计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2024,33(3):111−117 [doi: 10.15888/j.cnki.csa.009429]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

融合深度信息的室内场景分割算法①

王　柳[1], 梁铭炬[2]

[1](郑州大学 电气与信息工程学院, 郑州 450001)
[2](广东佛山联创工程研究生院 软件专业部, 佛山 528300)
通信作者: 王　柳, E-mail: 1718194513@qq.com

摘　要: 针对室内复杂场景中, 图像语义分割存在的特征损失和双模态有效融合等问题, 提出了一种基于编码器-解码器架构的融合注意力机制的轻量级语义分割网络. 首先采用两个残差网络作为主干网络分别对 RGB 和深度图像进行特征提取, 并在编码器中引入极化自注意力机制, 然后设计引入双模态融合模块在不同阶段对 RGB 特征和深度特征进行有效融合, 接着引入并行聚合金字塔池化以获取区域之间的依赖性. 最后, 采用 3 个不同尺寸的解码器将前面的多尺度特征图进行跳跃连接并融合解码, 使分割结果含有更多的细节纹理. 将本文提出的网络模型在 NYUDv2 数据集上进行训练和测试, 并与一些较先进 RGB-D 语义分割网络对比, 实验证明本文网络具有较好分割性能.

关键词: RGB-D 图像; 注意力机制; 多模态融合; 上下文聚合

引用格式: 王柳,梁铭炬.融合深度信息的室内场景分割算法.计算机系统应用,2024,33(3):111–117. http://www.c-s-a.org.cn/1003-3254/9429.html

Indoor Scene Segmentation Algorithm Based on Fusion of Deep Information

WANG Liu[1], LIANG Ming-Ju[2]

[1](School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou 450001, China)
[2](Software Department, Guangdong Foshan Lianchuang Engineering Graduate School, Foshan 528300, China)

Abstract: A lightweight semantic segmentation network based on encoder-decoder architecture with fusion attention mechanism is proposed to address the issues of feature loss and effective bimodal fusion in image semantic segmentation in complex indoor scenes. Firstly, two residual networks are used as backbone networks to extract features from RGB and depth images, and a polarized self-attention (PSA) module is introduced into the encoder. Then, a bimodal fusion module is designed and introduced to effectively fuse RGB and depth features at different stages. A context module is introduced to obtain dependencies between regions. Finally, three decoders of different sizes are applied to skip connect and fuse the previous multi-scale feature maps to improve the segmentation accuracy of small targets. The proposed network model is trained and tested on the NYUDv2 datasets and compared with more advanced RGB-D semantic segmentation networks. The experiments show that the proposed network has good segmentation performance.

Key words: RGB-D image; attention mechanism; multimodal fusion; context aggregation

图像语义分割是一种像素级别的分类任务, 旨在根据图像中每个像素点的语义含义进行分类, 并对各个类别赋予不同的语义标签[1]. 室内场景分割存在语义类别较多、物体特征不突出、光照不均匀等问题[2]. 随着深度传感器的普及, 采集的图像信息从彩色信息扩展到了 RGB-D 信息. RGB 图像能够描述物体颜色和纹理之类的外观信息, 而深度图像能够描述物体的三维几何信息, 这种信息很难在 RGB 特征中提取并且不

① 基金项目: 广东省科技创新战略专项资金 (纵向协同管理方向)(2018FS05020102); 佛山市高质量专利培育项目 (1920025003148)
收稿时间: 2023-08-31; 修改时间: 2023-09-26; 采用时间: 2023-10-25; csa 在线出版时间: 2024-01-02
CNKI 网络首发时间: 2024-01-03

会随着光照的变化而变化. 近年来, 很多研究将颜色信息和深度信息结合起来用于语义分割任务, RGB-D 图像语义分割引入了对 RGB 数据进行补充的深度信息, 可以有效提高室内场景语义分割的性能[3].

1 相关工作

随着深度学习的发展, 在 RGB-D 语义分割网络领域中也涌现了很多优秀的算法. 早期 Gupta 等人[4,5]将深度图编码成水平视差、地面高度和重力与表面法线之间的角度 3 个通道, 这种编码方式后来在其他算法中得到了广泛的应用, 但是该方法忽略了各个通道的独立成分, 具有一定局限性. 后来, 多层残差特征融合网络 RedNet 采用双分支架构, 将残差模块作为基本的模块应用于编码器和解码器, 并用元素求和的方式进行特征融合, 取得了不错的效果[6]. 而在 2019 年发布的注意力互补网络 ACNet[7]则首次利用了三平行分支的网络架构, 较好地平衡了 RGB-D 图像中 RGB 图像特征和深度图像特征, 有效利用了融合特征. Wu 等人[8]在 2022 年提出了一种新的框架, 将深度信息纳入 RGB 卷积神经网络之中, 该网络能够生成一个 2D 的深度偏移量以指导 RGB 图像上的特征提取, 并引入了深度适应卷积和深度适应平均池化来取代基本的 CNN 运算符. Bai 等人[9]则考虑深度数据的几何信息和局部相关性, 提出了一个像素差分卷积注意模块, 并将其扩展到集成差分卷积注意力, 使得网络能够传播长距离上下文依赖关系. 赵经阳等人[10]提出了一个可以自适应调整深度信息嵌入程度的特征提取模块, 能够更加合理地利用深度信息.

上述方法虽然在 RGB-D 语义分割任务中取得了不错的效果, 但仍然存在着一些不足.

第一, 大多数采用编码器-解码器架构的语义分割算法中, 编码器降低了空间维度, 提高了通道维度. 这种结构在编码器阶段就会因为空间维度的降低而造成部分特征损失, 进而造成分割性能上的降低.

第二, 由于深度传感器的物理特性, 深度信息中包含大量异常噪声特征, 直接利用时将会带来负面影响, 产生的不精确性可能会抵消引入深度图所带来的性能提高[10].

第三, 由于深度特征和 RGB 特征描述的信息差异很大, 因此对语义分割的贡献也有主要和次要之分, 过于简单的元素相加或者权值拼接等方式会融入过多的深度信息, 可能会对网络产生干扰.

基于上述, 为进一步有效融合深度信息, 提高室内复杂场景分割的性能, 本文提出了一种基于编码-解码架构的 RGB-D 分割网络, 并在 NYUDv2 数据集上进行训练和测试, 取得了不错的效果.

2 网络模型

本文提出的融合深度信息的 RGB-D 语义分割网络模型的整体结构如图 1 所示, 主干网络为两个并行的卷积分支, 分别从 RGB 图和深度图提取特征, 并在编码器中添加极化注意力机制和双模态融合模块以减少特征损失并有效融合多模态信息. 将特征提取网络提取的特征送入深度聚合金字塔池化模块以聚合上下文信息, 获取区域之间的依赖关系. 最后, 通过跳跃连接聚合不同尺度特征, 使用 3 个解码器将特征图映射成要判别的类别, 再通过上采样操作逐步恢复图像分辨率. 就其内部 4 个主要模块介绍如下.

2.1 PSA 极化注意力机制

为了解决在编码器-解码器结构下语义分割的空间特征损失问题, 本文引入了一个极化注意力模块 (polarized self-attention, PSA)[11], 参考文献[11]中介绍的使用方法, 仅在 RGB 分支中的 3×3 卷积后添加 PSA 注意力. 实验证明, 网络仅增加了少量的计算开支和参数数量, 却有效提高网络分割性能.

PSA 是针对像素级回归任务提出的一种双重注意力机制, 用于增强或者抑制部分特征信息. 受极化滤波思想的启发, PSA 具体可分为两个点: 第 1 点是极化滤波, 使图像的特征在一个维度方向上完全折叠, 同时让其正交方向的维度上保持高分辨率, 其主要作用就是在空间和维度两个分支上分别保持较高的分辨率, 降低由空间维度降低带来的特征损失. 第 2 点是特征增强, 采用 Softmax 进行归一化, 再使用 Sigmoid 函数增加注意力的动态范围, 能够更加真实地拟合输出分布.

图 2 所示为本文所采用的顺序结构的 PSA, 上边为通道分支, 下边为空间分支, 输入和输出均为 $C\times H\times W$. 在通道分支上, 先将输入特征 X 经过 1×1 卷积后转换为 Q 和 V, Q 的通道被全部压缩, 而 V 的通道则保持 $C/2$. 经过 Softmax 函数后对 Q 的信息进行增强, 并增加注意力的范围, 然后将 Q 和 V 做矩阵乘法运算, 再经过 1×1 卷积, LN 层将通道数量还原为 C, 最后使用 Sigmoid 函数进行动态映射. 经过通道分支之后, 输出 Z^{ch} 进入空间分支, 过程与通道分支相反, 最后输出经过注意力变换后的特征图.

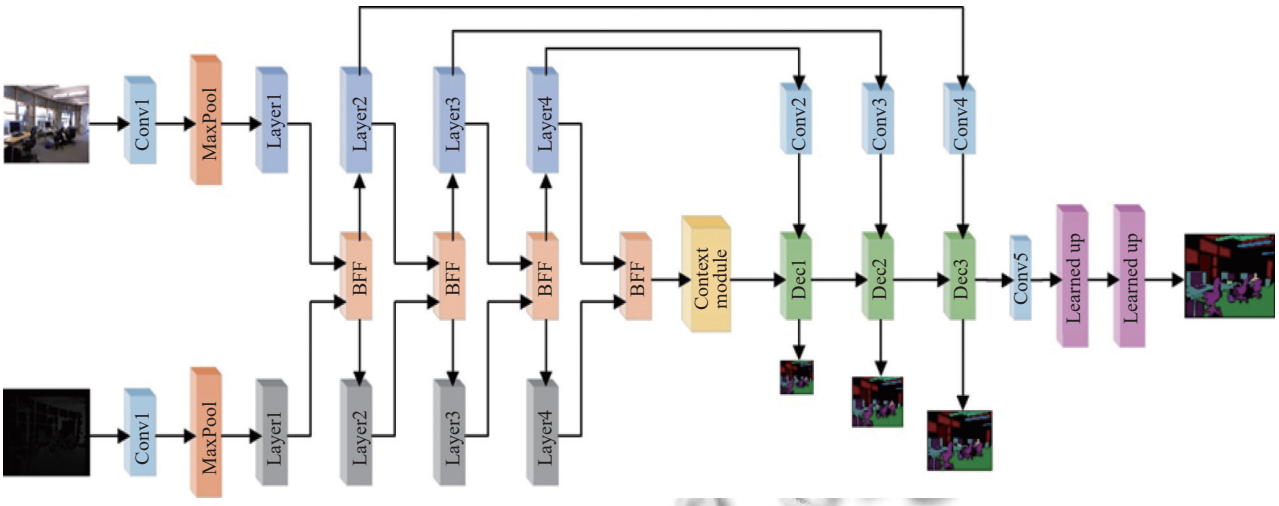

图 1 网络总体框架图

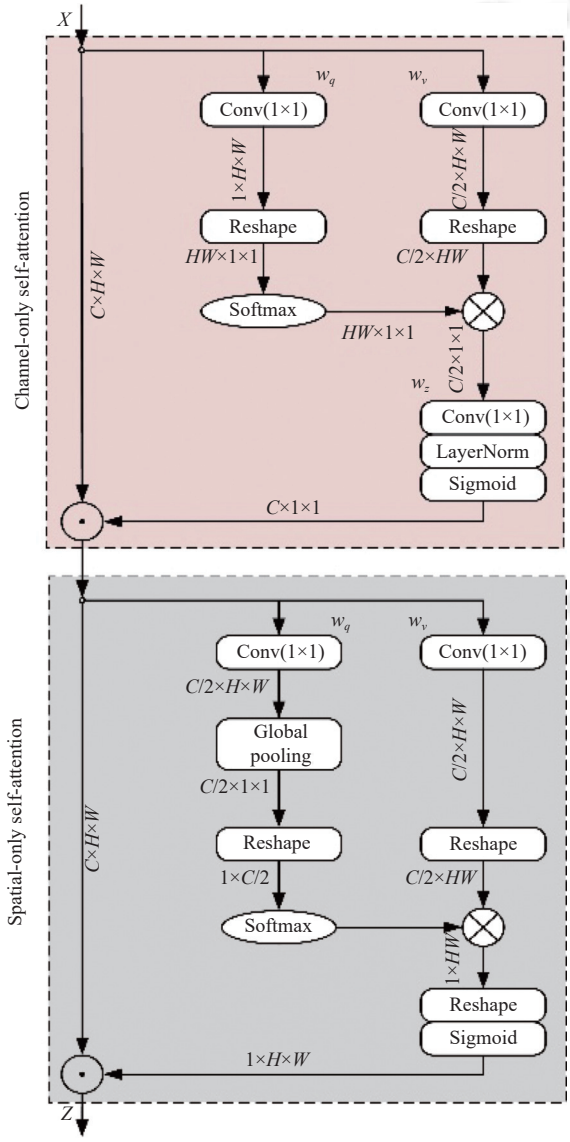

图 2 PSA 顺序结构

2.2 双模态特征融合模块 BFM

在网络模型中, 由于深度图像与 RGB 图像之间的差异性, 两种图像特征信息的融合是一个极大的挑战. 一个有效的双模态特征融合方法应该能从每个特征中识别出它们的优势, 将信息量最大的跨模态特征统一到有效的表示中. 受 SA-Gate[12]和空间注意力机制的启发, 为了充分利用深度信息, 本文设计了一个双模态特征融合模块 (bimodal feature fusion module, BFM), 该模块不仅可以有效地校准 RGB 特征, 而且还可以通过多个阶段来提取准确的深度信息, 将两者交替合并生成融合特征. 为了确保模态之间的信息特征传播, BFM 被设计为具有两个操作, 即单一模态的特征校准和跨模态特征融合, 如图 3 所示.

特征校准: 该部分借鉴 SA-Gate 中的特征校准操作, 首先将两种模态的特征图RGB_{in}和Dep_{in}进行通道到维度上的拼接, 然后进行全局平均池化操作F_{gp}, 再通过F_{mlp}中的卷积操作, 还原通道维度, 接着通过 Sigmoid 函数, 得到两个通道上对应的权重W_{rgb}和W_{depth}, 将其与对应的原始通道特征进行空间相乘, 并与另一原始通道的特征图进行像素相加, 得到去噪校准后的特征RGB_{rec}和Dep_{rec}, 该部分操作可以用如下公式表示:

$$W=\sigma\left(F_{\text{mlp}}\left(F_{\text{gp}}\left(RGB_{\text{in}}\|Dep_{\text{in}}\right)\right)\right) \tag{1}$$

$$RGB_{\text{rec}}=Dep_{\text{in}}\times W_{\text{depth}}+RGB_{\text{in}} \tag{2}$$

$$Dep_{\text{rec}}=RGB_{\text{in}}\times W_{\text{rgb}}+Dep_{\text{in}} \tag{3}$$

特征融合: 为了充分利用 RGB 特征和深度特征的

互补性, 需要根据它们的表征能力在空间的某个位置互补地聚合双模态特征. 该部分首先将校准后的特征进行拼接操作, 得到一个新的特征模块, 接着对其进行平均池化 AvgPool 和最大池化 MaxPool 操作, 得到两个维度为 1×H×W 的空间注意力向量, 然后对两个向量进行拼接, 拼接后的向量维度为 2×H×W, 最后对该向量进行卷积和 Sigmoid 激活函数操作, 并与校准后的特征RGB_{rec}和Dep_{rec}进行元素相乘, 该部分操作可用如下公式表示:

$$F = [Avg(RGB_{\text{rec}} \| Dep_{\text{rec}}), Max(RGB_{\text{rec}} \| Dep_{\text{rec}})] \quad (4)$$

$$RGB_{\text{out}} = RGB_{\text{rec}} \otimes \sigma(convF) \quad (5)$$

$$Depth_{\text{out}} = Dep_{\text{rec}} \otimes \sigma(convF) \quad (6)$$

输出的RGB_{out}和$Depth_{\text{out}}$会在不同阶段送入主干网络中进一步提取特征. BFM 模块通过特征校准和特征融合, 得到 RGB 特征和深度特征的注意力权重, 并利用得到的权重信息对特征进行重新分配和有效融合, 最终提高模型的分割效果.

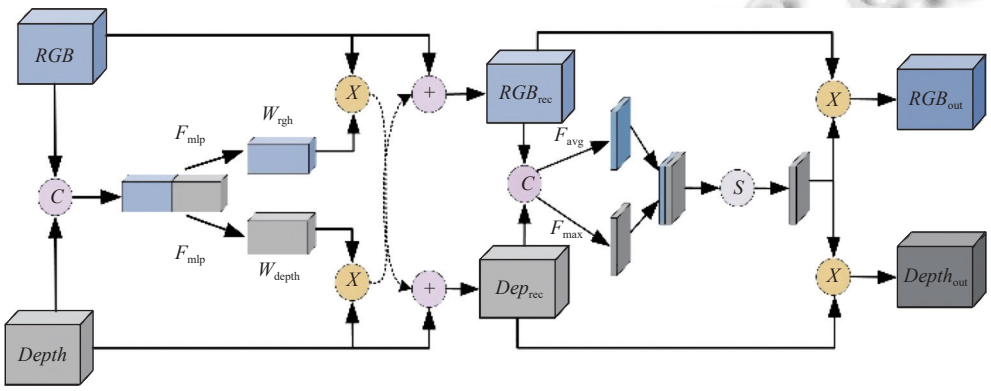

图 3 双模态融合模块 BFM

2.3 上下文模块

一幅图像中任何一个像素都与周围像素存在一定的关系, 大量像素的关联才形成了图像中的各种物体, 但是在经过残差网络对特征的提取和双模态特征融合之后, 由于不同感受野的语境关系和部分上下文信息的缺失会导致易混淆的物体和小尺寸物体的错误分类[13], 针对这个问题, 本文在模型的编码器和解码器之间引入一个并行聚合金字塔池化模块 (parallel aggregation pyramid pooling module, PAPPM)[14], 该模块将特征聚合与金字塔池化相结合, 输入低分辨率的特征图, 利用不同的池化方式增加感受野, 以融合不同尺度的特征信息, 捕获更加丰富的上下文信息. PAPPM 结构如图 4 所示, 在模块内部, 由较大的池化核提取的上下文与更深的信息流集成, 并且通过将不同深度与不同大小的池化核集成而形成多尺度性质. 该模块是用序列 BN-ReLU-Conv 实现的, 并且对深度聚合金字塔池化模块的连接进行改进使其并行化, 减少了每个尺度的通道数, 因此几乎不影响推理速度.

2.4 解码器

虽然经过双模态融合和上下文模块之后, 网络得到的特征图已经包含了丰富的语义信息和全局信息, 但将特征图直接送入编码器中进行下采样的时候会丢失细粒度细节, 因此, 参考文献[15,16] , 我们也采用了 3 个解码器模块. 解码器以上下文模块输出的 512 通道的特征图作为第 1 层的输入, 在后续解码器层中采用多尺度跳跃连接, 用 1×1 卷积将来自编码器子网的浅层、低层、细粒度特征与来自解码器子网的深度、语义、粗粒度的相同尺度特征图结合在一起 , 从而获取包含低级空间信息和高级语义信息的全局特征, 有效地恢复目标对象的细节纹理[17]. 同时, 在每一次跨层融合之后, 对特征图结果采用上采样方式来扩大分辨率, 再使用 3×3 的深度卷积来组合相邻特征.

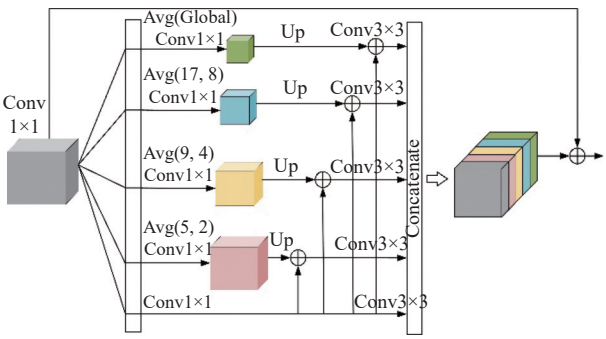

图 4 并行聚合金字塔池化模块 PAPPM

3 实验

3.1 实验数据集

本文使用常用的 RGB-D 语义分割数据集 NYUDv2 评估所提出的模型. NYUDv2 数据集包含 1 449 张来自室内场景的 RGB-D 图像. 本文使用的语义标签划分为 40 类, 在实验中, 使用 795 张图像训练, 654 个图像测试.

3.2 实验细节

本文采用交叉熵函数作为损失函数, 并在每一个解码器模块后加一个输出, 将不同分辨率的输出和最后结果输入到网络末端得到最终损失函数:

$$Loss(x,y)=\frac{1}{N}\sum_i\left(1-\ln\frac{\exp(x_i[y_i])}{\sum_k\exp(x_i[k])}\right) \tag{7}$$

其中, x_i表示像素在位置i上输出中相应类别所得分数; y_i表示在位置i上标签语义映射的类索引; N表示输出的整体分辨率; k表示数据集中的类别数. 总损失函数为 3 个解码器模块和最后的损失之和.

我们采用各种基准数据集最常用的标准, 像素精度 PA、平均交并比 $mIoU$ 以及各个类别交并比 IoU 作为评价指标. 计算公式如下:

$$PA=\frac{\sum_{i=0}^{k}P_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k}P_{ij}} \tag{8}$$

$$IoU=\frac{P_{ii}}{\sum_{j=0}^{k}P_{ij}+\sum_{j=0}^{k}P_{ji}-P_{ii}} \tag{9}$$

$$mIoU=\frac{1}{k+1}\sum_{i=0}^{k}\frac{P_{ii}}{\sum_{j=0}^{k}P_{ij}+\sum_{j=0}^{k}P_{ji}-P_{ii}} \tag{10}$$

本文实验环境为 64 位 Ubuntu 18.04 操作系统, NVIDIA GeForce RTX 2080 TI 显卡, 在 CUDA 10.2、PyTorch 1.8.1 下进行模型的训练与测试. 在训练过程中使用的优化器为 SGD 优化器, 训练批次大小为 4, 初始学习率设置为 0.01, 迭代次数为 500 个 epoch. 图 5 所示黑色代表训练时损失函数变化曲线, 红色代表验证集的 $mIoU$ 指标的变化曲线, 横坐标为迭代次数 epoch. 从图 5 中可以看出, 随着 epoch 的增长, 损失函数在 300 个 epoch 左右趋于收敛, 随后有小幅下降, 指标 $mIoU$ 也趋近收敛, 最终达到 50.3%, 证明模型训练的可行性和算法的有效性.

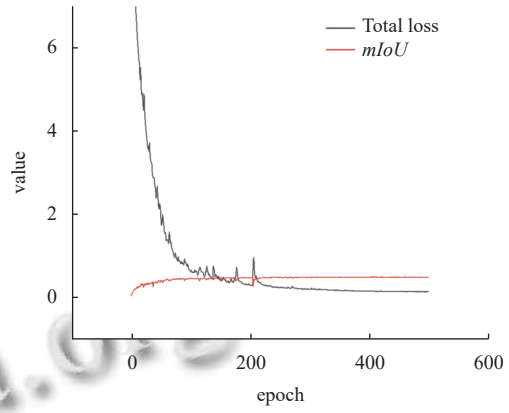

图 5 总损失函数和 $mIoU$ 变化曲线

3.3 实验分析

3.3.1 消融实验

实验通过记录单独减少或使用简单操作替换每个模块后的网络性能, 证明本文构建的各个模块在整个网络中的优势. 消融实验如表 1 所示. 本文网络在 NYUDv2 数据集上 40 个类别的 IoU 值如表 2 所示, 图 6 所示为算法在 NYUDv2 数据集上的分割结果示例.

表 1 在 NYUDv2 数据集上的消融实验 (%)

模块	PSA	BFM	PAPPM	*PA*	*mIoU*
All	√	√	√	77.6	50.3
PSA	—	√	√	76.5	48.7
BFM	√	—	√	76.8	48.9
PAPPM	√	√	—	77.4	49.6

表 2 本文算法在 NYUDv2 数据集上各类别 IoU (%)

类别	*IoU*	类别	*IoU*	类别	*IoU*	类别	*IoU*
wall	80.0	picture	61.4	clothes	24.9	person	71.8
floor	87.8	counter	69.5	ceiling	76.0	night stand	50.5
cabinet	63.5	blinds	58.7	books	31.4	toilet	78.5
bed	71.3	desk	26.0	refridgerator	60.1	sink	59.3
chair	66.1	shelves	15.1	television	55.4	lamp	52.8
sofa	63.6	curtain	60.1	paper	32.1	bathtub	48.9
table	46.8	dresser	49.9	towel	41.5	bag	11.5
door	39.3	pillow	51.9	shower curtain	56.9	others1	32.3
window	46.9	mirror	52.2	box	14.8	others2	20.0
bookshelf	47.1	floormat	38.9	whiteboard	57.5	others3	39.8

(1) PSA 极化注意力机制

从表 1 第 2 行可以看出, 当编码器部分不再使用 PSA 极化注意力机制, 在 NYUDv2 数据集上网络交并比和平均精度分别下降了 1.6% 和 1.1%. 图 6 第 3 列

为主干网络没有添加 PSA 极化注意力机制的分割结果. 而最后一列则是完整的网络模型的分割结果. 对比可以看出, 没有添加注意力机制的网络分割比较杂乱, 存在很多误匹配的情况.

图 6 分割结果

(2) 双模态融合 BFM 模块

在两个主干网络提取特征之后对 RGB 特征和深度特征采用简单的元素相加的方式, 相比于原始网络在 NYUDv2 数据集上平均交并比和平均精度分别下降了 1.4% 和 0.8%. 由图 6 第 4 列和最后一列分割图像的对比可以看出, 像灯光下的墙壁和墙画之类的物体, 在边界不清、颜色相近和易混淆物体存在的场景下, 本文网络能更准确提取边界特征区分出不同物体种类, 同时对光照条件也具有更好的鲁棒性.

(3) 并行聚合金字塔池化 PAPPM 模块

PAPPM 模块去掉之后, 在 NYUDv2 数据集上网络平均交并比降低了 0.7% 和 0.2%. 由图 6 的可视化结果可以看出, 去除并行金字塔池化模块后, 网络对于小目标物体的分割不再敏感, 会将一些如书本、盒子之类的小目标物体划分到与之相邻的大目标中; 而加入上下文模块之后加强了特征的全局信息, 强化了整个网络的分割能力.

3.3.2 模型对比

本文采用的主干网络虽然是轻量级的 ResNet34, 但性能超过了部分主干网络为 ResNet101 的算法. 而与主干为双分支的 ResNet50 的算法相比, 本文网络参数量较小, 但是取得了具有竞争力的效果. AFACNet 与 AMBFNet 与本文网络采用的是相同的特征提取网络, 但是这两个模型的分割精度分别为 48.2% 和 50.0%, 略低于本文算法. 具体算法对比见表 3.

表 3 算法对比 (%)

网络结构	Backbone	*mIoU*	*PA*
2.5Dconv[18]	ResNet101	49.1	75.9
ASNet[19]	ResNet101	49.5	76.4
SGNet[20]	ResNet101	50.2	76.1
Malleable 2.5D[21]	ResNet101	50.9	76.9
ACNet[7]	ResNet50×3	48.3	69.2
SEANet[22]	ResNet50×3	49.3	83.3
CMANet[23]	ResNet50×2	47.3	73.9
ShapeConv[24]	ResNet50×2	48.8	75.0
SA-Gate[12]	ResNet50×2	50.4	—
AFACNet[25]	ResNet34×2	48.2	—
AMBFNet[17]	ResNet34×2	50.0	—
Ours	ResNet34×2	50.3	77.6

4 结论

针对室内复杂场景中, 图像语义分割存在的特征损失和双模态有效融合等问题, 本文提出了一种 RGB-D 室内场景语义分割网络. 该网络在编码器阶段引入了 PSA 极化注意力机制以减少空间维度上的特征损失, 构建了双模态特征融合模块以校准特征和融合多模态

信息, 引入了并行聚合金字塔池化模块以捕获全局信息. 在常用的室内分割数据集上进行了实验, 验证了本文网络的有效性, 并展示了在 NYUDv2 数据集上的分割示例. 结果表明, 本文采用参数量较少的 ResNet34 作为主干网络, 可以在占用较少资源的情况下取得更好的效果, 因此具有一定的实际应用价值. 下一步将在该算法的基础上进行改进, 增加分类和方向估计等其他任务, 构建一种多任务的室内场景解析算法并在嵌入式系统上进行部署.

参考文献

1 伏娜娜, 许钢, 陈玲, 等. 基于通道特征融合的 RGB-D 图像语义分割方法. 四川轻化工大学学报 (自然科学版), 2022, 35(4): 42–48.

2 梁博, 于蕾, 李爽. 基于卷积神经网络的多任务图像语义分割. 无线电工程, 2019, 49(7): 575–580.

3 贺照蒙, 孔广黔, 吴云. 一种改进的室内场景语义分割网络. 计算机工程与应用, 2021, 57(16): 197–202.

4 Gupta S, Arbelaez P, Malik J. Perceptual organization and recognition of indoor scenes from RGB-D images. Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland: IEEE, 2013. 564–571. doi: 10.1109/CVPR.2013.79.

5 Gupta S, Girshick R, Arbeláez P, *et al*. Learning rich features from RGB-D images for object detection and segmentation. Proceedings of the 13th European Conference on Computer Vision. Zurich: Springer, 2014. 345–360.

6 Jiang JD, Zheng LN, Luo F, *et al*. RedNet: Residual encoder-decoder network for indoor RGB-D semantic segmentation. arXiv:1806.01054, 2018.

7 Hu XX, Yang KL, Lei F, *et al*. ACNet: Attention based network to exploit complementary features for RGBD semantic segmentation. Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP). Taipei: IEEE, 2019. 1440–1444. [doi: 10.1109/ICIP.2019.8803025]

8 Wu ZW, Allibert G, Stolz C, *et al*. Depth-adapted CNNs for RGB-D semantic segmentation. arXiv:2206.03939, 2022.

9 Bai LZ, Yang J, Tian CQ, *et al*. DCANet: Differential convolution attention network for RGB-D semantic segmentation. arXiv:2210.06747, 2022.

10 赵经阳, 余昌黔, 桑农. RGB-D 语义分割: 深度信息的选择使用. 中国图象图形学报, 2022, 27(8): 2473–2486.

11 Liu HJ, Liu FQ, Fan XY, *et al*. Polarized self-attention: Towards high-quality pixel-wise regression. arXiv: 2107.00782, 2021.

12 Chen XK, Lin KY, Wang JB, *et al*. Bi-directional cross-modality feature propagation with separation-and-aggregation gate for RGB-D semantic segmentation. Proceedings of the 19th European Conference on Computer Vision. Glasgow: Springer, 2020. 561–577.

13 范润泽, 刘宇红, 张荣芬, 等. 基于多尺度注意力机制的道路场景语义分割模型. 计算机工程, 2023, 49(2): 288–295. [doi: 10.19678/j.issn.1000-3428.0063257]

14 Xu JC, Xiong ZX, Bhattacharyya SP. PIDNet: A real-time semantic segmentation network inspired by PID controllers. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 19529–19539.

15 Seichter D, Köhler M, Lewandowski B, *et al*. Efficient RGB-D semantic segmentation for indoor scene analysis. Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA). Xi'an: IEEE, 2021. 13525–13531. [doi: 10.1109/ICRA48506.2021.9561675]

16 张吉友, 张荣芬, 刘宇红, 等. 基于注意力机制的多模态图像语义分割. 液晶与显示, 2023, 38(7): 975–984.

17 罗盆琳, 方艳红, 李鑫, 等. RGB-D 双模态特征融合语义分割. 计算机工程与应用, 2023, 59(7): 222–231.

18 Xing YJ, Wang JB, Chen XK, *et al*. 2.5D convolution for RGB-D semantic segmentation. Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP). Taipei: IEEE, 2019. 1410–1414.

19 段立娟, 孙启超, 乔元华, 等. 基于注意力感知和语义感知的 RGB-D 室内图像语义分割算法. 计算机学报, 2021, 44(2): 275–291.

20 Chen LZ, Lin Z, Wang ZQ, *et al*. Spatial information guided convolution for real-time RGBD semantic segmentation. IEEE Transactions on Image Processing, 2021, 30: 2313–2324. [doi: 10.1109/TIP.2021.3049332]

21 Xing YJ, Wang JB, Zeng G. Malleable 2.5D convolution: Learning receptive fields along the depth-axis for RGB-D scene parsing. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 555–571.

22 顾嘉城, 龙英文, 吉明明, 等. 一种面向室内场景的语义分割网络. 激光与红外, 2023, 53(4): 615–625.

23 Zhu LZ, Kang ZZ, Zhou M, *et al*. CMANet: Cross-modality attention network for indoor-scene semantic segmentation. Sensors, 2022, 22(21): 8520. [doi: 10.3390/s22218520]

24 Cao JM, Leng HC, Lischinski D, *et al*. ShapeConv: Shape-aware convolutional layer for indoor RGB-D semantic segmentation. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021. 7068–7077.

25 张静怡. 基于非对称融合和关联上下文的 RGBD 语义分割算法研究. 现代计算机, 2022, 28(1): 96–100.

(校对责编: 孙君艳)