计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
2025,34(9):22−30 [doi: 10.15888/j.cnki.csa.009917] [CSTR: 32024.14.csa.009917]

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

ARD-UNet++: 基于 UNet++的城市遥感图像分割增强模型①

蒋春鸿[1], 陈宇宸[2], 洪泽泓[1], 邹雲宇[1], 潘家辉[1], 梁　军[1]

[1](华南师范大学 人工智能学院, 佛山 528225)
[2](华南师范大学 阿伯丁数据科学与人工智能学院, 佛山 528225)
通信作者: 潘家辉, E-mail: panjh82@qq.com

摘　要: 城市遥感图像具有分辨率高、背景多样、纹理复杂等特点, 这对边界分割提出了挑战. 当前主流的语义分割模型遇到了一些困难, 包括边缘模糊、平滑角等缺陷和无法捕获远程依赖关系. 为了解决这些挑战, 本研究提出了一种基于 UNet++的增强模型 ARD-UNet++. 采用 7×7 深度可分离卷积来减少参数计数, 促进更密集的特征提取和更全面的上下文信息捕获; 引入 SimAM 非参数注意力机制, 在不引入额外参数的情况下选择性地关注关键特征, 有效地抑制无关信息; 集成了残差连接以防止局部最优, 其中 Res-SimAM 模块取代了上采样节点中的标准卷积块. 与 UNet++相比, 本增强模型在 UAVid 和 Potsdam 数据集上表现出显著的提升效果, *mIoU* 分别提高了 6.77% 和 1.79%, *F*1 分别提高了 4.71% 和 1.17%, *OA* 分别提高了 4.99% 和 0.98%. 通过与当前主流模型的对比分析, ARD-UNet++具有优越的性能, 是城市遥感图像精确分割的理想解决方案.

关键词: SimAM 注意力机制; 残差连接; 深度可分离卷积; 城市遥感图像

引用格式: 蒋春鸿,陈宇宸,洪泽泓,邹雲宇,潘家辉,梁军.ARD-UNet++: 基于 UNet++的城市遥感图像分割增强模型.计算机系统应用,2025,34(9): 22–30. http://www.c-s-a.org.cn/1003-3254/9917.html

ARD-UNet++: Enhanced Image Segmentation Model for Urban Remote Sensing Images Based on UNet++

JIANG Chun-Hong[1], CHEN Yu-Chen[2], HONG Ze-Hong[1], ZOU Yun-Yu[1], PAN Jia-Hui[1], LIANG Jun[1]

[1](School of Artificial Intelligence, South China Normal University, Foshan 528225, China)
[2](Aberdeen Institute of Data Science and Artificial Intelligence, South China Normal University, Foshan 528225, China)

Abstract: Urban remote sensing images pose challenges in boundary segmentation due to their high resolution, diverse backgrounds, and intricate textures. Mainstream semantic segmentation models encounter difficulties, including edge blurring, smooth corners, and the inability to capture long-range dependencies.To address these challenges, ARD-UNet++, an enhanced model based on UNet++, is introduced. A 7×7 depthwise separable convolution is employed to reduce the parameter count, facilitating denser feature extraction and comprehensive contextual information capture. The SimAM non-parametric attention mechanism is introduced to selectively focus on crucial features without introducing additional parameters, effectively suppressing irrelevant information. Residual connections are integrated to prevent local optima, with the Res-SimAM module replacing the standard convolution block in upsampling nodes. In comparison to UNet++, the proposed enhanced model demonstrates significant improvements in UAVid and Potsdam datasets, achieving a 6.77% and 1.79% increase in *mIoU*, 4.71% and 1.17% in *F*1, and 4.99% and 0.98% in *OA*, respectively. A comparative

① 基金项目: 国家自然科学基金面上项目 (62076103); 广东省普通高校特色创新项目 (2022KTSCX035); 佛山市高等教育高层次人才项目 (303480)
蒋春鸿与陈宇宸为共同第一作者.
收稿时间: 2025-01-02; 修改时间: 2025-01-21; 采用时间: 2025-02-18; csa 在线出版时间: 2025-06-13
CNKI 网络首发时间: 2025-06-16

analysis against recent mainstream models underscores its superior performance, positioning ARD-UNet++ as a promising solution for precise urban remote sensing image segmentation.

Key words: SimAM attention mechanism; residual connection; depthwise separable convolution; urban remote sensing image

1 引言

1.1 研究背景

遥感图像的语义分割是遥感图像处理领域的一项关键的垂类应用, 通过将遥感图像中的每个像素分配给特定类别, 精确地描绘和标记高分辨率遥感图像中代表各种物体或物体类别的区域. 遥感图像分割的准确性在土地利用管理、环境监测和城市规划等实际应用中具有至关重要的意义[1].

然而, 随着遥感技术的不断进步, 城市遥感图像呈现出高分辨率、动态背景、复杂纹理信息和密切相关的上下文关系等独特的挑战. 遥感图像在拍摄成像时常常具有亚米级的高分辨率, 能够捕捉到地面上非常精细的细节, 庞大的数据量对模型的计算能力有严格要求; 细节丰富的小目标 (如道路、车辆等) 占据较少的像素, 在分割中常会被忽略或错误分类, 分类结果常被常见类别主导模型的训练而导致稀有类别分类结果准确性较差; 同一类别受地物多样性的影响会呈现出不同的纹理信息, 其中边缘模糊、平滑角对结果准确率影响较大, 需要有更准确处理复杂纹理特征的结构.

尽管已有研究历程对城市遥感图像的语义分割已有不少思路贡献, 但对如何高效识别边界做好图像分割及如何提高准确率的探索仍较为有限, 在处理复杂和动态的城市遥感图像方面仍面临挑战. 早期阶段, 语义分割依赖于像素级分类方法, 利用统计和传统机器学习算法, 如最大似然分类器[2]、支持向量机 (SVM)[3]、决策树[4], 对遥感图像中的单个像素进行分类. 这些方法主要关注光谱信息, 并结合不同土地覆盖类别的统计特征, 通常依赖于基于规则和人工设计的特征. 随后, 深度学习方法的出现标志着遥感图像语义分割的革命性突破. 这些方法, 如全卷积网络 (FCN)[5]、DeepLab[6]和 SegNet[7], 通过自动学习数据中的特征表示来实现图像语义信息的端到端学习. 这些特征通常会在边界分割中产生问题, 如不连续、误分类、遗漏和拓扑结构 (如孔洞) 的异常.

1.2 研究现状

在早期阶段, 城市遥感图像的语义分割严重依赖于传统的机器学习方法, 特别是手动设计特征. 然而, 这些基于像元的传统方法[8]难以充分利用高分辨率遥感图像的特征信息, 且受同物异谱、异物同谱现象影响, 易出现错分、漏分情况.

深度学习的出现引入了 Long 等人[5]提出的全卷积网络 (FCN), 利用了包含全卷积-反卷积的结构. 虽然全卷积组件集成了 VGG[9]、ResNet[10]和 GoogLeNet[11]等经典网络, 但反卷积部分通过上采样生成原始大小的语义分割图像. 虽然 FCN 适用于任意大小的图像, 但涉及多次下采样的深度卷积过程会导致城市遥感图像分辨率降低, 并丢失细节和空间信息. 这种限制可能导致分割结果缺乏细粒度结构.

在 FCN 概念的基础上, Ronneberger 等人[12]提出了 U-Net, 引入了一个编码器-解码器结构, 通过对称的上采样路径来抵消空间信息丢失. 然而, U-Net 缺乏多尺度特征融合机制, 在面对多尺度信息时表现相对较弱. 为了解决这个问题, 在 U-Net 网络模型上做出改进的 UNet++[13]在编码器和解码器中加入了额外的分支连接, 允许在不同尺度上更好地捕获特征. 然而, 由于参数的大量增加, 需要更大的计算资源和存储空间, 使得模型容易受到无关特征的干扰.

FusionU-Net[14]引入了一种多尺度特征融合的替代解决方案, 在编码器中采用从浅到深的相邻特征图之间的特征融合方法, 然后进行反向操作. 这种设计考虑了深层和浅层生成的特征图之间的语义差异. 在特征融合过程中, 不适当的融合可能会对模型产生负面影响, 忽略浅层和深层之间的隐含相关性.

在图像分割领域, 基于特征金字塔结构的特征金字塔网络 (FPN)[15]也发挥了重要作用. 然而, FPN 在特征提取和融合方面面临挑战, 难以聚合大量的判别特征. A^2-FPN[16]提出了一种利用注意力机制指导特征提取和融合的注意力聚合模块 (AAM). 然而, 这种方法并没有从根本上解决特征边缘模糊和平滑角等缺陷.

相比于自然图像的语义分割, 遥感图像语义分割由于有其自身的特殊性和挑战性, 如存在大量微小目标, 对分割方法和结果的精细性提出了更高的要求. 近

年来, 对遥感图像语义分割的研究持续深入, 新模型不断涌现. 张静等人[17]提出多尺度信息融合模型, 在编码阶段基于 DenseNet 网络融合多尺度特征, 解码阶段设计短解码器恢复细节信息, 并采用分层监督机制训练网络. 张哲晗等人[18]设计 SegProNet 网络, 利用池化索引与卷积融合信息, 构建 Bottleneck 层减少参数量, 改进激活函数提升性能, 为后续研究奠定基础. 随后, 梁敏等人[19]提出 SSMSRDA 模型缩小源域和目标域间差异, 高梁等人[20]融合高度信息网络提升性能. 近期, 梁龙学等人[21]提出全局信息重建网络 MAGIFormer, 结合多尺度注意力提取, 编码器引入多尺度注意力骨干, 解码器设计全局多分支局部 Transformer 块和极化特征精炼投, 提高了图像分割的平均交并比. 以上研究均为本研究所述结构网络提供了思路.

目前, 该领域研究虽不断取得进展, 但面对复杂应用场景仍需探索更有效方法. 因此, 在上述研究的基础上, 本研究强调了局部关键信息和上下文细节, 提出了一个基于 UNet++的增强模型, 并命名为 ARD-UNet++ (A: 注意力机制; R: 残差连接; D: 深度可分离卷积), 有效地缓解了边缘模糊和平滑角等问题, 在小感受野内能有效学习到全局信息. “ARD”的名称标识对应于 3 个关键的改进方法, 依次对应于本研究的贡献.

(1) SimAM 注意力机制. 为了解决城市遥感图像在复杂纹理信息和动态背景下对关键特征的关注需求, 本文引入了一种无参数的注意力机制 SimAM[22]. SimAM 使模型能够强调特征图中的重要特征, 在不引入额外推理时间的情况下有效地抑制不相关的特征污染.

(2) 残差连接. 受 ResNet 中的 BasicBlock 残差结构的启发, 我们引入残差连接, 将浅层特征转移到深层. 本研究提出的 PFC 策略块和 Res-SimAM 模块作为图像特征提取模块, Res-SimAM 模块取代上采样节点中的普通卷积块. 这减轻了过度使用注意力机制引起的过拟合问题, 并促进了表层和深层之间的信息融合, 解决了不连续边界分割等挑战.

(3) 深度可分离卷积. 考虑到城市遥感图像中密切的上下文关系和 UNet++的大量参数计数, 我们将传统的 3×3 卷积替换为 7×7 深度可分离卷积. 这种修改使得网络捕获密集包装的特征, 便于有效地提取上下文信息.

在 UAVid 和 Potsdam 数据集上进行的大量实验表明, ARD-UNet++在城市遥感图像分割任务中具有优越的性能. 该模型有效地缓解了边缘模糊和过于平滑的角落等缺陷问题, 为解决城市遥感图像处理的复杂挑战提供了有效的解决方案.

2 基于 UNet++改进的增强模型

本研究提出的增强模型 ARD-UNet++的结构与 UNet++结构类似, 如图 1 所示. 它从上到下分为 5 个阶段, Backbone 骨干代表编码阶段, 其余节点作为上采样节点. 同一层的卷积张量具有相同的维数, 池化层为 2D 最大池化层, 窗口大小为 2×2. 从本质上讲, 模型的网络包括一个编码部分和一个具有密集跳跃连接的解码部分. 在编码阶段, 采用 PFC 策略块和 Res-SimAM 模块精心提取图像信息. 在具有密集跳跃连接的解码部分, Res-SimAM 模块取代普通卷积块, 优先处理关键纹理信息.

2.1 PFC 策略块

如图 1 所示, PFC 策略块的核心元素是深度可分离卷积, 包括两个步骤: 深度可分离卷积和点卷积. 深度可分离卷积的有效性在 Xception[23]和 MobileNet[24]中得到了很好的证明, 已经被广泛采用. 在编码阶段, 用 PFC 策略块替换常规卷积块. 具体来说, 通过输入层对图像进行处理后, 对结果进行深度可分离卷积, 卷积核大小为 7, 输出通道为 64, 填充为 3, 无偏置项. 随后, 我们将输入层添加到深度可分离卷积结果中, 合并后的输出通过批归一化层 (图中为 BN) 和 ReLU 激活函数进行处理, 作为点卷积的输入. 在点卷积中, 我们省略了偏置项, 从而得到 PFC 策略块的最终输出. PFC 策略块定义如下:

$$f_1 = x + D(\sigma\{B(x)\}) \tag{1}$$

$$f_2 = P(\sigma\{B(f_1)\}) \tag{2}$$

其中, f_1表示 PFC 策略块的输出, σ表示 ReLU 激活函数, B表示批归一化层, x表示输入层的输出, D表示深度可分离卷积, P表示点卷积. 这种设计使模型能够在编码部分的开始就彻底提取图像特征, 为后续的 Res-SimAM 模块提供丰富的信息. 与常规的 3×3 卷积相比, 该方法的接受域更大, 可以提取更密集的特征信息, 更好地平衡全局信息, 更适合处理城市遥感数据集. 此外, 它有助于将浅层信息传递到深层, 有效降低过拟合的风险.

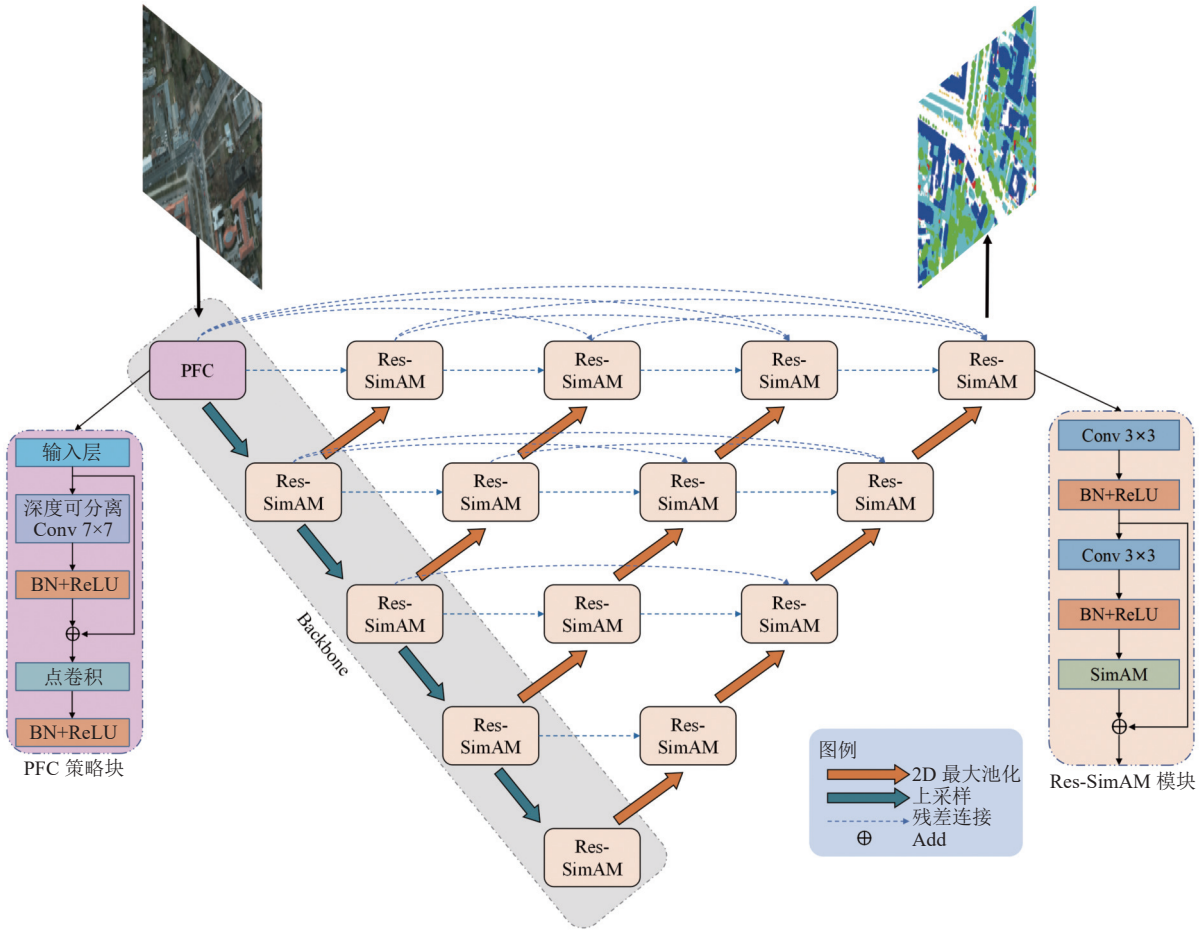

图 1 ARD-UNet++概述

2.2 Res-SimAM 模块

在 Res-SimAM 模块中, 最核心的设计是集成了 SimAM 注意力机制. SimAM 注意力机制是一种三维无参数机制, 其灵感来自于神经科学理论中的能量函数公式. 这种注意力机制在声纹识别领域取得了巨大成功.

如图 1 所示, Res-SimAM 模块由两个连续的卷积块组成. 这种设计选择建立在两个卷积的叠加上, 使网络能够更深入地理解图像. 它有助于在更高的语义级别捕获更多信息, 从而有助于提高分割模型的整体性能. 两个卷积块都包括卷积层、批归一化层和 ReLU 激活函数. 卷积层的核大小为 3×3, 步长为 1, 填充为 1. 关键的区别在于第 2 个卷积块, 其中在最后附加了一个无参数的 SimAM 注意力机制. 将注意力机制之前的卷积张量的输出添加到注意力机制之后的卷积张量中, 并将 Add 方法的特征融合作为 Res-SimAM 模块的输出. Res-SimAM 模块的定义如下.

$$f_1 = \sigma(B\{Conv(x)\}) \tag{3}$$

$$f_2 = S\{\sigma(B\{Conv(f_1)\})\} \tag{4}$$

$$f_3 = f_1 + f_2 \tag{5}$$

其中, f_3表示 Res-SimAM 模块的输出, σ表示 ReLU 激活函数, B表示批归一化层, $Conv$表示 3×3 卷积层, S 表示 SimAM 注意力机制. 在编码阶段, 将下采样节点中的普通卷积块替换为 Res-SimAM 模块. 该模块能够有效地过滤和选择复杂的特征信息, 使模型能够更加专注于特征图中的关键特征, 同时防止模型陷入局部最优. 在解码阶段, 上采样节点中的卷积块遵循 UNet++框架, 保持紧密连接, 且保持与 UNet++一致的上采样过程.

2.3 残差连接操作

残差连接通过在卷积网络中引入跳跃连接, 将输入特征直接加到深层特征的输出上, 形成一种短路路径. 这种设计能够缓解深层网络中梯度消失的问题, 同

时保留浅层特征的信息, 提升模型的表达能力. 假设输入特征为x, 通过两层卷积后提取的特征为$F(x)$, 则残差连接的输出可以表示为:

$$y = F(x) + x \tag{6}$$

其中, $F(x)$表示卷积层提取的非线性特征, 而x是直接通过跳跃连接传递的输入特征. 通过这种结构, 模型不仅能够捕捉深层次的复杂特征, 还能保留输入的原始信息, 增强了特征的多样性. 此外, 残差连接的 Add 方法无需额外的参数或复杂计算, 大大提高了网络的训练效率. 这种操作被用于将模型浅层提取的边缘信息和纹理特征直接传递到深层结构中, 以更好地服务于语义分割任务的像素级分类需求.

3 实验与结果

3.1 实验数据及说明

本研究选用 UAVid 和 Potsdam 两个数据集进行实验结果的评估.

UAVid 数据集专门为城市街道量身定制, 由 University of Twente 于 2020 年发布[25], 具有两种空间分辨率 (3840×2160 和 4096×2160). 它包含 8 个对象类: “Clutter” “Human”“Static_Car”“Moving_Car”“LowVeg”“Tree” “Road”和“Building”. 为了提高实验的严谨性, 我们将训练集和验证集重组成 3 个不同的部分: 训练集、验证集和测试集. 在图 2 中, 官方未标记的测试集仅用于视觉表示.

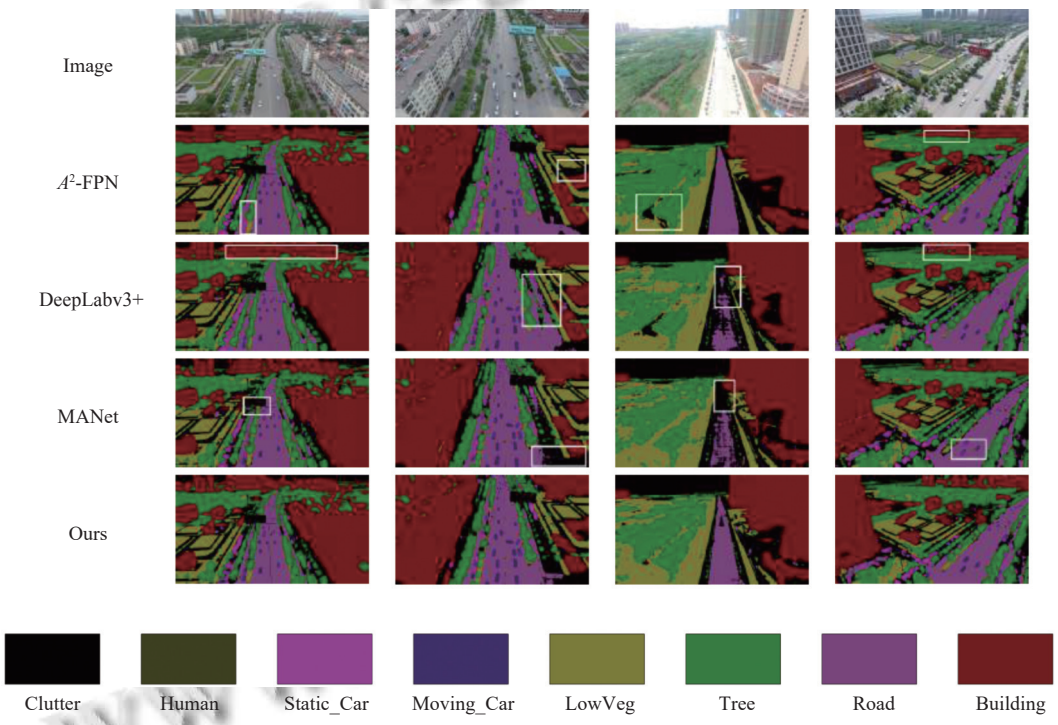

图 2 UAvid 数据集上的分割结果

Potsdam 数据集是另一个城市地区的语义分割数据集, 包括 6 类: “Clutter”“Car”“Tree”“LowVeg”“Building”和“ImSurf”. 它由 38 张图像组成, 每张图像的分辨率为 6000×6000, 并使用分割标签进行注释. 实验过程中, 我们选用 16 张图像用于训练, 7 张用于验证, 15 张用于测试.

3.2 实现细节

实验中使用带有 24 GB VRAM 的 RTX 3090 GPU, 辅以运行在 2.50 GHz 的 Intel Xeon(R) Platinum 8255C CPU 上. 带有 CUDA 加速的 PyTorch 框架促进了有效实验. 为确保可比性, 在训练期间批量大小 (BatchSize) 为 3, 在验证期间批量大小为 2. 对比实验中, 训练时间超过 200 个 epoch, 而消融实验为了快速收敛, 缩短了 40 个 epoch. 采用基本学习率为 6E–4 的 AdamW 优化器, 结合余弦退火策略动态调整学习率. 所有实验采用的损失函数为 UNetFormer 损失[26], 不含辅助损失. 在

实验之前, 对输入图像进行预处理, 包括将图像和相应的标签均匀裁剪到 1024×1024. 在训练过程中, 采用随机垂直翻转、随机水平翻转和随机亮度调整等数据增强技术来增强模型的鲁棒性.

3.3 实验结果与分析

为了全面评估本文模型的性能, 我们将其分别与卷积神经网络 (CNN) 模型[27]、基于注意力机制的视觉变形 (ViT) 模型[28]进行比较. 其中, 选取指标的对应计算公式分别为:

$$mIoU = \frac{1}{N}\sum \frac{TP}{TP+FP+FN} \tag{7}$$

$$F1 = \frac{1}{N}\sum \frac{2TP}{2TP+FP+FN} \tag{8}$$

$$OA = \frac{\sum TP}{P} \tag{9}$$

其中, TP为真阳性数据 (预测为正例实际为正例), TN为真阴性数据 (预测为反例实际为反例), FP为假阳性数据 (预测为正例实际为反例), FN为假阴性数据 (预测为反例实际为正例), N为分割类别总数, P为像素总数.

基于 CNN 的模型包括 MANet、A^2-FPN、FusionU-Net、CMUNeXt 和经典的 DeepLabv3+. 基于注意力机制的 ViT 模型包括 BANet、UNetFormer 和 FTUNet-Former. 表 1 和表 2 分别展示了本文模型在 UAVid 和 Potsdam 数据集上的性能, 显示了其相较于其他模型的优势.

表 1 UAVid 数据集上模型对比实验结果

模型	*mIoU*	*F*1	*OA*
DeepLabv3+[29]	0.4548	0.6845	0.7268
BANet[30]	0.4024	0.5447	0.6796
MANet[31]	0.4628	0.6060	0.7180
A^2-FPN[16]	0.5605	0.6988	**0.7830**
FTUNetFormer[26]	0.5171	0.6605	0.7532
UNetFormer[26]	0.4987	0.6434	0.7355
FusionU-Net[14]	0.5602	0.7050	0.7690
CMUNeXt[32]	0.5203	0.6643	0.7545
本模型	**0.5689**	**0.7105**	0.7773

ARD-UNet++不仅获得了更高的性能分数, 而且还产生了更好的质量结果. 图 2 给出了 UAVid 数据集上的分割结果, 突出显示了 A^2-FPN 对目标边界的不敏感、DeepLabv3+的小目标分割变形问题, 以及 MANet 明显的噪声导致严重的道路脱节和碎片化结果. 相比之下, 我们的模型具有更平滑的分割边缘、更清晰的目标边界, 有效地解决了边缘模糊和平滑角等问题.

表 2 Potsdam 数据集上模型对比实验结果

模型	*mIoU*	*F*1	*OA*
DeepLabv3+	0.6275	0.7687	0.7750
BANet	0.5683	0.7225	0.7187
MANet	0.6448	0.7827	0.7754
A^2-FPN	0.6933	0.8180	0.8134
FTUNetFormer	0.6454	0.7832	0.7738
UNetFormer	0.5627	0.7194	0.7154
FusionU-Net	0.6336	0.7747	0.7679
CMUNeXt	0.7328	0.8444	0.8328
本模型	**0.7341**	**0.8450**	**0.8346**

图 3 显示了 Potsdam 数据集上的分割结果. Fusion-U-Net 在边缘处理方面具有优势, 但也存在目标遗漏问题. CMUNeXt 加剧了这一问题, 导致整个森林、汽车和中型建筑的消失. 相比之下, ARD-UNet++在两个数据集上都展示了更好的分割结果, 显示了更清晰的目标边界, 并减轻了边缘模糊等问题.

表 3 提供了 Potsdam 数据集上实验结果的细化. ARD-UNet++在分割“Building”“Car”和“Clutter”上得分最高. 尽管 CMUNeXt 在对“ImSurf”“LowVeg”和“Tree”的分割中获得了最高分, 但本文模型的第 2 好表现表明了其竞争力. 值得注意的是, 在本文模型表现出色的细分类别中, 它的 *mIoU* 比第 2 好的模型高出约 1%. 对于“Clutter”分割, A^2-FPN 获得第 2 名, 而 CMUNeXt 表现一般, 排名第 4.

表 4 给出了在 UAVid 数据集上的详细实验结果. ARD-UNet++在所有 5 个类别中都获得了最高分, 并在其余 3 个类别中获得了第 2 好的表现. 这些实验突出了我们模型的整体卓越性和稳定性, 将其定位为最优选择.

3.4 消融实验

为了评估我们的模块设计的有效性, 我们进行了消融实验, 深入分析不同模块和策略对模型性能的影响. 这些实验均在严格统一的实验环境设置下进行, 以确保结果的公正性和可比性. 我们通过对比在不同模块组合和改进策略下模型的表现, 从多个角度验证了每个改进的贡献和重要性. 表 5 给出了两个数据集上的消融实验结果. 其中, PFC-L 策略块在 PFC 策略块基础上引入长程注意力机制 SimAM, 通过全局关系建模捕获特征图中的远程依赖关系.

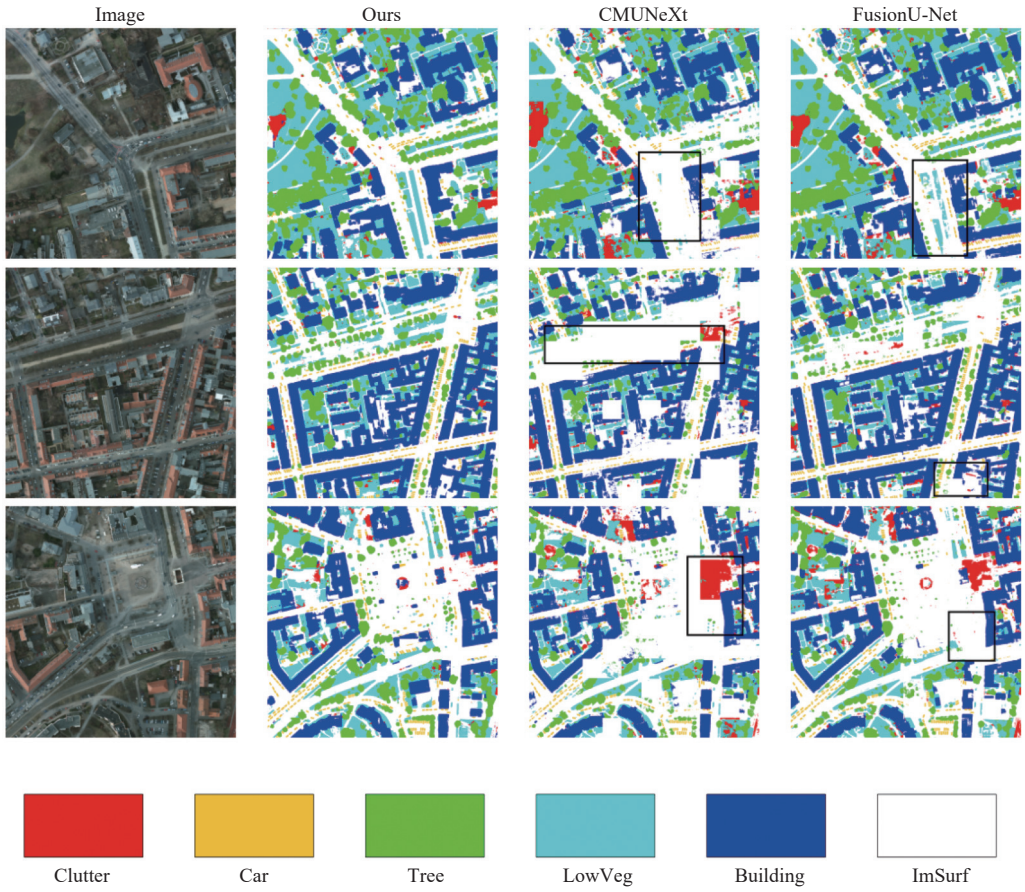

图 3 Potsdam 数据集上的分割结果

表 3 Potsdam 数据集上分数的细化, 使用 *mIoU* 测量

模型	ImSurf	Building	LowVeg	Tree	Car	Clutter
DeepLabv3+	0.6335	0.7515	0.5919	0.6326	0.5281	0.0
BANet	0.5937	0.6353	0.5361	0.4591	0.6172	0.2592
MANet	0.6417	0.7355	0.6088	0.5726	0.6652	0.2621
A^2-FPN	0.6886	0.7846	0.6568	0.6497	0.6870	0.3327
FTUNetFormer	0.6386	0.7315	0.6013	0.5803	0.6754	0.2720
UNetFormer	0.5604	0.6357	0.5289	0.5304	0.5580	0.2619
FusionU-Net	0.6216	0.7207	0.5956	0.5886	0.6414	0.3100
CMUNeXt	**0.7272**	0.8193	**0.6594**	**0.6727**	0.7856	0.2967
本模型	0.7269	**0.8284**	0.6587	0.6704	**0.7860**	**0.3412**

表 4 UAVid 数据集上分数的细化, 使用 *F*1 测量

模型	Building	Road	Tree	LowVeg	Moving_Car	Static_Car	Human	Clutter
DeepLabv3+	0.8076	0.7886	0.7514	0.6554	0.6042	0.5000	0.0	0.6464
BANet	0.7456	0.7280	0.6955	0.6607	0.5536	0.2201	0.2092	0.6115
MANet	0.7707	0.7673	0.7582	0.7013	0.6530	0.3909	0.2009	0.6285
A^2-FPN	**0.8570**	**0.8436**	0.7979	0.7371	0.7384	0.5795	0.3382	**0.7072**
FTUNetFormer	0.8192	0.7861	0.7815	0.7438	0.7067	0.4924	0.2937	0.6629
UNetFormer	0.7939	0.7910	0.7694	0.7023	0.7136	0.4625	0.2708	0.6446
FusionU-Net	0.8557	0.7967	0.7921	0.7521	0.7252	0.5737	0.4394	0.6753
CMUNeXt	0.8350	0.8002	0.7824	0.7210	0.6822	0.5198	0.3092	0.6641
本模型	0.8546	0.8106	**0.8012**	**0.7590**	**0.7470**	**0.5949**	**0.4060**	0.6892

表 5 UAVid 数据集和 Potsdam 数据集的消融实验结果

模型	UAVid			Potsdam		
	mIoU	*F*1	*OA*	*mIoU*	*F*1	*OA*
UNet++	0.4370	0.5865	0.6672	0.7084	0.8271	0.8126
Res-SimAM和UNet++	0.4784	0.2634	0.7189	0.7234	0.8386	0.8232
PFC和UNet++	0.4214	0.5692	0.6623	0.7172	0.8337	0.8213
PFC-L和UNet++	0.5011	0.6453	0.7372	0.7296	0.8401	0.8243
本模型	0.4666	0.6141	0.7005	0.7211	0.8368	0.8206

在 Potsdam 数据集的评估结果中, 与 UNet++相比, Res-SimAM 模块的引入使 *mIoU*、*F*1 和 *OA* 指标分别提高了 6.23%、4.12% 和 3.56%. 如前所述, UNet++容易受到不相关特性的干扰. SimAM 注意力机制的引入增强了模型“过滤”和“净化”不相关特征的能力. 此外, 残差连接的引入防止了过度使用注意力机制, 降低了模型收敛到局部最优的风险. PFC 策略块通过扩大接受野, 从而更好地整合上下文信息并防止过拟合. 仅在编码阶段使用单个 PFC 策略块导致精度提高 0.26%. 为了进一步探索 PFC 策略块的性能, 本研究将编码阶段的所有节点替换为 PFC 策略块. 这将参数数量从原来的 47.2M 减少到 36.1M, 准确度相应提高 14.7%. 这些结果在 UAVid 数据集上得到了类似的验证.

与 UNet++相比, 最终的 ARD-UNet++在 UAVid 数据集上的 *mIoU*、*F*1 和 *OA* 指标分别提高了 6.77%、4.71% 和 4.99%, 在 Potsdam 数据集上分别提高了 1.79%、1.17% 和 0.98%. 这些结果证实了我们设计的模块在提高模型性能方面的有效性.

4 总结

在本研究中, 主要解决了主流模型在处理遥感图像时遇到的边缘模糊、平滑角以及难以捕获远程依赖关系的普遍挑战. 我们提出的解决方案 ARD-UNet++在 UNet++框架上引入了 3 个关键的改进策略. 首先, 将传统的 3×3 卷积替换为 7×7 深度可分离卷积, 减少了模型参数并增强了上下文信息提取. 其次, SimAM 非参数注意力机制的引入使模型更加关注特征图中的关键特征, 在不增加参数数量的情况下有效抑制无关特征污染. 最后, 引入残差连接, 表现在 Res-SimAM 模块和 PFC 策略块中, 作为图像特征提取模块, 前者取代了上采样节点中的标准卷积块.

大量的实验证实了本文模型的有效性, 展示出其在处理图像中与边缘模糊和平滑角相关问题的巨大潜力. 未来, 我们将进一步优化模型结构, 探索更多轻量化和高效的特征提取方式, 同时结合多模态数据和自监督学习, 提升模型在遥感图像处理中的鲁棒性与泛化能力, 以应对更加复杂的实际应用场景. 同时, 将研究关于遥感图像的超分辨率重建, 尝试从低分辨率图像中恢复出高分辨率图像, 提高图像的清晰度和细节表现力, 为城市建设、资源管理等实际应用提供更优质的图像数据和技术支持.

参考文献

1 王玮哲. 高分辨遥感图像目标识别技术综述. 通讯世界, 2017(16): 276–277. [doi: 10.3969/j.issn.1006-4222.2017.16.200]

2 Paola JD, Schowengerdt RA. A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification. IEEE Transactions on Geoscience and Remote Sensing, 1995, 33(4): 981–996. [doi: 10.1109/36.406684]

3 Hearst MA, Dumais ST, Osuna E, *et al*. Support vector machines. IEEE Intelligent Systems and Their Applications, 1998, 13(4): 18–28.

4 Song YY, Lu Y. Decision tree methods: Applications for classification and prediction. Shanghai Archives of Psychiatry, 2015, 27(2): 130–135.

5 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 3431–3440.

6 Chen LC, Papandreou G, Kokkinos I, *et al*. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834–848.

7 Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481–2495.

8 梁艳. 面向对象与基于像素的高分辨率遥感影像分类在土地利用分类中的应用比较 [硕士学位论文]. 太原: 太原理工大学, 2015.

9 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. Proceedings of the 3rd International Conference on Learning Representations. San Diego, 2015.

10 He KM, Zhang XY, Ren SQ, *et al*. Deep residual learning for image recognition. Proceedings of the 2016 IEEE

Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778.
11 Szegedy C, Liu W, Jia YQ, *et al*. Going deeper with convolutions. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 1–9.
12 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention. Munich: Springer, 2015. 234–241.
13 Zhou ZW, Rahman Siddiquee M, Tajbakhsh N, *et al*. UNet++: A nested U-Net architecture for medical image segmentation. Proceedings of the 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018 Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Granada: Springer, 2018. 3–11.
14 Li ZY, Lyu HB, Wang J. FusionU-Net: U-Net with enhanced skip connection for pathology image segmentation. Proceedings of the 15th Asian Conference on Machine Learning. Istanbul: PMLR, 2023. 694–706.
15 Lin TY, Dollár P, Girshick R, *et al*. Feature pyramid networks for object detection. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 2117–2125.
16 Li R, Wang LB, Zhang C, *et al*. A^2-FPN for semantic segmentation of fine-resolution remotely sensed images. International Journal of Remote Sensing, 2022, 43(3): 1131–1155.
17 张静, 靳淇兆, 王洪振, 等. 多尺度信息融合的遥感图像语义分割模型. 计算机辅助设计与图形学学报, 2019, 31(9): 1509–1517.
18 张哲晗, 方薇, 杜丽丽, 等. 基于编码-解码卷积神经网络的遥感图像语义分割. 光学学报, 2020, 40(3): 0310001.
19 梁敏, 汪西莉. 结合超分辨率和域适应的遥感图像语义分割方法. 计算机学报, 2022, 45(12): 2619–2636. [doi: 10.11897/SP.J.1016.2022.02619]
20 高梁, 钱育蓉, 刘慧. 融合高度信息的遥感图像语义分割网络. 计算机工程与设计, 2023, 44(8): 2417–2424.
21 梁龙学, 贺成龙, 吴小所, 等. 全局信息提取与重建的遥感图像语义分割网络. 浙江大学学报 (工学版), 2024, 58(11): 2270–2279, 2319.
22 Yang LX, Zhang RY, Li LD, *et al*. SimAM: A simple, parameter-free attention module for convolutional neural networks. Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021. 11863–11874.
23 Chollet F. Xception: Deep learning with depthwise separable convolutions. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 1251–1258.
24 Howard AG, Zhu ML, Chen B, *et al*. MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861, 2017.
25 Lyu Y, Vosselman G, Xia GS, *et al*. UAVid: A semantic segmentation dataset for UAV imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 165: 108–119
26 Wang LB, Li R, Zhang C, *et al*. UNetFormer: A UNet-like Transformer for efficient semantic segmentation of remote sensing urban scene imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, 190: 196–214.
27 LeCun Y, Bottou L, Bengio Y, *et al*. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278–2324. [doi: 10.1109/5.726791]
28 Dosovitskiy A, Beyer L, Kolesnikov A, *et al*. An image is worth 16x16 words: Transformers for image recognition at scale. Proceedings of the 9th International Conference on Learning Representations. OpenReview.net, 2021.
29 Chen LC, Zhu YK, Papandreou G, *et al*. Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 801–818.
30 Wang LB, Li R, Wang DZ, *et al*. Transformer meets convolution: A bilateral awareness network for semantic segmentation of very fine resolution urban scene images. Remote Sensing, 2021, 13(16): 3065. [doi: 10.3390/rs13163065]
31 Li R, Zheng SY, Zhang C, *et al*. Multiattention network for semantic segmentation of fine-resolution remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5607713.
32 Tang FH, Ding JR, Quan Q, *et al*. CMUNeXt: An efficient medical image segmentation network based on large kernel and skip fusion. Proceedings of the 2024 IEEE International Symposium on Biomedical Imaging. Athens: IEEE, 2024. 1–5.

(校对责编: 张重毅)