

DOI:10.19651/j.cnki.emt.2211645

# 融合多尺度卷积和注意力机制的场景提取方法

闫祎巧<sup>1</sup> 王宏生<sup>1</sup> 赵怀慈<sup>2</sup> 刘鹏飞<sup>2</sup>

(1. 沈阳工业大学信息科学与工程学院 沈阳 110870; 2. 中国科学院沈阳自动化研究所光电信息处理重点实验室 沈阳 110016)

**摘要:** 复杂背景下,不同尺度建筑物的特征差异较大,现有算法对多尺度建筑物分割存在分割不均以及误判等问题。为了解决上述问题,本文设计了一种适应多尺度变化的新型网络结构。首先,针对遥感图像场景提分割精度低的问题,引入坐标注意力机制,嵌入到基础网络中增强上下文信息捕获能力,消除噪声的同时增强网络对于空间特征的提取能力。引入了新型递归残差卷积模块,加深网络层次的同时减少信息丢失,提高特征提取效率。最后,在跳跃连接中引入了空洞空间卷积池化金字塔增大网络感受野,增强有效特征,抑制无用特征。设计系统验证模型的实用性。实验结果表明,本文方法在精确率、召回率、F1-score 和 IoU 指标中比 U-Net 网络分别提高了 3.05%、1.56%、1.3%、3.08%。

**关键词:** 语义分割;建筑物提取;遥感图像;注意力机制

**中图分类号:** TP391 **文献标识码:** A **国家标准学科分类代码:** 520.1040

## Scene extraction methods incorporating multi-scale convolution and attention mechanisms

Yan Yiqiao<sup>1</sup> Wang Hongsheng<sup>1</sup> Zhao Huaici<sup>2</sup> Liu Pengfei<sup>2</sup>

(1. School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China;

2. Key Laboratory of Optical-Electronics Information Processing, Shenyang Institute of Automation Chinese Academy of Sciences, Shenyang 110016, China)

**Abstract:** Under complex background, the characteristics of buildings at different scales are quite different, and the existing algorithms have problems such as uneven segmentation and misjudgment of multi-scale building segmentation. To solve the above problems, we design a new network structure adapted to scale changes. Firstly, aiming at the problem of low segmentation accuracy in remote sensing image scenes, we introduce and embed a coordinate attention mechanism in the basic network to enhance the context information capture ability, eliminate noise and enhance the network's ability to extract spatial features. we introduce A new recursive residual convolution module ed to deepen the network layer, reduce information loss, and improve the efficiency of feature extraction. Finally, we introduce a hollow space convolutional pooled pyramid in the hop connection to increase the network receptive field, enhance the effective features, and suppress the useless features. Design the system to verify the usefulness of the model. Experimental results show that the proposed method improves the accuracy, recall, F1-score, and IoU indicators by 3.05%, 1.56%, 1.3%, and 3.08% compared with the U-Net network, respectively.

**Keywords:** semantic segmentation; building extraction; remote sensing image; attention mechanism

## 0 引言

高分辨率遥感图像是农业、自动驾驶、军事侦察、遥感探测目标识别等领域的重要信息来源。对帮助城市建设、灾害管理、城市动态监测、地图服务等<sup>[1-4]</sup>方面具有重要的科研意义和经济价值。

语义分割是场景理解与计算机视觉领域中的研究热点。遥感图像会受到天气环境、分辨率大小、采集角度和高

度影响,往往包含更多的噪音,类间差异较小,类内差异大,背景信息复杂。这些特点导致难以从复杂背景中准确地区分建筑物轮廓,使得遥感图像解译工作难度大。如何准确地提取不同尺度建筑物的模糊边界是一个具有挑战性的任务。Huang等<sup>[5]</sup>根据建筑物光谱结构特性设计了建筑物形态指数。Hu等<sup>[6]</sup>提出了一种基于形态学特征的建筑物自动提取方法。Ok等<sup>[7]</sup>开发了利用建筑阴影信息的建筑物检测方法。上述方法都需要对图像进行复杂变换获得

收稿日期:2022-10-10

建筑物的特征,在不同数据集下,性能有较大差异。

基于深度学习的语义分割方法极大地提高了算法对遥感图像的分割效果。全卷积神经网络 (fully convolutional network, FCN)<sup>[8]</sup> 是将卷积神经网络 (convolutional neural networks, CNN) 中的全连接层替换为卷积层所形成的,全卷积神经网络可以减少冗余计算、提高大范围土地利用覆盖制图的效率。在图像语义分割中取得了理想效果。经典的全卷积神经网络包括 U-Net<sup>[9]</sup>、SegNet<sup>[10]</sup>、DeepLab<sup>[11]</sup> 等模型。受 U-Net 思想的启发,一些方法在网络结构上进行了改进,提出了基于 U-Net 的变体结构, Ji 等<sup>[12]</sup> 使用了原图和双下采样图像同时训练一个联合 U-Net (SiU-Net) 模型,两个网络共享权值,以获得更多尺度的特征,但训练资源是双重的,大大降低了训练效率。在文献[13]中提出了一种基于多条平行路径学习的 MAP-Net 网络提取多尺度建筑足迹和精确边界,以保留多尺度特征的空间位置,并且每个阶段都以固定频率提取高层语义特征。Liu 等<sup>[14]</sup> 引入了一个空间剩余初始模块,减少丢失的全局语义信息,聚合多尺度上下文来捕获不同尺度的建筑。为了保障建筑物的空间连续性和获得准确轮廓,文献[15]提出了一种密集空间金字塔池化分割方法 EU-Net,反向利用焦点损失来抑制地面真实中误差标签的影响,使训练阶段更加稳定。Alshehhi 等<sup>[16]</sup> 使用基于单一补丁的卷积神经网络架构提取特征并融合邻域特征来提高分割精度; Li 等<sup>[17]</sup> 采用了一种稳定的学习策略来训练 GAN 模型,虽然模型在特定数据集上取得了不错的效果,但是训练时间较长,并且模型非常不稳定。Wang 等<sup>[18]</sup> 介绍了一种通过跟踪边缘和线性特

征的半自动建筑提取方法,但是需要特定的已知条件,限制了工程应用。

针对复杂背景下小目标的边界分割不均以及误判等问题,本文提出一种基于坐标注意力和空洞空间卷积池化金字的语义分割方法。首先,将坐标注意力机制融合到 U-Net 网络的卷积层中,捕获更多全局上下文语义信息,适应多尺度变化。其次,针对普通跳跃连接直接将不同层级的特征相连问题,构建了基于空洞卷积的新跳跃连接模块,分别提取不同尺度的网络特征,减小下采样造成的空间信息损失。最后,利用递归残差卷积层进行特征积累,提高网络对边界的识别能力,提高了对多尺度建筑物的分割精度,降低了误判率。并设计系统以便于方法的应用。

## 1 本文方法

### 1.1 坐标注意力

遥感图像蕴含着丰富的空间信息,同时也具有背景复杂、干扰源多的特点,关键是要提高特征提取能力。引入坐标注意力结构 (coordinate attention, CA),既强调位置信息的重要性,又获取长程依赖信息。同时捕获跨通道的维度信息,包含方向感知和位置感知,使得模型更准确地定位并识别到目标区域。此种操作使网络获取更大区域信息的同时,避免了引入大的开销。

坐标注意力机制采用两个步骤进行精确位置信息建模,分别是坐标信息嵌入 (coordinate information embedding, CIE) 和坐标注意力生成 (coordinate attention generation, CAG)。坐标注意力机制示意图如图 1 所示。

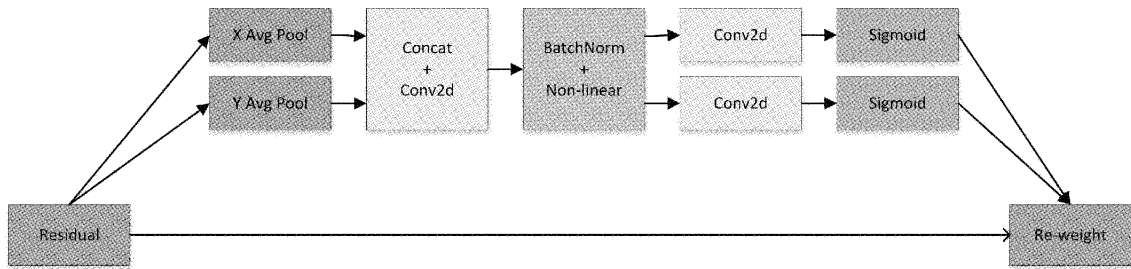


图 1 坐标注意力机制示意图

在 CIE 中,将全局池化拆分成两个特征编码结构。给定输入  $\mathbf{X}$ ,采用两个池化核的空间扩展方式对每个通道沿着水平与垂直两个方向坐标分别编码,输出可以表示为:

$$z_c^h(h) = \frac{1}{w} \sum_{0 \leq i \leq w} x_c(h, i) \quad (1)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq i \leq H} x_c(j, w) \quad (2)$$

两个变换可以沿着两个空间方向提取特征,并生成方向相关特征图。这两种变换使得坐标注意力模块可以更好地捕获远程依赖性,有助于进行更精确地目标定位。CIE 具有全局感受野,同时编码了精确位置信息。

在 CAG 中,对给定通道的 CIE 模块特征进行拼接,

CAG 使用了新的变换,运算所耗资源少,能够适配移动端应用;它可以充分利用所捕获的位置信息,以促进感兴趣区域的精确定位;CAG 模块更有助于捕获通道维度相关性信息,并送入到共享卷积中。得到编码空间信息的中间特征:

$$f = \delta(F_1([z^h, z^w])) \quad (3)$$

$r$  表示用于控制 SE 模块大小的压缩比例参数。然后,将  $f$  沿着空间维度拆分为两个独立张量  $f^h \in \mathbf{R}^{R/r \times B}$ ,  $f^w \in \mathbf{R}^{C/r \times W}$ ; 并采用两个额外的卷积,对其进行匹配为具有与  $X$  相同维度的张量:

$$g^h = \delta(F_h(f^h)) \quad (4)$$

$$g^w = \delta(F_w(f^w)) \quad (5)$$

对输出  $g^h$ , 进行维数扩展并作为注意力权重, 作用于输入得到注意力模块的输出  $Y$ :

$$y_c(i, j) = x_c(i, j) \times g_c^h(j) \quad (6)$$

注意力模块分别对输入的水平 and 垂直两个方向进行注意力特征编码, 有助于反应区域的目标在两个方向上是否存在。这种特性使得注意力机制可以进行更精确的目标定位, 进而有助于更好的识别决策。

### 1.2 空洞空间卷积池化金字塔

空洞空间卷积池化金字塔 (atrous spatial pyramid pooling, ASPP) 对所给定的输入以不同膨胀因子的空洞卷积进行采样, 融合成最终结果。空洞空间卷积池化金字塔在保留空间位置信息的同时, 能够扩大感受野, 捕获多尺度的上下文信息, 相当于以多个比例捕捉图像的上下文, 并融合多尺度信息。此时输入通道为原始输入通道的五倍。为了加强基础网络的信息利用率, 空洞卷积能够在扩大感受野的同时避免特征图分辨率的下降。使用空洞卷积替代传统的下采样操作, 能够保留图像的原始信息, 不同的膨胀因子使得网络具有了多尺度网络辨别空间信息的能力, 能够减少尺度噪声。

### 1.3 密集递归卷积层特性

卷积神经网络 CNN 在特征提取过程中, 易出现网络退化问题, 明显丢失浅层信息, 难以适用于依赖不同尺度获取信息的遥感图像语义分割任务。因此, 设计密集递归卷积层 (dense recurrent convolutional layer, DRCL) 强化网络模型提取多尺度特征的能力, 防止网络退化, 模块主要由递归卷积层 (recurrent convolutional layer, RCL) 模块和 (dense block, DB) 模块构成。

RCL 模块通过合并循环连接到每一个卷积层, 加深网络结构。图 2 左半部分是一个 RCL 单元, 右半部分是 RCL 结构循环操作按时间线的展开图。

$$z_{ijk}(t) = (w_k^l)^T u^{(i,j)}(t) + (w_k^r)^T x^{(i,j)}(t-1) + b_k \quad (7)$$

式(7)为 RCL 某层网络计算方式,  $i, j$  是像素坐标,  $k$  指的是第  $k$  个特征图。  $t = 3 (0 \sim 3)$  指的是一个单元含有  $t$  次递归的卷积层。  $z_{ijk}(t)$  表示在第  $t$  次循环隐藏层的值, 时间第  $t$  步的输入  $u^{(i,j)}(t)$  表示普通反馈网络的输入,  $x^{(i,j)}(t-1)$  表示某一个单元内, 循环输入。在一个 DRCL 中,  $u^{(i,j)}(t)$  是不变的,  $x(t)$  随时间步数在变化。是上一时刻的隐藏层的值传入到下一时刻的隐藏层时的权重矩阵。 DRCL 示意图如图 3 所示。

DB 模块加强了特征重用和特征传递, 缓解了梯度消失又减少网络参数, 保证每一层的特征信息获得充分利用。具体将本层输出与前面所有的输出合并作为下一层的输入, 即:

$$z^l = H_l([z^0, z^1, z^2, z^3, \dots, z^{j-1}]) \quad (8)$$

式(8)中  $[z^0, z^1, z^2, z^3, \dots, z^{j-1}]$  表示将  $0 \sim (j-1)$  层的特征图合并。

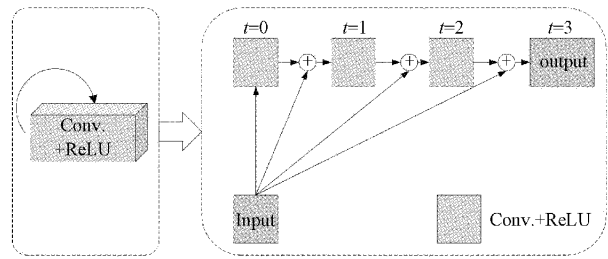


图 2 步长为 3 的 RCL 展开示意图

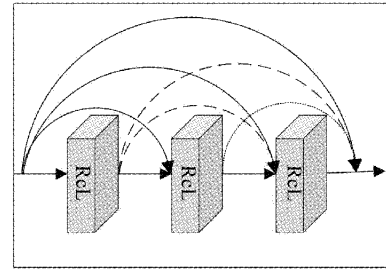


图 3 DRCL 示意图

将 RCL 模块传入 DB 模块中进行特征积累, 确保了更强的特征表示, 并且不会增加网络参数数量, 同时有助于保留遥感图像分割任务必不可少的角点、边缘等低层次特征信息。使用循环结构比传统基于窗口的神经网络噪声低, 能够最大化地提取上下文信息。

### 1.4 本文网络模型

本文提出的网络结构在 U-Net 编码器解码器的基础上进行改进, 如图 4 所示。

输入的遥感图像在编码器中, 首先经过  $3 \times 3$  的卷积层, 获得通道数为 64 的粗特征; 然后经过四层循环坐标注意力、DRCL 和下采样, 其中卷积核大小为  $2 \times 2$ , 对应通道数为  $[64, 128, 256, 512]$ ; 解码器结构与编码器对称, 实现不同模式下的复杂遥感图像语义分割。在上采样和下采样层嵌入坐标注意力模块和 DRCL 模块, 聚焦更丰富的上下文信息, 并提高特征提取精度, 减少特征损耗。原始 U-Net 在每层编码器和解码器之间建立简单的跳跃连接, 网络模型整合多尺度特征信息的能力不足。因此引入 ASPP 结构, 利用空洞卷积保持住不同感受野的信息, 将不同尺度的特征图融合, 充分利用低分辨率特征图的语义信息和高分辨率特征图的空间信息, 最大程度保留特征信息, 膨胀卷积膨胀率分别设置为  $[6, 12, 18]$ 。通过残差连接的方式嵌入 DRCL 模块, 联合密集与残差网络优势, 解决深层网络的梯度消失问题, 并提高网络对多尺度遥感图像重要通道特征的学习能力。

### 1.5 损失函数

网络使用二值交叉熵 (binary cross entropy, BCE) 损失函数来衡量图像语义分割任务中卷积神经网络模型的预测值和实际值之间的相似度。采用了类间竞争机制, 擅长学习类间的信息。

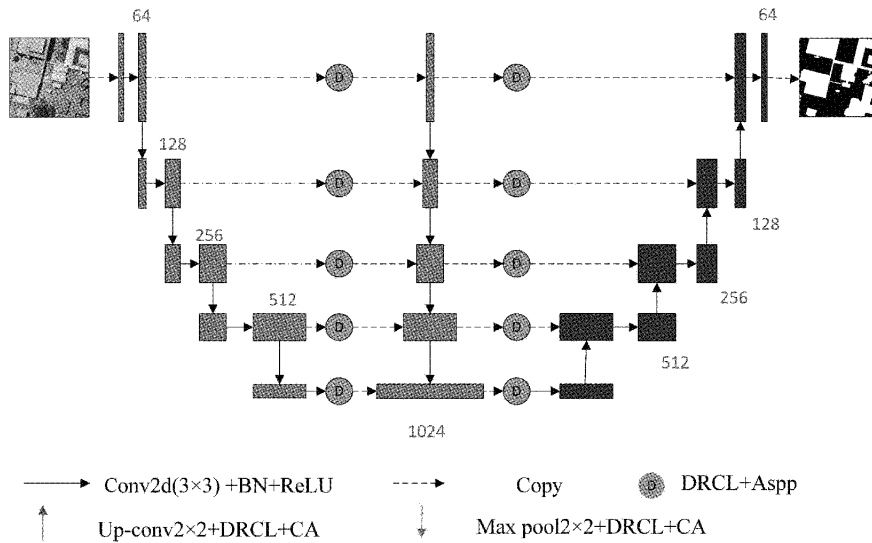


图 4 网络结构示意图

BCE 损失函数的表达式如下:

$$L_{BCE} = -\frac{1}{N} \sum_{i=1}^N (g_i \times \ln(p_i) + (1 - g_i) \times \ln(1 - p_i)) \quad (9)$$

式中:  $g_i$  为像素点  $i$  的真实类别,  $p_i$  为网络对像素点  $i$  的预测概率,  $N$  表示像素点数。

交叉熵作为损失函数,在梯度下降时,可以避免均方误差损失函数学习速率下降的问题。提高像素之间的空间连续性。通过最小化损失函数,引导网络逐步达到收敛状态,减少网络的预测误差,提升卷积神经网络模型的分割性能。

## 2 实验结果与分析

### 2.1 评价指标以及数据集

采用 WHU 航空建筑数据集和 Massachusetts 建筑物数据集对算法的性能进行评估。其中 WHU 航空建筑物数据集包含 8 189 张分辨率为 0.3 m/pixel、尺寸为  $512 \times 512$  的图片。Massachusetts 建筑物数据集由 151 张、大小为  $1\,500 \times 1\,500$  像素的波士顿地区航拍图像组成。首先将原图随机裁剪为  $512 \times 512$  像素的图像块,然后进行水平翻转、旋转等操作,扩展数据量。实验将数据集划分为训练集、验证集和测试集,所占比例分别为 80%、10% 和 10%。

为了准确评估本文算法的分割性能,选取了四项客观评估指标,包括:精确率(precision, Prec)、召回率(recall, Rec)、F1 分数和交并比(intersection over union, IoU)。

### 2.2 实验设置

实验环境配置为 Intel Xeon E5 CPU@2.5 GHz,显卡为 4 块 GeForce RTX 2080Ti 11 GB。本文实验基于 Pytorch 框架上进行实验,批处理设置为 16;初始的学习率

设置为 0.001;权重衰减系数设置为 0.001, epoch = 200。使用二值交叉熵损失函数,当训练损失在固定间隔的迭代次数内保持不变时,采取学习率下降 10% 的动态更新策略,迭代次数间隔设置为 30 次,固定阈值 0.5。

### 2.3 实验结果

#### 1) WHU 数据集分割实验对比

为验证本文所提算法的分割效果,将本文方法与文献[8-10]中的算法在基准数据集 WHU 上进行对比实验,结果如图 5 所示。从图中可以看出,U-Net 网络精细程度低,导致特征提取粗糙,建筑物信息获取不完整;SegNet 可以捕获清晰边缘,但图像纹理细节保留不足且存在错分情况;PSPNet 包含的金字塔池化模块能够保留足够多的深层特征信息,但也存在明显的忽视小目标、大建筑物分割不连续等情况。Deeplab v3+网络依赖金字塔结构,可以相对完整分割出建筑物轮廓,但是误分概率较大;MAP-Net 具有多并行路径学习空间定位结构,重点保留了高层语义特征信息,但融合浅层特征不足,导致部分细节存在错分情况。本文方法在提取信息和多尺度特征融合方面做出了改进,获得了分割边缘明确、纹理细节清晰的分割结果,即使在亮度不均等复杂背景下,仍能准确地提取目标信息。

客观评价指标如表 1 所示,本文方法在 Precoson、F1-score 和 IoU 指标分别为 95.66%、91.62%,均优于其他对比算法,与主观评价结果吻合。其中相比于 U-Net 网络,本文方法的 F1-score 和 IoU 分别提升了 2.51%、5.14%。客观指标评价结果证明,提出的改进方法在建筑物尺度变化、形状不规则以及边缘模糊等复杂场景目标分割任务中性能更加优异。

#### 2) Massachusetts 数据集分割实验对比

图 6 为在 Massachusetts 数据集中分割可视化结果,

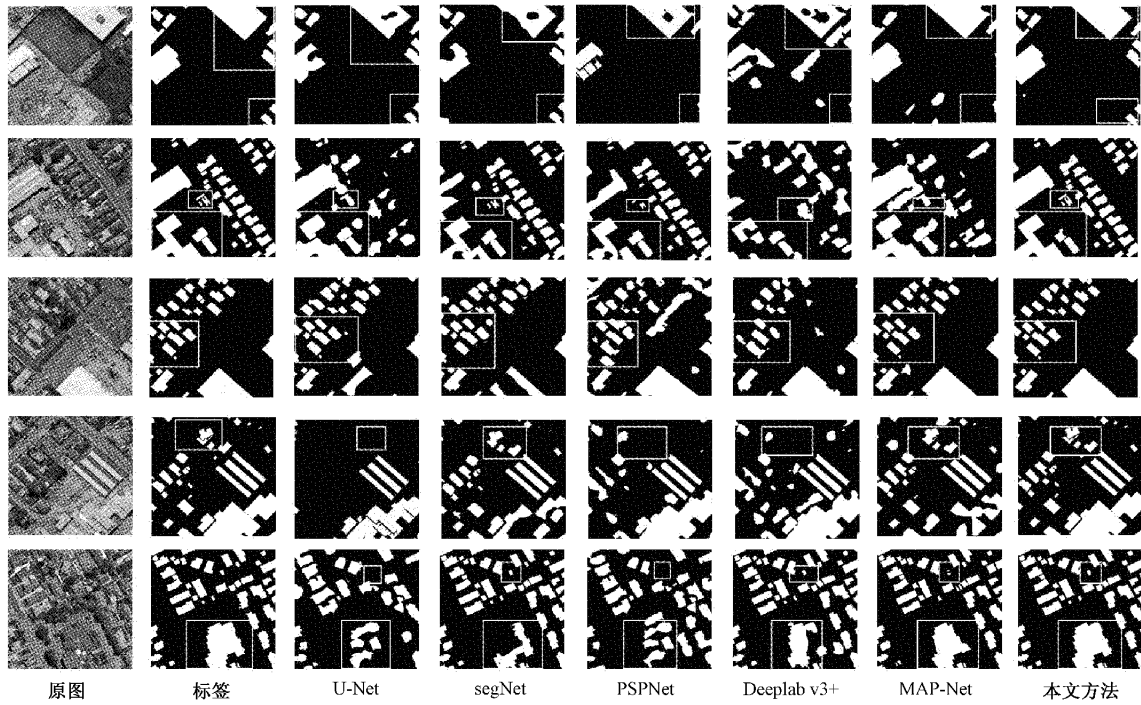


图 5 WHU 数据集对比实验结果

表 1 whu 数据集对比实验客观评价结果

方法	Prec	Rec	F1-Score	IoU
U-Net	0.912 7	0.945 1	0.928 9	0.864 8
SegNet	0.921 4	0.899 5	0.928 4	0.852 8
PSPNet	0.916 7	0.936 9	0.921 0	0.862 3
Deeplab v3+	0.929 4	0.939 2	0.934 3	0.876 7
MAP-Net	0.956 2	<b>0.948 1</b>	0.952 1	0.908 6
本文方法	<b>0.956 6</b>	0.930 8	<b>0.954 0</b>	<b>0.916 2</b>

可以看出,U-Net 和 SegNet 方法都是端到端的编码器-解码器结构,对空间位置不敏感,不利于对分类的细节刻画。对于图中建筑物的边缘特征提取不够完整。Deeplab v3+网络采用改进的 Xception 作为骨干网络,提高了分割效果。未能聚合不同区域的上下文信息,对位置相近的房屋等无法进行很好的进行区分。EU-Net 方法引入 DSPP 模块关注多尺度建筑物的提取,存在纹理细节不丰富的分界处分割效果不好的问题。本文方法通过引入坐标注意力机制,注意全局信息,可以准确分割出图像中位置相近的小目标建筑物和其他连续大型建筑物,纹理细节清晰,轮廓连续且完整,不存在大面积的过分割问题,整体结果最接近标签数据,分割结果可以直接应用于建筑物提取领域。

表 2 为定量比较结果,P1 表示原 U-Net 的分割结果。P2 表示 SegNet 的分割结果。P3 表示 Deeplab v3+ 分割结果。P4 表示 EU-Net 的分割结果。P5 表示本文方法。可以看出,本文方法在精确率、F1 指数、IoU 上均取得了最

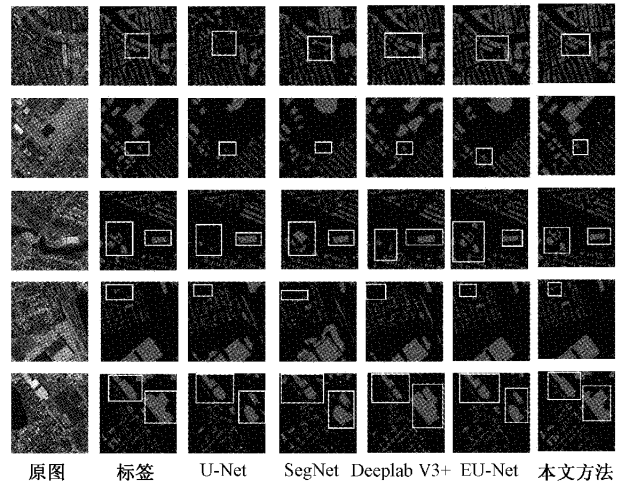


图 6 Massachusetts 数据集对比实验分割结果图

优,仅在 Recall 指标上略低于 E-UNet 算法。此外,相较于 deeplabv3+ 算法,因增加了网络复杂度,本文算法在 Precision 和 F1-Sore 获得较明显提升。

表 2 Massachusetts 数据集对比实验客观结果

方法	Prec	Rec	F1-Score	IoU
P1	0.855 1	0.808 7	0.831 2	0.711 2
P2	0.884 2	0.683 2	0.785 1	0.655 3
P3	0.847 5	0.787 7	0.815 7	0.692 3
P4	0.867 0	<b>0.834 0</b>	0.850 1	0.739 3
P5	<b>0.885 6</b>	0.824 3	<b>0.854 2</b>	<b>0.742 0</b>

为了进一步验证本研究所提出网络结构的性能,将本文的分割结果与同类的分割模型进行比较。由表 3 可知,与对比网络相比,本文扩大了感受野的同时加强位置语义信息和空间语义信息的处理,各类信息相互补充,提升分割性能。

表 3 不同算法在 Massachusetts 数据集的性能评价

方法	Prec	Rec	F1-Score	IoU
SiU-Net <sup>[12]</sup>	0.681 0	0.746 0	0.712 0	0.552 0
Joint-Net <sup>[19]</sup>	0.862 1	0.812 9	0.836 8	0.719 9
HRNet <sup>[20]</sup>	0.858 2	0.820 1	0.838 7	0.722 2
MA-FCN <sup>[21]</sup>	0.870 7	<b>0.828 9</b>	0.849 3	0.738 0
本文方法	<b>0.885 6</b>	0.824 3	<b>0.854 2</b>	<b>0.742 0</b>

由于遥感图像特征信息丰富,文献[12]采用了联合训练方法模型,两个网络共享权值。但同时也降低了训练效率。文献[19]、[20]对网络模块进行了优化,提高了分割效果,但使用的卷积层较浅,不可避免的丢失了图像中深层的语义信息。文献[21]更注重细节的刻画。本文通过构建结合坐标注意力与带有空洞池化金字塔的递归残差卷积层的分割网络。充分利用空间和通道信息,充分利用显著信息和非显著信息,扩大感受野的同时减少漏分错分。除文献[21]在召回率略高于本文,本文其他指标均高于其他文献。

### 3) 消融实验

为了充分说明本文提出的各个模块的有效性,表 4 展示了在 Massachusetts 数据集上依次进行结构改进后的结果。N1 表示原 U-Net 的分割结果。N2 表示基础网络+DCRL+ASPP 的分割结果。N3 表示基础网络+DCRL+CA 分割结果。N4 表示基础网络+CA+ASPP 的分割结果。N5 表示本文的分割结果。

表 4 消融实验客观评价结果

方法	Prec	Rec	F1-Score	IoU
N1	0.855 1	0.808 7	0.831 2	0.711 2
N2	0.864 9	0.809 7	0.838 2	0.728 0
N3	0.879 6	0.812 2	<b>0.851 3</b>	0.739 4
N4	0.885 3	0.816 2	0.849 8	0.738 2
N5	<b>0.885 6</b>	<b>0.824 3</b>	0.844 2	<b>0.742 0</b>

可以看出,当加入坐标注意力块之后,充分整合了上下文信息,准确率 N5 较 N2 了 2.07%。引入 Aspp 模块后,加强了对建筑物目标的多尺度特征提取;加入 DRCL 模块优化网络后,更深层次的卷积结构能够更全面的处理语义信息,有效区分建筑物与其他背景区域,提升分割效果;本文提出的方法的精确率、召回率、F1 指数和 IoU 每个指标较 U-Net 网络分别提高了 3.05%、1.56%、1.3%和

3.08%,整体分割性能优于原始网络。

### 4) 应用

由于天气、地理位置、光照等条件不同,不同数据集图片存在差异较大。为获得更好的分割结果,设计系统调用本文模型。系统功能主要包含两方面,即支持用户训练新的遥感图像数据集,根据需求设置批量大小、学习率等相关参数,调用模型对遥感图像进行语义分割。以及调用训练好的遥感图像语义分割模型进行分割,以快速的获得分割结果。

对系统的各项功能进行测试,已经训练好的数据集,能够实现图像快速语义分割。同时支持新数据集的训练。通过测试,系统能够提供满足需求的结果。证明了算法的可行性和实用性。

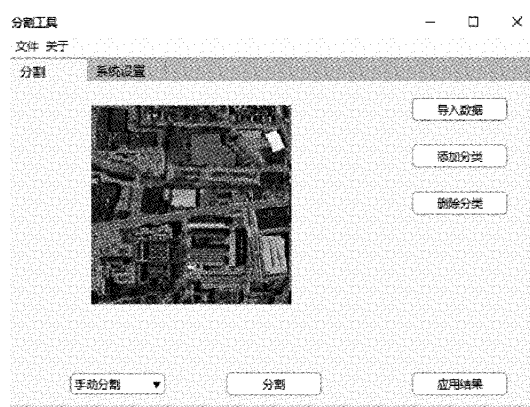


图 7 遥感图像分割系统

## 3 结 论

本文针对多尺度遥感图像建筑物难以准确分割的问题,提出了一种融合多尺度卷积和坐标注意力机制的场景提取方法。首先,引入空洞池化金字塔的递归残差卷积层,提升感受野,提高分割性能并减少空间细节损失;其次,在网络的上下采样过程中,加入坐标注意力模块,利用通道相关性与远程依赖性,充分获取上下文语义信息和全文语义信息。通过在多个数据集上的实验结果证明,改进方法在边缘、细节纹理和客观指标数据上均达到了更优的效果,具有更准确的图像分割精度。为验证本文的实用性,设计系统支持模型训练以及快速分割,以便实际应用。另外,本文在运行效率和网络泛化能力等方面仍存在部分提升空间,在未来的研究中将进一步完善。

### 参考文献

[1] 王溪波,曹士彭,赵怀慈,等. 双边特征聚合与注意力机制点云语义分割[J]. 仪器仪表学报, 2021,42(12): 175-183.

[2] BOUKRI M, FARSI M N, MEBARKI A, et al. Seismic vulnerability assessment at urban scale: Case

- of algerian buildings [J]. International Journal of Disaster Risk Reduction, 2018, 31: 555-575.
- [3] LI X, LI Z, YANG J, et al. Spatiotemporal characteristics of earthquake disaster losses in China from 1993 to 2016 [J]. Natural Hazards, 2018, 94(2): 843-865.
- [4] ZHANG B, CHEN Z, PENG D, et al. Remotely sensed big data: Evolution in model development for information extraction[J]. Proceedings of the IEEE, 2019, 107(12):2294-2301.
- [5] HUANG X, ZHANG L. Morphological building/shadow index for building extraction from high-resolution imagery over urban areas[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2011, 5(1): 161-172.
- [6] HU L, ZHENG J, GAO F. A building extraction method using shadow in high resolution multispectral images[C]. 2011 IEEE International Geoscience and Remote Sensing Symposium, IEEE, 2011: 1862-1865.
- [7] OK A O, SENARAS C, YUKSEL B. Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2012, 51(3): 1701-1717.
- [8] LONG J, SHELHAMER E, DARREL T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014,39(4):640-651.
- [9] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [C]. Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, IEEE,2015:234-241.
- [10] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(12): 2481-2495.
- [11] CHEN L C, PAPANDEOU 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] JI S, WEI S, LU M. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 57 (1): 574-586.
- [13] ZHU Q, LIAO C, HU H, et al. MAP-Net: Multiple attending path neural network for building footprint extraction from remote sensed imagery [J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(7): 6169-6181.
- [14] LIU P, LIU X, LIU M, et al. Building footprint extraction from high-resolution images via spatial residual inception convolutional neural network [J]. Remote Sensing, 2019, 11(7): 830.
- [15] KANG W, XIANG Y, WANG F, et al. EU-Net: An efficient fully convolutional network for building extraction from optical remote sensing images [J]. Remote Sensing, 2019, 11(23): 2813.
- [16] ALSHEHHI R, MARPU P R, WOON W L, et al. Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2017, 130: 139-149.
- [17] LI X, YAO X, FANG Y. Building-A-Nets: Robust building extraction from high-resolution remote sensing images with adversarial networks[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11:3680-3687.
- [18] WANG D, SONG W D. A method of building edge extraction from very high resolution remote sensing images[J]. Environ. Prot. Circ. Econ, 2009, 29: 26-28.
- [19] ZHANG Z, WANG Y. JointNet: A common neural network for road and building extraction[J]. Remote Sensing, 2019, 11(6): 696.
- [20] SUN K, ZHAO Y, JIANG B, et al. High-resolution representations for labeling pixels and regions [J]. ArXiv Preprint, 2019, ArXiv:1904.04514.
- [21] WEI S, JI S, LU M. Toward automatic building footprint delineation from aerial images using CNN and regularization [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 58(3): 2178-2189.

### 作者简介

闫祎巧, 硕士研究生, 主要研究方向为图像处理等。

王宏生, 硕士, 副教授, 主要研究方向为智能搜索引擎, 人工智能。

赵怀慈, 博士, 研究员, 博导, 主要研究方向为图像处理、复杂系统建模与仿真技术, 指挥、控制、通信与信息处理技术。  
E-mail: hczhao@sia.cn