ISSN 1003-501X CN 51-1346/O4

光电工程

Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

基于BiLevelNet的实时语义分割算法

吴马靖，张永爱，林珊玲，林志贤，林坚普

引用本文:

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.

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.

https://doi.org/10.12086/oee.2024.240030

收稿日期: 2024-01-30; 修改日期: 2024-03-13; 录用日期: 2024-03-13

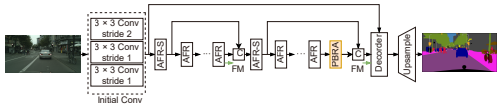
相关论文

Opto-Electronic Engineering
光 电 工 程

Article
2024 年, 第 51 卷, 第 5 期

DOI: 10.12086/oee.2024.240030

基于 BiLevelNet 的实时语义分割算法



吴马靖[1]，张永爱[1,2]，林珊玲[1,2]，林志贤[1,2]，林坚普[1,2*]

[1]福州大学 先进制造学院，福建 泉州 362200；
[2]中国福建光电信息科学与技术创新实验室，福建 福州 350116

摘要：针对语义分割网络参数量过大导致其难以部署在内存受限的边缘设备等问题，本文提出一种基于 BiLevelNet 的轻量级实时语义分割算法。首先，利用空洞卷积扩大感受野，并结合特征复用策略增强网络的区域感知能力。接着，嵌入两阶段的 PBRA 注意力机制，建立远距离相关物体之间的依赖关系以增强网络的全局感知能力。最后，引入结合浅层特征的 FADE 算子以改善图像上采样效果。实验结果表明，在输入图像分辨率为 512×1024 的情况下，本文网络在 Cityscapes 数据集上以 121 f/s 的速率获得了 75.1% 的平均交并比，模型大小仅为 0.7 M。同时在输入图像分辨率为 360×480 的情况下，在 Camvid 数据集上取得 68.2% 的平均交并比。同当前其他实时语义分割方法相比，该网络性能取得速度与精度的均衡，符合自动驾驶应用场景对实时性的要求。

关键词：实时语义分割；自动驾驶；深度学习；自注意力；上采样

中图分类号：TP394.1; TH691.9　　**文献标志码：**A

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
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

Real-time semantic segmentation algorithm based on BiLevelNet

Wu Majing[1], Zhang Yong'ai[1,2], Lin Shanling[1,2], Lin Zhixian[1,2], Lin Jianpu[1,2*]

[1]School of Advanced Manufacturing, Fuzhou University, Quanzhou, Fujian 362200, China;
[2]Fujian Science & Technology Innovation Laboratory for Optoelectronic Information of China, Fuzhou, Fujian 350116, China

Abstract: In response to the problem of the large parameter size of semantic segmentation networks, making it difficult to deploy on memory-constrained edge devices, a lightweight real-time semantic segmentation algorithm is proposed based on BiLevelNet. Firstly, dilated convolutions are employed to augment the receptive field, and feature reuse strategies are integrated to enhance the network's region awareness. Next, a two-stage PBRA (Partial Bi-Level Route Attention) mechanism is incorporated to establish dependencies between distant objects, thereby augmenting the network's global perception capability. Finally, the FADE operator is introduced to combine shallow features to improve the effectiveness of image upsampling. Experimental results show that, at an input

收稿日期：2024-01-30；修回日期：2024-03-13；录用日期：2024-03-13
基金项目：国家重点研发计划资助项目 (2023YFB3609400)；福建省自然科学基金资助项目 (2020J01468)；国家自然科学基金青年科学基金资助项目 (62101132)
*通信作者：林坚普，ljp@fzu.edu.cn。

image resolution of 512×1024, the proposed network achieves an average Intersection over Union (IoU) of 75.1% on the Cityscapes dataset at a speed of 121 frames per second, with a model size of only 0.7 M. Additionally, at an input image resolution of 360×480, the network achieves an average IoU of 68.2% on the CamVid dataset. Compared with other real-time semantic segmentation methods, this network achieves a balance between speed and accuracy, meeting the real-time requirements for applications like autonomous driving.

Keywords: real-time semantic segmentation; autonomous driving; deep learning; self-attention; upsampling

1 引　言

语义分割作为计算机视觉领域的基础研究任务，因其具有逐像素分类能力从而在自动驾驶[1]、医学图像[2]、遥感[3]等领域得到广泛应用。作为像素级分类任务，其目标是为图像中的每个像素赋予相应的类别标签，从而实现对图像中不同物体或区域的精细划分和识别。在以自动驾驶为代表的应用中，语义分割可以提供强大的场景解析能力，并为视觉理解提供细粒度和深层的语义信息。

语义分割模型通常采用 ImageNet 预训练骨干网络提取对象的局部信息，并使用上下文模块挖掘语义信息，例如 PPM (pyramid pooling module)[4]、或 ASPP (atrous spatial pyramid pooling)[5]。然而，Andrew 等[6]指出减半最后一个卷积层的通道数量不会降低分割精度，从而表明 ImageNet 骨干在分割任务存在通道冗余的现象。此外，ImageNet 数据集中 224×224 的尺寸远小于语义分割的图像。例如，Cityscapes[7] 的图像分辨率为 1024×2048，而 CamVid[8] 的图像分辨率为 720×960，ImageNet 模型处理大尺寸图片时面临感受野不足的问题，并且存在参数规模较大和推理速度慢等问题。因此需要构建一个轻量级的语义分割网络，以满足自动驾驶场景的实际需求。

当前的轻量级语义分割模型，大致分为双分支架构[9]和基于多次下采样方法[10]。这两类方法在权衡精度和速度方面取得不错的表现，但双分支架构在初始下采样阶段拆分支路会阻碍信息交互，重复提取浅层特征导致参数冗余。多次下采样会逐渐丢失空间信息，在上采样阶段难以恢复缺失的空间信息，从而导致分割性能下降。

相比于上述两种结构，基于轻量化特征提取模块的编码—解码结构[11-14]更适用于资源受限场景。这种结构通过多次降低分辨率，大幅减少网络训练和推理中的信息流参数。并结合分解卷积、深度可分离卷积[12]、划分—重排操作[14]配置基础特征提取单元。然而，由于较小的感受野限制，这种模型无法获取更多的上下文信息，从而限制了分割精度。虽然使用空洞卷积在一定程度上缓解了这个问题，但堆叠空洞卷积会导致棋盘效应，进而导致局部信息丢失，这对于像素级密集预测任务难以接受。

为了解决上述问题，本文提出一种基于两阶段路由自注意力的实时分割网络 (bi-level attention real-time semantic segmentation network, BiLevelNet)，主要工作如下：

1) 为了实现高效的特征提取，基于多层次特征提取策略设计了非对称特征复用卷积模块 (asymmetric feature reuse convolution module，AFR module)，并进一步搭建出整体的轻量架构。

2) 鉴于自动驾驶场景物体的尺寸存在显著变化，本文采用两阶段路由自注意力模块 (bi-level route attention module, BRA module)，用于捕捉远距离物体间的依赖关系，从而增强特征表达能力。此外，结合通道缩减的思想，减少冗余特征的影响。通过实验得到最佳通道缩减因子。

3) 针对上采样阶段难以恢复空间信息的问题，本文采用了 FADE 上采样模块 (fuses the assets of decoder and encoder upsampling operator，FADE)，该方法在保持轻量化的同时，同时参考空间信息和语义信息进行上采样，从而完成图像空间细节的恢复。

4) 在 Cityscapes 和 Camvid 数据集上验证本文算法的有效性。实验结果显示，mIoU 分别达到 75.1% 和 68.2%。当输入图像大小为 512×1024 时，实现了 121 帧/秒的速度。与当前轻量级算法相比，本文方法取得了更好的分割性能。

2 相关工作

2.1 轻量级特征提取模块

RELAXNet[13] 和 LEDNet[15] 展示了通道分离和通道打乱在压缩网络参数的积极影响，认为其有利于促

进通道间的信息交互。DWGNet[16] 认为随着特征语义表示的增强，更高的阶段需要更大的感受野，以及恰当空洞率的膨胀卷积有助于建立长距离特征之间的连接。Fasternet[17] 在最小卷积单元中只对部分通道进行特征提取，同时保留其余通道的完整信息，提出了应对 shufflenetV2[18] 中论述的通道冗余现象的一种解决思路。综上，本文结合空洞卷积和特征复用策略，将初步提取的特征进行复用，并通过单元级的长短残差连接，进一步增强对图像不同区域的感知能力。

2.2 注意力机制

注意力机制在提高分割准确性方面至关重要。CBAM[19] 简单结合通道与空间注意力的做法，无法捕捉远距离目标的依赖关系。从通道层面切入，文献 [20] 考虑到每个类别与输入特征逐通道之间的关系，从而增强类别信息感知能力。从稀疏性角度切入，Huang 等[21] 提出一种交叉注意力模块，对局部窗口的注意力计算只考虑十字路径，以稀疏地建模像素间的长距离依赖关系，相比逐像素计算的自注意力机制更加高效和轻量。池化注意力[22] 通过条带形状的池化窗进一步简化了计算。可变形注意力[23] 通过不规则网格实现图像自适应稀疏性，但引入了与内容无关的序列向量。为此，本文采用具备内容感知和稀疏性的两阶段自注意力模块[24]，以较小代价有效地捕获远距离对象之间的关系。

2.3 上采样算子

在解码器中，通过临近插值法或双线性插值法对特征图进行上采样操作是一种常见做法。前者速度快但效果较差，后者在图像边缘过渡平滑导致局部高频信息丢失。基于反卷积的上采样忽略了低层的内容，无法适应局部信息的变化。为解决这些问题，CARAFE[25] 上采样算子被提出，在较大的感知领域内聚合上下文信息，同时实时感知特定的内容，并保持计算效率。文献 [26] 考虑到空间细节信息的影响，提出编码指导上采样模块，但文中细节块的设计未能充分提取空间信息。Lu 等人观察到使用解码器深层特征可以增强区域连续性，而使用编码器浅层特征有助于恢复细节，并结合这两种优点提出通用的 FADE 算子[27]。FADE 算子在保存语义信息的同时补偿由于下采样导致的细节信息丢失，在区域和细节敏感的密集预测任务上表现良好。因此，在本文中，FADE 算子被引入解码器，用于接收编码器中的区域信息和解码器中的语义信息，完成图像空间细节的恢复。

3 本文模型及网络结构

3.1 网络结构设计

本文目的是设计具备快速推理与高精度的轻量化网络。基于编码器-解码器架构，设计出了一个实时分割网络。目前，大多数方法为了获取更深层次的语义信息，对输入图片进行了多次下采样以扩大卷积层的有效感受野，在恢复至原始分辨率时容易造成空间信息丢失。本文网络设计细节如表 1 所示，整体架构如图 1 所示。

表 1 本文网络框架

Table 1 Network framework of BiLevelNet

Stage	Operator	Mode	Output size
Stage 1	3 × 3 Conv	Stride 2	32 × 256 × 512
	3 × 3 Conv	Stride 1	32 × 256 × 512
	3 × 3 Conv	Stride 1	32 × 256 × 512
Stage 2	AFR-S		64 × 128 × 256
	2 × ARF	Dilated 2	64 × 128 × 256
Stage 3	AFR-S		128 × 64 × 128
	4 × AFR	Dilated 4	128 × 64 × 128
	5 × AFR	Dilated 8	128 × 64 × 128
Decoder	DAF		32 × 256 × 512
	1 × 1 Conv	Stride 1	19 × 256 × 512
	Bilinear		19 × 512 × 1024

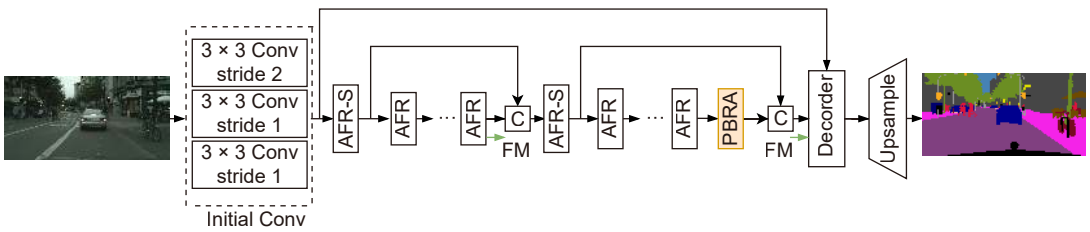


图 1 BiLevelNet 的网络框架

Fig. 1 Network framework of BiLevelNet

该网络的编码部分可分为三个阶段。首先使用三个级联的卷积提取初始特征，其中步长为 2 的 3×3 卷积用于缩小输入图像的尺寸。步长为 1 的 3×3 卷积用于提取初始特征。本文采用步长为 2 的 3×3 卷积缩小输入特征尺寸，相比结合了最大池化和步长为 2 的卷积的常规做法，可以捕获更多的局部信息，同时减少细节信息丢失。

对输入图像送入 AFR-S 进行通道变换和下采样处理之后，将信息传递至第二阶段。由 2 个堆叠的 AFR 模块组成，空洞率设置为 2，可有效地提取浅层局部信息。将第二阶段的输入拼接到输出端，作为第三阶段的输入。在第三阶段，为扩大网络的感受野，堆叠 4 个空洞率为 4 和 5 个空洞率为 8 的 AFR 模块，同样使用残差将输入拼接到输出端。

最后，为确保网络各阶段的语义一致性，将三个阶段的输出特征图送入解码器进行特征融合，完成原始分辨率的恢复。

3.2 非对称特征复用模块

其中，在模块设计过程中，本文参考了瓶颈残差结构 (图 2(a) 与非瓶颈残差结构 (图 2(b))，扩大感受野对于提升上下文信息获取能力至关重要，采用恰当膨胀率的空洞卷积以增强网络的上下文感知能力是一种常见做法。故本文结合 ERFNet[11] 中的分解卷积和 DFANet[28] 中特征复用的思想，提出了非对称特征复用单元，作为网络的基础特征提取单元，如图 3 所示。

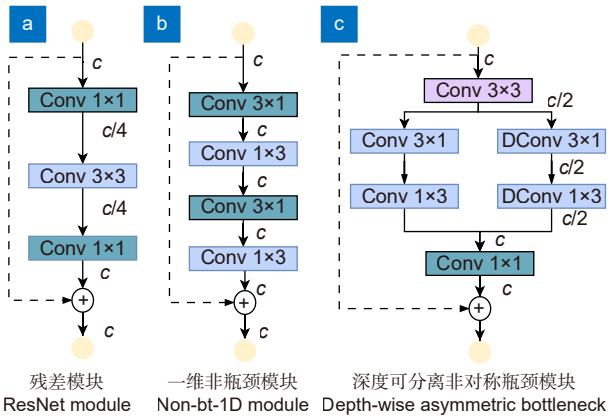


图 2 不同特征提取模块的比较

Fig. 2 Comparison of different feature extraction modules

本文提出 AFR 模块，使用特征复用策略有效实现了多层次特征提取。首先，对输入特征 X 进行 1×1 卷积操作实现降维并促进通道间的信息交互。接下来，本文设计了一个双分支卷积支路来捕捉更多的上下文信息。其中，局部特征通过 3×3 的深度可分离卷积获取。另一支路在不同阶段使用不同膨胀率的空洞卷积来获取候选区域周围的上下文信息。为了降低计算成本并保证实时推理速度，该分支同样采用深度可分离卷积实现。将第一个双支路卷积的两个支路的输出特征相加后与输入特征进行拼接，得到初步提取的特征。

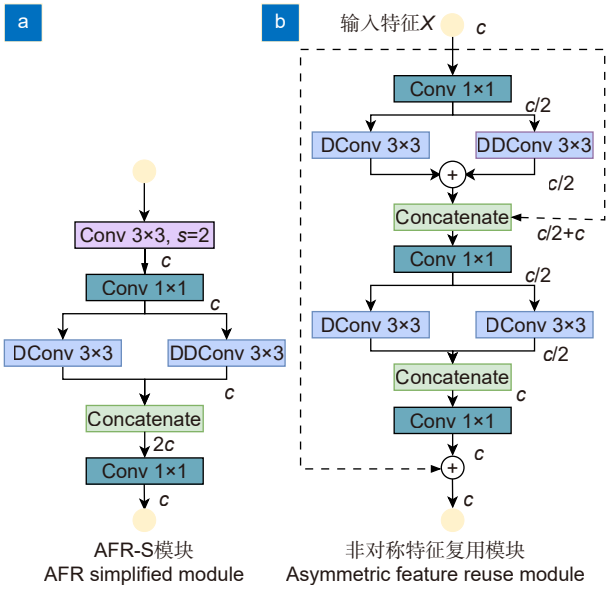


图 3 AFR-S 模块和 AFR 模块

Fig. 3 AFR-S module and AFR module

为了进一步提取更复杂的语义特征，将初步提取的特征重新送入 1×1 卷积和双支路卷积层，进一步增强特征表达能力并提取更丰富的语义信息。最后，使用 1×1 卷积来增强通道间的信息交互，再使用长距离残差与输入特征图相加以恢复有用的丢失信息并加快训练过程，构建出更全面的特征表达。

图 4 展示了 AFR 模块中不同层级的特征图。图 4(b) 为输入特征，整体语境的响应较好，但在方框处杆子和道路边界的响应较弱。通过 3×3 深度卷积获取的局部特征 4(d) 细节信息丰富，如道路边界和车辆轮廓。而另一分支通过空洞卷积获取的上下文信息 4(e) 更侧重语义信息，在道路上有连续和完整的激活。通过双分支相加融合的特征在 4(f) 中，对局部和整体场景均展现出较好的激活。

将图 4(f) 与输入特征 4(b) 相拼接后得到 4(g)，由于输入特征通道数是 4(f) 的两倍，初步观察 4(g) 和原始特征似乎有一致的响应。然而，经过第二次 1×1 卷积得到的特征 4(h) 中，展现了完整的道路响应，同时局部杆子的特征也未丢失，且无噪声干扰。

与对图 4(c) 进行单次特征提取相比，复用特征

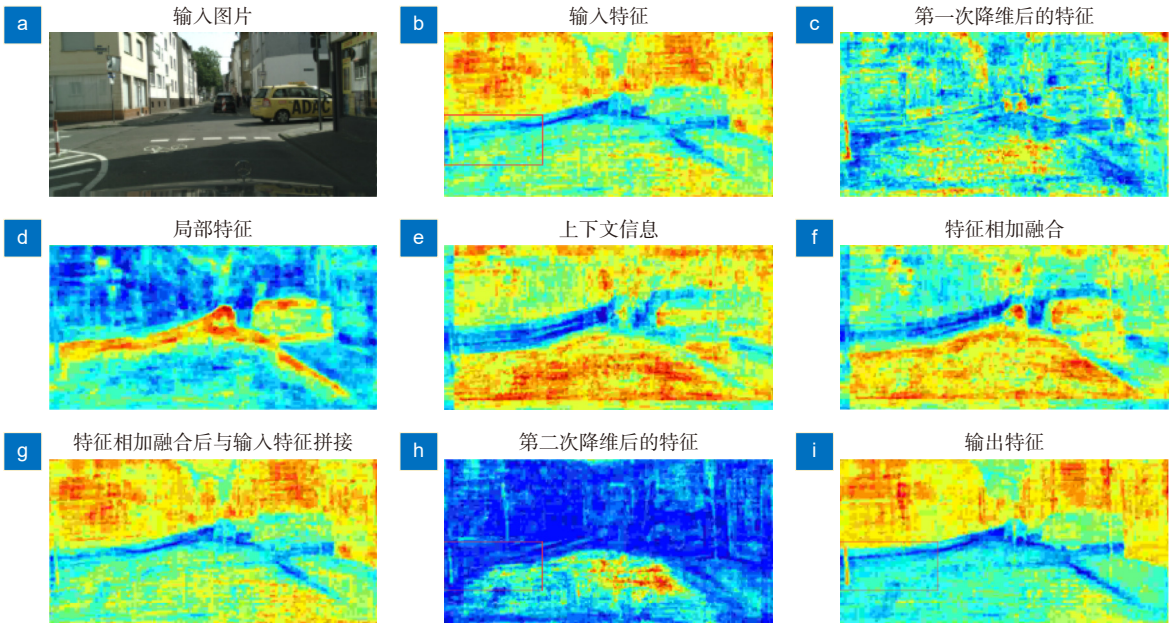


图 4　区域感知与特征复用示意图

Fig. 4　Schematic diagram of region perception and feature reuse

4(h) 有助于强化模型对细节和整体场景的理解，同时减少背景干扰，这一点在输出特征 4(i) 杆子和道路边界处更连续的响应中得到验证。

3.3 基于通道缩减的两阶段自注意力模块

在自动驾驶场景中，由于物体尺度变化较大，为了提升网络对不同尺度物体信息的获取能力，本文引入了注意力机制。基于卷积的注意力模块感受野较小，难以建立长距离依赖关系。自注意力具有全局感受野，但却带来高额计算量。为了克服这些问题，本文采用一种具有动态稀疏特性的两阶段路由自注意力模块，称为 BRA 模块。

BRA 模块在预先划分的窗口中过滤掉大部分不相关的键值对，只保留少量的相关区域。对于输入的特征图，通过线性映射获得 QKV 序列。之后通过邻阶矩阵构建有向图得到不同键值对之间的相关度，可以理解为每个区域与其他区域之间的相关程度。最后，基于区域级的路径查询矩阵，应用细粒度的自注意力计算得到权重再分配后的特征图。其结构如图 5 所示。

首先将一张二维的特征图$x \in R^{H\times W\times C}$分割成$S\times S$个大小为$H/S$个互不重叠的窗口，得到$x^r \in R^{S^2\times HW/S^2\times C}$，其中 C 为原特征图的通道，H 和 W 分别为特征图的高度和宽度。随后进行线性投影操作，得到 $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V} \in R^{S^2\times HW/S^2\times C}$序列向量。

然后，通过构建一个有向图确定每个给定区域应该关注的其他区域 (即参与细粒度注意力计算的区域)。

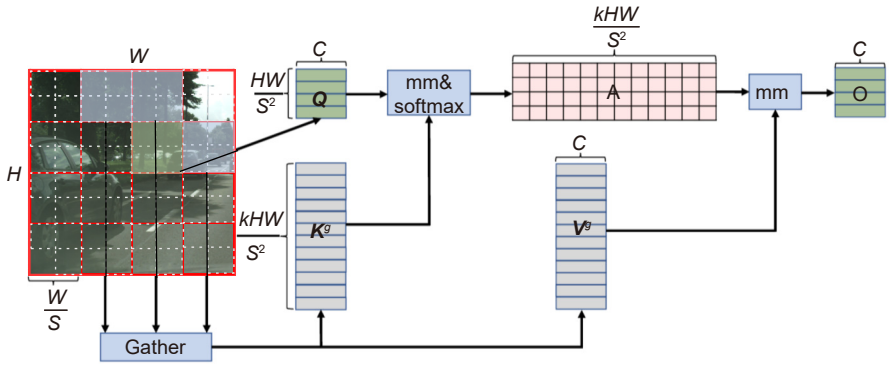


图 5　BRA 模块

Fig. 5　BRA module

首先分别对 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 进行区域级的均值操作得到 $\boldsymbol{Q}^{\mathrm{r}},\boldsymbol{K}^{\mathrm{r}}\in R^{S^2\times C}$，其作用是降低计算量。然后通过将$\boldsymbol{Q}^{\mathrm{r}}$和转置的$\boldsymbol{K}^{\mathrm{r}}$进行矩阵乘法，得到区域之间关联图的邻接矩阵$\boldsymbol{A}^{\mathrm{r}}=\boldsymbol{Q}^{\mathrm{r}}(\boldsymbol{K}^{\mathrm{r}})^{\mathrm{T}}$，$\boldsymbol{A}^{\mathrm{r}}\in R^{S^2\times S^2}$衡量了两个区域之间的语义相关性。为了保留对每个区域最重要的前 k 个区域，对$\boldsymbol{A}^{\mathrm{r}}$逐行计算$top-k$区域并得到相关矩阵 $\boldsymbol{I}\in R^{S^2\times k}$，得到前 k 个最相关的$\boldsymbol{KQ}$对，公式如下：

$$\boldsymbol{I}^{\mathrm{r}}=topkIndex(\boldsymbol{A}^{\mathrm{r}}), \tag{1}$$

其中：第 i 行包含了与第 i 个区域最相关的 k 个区域的索引。

使用区域间的索引矩阵$\boldsymbol{I}^{\mathrm{r}}$，可以指定区域 i 中的每个查询向量 $\boldsymbol{Q}$ 只关注索引$\boldsymbol{I}^{\mathrm{r}}_{(i,1)},\boldsymbol{I}^{\mathrm{r}}_{(i,2)},\ldots,\boldsymbol{I}^{\mathrm{r}}_{(i,k)}$所对应的 k 个区域中的键值对。使用式 (2) 收集这些分散在整个特征图上的区域：

$$\boldsymbol{K}^{\mathrm{g}}=gather(\boldsymbol{K},\boldsymbol{I}^{\mathrm{r}}),\boldsymbol{V}^{\mathrm{g}}=gather(\boldsymbol{V},\boldsymbol{I}^{\mathrm{r}}), \tag{2}$$

其中：$\boldsymbol{Q}$，$\boldsymbol{K}$，$\boldsymbol{V}\in R^{S^2\times kHW/S^2\times C}$代表收集到的键值对。使用式 (3) 计算得到注意力权重信息：

$$Att=softmax(\boldsymbol{Q}\otimes\boldsymbol{K}^{\mathrm{g}})\otimes\boldsymbol{V}^{\mathrm{g}}, \tag{3}$$

其中：通过 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 之间的矩阵乘法计算得到注意权重，而没有引入额外的可学习参数，节省了内存和计算资源。

在此基础上，本文结合 Fasternet 中通道缩减的思想，对送入 BRA 模块的特征通道进行调控，形成了 PBRA 模块，这一设计避免通道数较多时受到冗余特征的影响，有助于均衡重要区域和非重要区域的注意力权重。PBRA 模块如图 6 所示。

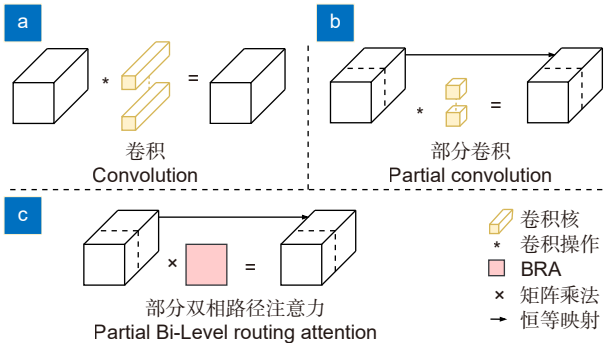


图 6 PBRA 模块

Fig. 6 Partial Bi-Level route attention module

3.4 跨尺度特征聚合

常用的双线性插值在上采样时存在难以恢复丢失的图像细节信息的问题，为此本文在解码器中引入 FADE 上采样算子，能够有效地增强图像的空间信息。

其详细结构如图 7 所示。首先，特征图F_{L}和F_{M}分别输入，经过 Refine 模块进行特征精炼和通道对齐。该模块串联了 3×3 卷积和 1×1 卷积。随后，经过处理的特征进行求和，再次输入 Refine 模块以进一步对齐信息，生成特征F_{1}。接着，特征图F_{M}和F_{H}通过 FADE 进行上采样得到特征F_2，与F_{1}一同输入 DAF 模块得到输出 Y，如式 (4) 所示。

$$Y=DAF(Refine(Refine(F_{\mathrm{L}})+Refine(F_{\mathrm{M}}))). \tag{4}$$

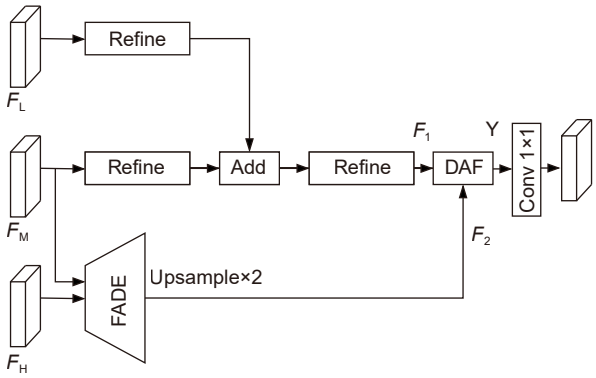


图 7 解码器模块

Fig. 7 Decorder module

4 实验与结果分析

4.1 数据集

本文采用 Cityscapes 数据集和 CamVid 数据集验证算法的有效性。Cityscapes 常用于语义分割网络的评估。本文仅使用其中的 5000 张精细标注图像，分辨率为 1024×2048，标签分为 19 个类别，包括道路、行人、车辆、建筑物等，覆盖了城市场景中可能出现的各种对象和区域。包含 2975 张训练图像、500 张验证图像和 1525 张测试图像。CamVid 是采用汽车驾驶的视角拍摄的道路场景数据集。所有图像的分辨率为 960×720，语义标签分为 11 个类别。该数据集一共有 701 张图像，分别包含 367 张训练集合图像，101 张验证图像和 233 张测试图像。

4.2 评价指标

本文使用平均交并比 (mIoU) 作为分割精确度衡量指标，如式 (5) 所示:

$$mIoU=\frac{1}{1+k}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}}, \tag{5}$$

其中：p_{ii},p_{ij},p_{ji}分别表示预测正确的像素数量，将 i 类预测为 j 类的像素数量，以及将 j 类预测为 i 类的

概率的像素数量。k 表示类别数。分割速度采用每秒处理帧数 (frames per second, FPS)，内存消耗则以参数量 (parameters, Params) 来衡量，而计算复杂度则使用每秒浮点运算次数 (floating-point operations per second, FLOPs) 来衡量。

4.3 实验设置

对于 Cityscapes 数据集，实验采用随机梯度下降法 (SGD)，动量设置为 0.9，权重衰减为 2e-4，使用“poly”作为学习率衰减策略，其中将初始学习率设置为 4.5e-2，指数系数为 0.9，最大迭代次数设置为 1000 个 epoch，批量大小设置为 8。对于 Camvid 数据集，实验使用 Adam 算法进行优化，初始学习率为 1.5e-3，权重衰减为 2e-4。学习率使用热身策略 (warm-up)。热身因子、热身迭代次数和指数系数分别设置为 1/3、500 和 0.9。最大迭代次数为 1000 个 epoch，批量大小设置为 8。

数据增强方面，对输入图像进行随机水平翻转、均值减法和随机缩放。随机缩放比例包含{0.75、1.0、1.25、1.5、1.75、2.0}。最后，对图像随机裁剪为固定大小以进行训练。

所有的实验都是基于 PyTorch-1.10,CUDA 11.3, CUDNN 8.2.0 进行。使用单个 NVIDIA RTX 3090 GPU 进行训练和推理。数据集类别像素数量统计如图 8 所示，存在严重的类别分布不均问题。

对于 Cityscapes 数据集，采用 OHEM (online hard example mining) 损失函数进行优化，本文采用的交叉熵损失:

$$loss=-\sum_{i=1}^{m}\sum_{j=1}^{n}\hat{y}_{i,j}\lg\frac{\exp(x_{i,j})}{\sum_{i=1}^{m}\sum_{j=1}^{n}\exp(x_{i,j})}, \tag{6}$$

其中：$x_{i,j}$表示$m\times n$大小的图像第i,j个像素值，$\hat{y}_{i,j}$表示模型在输入$x_{i,j}$为时的预测值。

OHEM 通过在损失计算后，选取损失超过阈值 thresh 的像素作为难负样本引入训练，以便网络更加关注难以学习的样本，从而增强分割精度。

对于 CamVid 数据集，采用类别加权策略：

$$\begin{cases} W_{\text{class}}=\dfrac{1}{\ln(c+P_{\text{class}})} \\ loss=-\sum\limits_{i=1}^{i}\sum\limits_{j=1}^{j}W_{\text{class}}\hat{y}_{i,j}\lg\dfrac{\exp(x_{i,j})}{\sum\limits_{i=1}^{i}\sum\limits_{j=1}^{j}\exp(x_{i,j})} \end{cases}, \tag{7}$$

其中：W_{class}指某个类别的权重，P_{class}指该类别样本分布。c 为可超参数，设置为 1.10。

4.4 消融实验

4.4.1 不同特征提取模块的消融实验

为验证 AFR 模块有效性，分别采用 DABNet 中的 DAB 模块、LEDNet 的 SSnbt 模块以及本文中的 AFR 模块构建网络。各网络性能对比如表 2 所示，可见由 DAB 模块构成的网络取得最佳推理速度，但 mIoU 相比本文模型从 75.5% 下降至 71.8%。使用 SSnbt 模块替换 AFR 模块后，精度下降了 8.4%，同时参数量提升了 0.15 M。

4.4.2 PBRA 模块的缩减因子 r 的消融实验

当输入通道的数量较大时，模型需要处理更多的信息，其中可能包含一些冗余信息，从而降低模型的表达能力和准确性。当减小输入通道的数量时，模型可以更加集中地学习每个通道所包含的信息，并更好区分有用的特征和噪声或无用的信息。但通道数量过小，可能会导致信息丢失或信息瓶颈，因此需要在减小通道数量和保持足够信息量之间找到平衡。

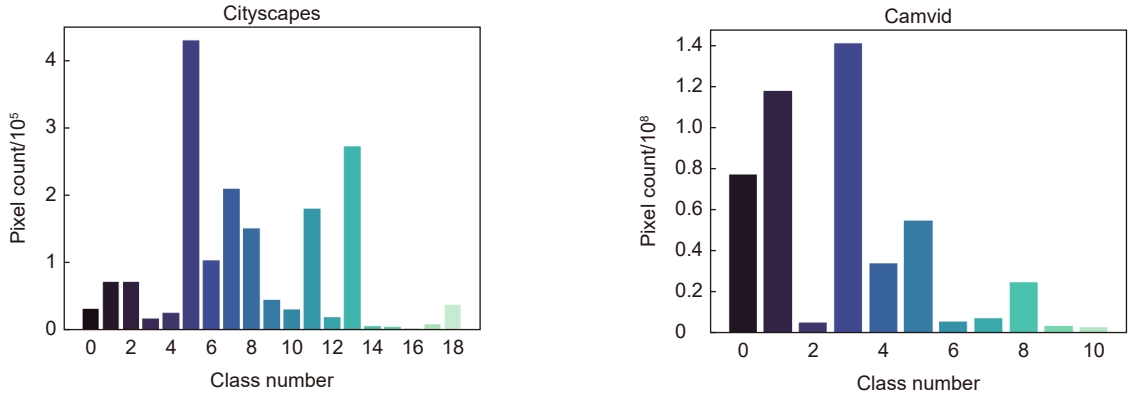


图 8 Cityscapes 数据集与 Camvid 数据集的样本分布

Fig. 8 Sample distribution of Cityscapes dataset and Camvid dataset

表 2 不同特征提取模块在 Cityscapes 数据集的性能对比

Table 2 Performance comparison of different feature extraction modules on the Cityscapes dataset

	Params/M	FLOPs/G	FPS	mIoU/%
SSnbt	0.83	11.61	132	67.1
DAB	0.75	10.78	140	71.8
AFR	0.68	9.64	128	**75.5**

本实验采用逐步减少的通道缩减因子，具体见表 3。其中 r=0 代表未采用 BRA 模块，而当 r=1/16，仅将极少数通道送入 BRA 模块，限制了其全局感知能力。另一方面，r=1 即所有通道送入 BRA 模块时，模型 mIoU 提升仅为 0.2 个百分点，原因有两点，一是通道中的冗余信息降低了对关键特征的有效集中，二是不同通道间较弱相关性的特征产生较弱的权重分布，限制了对于关键关系的建模能力。为克服上述问题，本文从通道调制入手进行特征选择，实验结果表明，当缩减因子 r=1/4 时，模型能更好地适应数据的特征分布，从而显著提升了性能，在参数量方面仅有 0.01 M 增加。因此，选取 r=1/4，即为本文所提的 PBRA 模块，相较于 BRA，PBRA 在 mIoU 上提升 1.5%，速度提升 12 f/s。

表 3 不同缩减因子在 Cityscapes 验证集的实验结果

Table 3 Experimental results of different reduction factor modules in Cityscapes validation set

Ratio	Params/M	FLOPs/G	FPS	mIoU/%
0	0.67	9.59	135	74.0
1	0.74	10.25	116	74.2
1/2	0.69	9.75	120	75.0
1/4	**0.68**	**9.64**	**128**	**75.5**
1/8	0.67	9.61	130	75.1
1/16	0.67	9.6	131	74.1

图 9 展示了引入 PBRA 模块前后的分割结果对比。结果显示，PBRA 提升了全局感知，对于基础模型在分割车辆时存在的类内不一致和边界外像素点误分类的现象得到改善。

4.4.3 FADE 上采样算子的消融实验

观察采用双线性插值进行上采样的分割结果 (如图 10)，注意到在道路边缘存在不连续和像素点缺失的问题。可视化双线性插值上采样前的深度特征图 (如图 11(a)，我们观察到道路和人行道的特征响应相似，容易被误判为同一类别。在图 11(c) 中，这一类别模糊现象更加直观。然而，在浅层特征图中，边缘信息得到了清晰的捕获，如图 11(b) 所示。因此，本

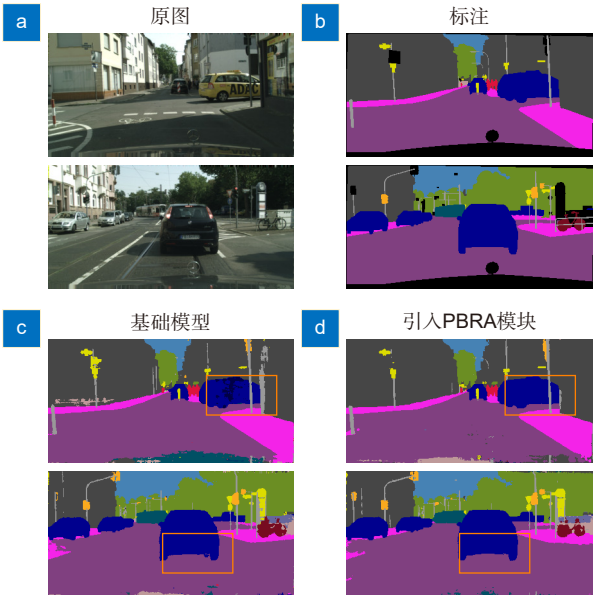


图 9 引入 PBRA 模块前后的分割结果对比

Fig. 9 Comparison of results with using the PBRA modules

文引入了边缘信息完整保留的浅层特征，将其融入 FADE 上采样时的参考内容，以有效弥补空间信息的丢失。这样的处理产生了更为连续和清晰的边缘分割效果，如图 12(a) 所示。在图 12(b) 中展示的浅层特征中的边缘信息与 FADE 上采样分割结果中的道路边界高度重合，从而验证了浅层特征能改善上采样时的边缘分割效果。

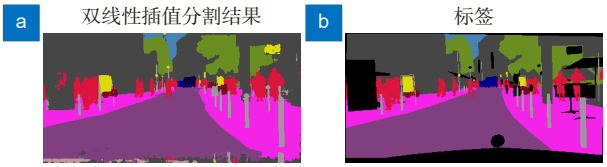


图 10 双线性插值分割结果对比

Fig. 10 Comparison of segmentation results using bilinear interpolation

从表 4 中可以看出，FADE 上采样算子有效提升了分割精度，并保持了轻量的参数和实时性能。与固定上采样策略的双线性插值相比，FADE 上采样算子在 mIoU 上取得了 0.4 个百分点的提升。验证了结合编码器中浅层信息可以高效聚合空间信息，提升上采样的可学习性，从而提高物体分割的准确度。

4.5 网络性能的定量对比

为验证本文方法的高效性，本节选取了现阶段几种优秀算法进行对比分析，实验结果如表 5。

结果表明，本文方法能较好地兼顾精度和推理速

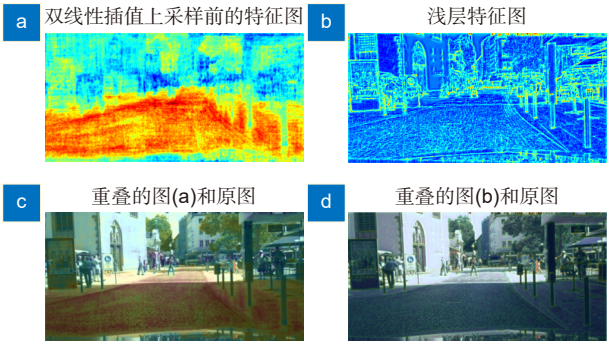


图 11 双线性插值前的特征图与浅层特征图

Fig. 11 Feature map before bilinear interpolation and the shallow feature map

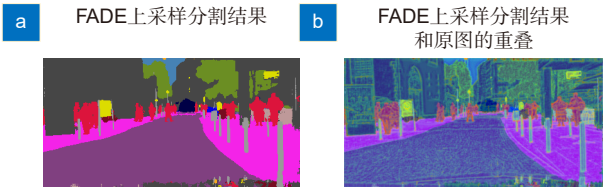


图 12 FADE 上采样分割结果

Fig. 12 FADE upsampling segmentation results

表 4 FADE 上采样算子在 Cityscapes 验证集的消融实验

Table 4 Experimental results of FADE modules on the Cityscapes validation dataset

	Params/M	FLOPs/G	FPS	mIoU/%
Bilinear	0.68	9.64	128	75.5
FADE	0.7	10.4	121	75.9

度。精度方面，本文方法获得了 75.1% 的平均交并比，高于所有的对比模型。推理速度方面，取得 121 f/s，符合自动驾驶中的实时性要求，仅低于 Bisenet-v2，并且参数量上只占 Bisenet-v2 的 1/8。参数量方面，虽然 ENet 和 DALNet 的参数量少于本文方法，但其对应的精度和推理速度明显低于本文方法。

表 6 进一步给出本文方法在 Cityscapes 测试集上的各类别交并比。由表中数据可知，大面积物体类别的分割准确率较高，例如建筑物、火车、公交车等分别达到了 91.8%、73.9%、78.8%。而对于其它类别，本文方法也取得了较好的分割精度，如道路、人行道、植被等分别达到 98.0%、82.2%、92.8%。主要归因于两点：首先，AFR 模块充分获取多尺度信息并复用深层特征。其次，PBRA 模块成功捕捉了场景中分散区域之间的远距离关联，权衡关键与次要特征间的注意力权重，协同提升分割结果的类内一致性。

表 5 不同模型在 Cityscapes 数据集的性能对比

Table 5 Performance comparison of different models on the Cityscapes dataset

Algorithm	Size	Params/M	FLOPs/G	FPS	mIoU/%
ENet	512×1024	**0.36**	4.35	42	58.3
ERFNet	512×1024	2.10	26.8	59	68.0
LEDNet	512×1024	0.94	11.5	71	69.2
DABNet	512×1024	0.76	-	104	70.1
ELANet[29]	512×1024	0.67	**9.7**	93	74.7
RELAXNet	512×1024	1.90	22.84	64	74.8
DALNet[30]	512×1 024	0.48	-	74	71.1
BiseNet-v2	512×1024	3.40	21.2	**156**	72.6
MIFNet[31]	512×1024	0.82	12.03	74	73.1
文献[32]	512×1024	6.22	12.5	154.7	74.2
Ours	512×1024	0.70	10.4	121	**75.1**

4.6 网络性能的定性对比

本文及其他网络在 Cityscapses 数据集上的分割效果对比如图 13 所示。

如图 13 所示的第一行和第二行，本文方法能准确辨别卡车和公交车类别区域内的所有像素点，而其他方法的结果倾向于将这类大物体分割成若干个区域。其主要原因在于大物体的尺寸可能超出了卷积操作所能捕捉到的局部范围，较小的滑动窗口无法充分利用相邻区域的上下文信息。而本文引入的 PBRA 模块通

表 6 不同模型在 Cityscapes 数据集上的各类别交并比

Table 6 Evaluation results of per-class IoU /% on the Cityscapes dataset

Class	ERFNet	DABNet	LEDNet	FDDWNet	Ours
Roa	97.9	96.8	97.1	**98.0**	**98.0**
Sid	82.1	78.5	78.6	82.4	**82.2**
Bui	90.7	90.9	90.4	91.1	**91.8**
Wal	45.2	45.3	46.5	52.5	**54.8**
Fen	50.4	50.1	48.1	51.2	**56.5**
Pol	59.0	59.1	60.9	59.9	**63.2**
Tli	62.6	65.2	60.4	64.4	**68.4**
TSi	68.4	70.7	71.1	68.9	**72.1**
Veg	91.9	92.5	91.2	92.5	**92.8**
Ter	69.4	68.1	60.0	70.3	**70.5**
Sky	94.2	**94.6**	93.2	94.4	94.5
Ped	78.5	80.5	74.3	80.8	**82.3**
Rid	59.8	58.5	51.8	59.8	**65.2**
Car	93.4	92.7	92.3	94.0	**94.3**
Tru	52.5	52.7	**61.0**	56.5	59.2
Bus	60.8	67.2	72.4	68.9	**78.5**
Tra	53.7	50.9	51.0	48.6	**73.9**
Mot	49.9	50.4	43.3	55.7	**57.9**
Bic	64.2	65.7	70.2	67.7	**70.2**

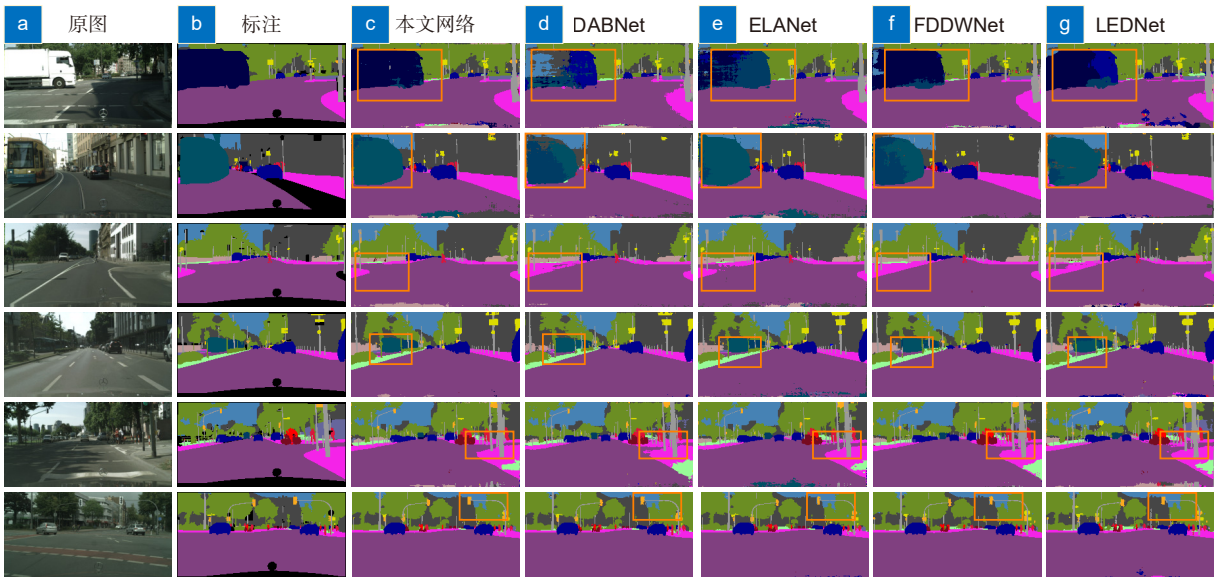


图 13 各网络在 Cityscapes 数据集上的可视化结果

Fig. 13 Visualization results of networks on Cityscapes dataset

过计算每个像素位置与高相关度区域之间的关联性权重，有效捕获全局范围内的信息，摆脱了局部窗口的视野局限，能够更好地理解大物体的结构和边界。从第三行和第五行中可看出 DABNet、FDDWNet 和 LEDNet 对道路像素的判别受到其他像素的干扰，而本文方法较好地预测了道路的边界。本文方法能够准确地划分大型车辆及道路的边界，达到较好的分割效果；对于第六行中的栏杆和路灯等小物体，分割效果更为完整。在第四行中本文方法准确地划分出树枝的像素区域，而其余方法均出现了局部误判为电线杆的现象。综上，本文在自动驾驶场景中的关键类别 (如行人、车辆、道路和信号灯) 分割结果最优。

为验证本文方法的泛化能力，在另一常用数据集 CamVid 上进行了对比实验。实验结果如表 7 所示。精度上，本文方法仅低于采用 ImageNet 预训练的 Bisenet-v2 模型，但值得注意的是，本文采用从头训练的方法，并且参数量上只占其 1/8。在参数量上，尽管本文方法的参数量是 ENet 的两倍，但本文方法精度指标更优。综上所述，本文方法更适用于资源受限的硬件设备。

图 14 的可视化结果对比也可看出，本文方法在围墙及建筑物 (图中黄色虚线框区域) 等类别上的分割效果更好，PBRA 模块建立的长距离依赖有效捕捉到围墙边界及内部之间的相互联系，显著提升了围墙类别的类内一致性，有效改善了对道路状况的感知能力。

表 7 CamVid 数据集上的性能对比

Table 7 Performance comparison on the CamVid dataset

Algorithm	Size	Pretrain	Params/M	mIoU/%
ENet	360×480	N	0.36	51.3
CGNet	360×480	N	0.5	64.7
DALNet	360×480	N	0.47	66.1
LEDNet	360×480	N	0.94	66.6
DABNet	360×480	N	0.76	66.4
MIFNet	360×480	N	0.81	67.7
ELANet	360×480	N	0.67	67.9
BiseNet-v2	360×480	Y	5.8	**68.7**
Ours	360×480	N	0.7	68.2

5 结束语

本文提出一种高效轻量级的实时语义分割方法，专注于平衡推理速度和准确性。通过轻量级特征提取 (AFR) 模块，取得高效的特征提取，并采用两阶段自注意力 (PBRA) 模块，解决了同一物体类内不一致的问题。在解码器部分，采用 FADE 上采样算子接收编码器中的区域信息和解码器中的语义信息，完成图像空间细节的恢复，改善了由于多次下采样导致的空间信息损失的问题。在 Cityscapes 上本文取得了 75.1% 的分割精度的同时，以 0.7 M 的参数量实现 121 f/s 的推理速度。在参数、准确性和速度的综合考虑下，BiLevelNet 优于其他轻量级网络，更适用于自动驾驶场景。

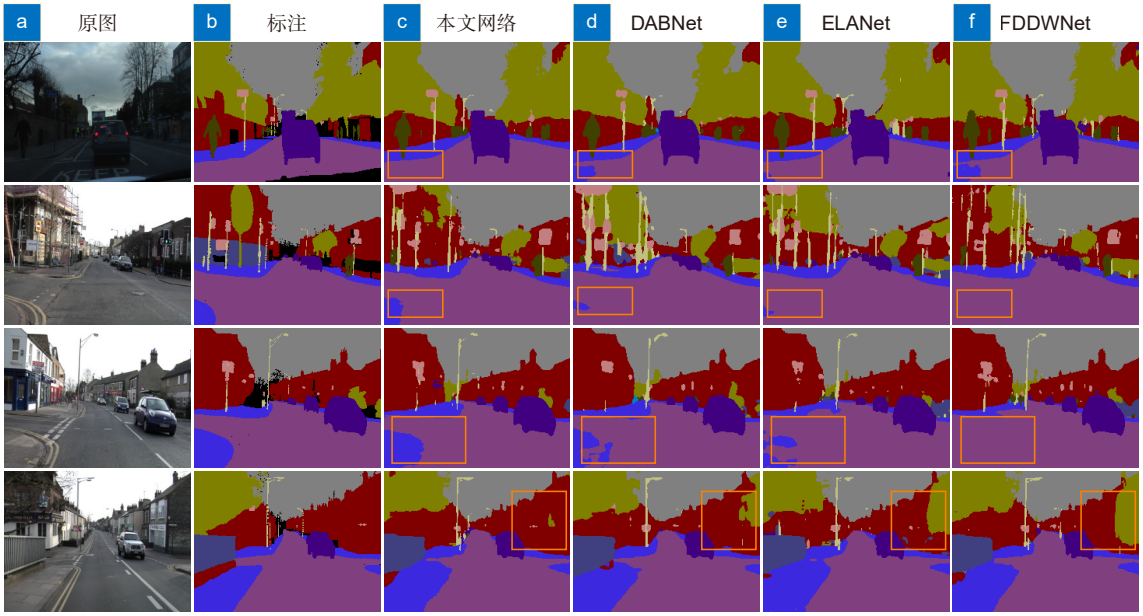


图 14 各网络在 Camvid 数据集上的可视化结果

Fig. 14 Visualization results of networks on the Camvid dataset

参考文献

[1] Li L H, Qian B, Lian J, et al. Traffic scene segmentation based on RGB-D image and deep learning[J]. *IEEE Trans Intell Transp Syst*, 2017, **19**(5): 1664−1669.

[2] Liang L M, Lu B H, Long P W, et al. Adaptive feature fusion cascade transformer retinal vessel segmentation algorithm[J]. *Opto-Electron Eng*, 2023, **50**(10): 230161.
梁礼明, 卢宝贺, 龙鹏威, 等. 自适应特征融合级联Transformer视网膜血管分割算法[J]. 光电工程, 2023, **50**(10): 230161.

[3] Min F, Peng W M, Kuang Y G, et al. A remote sensing ground object segmentation algorithm based on non-subsampled contourlet transform[J]. *Electron Opt Control*, 2023, **30**(11): 49−55.
闵锋, 彭伟明, 况永刚, 等. 基于非下采样轮廓波变换的遥感地物分割算法[J]. 电光与控制, 2023, **30**(11): 49−55.

[4] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 2881–2890. https://doi.org/10.1109/CVPR.2017.660.

[5] Zhang W B, Qu J, Wang W, et al. An improved Deeplab v3+ image semantic segmentation algorithm incorporating multi-scale features[J]. *Electron Opt Control*, 2022, **29**(11): 12−16,30.
张文博, 瞿珏, 王崴, 等. 融合多尺度特征的改进Deeplab v3+图像语义分割算法[J]. 电光与控制, 2022, **29**(11): 12−16,30.

[6] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3[C]//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*, 2019: 1314–1324. https://doi.org/10.1109/ICCV.2019.00140.

[7] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 3213–3223. https://doi.org/10.1109/CVPR.2016.350.

[8] Brostow G J, Fauqueur J, Cipolla R. Semantic object classes in video: a high-definition ground truth database[J]. *Pattern Recognit Lett*, 2009, **30**(2): 88−97.

[9] 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.

[10] Zhuang M X, Zhong X Y, Gu D B, et al. LRDNet: a lightweight and efficient network with refined dual attention decorder for real-time semantic segmentation[J]. *Neurocomputing*, 2021, **459**: 349−360.

[11] Romera E, Álvarez J M, Bergasa L M, et al. ERFNet: efficient residual factorized ConvNet for real-time semantic segmentation[J]. *IEEE Trans Intell Transp Syst*, 2018, **19**(1): 263−272.

[12] Liu J, Zhou Q, Qiang Y, et al. FDDWNet: a lightweight convolutional neural network for real-time semantic segmentation[C]//*Proceedings of the ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2020: 2373–2377. https://doi.org/10.1109/ICASSP40776.2020.9053838.

[13] Liu J, Xu X Q, Shi Y Q, et al. RELAXNet: residual efficient learning and attention expected fusion network for real-time semantic segmentation[J]. *Neurocomputing*, 2022, **474**: 115−127.

[14] Lin S L, Peng X L, Lin J P, et al. Object detection of steel surface defect based on multi-scale enhanced feature fusion[J]. *Opt Precision Eng*, 2024, **32**(7): 1076−1086.
林珊玲, 彭雪玲, 林坚普, 等. 多尺度增强特征融合的钢表面缺陷目标检测[J]. 光学精密工程, 2024, **32**(7): 1076−1086.

[15] Wang Y, Zhou Q, Liu J, et al. Lednet: a lightweight encoder-decoder network for real-time semantic segmentation[C]//*Proceedings of 2019 IEEE International Conference on Image Processing*, 2019: 1860–1864. https://doi.org/10.1109/ICIP.2019.8803154.

[16] Wei H R, Liu X, Xu S C, et al. DWRSeg: dilation-wise residual network for real-time semantic segmentation[Z]. arXiv: 2212.01173, 2023. https://arxiv.org/abs/2212.01173v1.

[17] Chen J R, Kao S H, He H, et al. Run, don't walk: chasing

higher FLOPS for faster neural networks[C]//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 12021–12031. https://doi.org/10.1109/CVPR52729.2023.01157.

[18] Ma N N, Zhang X Y, Zheng H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 116–131. https://doi.org/10.1007/978-3-030-01264-9_8.

[19] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.

[20] Zhang C, Huang Y P, Guo Z Y, et al. Real-time lane detection method based on semantic segmentation[J]. *Opto-Electron Eng*, 2022, **49**(5): 210378.
张冲, 黄影平, 郭志阳, 等. 基于语义分割的实时车道线检测方法[J]. 光电工程, 2022, **49**(5): 210378.

[21] Huang Z L, Wang X G, Huang L C, et al. CCNet: criss-cross attention for semantic segmentation[C]//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*, 2019: 603–612. https://doi.org/10.1109/ICCV.2019.00069.

[22] Wu G, Ge Y, Chu J, et al. Cascade pooling self-attention research for remote sensing image retrieval[J]. *Opto-Electron Eng*, 2022, **49**(12): 220029.
吴刚, 葛芸, 储珺, 等. 面向遥感图像检索的级联池化自注意力研究[J]. 光电工程, 2022, **49**(12): 220029.

[23] Xia Z F, Pan X R, Song S J, et al. Vision transformer with deformable attention[C]//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 4794–4803. https://doi.org/10.1109/CVPR52688.2022.00475.

[24] Zhu L, Wang X J, Ke Z H, et al. BiFormer: vision transformer with Bi-level routing attention[C]//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 10323–10333. https://doi.org/10.1109/CVPR52729.2023.00995.

[25] Wang J Q, Chen K, Xu R, et al. CARAFE: content-aware ReAssembly of FEatures[C]//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*, 2019: 3007–3016. https://doi.org/10.1109/ICCV.2019.00310.

[26] Liu C J, Qiao Z, Yan H W, et al. Semantic segmentation network for remote sensing image based on multi-scale mutual attention[J]. *J Zhejiang Univ (Eng Sci)*, 2023, **57**(7): 1335−1344.
刘春娟, 乔泽, 闫浩文, 等. 基于多尺度互注意力的遥感图像语义分割网络[J]. 浙江大学学报 (工学版), 2023, **57**(7): 1335−1344.

[27] Lu H, Liu W Z, Fu H T, et al. FADE: fusing the assets of decoder and encoder for task-agnostic upsampling[C]//*Proceedings of the 17th European Conference on Computer Vision*, 2022: 231–247. https://doi.org/10.1007/978-3-031-19812-0_14.

[28] Li H C, Xiong P F, Fan H Q, et al. DFANet: deep feature aggregation for real-time semantic segmentation[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 9522–9531. https://doi.org/10.1109/CVPR.2019.00975.

[29] Yi Q M, Dai G S, Shi M, et al. ELANet: effective lightweight attention-guided network for real-time semantic segmentation[J]. *Neural Process Lett*, 2023, **55**(5): 6425−6442.

[30] Shi M, Shen J L, Yi Q M, et al. Rapid and ultra-lightweight semantic segmentation in urban traffic scene[J]. *J Front Comput Sci Technol*, 2022, **16**(10): 2377−2386.
石敏, 沈佳林, 易清明, 等. 快速超轻量城市交通场景语义分割[J]. 计算机科学与探索, 2022, **16**(10): 2377−2386.

[31] Yi Q M, Zhang W T, Shi M, et al. Semantic segmentation for road scene based on multiscale feature fusion[J]. *Laser Optoelectron Prog*, 2023, **60**(12): 1210006.
易清明, 张文婷, 石敏, 等. 多尺度特征融合的道路场景语义分割[J]. 激光与光电子学进展, 2023, **60**(12): 1210006.

[32] Lan J P, Dong F L, Yang Y H, et al. Real-time image semantic segmentation network algorithm based on improved STDC-Seg[J]. *Transducer Microsyst Technol*, 2023, **42**(11): 110−113,118.
兰建平, 董冯雷, 杨亚会, 等. 改进STDC-Seg的实时图像语义分割网络算法[J]. 传感器与微系统, 2023, **42**(11): 110−113,118.

作者简介

吴马靖 (1997-)，男，福建泉州人，硕士研究生，2019 年于桂林电子科技大学获得学士学位，主要从事图像处理方面的研究。

E-mail：442683978@qq.com

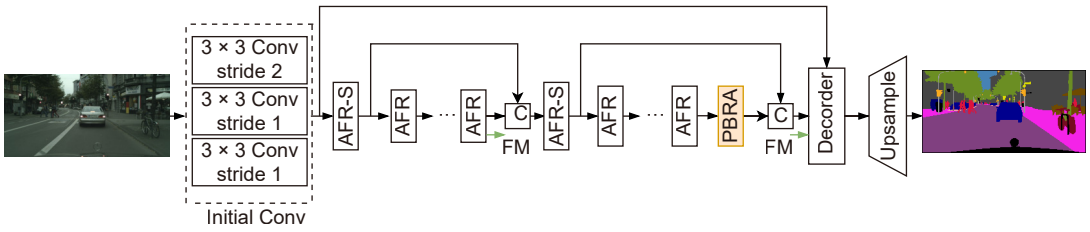
【通信作者】林坚普 (1989-)，男，福建泉州人，博士，讲师，硕士研究生导师，福州大学先进制造学院电子信息系教师，主要从事新型显示技术、图像处理技术、电子纸驱动与集成等方面的研究。

E-mail：ljp@fzu.edu.cn

扫描二维码，获取PDF全文

Real-time semantic segmentation algorithm based on BiLevelNet

Wu Majing[1], Zhang Yong'ai[1,2], Lin Shanling[1,2], Lin Zhixian[1,2], Lin Jianpu[1,2*]



Network framework of BiLevelNet

Overview: In response to the challenge posed by the large parameter sizes of semantic segmentation networks, which complicate deployment on memory-constrained edge devices, a lightweight real-time semantic segmentation algorithm based on BiLevelNet is proposed. Initially, dilated convolutions are utilized to broaden the receptive field, and strategies for reusing features are integrated to bolster the network's awareness of regions. Subsequently, a two-stage PBRA (Partial Bi-Level Route Attention) mechanism is adopted to form connections between distant objects, thereby enhancing the network's capability to perceive global contexts. Moreover, the FADE operator is introduced for merging shallow features, thereby augmenting the efficacy of image upsampling.

Within the depicted AFR module in Fig. 4, a variety of hierarchical feature maps are presented, along with descriptions of their characteristics and roles. The distinctions and connections between the input feature map, the local feature map achieved through 3×3 depth convolution, and the context information feature map acquired through dilated convolution are clarified. It is further emphasized how these features are effectively amalgamated in the final fused feature map, showcasing strong activation across both local and global contexts. Additionally, a gradually decreasing channel reduction factor is employed, as elaborated in Table 3. Through the gradual adjustment of the channel reduction factor, it is observed that with a reduction factor of r=1/4, the PBRA module enhances mIoU by 1.5% and boosts speed by 12FPS in comparison to BRA.

Moreover, discontinuities and missing pixels are noted in segmentation results when bilinear interpolation is used for upsampling. Observations of the depth feature maps prior to bilinear upsampling reveal that features corresponding to roads and sidewalks bear similarities, leading to potential misclassifications. To counteract this issue, shallow features that preserve edge information are introduced and merged into the FADE upsampling process, thereby improving edge segmentation. This method effectively addresses the loss of spatial information, resulting in smoother and more defined edge segmentation outcomes.

Experimental outcomes indicate that, at an input image resolution of 512×1024, the network attains an average Intersection over Union (IoU) of 75.1% on the Cityscapes dataset, operating at a speed of 121 frames per second, while maintaining a modest model size of only 0.7M. Furthermore, at an input image resolution of 360×480, the network secures an average IoU of 68.2% on the CamVid dataset. Compared with other real-time semantic segmentation methods, this network maintains an optimal balance between speed and accuracy, fulfilling the real-time operation requirements for applications such as autonomous driving.

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; DOI: 10.12086/oee.2024.240030

Foundation item: Project supported by the National Key R&D Program of China (2023YFB3609400), Fujian Province Natural Science Foundation of China (2020J01468), and Youth Science Foundation of the National Natural Science Foundation of China (62101132)

[1]School of Advanced Manufacturing, Fuzhou University, Quanzhou, Fujian 362200, China; [2]Fujian Science & Technology Innovation Laboratory for Optoelectronic Information of China, Fuzhou, Fujian 350116, China

* E-mail: ljp@fzu.edu.cn