

多尺度特征增强的 ResUNet+ 遥感影像建筑物提取

罗松强, 李浩*, 陈仁喜

河海大学地球科学与工程学院, 江苏 南京 211100

摘要 针对 ResUNet 网络从遥感影像中提取小型和不规则建筑物时存在分割精度低和边界粗糙的问题, 提出一种多尺度特征增强的残差 U 形网络 ResUNet+。该网络以 ResUNet 网络结构为基础, 在编码器内添加特征压缩激活模块以提升网络对有效特征的学习能力, 在编码网络的最后一层使用空洞空间金字塔池化模块来获取不同尺度的建筑物上下文信息。在两个广泛公开使用的 WHU 航空图像数据集和 INRIA 建筑数据集上进行实验, 并将其与 SE-UNet、DeepLabv3+、DenseASPP 和 ResUNet 语义分割网络进行对比。实验结果表明, ResUNet+ 在精确率、召回率和 F_1 分数 3 项精度指标中均表现最优, 对测试影像中大小各异和形状不规则的建筑物具有更精确的分割结果。

关键词 遥感; 建筑物提取; 残差网络; 空洞卷积; 多尺度特征增强

中图分类号 P237

文献标志码 A

doi: 10.3788/LOP202259.0828007

Building Extraction of Remote Sensing Images Using ResUNet+ with Enhanced Multiscale Features

Luo Songqiang, Li Hao*, Chen Renxi

School of Earth Science and Engineering, Hohai University, Nanjing, Jiangsu 211100, China

Abstract We propose ResUNet+, an enhanced multiscale features residual U-shape network, to address issues in the extraction of small and irregular buildings from remote sensing images using the ResUNet, such as low segmentation accuracy and rough boundaries. Based on the ResUNet architecture, the squeeze and excitation module is used in the encoder to improve the network's ability to learn effective features, and the atrous spatial pyramid pooling module is selected as the last layer of the encoding network to obtain context information of buildings at various scales. We evaluate the proposed ResUNet+ and compare it with SE-UNet, DeepLabv3+, DenseASPP, and ResUNet semantic segmentation networks on two commonly used public datasets: the WHU Aerial Imagery Dataset and INRIA Buildings Dataset. The results of the experiments show that ResUNet+ outperforms other networks in terms of precision, recall, and F_1 -score. The segmentation results also show that ResUNet+ excels at extracting buildings of various sizes and irregular shapes.

Key words remote sensing; building extraction; residual network; atrous convolution; multiscale features enhancement

1 引言

遥感影像是获取建筑物信息的重要途径之一。

实现对遥感影像中建筑物的精确提取, 在城市规划、灾害评估、军事侦察及导航等领域具有极其重要的意义^[1]。但从高分辨率遥感影像中高精度自动

收稿日期: 2021-08-12; 修回日期: 2021-09-30; 录用日期: 2021-10-19

基金项目: 中国科学院太空应用重点实验室开放基金(LSU-KFJJ-2018-10)、国家自然科学基金(41471276)

通信作者: lihao@hhu.edu.cn

化提取建筑物信息仍是一项艰巨的任务,其主要原因包括:1)遥感影像中存在噪声阴影和地物遮挡等问题;2)建筑物往往由不同的材料建造而成,在外观、大小、形状和纹理方面存在多样性;3)一些其他地物表现出与建筑物相似的特征,如水泥道路等。这些都增加了建筑物提取的难度^[2]。

传统遥感影像提取建筑物的方法主要依赖于人工设计的特征,如局部特征(边缘角点^[3]等)、形态学特征^[4]、光谱特征^[5]等。这些方法虽然在建筑物提取方面取得了较高的精度,但其特征表达能力有限,往往适用于特定小范围的高分辨率影像。另外,对于不同时间和空间的影像不能鲁棒地适应,无法满足复杂场景下遥感影像建筑物精确提取的需求。

近年来,以卷积神经网络为代表的深度学习技术能够自动学习并生成复杂度极高的非线性特征,在图像分割领域得到了广泛的应用^[6]。目前,用于从高分辨率遥感影像中提取建筑物的大多数网络是全卷积神经网络的变体。全卷积神经网络是 Long 等^[7]提出的一个应用于图像语义分割的端到端深度学习网络,其基本思想是用卷积层替换全连接层,以便在最终的特征图中包含位置信息。此外,为了提高特征图的空间分辨率,卷积神经网络的最后一层被上采样到与输入图像相同的大小,但图像在下采样过程中会丢失空间细节,故卷积神经网络用于建筑物提取时其分割精度还有待进一步提高。一些研究者提出了基于编码器和解码器的 UNet^[8]和 SegNet^[9]网络,通过编码器和解码器之间的多个跳跃连接,将卷积网络提取到的浅层和深层特征进行融合,改善了建筑物提取的结果。刘浩等^[10]在 UNet 网络基础上提出了 SE-UNet,通过在编码器内使用特征压缩激活(SE)模块来增强网络的特征学习能力,同时在解码的过程中通过连接编码特征实现对空间细节信息的恢复,提高了网络对建筑物的分割精度。文献^[11]表明更深层次的网络能够提升模型的表达能力,获得更好的分割结果。但当网络的层数过多时,模型训练的准确率快速下降,网络出现“退化”问题。He 等^[12]提出深度残差网络 ResNet,解决了深层网络的退化问题并使网络更易于优化,利用更深层的网络实现了更复杂特征的提取。朱淑鑫等^[13]通过在 ResNet 网络中引入注意力机制来提升网络对目标对象的特征表示能力,实现了对遥感影像中复杂场景的正确分类。刘

亦凡等^[14]在 UNet 网络的编码器和解码器中引入残差模块,构建了 ResUNet 网络,该网络能有效提取建筑物的多级特征和边缘信息,但并没有对影像进行多尺度特征提取。由于没有足够的多尺度上下文信息,因此对不同尺度下建筑物的提取能力较弱,不能完整地得到建筑物的分割结果。尤其是在大规模提取建筑物的情况下,分割精度较低。罗李焱等^[15]基于 DeepLabv3+ 网络对遥感影像中的建筑物进行提取,通过空洞空间金字塔池化(ASPP)模块获得图像的多尺度上下文信息,该网络适用于复杂场景下的建筑物分割,但其提取到的多尺度特征中包含无效特征,未能充分利用有效特征。

尽管上述研究都能实现对高分辨率遥感影像中建筑物的提取,但在提取形状不规则或尺寸较小建筑物上其提取精度仍然有提升的空间^[16]。本文对 ResUNet 网络的改进点主要有以下两个:1)在 SE 模块中并行添加最大池化层,以提升网络对建筑物突出特征的保留能力,并将其与编码器中的残差模块结合在一起,对每一个卷积得到的特征进行压缩激活处理,以促进有效特征的传播,抑制无用特征;2)在编码网络的最后一层中引入 ASPP 模块,并将前一层的输出作为后一层的输入,以更密集的方式连接一组空洞卷积,最终获得更大范围的感受野,提高了网络对小型及不规则建筑物的分割精度。

2 原理与方法

2.1 残差模块

残差神经网络^[17]是由一系列的残差模块组成,其通过引入恒等映射实现跳层连接,以促进网络的训练和缓解网络的退化问题。每个残差模块包含两个卷积块和一个恒等映射,如图 1 所示。其中卷积块由批归一化层(BN)、线性整流(ReLU)函数和 3×3 卷积层组成;恒等映射将上一层的特征信息与本层处理后的特征信息进行相加并进行线性激活。残差模块的计算公式为

$$\mathbf{X}_{l+1} = \delta(H(\mathbf{X}_l, \mathbf{W}_l) + \mathbf{X}_l), \quad (1)$$

式中, \mathbf{X}_l 和 \mathbf{X}_{l+1} 分别是第 l 个残差单元的输入和输出特征, \mathbf{W}_l 是一组与第 l 个残差单元相关的权重,

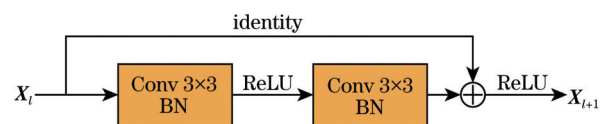


图 1 残差模块

Fig. 1 Residual module

$H(\cdot, \cdot)$ 代表残差函数, $\delta(\cdot)$ 为ReLU函数。

2.2 压缩激活模块

对于卷积神经网络编码过程中提取到的特征, 注意力机制能够增大依赖性较强的特征权重, 降低噪声干扰, 提高对有效信息的利用率, 在图像分割中

已得到广泛的应用。Hu等^[18]提出的SE模块是注意力机制的一个分支, 其主要思想是通过计算不同通道的特征权重, 进而有效地构建通道之间的相互依赖关系。SE模块由压缩模块和激活模块两部分组成, 如图2所示, 可以直接与残差网络一起使用。

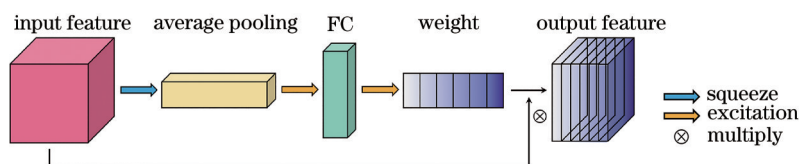


图2 压缩激活模块

Fig. 2 Squeeze and excitation module

SE模块的输入是各个编码模块的输出特征图 $F \in \mathbb{R}^{c \times h \times w}$ 其中 c, h, w 分别为输出特征图的通道数、高度和宽度。首先压缩模块对 F 进行全局平均池化得到 $F_{Avg} \in \mathbb{R}^{c \times 1 \times 1}$, 随后激活模块使用两个全连接层进一步对 F_{Avg} 进行压缩和重构, 经Sigmoid函数得到各通道的权重向量 $W \in \mathbb{R}^{c \times 1 \times 1}$, 最后将 F 与 W 两个矩阵的对应元素相乘, 得到对各个通道赋予不同权重的输出特征图 $F' \in \mathbb{R}^{c \times h \times w}$ 。其中全连接层用于捕获非线性跨通道交互信息, 为了降低计算复杂度, 首先将通道数压缩为原来的 $1/r$, 再将通道维度恢复为原本的大小, 在本文中令 $r=8$ 。SE的计算公式为

$$F' = \left[\sigma \left(W_2 \left(\delta \left(W_1 \left(\text{AvgPool}(F) \right) \right) \right) \right) \right] \otimes F, \quad (2)$$

式中: F 和 F' 分别为输入和输出特征图; $\text{AvgPool}(\cdot)$ 为全局平均池化函数; $W_1(\cdot)$ 和 $W_2(\cdot)$ 分别代表

两个全连接函数; $\sigma(\cdot)$ 为Sigmoid函数; \otimes 为哈达玛积。

2.3 空洞空间金字塔池化模块

针对不同尺度下的目标分割问题, 多尺度的上下文信息有助于解决歧义情况, 并能获得更可靠的分割结果。与标准卷积运算相比, 空洞卷积能够在不增加参数数量的情况下实现更大的感受野, 因此可以获取更高级别的语义信息。空洞卷积引入了一个“扩张率”(DR)的新参数, 该参数定义了卷积核处理数据时各值的间距。设置不同扩张率时, 感受野也会不同。DeepLab^[19]基于空洞卷积提出了ASPP模块, ASPP通过将具有不同扩张率的空间卷积生成的特征图进行级联, 以使输出特征图具有多尺度信息并最终提升网络性能。ASPP包括1个 3×3 普通卷积层、3个 3×3 空洞卷积层(DR分别取6、12、18)和平均池化层, 同时每个卷积层之后加

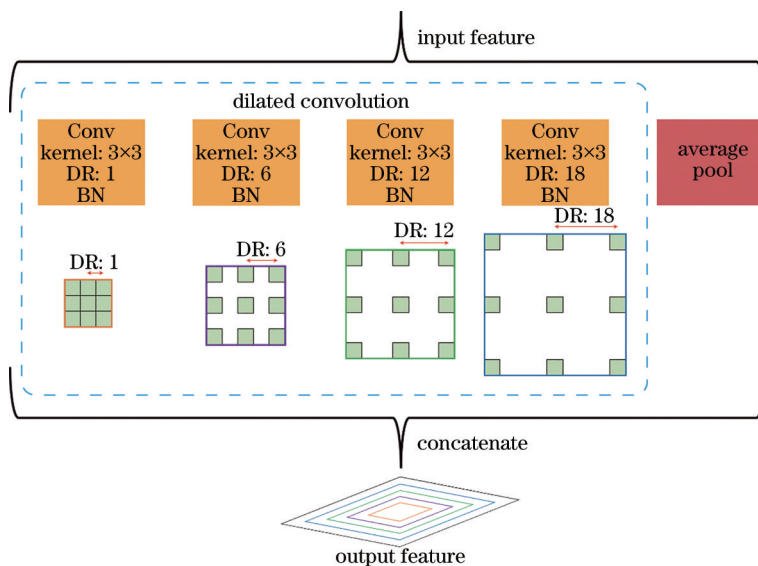


图3 空洞空间金字塔池化模块

Fig. 3 Atrous spatial pyramid pooling module

入批归一化层以提高网络的训练精度,如图 3 所示。

ASPP 能有效地提取出不同尺度目标及其边缘,较好地解决了高分辨率遥感影像中存在的多尺度目标问题,其计算公式为

$$Y = \text{Concat}(\text{AvgPool}(X), H_{1,3}(X), H_{6,3}(X), H_{12,3}(X), H_{18,3}(X)), \quad (3)$$

式中: X 和 Y 分别为输入和输出特征图; $\text{Concat}(\cdot)$ 表示对特征图进行第 1 维度上的拼接操作; $H_{r,n}(\cdot)$ 表示扩张率为 r , 卷积核大小为 n 的带孔卷积。

2.4 多尺度特征增强的 ResUNet+

本文提出的多尺度特征增强的 ResUNet+ 是一个端到端的像素级建筑物分割网络,输入为高分辨率遥感影像,输出为以黑色为背景、白色为建筑物的分割图。ResUNet+ 采用 U 型结构,共有 3 个

编码器块、Dense ASPP 和 3 个解码器块,网络结构如图 4 所示。

上侧的编码部分以 ResNet 为骨干网络来获取建筑物的基本特征,残差块在不同的网络层间传播信息,可以解决每个编码器中的退化问题,同时降低计算成本。为了在最终特征图中保留更多的细节,将 SE 模块与残差模块结合在一起,使其能根据特征的有效性自适应地对特征进行激活,提升了对不同数据集的泛化能力并改善了网络性能。在 SE 中,采用全局平均池化来汇总信息以计算不同通道特征的权重,虽能有效地获取目标物体的范围,但会丢失目标的突出特征。因此在 SE 模块中通过并行添加最大池化层来保留目标物体的突出特征,减少无用信息的影响,提升了网络的表达能力,如图 4(a)所示。

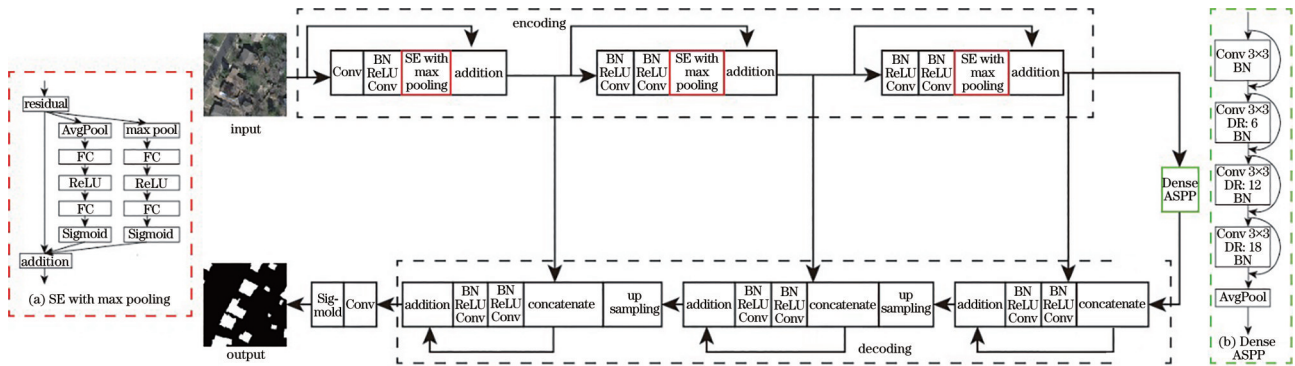


图 4 多尺度特征增强的 ResUNet+

Fig. 4 Multiscale features enhancement ResUNet+

在编码网络的最后一层采用 ASPP 模块,尽管 ASPP 能够生成多尺度特征,但在尺度轴上其特征分辨率还不够密集,获取的感受野还不够大。为了进一步捕获更密集的多尺度特征,在 ASPP 中,将前一层的输出作为后一层的输入,以更密集的方式连接一组空洞卷积,以此获得更大范围的感受野,如图 4(b)所示。根据文献[20],在没有显著增加网络大小的情况下,ASPP 中每一层的感受野的大小分别为 1、13、25 和 37,而改进后的 Dense ASPP 中每一层的感受野分别为 1、13、37 和 61,这意味着 Dense ASPP 能捕获密集的全局上下文信息,能更好地应对目标物体尺度的变化。

下侧的解码部分也由残差块组成,对深层的特征图进行最近邻上采样,并与来自其相应编码路径的特征图相连接。在网络的最后,通过一个 1×1 卷积层和 Sigmoid 函数输出影像中每个像素属于建筑物的概率,可得到建筑物的分割结果。

3 实验与分析

3.1 数据集及预处理

为了验证 ResUNet+ 网络对小型和不规则建筑物提取的准确性,本文采用的 WHU 和 INRIA 建筑物数据集涵盖了城市和郊区,包含各种大小和不规则的建筑物及地表覆盖物,场景的复杂性和建筑物的多样性增加了提取难度。图 5 是两个数据集的示例,其中白色区域代表建筑物。

1) WHU 数据集^[21]由 8189 张大小为 $512 \text{ pixel} \times 512 \text{ pixel}$ 的图像组成,空间分辨率为 0.3 m ,范围覆盖新西兰克萊斯特彻奇的 450 km^2 的面积,其中包含超过 220000 座建筑物。该数据集分为训练集 4736 张、验证集 1036 张和测试集 2416 张。WHU 数据集集中的所有建筑物标签均经过人工校准,因此适合于评估所提方法性能。由于训练样本相对较少,采用中心裁剪、随机裁剪、水平翻转、垂直翻转、比



图 5 数据集中的影像和相应标签

Fig. 5 Images and corresponding labels of the datasets

例缩放、随机旋转(旋转角度从 0° 到 90° 中随机选择)和亮度增强等数据增强手段来增加图像数量,最终获得训练集 10724 张、验证集 3574 张、测试集 3574 张。

2) INRIA 数据集^[22]包含 810 km^2 的遥感影像,空间分辨率为 0.3 m ,由美国的奥斯汀、芝加哥和基特萨普,奥地利的蒂罗尔西部和维也纳 5 个风格迥异地区的建筑组成。每个地区有 36 张大小为 $5000 \text{ pixel} \times 5000 \text{ pixel}$ 的图像。受实验平台 GPU 存储单元的限制,本文将每张影像无缝裁剪成大小为 $512 \text{ pixel} \times 512 \text{ pixel}$ 的图像,共计 14580 张;并随机从中选出 8748 张图像组成训练集,2916 张图像组

成验证集,余下的 2916 张图像组成测试集。

3.2 分割性能评价指标

为了量化评价建筑物的分割结果,本文选取精确率(P)、召回率(R)和 F_1 分数(F_1)作为像素级语义分割的评价指标。计算公式分别为:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad (4)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (5)$$

$$F_1 = 2 \times \frac{P \times R}{P + R}, \quad (6)$$

式中 N_{TP} 代表正确识别的建筑物像素数, N_{FP} 代表错提成建筑物的像素数, N_{FN} 代表漏提建筑物的像素数。

3.3 网络参数设置

实验使用 TensorFlow 作为后端的 Keras 深度学习框架,在带有 8 GB 内存的 NVIDIA GeForce GTX 2060 Super GPU 上进行训练。具体参数设置如下:批大小为 4、迭代次数为 50、优化器为 Nadam、损失函数为二分类交叉熵、激活函数为 ReLU,初始学习率设为 0.0001,且在训练过程中动态调整学习率的大小以使模型更好地收敛。

3.4 结果与分析

针对本文所提出的网络,为了验证改进后的 SE 和 Dense ASPP 模块对网络整体性能的影响,采用以下方式进行消融实验分析:1)删除本文网络中的 Dense ASPP 模块,保留改进后的 SE 模块,将该消融模型记为 A;2)删除本文网络中改进后的 SE 模块,保留 Dense ASPP 模块,将该消融模型记为 B。表 1 和图 6 分别给出了消融模型在 WHU 和 INRIA 测试集上的精度评价结果和建筑物分割结果。

表 1 消融模型在 WHU 和 INRIA 测试集上的建筑物分割精度

Table 1 Building segmentation accuracy of ablation models on WHU and INRIA testing datasets

unit: %

Dataset	Network	SE with max pooling	Dense ASPP	P	R	F_1
WHU	A	✓	×	95.64	93.07	94.34
	B	×	✓	95.83	93.71	94.76
	ResUNet+	✓	✓	96.13	95.14	95.63
INRIA	A	✓	×	86.33	90.13	88.19
	B	×	✓	90.66	87.12	88.85
	ResUNet+	✓	✓	90.81	90.11	90.46

从表 1 可以看出,本文提出的 ResUNet+ 网络中使用到的两个模块在三项评价指标中均发挥了作用,其中 F_1 值是精确率与召回率的加权调和平均值,更能全面体现网络的分割性能,通常选 F_1 值最

高的网络作为最佳网络。当消融实验中缺少 Dense ASPP 模块时,网络性能下降最明显,WHU 和 INRIA 测试集上的 F_1 值分别降低了 1.29 个百分点和 2.27 个百分点,该情况对精度的影响最大;当缺

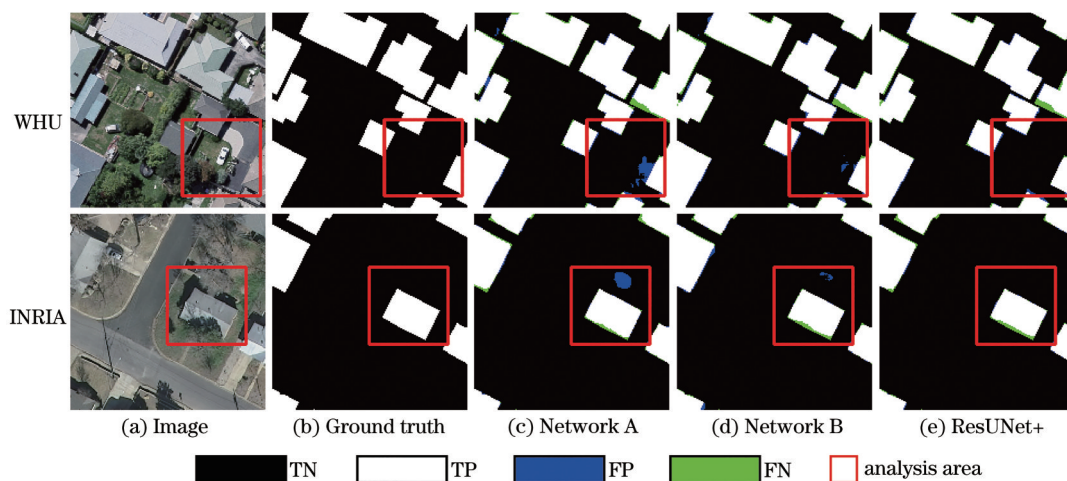


图6 消融模型在WHU和INRIA测试集上的分割结果

Fig. 6 Building segmentation results of ablation models on WHU and INRIA testing datasets

少改进后的SE模块时,WHU和INRIA测试集上的 F_1 值分别降低了0.87个百分点和1.61个百分点,该情况对精度的影响较小。消融实验结果表明改进后的SE和Dense ASPP模块在多尺度建筑分割任务中的重要性,且两个模块对网络整体性能的影响程度依次递增。

图6为消融模型在测试集上的分割结果。为了更加清晰直观地对比各模型的性能,用不同颜色对建筑物的分割结果进行区分。其中TP代表正确分割的建筑物像素,TN代表背景像素,FP代表错提建筑物的像素,FN代表漏提建筑物的像素,红色矩形框表示着重分析的区域。可以看出,去除Dense ASPP模块的消融模型A的分割结果最差,存在将道路错提取成建筑的现象,此外对小型建筑物的分割也不够精确,这说明Dense ASPP模块对网络整体性能的影响最大。导致上述现象的可能原因是删除Dense ASPP模块的模型A难以在多个

尺度上捕获到足够的上下文信息,因此未能对不同大小的建筑物实现准确分割。去除改进后的SE模块的消融模型B的错提现象得到了缓解,但与ResUNet+网络相比仍存在差距,导致该现象可能原因是改进后的SE模块通过对提取的特征进行加权,强化了与建筑物相关的显著特征且抑制了无用信息在网络中的传播,使得建筑物的提取结果更加完整。本文网络引入的改进后的SE和Dense ASPP模块,提高了对多尺度建筑物的分割精度,解决了不规则建筑物边界分割粗糙的问题,综合指标 F_1 值达到最优,由此说明了两个模块结合是有效的。

为了评估ResUNet+网络的性能,将其与目前主流的语义分割网络(如SE-UNet、DeepLabv3+、Dense ASPP和ResUNet)进行对比,同时所有网络均采用相同的超参数进行训练以确保实验的公平性。

图7列出了5种网络在测试集上对部分输入影

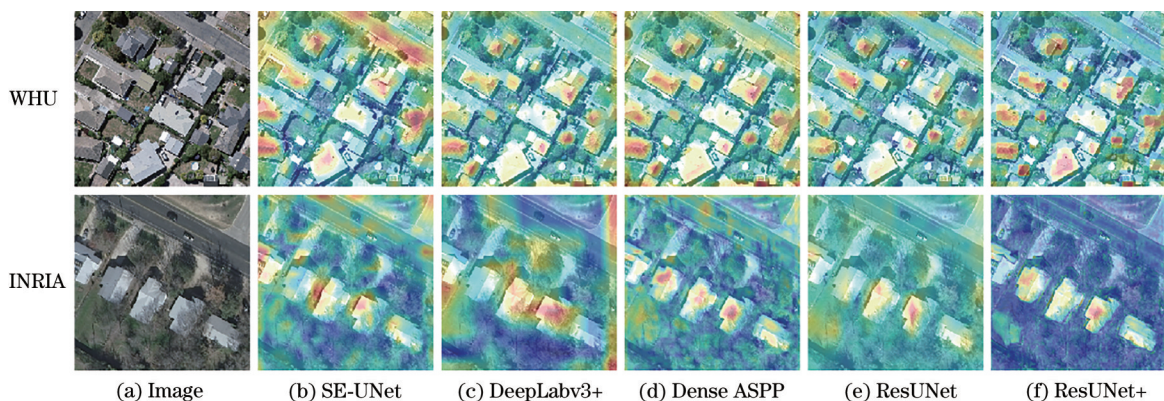


图7 不同网络在WHU和INRIA测试集上提取特征的热图

Fig. 7 Heat maps of different networks extracting features on WHU and INRIA testing datasets

像提取特征的热图(Heat map)。从图中可以看出, SE-UNet 网络获取的特征热图中包含一些非建筑物,如水泥路面等,这会导致将水泥路面错提取成建筑物的现象;DeepLabv3+和 Dense ASPP 获取的特征热图中难以准确地描述影像中建筑物,因此将导致建筑物的分割精度不佳;ResUNet 网络获取的特征热图能准确描述大中型建筑物,但对于小型建筑物获取的特征较少,且建筑物的边界信息丢失严重。对比发现,ResUNet+网络获取的特征热图可以更加准确地凸显不同大小建筑物的主体,边界信息较为明显清晰,这表明网络中改进后的 SE 模块能够对表示建筑物的特征进行加权强化, Dense ASPP 模块也能够获得不同尺度建筑物的上下文信息,从而对小型和不规则建筑物提取出指向性强的特征。

表 2 为 5 种网络在测试集上的建筑物分割精度。从表中可以看出,ResUNet+在各项精度指标中均表现最优,其次是 Dense ASPP。与 ResUNet 相比,ResUNet+的精确率、召回率和 F_1 值在 WHU 测试集上分别提高了 2.98 个百分点、3.37 个百分点和 3.18 个百分点,在 INRIA 测试集上分别提高了 4.36 个百分点、6.00 个百分点和 5.20 个百分点,这说明 ResUNet+网络的性能得到了提升。

图 8 为 5 种网络在测试集上对部分影像的建筑物

表 2 不同网络在 WHU 和 INRIA 测试集上的建筑物分割精度

Table 2 Building segmentation accuracy of different networks on WHU and INRIA testing datasets unit: %

Dataset	Network	P	R	F_1
WHU	SE-UNet	90.95	87.63	89.26
	DeepLabv3+	94.42	92.59	93.50
	Dense ASPP	94.81	93.14	93.97
	ResUNet	93.15	91.77	92.45
	ResUNet+	96.13	95.14	95.63
INRIA	SE-UNet	83.47	82.08	82.77
	DeepLabv3+	87.85	86.22	87.03
	Dense ASPP	88.66	87.30	87.97
	ResUNet	86.45	84.11	85.26
	ResUNet+	90.81	90.11	90.46

物分割结果。对于第 1 行影像右上方的小型建筑物, SE-UNet 和 ResUNet 均未能完整地提取, DeepLabv3+、Dense ASPP 和 ResUNet+ 正确提取出了建筑物的形状,但 DeepLabv3+ 和 Dense ASPP 错误地将部分道路提取成建筑物。此外对于影像中的不规则建筑物,只有 ResUNet+ 在完整性上做得最好。第 2 行影像中存在不同尺度大小的建筑物, SE-UNet、DeepLabv3+、Dense ASPP 和 ResUNet 均漏提了小型建筑物,只有 ResUNet+ 对尺度各异的建筑实现了较为准确的分割。对于第 3 行影像中

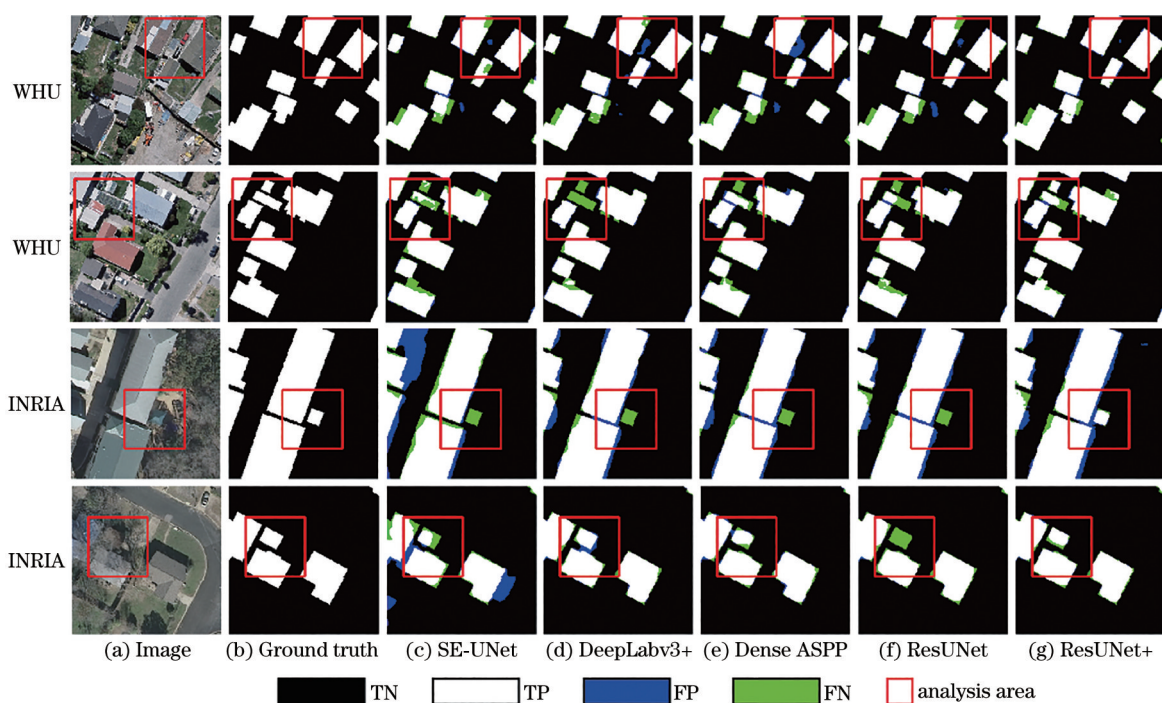


图 8 不同网络在 WHU 和 INRIA 测试集上的建筑物分割结果

Fig. 8 Building segmentation results of different networks on WHU and INRIA testing datasets

的建筑物,SE-UNet存在较为严重的错提现象,其余4种网络对大型规则的建筑物都得到了较好的分割结果,但只有ResUNet+准确地提取到了旁边的小型建筑物。第4行影像中受树木遮挡的影响,SE-UNet、DeepLabv3+和ResUNet都存在不同程度上的漏提和错提现象,Dense ASPP和ResUNet+均正确提取出了建筑物的形状,但ResUNet+仍然具有更高的精确度。

从图8的结果可以看出,高分辨率遥感影像中的建筑物屋顶在外观上与道路非常相似,由于无法获得丰富的全局上下文信息,SE-UNet模型对建筑物的准确分割能力较差,将大部分的道路错误分割成建筑物。DeepLabv3+采用ASPP将多个不同扩张率的空洞卷积层特征合并来扩大感受野,提高了对不同尺度建筑物的分割精度,但在建筑物的边界上产生更多的错误分割,特别是在小型建筑物上。Dense ASPP也是采用了空洞卷积金字塔结构,其性能与DeepLabv3+网络相近。由于其以并行和级联的方式使用空洞卷积,故可以在更大范围内生成更多尺度的特征,但错提现象依旧存在。ResUNet无法准确识别密集建筑区域内的小型建筑物。与其他4个网络相比,ResUNet+在非矩形和小型建筑物上都实现了更为准确的分割,网络的分割结果与标签具有较强的一致性。这是因为高分辨率影像中建筑物的大小和形状各异,要完整描述建筑的特征需要较大的感受野,需要网络对深层特征具有更强的表达能力。Dense ASPP模块的引入能够获取更为丰富的多尺度上下文信息,且对大小各异和形状不规则的建筑物具有更好的识别效果,故错提和漏提现象被极大地减少;通过在编码器内使用改进后的SE模块,能够有效地增强网络的特征提取能力,且建筑物的边缘信息也较为完整。这充分表明通过在网络中添加改进后的SE和Dense ASPP模块,ResUNet+网络能够更加有针对性地提取到具有判别力的特征信息,从而适应不同大小和不规则建筑物的提取。

4 结 论

针对高空间分辨率遥感影像中小型和不规则建筑物的精确提取问题,在ResUNet残差网络中引入改进后的SE和Dense ASPP模块,提出一种多尺度特征增强的ResUNet+网络。改进后的SE模块能够根据网络损失值学习特征权重,自适应地

对特征进行激活,强化对有效特征的学习;同时引入的Dense ASPP模块能够在多个层次的特征图上获得更大范围的上下文信息,提升了网络对不同尺度建筑物的特征提取能力。ResUNet+在WHU测试集上的精确率、召回率和 F_1 值分别达到96.13%、95.14%和95.63%,在INRIA测试集上的精确率、召回率和 F_1 值分别达到90.81%、90.11%和90.46%,均超过了本文引用的SE-UNet、DeepLabv3+、Dense ASPP和ResUNet;从建筑物的分割结果来看,各个网络都可以对大中型建筑物实现较准确的提取,而Dense ASPP和ResUNet+的分割结果总体上优于SE-UNet、DeepLabv3+和ResUNet,且只有ResUNet+能同时准确地提取出小型的和形状不规则的建筑物。然而本文提出的算法用于高分辨率影像中建筑物分割时分割形状并不是建筑物通常具有的直线边界,在后续的工作中将重点研究如何提升建筑物边界分割的准确性,进一步提高建筑物的分割精淮度。

参 考 文 献

- [1] Wang Y, Yang Y, Wang B S, et al. Building segmentation in high-resolution remote sensing image through deep neural network and conditional random fields[J]. *Journal of Remote Sensing*, 2019, 23(6): 1194-1208.
王宇, 杨艺, 王宝山, 等. 深度神经网络条件随机场高分辨率遥感图像建筑物分割[J]. *遥感学报*, 2019, 23(6): 1194-1208.
- [2] Diakogiannis F I, Waldner F, Caccetta P, et al. ResUNet-A: a deep learning framework for semantic segmentation of remotely sensed data[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 162: 94-114.
- [3] Wang J X, Feng D J, Chen J F. A building boundary regularization method by contrasting Harris operator and Susan operator[J]. *Bulletin of Surveying and Mapping*, 2020(4): 11-15.
王杰茜, 冯德俊, 陈建飞. 对比Harris算子和Susan算子的建筑物边界规则化方法[J]. *测绘通报*, 2020(4): 11-15.
- [4] Lin X G, Zhang J X. Object-based morphological building index for building extraction from high resolution remote sensing imagery[J]. *Acta Geodaetica et Cartographica Sinica*, 2017, 46(6): 724-733.
林祥国, 张继贤. 面向对象的形态学建筑物指数及其

- 高分辨率遥感影像建筑物提取应用[J]. 测绘学报, 2017, 46(6): 724-733.
- [5] Lü D S, Lin N, Zhang X Q. Object-oriented multi-scale multi-feature high-resolution remote sensing image building extraction[J]. Beijing Surveying and Mapping, 2019, 33(2): 191-195.
吕道双, 林娜, 张小青. 面向对象的多尺度多特征高分遥感影像建筑物提取[J]. 北京测绘, 2019, 33(2): 191-195.
- [6] Jin Y W, Xu W B, Zhang C, et al. Boundary-aware refined network for automatic building extraction in very high-resolution urban aerial images[J]. Remote Sensing, 2021, 13(4): 692-712.
- [7] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3431-3440.
- [8] Ren X L, Wang Y P, Yang J Y, et al. Building detection from remote sensing images based on improved U-net[J]. Laser & Optoelectronics Progress, 2019, 56(22): 222801.
任欣磊, 王阳萍, 杨景玉, 等. 基于改进 U-net 的遥感影像建筑物提取[J]. 激光与光电子学进展, 2019, 56(22): 222801.
- [9] Zhang Z H, Fang W, Du L L, et al. Semantic segmentation of remote sensing image based on encoder-decoder convolutional neural network[J]. Acta Optica Sinica, 2020, 40(3): 0310001.
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[J]. 光学学报, 2020, 40(3): 0310001.
- [10] Liu H, Luo J C, Huang B, et al. Building extraction based on SE-Unet[J]. Journal of Geo-Information Science, 2019, 21(11): 1779-1789.
刘浩, 骆剑承, 黄波, 等. 基于特征压缩激活 Unet 网络的建筑物提取[J]. 地球信息科学学报, 2019, 21(11): 1779-1789.
- [11] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA. New York: IEEE Press, 2015: 1-9.
- [12] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [13] Zhu S X, Zhou Z J, Gu X J, et al. Scene classification of remote sensing images based on RCF network[J]. Laser & Optoelectronics Progress, 2021, 58(14): 1401001.
朱淑鑫, 周子俊, 顾兴健, 等. 基于 RCF 网络的遥感图像场景分类研究[J]. 激光与光电子学进展, 2021, 58(14): 1401001.
- [14] Liu Y F, Zhang Q Z, Wang G H, et al. Building extraction in remote sensing imagery based on deep residual network[J]. Remote Sensing Information, 2020, 35(2): 59-64.
刘亦凡, 张秋昭, 王光辉, 等. 利用深度残差网络的遥感影像建筑物提取[J]. 遥感信息, 2020, 35(2): 59-64.
- [15] Luo L Y, Zhu T E, Zhang G M, et al. Building extraction on high spatial resolution remote sensing image based on DeepLabv3+ segmentation model [J]. Electronic Information Warfare Technology, 2021, 36(4): 65-69, 73.
罗李焱, 朱婷婷, 张桂铭, 等. 基于 DeepLabv3+ 模型的高分辨率遥感影像建筑物提取[J]. 电子信息对抗技术, 2021, 36(4): 65-69, 73.
- [16] Tian Q L, Qin K, Chen J, et al. Building change detection for aerial images based on attention pyramid network[J]. Acta Optica Sinica, 2020, 40(21): 2110002.
田青林, 秦凯, 陈俊, 等. 基于注意力金字塔网络的航空影像建筑物变化检测[J]. 光学学报, 2020, 40(21): 2110002.
- [17] Zhang Z X, Liu Q J, Wang Y H. Road extraction by deep residual U-net[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5): 749-753.
- [18] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [19] 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.
- [20] Yang M K, Yu K, Zhang C, et al. DenseASPP for semantic segmentation in street scenes[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake

- City, UT, USA. New York: IEEE Press, 2018: 3684-3692.
- [21] Ji S P, Wei S Q, 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, 2019, 57(1): 574-586.
- [22] Maggiori E, Tarabalka Y, Charpiat G, et al. Can semantic labeling methods generalize to any city? The inria aerial image labeling benchmark[C]//2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), July 23-28, 2017, Fort Worth, TX, USA. New York: IEEE Press, 2017: 3226-3229.