中图法分类号:TP751　　**文献标识码**: A　　**文章编号**: 1006-8961(XXXX)XX-0001-13

论文引用格式:Hou Zhiqiang, Li Junge, Ma Sugang, Liu Yang, Liu Lu, YU Wangsheng. Edge and Body Feature Enhancement Network for Real-time Semantic Segmentation[J/OL]. Journal of Image and Graphics, XXXX:1-13. DOI: 10.11834/jig.260252. (侯志强, 李俊歌, 马素刚, 刘洋, 刘璐, 余旺盛. 边缘与主体特征增强的实时语义分割算法[J/OL]. 中国图象图形学报,XXXX:1-13. DOI: 10.11834/jig.260252.)[DOI:10.11834/jig.260252]

边缘与主体特征增强的实时语义分割算法

侯志强[1,2],李俊歌[1,2],马素刚[1,2],刘洋[1],刘璐[3],余旺盛[4]

1. 西安邮电大学计算机学院　西安　710121; 2. 西安邮电大学陕西省网络数据分析与智能处理实验室　西安　710121;
3. 西安邮电大学网络空间安全学院　西安　710121; 4. 空军工程大学信息与导航学院　西安　710038

摘　要: **目的**　实时语义分割在自动驾驶、视频监控和增强现实等场景中具有重要应用价值,但现有方法在保持高推理速度的同时,往往难以同时保证边界细节的精确刻画和全局语义的一致性。针对这一问题,该文提出一种边缘-主体特征增强网络。**方法**　首先通过多分支协同架构与模块化注意力机制,有效提升模型对边界区域、语义信息及多尺度结构的建模能力。首先,提出边缘-主体特征解耦模块,实现边缘与主体特征的结构性分离与互补增强;其次,设计多尺度感知金字塔模块,结合多尺度平均池化与深度可分离空洞卷积,显著提升多尺度上下文建模能力并缓解空洞卷积伪影;最后,构建特征注意力融合模块,融合细粒度结构特征与语义特征,通过注意力机制强化关键区域表达。**结果**　实验结果表明,在Cityscapes数据集上,所提算法在保持65 frame/s(FPS)推理速度的同时,平均交并比(mean intersection over union,mIoU)达到79.5%;在CamVid数据集上,mIoU达到83.0%,速度保持在108.1FPS,显著优于现有主流方法。**结论**　所提方法在精度与效率之间实现了良好平衡,适用于对实时性和高精度均有较高要求的语义分割任务。https://www.scidb.cn/s/RJnYNr

关键词: 实时语义分割;多分支;信息分解;多尺度上下文;注意力机制

Edge and Body Feature Enhancement Network for Real-time Semantic Segmentation

Hou Zhiqiang[1,2], Li Junge[1,2], Ma Sugang[1,2], Liu Yang[1], Liu Lu[3], YU Wangsheng[4]

1. Institute of Computer, Xi'an University of Posts and Telecommunications, Xi'an 710121, China; 2. Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an University of Posts and Telecommunications, Xi'an 710121, China; 3. Institute of Cyberspace Security, Xi'an University of Posts and Telecommunications, Xi'an 710121, China; 4. Institute of Information and Navigation, Air Force Engineering University, Xi'an 710038, China

Abstract: **Objective** Real-time semantic segmentation plays a pivotal role in a wide range of applications, including autonomous driving, video surveillance, and augmented reality, where accurate and timely scene understanding is critical for safety, efficiency, and user experience. Despite considerable advances in this field, existing segmentation methods often face a persistent challenge: the trade-off between maintaining high inference speed and achieving fine-grained boundary precision. This difficulty arises from the inherent differences between edge features and object body features, which require fundamentally different processing strategies. Edge regions typically contain abundant high-frequency details and

收稿日期:2026-05-06;**修回日期**:2026-11-02

基金项目:国家自然科学基金(62072370);陕西省自然科学基金(2023-JC-YB-598);装备智能运用教育部重点实验室开放基金(2024-AAIE-KF02-04)

Supported by: National Natural Science Foundation of China(62072370); Natural Science Foundation of Shaanxi Province (2023-JC-YB-598); Open Fund of Key Laboratory of the Ministry of Education on Artificial Intelligence in Equipment(2024-AAIE-KF02-04)

© 中国图象图形学报版权所有

are highly sensitive to local structural variations, making them crucial for delineating object boundaries accurately and preserving small or thin structures. In contrast, the interiors of objects rely predominantly on global semantic consistency, contextual understanding, and the integration of surrounding scene information to ensure coherent object representation. These contrasting requirements frequently result in common issues, such as blurred or jagged boundaries, loss of small structures, fragmented object representation, and inaccurate segmentation under complex lighting, occlusion, or background clutter. These problems are particularly pronounced in dense urban scenes, where objects vary widely in shape, size, appearance, and spatial relationships. To tackle these limitations and address both local and global feature requirements, this study introduces the Edge-Body Feature Enhancement Network (EBFENet), a novel architecture designed to explicitly model, disentangle, and enhance edge and body features simultaneously. By leveraging a multi-branch collaborative design combined with modular attention mechanisms, EBFENet not only enables efficient feature decomposition and enhancement but also facilitates effective cross-feature interactions and information propagation, aiming to preserve both structural detail and semantic consistency. The network is optimized to maintain high inference speed for real-time deployment, providing a robust, practical, and scalable solution for high-precision semantic segmentation tasks in dynamic, complex, and challenging environments. **Method** The proposed Edge-Body Feature Enhancement Network (EBFENet) is designed around a multi-branch collaborative architecture that enables the network to effectively capture, disentangle, and integrate complementary information from edge, semantic, and contextual features while preserving real-time efficiency. At the core of the encoder, the Edge-Body Information Module (EBIM) explicitly separates edge and body features, addressing their fundamental differences: edge features contain rich high-frequency details critical for delineating object boundaries, whereas body features provide low-frequency but semantically consistent information for coherent object representation. Within EBIM, input features are processed through convolutional layers and routed into two parallel pathways. One pathway focuses on capturing high-frequency structural details through fine-grained convolutions and boundary-sensitive operations, enhancing edge clarity. The other emphasizes global semantic consistency using larger receptive fields and context-aware convolutions, ensuring object interiors are coherent. To facilitate complementary enhancement, EBIM incorporates a feature reconstruction mechanism in which edge and body representations are cross-modulated through lightweight attention operations, allowing edge information to refine boundaries while semantic features maintain context. To strengthen multi-scale understanding, a Scale-Aware Pyramid Module (SAPM) combines multi-scale average pooling with depthwise separable dilated convolutions. This design expands the receptive field to capture long-range dependencies while mitigating gridding artifacts and preserving fine-scale structures. Following feature extraction, a Feature Attention Fusion Module (FAM) adaptively fuses edge-enhanced and semantic maps using channel-wise attention to emphasize informative channels while suppressing less relevant ones. Residual connections facilitate information propagation and improve gradient flow. Collectively, EBIM, SAPM, and FAM work synergistically: EBIM disentangles and enhances features, SAPM provides scale-aware contextual understanding, and FAM selectively integrates them, balancing boundary precision and semantic consistency. The multi-branch design allows parallel processing of specialized features, reducing redundancy, while efficient operations such as depthwise separable convolutions, lightweight attention mechanisms, and pooling strategies maintain high inference speed suitable for real-time deployment. Through this combination of modular components and attention-driven feature interaction, EBFENet effectively addresses real-time semantic segmentation challenges in complex urban environments, achieving superior accuracy without sacrificing efficiency. **Result** Extensive experiments demonstrate that EBFENet achieves state-of-the-art performance on multiple benchmarks while maintaining high inference speed. On the Cityscapes dataset, EBFENet reached 79.5%/79.2% mIoU on the validation and test sets, respectively, with an inference speed of 65 FPS, outperforming competitive methods such as BiSeNet, DDRNet-23-S, and PIDNet-S across multiple metrics (Tables 4 and 5). Similarly, on the CamVid dataset, EBFENet attained an mIoU of 83.0% at 108.1 FPS, surpassing HSNet-L and PIDNet-S in both accuracy and efficiency. Qualitative evaluations reveal that the proposed network produces sharper object boundaries, more precise contours, and better segmentation of small, thin, or partially occluded objects compared to baseline approaches (Figs. 7 and 8). Detailed ablation studies confirm that each module—EBIM, SAPM, and FAM—contributes significantly to the overall improvement in both segmentation accuracy and real-time performance, demonstrating the effectiveness of the proposed multi-branch, attention-driven design (Tables 1

© 中国图象图形学报版权所有

and 2). **Conclusion** In summary, this work presents a novel multi-branch feature attention and information decomposition network tailored for real-time semantic segmentation. The Edge-Body Information Module (EBIM) decouples and complementarily enhances edge and body features, improving the representation of boundaries and local structures. The Scale-Aware Pyramid Module (SAPM) aggregates multi-scale contextual information using a combination of average pooling and depthwise separable atrous convolutions, thereby enhancing the network's receptive field and reducing gridding artifacts. Finally, the Feature Attention Fusion Module (FAM) strengthens the global semantic representation and selectively emphasizes key regions during cross-branch feature fusion, leading to superior structural integrity and semantic consistency in segmentation outputs. Compared with existing real-time segmentation methods, EBFENet achieves a favorable balance between accuracy and inference speed, demonstrating its potential for deployment in practical scenarios that require high-precision and efficient scene understanding. Future work will focus on exploring dynamic cross-branch feature interactions and adaptive weight allocation strategies to further improve generalization, robustness, and performance in highly complex or previously unseen urban environments.

Key words: Real-time semantic segmentation; Multi-branch network; Information decomposition; Multi-scale context; Attention mechanism

论文引用格式：Hou Zhiqiang, Li Junge, Ma Sugang, Liu Yang, Liu Lu, Yu Wangsheng. Edge and Body Feature Enhancement Network for Real-time Semantic Segmentation[J/OL]. Journal of Image and Graphics. DOI: 10.11834/jig.260252.（侯志强，李俊歌，马素刚，刘洋，刘璐，余旺盛. 边缘与主体特征增强的实时语义分割算法[J/OL]. 中国图象图形学报. DOI: 10.11834/jig.260252.）

0 引 言

语义分割是计算机视觉中的一项基础且关键的任务，目的是为图像中的每一个像素赋予精确的语义类别标签。广泛应用于自动驾驶(Zhao等，2024)、遥感测绘(Cheng等，2024)、医学图像分析(Azad等，2024)等多个领域，为环境感知与智能决策提供核心支持。

在街景理解等复杂场景中常存在背景多样、物体尺度差异显著以及目标边界模糊等问题，对分割的精度与推理效率均提出了很高要求。为了提高分割精度，Long等人(Long等，2015)提出的全卷积网络开创性地实现了端到端的像素级预测。SegNet(Badrinarayanan等，2017)利用最大池化索引实现精确上采样，DeepLab系列(Chen等，2017)(Chen等，2018)通过空洞卷积与空洞空间金字塔池化有效捕获多尺度上下文信息，PSPNet(Zhao等，2017)提出空间金字塔池化模块融合全局上下文，从而提升了多尺度特征建模能力。HRNet(Yu等，2021)通过多分辨率并行结构保留了高分辨率特征表示。虽然这些方法在精度上表现优异，但普遍依赖深层重型骨干网络与复杂的特征融合结构，计算成本较高，限制了其在实时场景和边缘设备上的部署。为提升推理速度与轻量化特性，ICNet(Zhao等，2018)利用多分辨率输入与级联预测实现高效推理；BiSeNet系列(Yu等，2018)(Yu等，2021)采用双分支结构分别建模空间细节与语义上下文，使网络在实时推理条件下仍能获得较高精度；STDCNet(Fan等，2021)引入短期密集连接模块在保持轻量化的同时扩展感受野；之后，DDRNet(Pan等，2023)基于双分辨率并行分支构建高效的主干结构并结合深度聚合金字塔池化模块增强多尺度感受野；PIDNet(Xu等，2023)在双分支框架的基础上引入边界感知分支，在保证推理速度的同时显著提升了整体语义分割性能；RegSeg(Gao，2023)通过结合基于分组卷积与轻量级解码器，实现长程依赖建模与局部细节恢复。尽管这些方法有效平衡了速度与精度，但在复杂街景场景中仍面临语义一致性与边界锐度难以兼顾的问题。

近年来，注意力机制在语义分割中被广泛应用，用以增强关键区域的特征响应与判别能力。DANet(Fu等，2019)率先引入双重注意力机制，通过空间与通道注意力并行建模实现了全局与局部信息的协同增强，LEDNet(Wang等，2019)提出上下文嵌入模块，将通道压缩与空间注意相结合，在保证实时性的同时提升浅层特征的判别力。GCNet(Cao等，2019)通过上下文引导模块实现局部注意与全局一致性的联合建模。Chen等人(Chen等，2024)提出自适应注

© 中国图象图形学报版权所有

意力机制融合模块，用于复杂场景的实时语义分割，实现不同特征间的动态权重分配与关键区域增强建模。LCTNet（Hou 等，2025）提出了轻量级卷积神经网络-变换器（convolutional neural network-Transformer，CNN-Transformer）结合的实时语义分割网络，通过小波增强卷积模块、多尺度融合模块和轻量级 Transformer，有效整合局部与全局语义信息，实现了城市街景场景下高精度、实时且参数量小的分割性能。尽管现有注意力方法在特征增强方面取得显著进展，但其大多侧重于全局或通道维度的特征重加权，对于边界细粒度特征的保持与跨层信息中的冗余抑制仍显不足。

在语义分割任务中，边界区域与主体区域在表征需求上具有天然差异：边界区域富含高频信息和精确轮廓，要求高空间分辨率与敏锐的局部结构建模；而主体区域更依赖语义一致性与全局上下文。若采用单一特征流进行统一建模，常导致边界模糊、小目标漏检等问题。因此，如何在保持实时性的同时兼顾边界细节恢复与语义一致性建模，仍是当前实时语义分割方法面临的重要挑战。

针对上述挑战，本文提出了一种边缘-主体特征增强网络（Edge-Body Feature Enhancement Network，EBFENet），在多分支结构的基础上引入模块化注意力机制，有效提升模型在边界区域、语义信息与多尺度方面的能力，兼顾精度与速度。主要贡献如下：

1）提出边缘-主体信息模块，通过信息解耦与互补增强机制实现边缘与主体特征的结构性分离与融合，有效提升边界定位精度与区域判别能力。

2）提出多尺度感知金字塔模块，结合多尺度平均池化与深度可分离空洞卷积，实现丰富上下文信息建模并缓解空洞卷积的伪影问题。

3）提出特征注意力融合模块，融合边缘与边界分支的细粒度特征与语义特征，并通过通道注意力机制增强关键区域特征表达能力。

1 方　法

本文所提出的用于实时语义分割的边缘-主体特征增强网络的整体架构如图 1 所示。网络主要由语义分支、细节分支和边界分支构成。初始输入图像经过多层卷积处理后，生成空间分辨率为原图 1/8

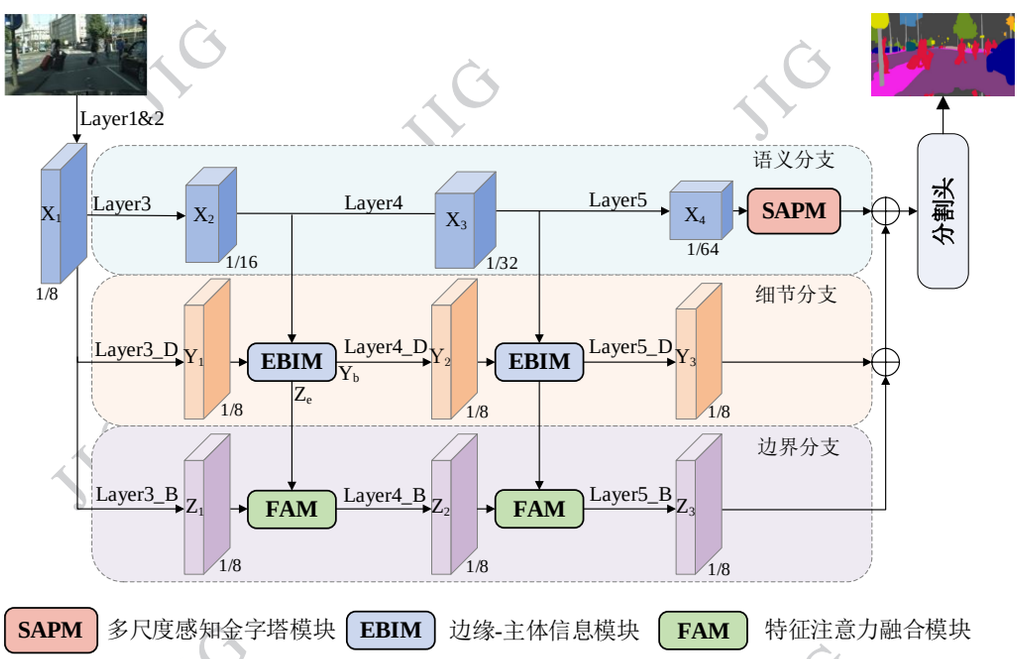


图 1　EBFENet 整体架构示意图

Fig. 1　Overall structure of EBFENet

的特征图 X_1，并分别输入至三个分支以进行后续处理。在语义分支中，X_1 经逐步下采样生成多个尺度的特征表示 X_2、X_3 和 X_4，其空间分辨率分别为原图的 1/16、1/32 和 1/64，以捕获深层次的语义上下文

© 中国图象图形学报版权所有

信息。为进一步增强语义表达能力，本文在语义分支最后阶段引入了多尺度感知金字塔模块（scale-aware pyramid module，SAPM），融合多尺度全局上下文信息。细节分支与边界分支保持特征图的空间分辨率为1/8，从而更好地保留图像中的局部结构和边缘细节。在细节增强路径中，本文提出了边缘-主体信息模块（edge-body information module，EBIM），该模块融合来自语义分支中间层特征X_2与细节分支的初始特征Y_1，生成物体边界特征Z_e与物体内部特征Y_b。其中Y_b经由细节分支主干网络进一步处理后，作为输入进入下一个阶段的EBIM模块，用以挖掘更丰富的内部结构信息。在边界分支中，本文设计了特征注意力融合模块（feature attention module，FAM），以实现对边缘信息的精细化建模。边界分支的初始特征Z_1与EBIM输出的边缘特征Z_e一同输入至FAM模块，输出融合后的边界特征Z_2，以进一步提升边界定位的准确性。最后，来自语义分支的深层特征X_4经SAPM处理后，与融合后的细节特征Y_3和边缘特征Z_3进行加权融合，并通过语义分割头进行像素级分类预测，生成最终的语义分割结果。其中，EBIM模块和FAM模块在网络中被重复应用两次，用于捕获不同阶段的边缘与内部特征信息。

以下将分别详细介绍本文提出的三个核心模块：边缘-主体信息模块（EBIM）、多尺度感知金字塔模块（SAPM）以及特征注意力融合模块（FAM）。

1.1 边缘–主体信息模块

在语义分割任务中，边缘区域与主体区域存在显著差异：前者包含丰富的高频结构信息，关注定位精度；后者侧重全局语义与上下文一致性。然而，现有分割网络往往难以兼顾两者建模，导致边界模糊或语义表达不足。为此受Li等人（Li等，2020）的启发，本文提出了一种边缘-主体信息模块（EBIM），通过结构性设计与信息解耦机制，对边缘与主体区域间的特征差异进行建模，实现高效的互补增强。EBIM的整体结构如图2所示。

输入特征图X_2首先经过一次下采样操作，压缩其空间分辨率，以聚合更丰富的上下文语义信息，并抑制局部纹理干扰。随后通过上采样将特征恢复至原始尺寸，构造一个具有上下文但相对模糊的辅助特征表示。为了融合上下文感知与原始细节信息，得到的辅助特征表示与X_2在通道维度上进行拼接，得到融合后的特征，使模型能够同时感知边缘的高频信息与区域的语义背景。该融合特征随后输入一个3×3的卷积层，用以提取局部空间结构与纹理边界信息，并初步显式建模边缘与区域的混合特征分布得到。

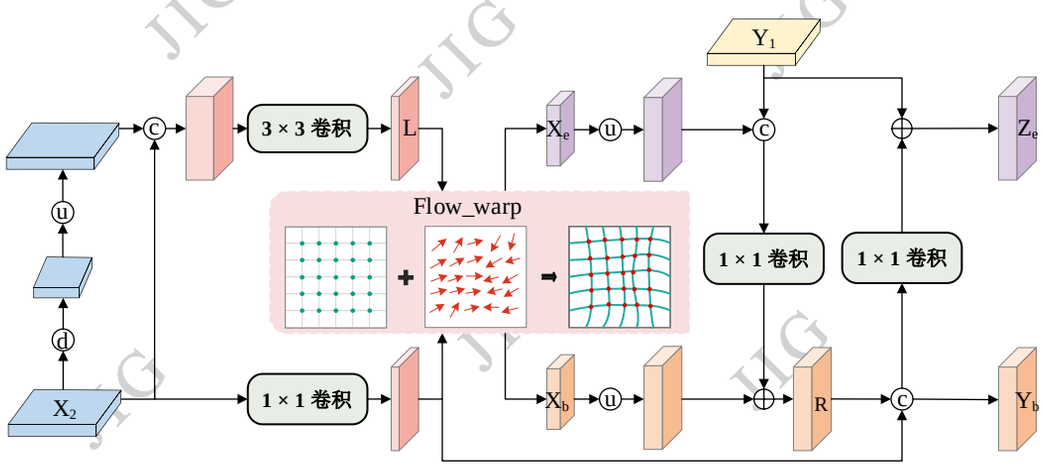


图2　EBIM结构图

Fig. 2　Overall structure of EBIM

具体操作步骤可表示为：

$$\boldsymbol{L}=\mathrm{Con}\,v_3\left(\mathrm{Concat}\left(\mathrm{Up}\left(\mathrm{Down}\left(\boldsymbol{X}_2\right),\boldsymbol{X}_2\right)\right)\right) \tag{1}$$

式中，Up和Down分别指上采样和下采样操作，Concat指拼接操作，$\mathrm{Con}\,v_3$指3×3的卷积操作。

接着，引入Flow_warp模块（Li等，2020），该模块以建模特征内部的显式流动关系为核心，模拟特征从边缘到区域之间的过渡过程，生成一个流动场，基于该流动场进行反向采样，得到区域特征X_b，X_b再与原始特征相减以得到边缘响应X_e。之后为实现边缘

© 中国图象图形学报版权所有

精细定位与语义补偿，X_e被上采样至原尺寸，并与细节分支特征Y_1进行拼接，整合两者的边缘信息，拼接后的边缘增强特征通过1×1的卷积与上采样后的特征X_b进行特征相加操作，提升边缘识别过程中的语义完整性，得到特征R。最终再将R与经1×1卷积细化后的原始输入特征X_2拼接，得到包含区域结构语义与边缘细节联合表示的特征Y_b。该特征经过1×1卷积后与Y_1进行逐元素加法操作，输出边缘增强特征Z_e。

实现过程可表示为：

$$\boldsymbol{X}_e,\boldsymbol{X}_b = \text{Flow_warp}\left(\left(\text{Con v}_1\left(\boldsymbol{X}_2\right),\boldsymbol{L}\right)\right) \tag{2}$$

$$\boldsymbol{R} = \text{Up}\left(\boldsymbol{X}_b\right)\oplus \text{Con v}_1\left(\text{Concat}\left(\text{Up}\left(\boldsymbol{X}_e\right),\boldsymbol{Y}_1\right)\right) \tag{3}$$

$$\boldsymbol{Y}_b = \text{Concat}\left(\boldsymbol{R},\text{Con v}_1\left(\boldsymbol{X}_2\right)\right) \tag{4}$$

$$\boldsymbol{Z}_e = \text{Con v}_1\left(\boldsymbol{Y}_b\right)\oplus \boldsymbol{Y}_1 \tag{5}$$

式中，Flow_warp是自定义的操作，⊕表示逐元素相加操作，Con v_1指1×1的卷积操作。

通过结构性的信息分离与重构机制，EBIM能有效缓解边缘-区域建模失衡问题，提升边界定位的精确度与区域判别的鲁棒性。同时，模块将区域语义特征反馈至细节分支，实现边缘识别的语义支撑，增强两个子分支间的信息流动与任务协同。

1.2 多尺度感知金字塔模块

受MLFNet（Fan等，2023）中提出的多尺度感知思想及PIDNet（Xu等，2023）中PAPPM模块的启发，本文提出了一种多尺度感知金字塔模块（SAPM），其结构如图3所示，在提升多尺度上下文建模能力的同时，缓解空洞卷积带来的伪影问题并兼顾计算效率。

在SAPM模块中，首先对输入特征图X进行1×1卷积以调整通道维度。同时特征图X通过三个并行分支进行处理，每个分支均先进行3×3平均池化操作以聚合局部上下文信息，然后分别引入不同感受野的深度可分离空洞卷积操作，卷积核大小分别设为3×3、5×5、7×7，膨胀率依次为6、12、18，得到三个包含不同尺度信息的特征图，该设计能够在保证有效感受野扩展的同时缓解大膨胀率导致的特征不连续问题。与传统的金字塔结构不同，SAPM在多尺度分支设计中引入了更紧密的跨尺度信息融合机制。三个空洞卷积分支在处理由3×3平均池化得到的局部上下文特征的基础上，均与第一个经1×1卷积分支的输出特征进行融合得到特征图F_2、F_3、F_4，即通过逐元素相加的方式增强语义表达的一致性与稳定性。这种设计不仅保留了初始特征图中的基本结构信息，还显式引导各尺度分支关注语义主干的全局上下文，从而提高多尺度特征之间的协同性与语义一致性。为了进一步补充全局上下文信息，SAPM还引入了一个自适应平均池化分支，将输入特征图X进行自适应平均池化后通过1×1卷积压缩通道数，再上采样至原始空间分辨率，得到全局特征图。

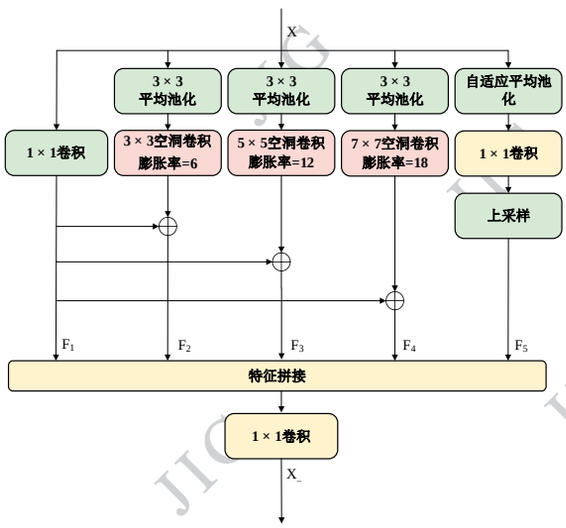


图3 SAPM结构图

Fig. 3 Overall structure of SAPM

上述过程可形式化表示为：

$$\boldsymbol{F}_1 = \text{Conv}_1(\boldsymbol{X}) \tag{6}$$

$$\boldsymbol{F}_i = \text{DConv}_{k_i}^{r_i}(\text{AvgPool}(\boldsymbol{X}))\oplus \boldsymbol{F}_1, i = 2,3,4 \tag{7}$$

$$\boldsymbol{F}_5 = \text{Up}(\text{Con v}_1(\text{AAP}(\boldsymbol{X}))) \tag{8}$$

式中，$\text{DConv}_{k_i}^{r_i}$表示核大小为k_i、膨胀率为r_i的深度可分离空洞卷积操作，AvgPool表示3×3平均池化操作，AAP为自适应平均池化。

接着，将三个分支输出的局部特征图F_2、F_3、F_4与全局特征F_5以及特征F_1进行通道维度上的拼接，再通过一个1×1卷积进行通道压缩，得到最终输出特征图：

$$\boldsymbol{X}_- = \text{Con v}_1(\text{Concat}(\boldsymbol{F}_1,\boldsymbol{F}_2,\boldsymbol{F}_3,\boldsymbol{F}_4,\boldsymbol{F}_5)) \tag{9}$$

1.3 特征注意力融合模块

为了进一步提升模型对边界区域结构信息的感知与表达能力，本文设计了一种特征注意力融合模块（FAM），充分融合边缘分支所提取的细粒度结构信息与边界分支中的语义特征，利用注意力机制实

© 中国图象图形学报版权所有

现更具选择性的特征增强，从而提高模型对目标轮廓和细节区域的响应能力，其结构如图4所示。

在FAM模块中，输入包括边界分支提取的原始特征图Z_I以及来自EBIM模块的边缘特征图Z_e。首先，二者通过逐元素加法操作进行初步融合，得到包含语义与结构信息的中间特征，有效缓解了边界分支可能因下采样或语义压缩而导致的细节缺失问题，强化了边界处的结构表达能力。

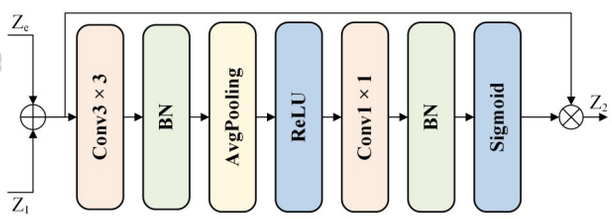


图4　FAM结构图

Fig. 4　Overall structure of FAM

随后融合特征输入一个特征注意力生成分支，用于动态建模不同通道的重要性。该分支首先通过一个3×3卷积层提取上下文特征，随后依次通过批归一化、平均池化与ReLU激活函数，对特征进行压缩与非线性转换，再利用一个1×1卷积与第二个批归一化层进行降维与归一化处理，最后通过Sigmoid函数生成通道注意力权重A：

$$Z_f = Z_I \oplus Z_e \tag{10}$$

$$V = \mathrm{ReLU}(\mathrm{AvgPool}(\mathrm{BN}(\mathrm{Con}\,v_3(Z_f)))) \tag{11}$$

$$A = \mathrm{Sigmoid}(\mathrm{BN}(\mathrm{Con}\,v_3(V))) \tag{12}$$

式中，BN表示批归一化，AvgPool为平均池化。

该注意力向量被用于对融合后的特征进行逐通道加权操作，以实现通道维度上的特征选择与增强：

$$Z_2 = Z_I \otimes A \tag{13}$$

其中，⊗表示逐通道乘法操作。

2　实　验

2.1　数据集

为验证所提出算法的有效性，本文在Cityscapes（Cordts等，2016）和CamVid（Brostow等，2009）两个语义分割数据集上进行了实验。Cityscapes是一个广泛用于城市街景语义分割任务的公开数据集。本文实验仅使用精细标注图像，图像分辨率均为2048×1024。数据集中包括2，975张用于训练的图像、500张验证图像和1，525张测试图像。CamVid是专门用于场景理解与语义分割的数据集，该数据集共包含701张分辨率为960×720的图像，其中367张用于训练，101张用于验证，233张用于测试。

2.2　实验环境

本文的实验在Ubuntu 16.04系统环境下进行，使用一块NVIDIA GeForce RTX 3090显卡进行模型训练，实验平台基于Python 3.7、PyTorch 1.8.1和CUDA 11.1构建。

模型训练采用带有动量的随机梯度下降（SGD）优化算法，其中初始学习率设为0.01，权重衰减系数为0.0005，动量参数为0.9。为了增强模型的泛化能力，对训练图像进行了数据增强，包括随机裁剪、随机水平翻转以及缩放比例范围为[0.5，2.0]的随机缩放。在训练过程中，对于Cityscapes数据集，图像被随机裁剪为1024×1024，训练迭代次数为520；对于CamVid数据集，图像被裁剪为960×720，训练迭代次数为200。在推理阶段，Cityscapes和CamVid数据集分别采用2048×1024和960×720分辨率的图像作为输入。推理速度在单张GTX 3090 GPU上测量，batch size设置为1。

2.3　消融实验

2.3.1　模块性能分析

本文在Cityscapes数据集上对三个模块进行了消融实验，包括边缘-主体信息模块（EBIM）、多尺度感知金字塔模块（SAPM）以及特征注意力融合模块（FAM）。采用在保持65 frame/s（FPS）平均交并比（mean intersection over union，mIoU）、每秒帧数（frames per second，FPS）、浮点运算次数（FLOP）和模型参数量（Parameter）指标评估相应的模型，分析、验证各模块对提升算法整体的性能表现。为了更好地说明的有效性，本文进行了如表1所示的消融实验。

在第一组实验中，仅采用Baseline网络结构，未引入任何增强模块，mIoU为78.3%，参数量为7.6M，FLOPs为47.6G，推理速度为85FPS。第二组实验在Baseline基础上引入EBIM模块，通过强化边缘与主体区域之间的信息交互，使mIoU提升至78.8%。第三组实验引入SAPM模块，进一步提升了模型的多尺度上下文建模能力，mIoU达78.6%。第四组实验引入FAM模块，增强了不同尺度特征间的融合能力，mIoU提升至78.5%。

© 中国图象图形学报版权所有

表1　本文算法各个模块在Cityscapes验证集上的消融实验研究

Table 1　This paper studies the ablation experiments of each module of the algorithm on the Cityscapes validation set.

Baseline	EBIM	SAPM	FAM	mIoU (%)	Params (M)	FLOPs (G)	Speed (FPS)
✓				78.3	7.6	47.6	85
✓	✓			78.8	9.2	114.6	75
✓		✓		78.6	8.7	106.4	80
✓			✓	78.5	9.3	112.6	67

注：✓表示启用了对应模块。

在第五组至第七组实验中，分别组合使用了任意两个模块。第五组同时引入EBIM与SAPM，mIoU进一步提升至79.2%，验证了EBIM与SAPM在边缘细节还原与语义建模方面的互补性；第六组结合EBIM与FAM，mIoU为78.9%；第七组联合使用SAPM与FAM，mIoU为78.8%。第八组实验集成了全部三个模块，模型在保持较低参数量与可接受计算开销的前提下，实现了79.5%的mIoU，显著优于Baseline，充分证明了EBIM、SAPM与FAM三者在提升语义分割精度方面的协同增益作用。

2.3.2　EBIM模块性能分析

为了增强模型对边缘信息的建模能力，提升边界区域的结构感知与细节表达，本文设计了边缘-主体信息交互模块（EBIM）。该模块通过引导语义特征与边缘特征之间的信息交互，有效提升边界感知能力和目标区域的分割精度。为进一步验证EBIM模块的有效性，本文设计了表2的对比实验。

在第一组实验中，模型未使用任何EBIM模块（记作N/A），mIoU为78.3%，参数量为7.6M，FLOPs为47.6G，推理速度为85FPS。第二组实验中，在语义分支1/16特征图位置引入了EBIM模块（EBIM-stage1），该阶段主要针对中等尺度的边缘区域进行感知与补偿，mIoU提升至78.6%，说明该模块能够有效增强中尺度语义特征与边缘信息的融合。第三组实验在语义分支1/32特征图位置引入EBIM模块（EBIM-stage2），着重捕捉较大感受野下的结构边界信息，mIoU为78.4%。尽管提升略小于EBIM-stage1，但仍验证了在高层语义中引入边缘信息具有一定的效果。第四组实验则同时在两个阶段均引入EBIM模块，mIoU进一步提升至78.8%，且整体推理速度仍保持在75FPS，达到了精度与效率的良好平衡。

表2　EBIM模块在Cityscapes验证集上的消融实验研究

Table 2　Ablation experiment of EBIM module on Cityscapes validation set

Method	MIoU (%)	Params (M)	FLOPs (G)	Speed (FPS)
N/A	78.3	7.6	47.6	85
EBIM-stage1	78.6	8.7	84.7	80
EBIM-stage2	78.4	9.8	85.2	78
EBIM-stage1&2	**78.8**	**9.2**	**114.6**	**75**

注：粗体为最优结果。

2.4　定量分析

2.4.1　Cityscapes数据集

为进一步验证本文提出模型在实时语义分割任务中的性能优势，本文在Cityscapes数据集上选取近年来具有代表性的高性能分割算法进行对比实验，结果如表3所示。

从表3中结果可以看出，本文模型在保持较快速度的同时，在准确性方面也达到了较高水平。其中，本文模型在Cityscapes验证集与测试集上分别取得了79.5%和79.2%的mIoU，推理速度达到65 FPS，表现出良好的速度-精度权衡性能。与BiSeNet、BiSeNetV2等经典实时分割算法相比，本文方法在

准确率上均实现了显著提升，mIoU提升幅度达到4%以上，且在推理速度方面仍保持较强优势，优于BiSeNetV2等方法，显示出良好的实用性。与近年高精度实时分割模型如DDRNet-23-S、STDC2-Seg75、RTFormer-S、PIDNet-S等方法相比，本文方法在准确率上具有明显优势，尤其在测试集上表现优异，mIoU达到79.2%，高于上述所有模型。此外，本文方法的推理速度达到65 FPS，远高于SegNext-T与AFFormer-B等算法，体现出本文模型更强的部署效率。值得一提的是，尽管RegSeg、HSNet-L等方法在精度上亦有不错表现，但在推理速度上大多不具备实时性能。而本文模型在精度与速度间取得了更加均衡的表现，为直观地展示分析结果，本文给出算法在Cityscapes测试集上的准确性-速度比较的散点图，如图5所示。

© 中国图象图形学报版权所有

表3　Cityscapes数据集上的准确性和速度比较

Table 3　Accuracy and speed comparison on the Cityscapes dataset

Method	Ref	Resolution	mIoU(%)		Speed(FPS)
			val	test	
BiSeNet(Yu等,2018)	ECCV2018	2048×1024	74.8	74.4	65.5
BiSeNetV2(Yu等,2021)	IJCV2020	1024×512	75.8	75.3	47.3
STDC2-Seg75(Fan等,2021)	CVPR2021	1536×768	77.0	76.8	97.0
DDRNet-23-S(Pan等,2023)	CVPR2021	2048×1024	77.2	76.7	102.0
RTFormer-S(Wang等,2022)	NeurIPS2022	2048×1024	76.3	-	110.0
SegNext-T(Guo等,2022)	NeurIPS2022b	2048×1024	79.8	-	28.1
PIDNet-S(Xu等,2023)	CVPR2023	2048×1024	78.8	78.6	93.2
AFFormer-B(Dong等,2023)	AAAI 2023	2048×1024	78.7	-	28.4
RegSeg(Gao,2023)	CVPR2023	1024×2048	78.5	78.3	37.0
HSNet-L(Peng等,2024)	IEEE TSMC-S2024	1024×512	-	80.0	33.0
DMANet-B(Dong等,2025)	Neurocomputing2024	1024×512	78.5	-	90.3
EBFENet (Ours)	-	**2048　1024**	**79.5**	**79.2**	**65.0**

注:粗体为本论文结果;val与test分别为Cityscapes验证集与测试集上的mIoU;“-”表示原文献未报告。

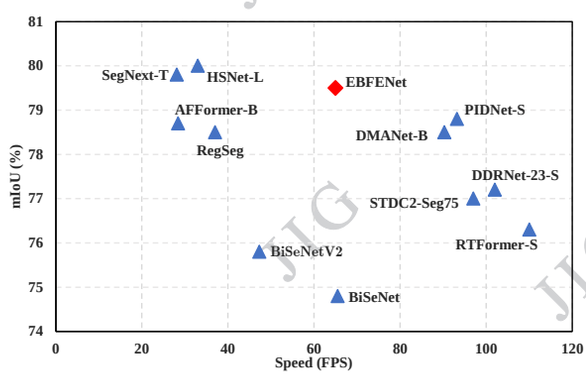


图5　在Cityscapes数据集上的准确-速度比较示意图

Fig. 5　A diagram illustrating the accuracy-speed comparison on the Cityscapes dataset.

2.4.2　CamVid数据集

如表4所示,本文方法与近年来在CamVid测试集上具有代表性的实时语义分割算法进行了比较,为直观地展示分析结果,本文给出算法在CamVid测试集上的准确性-速度比较的散点图,如图6所示。从表中结果可以看出,早期方法如ICNet、BiSeNet在速度方面表现尚可,但精度相对较低,mIoU均未超过68%。随后提出的BiSeNetV2及其改进版本在精度上有所提升,其中BiSeNetV2-L在mIoU上达到73.2%,但推理速度明显下降,仅为33 FPS。CSRNet-heavy、Cross-CBAM-L等方法在精度上进一步提升,但速度均处于中等水平。

高精度模型方面,DDRNet-23及PIDNet-S等方法的mIoU均超过79%,其中DDRNet-23达到80.6%,PIDNet-S则在保持80.1%精度的同时实现了153.7 FPS的高速度,具备较好的平衡性。

与上述方法相比,本文提出的EBFENet在CamVid测试集上取得了83.0%的mIoU,显著高于现有最优方法HSNet-L(80.7%)与PIDNet-S(80.1%),精度提升幅度分别为2.3%和2.9%;同时保持了108.1 FPS的推理速度,能够满足实时分割需求。这表明本文方法在保证高精度的同时兼顾了推理效率,达到了速度与精度的优良平衡。综上本文方法在CamVid数据集上相较于现有主流实时语义分割网络表现出更为优越的综合性能,适用于对分割精度和处理速度均有较高要求的实际场景。

2.5　定性分析

在Cityscapes和CamVid数据集上,图7与图8呈现了本文算法同基准算法的定性对比结论,并利用白色矩形框对复杂区域进行了标注。

2.5.1　Cityscapes数据集

图7展示了本章提出的网络与基线方法在定性结果上的对比,其中白色方框标注了具有挑战性的区域。可以观察到,基线方法在部分场景中存在错

© 中国图象图形学报版权所有

表4　CamVid测试集的准确性和速度比较

Table 4　Accuracy and speed comparison of the CamVid test set

Method		Ref	Resolution	Speed(FPS)
ICNet(Zhao等,2018)	ECCV2018	720×960	67.1	27
BiSeNet(Yu等,2018)	ECCV2018	960×720	65.6	175
BiSeNetV2-L(Yu等,2021)	IJCV2020	960×720	73.2	33
DDRNet-23(Pan等,2023)	CVPR2021	960×720	80.6	94.0
MFNet(Lu等,2022)	T-ITS2022	960×720	71.5	145
CSRNet-heavy(Xiong等,2023)	ESWA2022	960×720	74.5	46.4
Cross-CBAM-L(Zhang等,2023)	arXiv2023	960×720	75.6	146.8
PIDNet-S(Xu等,2023)	CVPR2023	960×720	80.1	153.7
HSNet-L(Peng等,2024)	IEEE TSMC-S2024	960×720	80.7	74.2
DMANet-B(Dong等,2025)	Neurocomputing2024	960×720	78.7	74.6
EBFENet (Ours)	-	**960　720**	**83.0**	**108.1**

注:粗体为本论文结果;该表为CamVid测试集结果。

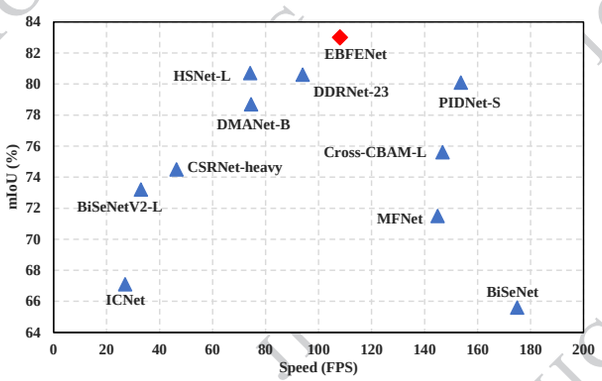


图6　在CamVid数据集上的准确-速度比较示意图

Fig. 6　A diagram illustrating the accuracy-speed comparison on the CamVid dataset

误标注,而本章方法能够实现更加精确的分割,充分体现了其有效性。例如,在前两行图像中,本章方法在公交车与两名行人之间被部分遮挡的摩托车区域上,分割结果更为精细,而基线方法的结果相对粗糙;在第三行与第五行图像中,基线方法在行人边缘的分割表现不佳,而本章方法能够准确保留边缘细节;在第四行图像中,本章方法对交通标识牌的内部区域分割更加完整且形状准确,而基线方法未能正确分割出标识牌的形状。

2.5.2　CamVid数据集

图8展示了本章提出的算法与基线方法在CamVid数据集上的可视化对比,其中白色矩形框选定了难度较大的局部区域。较之基准算法,EBFENet可实现更精准的类别辨析并达成分割任务。在前两行样本中,针对房屋及墙面这类大尺寸目标,EBFENet的分割效能更具精确性,且边缘轮廓保持得更为完备。

而在第三至第五行图片中,EBFENet在柱状物体、草丛以及路面等区域的边缘分割上表现出更高的精细度与准确性,明显优于基线算法。

3　结　论

本文提出了一种多分支协同结构的特征注意信息分解网络,边缘-主体信息模块通过对边缘与主体特征进行结构性解耦与互补增强,提升了模型对边界细节和局部结构的表征能力。多尺度感知金字塔模块融合多尺度平均池化与深度可分离空洞卷积,有效增强了多尺度上下文感知能力,缓解了空洞卷积带来的伪影问题。特征注意力融合模块在融合不同分支特征的同时,强化了全局语义与关键区域的特征表达,使最终分割结果在结构完整性与语义一致性方面得到显著提升。与现有主流实时分割算法相比,本文在分割精度与推理速度上均取得了良好平衡,能够实现实时分割效果。未来的研究将重点探索跨分支特征交互与动态权重分配机制,以进一步提升模型的泛化性能与鲁棒性。

© 中国图象图形学报版权所有

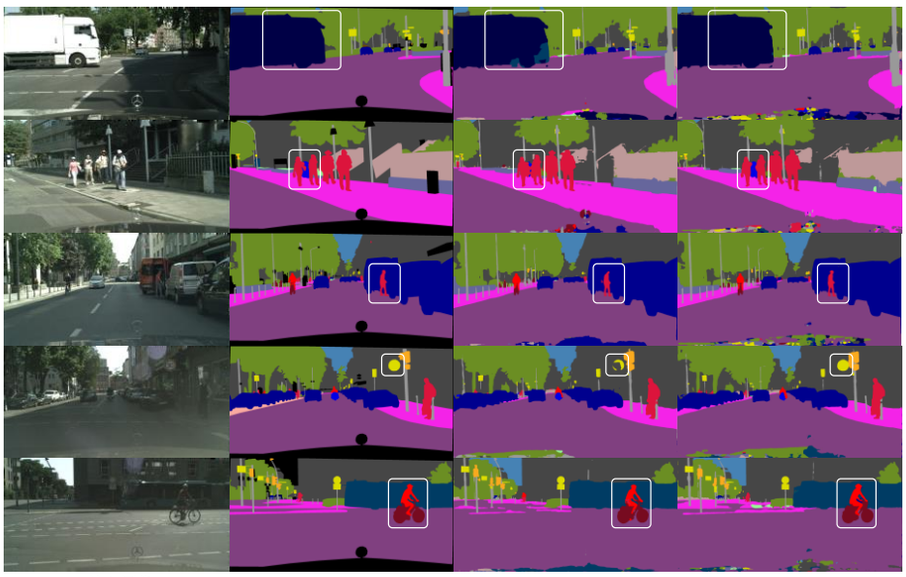


图7　Cityscapes数据集的分割效果可视化展示。

Fig. 7　Visualization of segmentation results from the Cityscapes dataset

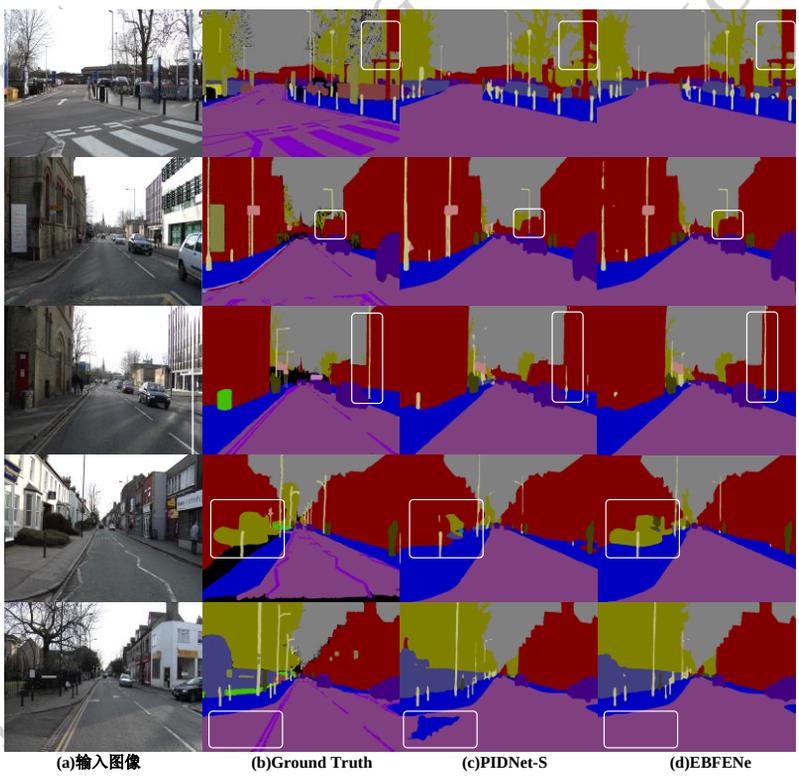


图8　CamVid数据集的分割效果可视化展示。

Fig. 8　Visualization of segmentation results from the CamVid dataset

参考文献(References)

Azad R, Aghdam E K, Rauland A, Jia Y, Avval A H, Bozorgpour A, et al. 2024. Medical image segmentation review: The success of U-Net. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12): 10076-10095[DOI: 10.1109/TPAMI.2024.3435571]

Badrinarayanan V, Kendall A and Cipolla R. 2017. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence,

© 中国图象图形学报版权所有

39(12): 2481-2495 [DOI: 10.1109/TPAMI.2016.2644615]

Brostow G J, Fauqueur J and Cipolla R. 2009. Semantic object classes in video: A high-definition ground truth database. Pattern Recognition Letters, 30(2): 88-97 [DOI: 10.1016/j.patrec.2008.04.005]

Cao Y, Xu J, Lin S, Wei F and Hu H. 2019. GCNet: Non-local networks meet squeeze-excitation networks and beyond//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Seoul: IEEE: 1971-1980 [DOI: 10.1109/ICCVW.2019.00246]

Chen D, Liu L, Wang C H, Bai X R and Wang Z C. 2024. Adaptive attention mechanism fusion for real-time semantic segmentation in complex scenes. Journal of Electronics & Information Technology, 46(8): 3334-3342 (陈丹, 刘乐, 王晨昊, 白熙茹, 王子晨. 2024. 复杂场景下自适应注意力机制融合实时语义分割. 电子与信息学报, 46(8): 3334-3342) [DOI: 10.11999/JEIT231338]

Chen L C, Papandreou G, Kokkinos I, Murphy K and Yuille A L. 2018. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4): 834-848 [DOI: 10.1109/TPAMI.2017.2699184]

Chen L C, Papandreou G, Schroff F and Adam H. 2017. Rethinking atrous convolution for semantic image segmentation [EB/OL]. [2017-06-30]. https://arxiv. org/abs/1706.05587 [DOI: 10.48550/arXiv. 1706.05587]

Chen L C, Zhu Y, Papandreou G, Schroff F and Adam H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation//Computer Vision - ECCV 2018. Cham: Springer: 833-851 [DOI: 10.1007/978-3-030-01234-2_49]

Cheng J, Deng C J, Su Y Z, An Z Y and Wang Q. 2024. Methods and datasets on semantic segmentation for unmanned aerial vehicle remote sensing images: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 211: 1-34 [DOI: 10.1016/j. isprsjprs.2024.03.012]

Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, et al. 2016. The Cityscapes dataset for semantic urban scene understanding//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE: 3213-3223 [DOI: 10.1109/CVPR.2016.350]

Dong B, Wang P and Wang F. 2023. Head-free lightweight semantic segmentation with linear transformer. Proceedings of the AAAI Conference on Artificial Intelligence, 37(1): 516-524 [DOI: 10.1609/aaai.v37i1.25126]

Dong Y, Mao C, Zheng L and Wu Q. 2025. DMANet: Dual-branch multiscale attention network for real-time semantic segmentation. Neurocomputing, 617: 128991 [DOI: 10.1016/j. neucom. 2024.128991]

Fan J, Wang F, Chu H, Hu X, Cheng Y and Gao B. 2023. MLFNet: Multi-level fusion network for real-time semantic segmentation of autonomous driving. IEEE Transactions on Intelligent Vehicles, 8(1): 756-767 [DOI: 10.1109/TIV.2022.3176860]

Fan M, Lai S, Huang J, Wei X, Chai Z, Luo J, et al. 2021. Rethinking BiSeNet for real-time semantic segmentation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE: 9711-9720 [DOI: 10.1109/CVPR46437.2021.00959]

Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, et al. 2019. Dual attention network for scene segmentation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE: 3141-3149 [DOI: 10.1109/CVPR.2019.00326]

Gao R. 2023. Rethinking dilated convolution for real-time semantic segmentation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Vancouver: IEEE: 4675-4684 [DOI: 10.1109/CVPRW59228.2023.00493]

Guo M H, Lu C Z, Hou Q, Liu Z N, Cheng M M and Hu S M. 2022. SegNeXt: Rethinking convolutional attention design for semantic segmentation. Advances in Neural Information Processing Systems, 35: 1140-1156 [DOI: 10.52202/068431-0084]

Hou Z Q, Qu M J, Li J G, Ma S G, Wang Y C and Yang X B. 2025. Lightweight CNN-Transformer combined network for real-time semantic segmentation. Journal of Image and Graphics, 30(7): 2437-2450 (侯志强, 屈敏杰, 李俊歌, 马素刚, 王昀琛, 杨小宝. 2025. 轻量级 CNN-Transformer 相结合的实时语义分割网络. 中国图象图形学报, 30(7): 2437-2450) [DOI: 10.11834/jig.240527]

Li X, Li X, Zhang L, Cheng G, Shi J, Lin Z, et al. 2020. Improving semantic segmentation via decoupled body and edge supervision//Computer Vision - ECCV 2020. Cham: Springer: 435-452 [DOI: 10.1007/978-3-030-58520-4_26]

Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]

Lu M, Chen Z, Liu C, Ma S, Cai L and Qin H. 2022. MFNet: Multi-feature fusion network for real-time semantic segmentation in road scenes. IEEE Transactions on Intelligent Transportation Systems, 23(11): 20991-21003 [DOI: 10.1109/TITS.2022.3182311]

Pan H, Hong Y, Sun W and Jia Y. 2023. Deep dual-resolution networks for real-time and accurate semantic segmentation of traffic scenes. IEEE Transactions on Intelligent Transportation Systems, 24(3): 3448-3460 [DOI: 10.1109/TITS.2022.3228042]

Peng X, Cheng J, Tang X, Deng Z, Tu W and Xiong N. 2024. HSNet: An intelligent hierarchical semantic-aware network system for real-time semantic segmentation. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 54(7): 4318-4330 [DOI: 10.1109/TSMC.2024.3377280]

Wang J, Gou C, Wu Q, Feng H, Han J, Ding E, et al. 2022. RTFormer: Efficient design for real-time semantic segmentation

© 中国图象图形学报版权所有

with transformer. Advances in Neural Information Processing Systems, 35: 7423-7436 [DOI: 10.52202/068431-0539]

Wang Y, Zhou Q, Liu J, Xiong J, Gao G, Wu X, et al. 2019. LEDNet: A lightweight encoder-decoder network for real-time semantic segmentation//2019 IEEE International Conference on Image Processing (ICIP). Taipei: IEEE: 1860-1864 [DOI: 10.1109/ICIP.2019.8803154]

Xiong J, Po L M, Yu W Y, Zhou C, Xian P and Ou W. 2023. CSRNet: Cascaded selective resolution network for real-time semantic segmentation. Expert Systems with Applications, 211: 118537 [DOI: 10.1016/j.eswa.2022.118537]

Xu J, Xiong Z and Bhattacharyya S P. 2023. PIDNet: A real-time semantic segmentation network inspired by PID controllers//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE: 19529-19539 [DOI: 10.1109/CVPR52729.2023.01871]

Yu C, Gao C, Wang J, Yu G, Shen C and Sang N. 2021. BiSeNet V2: Bilateral network with guided aggregation for real-time semantic segmentation. International Journal of Computer Vision, 129(11): 3051-3068 [DOI: 10.1007/s11263-021-01515-2]

Yu C, Wang J, Peng C, Gao C, Yu G and Sang N. 2018. BiSeNet: Bilateral segmentation network for real-time semantic segmentation//Computer Vision – ECCV 2018. Cham: Springer: 334-349 [DOI: 10.1007/978-3-030-01261-8_20]

Yu C, Xiao B, Gao C, Yuan L, Zhang L, Sang N, et al. 2021. Lite-HRNet: A lightweight high-resolution network//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE: 10435-10445 [DOI: 10.1109/CVPR46437.2021.01030]

Zhang Z, Xu Z, Gu X and Xiong J. 2023. Cross-CBAM: A lightweight network for scene segmentation[EB/OL]. [2023-06-01]. https://arxiv. org/abs/2306.02306 [DOI: 10.48550/arXiv. 2306.02306]

Zhao H, Qi X, Shen X, Shi J and Jia J. 2018. ICNet for real-time semantic segmentation on high-resolution images//Computer Vision – ECCV 2018. Cham: Springer: 418-434 [DOI: 10.1007/978-3-030-01219-9_25]

Zhao H, Shi J, Qi X, Wang X and Jia J. 2017. Pyramid scene parsing network//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 6230-6239 [DOI: 10.1109/CVPR.2017.660]

Zhao J, Zhao W, Deng B, Wang Z, Zhang F, Zheng W, et al. 2024. Autonomous driving system: A comprehensive survey. Expert Systems with Applications, 242: 122836 [DOI: 10.1016/j.eswa.2023.122836]

作者简介

侯志强，男，教授，研究方向为计算机视觉、目标跟踪等。E-mail: hou-zhq@sohu.com

李俊歌，女，硕士生，研究方向为计算机视觉、语义分割。E-mail:ljg163yx@163.com

马素刚，男，教授，研究方向为计算机视觉、机器学习等。E-mail: msg@xupt.edu.cn

刘洋，男，讲师，研究方向为智能优化、进化计算、计算机视觉。E-mail: yliu@xupt.edu.cn

刘璐，女，高级工程师，研究方向为计算机科学与应用。E-mail: alice_1213@xupt.edu.cn

余旺盛，男，副教授，研究方向为图像处理、视觉跟踪。E-mail: xing_fu_yu@sina.com

© 中国图象图形学报版权所有