- [24] Perazzi F, Pont-Tuset J, McWilliams B, et al. A benchmark dataset and evaluation methodology for video object segmentation[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 724-732.
- [25] Yu C, Gao C, Wang J, et al. Bisenet v2: bilateral network with guided aggregation for real-time semantic segmentation[J]. International Journal of Computer Vision, 2021, 129: 3051-3068.
- [26] Han H Y, Chen Y C, Hsiao P Y, et al. Using channel-wise attention for deep CNN based real-time semantic segmentation with class-aware edge information[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(2): 1041-1051.
- [27] Tsai T H, Tseng Y W. BiSeNet V3: bilateral segmentation network with coordinate attention for real-time semantic segmentation[J]. Neurocomputing, 2023, 532: 33-42.
- [28] Liu Q, Wang C, Li Z, et al. Attention based lightweight asymmetric network for real-time semantic segmentation[J]. Engineering Applications of Artificial Intelligence, 2024, 130: No. 107736.