

北京航空航天大学学报
Journal of Beijing University of Aeronautics and Astronautics
ISSN 1001-5965,CN 11-2625/V

《北京航空航天大学学报》网络首发论文

题目：融合边界感知与权重稀疏优化的语义分割网络
作者：陶小苗，王永顺，邬开俊，白晓凤
DOI：10.13700/j.bh.1001-5965.2026.0404
收稿日期：2026-06-22
网络首发日期：2026-11-09
引用格式：陶小苗，王永顺，邬开俊，白晓凤．融合边界感知与权重稀疏优化的语义分割网络[J/OL]．北京航空航天大学学报.
https://doi.org/10.13700/j.bh.1001-5965.2026.0404



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

融合边界感知与权重稀疏优化的语义分割网络

陶小苗✉，王永顺，邬开俊，白晓凤

（兰州交通大学 电子与信息工程学院，兰州 730070）
*通信作者 E-mail： taoxm@lzjtu.edu.cn

摘要 针对语义分割中跨分辨率特征融合存在信息冗余、边界细节易丢失及模型参数冗余等问题，提出一种融合边界感知与权重稀疏优化的语义分割网络（Boundary-Aware Sparse Segmentation Network，BASS-Net）。首先，在跨分辨率特征融合中引入低开销通道重标定机制，自适应调整低分辨率语义特征的贡献，增强有效语义信息并抑制冗余响应；其次，设计边界感知特征精炼模块，结合全局语义约束与局部卷积特征，改善复杂边界与细粒度结构的预测质量；此外，采用基于 L1 范数的渐进式非结构化剪枝，对低重要性权重进行稀疏化，减少模型冗余连接。实验结果表明，BASS-Net 在 Cityscapes、COCO-Stuff 10K 和 CamVid 数据集上的平均交并比（mean Intersection over Union，mIoU）分别达到 81.30%、35.97%和 76.53%。结果验证了所提方法在提高分割精度、保持边界细节及控制模型冗余方面的有效性，并为后续模型压缩与部署优化提供基础。

关键词 语义分割；跨分辨率特征融合；通道重标定；边界感知；权重稀疏化
中图分类号 TP391.4 文献标志码 A DOI: 10.13700/j.bh.1001-5965.2026.0404

Semantic segmentation network integrating boundary awareness and weight sparsity optimization

Tao Xiaomiao✉, Wang Yongshun, Wu Kaijun, Bai Xiaofeng

（School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China）
*E-mail: taoxm@lzjtu.edu.cn

Abstract To address the problems of information redundancy in cross-resolution feature fusion, loss of boundary details, and redundant model parameters in semantic segmentation, a Boundary-Aware Sparse Segmentation Network (BASS-Net) is proposed. First, a low-overhead channel recalibration mechanism is introduced into cross-resolution feature fusion to adaptively adjust the contributions of low-resolution semantic features, enhance effective semantic information, and suppress redundant responses. Second, a boundary-aware feature refinement module is designed to integrate global semantic constraints with local convolutional features, thereby improving prediction quality in complex boundary regions and fine-grained structures. Furthermore, progressive unstructured pruning based on the L1 norm is employed to sparsify low-importance weights and reduce redundant connections in the model. Experimental results show that BASS-Net achieves mean Intersection over Union (mIoU) scores of 81.30%, 35.97%, and 76.53% on the Cityscapes, COCO-Stuff 10K, and CamVid datasets, respectively. The results demonstrate the effectiveness of the proposed method in improving segmentation accuracy, preserving boundary details, and controlling model redundancy, while providing a foundation for subsequent model compression and deployment optimization.

Key words semantic segmentation; cross-resolution feature fusion; channel recalibration; boundary awareness; weight sparsification

语义分割是像素级场景理解的重要技术，广泛应用于自动驾驶、移动机器人和智能感知等领域[1]。与图像分类和目标检测相比，语义分割不仅需要准确识别像素类别，还需保留目标轮廓及局部

收稿日期：2026-06-22
基金项目：甘肃省重点研发项目（25YFGA047）；甘肃省高校产业支撑计划项目（2026CYZC-032）
Fund：The Key Research and Development Program Project of Gansu Province(25YFGA047)，Gansu University Industry Support Program(2026CYZC-032)
网络首发时间：2026-11-09 09:15:40 网络首发地址：https://link.cnki.net/urlid/11.2625.V.20261106.1734.003

细节[2]。然而，提高特征分辨率和增强上下文建模会增加计算开销，过度压缩网络则易造成细节丢失。因此，兼顾分割精度、边界质量与模型复杂度仍是重要研究问题[3]。

随着卷积神经网络（Convolutional Neural Network，CNN）的发展，多种高效语义分割方法相继提出。图像级联网络（Image Cascade Network，ICNet）[4]通过多分辨率级联降低计算开销；双边分割网络（Bilateral Segmentation Network，BiSeNet）[5]利用双路径结构分别提取空间细节与语义信息；高效空间金字塔网络（Efficient Spatial Pyramid Network，ESPNet）[6]通过高效卷积扩大感受野；快速语义分割卷积神经网络（Fast Semantic Segmentation Convolutional Neural Network，Fast-SCNN）[7]采用学习降采样与多分支结构提高推理效率。这些方法在一定程度上改善了分割精度与计算效率的平衡。

然而，现有双分支及多尺度分割网络仍存在跨分辨率特征融合冗余、边界细节易丢失等问题，直接融合易引入无效语义响应，连续下采样则会削弱细粒度结构表达。此外，过度压缩网络可能削弱特征表达能力，影响分割性能[8]。因此，仍需进一步优化特征融合、边界恢复与参数冗余控制。

针对上述问题，在 PCLD-Net 基础上构建统一基线模型 PCLD-Base，提出融合边界感知与权重稀疏优化的语义分割网络（BASS-Net）。主要贡献如下：

（1）提出低开销通道重标定机制，自适应筛选低分辨率语义特征，抑制冗余响应，提高跨分辨率特征融合质量。

（2）设计边界感知特征精炼模块，结合全局语义约束与局部卷积特征，改善复杂边界及细粒度结构的预测质量。

（3）采用基于 L1 范数的渐进式非结构化剪枝，对低重要性权重进行稀疏化，在保持分割精度的同时减少冗余连接，提高模型压缩潜力。

1 相关工作

1.1 语义分割网络架构与特征增强方法

双分支语义分割通过空间与上下文分支分别提取细节和高级语义信息。ICNet 通过多分辨率级联降低计算开销[4]；BiSeNet 采用双路径结构实现信息互补[5]；ESPNet 利用分解卷积与空洞卷积扩大感受野[6]；Fast-SCNN 通过学习降采样和多分支结构提高推理效率[7]。相关研究通过细节增强、注意力与边界感知优化分割性能[9-10]；ENet 采用轻量化编码器-解码器结构，但边界表达有限[11]；STDCNet 通过短程密集连接增强特征复用[12]；DDRNet 利用双分辨率并行结构实现特征交互[13]；PIDNet 对语义、细节和边界信息进行差异化建模[14]。然而，直接融合跨分辨率特征易引入冗余响应，过度压缩网络则可能削弱细节表达。

多尺度融合与注意力机制可增强语义及边界特征。DeepLab 系列通过空洞卷积与编码器—解码器结构增强多尺度特征表达[15-17]；相关研究结合多尺度特征与全局信息改善分割性能[18]，并通过多路径特征融合增强语义表达[19]；PointRend 利用不确定区域采样精细预测边界[20]；SENet 通过通道重标定增强重要特征[21]；DANet 联合建模空间与通道依赖[22]；Transformer 相关方法通过层次特征建模增强语义表达[23]；多层级特征与注意力融合进一步改善分割性能[24]。部分方法计算开销较大，边界细节恢复仍受限制。因此，本文结合低开销通道重标定与边界感知特征精炼，改善跨分辨率特征融合与边界预测。

1.2 模型剪枝与参数稀疏化方法

模型剪枝通过减少冗余参数降低复杂度。网络瘦身通过识别低贡献结构减少冗余计算[25]；滤波器剪枝通过移除低重要性卷积核压缩网络[26]；相关研究通过编码器-解码器压缩及任务相关剪枝平衡模型规模与分割精度[27-28]。然而，过度剪枝易造成边界及小目标信息丢失。渐进式剪枝通过逐步筛除低重要性权重减小性能波动。因此，本文采用训练后基于 L1 范数的渐进式非结构化剪枝减少冗余连接，提高模型压缩潜力，实际加速仍依赖稀疏算子及硬件支持。

2　BASS-Net 模型

BASS-Net 整体架构如图 1 所示，以 PCLD-Base 为基础，由双分辨率主干、多尺度上下文特征匹配模块（Multi-scale Context Feature Matching，MsCFM）、深度聚合金字塔池化模块（Deep Aggregation Pyramid Pooling Module，DAPPM）、Transformer 上下文增强编码单元、通道重标定模块、边界感知特征精炼模块及 DDRHead 分割头组成。文中“1/s 尺度”表示相对于输入图像的空间分辨率比例。若输入尺寸为 H×W，则 1/4、1/8、1/16、1/32 和 1/64 尺度分别对应 H/4×W/4、H/8×W/8、H/16×W/16、H/32×W/32 和 H/64×W/64，而非相邻层的累计下采样倍数。

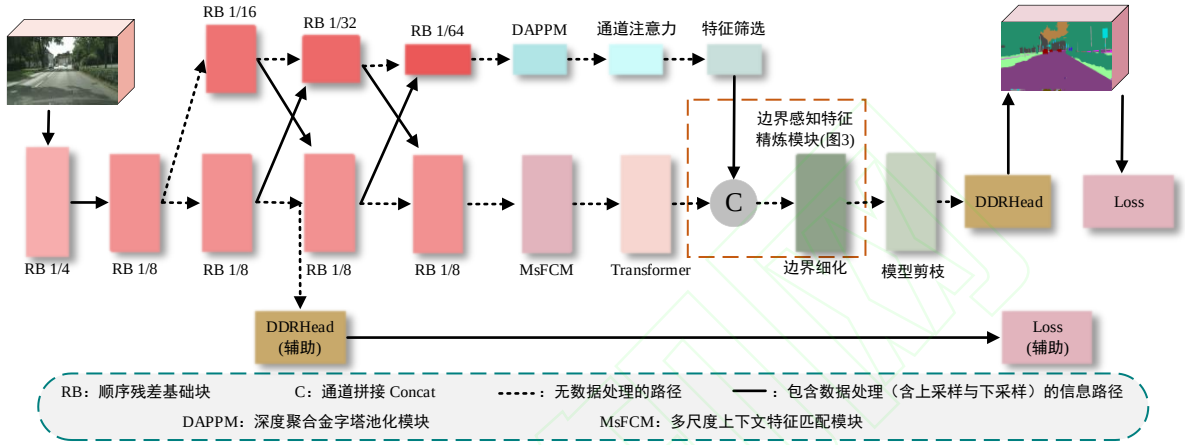


图 1 BASS-Net 模型图
Fig.1 Overall architecture of BASS-Net

高分辨率空间分支经初始下采样获得 1/4 尺度特征，再经 2 倍下采样得到 1/8 尺度特征，通过残差块（Residual Block，RB）保留空间细节，并保留 PCLD-Base 中的 MsCFM 及 Transformer 上下文增强路径。低分辨率上下文分支由 1/8 尺度特征逐级下采样至 1/16、1/32 和 1/64 尺度，并通过 DAPPM 聚合多尺度语义信息。

跨分辨率融合前，利用通道重标定模块筛选低分辨率语义特征，抑制冗余响应；融合后，通过边界感知特征精炼模块结合全局语义与局部卷积信息，改善边界预测。DDRHead 分别通过辅助预测支路和最终预测支路计算辅助损失与主损失。主体网络训练完成后，采用基于 L1 范数的渐进式非结构化剪枝减少冗余权重，提高模型压缩潜力。

2.1 通道重标定模块

在双分支语义分割网络中，低分辨率上下文分支虽然能够提供高级语义信息，但也包含与边界恢复和细节表达关联较弱的冗余通道。直接进行跨分辨率融合易对高分辨率空间特征造成干扰，削弱边界及细粒度结构的表达能力。为此，在低分辨率语义特征进入融合路径前引入低开销通道重标定模块，对语义通道进行自适应筛选，其结构如图 2 所示。

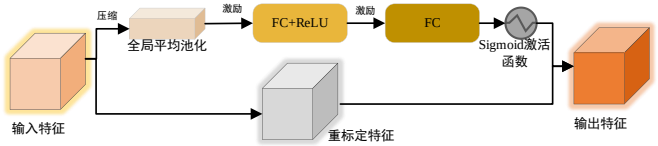


图 2　低开销通道重标定模块结构
Fig.2 Structure of the low-overhead channel recalibration module

该模块采用“全局平均池化+两层全连接层+Sigmoid 激活”结构，与压缩与激励网络（Squeeze-and-Excitation Network，SENet）的通道重标定思想相近，但作用位置和对象不同。SENet 通常用于主干网络内部的通用特征增强，而该模块仅对跨分辨率融合前的低分辨率语义特征进行筛选，以减少冗余

响应对空间细节的干扰。相比卷积块注意力模块（Convolutional Block Attention Module，CBAM）等同时建模空间注意力的方法，该模块不引入额外的空间注意力计算，以控制特征筛选开销。

设输入特征为 $X_c \in \mathbb{R}^{C\times H\times W}$，其中 C、H 和 W 分别表示特征通道数、高度和宽度。通道重标定包括压缩、激励和重标定三个阶段。

在压缩阶段，通过全局平均池化聚合输入特征 X 的空间信息，获得各通道的全局统计描述，如式(1)所示。

$$z_c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_c(i,j) \tag{1}$$

在激励阶段，利用两层全连接层（Fully Connected Layer，FC）建模通道间的依赖关系。第一层将通道维度压缩至原来的 1/r，并通过修正线性单元（Rectified Linear Unit，ReLU）引入非线性表示；第二层恢复至原始通道数，经 Sigmoid 函数生成归一化通道权重，如式(2)和式(3)所示。

$$u = \mathrm{ReLU}(\omega_1 z + b_1) \tag{2}$$

$$s = \mathrm{Sigmoid}(\omega_2 u + b_2) \tag{3}$$

式中，式(2)的输入为式(1)得到的通道统计向量，输出为第一层 FC 后的中间表示；式(3)的输入为该中间表示，输出为经 Sigmoid 归一化后的通道权重向量。ω_1 和 ω_2 为可学习权重矩阵，b_1 和 b_2 为对应偏置项。

在重标定阶段，将归一化通道权重与原始输入特征逐通道相乘，得到重标定后的特征，如式(4)所示。

$$X_c = s_c \cdot X_c \tag{4}$$

该模块不直接生成边缘细节，而是通过抑制冗余语义干扰，提高高分辨率空间信息在融合过程中的保留稳定性。同时，模块仅采用全局平均池化与两层全连接变换，额外计算开销较小，可为后续边界感知特征精炼提供更有效的语义输入。

2.2 边界感知特征精炼模块

语义分割网络在目标边界区域易出现类别混淆、边缘断裂及细小结构丢失等问题。局部卷积特征虽能保留纹理与空间细节，但缺乏全局语义约束；全局上下文信息有助于保持区域语义一致性，却难以精确定位边界。为此，设计边界感知特征精炼模块，通过全局语义与局部特征的协同建模，改善复杂边界区域的预测质量。模块结构如图 3 所示，其在整体网络中的对应位置已在图 1 中以虚线框标出。

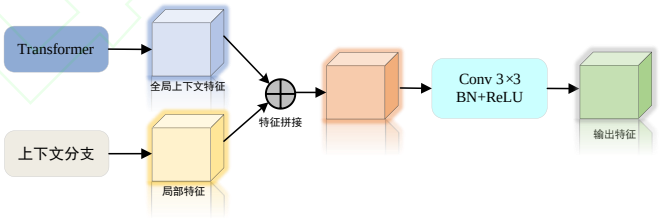


图 3 边界感知特征精炼模块结构
Fig.3 Structure of the boundary-aware feature refinement module

该模块利用既有 Transformer 上下文增强路径提取的全局语义特征提供区域一致性约束，并结合局部卷积特征保留边界邻域的纹理与结构信息，无需额外构建独立的 Transformer 边界建模分支。局部精炼单元由 3×3 卷积、批归一化（Batch Normalization，BN）和修正线性单元（Rectified Linear Unit，ReLU）激活函数组成，通过可学习的局部特征变换细化边界表示。

设上下文增强路径输出的全局语义特征 $F_{\text{transformer}}$ 和上下文分支提取的局部语义特征 F_{context} 已完成空间尺寸对齐。首先沿通道维度拼接两类特征，获得融合特征，如式(5)所示。

$$F_{\text{fuse}} = \text{Concat}(F_{\text{transformer}}, F_{\text{context}}) \tag{5}$$

式中，$F_{\text{transformer}}$ 用于提供长距离依赖与语义一致性约束，F_{context} 用于保留边界邻域的局部响应。

随后，通过局部卷积精炼单元对融合特征进行处理，在保持预设输出通道数的条件下获得精炼特征，其过程如式(6)所示。

$$F_{\text{fused}} = \text{ConvModule}(F_{\text{fuse}}) \tag{6}$$

其中，$\text{ConvModule}(\cdot)$ 由 3×3 卷积、BN 和 ReLU 激活函数组成，在保持预设输出通道数的条件下，对融合后的边界相关特征进行局部细化，其具体形式如式(7)所示。

$$F_{\text{refined}} = \text{ReLU}(\text{BN}(\omega_{\text{conv}} * F_{\text{fused}} + b_{\text{conv}})) \tag{7}$$

式中，ω_{conv} 和 b_{conv} 分别表示 3×3 卷积层的权重和偏置参数。

该模块通过全局语义约束缓解边界两侧的类别混淆，并利用局部卷积细化边缘位置，提高边界预测的连续性与细粒度结构的表达能力。与额外引入独立边界预测分支或高分辨率注意力计算的方法相比，该模块仅在融合后增加局部卷积精炼单元，能够控制额外计算开销，并参与网络端到端训练。

2.3 渐进式剪枝与稀疏化策略

为减少模型参数冗余及部署开销，在主体网络训练完成后引入基于 L1 范数的渐进式非结构化剪枝策略，对卷积层和全连接层中的低重要性权重进行稀疏化。该策略作为训练后的参数优化步骤，不改变主体网络结构，主要用于减少冗余连接并提高模型压缩潜力，其流程如图 4 所示。

设待剪枝层的原始权重为 ω_{nct}，引入与其形状一致的二值掩码矩阵 M，掩码值为 0 和 1 分别对应剪除与保留的权重连接，二者通过逐元素相乘得到稀疏权重。剪枝后的稀疏权重由式(8)定义。通过筛除低重要性权重，在适度剪枝条件下减少参数冗余并控制分割精度损失。

$$\omega_{\text{nct}} = M \odot \omega_{\text{nct}} \tag{8}$$

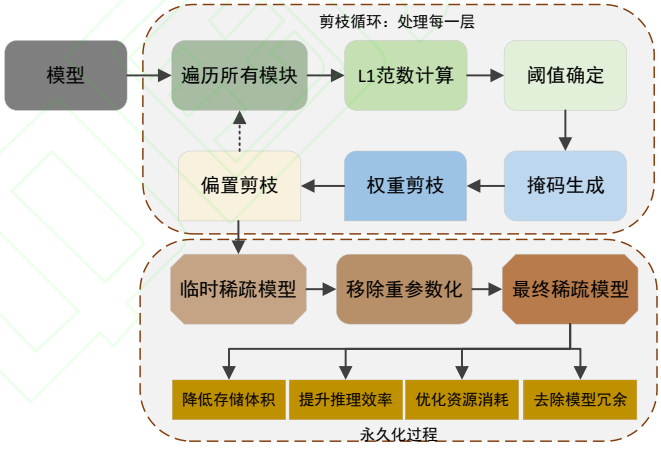


图 4　模型剪枝与稀疏化流程

Fig.4 Pipeline of model pruning and sparsification

剪枝过程包括逐层稀疏化和参数固化两个阶段。在逐层稀疏化阶段，依次遍历卷积层和全连接层，根据各层权重绝对值评估连接重要性并确定层内剪枝阈值，生成对应的二值掩码，逐步筛除低重要性权重。各层分别计算剪枝阈值，以避免统一全局阈值造成层间剪枝不平衡。

在参数固化阶段，移除剪枝过程中的临时重参数化表示，将稀疏权重写回模型参数，得到可用于保存与部署的稀疏模型。

该策略通过逐步减少低重要性权重连接，在控制分割精度损失的同时提高权重稀疏度。由于非结构化剪枝不改变网络的卷积层、通道数及标准稠密计算图，其直接收益主要体现为冗余连接减少与模型压缩潜力提升，实际推理加速仍取决于稀疏算子、内存访问机制及硬件支持。

3 实验

3.1 数据集

为验证所提方法在不同场景及类别规模下的分割性能，选用 Cityscapes、COCO-Stuff 10K 和 CamVid 三个数据集进行实验，分别评价模型在道路场景、复杂多类别场景及有限样本道路场景下的分割能力。

Cityscapes 包含 5000 张分辨率为 2048×1024 的精细标注图像，涉及 30 个标注类别，其中 19 类用于评价。训练集、验证集和测试集分别包含 2975、500 和 1525 张图像，实验采用训练集训练，验证集进行性能评价、对比及消融实验。

COCO-Stuff 10K 包含 10000 张标注图像，其中 9000 张用于训练、1000 张用于评价，涵盖 80 个物体类别、91 个背景类别及 1 个未标注类别，实验采用 171 个有效类别进行评价。

CamVid 包含 701 张分辨率为 960×720 的标注图像，采用 11 类语义分割设置，训练集、验证集和测试集分别包含 367、101 和 233 张图像，分别用于模型训练、模型选择及最终性能评价。

3.2 实验设置

实验基于 PyTorch 框架，在 NVIDIA RTX 4090 GPU 平台上进行。PCLD-Base、BASS-Net 及各消融模型采用统一的数据划分、硬件环境和训练策略，主要实验参数见表 1。

表 1 实验参数配置

Table 1 Configuration of Experimental Parameters

参数名称	参数值
解码头	DDRHead
解码头损失函数	[OhemCrossEntropy]
解码头损失函数权重	[1.0, 0.4]
优化器类型	SGD
学习率	0.01
动量	0.9
权重衰减	0.0005
学习率调度器	PolyLR
学习率调度器幂	0.9
最大迭代次数	120000
每轮验证间隔	12000
每批次大小	4

Cityscapes 和 COCO-Stuff 10K 训练图像经随机尺度变换后，分别裁剪为 1024×1024 和 512×512，并采用随机水平翻转及光度扰动增强。验证阶段采用完整图像，其中 COCO-Stuff 10K 保持原始长宽比进行尺度调整。CamVid 采用 960×720 输入尺寸，验证和测试阶段均使用完整图像。

部分对比方法的实验结果来源于公开文献，因硬件平台、输入尺度及测试条件存在差异，仅用于总体性能参考；基线模型、BASS-Net 及各消融模型均在统一条件下进行比较。

3.3 评价指标

采用平均交并比（mean Intersection over Union，mIoU）、平均准确率（mean Accuracy，mAcc）、帧率（Frames Per Second，FPS）、参数量（Params）和十亿次浮点运算量（Giga Floating Point Operations，GFLOPs）评价模型的分割精度、推理效率及复杂度。

mIoU 为各类别交并比（Intersection over Union，IoU）的平均值，计算如式(9)和式(10)所示。

$$IoU_{\xi}=\frac{TP_{\xi}}{TP_{\xi}+FP_{\xi}+FN_{\xi}} \tag{9}$$

$$mIoU=\frac{1}{K}\sum_{\xi=1}^{K}IoU_{\xi} \tag{10}$$

式中，TP_{ξ}、FP_{ξ}、FN_{ξ} 和 K 各符号分别表示第 ξ 类的真正例、假正例、假负例像素数及评价类别总数。mAcc 为各类别像素准确率的平均值，如式(11)所示。

$$mAcc = \frac{1}{K}\sum_{\xi=1}^{K}\frac{TP_{\xi}}{TP_{\xi}+FN_{\xi}} \tag{11}$$

在推理效率方面，采用 FPS 衡量模型单位时间内能够处理的图像数量，如式(12)所示。

$$FPS = \frac{N}{T} \tag{12}$$

式中，N 表示测试过程中完成前向推理的图像数量，T 表示处理这些图像所消耗的总推理时间。

在模型复杂度方面，采用 Params 和 GFLOPs 衡量模型可学习参数总量及单次前向推理的浮点运算量，表示如式(13)和式(14)所示。

$$Params = \sum_{l=1}^{L} P_l \tag{13}$$

$$GFLOPs = \frac{\sum_{l=1}^{L} O_l}{10^9} \tag{14}$$

式中，L 表示网络中包含可学习参数的层数，P_l 表示第 l 层的可学习参数数量。Params 越小，模型所需的参数存储空间通常越低。O_l 表示第 l 个计算层在一次前向传播过程中产生的浮点运算次数，L 表示参与计算复杂度统计的网络层数。

FPS 在相同硬件与测试条件下统计，GFLOPs 采用相同输入尺寸和统计方式进行比较；公开文献中的 FPS 仅作参考。

3.4 对比实验

为评价 BASS-Net 的分割精度、推理效率和计算复杂度，在 Cityscapes、COCO-Stuff 10K 和 CamVid 数据集上与代表性方法进行对比，结果见表 2-表 4。本文不以“参数量最小”或“最高 FPS”作为方法定位，而是重点考察在现有双分辨率上下文框架基础上提高边界质量和分割精度时的精度-效率权衡。

Cityscapes 数据集的对比结果见表 2。统一实验条件下，BASS-Net 在 1024×1024 输入时取得 81.30%的 mIoU 和 37.78 FPS，较 PCLD-Base 和 PCLD-Net 分别提高 1.16 和 0.38 个百分点。

统一输入尺度与硬件平台下，部分对比方法在模型规模或推理速度方面具有优势，如 PIDNet-S[52]达到 93.2 FPS。相比之下，BASS-Net 具有较高的分割精度，但参数量和计算开销较大，因此不将其定位为轻量化模型，而侧重于双分辨率上下文框架基础上的精度与边界建模收益。

在 768×768、832×832 和 1024×1024 输入下，BASS-Net 的 mIoU 分别为 80.55%、80.85%和 81.30%，FPS 分别为 67.10、57.20 和 37.78。随着输入尺度增大，分割精度提高，但计算量增加、推理速度下降，体现了精度与效率之间的权衡。后续实验采用 mIoU 最高的 1024×1024 输入尺度。

表 2 cityscapes对比实验结果

Table 2 Cityscapes Comparison Experiment Results

Model\Metrics	mIoU↑	GPU	FPS↑	Resolution	Params/M↓	GFLOPs↓
SegNet[29]	57.00	TitanX	16.70	640×360	-	-
ENet[11]	57.00	TitanX	135.40	640×360	-	-
SQ[30]	59.80	TitanX	16.70	2048×1024	-	-
ICNet[4]	69.50	TitanX M	30.00	2048×1024	-	-
ESPNet[6]	60.30	TitanX	113.00	1024×512	-	-
ERFNet[31]	68.00	TitanX M	41.70	1024×512	-	-
Fast-SCNN[7]	68.00	TitanXp	123.50	2048×1024	-	-
DFANet B[32]	67.10	TitanX	120.00	1024×1024	-	-
SwiftNetRN-18 ens[33]	76.50	GTX 1080Ti	18.40	2048×1024	-	-
BiSeNetV2-L[34]	75.30	GTX 1080Ti	47.30	1024×512	-	-
SFNet(DF2)[35]	77.80	GTX 1080Ti	53.00	2048×1024	-	-
SFNet(ResNet-18)[35]	78.90	GTX 1080Ti	18.00	2048×1024	-	-
MSFNet[36]	77.10	GTX 2080Ti	41.00	2048×1024	-	-
CABiNet[37]	75.90	GTX 2080Ti	76.50	2048×1024	-	-
DDRNet[13]	79.05	RTX 4090	37.33	1024×1024	77.44	378.1
PCLD-Net[38]	80.92	RTX 4090	14.50	1024×1024	90.65	461.1
PCLD-Base	80.14	RTX 4090	37.89	1024×1024	88.22	447.0
PIDNet-S[14]	78.60	RTX 3090	93.20	2048×1024	7.6	47.6
STDC2-Seg75[12]	76.80	GTX 1080Ti	97.00	1536×768	-	-
SeaFormer-B-Seg100[37]	77.70	-	-	2048×1024	8.6	13.7
PIDNet-S[14]	78.60	RTX 4090	93.2	1024×1024	7.6	11.9
STDC2-Seg75[12]	76.80	RTX 4090	97.0	1024×1024	22.2	-
SeaFormer-B[39]	77.70	RTX 4090	37.5	1024×1024	8.6	3.4
SegFormer-B0[40]	76.20	RTX 4090	15.2	1024×1024	3.8	125.5
DDRNet-23-Slim[13]	77.40	RTX 4090	102.0	1024×1024	5.7	36.3
BiSeNetV2-L[34]	75.80	RTX 4090	47.3	1024×1024	4.59	-
BASS-Net	80.55	RTX 4090	67.10	768×768	90.49	257.9
BASS-Net	80.85	RTX 4090	57.20	832×832	90.49	302.7
BASS-Net	81.30	RTX 4090	37.78	1024×1024	90.49	458.6

COCO-Stuff 10K 对比结果见表 3。BASS-Net 的 mIoU 为 35.97%，较 PCLD-Base 和 PCLD-Net 分别提高 1.91 和 1.70 个百分点。相比基线，FPS 由 71.50 降至 68.75，参数量由 88.22 M 增至 90.64 M，GFLOPs 由 162.3 增至 166.7，表明精度提升伴随一定计算代价。统一测试条件下，BASS-Net 的 mIoU 高于 SegFormer-B0 的 35.60%和 SeaFormer-B 的 34.10%。PIDNet-S、STDC2-Seg75 及 DDRNet-23-Slim 未报告 mIoU，不参与精度排序；上述部分模型及 BiSeNetV2-L 参数量更小，部分推理速度更高。因此，该结果主要体现复杂多类别场景下的总体分割性能提升，而非参数规模优势。表 3 仅报告总体 mIoU 等指标，无法确定类别级增益，故结论限于总体分割性能及跨场景泛化能力的改善。综上，BASS-Net 以较大的模型规模换取更高的总体 mIoU，结合 Params、GFLOPs 和 FPS，从精度与复杂度两方面评价新增通道筛选与边界精炼模块的有效性及其计算代价。

表 3 coco-stuff10k对比实验结果

Table 3 coco-stuff10k Comparison Experiment Results

Model\Metrics	mIoU↑	GPU	Resolution	FPS↑	Params/M↓	GFLOPs↓
AnnNet[41]	32.14	GTX 1080Ti	768×768	-	-	-
ApcNet[42]	32.78	GTX 1080Ti	768×768	-	-	-
CcNet[43]	33.20	GTX 1080Ti	768×768	-	-	-
DeepLabV3[17]	32.28	GTX 1080Ti	768×768	-	-	-
EmaNet[44]	32.90	RTX 2080Ti	512×512	-	-	-
Upernet[45]	33.15	RTX 2080Ti	512×512	-	-	-
EncNet[46]	32.23	RTX 2080Ti	512×512	-	-	-
GcNet[47]	32.27	RTX 3080	512×512	-	-	-
PspNet[48]	33.07	RTX 3090	512×512	-	-	-
HRNet[49]	32.75	RTX 3090	768×768	-	-	-
DDRNet[13]	32.10	RTX 4090	512×512	-	77.59	134.1
PCLD-Net[38]	34.27	RTX 4090	512×512	71.78	88.37	165.9
PCLD-Base	34.06	RTX 4090	512×512	71.5	88.22	162.3
PIDNet-S[14]	-	RTX 4090	512×512	-	7.6	-
STDC2-Seg75[12]	-	RTX 4090	512×512	-	22.2	-
SeaFormer-B[39]	34.10	RTX 4090	512×512	41.9	8.6	-
SegFormer-B0[40]	35.60	RTX 4090	512×512	84.4	3.8	8.4
DDRNet-23-Slim[13]	-	RTX 4090	512×512	-	5.7	-
BiSeNetV2-L[41]	28.70	RTX 4090	512×512	42.5	4.59	-
BASS-Net	35.97	RTX 4090	512×512	68.75	90.64	166.7

表 4 CamVid对比实验结果

Table 4　CamVid Comparison Experiment Results

Model\Metrics	mIoU↑	mAcc↑	FPS↑	Params/M↓	GFLOPs↓
ENet[11]	51.34	63.87	92.43	-	-
ERFNet[31]	68.12	79.26	41.73	-	-
Fast-SCNN[7]	68.00	-	-	-	-
BiSeNet[5]	68.40	-	-	-	-
CGNet[50]	65.60	-	96.02	-	-
DABNet[51]	69.10	-	-	-	-
BiSeNetV2[34]	72.40	-	-	-	-
STDC2-Seg[12]	73.90	-	-	-	-
BiSeNetV2-L[34]	73.58	84.17	47.46	-	-
DDRNet[13]	74.83	85.64	38.57	-	-
PCLD-Base	75.41	86.18	38.24	88.22	297.5
PIDNet-S[14]	80.10	-	153.70	7.6	-
PIDNet-M[14]	82.00	-	85.60	34.4	-
STDC2-Seg[12]	73.90	-	152.2	22.2	-
SeaFormer-B[39]	-	-	-	8.6	-
SegFormer-B0[40]	-	-	-	3.8	-
DDRNet-23-Slim[13]	74.70	-	230.0	5.7	-
BiSeNetV2-L[34]	73.20	-	32.70	4.59	-
BASS-Net	76.53	87.06	37.91	90.49	302.0

为验证 BASS-Net 在不同道路场景中的泛化能力，在 CamVid 数据集上采用相同训练策略及优化参数进行补充实验，输入分辨率为 960×720。由表 4 可知，BASS-Net 的 mIoU 和 mAcc 分别为 76.53%和 87.06%，较 PCLD-Base 提高 1.12 和 0.88 个百分点，FPS 由 38.24 降至 37.91。相比 PIDNet-S 和 PIDNet-M，BASS-Net 在 mIoU 及模型规模方面不占优势，但相较基线改善了分割精度与边界表达。在统一输入尺度与硬件平台下的补充对比中，DDRNet-23-Slim 和 STDC2-Seg 分别达到 230.0 和 152.2 FPS，表明 BASS-Net 不以绝对速度或模型紧凑性作为主要优势。

3.5 消融实验

为全面评估各创新模块的贡献，本小节在 Cityscapes 和 COCO-Stuff 10K 数据集上开展消融研究。所有结构消融均以 PCLD-Base 作为共同起点，并在统一的 NVIDIA RTX 4090 平台和相同训练配置下进行。考虑到权重稀疏化属于主体网络训练完成后的部署优化步骤，本文将其从结构模块消融表中独立出来，并在表 7 中单独分析不同权重稀疏率的影响。

表 5 cityscapes消融实验结果

Table 5　Cityscapes Ablation Experiment Results

实验	PCLD-Base	通道重标定	边界精炼	mIoU↑	mAcc(OA)↑	FPS↑	Param/M↓	GFLOPs↓
1	√	×	×	80.14	87.33	37.89	88.22	447.0
2	√	√	×	80.57	87.82	37.85	88.23	447.1
3	√	×	√	80.79	87.78	37.80	90.48	458.5
4	√	√	√	81.30	88.22	37.78	90.49	458.6

Cityscapes 消融结果见表 5。通道重标定和边界精炼模块分别使 mIoU 提高 0.43 和 0.65 个百分点，联合后达到 81.30%，较 PCLD-Base 提高 1.16 个百分点。其中，通道重标定的额外计算开销较小，边界精炼承担主要新增开销，两者在特征筛选与边界建模方面具有互补作用。

权重稀疏化属于训练后非结构化剪枝，不改变网络结构及标准稠密计算图，不纳入结构消融，其稀疏率、非零权重数及性能变化在表 7 中单独分析，实际加速仍依赖稀疏算子和硬件支持。结构模块是提升分割精度，权重稀疏化则用于减少冗余连接并提高模型压缩潜力。

表 6 coco-stuff10k消融实验结果

Table 6 COCO-Stuff 10K Ablation Experiment Results

实验	PCLD-Base	通道重标定	边界精炼	mIoU↑	mAcc(OA)↑	FPS↑	Param/M↓	GFLOPs↓
1	√	×	×	34.06	39.88	71.5	88.22	162.3
2	√	√	×	34.49	40.30	71.4	88.23	162.4
3	√	×	√	34.69	40.50	71.2	90.48	166.6
4	√	√	√	35.97	42.21	68.75	90.64	166.7

COCO-Stuff 10K 消融结果见表 6。通道重标定和边界精炼模块分别使 mIoU 提高 0.43 和 0.63 个百分点，联合后达到 35.97%，较 PCLD-Base 提高 1.91 个百分点，表明两模块在复杂多类别场景下具有协同作用。其中，通道重标定的额外计算开销较小，边界精炼带来一定的参数量与计算量增加，完整网络的 FPS 由 71.5 降至 68.75，体现了分割精度与计算复杂度之间的权衡。权重稀疏化作为训练后

的独立优化步骤，在表 7 中单独分析。

表 7 不同权重稀疏率下模型性能变化

Table 7 Model Performance under Different Weight Sparsity Ratios

权重稀疏率	mIoU↑	FPS↑	非零权重数/M↓
0%	81.30	37.78	90.49
10%	81.17	37.90	84.67
20%	80.94	38.05	79.28
30%	80.26	38.20	72.93
40%	79.53	38.35	65.24

为分析权重稀疏化对模型性能的影响，在 Cityscapes 数据集上设置 10%、20%、30%和 40%的权重稀疏率，结果见表 7。随着稀疏率提高，非零权重数减少，mIoU 逐渐下降。其中，20%稀疏率下 mIoU 由 81.30%降至 80.94%，非零权重数由 90.49 M 降至 79.28 M；稀疏率超过 30%时，精度损失明显增大。FPS 仅小幅波动，不能直接归因于稀疏计算加速，实际加速仍依赖稀疏算子与硬件支持。由于稀疏率按参与剪枝的目标权重统计，而非零权重数按全模型统计，二者不呈同比例变化。综合分割精度与稀疏程度，选取 20%作为后续部署分析的参考稀疏率。

表 8 模型性能稳定性分析

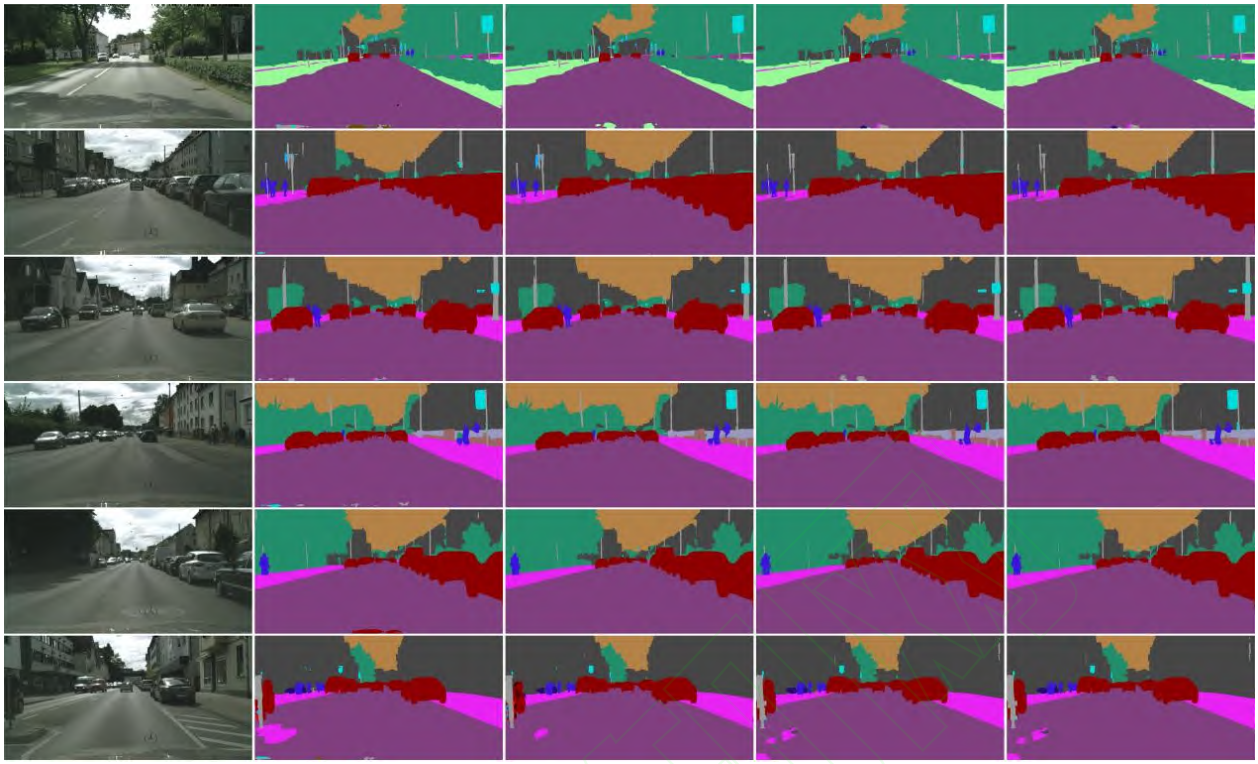
Table 8 Model Performance Stability Analysis

Model\Metrics	平均 mIoU↑	标准差↓	RSD/%↓
PCLD-Base	80.11	0.24	0.30
BASS-Net（稀疏化前）	81.26	0.19	0.23
BASS-Net（20%权重稀疏化）	80.83	0.31	0.38

为验证权重稀疏化前后的模型稳定性，采用 5 组不同随机初始化进行重复实验，统计 mIoU 均值、标准差及相对标准差（relative standard deviation，RSD），RSD 越小表示不同随机初始化下的性能波动越小。结果见表 8。PCLD-Base、BASS-Net 稀疏化前及 20%稀疏化后的平均 mIoU 分别为 80.11%、81.26%和 80.83%，RSD 分别为 0.30%、0.23%和 0.38%。结果表明，通道重标定与边界精炼未增加训练结果的随机波动；稀疏化后波动略有增加，但整体仍处于较低水平，表明模型具有较好的重复性。

3.6 可视化分析

为分析不同环境下的边界分割效果，从 Cityscapes 验证集中选取 6 组典型场景，涵盖光照变化、车辆遮挡、远距离小目标及细长结构，可视化对比如图 5 所示。与 PCLD-Base 相比，加入通道重标定模块后，冗余语义干扰减弱，边界预测有所改善；进一步加入边界精炼模块后，道路与路缘、植被交界处的类别过渡更连续，车辆轮廓更清晰，小目标及细长结构的边界保持更稳定。完整 BASS-Net 体现了两模块的协同作用，有效改善了复杂场景下的边界预测与细节表达。



(a) 原图　(b) PCLD-Base　(c) PCLD-Base-CR　(d) PCLD-Base-CR-Boundary　(e) BASS-Net

图 5 不同环境条件下边界分割可视化对比
Fig.5 Boundary segmentation visualization under different scene conditions

4 结 论

本文围绕复杂场景语义分割中跨分辨率特征冗余、边界细节易丢失与训练后参数冗余等问题，在前期 PCLD-Net 研究基础上构建统一基础网络 PCLD-Base，并从跨分辨率通道筛选、边界感知特征精炼与训练后权重稀疏化三个方面进行改进。实验结果表明，BASS-Net 在 Cityscapes、COCO-Stuff 10K 与 CamVid 数据集上的 mIoU 分别达到 81.30%、35.97%和 76.53%，相较 PCLD-Base 分别提高 1.16、1.91 和 1.12 个百分点。统一测试结果显示，BASS-Net 的推理速度与 PCLD-Base 接近，但其参数规模并非所有对比方法中最小，因此本文将其定位为注重分割精度和边界质量并兼顾计算效率的语义分割网络，而非轻量或实时模型。训练后的非结构化权重稀疏化主要用于减少冗余连接和提高模型压缩潜力，其实际部署加速仍取决于稀疏算子与硬件支持。未来将进一步降低边界精炼与上下文建模的计算开销，在保持精度的同时优化参数量与计算量，并结合结构化通道裁剪、稀疏算子及硬件感知部署，验证资源受限平台上的实际压缩与加速效果。

参考文献（References）

[1] ZHOU Y, ZHANG H, PARK S I, et al. Object-Centric Representation Learning for Video Scene Understanding[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 8410-8423.

[2] LUO Z, YANG W, YUAN Y, et al. Semantic segmentation of agricultural images: A survey[J]. Information Processing in Agriculture, 2024, 11(2): 172-186.

[3] ZHU Z, ZHANG Z, QI G, et al. A dual-branch network for ultrasound image segmentation[J]. Biomedical Signal Processing and Control, 2025, 103: 107368-107382.

[4] ZHAO H, QI X, SHEN X, et al. Icnet for real-time semantic segmentation on high-resolution images[C]//Proceedings of the European conference on computer vision(ECCV). Cham: Springer, 2018: 405-420.

[5] 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). Cham: Springer, 2018: 325-341.

[6] MEHTA S, RASTEGARI M, CASPI A, et al. Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation[C]//Proceedings of the european conference on computer vision (ECCV). Cham: Springer, 2018: 552-568.

[7] POUDEL R P K, LIWICKI S, CIPOLLA R. Fast-scnn: Fast semantic segmentation network[J]. arXiv preprint arXiv:1902.04502, 2019.

[8] XU R, ZHANG S, DONG C, et al. Lightweight attention refined and complex-valued BiSeNetV2 for semantic segmentation of polarimetric

SAR image[J]. Remote Sensing, 2025, 17(21): 3527.
[9] 郑秋梅， 牛薇薇， 王风华， 等. 基于细节增强的双分支实时语义分割网络[J]. 计算机应用， 2024， 44（10）： 3058-3066.
ZHENG Q M, NIU W W, WANG F H, et al. Dual-branch real-time semantic segmentation network based on detail enhancement[J]. Journal of Computer Applications, 2024, 44(10): 3058 - 3066 (in Chinese).
[10] LUO H, LIU C, SHARK L K. SABA: Scale-adaptive attention and boundary aware network for real-time semantic segmentation[J]. Expert Systems with Applications, 2025, 282: 127680.
[11] PASZKE A, CHAURASIA A, KIM S, et al. ENet: A deep neural network architecture for real-time semantic segmentation[EB/OL]. 2016. https://arxiv.org/abs/1606.02147.
[12] 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 (CVPR). Nashville: IEEE, 2021: 9716-9725.
[13] PAN H, HONG Y, SUN W, et al. Deep dual-resolution networks for real-time and accurate semantic segmentation of traffic scenes[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(3): 3448-3460.
[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 (CVPR). Vancouver: IEEE, 2023: 19529-19539.
[15] 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, 2018, 40(4): 834-848.
[16] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:1706.05587, 2017.
[17] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 801-818.
[18] GAO Y, LUO X, GAO X, et al. Semantic segmentation of remote sensing images based on multiscale features and global information modeling[J]. Expert Systems with Applications, 2024, 249: 123616.
[19] JIA S, DONG Y, MAO C, et al. Multi-path feature enhancement network for real-time semantic segmentation[J]. Neurocomputing, 2025, 637: 130099.
[20] KIRILLOV A, WU Y, HE K, et al. PointRend: Image segmentation as rendering[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 9799-9808.
[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.
[22] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 3146-3154.
[23] XU G, JIA W, WU T, et al. HAFormer: Unleashing the power of hierarchy-aware features for lightweight semantic segmentation[J]. IEEE Transactions on Image Processing, 2024, 33: 4202-4214.
[24] 郑仕敏， 毕建鹏， 于潇雁. 融合多层级特征与注意力机制的高效语义分割算法[J]. 福州大学学报（自然科学版）， 2026， 54（2）： 137-144.
ZHENG S M, BI J P, YU X Y. An efficient semantic segmentation algorithm fusing multi-level features and attention mechanism[J]. Journal of Fuzhou University (Natural Science Edition), 2026, 54(2): 137-144 (in Chinese).
[25] FANG Z, YIN B. A low functional redundancy-based network slimming method for accelerating deep neural networks[J]. Alexandria Engineering Journal, 2025, 119: 437-450.
[26] LÓPEZ-GONZÁLEZ C I, GASCÓ E, BARRIENTOS-ESPILLCO F, et al. Filter pruning for convolutional neural networks in semantic image segmentation[J]. Neural Networks, 2024, 169: 713-732.
[27] JAMSHIDI M, KALHOR A, VAHABIE A H. Efficient compression of encoder-decoder models for semantic segmentation using the separation index[J]. Scientific Reports, 2025, 15: 24639.
[28] XIE B, CAO J, ANWER R M, et al. Multi-query and multi-level enhanced network for semantic segmentation[J]. Pattern Recognition, 2024, 156: 110777.
[29] BADRINARAYANAN V, KENDALL A, CIPOLLA R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12): 2481-2495.
[30] TREML M, ARJONA-MEDINA J, UNTERTHINER T, et al. Speeding up semantic segmentation for autonomous driving[J]. 2016.
[31] ROMERA E, ALVAREZ J M, BERGASA L M, et al. Erfnet: Efficient residual factorized convnet for real-time semantic segmentation[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 19(1): 263-272.
[32] LI H, XIONG P, FAN H, et al. Dfanet: Deep feature aggregation for real-time semantic segmentation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Piscataway: IEEE, 2019: 9522-9531.
[33] ORSIC M, KRESO I, BEVANDIC P, et al. In defense of pre-trained imagenet architectures for real-time semantic segmentation of road-driving images[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Piscataway: IEEE, 2019: 12607-12616.
[34] 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(11): 3051-3068.
[35] LI X, YOU A, ZHU Z, et al. Semantic flow for fast and accurate scene parsing[C]//European conference on computer vision. Cham: Springer, 2020: 775-793.
[36] SI H, ZHANG Z, LV F, et al. Real-time semantic segmentation via multiply spatial fusion network[J]. arXiv preprint arXiv:1911.07217, 2019.
[37] KUMAAR S, LYU Y, NEX F, et al. Cabinet: Efficient context aggregation network for low-latency semantic segmentation[C]//2021 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE, 2021: 13517-13524.
[38] 陶小苗，王永顺，邬开俊，等. 面向语义分割的层次化上下文聚合与细节重建网络[J/OL]. 哈尔滨工业大学学报， [2026-10-25]. https：//link. cnki. net/urlid/23. 1235. T. 20260126. 1051. 005.
TAO X M, WANG Y S, WU K J,et al. Hierarchical context aggregation and detail reconstruction network for semantic segmentation[J/OL]. Journal of Harbin Institute of Technology, [2026-10-25].https://link.cnki.net/urlid/23.1235.T.20260126.1051.005. (in Chinese).
[39] WAN Q, HUANG Z, LU J, et al. SeaFormer: Squeeze-enhanced Axial Transformer for Mobile Semantic Segmentation[C]//International Conference on Learning Representations (ICLR). 2023.

[40] XIE E, WANG W, YU Z, et al. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers[J]. Advances in Neural Information Processing Systems, 2021, 34: 12077-12090.
[41] ZHU Z, XU M, BAI S, et al. Asymmetric non-local neural networks for semantic segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision. Piscataway: IEEE, 2019: 593-602.
[42] HE J, DENG Z, ZHOU L, et al. Adaptive pyramid context network for semantic segmentation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Piscataway: IEEE, 2019: 7519-7528.
[43] HUANG Z, WANG X, HUANG L, et al. Ccnet: Criss-cross attention for semantic segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision. Piscataway: IEEE, 2019: 603-612.
[44] LI X, ZHONG Z, WU J, et al. Expectation-maximization attention networks for semantic segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision. Piscataway: IEEE, 2019: 9167-9176.
[45] XIAO T, LIU Y, ZHOU B, et al. Unified perceptual parsing for scene understanding[C]//Proceedings of the European conference on computer vision (ECCV). Cham: Springer, 2018: 418-434.
[46] ZHANG H, DANA K, SHI J, et al. Context encoding for semantic segmentation[C]//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7151-7160.
[47] CAO Y, XU J, LIN S, et al. Gcnet: Non-local networks meet squeeze-excitation networks and beyond[C]//Proceedings of the IEEE/CVF international conference on computer vision workshops. Piscataway: IEEE, 2019.
[48] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Piscataway: IEEE,2017: 2881-2890.
[49] SUN K, ZHAO Y, JIANG B, et al. High-resolution representations for labeling pixels and regions[J]. arXiv preprint arXiv:1904.04514, 2019.
[50] WU T, TANG S, ZHANG R, et al. CGNet: A light-weight context guided network for semantic segmentation[J]. IEEE Transactions on Image Processing, 2020, 30: 1169-1179.
[51] LI G, YUN I, KIM J, et al. Dabnet: Depth-wise asymmetric bottleneck for real-time semantic segmentation[J]. arXiv preprint arXiv:1907.11357, 2019.