#### doi: 10.6046/zrzyyg.2023295

引用格式:刘晨晨,葛小三,武永斌,等.基于混合注意力机制和 Deeplabv3+的遥感影像建筑物提取方法[J].自然资源遥感, 2025,37(1): 31-37. (Liu C C,Ge X S,Wu Y B,et al.A method for information extraction of buildings from remote sensing images based on hybrid attention mechanism and Deeplabv3+[J].Remote Sensing for Natural Resources, 2025, 37(1): 31-37. )

# 基于混合注意力机制和 Deeplabv3+的 遥感影像建筑物提取方法

刘晨晨1, 葛小三1, 武永斌1,2, 余海坤3, 张蓓蓓3

(1.河南理工大学测绘与国土信息工程学院,焦作 454003; 2.河南省测绘地理信息技术中心,
 郑州 450003; 3.河南省遥感院,郑州 450003)

摘要:在大量且复杂的遥感影像中提取建筑物信息是遥感智能应用的重要研究内容之一。针对复杂环境下的遥感 影像建筑物提取不精准及小型建筑物易被忽略等问题,文章提出了一种基于混合注意力机制和 Deeplabv3+的遥感 影像语义分割算法——SC-deep 网络。该网络采用编码-解码结构,利用主干残差注意力网络提取深层特征和浅 层特征,通过空洞空间金字塔池化模块和通道空间注意力模块聚合遥感影像的空间和通道信息权重,有效利用了 遥感影像建筑物的多尺度信息,从而减少影像细节在训练中的损失。实验结果表明,所提方法在 Aerial imagery dataset 数据集上的分割结果均优于其他主流分割网络,能够有效识别并提取复杂建筑物边缘和小型建筑物,表现 出更优异的建筑物提取性能。

关键词:多尺度信息;建筑物提取;语义分割;注意力机制;空洞卷积
中图法分类号: TP 79 文献标志码: A 文章编号: 2097-034X(2025)01-0031-07

## 0 引言

建筑物作为人类生活环境的重要组成部分,体 现了当地社会的经济发展状况,在遥感影像中提取 建筑物区域能够为城市规划、社会经济发展等方面 提供有力依据和基础地表信息,因此精准且高效地 提取建筑物是遥感信息处理与分析的重要研究 内容。

当前,遥感建筑物提取可分为传统方法和基于 深度学习的方法2种类型<sup>[1]</sup>。传统的建筑物提取方 法主要有阈值分割法、随机森林法和支持向量机法 等。吴炜等<sup>[2]</sup>综合分析了光谱和形状特征,提出的 建筑物自动提取方法获取了较高的识别率和较低的 误识别率;贾士军等<sup>[3]</sup>基于统计理论框架利用融合 图像的颜色和纹理信息,实现了图像的分割和建筑 物目标提取;Lagunas等<sup>[4]</sup>先独立提取建筑物的角 点,再基于图像的模式匹配法进行角点匹配,从而实 现建筑物的特征点提取。然而,以上方法都需要一 定人力的特征设计和先验知识,受人的主观认知影 响较大,还存在时间成本大,效率低下,人力财力资 源消耗大等问题,面对广泛复杂的建筑物属性及周 边环境依然有明显的局限性。

近年来,随着深度学习的快速发展,其强大的特征自动提取能力被国内外众多学者所关注,并在图像分类、语义分割和目标识别等领域取得不俗的成果。以深度学习为代表的人工智能新技术为遥感技术开辟了一条新的"学习"渠道,达到了较高的准确度<sup>[5]</sup>。基于全卷积神经网络(fully convolutional networks,FCN)<sup>[6]</sup>的语义分割方法是当今遥感影像建筑物提取的主流,通过对影像中每个像素点逐一划分对应的类别完成对建筑物的自动提取。U-Net<sup>[7]</sup>,Unet++<sup>[8]</sup>,SegNet<sup>[9]</sup>,PSPNet<sup>[10]</sup>,Deeplab系列<sup>[11]</sup>等基于FCN的经典网络也广泛应用于建筑物提取领域,并取得了显著效果。季顺平等<sup>[12]</sup>建立了用于建筑物提取的开源数据集,结合 U-Net 和特征金字塔

收稿日期: 2023-09-22;修订日期: 2024-01-26

基金项目:国家自然科学基金项目"面向矿区地理协同设计的空间信息语义服务模式研究"(编号:41572341)、河南省自然科学基金 项目"深度学习支持下的灾损建筑物提取与检测研究"(编号:222300420450)和河南省高等教育教学改革研究与实践项目 (学位与研究生教育)"面向学科前沿的研究生创新能力提升路径研究与实践"(编号:2021SJGLX100Y)共同资助。

第一作者:刘晨晨(1998-),男,硕士研究生,主要从事遥感影像处理与应用方面的研究。Email: 18203233036@163.com。

通信作者: 葛小三(1971-),男,博士,教授,主要从事时空数据智能处理与分析和地理信息服务方面的研究。Email: gexiaosan@163. com。

网络(feature pyramid network, FPN)提出了 SU-Net, 实现了简单的跨尺度信息聚合,从多个尺度提高了 对建筑物的分析能力; Yang 等<sup>[13]</sup>提出了密集注意 力网络,更好地利用了不同层次的特征,有效提升了 建筑物的提取效果;赵凌虎等<sup>[14]</sup>采用 MobileNetv2 作为改进 Deeplabv3+的主干网络,较好地平衡了提 取精度与速度的矛盾; Xia 等<sup>[15]</sup>基于卷积和 Transformer 构造了双流特征提取网络,实现了从局部和 全局 2 方面对遥感建筑物进行提取; 郭文等<sup>[16]</sup>构 建了多源卫星建筑物提取数据库,并基于注意力增 强的 FPN 实现自动提取; 吕少云等<sup>[17]</sup>以 Unet++网 络为基础,在编码器末端添加了残差结构的空洞卷 积模块,具有较强的泛化能力。

虽然以上方法在遥感影像建筑物提取方面已达 到了较高精度,但对于小型建筑物和复杂场景下的 建筑物仍存在提取不完整,边缘粗糙等问题。针对 上述问题,本文提出了一种基于注意力机制和 Deeplabv3+的建筑物提取网络,结合通道注意力模块和 空间注意力模块对特征图进行多尺度处理,实现对 建筑物分割精度的有效提升。

1 研究方法

#### 1.1 特征提取主干网络

本文将 Deeplabv3+原网络中的 Xception<sup>[18]</sup>特征 提取网络替换为 Resnet50 网络,并在 Resnet50 网络 上添加通道注意力模块,解决因 Xception 网络结构 较深而导致产生大量无效通道及特征图部分细节信 息丢失的问题。Resnet50网络在传统的卷积神经网 络(convolutional neural network, CNN)的基础上引入 批归一化层(batch normalization, BN), 弃用 Dropout, 同时引入2个基本的残差模块(Residual),名称分 别为 Conv Block 和 Identity Block, 解决梯度消失或 爆炸问题。改进后的注意力残差网络在 Resnet50 原有结构上的每一个残差模块后都添加了一个通道 注意力模块(channel attention module, CAM), 过滤 无效的通道信息,如图1所示。其中Conv Block 针 对输入和输出维度不同的特征图,用于改变网络的 维度; Identity Block 针对输入和输出相同维度的特 征图,用于加深网络。本文采用分级输出的方式分 别输出浅层特征图和深层特征图,具体结构如表1 所示,在第二层级卷积后输出浅层特征图,在网络末 端输出深层特征图。



Tab.1 CAM-Resnet50 network structure

层级名称	输入尺寸	操作模块	重复 次数	输出 通道数	
Conv1	$224^2 \times 3$	7×7 卷积	1	128	
C	112 <sup>2</sup> ×64	Conv Block	1	250	
Conv2_x		Identity Block	2	256	
C 2	56 <sup>2</sup> ×256	Conv Block	1	512	
Conv3_x		Identity Block	3		
Comut n	202512	Conv Block	1	1.024	
Conv4_x	28-×512	Identity Block	5	1 024	
Conv5_x	142.1 024	Conv Block	1	2 048	
	14 ×1 024	Identity Block	2		
Conv6	$7^2 \times 2048$	1×1 卷积	1	2 048	

#### 1.2 注意力机制

解码器的特征综合分析能力决定了网络对图像 分割的准确性,受卷积块注意力模块(convolutional block attention module,CBAM)<sup>[19]</sup>的启发,本文提出 了一种 CAM 和空间注意力模块(spatial attention module,SAM),并添加在网络的解码器中,用以加强 全局信息之间的关联,关注特征图中的有效信息,忽 略或剔除无关信息。

1.2.1 通道注意力

CNN 通常进行多次卷积来获取更大的感受野, 为了限制参数量和模型大小,没有学习特征图通道 之间的关系。本文添加了一种轻量级的通道注意力 模块,具体结构如图 2 所示,其中 C, H, W 分别为特 征图的通道维数、高度和宽度。该模块首先对输入 特征图分别进行最大池化和平均池化保持通道维度 不变,实现特征图大小的压缩;然后使用多层感知 机(multilayer perceptron, MLP)模块对其通道进行缩 放,将 2 个输出结果逐元素相加,并通过 Sigmoid 激 活函数后得到通道注意力权重  $M_c$ ;最后计算通道 权重值与输入特征图的内积,从而使输出特征图拥 有通道权重。 第1期



图 2 CAM 模块 Fig.2 Channel attention module

#### 1.2.2 空间注意力

空间注意力是一种专注于目标位置信息的注意 力机制,SAM 模块的结构如图 3 所示。首先,对输 入特征图沿通道维度进行最大池化和平均池化,将 其拼接后进行卷积操作结合 2 种空间特征,然后通 过 Sigmoid 激活函数得到最终的空间注意力权重 *M*<sub>s</sub>。最后将权重值与输入特征图进行内积计算,赋 予其相应的空间权重。



### 图 3 SAM 模块 Fig.3 Spatial attention module

## 1.3 SC-deep 网络

多层卷积网络的运用虽然提取了更多的高级特征,但获取的深层特征由于池化层的存在通常会丢失信息和忽略整体和部分之间的关联,导致对小目标分割不准确和边界粗糙等问题。为解决上述问题,本文提出了一种语义分割网络 SC-deep,网络结构如图 4 所示,其中 rate 代表空洞卷积的采样率。

该网络基于编码-解码(Encoder-Decoder)结构,主要由残差注意力网络、空洞卷积、CAM和SAM模块等组成,使用分级输出保留浅层特征,并与深层特征进行融合,其次使用空洞卷积来增大感受野,从而更好地捕捉目标物体的上下文信息,此外还引入注意力机制来加强对小目标的关注度。



Fig.4 SC-deep network

在编码阶段中,使用残差注意力网络(CAM-Resnet50)作为特征提取模块,该模块旨在提取遥感 影像建筑物的浅层特征和深层特征。其中,浅层特 征包括颜色、纹理、边缘和棱角等信息,而深层特征 则是粗粒度语义等信息。为进一步提升深层特征的 表达能力,在空洞空间金字塔池化(atrous spatial pyramid pooling,ASPP)模块中将不同采样率下生成 的结果进行拼接,并通过卷积层和上采样调整通道 维度和特征图尺寸。

在解码阶段,首先利用 CAM 对深层特征进行加 权处理,以获取具有通道权重信息的深层特征图; 同时,浅层特征经过卷积操作后利用 SAM 在空间位 置上进行加权,以自动捕获重要区域特征,并得到全 局关联的浅层特征图;然后将深层特征图与浅层特 征图进行拼接,得到融合空间和通道注意力权重的 多尺度信息聚合特征图;最后通过卷积和上采样操 作,将聚合的多尺度信息特征图恢复到原始图像大 小。这种编解码结构的设计思路能够有效地提取并 利用遥感影像建筑物的各种特征信息,为图像分割 任务提供有力支撑。

## 2 实验与分析

#### 2.1 数据集与预处理

本文选取武汉大学季顺平教授及其团队提供的 WHU建筑物语义分割数据集中的子数据集 Aerial imagery dataset 作为研究对象。该数据来自新西兰 土地信息服务网站,手动编辑了基督城约 22 000 栋 独立建筑的矢量数据,影像的空间分辨率为 0.3 m。 数据集中含有 8 187 张裁剪后大小为 512 像素×512 像素的遥感图像,剔除无语义信息或信息较少的图 像后,本文选取了 6 455 张图像,并以 7:1:2的比例 将其划分为训练集,验证集和测试集。为提升模型 的鲁棒性,对训练数据集作随机裁剪、缩放、旋转等 操作,结果如图 5 所示。



(a) 原始图像

(b) 原始标签

6 位示 (c) 预处理图像
8 5 图像预处理结果

(d) 预处理标签

Fig.5 Image preprocessing results

#### 2.2 实验环境及配置

为验证本文算法在建筑物提取方面的优越性, 建立对比实验,所有实验均在统一软硬件实验环境 下进行。实验环境如表2所示。

表 2 实验环境配置

Tab.2 Experiment environment configuration

实验环境	配置参数
CPU	Intel(R) Xeon(R) Gold 6330
内存/GB	80
GPU	RTX 3090
显存/GB	24
CUDA	CUDA11.3
学习框架	Pytorch1.10.0
编程语言	Python3.8

在 SC-deep 网络训练过程中,输入为大小 512 像素×512 像素的遥感图像,设置训练样本批次大小 为4,数据进程个数为4,迭代轮数为30,初始学习 率为0.01,最小学习率为1×10<sup>-4</sup>,采用随机梯度下 降(stochastic grident descent,SDG)算法优化器进行 优化模型,同时使用交叉熵损失函数来计算预测值 和真实值之间的损失值。

#### 2.3 评价指标

为对本文建筑物提取所用模型的性能进行定量 描述,实验选取交并比(intersection over union, IoU)、召回率(Recall)、精确率(Precision)和F1分 数(F1-score)4个指标作为评价指标,公式分别为:

$$IoU = \frac{TP}{TP + FP + FN} \quad , \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \quad , \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \quad , \tag{3}$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \quad , \qquad (4)$$

式中: TP 为正确检测出的建筑物像元数; TN 为正确检测出的非建筑物像元数; FP 为错误检测出的 建筑物像元数; FN 为错误检测出的非建筑物像元数。

## 2.4 消融实验

2.4.1 主干网络消融实验

为验证残差注意力网络作为主干网络的优越

性,基于 Deeplabv3+模型的基础框架,使用相同的数 据集和实验环境,对主干网络的选取进行消融实验, 分别使用原论文的 Xception 网络、轻量级 mobilenetv2 网络,视觉注意力 vit 网络和残差注意力 CAM-Resnet50 网络作为主干网络对 Aerial imagery dataset 数据集进行训练,并从 IoU, Precision, Recall 和 F1score 这 4 个指标评判分割结果。消融实验结果如 表 3 所示。

表 3 主干网络消融实验结果

Tab.3 Results of backbone network ablation experiments

				( 70 )
主干网络	IoU	Precision	Recall	F1-score
Xception	84.48	93.20	90.03	92.59
mobilenetv2	83.94	91.33	91.21	91.27
vit	72.58	88.65	80.01	84.11
CAM-Resnet50	88.75	94.86	93.23	94.04

分析表 3 可知,使用残差注意力网络作为主干 网络可以明显提升模型性能。对比原论文 Xception 网络分别在 IoU, Precision, Recall 和 F1-score 方面提 高了 4.27, 1.66, 3.2 和 1.45 百分点。此外与其他主 干网络相比, 残差注意力网络也展现出不同程度的 性能提升, 充分证明了残差注意力网络具有更强的 特征提取能力。

2.4.2 注意力模块消融实验

在网络中,注意力模块的位置会导致分割精度 差异。为确定2种注意力模块的最优位置达到最佳 的分割效果,本文在模型中添加了不同位置的2种 注意力模块,并对模型进行训练。分割结果指标如 表4所示,表中 Deep 和 Low 分别表示编码器输出的 深层特征图和浅层特征图。

表 4 注意力模块消融实验结果 Tab.4 Results of attention module ablation experiments

			(%		
注意力模块位置	IoU	Precision	Recall	F1-score	
ASPP+CAM+SAM	88.82	94.86	93.31	94.08	
ASPP+SAM+CAM	88.52	95.27	92.59	93.91	
SAM+ASPP+CAM	88.76	94.81	93.29	94.04	
SAM+CAM+ASPP	86.36	93.52	91.85	92.68	
Deep+CAM, Low+SAM	88.86	95.05	93.18	94.10	

分析表4可知,由于低特征的图片具有更多的 空间信息,将 SAM 模块放置在浅层特征之后更为合 适。如果放置位置过前,就会导致概括性太大,无法 捕捉一些具体特征,提取的空间注意力由于通道数 较少导致概括性不足,更容易造成负面影响。因此, 在使用 SAM 模块时需要平衡通道数和概括能力之 间的关系。除了空间注意力,通道注意力也是提高 模型性能的重要手段。对于高维特征图更需要通道 注意力去维护图片的上下文关系,但是网络过深时, 通道数过多容易引起过拟合,且特征图太小,使用卷 积操作不当反而会引入大量非像素信息。对于编码 器输出的深层特征,通道数量和特征图大小达到较 好的平衡状态,更适合在其后添加 CAM 模块。

#### 2.5 实验结果分析

为论证 SC-deep 网络结构在建筑物提取方面 的优势,本文基于 Aerial imagery dataset 数据集分别 训练 U-Net,FCN, Deeplabv3+和 SC-deep 这4种网 络模型,并将模型在测试集上进行测试和分割结果 对比,实验中对比网络参数一致且运行环境相同。 如表5所示为模型实际运行的部分提取结果,为直 观展示,将分割结果叠加在原图上并对性能提升部 分局部放大用红框标出,如表6所示为模型分割结 果在 IoU, Precision, Recall 和 F1-score 这4个指标 上的表现。

## 表 5 对比实验分割可视化结果 Tab.5 Comparative experiment segmentation visualization results

序号	真实图像	U-Net	FCN	Deeplabv3+	SC-deep
1	<b></b>				5
2					
3					

表 6 对比实验结果

Tab.6Compare experimental results(%)

模型	IoU	Precision	Recall	F1-score
U-Net	85.13	92.31	91.62	91.97
FCN	87.95	93.66	93.52	93.59
Deeplabv3+	84.48	93.20	90.03	92.59
CS-deep	88.86	95.05	93.18	94.10

根据表 5 显示,针对背景简单的大型建筑物,主 流网络和本文网络都展现了较为精准的提取能力。 然而,在提取小型建筑物时,主流网络会出现提取不 完全和漏检现象。以第 1 行影像为例,在所有主流 网络中都存在着提取不完整,建筑物边界不清晰的 情况。而在处理更小型建筑物时(如第 2 行影像), FCN 网络和 Deeplabv3+网络都出现了漏检的现象。 此外,当提取复杂环境下的遥感影像建筑物(如第 3 行影像)时,U-Net 网络受到树木阴影的影响,导致 建筑物顶部信息提取不完全; FCN 网络和 Deeplabv3+网络则存在细节边界分割不清晰的问题。相 比之下,本文方法的提取结果更接近真实值,对于受 · 36 ·

到各种因素影响的建筑物提取更加准确,并且在处理小型建筑物和复杂场景的细节部分也能得到更精确的分割结果。

由表 6 可知,相较其他 3 种主流分割网络,本文 方法在建筑物提取方面的精度得到进一步提高。与 U-Net 网络相比,SC-deep 网络的提取结果的 *IoU*, *Precision,Recall*和 F1-score 分别提升了 3.73,2.74, 1.56 和 2.13 百分点。与 FCN 网络相比,在 Recall 基 本保持一致的情况下,其他 3 个指标均有不同程度 的提升。与 Deeplabv3+相比,*IoU*, *Precision*, *Recall* 和 F1-score 分别提升了 4.38,1.85, 3.15 和 1.51 百 分点,这说明本文对 Deeplabv3+网络的改进对建筑 物提取性能产生了积极的影响。

综上所述,通过对比评价指标及实验分割结果的分析,可以得出以下结论:本文提出的 SC-deep 网络在 Aerial imagery dataset 数据集上展现出更优秀的分割性能。尤其在小型建筑物的分割方面,其精准度更高。此外,在处理复杂场景下的遥感影像时,该网络也呈现出较为清晰的分割效果。

## 3 结论

建筑物是遥感影像反映地理信息的重要地物目标,利用遥感影像提取建筑物对地表覆盖分类、城市规划、地理信息数据库更新等具有重要意义。本文提出了基于混合注意力机制和 Deeplabv3+的遥感影像建筑物提取网络 SC-deep,在编码阶段引入残差注意力网络,利用 ASPP 模块增大感受野,在解码阶段引入 CAM 和 SAM 模块,从多尺度综合分析建筑物特征,有效地利用了高分辨率遥感影像的多尺度信息。

实验结果表明,本文 SC-deep 网络总体性能均 优于其他主流网络,能够有效增强小型建筑物的提 取效果,改善复杂场景下的遥感影像建筑物的细节 提取。

值得注意的是,尽管本文网络能够对细小目标 进行有效提取,但由于其像素比例小、可利用特征少 等特点,仍存在边缘不清晰等问题,后续的研究将集 中在提高细小目标提取精度上。

#### 参考文献(References):

 [1] 胡明洪,李佳田,姚彦吉,等.结合多路径的高分辨率遥感影像
 建筑物提取 SER-UNet 算法[J].测绘学报,2023,52(5):808-817.

Hu M H,Li J T,Yao Y J,et al.SER-UNet algorithm for building extraction from high - resolution remote sensing image combined with multipath[J].Acta Geodaetica et Cartographica Sinica, 2023, 52(5):808-817.

[2] 吴 炜,骆剑承,沈占锋,等.光谱和形状特征相结合的高分辨率 遥感图像的建筑物提取方法[J].武汉大学学报(信息科学版),2012,37(7):800-805.

Wu W, Luo J C, Shen Z F, et al. Building extraction from high resolution remote sensing imagery based on spatial – spectral method [J]. Geomatics and Information Science of Wuhan University, 2012,37(7):800-805.

- [3] 贾士军,王 昆.融合颜色和纹理特征的彩色图像分割[J].测绘科学,2014,39(12):138-142,147.
  Jia S J, Wang K.Color image segmentation by integrating color and texture features[J].Science of Surveying and Mapping, 2014, 39 (12):138-142,147.
- [4] Lagunas E, Amin M G, Ahmad F, et al. Pattern matching for building feature extraction [J]. IEEE Geoscience and Remote Sensing Letters, 2014, 11(12):2193-2197.
- [5] Gong J, Ji S.Photogrammetry and deep learning [J]. Journal of Geodesy and Geoinformation Science, 2018(1):1-15.
- [6] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4):640-651.
- [7] Ronneberger O, Fischer P, Brox T. U-net; Convolutional networks for biomedical image segmentation [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham; Springer, 2015;234-241.
- [8] Zhuo Z W, Tajbakhsh N, Liang J M, et al.Unet++: A nested U-Net architecture for medical image segmentation [EB/OL]. (2018-09-20). [2022-05-20]. https://arxiv.org/abs/1807.10165.
- Badrinarayanan V, Kendall A, Cipolla R.SegNet; A deep convolutional encoder-decoder architecture for image segmentation [J].
   IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12):2481-2495.
- [10] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network [C]//
   2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. IEEE, 2017;6230–6239.
- [11] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4):834-848.
- [12] 季顺平,魏世清.遥感影像建筑物提取的卷积神经元网络与开源数据集方法[J].测绘学报,2019,48(4):448-459.
  Ji S P, Wei S Q.Building extraction via convolutional neural networks from an open remote sensing building dataset[J].Acta Geodaetica et Cartographica Sinica,2019,48(4):448-459.
- [13] Yang H, Wu P, Yao X, et al.Building extraction in very high resolution imagery by dense-attention networks [J]. Remote Sensing, 2018,10(11):1768.
- [14] 赵凌虎,袁希平,甘 淑,等.改进 Deeplabv3+的高分辨率遥感影 像道路提取模型[J].自然资源遥感,2023,35(1):107-114. doi:10.6046/zrzyyg.2021460.
  Zhao L H,Yuan X P,Gan S, et al.An information extraction model of roads from high-resolution remote sensing images based on improved Deeplabv3+[J]. Remote Sensing for Natural Resources, 2023,35(1):107-114.doi:10.6046/zrzyyg.2021460.
- [15] Xia L, Mi S, Zhang J, et al. Dual-stream feature extraction network

第1期

based on CNN and transformer for building extraction [J].Remote Sensing, 2023, 15(10);2689.

[16] 郭 文,张 荞.基于注意力增强全卷积神经网络的高分卫星影 像建筑物提取[J].国土资源遥感,2021,33(2):100-107.doi: 10.6046/gtzyg.2020230.

Guo W,Zhang Q.Building extraction using high-resolution satellite imagery based on an attention enhanced full convolution neural network[J].Remote Sensing for Land and Resources, 2021, 33(2): 100-107.doi:10.6046/gtzyyg.2020230.

[17] 吕少云,李佳田,阿晓荟,等.Res\_ASPP\_UNet++:结合分离卷积 与空洞金字塔的遥感影像建筑物提取网络[J].遥感学报, 2023,27(2):502-519. Lyu S Y, Li J T, A X H, et al.Res\_ASPP\_UNet++:Building an extraction network from remote sensing imagery combining depthwise separable convolution with atrous spatial pyramid pooling [J].National Remote Sensing Bulletin, 2023, 27(2):502-519.

- [18] Chollet F.Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. IEEE, 2017: 1800– 1807.
- [19] Woo S, Park J, Lee J Y, et al. Cbam; Convolutional block attention module [C]// Proceedings of the European conference on computer vision (ECCV).2018;3-19.

## A method for information extraction of buildings from remote sensing images based on hybrid attention mechanism and Deeplabv3+

LIU Chenchen<sup>1</sup>, GE Xiaosan<sup>1</sup>, WU Yongbin<sup>1,2</sup>, YU Haikun<sup>3</sup>, ZHANG Beibei<sup>3</sup>

(1. School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China; 2. Henan

Surveying and Mapping Geographic Information Technology Center, Zhengzhou 450003, China; 3. Henan

Remote Sensing Institute, Zhengzhou 450003, China)

Abstract: Extracting information about buildings from a large and complex set of remote sensing images has always been a hot research topic in the intelligent applications of remote sensing. To address issues such as inaccurate information extraction of buildings and the tendency to ignore small buildings within a complex environment in remote sensing images, this study proposed the SC-deep network-a semantic segmentation algorithm for remote sensing images based on a hybrid attention mechanism and Deeplabv3+. Utilizing an encoder-decoder structure, this network employs a backbone residual attention network to extract deep – and shallow – layer features. Meanwhile, this network aggregates the spatial and channel information weights in remote sensing images using a dilated space pyramid pool module and a channel-space attention module. These allow for effectively utilizing the multi-scale information of building structures in remote sensing images, thereby reducing the loss of image details during training. The experimental results indicate that the proposed method outperforms other mainstream segmentation networks on the Aerial imagery dataset. Overall, this method can effectively identify and extract the edges of complex buildings and small structures, exhibiting superior building extraction performance.

Keywords: multi-scale information; building extraction; semantic segmentation; attention mechanisms; dilated convolution

(责任编辑:李瑜)

