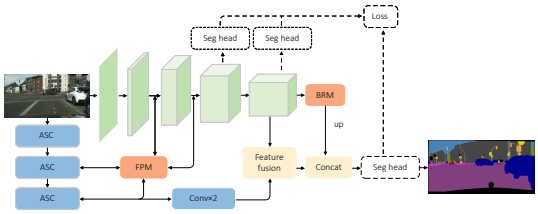
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 5 月, 第 53 卷, 第 5 期

基于双分支边界细化网络的街景语义分割方法



叶　涛[1,2], 林珊玲[1,2], 林坚普[1,2], 吕珊红[1,2], 林志贤[1,2,3*], 郭太良[2,3]

[1]福州大学先进制造学院，福建 泉州 362251；
[2]中国福建光电信息科学与技术实验室，福建 福州 350116；
[3]福州大学物理与信息工程学院，福建 福州 350116

摘要：针对城市街景语义分割中的边界模糊这一核心问题，本文提出一种基于双分支边界细化网络的街景语义分割方法(BRNet)。该方法通过边界细化模块重点强化边缘特征提取，能有效改善复杂场景下的边界分割质量；结合自适应采样卷积增强多尺度特征适应能力，并通过特征传递模块促进双分支信息互补。此外，引入洛瓦兹损失函数缓解类别不平衡问题。实验表明，BRNet 在 Cityscapes 数据集上的 mIoU 达 81.43%，较双边分割网络 V2 (bilateral segmentation network V2, BiSeNetV2) 提升 5.69%，在 WildDash2 和 BDD100K 数据集上展现良好泛化性，特别在边界清晰度和小目标分割方面表现优异，验证了其实用性。

关键词：语义分割；边界细化；自适应采样卷积；特征传递；双分支架构；城市街景

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250277 | CSTR: 32245.14.oee.2026.250277 | 中图分类号：TP183 | 文献标识码：A

叶涛，林珊玲，林坚普，等. 基于双分支边界细化网络的街景语义分割方法 [J]. 光电工程，2026，**53**(5): 250277

Ye T, Lin S L, Lin J P, et al. Semantic segmentation of street scene using a dual-branch boundary refinement network[J]. *Opto-Electron Eng*, 2026, **53**(5): 250277

1　引　言

随着人工智能技术的发展，智能驾驶系统已经逐渐普及。在智能驾驶系统中，车载视觉传感器通过感知周围环境，并利用语义分割技术实现场景理解，为决策提供关键信息[1]。然而，真实街景的复杂性为智能驾驶带来了巨大挑战。从大型的建筑到小型的路标，从行人到各类车辆，街景中的元素种类繁多，尺寸差异显著[2]。此外，多变的光照条件和视觉传感器的性能进一步增加了分割任务的难度。传统的分割方法面对复杂情况时，往往难以实现精确的分割[3]。因此，基于深度学习的街景语义分割技术迅速发展，其对于提升驾驶安全性、降低交通事故率具有重要意义。作为场景理解的核心技术，语义分割成为自动驾驶领域的研究热点之一[4-5]。

自全卷积网络 (Fully convolutional network，FCN)[6] 提出以来，研究者们不断致力于设计和优化更加高效、精确的分割模型。FCN 首次将卷积神经网络应用于像素级分类任务，奠定了现代语义分割技术的基础。随后，UNet[7] 凭借其独特的编码器-解码器结构和跳跃连接机制在医学图像分析中引起轰动。为了扩大感受野并在多

收稿日期：2025-09-16；修回日期：2025-11-21；录用日期：2025-11-21；网络出版日期：2026-05-25
基金项目：国家重点研发计划 (2021YFB3600603)
*通信作者：林志贤，lzx2005000@163.com
版权所有©2026 中国科学院光电技术研究所

个尺度上捕获上下文信息，DeepLab 系列模型应运而生。DeepLab[8] 首次引入了空洞卷积，有效扩大了感受野而不增加计算量，提升了模型对多尺度目标的识别能力，同时通过空洞空间金字塔池化 (atrous spatial pyramid pooling, ASPP) 模块，结合多尺度特征进行并行处理，显著提升了分割精度。其改进版 DeepLabv3+[9] 在此基础上巧妙地融合了编码器-解码器结构和 ASPP 模块，进一步提高了边界定位精度和多尺度目标的分割性能，成为语义分割领域的标杆模型之一。PSPNet[10] 引入的金字塔池化模块则大幅提升了复杂场景的分割能力，为细粒度的分割提供了有力支持。DenseASPP[11] 通过密集连接的 ASPP 模块进一步增强了多尺度特征捕获能力。Sun 等人提出的 HRNet[12] 通过多分辨率并行卷积和跨分辨率信息交换机制，显著增强了高分辨率特征捕获能力。Li 等人提出的 SegRet[13] 基于 RetNet 并集成轻量化残差解码器，减少参数量的同时保持了高精度。

近年来，针对街景图像中物体边界模糊、小目标及细长结构分割精度不足的问题，边界特征优化成为提升语义分割性能的关键研究方向之一。Gated-SCNN[14] 创新性地引入了一个独立的形状流分支，并利用门控机制强化对边界信息的建模，显著提升了物体轮廓的分割精度。PointRend[15] 则提出了一种迭代细分的思路，将分割视为在不确定区域进行点级预测的渲染问题，有效改善了物体边缘的细节。此外，Boundary Loss 等专门设计的损失函数也通过直接优化边界区域的预测，在提升分割边缘质量方面展现出良好效果。这些工作在边界精细化分割方面取得了显著进展。

与此同时，基于 Transformer 的语义分割方法迅速崛起[16]。Segformer[17] 设计了高效编码器和多尺度融合解码器，成功将 Transformer 引入视觉任务；AFFormer[18] 采用异构运算符协同建模细粒度与上下文特征；TSG[19] 提出动态多尺度特征选择机制优化融合；ContextFormer[20] 融合 CNN 与 Transformer 提升推理效率。2024 年，作为 Transformer 挑战者的 Mamba 架构[21] 出现，基于它的 Mamba Vision[22] 提出新型骨干网络，显著提升了视觉任务的计算和内存效率。然而，尽管上述模型在整体分割精度上取得了突破，它们在复杂街景场景下的边界精细化分割方面仍存在挑战，且普遍存在计算复杂度高和参数量大的问题[23]。

为了适应实际应用场景的需求和资源限制，高效且实时的语义分割算法成为了研究热点[24]。BiSeNet[25] 提出了空间路径和语义路径的双分支结构，空间路径用于捕获低层次的细节特征；语义路径则通过快速下采样获取高层次的语义信息，实现了对上下文信息和空间细节的有效平衡。BiSeNetV2[26] 在此基础上引入了细节分支和语义分支，细节分支使用浅层网络快速捕获空间细节，语义分支则采用深层网络提取丰富的语义信息。通过深度设计和高效的特征融合机制，DDRNet[27] 能在保持高精度的同时实现实时推理。PIDNet[28] 采用了独特的多分支结构，一个负责提取丰富语义信息的主分支和一个专注于捕获细粒度边界细节的辅助分支。尽管上述轻量级模型在速度与精度平衡方面表现优异，但在边界精细化分割方面仍存在局限[29]。例如，PIDNet 的辅助分支虽能提供边界细节信息，但其双分支间的信息融合机制往往不够充分，导致边界信息未能被主分支有效利用以显著提升最终分割图的边缘质量。同时，在追求高效的过程中，浅层网络或快速下采样策略容易损失部分精细的边界信息。

基于上述研究背景，如何有效捕获并利用精细边界信息以提升街景分割精度仍是一个亟待解决的关键问题。为进一步提升街景语义分割的精度，本文构建了一种融合边界细化模块的双分支分割网络 (BRNet)，主要贡献如下：1) 提出了一种面向边界细化的双分支协同优化机制。该机制的核心是边界细化模块 (boundary refinement module, BRM)，BRM 专注于在低分辨率分支增强边界表征，利用注意力机制强化边缘细节，确保图像的边界和轮廓信息得以保留和突出，提升分割结果的边界精度；2) 该机制的另一重要组成部分是自适应采样卷积 (adaptive sampling convolution, ASC)，作为关键支撑技术，其通过其动态感受野特性，提升了网络对复杂边界几何形态的适应能力，为精准的边界特征提取奠定了基础；3) 此外还引入了特征传递模块 (feature propagation module, FPM)，FPM 确保增强后的边界信息能有效、及时地反馈至高分辨率分支并进行融合，从而在整个网络中对边界区域实现持续性优化；4) 构建损失函数，融合交叉熵损失 (cross entropy loss) 和洛瓦兹损失 (Lovasz loss)，优化了交并比指标，使得模型在训练过程中更加关注分割的全面性能。

2 本文算法

2.1 网络结构

本文提出的街景语义分割网络结构 BRNet 如图 1 所示，整体结构主要由三部分构成：边界协同优化机制、

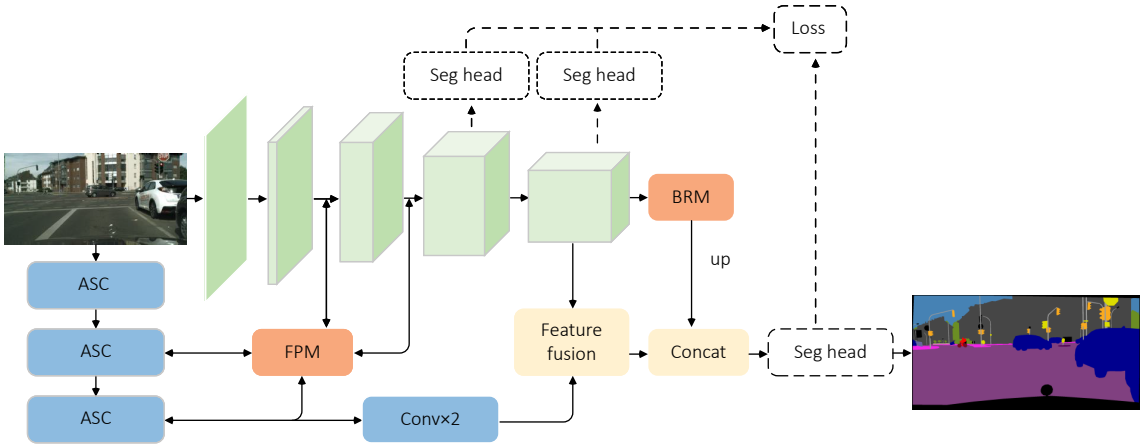


图 1 BRNet 整体结构

Fig. 1 Overall architecture of BRNet

特征传递模块以及融合损失函数。以下将结合图中各模块进行具体阐述。边界协同优化机制由 BRM 和 ASC 共同组成。如图 1 所示，高分辨率分支由三个自适应采样卷积模块构成，右侧标注“Conv×2”，表示通过两次卷积操作增强特征表达能力。该分支保留输入图像的细粒度空间信息，通过自适应采样卷积动态调整感受野，有效捕捉复杂场景中的细节特征。低分辨率分支由多个浅绿色模块组成，集成边界细化模块，用于提取高层次语义信息，并通过注意力机制强化边缘区域的细节，提升边界分割的准确性。FPM 位于高、低分辨率分支之间，通过通道压缩和下采样实现跨层次特征融合。损失函数设计方面，BRNet 融合了交叉熵损失和洛瓦兹损失，图中各模块上方标注的“Loss”表示对应部分的监督信号。该设计兼顾全局分类准确性与局部区域的分割质量，特别在应对类别不均衡问题时表现优异。最终，经过多级模块处理及特征融合后，网络输出精确的分割结果，如图 1 右下角所示。

2.2 边界协同优化机制

BiSeNetV2 在实时语义分割任务中表现出色，其采用了双分支网络结构，分别负责捕捉丰富的语义信息和精细的空间细节。这种设计不仅提升了分割的准确性，同时也保证了计算效率，适合在资源受限的环境中使用。然而，BiSeNetV2 采用的深度可分离卷积在处理具有大尺度变化的目标时表现不够理想。例如，在城市街景图像中，车辆、行人以及建筑物等目标的尺寸差异显著，单一的卷积方式难以适应这种多样性的需求。针对这一问题，本文借鉴了文献 [30-31] 的思路，并引入了 ASC，其卷积核被设置为 3×3，以实现计算复杂度和分割精度的平衡。如图 2 所示，该模块的核心创新在于其动态采样机制：首先，通过“动态邻接 (dynamic adj)”根据输入特征生成空间自适应的权重矩阵；进而，“几何变换 (geom trans)”模块利用该权重矩阵，动态地确定卷积采样窗口的形状、大小与空间位置，使其能够紧密贴合不同尺度的目标轮廓，如图 2 中精准聚焦于车辆区域的可变采样窗口。相比固定结构的卷积核，ASC 能够灵活调整感受野，从而更有效地聚合多尺度目标的关键特征信息。

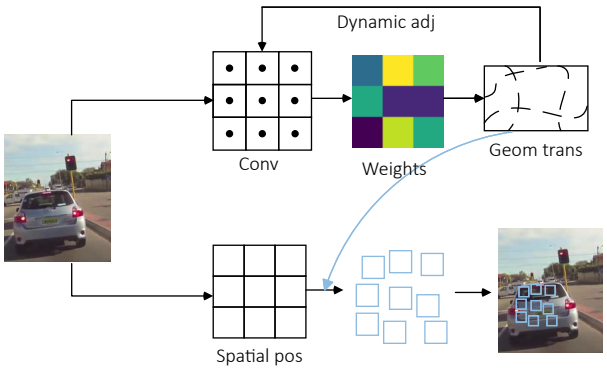


图 2 自适应采样卷积

Fig. 2 Adaptive sampling convolution

特征图首先通过自适应特征采样模块进行处理，该模块以输入特征图 $\boldsymbol{X} \in \mathbb{R}^{C\times H\times W}$ 为输入，通过动态生成调整参数实现卷积核采样策略的灵活调节。对于特征图中的任意空间位置 $p=(p_x, p_y) \in \mathbb{R}^2$，模块通过偏移量卷积层 *offsets* $(\cdot)$ 输出动态调整参数 $\Delta k \in \mathbb{R}^{2\times K^2}$ (其中 K 为卷积核尺寸)，其计算公式为式 (1)：

$$\Delta k = Conv_{\Delta}(\boldsymbol{X}), \tag{1}$$

式中：$Conv_{\Delta}(\cdot)$表示具有$2K^2$输出通道的标准卷积操作，与传统可变形卷积中独立于输入内容的固定偏移学习不同，ASC 模块的偏移量直接由输入特征图通过卷积动态预测，确保偏移量与局部内容高度相关，实现采样位置的空间精细化微调。

同时，若启用调制机制，模块通过掩码卷积层$masks(\cdot)$生成可学习权重$w \in \mathbb{R}^{K^2}$，并经 Sigmoid 函数归一化得到$w = \sigma(Conv_w(\boldsymbol{X})) \in (0,1]^{K^2}$，权重$w(k)$中的每个值均由输入特征图$\boldsymbol{X}$通过卷积操作动态预测并经非线性激活得到，其值与输入特征的局部内容高度相关。这意味着，对于不同的输入或同一输入的不同区域，$w(k)$会根据局部特征的显著性自适应调整，从而实现对重要特征区域的高权重分配。该权重并非全局固定，而是与输入特征的局部显著性直接关联。与传统静态卷积的固定采样式 (2) 不同，ASC 模块的自适应采样式 (3)：通过引入位置依赖的动态偏移$\Delta k(p)$和调制权重$w(k)$，实现了卷积核采样位置的空间自适应调整与采样重要性的动态分配，其中：$W(k)$ 表示基础权重，$p+k+\Delta k(p)$表示经几何变换微调后的采样坐标，通过双线性插值可将非整数坐标映射至特征图实际像素值。

$$F(p) = \sum_{k \in \mathcal{K}} W(k) \cdot \boldsymbol{X}(p+k), \tag{2}$$

$$F(p) = \sum_{k \in \mathcal{K}} w(k) \cdot W(k) \cdot \boldsymbol{X}(p+k+\Delta k(p)). \tag{3}$$

这种机制使感受野呈现出与输入内容相关的动态特性：$R(p) = \{p+k+\Delta k(p) \mid k \in \mathcal{K}\} \subseteq \mathbb{R}^2$，从而在捕捉复杂几何变换和细腻纹理特征时表现更优，同时通过权重$w(k)$对不同区域特征赋予差异化重要性权重，增强了模型对局部细节的敏感性与复杂背景下的目标识别鲁棒性，最终经形变卷积操作、归一化和激活函数输出变换后的特征图。

ASC 通过微调卷积核的采样策略，形成了灵活可变的感受野，使模型在捕捉复杂的几何变换和细腻的纹理特征方面表现得更加出色。本文算法将 ASC 嵌入到网络的高分辨率分支和自适应采样融合模块 (adaptive sampling fusion module, ASFM) 中，为后续的边界细化模块提供了高质量、强判别性的特征基础。此外该卷积所提取的丰富且空间精准的特征，通过 FPM 输送至 BRM，为其后续进行边界特征的聚焦与优化提供了关键的信息支撑，从而共同促成边界感知能力的系统提升。

在城市街景语义分割中，边界分割不清晰是普遍存在的一个挑战。这主要由于复杂的场景环境和多样化的目标形态，使得语义边界常常显得模糊且不连续。为解决这一问题，本文提出 BRM，其结构如图 3 所示。

与 Gated-SCNN 通过独立形状流分支建模边界不同，BRM 通过轻量级注意力实现类似效果，并获得更高的计算效率。BRM 首先通过两个连续的 3×3 卷积层处理输入特征图$\boldsymbol{X} \in \mathbb{R}^{C \times H \times W}$。第一个卷积层提取输入的局部特征并增强特征表达能力，为后续的边界细化提供更丰富的特征信息；第二个卷积则细化和整合特征信息，以保留更多的边缘细节。

$$\boldsymbol{X}_1 = ReLU(F_\theta(\boldsymbol{X}) + b_1), \tag{4}$$

$$\boldsymbol{X}_2 = F_\theta(\boldsymbol{X}_1) + b_2, \tag{5}$$

式中：F_θ表示卷积操作；b 为偏置项。通过双卷积层的设计，BRM 将感受野从 3×3 扩展到了 5×5[32]，其理论计算为

$$RF_l = RF_{l-1} + (k_l - 1) \times \prod_{i=1}^{l-1} s_i, \tag{6}$$

式中：$RF_{l-1} = 3$，表示经过第一层 3×3 卷积后的感受野；$k_l = 3$为第二层卷积核尺寸；步长$s_i = 1$，代入公式得$RF_l = 5$。这一设计有效地捕捉更广泛的上下文信息，从

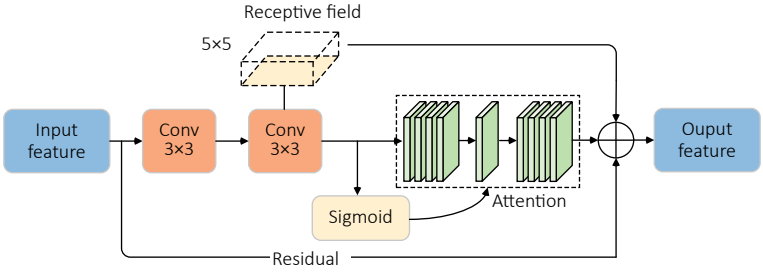


图 3 边界细化模块

Fig. 3 Boundary refinement module

而能够更准确地定位物体边界。此外，进一步引入的通道注意力机制是 BRM 的另一关键改进，通过全局平均池化生成通道描述：

$$z = GAP(\boldsymbol{X}_2) \in \mathbb{R}^C , \tag{7}$$

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \boldsymbol{X}_{2,c}(i,j) . \tag{8}$$

随后将通道数降至原来的 1/4，再通过升维卷积重建通道权重，并使用 Sigmoid 函数 (σ) 进行归一化处理，生成细粒度的通道注意力权重：

$$a = \sigma(F_\theta(ReLU(F_\theta(z)))) . \tag{9}$$

特征图按通道加权：$\boldsymbol{Xatt} = \boldsymbol{X}_2 \odot a$。注意力权重作用于双卷积后的特征图，而非原始输入，确保注意力聚焦于上下文富集的特征，而非噪声干扰的原始信号。此设计使模型能自适应地增强与边界相关的关键通道，同时抑制不相关的背景信息。为了进一步增强网络的表达能力和改善梯度流动，BRM 还采用了残差连接，缓解了深层网络中的梯度消失问题，也促进了特征的有效传递和融合。

不同于通用模块的全层泛化，BRM 专为低分辨率分支末端设计，解决了低分辨率分支在多次下采样中导致的边界模糊问题。提升了模型在边界分割方面的表现，增强了整体分割效果的精确性和鲁棒性。

2.3 特征传递

高分辨率分支能够获取图像的细节信息，提供精准的空间信息，但由于其感受野较小，限制了对上下文信息的捕获能力。低分辨率分支具有较大的感受野，能捕获丰富的语义信息，但其多次下采样操作会导致图像细节丢失，从而引发信息流失问题。常见的双分支网络通常在最后一层采用特征融合模块来缓解这一问题，但这种方法往往存在滞后性和信息整合不充分的缺陷。为更有效地缓解信息流失问题，本文引入了 FPM，其结构如图 4 所示。在不同场景中，输入特征图的具体尺寸是由网络结构本身决定的，即取决于高低分辨率分支的阶段性下采样率。

输入特征图为$\boldsymbol{X} \in \mathbb{R}^{C \times H \times W}$，经过高分辨率分支初始层$S_1$处理后生成特征图 $\boldsymbol{X}_S \in \mathbb{R}^{(H/2) \times (W/2) \times C_1}$，经低分辨率分支初始层$C_1$处理后，得到特征图 $\boldsymbol{X}_C \in \mathbb{R}^{(H/4) \times (W/4) \times C_2}$。为实现特征对齐与融合，模块首先对高分辨率特征执行通道压缩操作，通过 1×1 卷积将通道数压缩至C_2，得到$F_{\mathrm{C}}(\boldsymbol{X}_S) \in \mathbb{R}^{(H/2) \times (W/2) \times C_2}$；同时，对低分辨率特征$\boldsymbol{X}_C$执行上采样操作，将其空间尺寸从 ($H$/4,$W$/4) 提升至 ($H$/2,$W$/2)。特征交互过程通过双向残差连接实现，高分辨率分支特征更新公式如式 (10) 所示：

$$\boldsymbol{X}_{S1} = \boldsymbol{X}_S + F_{\mathrm{U}}(\boldsymbol{X}_c) , \tag{10}$$

式中：$\boldsymbol{X}_{S1}$为更新后的高分辨率特征，F_{U} 表示包含上采样与通道对齐的特征映射操作。此操作将低分辨率分支的丰富语义信息注入高分辨率特征，增强其上下文理解能力。

低分辨率分支特征更新公式如式 (11) 所示：

$$\boldsymbol{X}_{C1} = \boldsymbol{X}_C + F_{\mathrm{D}}(F_{\mathrm{C}}(\boldsymbol{X}_S)) , \tag{11}$$

式中：F_{D}表示下采样操作，F_{C} 表示压缩操作，将压缩后的高分辨率特征下采样至 (H/4)×(W/4) 尺度，使低分辨率分支能够捕获高分辨率分支的细节特征。

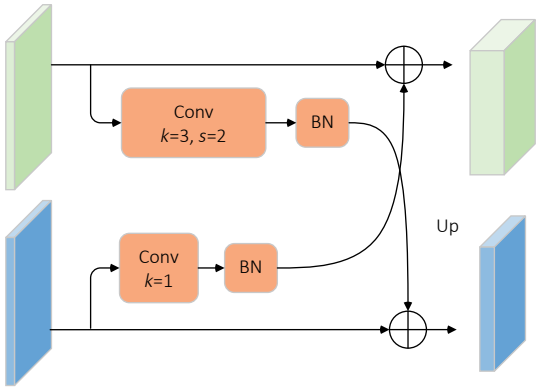


图 4 特征传递模块

Fig. 4 Feature propagation module

通过在每个分辨率层级构建动态特征传递路径，FPM 实现了跨分支、跨尺度信息的渐进式融合，且传递了自适应采样卷积捕获的高质量信息。假设网络有 N 个类似的关键节点，则在传统网络中，信息需要等待 N 个阶段的滞后，损耗是 $O(N^2)$。在引入 FPM 的网络中，每个 FPM 节点只处理当前阶段的微小信息滞后。由于每个阶段的信息差很小，融合的损耗也很小 (可视为常数 $O(1)$)。这样的融合操作进行了 N 次，因此总损耗是 $N \times O(1) = O(N)$。理论上将信息滞后损耗从 $O(N^2)$ 降低至 $O(N)$。同时通道压缩与分步融合的策略使整体计算量有所降低，通过引入 FPM，有效地缓解了信息流失问题，还增强了模型的特征表达能力。

2.4 融合损失函数

受拍摄效果和场景影响，城市街景图像中普遍存在

类别不平衡的问题，该问题使模型在训练过程中更倾向于预测占主导地位的类别，导致对稀有类别的分割性能显著下降。为有效解决这一问题，本文提出将交叉熵损失与洛瓦兹损失相结合，设计了一种融合损失函数，其数学表达如式 (12) 所示：

$$L_{\mathrm{BRNet}} = \alpha L_{\mathrm{Lovasz}} + \beta L_{\mathrm{Seg}} , \tag{12}$$

式中：α和β分别控制洛瓦兹损失和交叉熵损失的权重。

交叉熵损失用于衡量预测概率分布与真实标签分布之间的差异，其具体定义如式 (13) 所示：

$$L_{\mathrm{seg}} = -\sum_{i=1}^{C} y_i \cdot \log\left(\widehat{y_i}\right) , \tag{13}$$

式中：C 是类别总数；y_i是真实标签的 one-hot 编码；$\widehat{y_i}$是模型对第 i 类的预测概率。交叉熵损失函数在像素级别提供了精确的分类指导，但在类别不平衡时可能表现出局限性，容易导致模型对少数类的忽视。

洛瓦兹损失 (L_{Lovasz}) 通过直接优化交并比 (IoU) 指标，进一步提升了模型在类别不平衡分割上的表现。其具体定义如式 (14)：

$$L_{\mathrm{Lovasz}} = \frac{1}{C}\sum_{c=1}^{C}\sum_{i=1}^{n_c} \Delta_i^c \cdot m_i^c , \tag{14}$$

式中：C是类别总数；n_c是类别c的像素总数；m_i^c是类别c的第 i 个像素误差；Δ_i^c则为 IoU 梯度。洛瓦兹损失通过对预测像素的置信度进行排序，并将其与真实标签的理想排序进行对比，从而优化分割的整体结构性能。与传统的交叉熵损失不同，洛瓦兹损失将离散的 IoU 指标转化为一个连续且可微的函数，使得网络能够通过梯度下降直接优化分割的整体质量，而不仅仅是逐像素的分类准确性。

融合后的损失函数结合了交叉熵损失在像素级分类上的优势与洛瓦兹损失在优化整体分割质量上的能力。通过调整α和β的值，可以在训练过程中平衡这两种损失的影响，保证模型既能保持高精度的像素分类，又能有效处理类别不平衡分割的问题。

3 实验结果与分析

3.1 数据集和评价指标

为了评估本文所提出模型在街景图像语义分割任务中的表现，所有训练均在 Cityscapes 数据集[33] 上进行。Cityscapes 是语义分割领域最广泛使用的公共数据集之一，专注于城市街景场景，包含 5000 张精细标注的图像和 20000 张粗略标注的图像。本研究仅采用精细标注的部分数据，包括 2975 张训练集图像、500 张验证集图像和 1525 张测试集图像。该数据集涵盖了来自 50 个不同城市和多个季节的街景图像，图像内容丰富，场景复杂多变，适合用于模型性能评估。

为全面评估模型在复杂天气环境下的泛化能力与鲁棒性，本文进一步在 WildDash2[34] 和 BDD100K[35] 两个具有挑战性的数据集上进行了测试。WildDash2 是一个专注于自动驾驶场景鲁棒性评估的基准数据集，提供了覆盖全球多样化驾驶环境 (包括城市、高速公路及极端天气等) 的 4000 余张精细标注图像。BDD100K 则是一个大规模、多样化的自动驾驶数据集，包含 10 万张图像，具备像素级语义分割标注，覆盖多种天气条件、光照变化及不同地理场景，能够有效检验模型在未知场景中的适应能力。两个数据集的联合使用，为验证模型在真实复杂环境下的性能提供了充分依据。

本实验采用两个语义分割常用的评价指标平均交并比 (mIoU) 和平均像素精度 (mPA)。mIoU 越大，表明预测区域与真实区域的重叠程度越高，模型的预测准确性越高；mPA 越大，则表示模型在各类别上的像素级分类准确率越高。

mIoU 通过计算每个类别的真实标签与预测标签之间的交集与并集的比值，并取所有类别的平均值，计算公式如式 (15) 所示：

$$mIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} . \tag{15}$$

mPA 则通过计算每个类别正确分类的像素数与该类别总像素数的比值，并取所有类别的平均值，计算公式如式 (16) 所示：

$$mPA = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}} , \tag{16}$$

式中：k 表示类别数；p_{ii}表示被标注为第 i 类的样本中，模型正确预测为第 i 类的像素数；p_{ij}为真实类别为第 i 类的样本被错误预测为第 j 类的像素数。

3.2 实验环境与参数设置

本文的实验在 Ubuntu 20.04 操作系统上进行，硬件配备包括 NVIDIA RTX 2080 Ti GPU 和 Xeon (R)

Platinum 8255C CPU。深度学习框架选用 PyTorch 2.0.0，编程语言采用 Python 3.8。

在参数设置方面，采用小批量随机梯度下降 (mini-batch SGD) 作为优化器以更新网络参数，初始学习率设定为 0.025，动量 (momentum) 为 0.9，权重衰减 (weight decay) 为 5×10^{-4}。本研究还采用了翻转和随机裁剪等数据增强方法。其中，随机裁剪操作将输入图像与标签同步裁剪至固定尺寸 1024 pixel×1024 pixel，若原图尺寸不足则进行零填充；同时，为避免裁剪区域语义信息单一，设置了类别像素占比阈值 0.75。

3.3 消融实验

3.3.1 自适应卷积核尺寸分析

为在模型计算复杂度与分割精度之间实现良好平衡，本文选取了 3×3、5×5 和 7×7 三种卷积核尺寸，分别在 Cityscapes 数据集上对比测试了可变形卷积 (deformable convolution, DCN) 与 ASC 的性能表现，具体结果如表 1 所示。

从表 1 可以看出，卷积核尺寸的增大会显著增加模型复杂度和计算开销。对于 ASC，其参数量从 2.8 M 增至 5.0 M，计算量从 7.1 G 增至 12.0 G，然而其 mIoU 仅从 76.18% 微幅提升至 76.45%，增益十分有限。同时可变形卷积在核尺寸增至 7×7 时甚至出现了精度下降。这表明盲目增大卷积核并非有效策略。因此，为在计算效率与性能间取得最佳平衡，本文采取了一种针对性策略：仅在网络第一个 ASC 层使用 7×7 卷积核，以利用其更大的初始感受野来捕捉全局上下文信息；其余层均采用 3×3 卷积核，以高效地维持模型的整体性能并控制计算成本。

3.3.2 边界细化模块的对比验证

为验证本文方法中核心模块 BRM 的性能，分别采用 Gated-SCNN 中的 GS-Gate、FaPN[36] 中的 AlignFeat 以及 Sun 等提出的 BEM[37] 作为对比方法，单独构建网络并进行对比实验。实验结果如表 2 所示。

由表 2 可知，尽管 BRM 的分割精度略低于 GS-Gate，但其在模型复杂度与推理效率之间取得了更好的平衡。与现有边界细化模块相比，BRM 在显著减少参数规模和计算开销的同时，仍能保持具有竞争力的语义分割精度。

3.3.3 模块贡献度与协同效应分析

在确定了最优卷积核配置后，为进一步验证 BRNet 中各模块对分割性能的提升效果，本文在 Cityscapes 数据集上进行了系统的消融实验，具体结果见表 3。

第一组实验复现了 BiSeNetV2 的结果，作为本研究的基线 (Baseline)；第二组验证传统可变形卷积在网络中的效果；第三组至第八组实验则在此基础上逐步添加各模块，以观察各模块单独及组合使用时对整体性能的影响。

前四组实验结果显示，加入传统的可变形卷积导致模型的分割精度下降，这可能是因为偏移量的全局学习易受背景噪声干扰。而提出的 ASC 模块、BRM 模块和 FPM 模块均有助于提高分割精度。加入 ASC 模块后，

表 1 不同卷积核配置在 Cityscapes 数据集上的分割性能对比

Table 1 Performance comparison of different convolution kernels on Cityscapes

Kernel size	Params/M	FLOPs/G	mIoU/%
DCN (*k*=3)	2.5	6.5	74.67
DCN (*k*=5)	3.8	8.5	75.11
DCN (*k*=7)	4.5	10.8	74.05
ASC (*k*=3)	2.8	7.1	76.18
ASC (*k*=5)	4.2	9.5	76.07
ASC (*k*=7)	5.0	12.0	76.45

表 2 不同边界细化模块的性能对比

Table 2 Performance comparison of different boundary refinement modules

	Params/M	FLOPs/G	mIoU/%
GS-Gate	38.5	78.9	77.88
AlignFeat	16.8	45.2	75.09
BEM	5.1	12.5	75.65
BRM	8.3	21.7	77.85

mIoU 较 Baseline 提升了 0.44%。该模块通过微调卷积核的采样策略，形成了灵活可变的感受野，增强了模型对不同尺度目标的适应能力。BRM 模块的加入使 BRNet 的 mIoU 相比 Baseline 提高了 2.11%，通过注意力机制和残差连接强化了对边界信息的捕获能力，从而能保留并提取更多边界细节。FPM 模块单独使用时，mIoU 较 Baseline 提升了 0.23%，该模块在保留高分辨率分支细节信息的同时，充分融合了低分辨率分支的语义信息。

第六组至第八组实验分别结合上述模块，以评估其协同效应。第六组实验结合 ASC 模块和 BRM 模块，相较 Baseline 提升了 3.89%。第七组实验结合 ASC 模块和 FPM 模块，相较 Baseline 提升了 1.36%。第八组实验结合 BRM 模块和 FPM 模块，相较 Baseline 提升了 2.49%。第九组实验同时引入 ASC、BRM 和 FPM 三个模块，验证了各模块联合应用的有效性，BRNet 的 mIoU 较 Baseline 提高了 5.69%，证明了各模块协同作用对提升模型分割性能的贡献。

消融实验结果表明，本文提出的模块在不同层面上优化了模型的特征表达和信息融合能力，提升了城市街景图像语义分割的准确性。

3.3.4 边界协同优化机制分析

前述实验分别验证了各模块的有效性及内部协同效应。为更全面地评估本文所提出的 BRM、ASC、FPM 构成的边界协同优化机制的优势，本小节设计了与流行的同功能模块的对比组合实验。其中，边界增强模块 BEM 引导低层网络学习边界信息并优化全局信息；DCN 通过动态调整采样位置来增强模型对几何形变和复杂结构的适应性。具体结果见表 4。

由表 4 可知，本文提出的边界细化模块 (BRM) 在各项组合中均展现出核心优势。其与自适应采样卷积 (ASC) 的协同作用尤为突出：BRM+ASC 组合的性能显著优于对比组合，这得益于 BRM 在边界精细化与 ASC 在自适应感受野调整上的高效互补。相比之下，BEM+ASC 组合性能相对较低，说明 BRM 在边界处理上更具优势。当引入特征传递模块 (FPM) 构建 BRM+ASC+FPM 组合时，模型性能达到了最高的 81.43%，这表明 FPM 通过多尺度特征的有效传递与融合，显著增强了 BRM 与 ASC 的协同效应。实验充分证明，BRM 作为核心模块，与 ASC、FPM 等组件共同构

表 3　消融实验结果

Table 3　Results of ablation experiments

NO.	Baseline	DCN (3×3)	ASC (3×3)	BRM	FPM	mIoU/%
1	√					75.74
2	√	√				74.67
3	√		√			76.18
4	√			√		77.85
5	√				√	75.97
6	√		√	√		79.63
7	√		√		√	77.10
8	√			√	√	78.23
9	√		√	√	√	81.43

表 4　模块组合消融实验对比结果

Table 4　Comparison results of module combination ablation experiments

NO.	BRM	BEM	ASC (3×3)	DCN (3×3)	FPM	Module	mIoU/%
1		√		√		BEM+DCN	75.84
2		√	√			BEM+ASC	76.81
3	√			√		BRM+DCN	78.23
4	√		√			BRM+ASC	79.63
5		√		√	√	BEM+DCN+FPM	76.64
6		√	√		√	BEM+ASC+FPM	77.70
7	√			√	√	BRM+DCN+FPM	78.66
8	√		√		√	BRM+ASC+FPM	81.43

成了边界协同优化机制，能显著提升复杂场景下的分割精度。

3.4 定量分析

为了验证本文所提出的 BRNet 模型在不同环境下的综合性能和可靠性，本文将 BRNet 与多种具有代表性的语义分割网络进行了对比实验。所选对比模型包括轻量级网络 LMIINet[38]，多分支架构的 BiSeNetV2 和 DDRNet，以高精度为目标的模型 SegNeXt[39]，基于 Transformer 架构的 EfficientViT[40] 和 SCTNet[41] 等，以及融合 Mamba 架构的 SegMAN[42]。这些模型在语义分割领域具有较高的影响力和广泛的应用基础[43]。实验中采用的评价指标包括 mIoU、mPA 以及帧率 (FPS) 来衡量模型的分割精度和复杂度。mIoU 和 mPA 用于评估模型在分割任务中的性能表现，帧率则用于评估模型的推理速度 (FPS)。具体实验结果如表 5 所示。

实验结果表明，本文提出的 BRNet 在分割精度和推理速度之间实现了良好的平衡。BRNet 的 mIoU 达到 81.43%，在分割精度上优于大部分对比算法，仅次于 EfficientViT；同时，其 FPS 达到 67 f/s，展现出高效的实时处理能力。这一优势源于自适应采样卷积和灵活的采样策略，有效捕获了不同尺寸目标的特征。此外，BiSeNetV2 等多分支架构由于仅在最后一层进行特征交互导致无法有效利用双分支信息，本文通过引入特征传递模块有效缓解了这一问题。

融合 Mamba 架构的 SegMAN 注重局部细节，并使用滑动窗口局部注意力机制增强对局部特征的捕获能力。PEM 等模型通过交叉自注意力机制能够有效的获取上下文信息，但堆叠 Transformer 块使其推理速度较慢。相比之下，BRNet 采用的边界细化模块仅略微增加参数量，这得益于轻量级注意力机制，不仅延迟较小，还提供了丰富的边缘信息，这表明其在资源受限的实际应用环境中具有更高的适用性和实用价值。

与 BiSeNetV2 和 DDRNet 等注重实时的模型相比，BRNet 在分割精度略胜一筹，以较小的延迟换取了较大的提升。

图 5 展示了各算法在 Cityscapes 数据集上的 mIoU 与 FPS 的散点分布，从而更直观地分析不同算法的性能表现。从图上可以看出本文模型相较其他算法的优越性。

为了全面评估 BRNet 在处理小目标边界分割方面的表现，本文选取了 Cityscapes 数据集中多个常见且具有代表性的小目标类别进行详细分析，包括行人 (person)、自行车 (bicycle)、交通信号灯 (traffic light) 和交通标志 (traffic sign) 等。这些类别在实际交通场景中具有重要意义，然而由于其目标尺寸较小，分割难度较大。实验结果如表 6 所示，展示了 BRNet 与其他对比模型在这些小目标类别上的 mIoU。

结果表明，BRNet 在所有选定的小目标类别中均实现了性能提升。例如，在摩托车类别上，BRNet 相比于对比模型的 mIoU 提升了 5.65% 以上；在交通信号灯类别上，BRNet 的表现也较对比模型提升了超过 5.77%。这一成果主要归功于 ASC 和 BRM 的协同工作，使得模型即使对小尺度目标也能生成质量更高的边界预测。

3.5 定性分析

为了直观展示本文算法的提升效果，本文对模型、BiSeNetV2 和 DDRNet 进行了定性比较分析，结果如图 6 所示。图 6 的第一列展示了原始图像，第二列为对应的真实标签即标注值，而第三至第五列分别呈现了

表 5 对比试验结果

Table 5 Results of comparison experiments

Method	Reference	mIoU/%	mPA/%	FPS
LMIINet	[38]	72.00	—	87
BiSeNetV2	[26]	75.74	81.81	73
GCNet	[44]	77.30	—	62
DDRNet	[27]	77.97	86.01	58
PIDNet	[28]	78.74	86.44	60
SegNeXt-tiny	[39]	78.55	86.14	66
PEM	[45]	79.00	—	40
SCTNet	[41]	79.80	—	31
SegMAN	[42]	80.30	—	53
EfficientViT	[40]	82.10	—	49
BRNet (Ours)	—	81.43	88.82	67

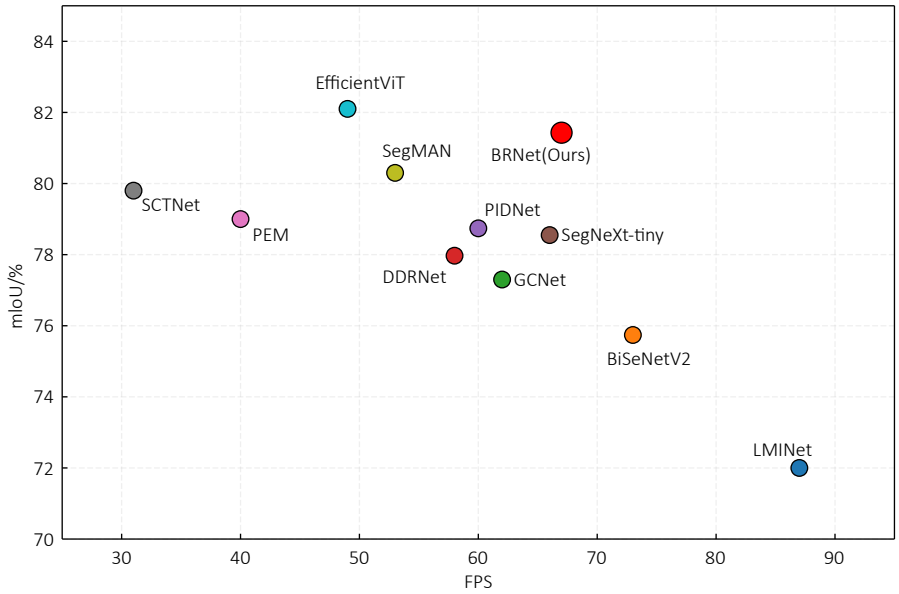


图 5 Cityscapes 数据集上的 mIoU-FPS 散点图

Fig. 5 mIoU-FPS scatter plot on cityscapes

表 6 小目标对比试验结果 IoU/%

Table 6 Results of small object comparison experiments IoU/%

Method	mIoU	Pole	Traffic light	Traffic sign	Person	Motorcycle	Bicycle
BiSeNetV2	75.74	61.09	67.30	76.21	78.32	53.45	73.81
DDRNet	77.97	64.43	69.99	77.94	80.95	59.08	76.29
PIDNet	78.74	65.14	71.48	79.06	81.75	62.33	77.25
BRNet (ours)	81.43	70.14	77.25	83.04	84.66	67.98	80.03

BiSeNetV2、DDRNet 以及 BRNet 的分割结果。选择 BiSeNetV2 和 DDRNet 作为定性对比对象，主要基于二者在实时语义分割领域的代表性和与本文工作的可比性。BiSeNetV2 作为公认的双分支基准模型，其架构与 BRNet 最为相似，便于进行公平且直观的对比，从而直接凸显 BRNet 的改进成效。DDRNet 则是另一类在实时性与精度平衡方面表现卓越的高效模型，其设计目标与本文高度一致，有助于展示 BRNet 在同类先进模型中的竞争力。从图 6 的分割结果可以发现：BRNet 在第一行示例中能够准确分割出车辆后方的行人以及远处阴影中的骑行者，这证明 BRNet 在处理遮挡和光照复杂的场景中具有较强的鲁棒性。第三行，BRNet 不仅准确分割出了远处楼梯上的行人，还完整地分割出交通标志，显示了其在处理远距离和小目标边界分割时的优越性能，能够有效捕捉边界信息。第四行，BRNet 在阴影处的遮挡人物分割上表现出色，同时对地形的分割也较为准确，表明其在复杂背景下依然能够保持高精度的分割效果，能够有效区分前景和背景。第五行则展示了 BRNet 能够完整地分割出被部分遮挡的车尾和远处的人物，即使在目标部分被遮挡的情况下，依然能够细致地识别目标边界并有效分离重叠或遮挡的物体。这些结果证明了 BRNet 在处理以边界模糊为代表的复杂场景时，仍能保持高精度的语义分割效果。这得益于 BRNet 的 BRM 和 ASC。

为验证本文模型在复杂城市环境下的泛化能力，使用 WildDash2 数据集对 Cityscapes 上训练的模型进行测试，所有模型未进行任何微调或域适应处理。结果如表 7 所示。

实验结果表明，在训练变量完全相同的情况下，本文提出的 BRNet 模型 WildDash2 数据集上展现出优异的跨数据集泛化能力。BRNet 以 35.81% 的 mIoU 和 50.27% 的 mPA 显著超越多数对比模型，其中 mIoU 指标较次优的 SegNeXt-tiny 提升 0.44%；较 DDRNet 则提升了 18.83%，验证了模型对恶劣天气干扰的鲁棒性。值得注意的是，所有模型在 Cityscapes 上训练后直接迁移测试的性能均出现显著下降，这凸显了天气因素对语

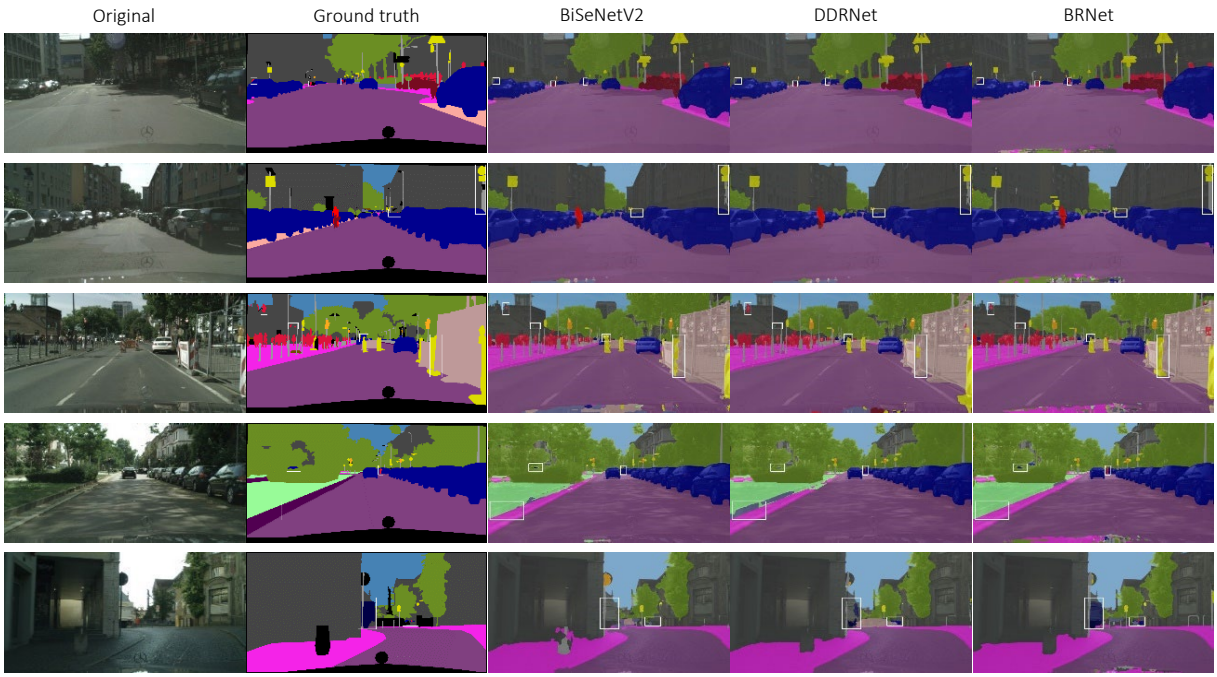


图 6 Cityscapes 数据集可视化分析结果

Fig. 6 Visual segmentation results of the Cityscapes dataset

义分割任务的挑战性。SegNeXt-tiny 的优异表现益于其多尺度卷积注意力，而本文方法通过边界细化模块解决边界模糊问题，提升了模型对雨雾等干扰的抑制能力。图 7 展示了模型在 WildDash2 数据集上的可视化结果。

表 7 WildDash2 数据集上的对比试验结果/%

Table 7 Comparative experimental results on the WildDash2 dataset/%

Method	mIoU	mPA
LMIINet	13.67	—
GCNet	25.35	—
BiSeNetV2	28.41	43.16
DDRNet	16.98	28.36
PIDNet	21.12	33.16
SegNeXt-tiny	35.37	50.19
BRNet (Ours)	35.81	50.27

如图 7 所示，在黄昏等低光照条件下，BiSeNetV2 和 DDRNet 等模型的分割结果出现明显的语义混淆和错误分类现象；而在雨雪天气干扰下，这些模型则表现出分割失效，部分目标甚至完全漏检。此外，在复杂城市场景中，对比模型的边缘分割结果模糊不清。这些问题主要源于特征提取能力受限，进而导致在极端光照、噪声干扰或复杂背景下出现边界信息丢失或特征退化。相比之下，得益于边界细化模块和自适应采样卷积带来的边界信息，BRNet 能够在上述环境下实现主体部分的准确分割分类，并展现出较好的泛化能力。

为进一步验证本文模型在更广泛场景下的泛化能力，在 BDD100K 数据集上对在 Cityscapes 上训练的模型进行了测试。所有模型同样未进行任何微调，以公平评估其跨域泛化性能。定量结果如表 8 所示。

实验结果表明，在 BDD100K 数据集上，BRNet 模型同样保持了领先的性能，其 mIoU 达到 43.18%，相较于对比模型表现良好。这证明了 BRNet 在应对不同城市环境、雨雪天气夜晚等场景下时具有较强的适应能力。

图 8 展示了各模型在 BDD100K 数据集上的定性对比结果。

从可视化效果可以看出，在光照良好的场景中(第一行示例)，BRNet 仍能保持精确的分割效果，说明模型在理想条件下具备良好的鲁棒性。在第二行的行人分割结果中，相较于 BiSeNetV2 和 DDRNet，BRNet 在目标结构的完整性方面表现更优，反映出边界协同优化机制的有效性。从第三行至第五行的示例可见，场景的亮度与能见度自雨天、桥下低光照至夜晚依次递减。相比对比模型出现的大范围漏分割与误分割现象，BRNet 能够有效识别并分割出图像中的关键场景要素，表明该模型在复杂环境下仍具备可靠的分割能力与较强的鲁棒性。

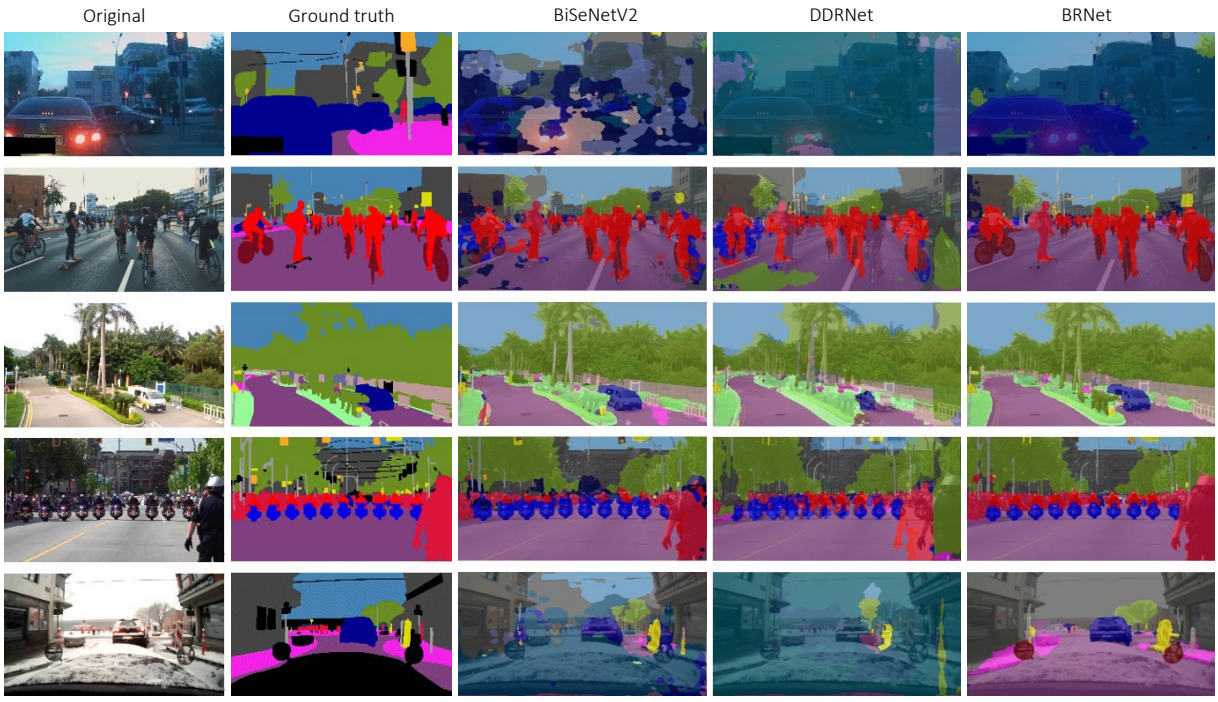


图 7 WildDash2 数据集可视化分析结果

Fig. 7 Visual segmentation results of the WildDash2 dataset

表 8 BDD100K 数据集上的对比试验结果/%

Table 8 Comparative experimental results on the BDD100K dataset/%

Method	mIoU	mPA
LMIINet	23.58	—
GCNet	33.49	—
BiSeNetV2	31.02	51.93
DDRNet	35.74	51.82
PIDNet	28.36	45.39
SegNeXt-tiny	43.27	60.08
BRNet (Ours)	43.18	61.12

通过类别激活图 (class activation map, CAM) 对模型的特征关注区域进行可视化分析，可进一步验证模型的分割能力和特征关注机制，相关结果见图 9。

针对“汽车”类别的 CAM 分析 (图 9 第一行) 表明，BRNet 的结果能够准确覆盖车辆的主体区域，并且对车辆边缘的响应更加清晰。在道路边缘区域的两辆汽车示例中，BRNet 不仅能够完整地捕捉到车辆的轮廓，还能有效区分车辆与周围背景的边界。相比之下，BiSeNetV2 的 CAM 结果虽然也能关注到车辆的主体，但在边缘区域的响应较为模糊，导致部分车辆的轮廓出现扩散现象。而 DDRNet 的 CAM 结果则存在一定的背景干扰，尤其是在车辆周围的区域，模型的关注点有部分偏离目标，导致车辆与背景的区分不够明确。

在“行人”类别的 CAM 图中 (图 9 的第二行)，BRNet 的结果能够清晰地分离每个行人，并且对行人边缘的响应非常集中。BiSeNetV2 的 CAM 结果在行人密集区域的表现略显不足，部分行人之间的特征响应出现了重叠，导致模型难以区分个体。DDRNet 的 CAM 结果虽然能够覆盖行人区域，但在行人边缘的关注上存在一定的模糊性，部分行人身体的特征响应较弱，导致细粒度特征丢失。

得益于边界细化模块配合自适应采样卷积，BRNet 在处理“汽车”和“行人”这两个目标类别时，展现出了更加精准和集中的特征关注区域。实验结果证明，BRNet 的 CAM 结果在目标边界的捕捉和区域完整性上具有明显优势。这些可视化证据从特征关注层面验证了 BRNet 在复杂场景分割任务中的优越性。

4 结 论

本文针对街景语义分割中边界模糊这一关键问题，提出了 BRNet 模型，其核心在于通过 BRM 与 ASC 的协同机制，实现了对物体边界的持续性优化。实验结果表明，本文设计的 BRNet 在 Cityscapes 数据集上的平均

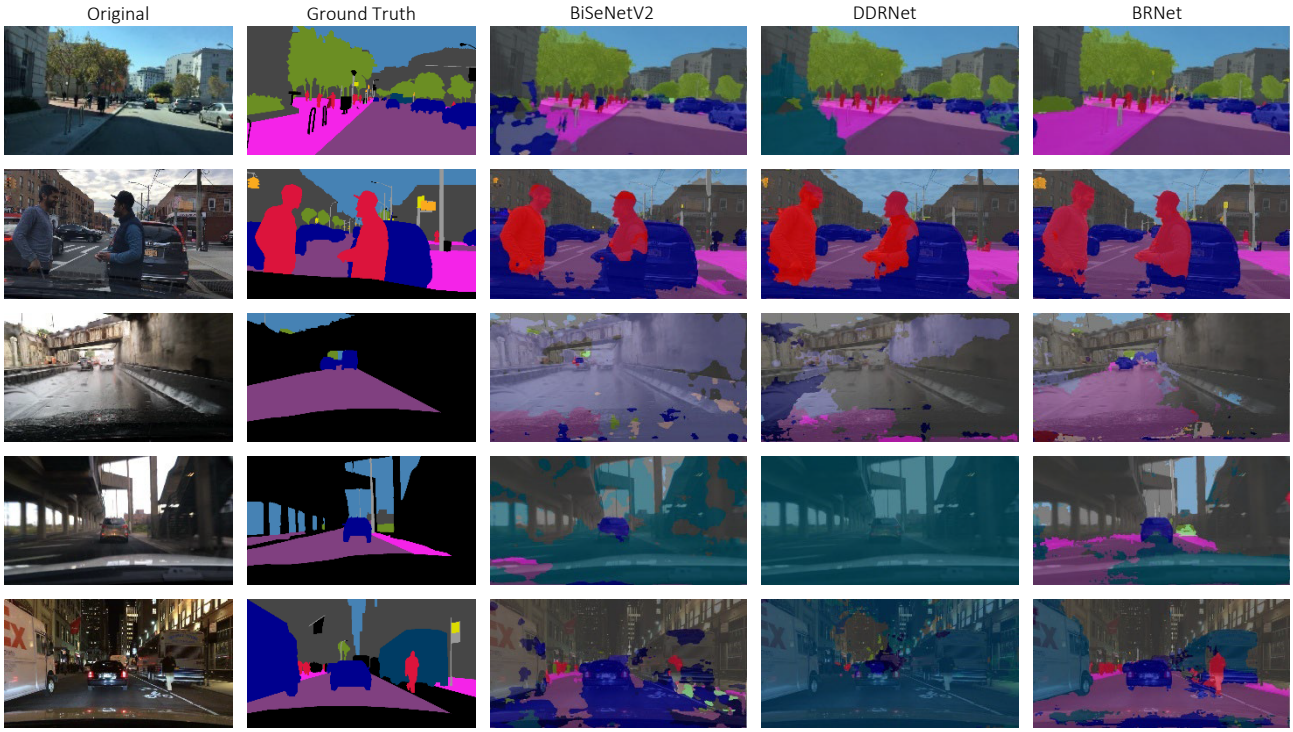


图 8 BDD100K 数据集可视化分析结果

Fig. 8 Visual segmentation results of the BDD100K dataset

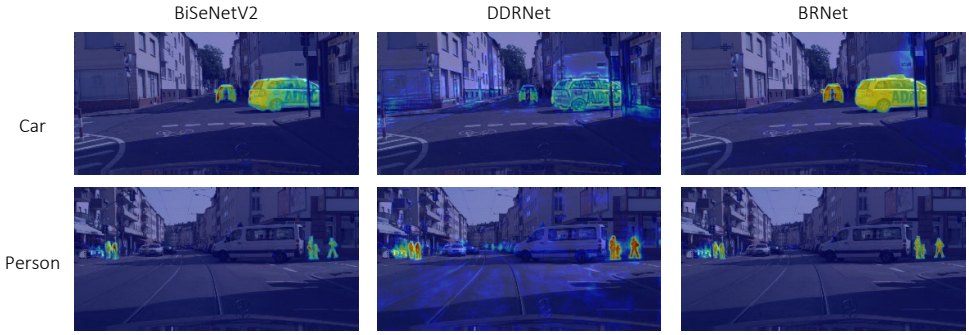


图 9 类别激活图实验结果

Fig. 9 Results of class activation map

交并比达到 81.43%，显著提升了对物体边界，特别是模糊边界的分割性能，在分割精度和推理速度上取得了良好的平衡。在 WildDash2 和 BDD100K 数据集上，包括 BRNet、BiSeNetV2、DDRNet、LMIINet、SegNeXt 在内的模型，其分割精度均较 Cityscapes 数据集上的结果出现了不同程度的下降，但 BRNet 的性能下降幅度相对较小，展现出了较好的泛化能力。未来工作将重点优化模型，进一步提升其在实际部署中的推理效率。

作者贡献

叶涛：改进方法的提出，论文构思和撰写，实验数据整理与分析；林珊玲：改进方法的研讨，论文审核与编辑写作；林坚普：论文审核与编辑写作；吕珊红：调查研究；林志贤：论文思路方案和实验执行进度总体指导，论文审核；郭太良：提供课题实验条件与资源。

利益冲突

所有作者声明无利益冲突

参考文献

1. Liu B C, Cai H Y, Yang S Y, et al. Lightweight semantic segmentation network for autonomous driving scenarios[J]. *J Xidian Univ*, 2023, **50** (1): 118−128.

刘博翀, 蔡怀宇, 杨诗远, 等. 一种用于自动驾驶场景的轻量级语义分割网络[J]. 西安电子科技大学学报, 2023, **50** (1): 118−128.

2. Wu M J, Zhang Y A, Lin S L, et al. Real-time semantic segmentation algorithm based on BiLevelNet[J]. *Opto-Electron Eng*, 2024, **51** (5): 240030.
吴马靖, 张永爱, 林珊玲, 等. 基于 BiLevelNet 的实时语义分割算法[J]. 光电工程, 2024, **51** (5): 240030.
3. Hou Z Q, Cheng M J, Ma S G, et al. Real-time urban street view semantic segmentation based on cross-layer aggregation network[J]. *Opt Prec Eng*, 2024, **32** (8): 1212−1226.
侯志强, 程敏婕, 马素刚, 等. 基于跨层次聚合网络的实时城市街景语义分割[J]. 光学 精密工程, 2024, **32** (8): 1212−1226.
4. Hang H, Huang Y P, Zhang X R, et al. Design of Swin Transformer for semantic segmentation of road scenes[J]. *Opto-Electron Eng*, 2024, **51** (1): 230304.
杭昊, 黄影平, 张栩瑞, 等. 面向道路场景语义分割的移动窗口变换神经网络设计[J]. 光电工程, 2024, **51** (1): 230304.
5. Hao M, Bai H, Xu T T. Remote sensing image road extraction by integrating ResNeSt and multi-scale feature fusion[J]. *Opto-Electron Eng*, 2025, **52** (1): 240236.
郝明, 白鹤, 徐婷婷. 融合 ResNeSt 和多尺度特征融合的遥感影像道路提取[J]. 光电工程, 2025, **52** (1): 240236.
6. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//*2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015: 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965.
7. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[C]//*18th International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015: 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
8. Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. arXiv: 1706.05587, 2017. https://doi.org/10.48550/arXiv.1706.05587.
9. Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//*15th European Conference on Computer Vision – ECCV 2018*, 2018: 833–851. https://doi.org/10.1007/978-3-030-01234-2_49.
10. Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network[C]//*2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017: 6230–6239. https://doi.org/10.1109/CVPR.2017.660.
11. Yang M K, Yu K, Zhang C, et al. DenseASPP for semantic segmentation in street scenes[C]//*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 3684–3692. https://doi.org/10.1109/CVPR.2018.00388.
12. Wang J D, Sun K, Cheng T H, et al. Deep high-resolution representation learning for visual recognition[J]. *IEEE Trans Pattern Anal Mach Intell*, 2021, **43** (10): 3349−3364.
13. Li Z Y, Chang Y, Wu Y. SegRet: an efficient design for semantic segmentation with retentive network[EB/OL]. arXiv: 2502.14014, 2025. https://doi.org/10.48550/arXiv.2502.14014.
14. Takikawa T, Acuna D, Jampani V, et al. Gated-SCNN: gated shape CNNs for semantic segmentation[C]//*2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019: 5228–5237. https://doi.org/10.1109/ICCV.2019.00533.
15. Kirillov A, Wu Y X, He K M, et al. PointRend: image segmentation as rendering[C]//*2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020: 9796–9805. https://doi.org/10.1109/CVPR42600.2020.00982.
16. Zhang Y, Ma C M, Liu S D, et al. Multi-scale feature enhanced Transformer network for efficient semantic segmentation[J]. *Opto-Electron Eng*, 2024, **51** (12): 240237.
张艳, 马春明, 刘树东, 等. 基于多尺度特征增强的高效 Transformer 语义分割网络[J]. 光电工程, 2024, **51** (12): 240237.
17. Xie E Z, Wang W H, Yu Z D, et al. SegFormer: simple and efficient design for semantic segmentation with transformers[C]//*Proceedings of the 35th International Conference on Neural Information Processing Systems*, 2021: 924. https://doi.org/10.5555/3540261.3541185.
18. Dong B, Wang P C, Wang F. Head-free lightweight semantic segmentation with linear transformer[C]//*Proceedings of the 37th AAAI Conference on Artificial Intelligence*, 2023: 516–524. https://doi.org/10.1609/aaai.v37i1.25126.
19. Shi H C, Hayat M, Cai J F. Transformer scale gate for semantic segmentation[C]//*2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023: 3051–3060. https://doi.org/10.1109/CVPR52729.2023.00298.
20. Abid M M N, Mehta N, Wu Z W, et al. ContextFormer: redefining efficiency in semantic segmentation[EB/OL]. arXiv: 2501.19255, 2025. https://doi.org/10.48550/arXiv.2501.19255.
21. Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces[EB/OL]. arXiv: 2312.00752, 2023. https://doi.org/10.48550/arXiv.2312.00752.
22. Hatamizadeh A, Kautz J. MambaVision: a hybrid mamba-transformer vision backbone[C]//*2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025: 25261–25270. https://doi.org/10.1109/CVPR52734.2025.02352.
23. Li X, Liu J, Zhu K F. A review of Transformer applications and acceleration optimization techniques in image segmentation[J]. *Comput Eng Appl*, 2026, **62**(5): 57−72.
李霞, 刘俊, 朱克峰. Transformer 在图像分割领域的应用与加速优化技术综述[J]. 计算机工程与应用, 2026, **62**(5): 57−72.
24. Ren F L, Yang L, Zhou H B, et al. Real-time semantic segmentation based on improved BiSeNet[J]. *Opt Pre Eng*, 2023, **31** (8): 1218−1227.
任凤雷, 杨璐, 周海波, 等. 基于改进 BiSeNet 的实时图像语义分割[J]. 光学 精密工程, 2023, **31** (8): 1218−1227.
25. Yu C Q, Wang J B, Peng C, et al. BiSeNet: bilateral segmentation network for real-time semantic segmentation[C]//*15th European Conference on Computer Vision – ECCV 2018*, 2018: 334–349. https://doi.org/10.1007/978-3-030-01261-8_20.
26. Yu C Q, Gao C X, Wang J B, et al. BiSeNet V2: bilateral network with guided aggregation for real-time semantic segmentation[J]. *Int J Comput Vis*, 2021, **129** (11): 3051−3068.
27. Hong Y D, Pan H H, Sun W C, et al. Deep dual-resolution networks for real-time and accurate semantic segmentation of road scenes[EB/OL]. arXiv: 2101.06085, 2021. https://doi.org/10.48550/arXiv.2101.06085.
28. Xu J C, Xiong Z X, Bhattacharyya S P. PIDNet: a real-time semantic segmentation network inspired by PID controllers[C]//*2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023: 19529–19539. https://doi.org/10.1109/CVPR52729.2023.01871.

29. Gao C X, Xu Z Z, Wu D Y, et al. Deep learning-based real-time semantic segmentation: a survey[J]. *J Image Graph*, 2024, **29** (5): 1119−1145.
高常鑫, 徐正泽, 吴东岳, 等. 深度学习实时语义分割综述[J]. 中国图象图形学报, 2024, **29** (5): 1119−1145.
30. Wang Z Y, Liu D Y, Wu L Y, et al. A deep semantic segmentation network with semantic and contextual refinements[J]. *IEEE Trans Multimedia*, 2025, **27**: 4856−4868.
31. Xiong Y W, Li Z Q, Chen Y T, et al. Efficient deformable ConvNets: rethinking dynamic and sparse operator for vision applications[C]//*2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024: 5652–5661.
https://doi.org/10.1109/CVPR52733.2024.00540.
32. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]//*3rd International Conference on Learning Representations*, 2015.
33. Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C]//*2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016: 3213–3223.
https://doi.org/10.1109/CVPR.2016.350.
34. Zendel O, Schörghuber M, Rainer B, et al. Unifying panoptic segmentation for autonomous driving[C]//*Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022: 21319–21328.
https://doi.org/10.1109/CVPR52688.2022.02066.
35. Yu F, Chen H F, Wang X, et al. BDD100K: a diverse driving dataset for heterogeneous multitask learning[C]//*Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020: 2633–2642.
https://doi.org/10.1109/CVPR42600.2020.00271.
36. Huang S H, Lu Z C, Cheng R, et al. FaPN: feature-aligned pyramid network for dense image prediction[C]//*Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021: 844–853.
https://doi.org/10.1109/ICCV48922.2021.00090.
37. Sun H P, Zhang Y W, Xu L M, et al. Ultra-high resolution segmentation via boundary-enhanced patch-merging transformer[C]//*Proceedings of the 39th AAAI Conference on Artificial Intelligence*, 2025: 7087–7095.
https://doi.org/10.1609/aaai.v39i7.32761.
38. Qiu Y Y, Xu G A, Gao G W, et al. Efficient semantic segmentation via lightweight multiple-information interaction network[EB/OL]. arXiv: 2410.02224, 2024.
https://doi.org/10.48550/arXiv.2410.02224.
39. Guo M H, Lu C Z, Hou Q, et al. Segnext: Rethinking convolutional attention design for semantic segmentation[C]//*Advances in neural information processing systems*, 2022, 35: 1140-1156.
https://doi.org/10.52202/068431-0084
40. Cai H, Li J Y, Hu M Y, et al. EfficientViT: lightweight multi-scale attention for high-resolution dense prediction[C]//*2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023: 17256–17267.
https://doi.org/10.1109/ICCV51070.2023.01587.
41. Xu Z Z, Wu D Y, Yu C Q, et al. SCTNet: single-branch CNN with transformer semantic information for real-time segmentation[C]//*Proceedings of the 38th AAAI Conference on Artificial Intelligence*, 2024: 6378–6386.
https://doi.org/10.1609/aaai.v38i6.28457.
42. Fu Y X, Lou M, Yu Y Z. SegMAN: omni-scale context modeling with state space models and local attention for semantic segmentation[C]//*2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025: 19077–19087.
https://doi.org/10.1109/CVPR52734.2025.01777.
43. Wang Z, Qu S J. Research progress and challenges in real-time semantic segmentation for deep learning[J]. *J Image Graph*, 2024, **29** (5): 1188−1220.
王卓, 瞿绍军. 深度学习实时语义分割研究进展和挑战[J]. 中国图象图形学报, 2024, **29** (5): 1188−1220.
44. Yang G Y, Wang Y, Shi D M, et al. Golden cudgel network for real-time semantic segmentation[C]//*2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025: 25367–25376.
https://doi.org/10.1109/CVPR52734.2025.02362.
45. Cavagnero N, Rosi G, Cuttano C, et al. PEM: prototype-based efficient MaskFormer for image segmentation[C]//*2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024: 15804–15813.
https://doi.org/10.1109/CVPR52733.2024.01496.

作者简介



叶涛 (2000-)，男，福建漳州人，硕士研究生，2023年于中南民族大学获得学士学位，主要从事深度学习和图像处理以及图像分割的研究。
E-mail：ytis8761@163.com



【通信作者】林志贤 (1975-)，男，福建泉州人，博士，教授，博士研究生导师，2010 年于福州大学获得博士学位，主要从事电子纸显示和平板显示器件驱动以及图像处理技术的研究。
E-mail：lzx2005000@163.com

Semantic segmentation of street scene using a dual-branch boundary refinement network

Ye Tao[1,2], Lin Shanling[1,2], Lin Jianpu[1,2], Lv Shanhong[1,2], Lin Zhixian[1,2,3*], Guo Tailiang[2,3]

[1] School of Advanced Manufacturing, Fuzhou University, Quanzhou, Fujian 362251, China;
[2] Fujian Science and Technology Innovation Laboratory for Photoelectric Information, Fuzhou,Fujian 350116, China;
[3] College of Physics and Information Engineering, Fuzhou University, Fuzhou, Fujian 350116, China

Abstract

Objective Semantic segmentation of urban street scenes plays a central role in perception systems for autonomous driving, where reliable pixel-level understanding of roads, vehicles, infrastructure, and vulnerable road users is essential for safe planning and control. Existing real-time networks still struggle to delineate object boundaries under occlusions, scale variation, cluttered backgrounds, and complex lighting, which leads to fragmented objects, blurred contours, and missed small instances such as poles, traffic signs, riders, and pedestrians. Many architectures emphasize global context or throughput, yet fail to model fine edges explicitly and therefore sacrifice boundary quality. This study aimed to enhance boundary precision and small-object segmentation while preserving high inference speed, through the design of a dual-branch network that exploits explicit boundary refinement, adaptive receptive fields, and cross-branch information propagation for street-scene understanding.

Methods A dual-branch boundary refinement network (BRNet) was developed for street-scene semantic segmentation. The architecture contained a high-resolution branch with three adaptive sampling convolution (ASC) blocks that preserved fine spatial structure and adjusted receptive fields dynamically to complex object geometry. A low-resolution branch integrated a Boundary Refinement Module (BRM) that expanded the effective kernel to a 5×5 region with double convolutions and employed lightweight channel and spatial attention to strengthen edge features and suppress background noise near object contours. A feature propagation module (FPM) linked the two branches through progressive cross-scale fusion, compressing channels and performing downsampling or upsampling so that enhanced semantic and boundary cues from the low-resolution path flowed back to the high-resolution path and complemented local details. This design formed a collaborative boundary optimization mechanism, where both branches contributed to a consistent refinement of contours at different resolutions. Supervision signals were attached to multiple stages, and a hybrid loss combined cross-entropy with Lovasz-Softmax loss to optimize intersection-over-union directly and alleviate the class-imbalance characteristic of street-scene datasets, in which background and road pixels dominate rare categories. The network was trained and evaluated on the Cityscapes benchmark with fine annotations. The training pipeline adopted stochastic gradient descent with a poly learning-rate schedule, and used common data augmentation operations such as random scaling, cropping, color jittering, and horizontal flipping to increase robustness to viewpoint and appearance variations. Implementation relied on a standard deep learning framework with synchronized batch normalization and mixed-precision computation to stabilize optimization and accelerate training. Performance was assessed by mean intersection over union (mIoU), mean pixel accuracy (mPA), and frames per second (FPS) to capture both segmentation quality and computational efficiency. Additional experiments on the WildDash2 and BDD100K datasets examined robustness under distribution shifts that involved different cities, illumination conditions, camera setups, and adverse weather.

Results and Discussions BRNet achieved an mIoU of 81.43% on the Cityscapes test set and delivered a frame rate of 67 f/s on a single GPU, which demonstrated a favorable balance between accuracy and real-time performance. The method outperformed the bilateral segmentation network BiSeNetV2 by 5.69 percentage points in mIoU and exceeded other multi-branch or lightweight baselines such as DDRNet, LMIINet, and SegNeXt on most evaluation metrics, while remaining close to transformer-based EfficientViT in accuracy with noticeably lower computational cost. Quantitative analysis of small-object categories showed consistent gains over BiSeNetV2, DDRNet, and PIDNet for poles, traffic lights, traffic signs, pedestrians, motorcycles, and bicycles, confirming that explicit boundary modeling and adaptive sampling improved recognition of thin structures and distant targets. A separate comparison with models that integrated Mamba or transformer blocks indicated that BRNet obtained comparable segmentation quality with fewer parameters and reduced latency, because boundary attention was introduced through compact convolutional modules instead of heavy global self-attention. Cross-dataset tests further validated the generalization ability of BRNet. On the BDD100K dataset the network reached an mIoU of 43.18% and maintained a smaller performance drop than competing models when transferred from Cityscapes to unfamiliar urban environments that contained night scenes, rain, and complex traffic patterns. On the WildDash2 benchmark BRNet preserved coherent road layouts and object masks in

Foundation item: National Key Research and Development Program of China (2021YFB3600603)
*Correspondence: Lin Z X, E-mail: lzx2005000@163.com

images with motion blur, low visibility, and camera distortions. Qualitative visualizations on WildDash2 and BDD100K revealed clearer object contours, better separation of overlapping vehicles and pedestrians, and more complete road and sidewalk regions, especially in regions with shadows, glare, or motion blur. Class activation maps indicated that the boundary refinement mechanism guided attention to compact regions along object outlines, which helped preserve structural integrity, reduced boundary ambiguity, and mitigated confusion between foreground and background.

Conclusions The dual-branch BRNet with collaborative boundary refinement improved semantic segmentation of urban street scenes, especially in boundary precision and small-object delineation, while keeping real-time inference capability suitable for on-board deployment. The high-resolution branch with adaptive sampling captured detailed spatial layouts and local textures, and the low-resolution branch with the boundary refinement module enhanced semantic discrimination along contours; the feature propagation module linked these components into a continuous optimization loop for boundary representation across scales. The hybrid loss function reduced the impact of class imbalance and supported stable optimization toward high IoU scores without sacrificing pixel-wise accuracy. Experiments on Cityscapes, WildDash2, and BDD100K confirmed strong generalization across datasets and robustness to challenging imaging conditions such as adverse weather and poor illumination. These characteristics make BRNet a competitive candidate for perception stacks of autonomous driving and advanced driver-assistance systems that require accurate, efficient, and reliable understanding of complex urban environments. Future work will refine the architecture and compression strategy to further reduce latency and memory footprint, and will explore integration with panoptic segmentation, instance-level boundary detection, temporal consistency modeling, and multi-task learning frameworks to obtain richer and more coherent scene understanding from a single unified model.

Keywords: semantic segmentation; boundary refinement; adaptive sampling convolution; feature propagation; dual-branch architecture; urban street view

Ye T, Lin S L, Lin J P, et al. Semantic segmentation of street scene using a dual-branch boundary refinement network[J]. *Opto-Electron Eng*, 2026, **53**(5): 250277; DOI: 10.12086/oee.2026.250277

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文