

融合 ASPP 与双注意力机制的建筑物提取模型

于明洋 徐海青* 张文焯 徐帅 周放亮
(山东建筑大学测绘地理信息学院, 济南 250101)

摘要 精准高效地从高分辨率遥感影像中提取建筑物信息对国土规划和地图制图意义重大, 近年来基于卷积神经网络进行建筑物信息提取已经取得了很大的进展, 然而在处理高分辨率遥感影像时仍存在影像的高级语义特征利用不够充分, 难以获得细节丰富高精度分割影像的问题。文章针对以上问题提出了一种用于建筑物自动提取的深度学习网络结构空洞空间与通道感知网络 (Atrous Space and Channel Perception Network, ASCP-Net)。该模型将空洞空间金字塔池化 (Atrous Spatial Pyramid Pooling, ASPP) 和空间与通道注意力 (Spatial and Channel Attention, SCA) 模块融入到编码器-解码器结构中, 通过 ASPP 模块来捕获和聚合多尺度上下文信息, 采用 SCA 模块选择性增强特定位置和通道中更有用的信息, 并将高低层特征信息输入解码网络完成建筑物信息的高效提取。在 WHU 建筑数据集 (WHU Building Dataset) 上进行实验, 结果表明: 文章提出的方法总体精度和 F_1 评分分别达到了 97.4% 和 94.6%, 相比其他模型能够获得更清晰的建筑物边界, 尤其对图像边缘不完整建筑的提取效果较好, 有效提升了建筑物提取的精度和完整性。

关键词 高分辨率遥感影像 双注意力机制 空洞卷积 建筑物提取

中图分类号: TP751.1

文献标志码: A

文章编号: 1009-8518(2024)01-0136-11

DOI: 10.3969/j.issn.1009-8518.2024.01.012

Building Extraction Model Based on ASPP and Dual Attention Mechanism

YU Mingyang XU Haiqing* ZHANG Wenzhuo XU Shuai ZHOU Fangliang

(School of Surveying and Mapping, Shandong Jianzhu University, Jinan 250101, China)

Abstract Accurate and efficient building information extraction from high-resolution remote sensing images is of great significance for land planning and mapping. In recent years, great progress has been made in building information extraction based on convolutional neural networks. However, there still exists the problems that the advanced semantic features of the images are not sufficiently utilized and it is difficult to obtain detailed and high-precision segmentation images when processing high-resolution remote sensing images. To solve the above problems, a deep learning network architecture, Atrous Space and Channel Perception Network (ASCP-Net), is proposed for automatic building extraction. The Atrous Spatial Pyramid Pooling (ASPP) and Spatial and channel-wise Attention (SCA) modules are integrated into the encoder-decoder structure. Multi-scale context information is captured and aggregated through the ASPP module. Meanwhile, the SCA module is used to

收稿日期: 2023-03-27

引用格式: 于明洋, 徐海青, 张文焯, 等. 融合 ASPP 与双注意力机制的建筑物提取模型[J]. 航天返回与遥感, 2024, 45(1): 136-146.

YU Mingyang, XU Haiqing, ZHANG Wenzhuo, et al. Building Extraction Model Based on ASPP and Dual Attention Mechanism[J]. Spacecraft Recovery & Remote Sensing, 2024, 45(1): 136-146. (in Chinese)

selectively enhance the more useful information in specific locations and channels, and the high and low-layer feature information is input into the decoding network to achieve efficient building information extraction. Experiments on the WHU Building Dataset show that, for the overall accuracy and F1 score, the proposed method reaches 97.4% and 94.6% respectively, and can obtain clearer building boundaries compared with other models, especially for the extraction of incomplete buildings at image edges, and effectively improving the accuracy and integrity of building extraction.

Keywords high-resolution remote sensing images; dual attention mechanism; atrous convolution; building extraction

0 引言

高效准确的从遥感影像中提取建筑物,对于城市发展与规划、地图制作与更新、地区人口预估等方面具有重要意义^[1-3]。随着卫星和传感器技术的不断进步,采集的数据空间分辨率不断提高,图像具有更多细节特征的同时其干扰信息的冗余程度和差异也在不断增加。如何利用高分辨率遥感影像更准确、更及时地检测并分割出建筑物,实时掌握建筑物动态信息具有十分重要的意义^[4]。

传统研究主要集中在结合机器学习算法和手工特征来解决建筑提取问题^[5]。文献 [6] 基于高分辨率多光谱航空图像和激光雷达数据进行空间、光谱和纹理等特征的提取,利用该多特征数据对支持向量机 (Support Vector Machine, SVM) 算法进行训练,并对建筑物和非建筑物进行分类;文献 [7] 结合数字地表模型 (Digital Surface Model, DSM) 数据来进行建筑物提取;文献 [8] 使用激光雷达 (Laser Radar, LiDAR) 和合成孔径雷达 (Synthetic Aperture Radar, SAR) 数据来提取建筑物。此外,诸如 boosting^[9] 和随机森林^[10] 等优秀的机器学习分类器也被用于建筑物的提取。这些基于手工特征的传统方法通常需要先验知识,适用于解决特定的任务,而且它们的提取效率无法保证,因此难以广泛应用于建筑物的自动提取。

随着计算机运算性能和算力的快速增长,利用深度学习算法进行建筑物提取取得诸多进展。文献 [11] 以 U-net 为基础,设计了新的交叉熵损失函数,并且引入形态学建筑物指数来进行建筑物提取;文献 [12] 提出基于注意力重新加权的 RFU-Net,在融合不同特征时通过注意力机制弥合特征之间的语义差距,并在三个公开航空影像数据集上验证了模型的改进性能;文献 [13] 等基于双线性插值上采样和多尺度特征组合提出一种多尺度建筑物提取网络,以解决语义分割网络中的连续下采样会损失特征中的细节信息,导致提取结果边缘模糊的问题;文献 [14] 同时使用膨胀卷积与金字塔池化来进行建筑物提取;文献 [15] 使用双注意力机制,并对比了不同的连接方式来进行建筑物提取。诸多学者针对建筑物语义分割做了相关研究,但大部分没有基于图像的高级特征对模型进行优化改进,图像高级特征的利用程度不够,直接影响到部分建筑的提取效果。

与许多现存方法不同,本文针对在建筑物提取任务中深层语义特征利用程度不够充分以及裁剪后图像边缘的建筑分割易混淆的问题,设计了一种新的深度学习建筑物提取网络 ASPP-Net。模型先经过 ASPP 模块在多个尺度上进行高级特征提取,再通过空间与通道注意力机制选择性地融合特定位置和通道中更有用的特征,进一步加强对图像高级特征的利用。本文方法可以同时维持模型对浅层特征和深层多尺度特征的代表能力,保证对不同尺度建筑物提取的完整性;同时,从空间和通道两种维度表示不同的语义特征,提升对影像边缘不完整建筑物的分割精度。最后,与五种经典算法进行对比,证明我们的方法可以有效改善提取过程中出现的漏提现象,提升建筑物的整体分割精度。

1 网络结构

本文提出网络的总体框架如图 1 所示。空洞空间与通道感知网络 (ASCP-Net) 是一种端到端的训练模型, 整体框架分为编码网络和解码网络。编码器部分通过深度卷积神经网络 (Deep Convolutional Neural Network, DCNN) + 空洞空间金字塔池化 (Atrous Spatial Pyramid Pooling, ASPP) + 空间与通道注意力 (Spatial and Channel Attention, SCA) 进行影像特征提取, 先在 DCNN 中通过设置不同的空洞卷积扩张率得到低层特征图和高层特征图, 高层特征经过 ASPP 进行多尺度特征提取, 更好地保留影像中的语义信息; 再经过 SCA 网络自适应地将位置上相似的特征进行关联, 同时选择性地强调相互依赖的通道图; 最后将两个注意力模块的输出相加, 以进一步改进特征表示, 低层特征与经过特征增强的高层特征将一起输入到解码器部分。解码网络负责对编码信息进行解码, 恢复特征图的语义特征信息。高层特征经过 4 倍双线性内插上采样再与经过 1×1 卷积运算后的低层特征进行融合, 聚合不同尺度的语义特征, 同时避免出现梯度消失问题。通过上采样将输出结果分为两类: 建筑物和非建筑物。

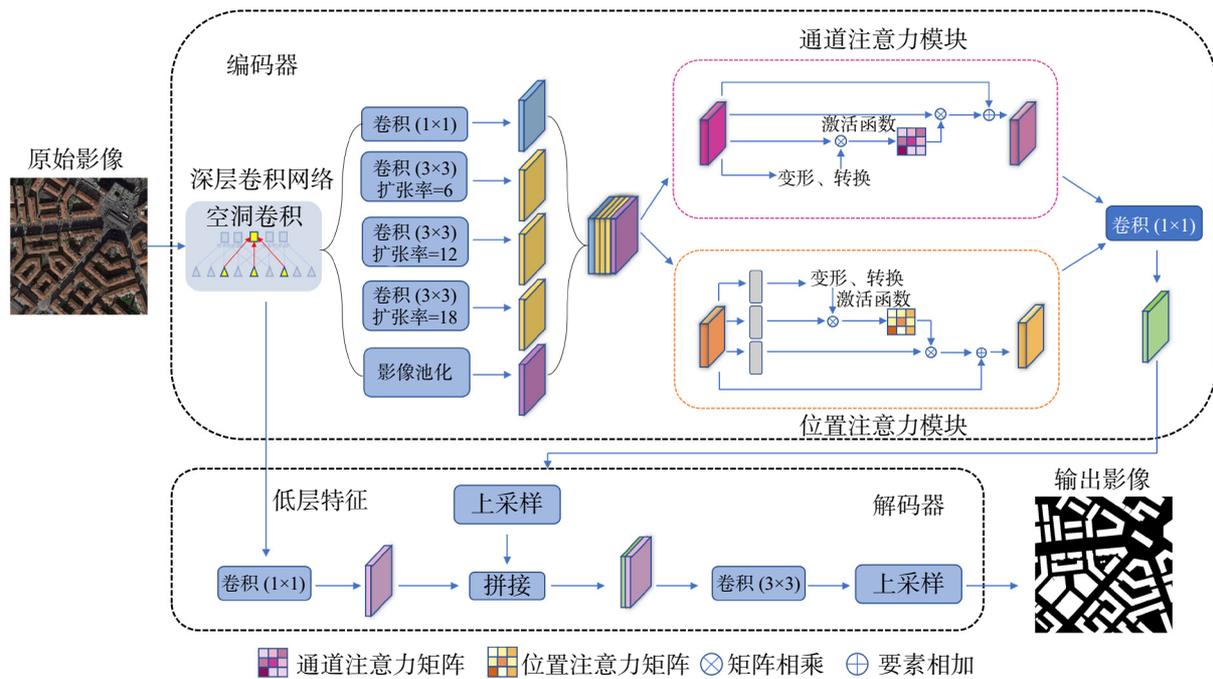


图 1 模型整体架构

Fig.1 Overall architecture of the model

1.1 DCNN 网络

DCNN 是在传统 CNN 的基础上通过重复堆叠多个卷积层来实现的一种网络结构, 通过增加神经网络的深度来直接影响模型的信息提取能力。本文使用深层网络模型 Xception^[16] 来分别提取图像的低级特征和高级特征, 其内部主要结构为残差卷积神经网络。ResNet^[17] 由何恺明等人于 2015 年提出, 它通过建立跳跃连接在很大程度上解决了梯度消失的问题, 从而可以训练到更深层的网络, ResNet 的残差计算方法如式 (1)

$$x_k = L(x_{k-1}) + x_{k-1} \quad (1)$$

式中 x_{k-1} 为第 $k-1$ 层的输出; L 表示对上一层的输入图进行卷积、归一化等操作; x_k 为第 k 层输出的结果; ResNet 的残差结构如图 2 (a) 所示。

1.2 ASPP 模块

ASPP 最开始在 Deeplab V2 中提出, 它将空洞卷积与金字塔池化 (Spatial Pyramid Pooling, SPP) 进行结合^[18]。SPP 的核心思想是利用多个不同尺度的池化层进行特征提取并融合成一个尺度统一的向量输入到全连接层, 以解决池化层产生的不同输出导致全连接层无法训练的问题。如图 2 (b) 所示, 对于给定的输入, ASPP 用不同扩张率的空洞卷积进行并行采样, 包括一个 1×1 的卷积层, 多个不同扩张率的 3×3 的卷积层以及一个池化层, 池化后的特征信息经过 1×1 的卷积和上采样将图像尺寸恢复至预期值, 最后将得到的所有结果拼接来扩大通道数, 再通过 1×1 的卷积来将通道数降低到预期值进行输出。在本文中, 输入为上一步深度神经网络提取出的高级语义信息, 经过 ASPP 进行多尺度特征提取, 从而有效聚合多个不同尺度的图像高级语义特征。

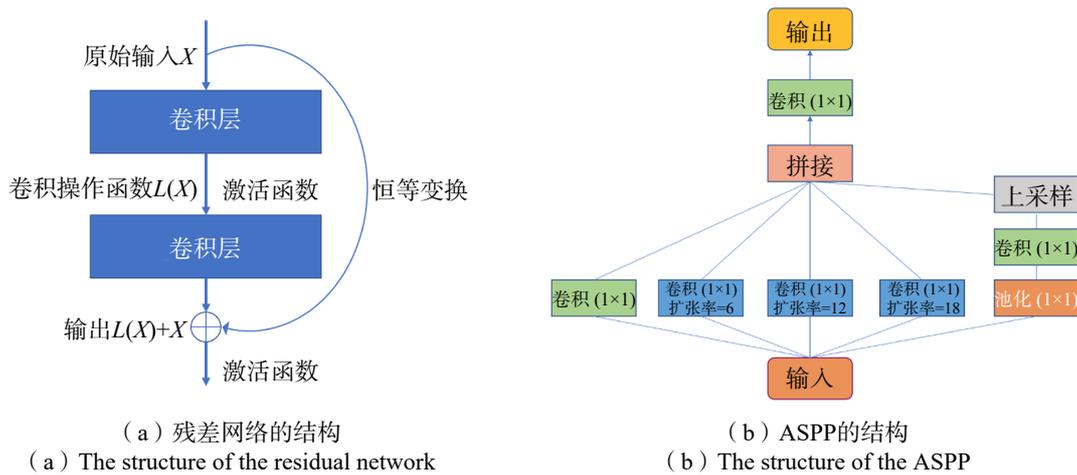


图 2 残差网络和 ASPP 的结构
Fig.2 Structure of the residual network and ASPP

1.3 SCA 网络

SCA 模块^[19]通过加入通道注意力机制和位置注意力机制进一步利用网络中的多种特征信息。位置注意力模块与特征图中哪些位置更为重要相关联, 即对应原始图像中哪些感受野更为重要; 通道注意力模块与特征图中哪些通道更为重要相关联, 它以一种自适应的方式让网络更好的学习输入影像中的重要信息, 从而提高模型的特征表示能力。在本文中, 输入为经过 ASPP 处理的图像高级语义特征, 通过空间注意力模块和通道注意力模块分别处理后将两者得到的结果进行累加; 再经过一个 1×1 的卷积后将最终结果输入到解码器中。

1.3.1 通道注意力模块

通道注意力模块通过计算输入图像各通道的权重来选择性的关注信息更为丰富的通道, 提高特征的代表能力。如图 3 (a) 所示, 原始输入 A 的形状为 $C \times H \times W$ (C 、 H 、 W 分别表示输入的通道数、高度和宽度), 先对其进行变形、转换得到大小为 $K \times C$ ($K=H \times W$) 的矩阵, 再对 A 进行变形得到 $C \times K$ ($K=H \times W$) 的矩阵, 将两个矩阵相乘再经过 Softmax 得到形状为 $C \times C$ 的权重矩阵 N ; 将 N 与变形后的原始输入 A 相乘得到形状为 $C \times K$ ($K=H \times W$) 的矩阵, 最后将得到的矩阵变形后与原始的 A 进行累加得到最终结果 $E=C \times H \times W$ 。

1.3.2 位置注意力模块

位置注意力模块重点关注图像关键信息更丰富的空间区域并赋予它们更大的权重。如图 3 (b) 所示, 位置注意力机制与通道注意力机制类似, 不同之处在于不是对原始输入 A 本身进行变形转换, 而是先通

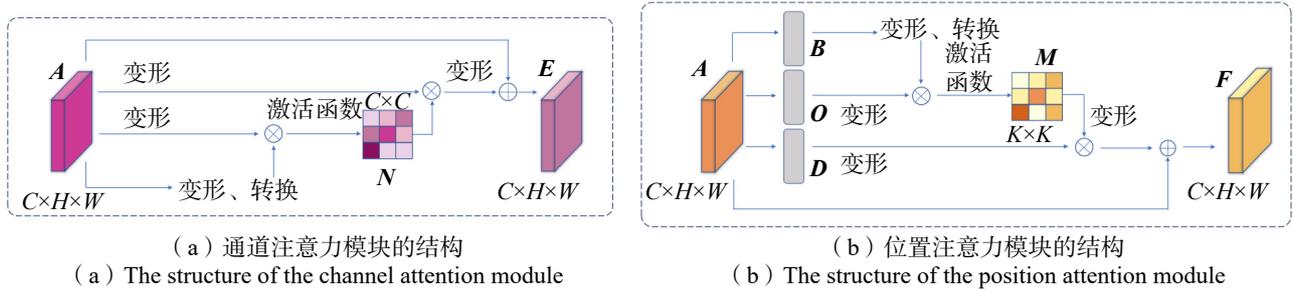


图3 通道与位置注意力模块
Fig.3 Channel and position attention modules

过神经网络对 A 进行处理得到形状相似的矩阵 B 、 O 、 D ，再分别对 B 、 O 、 D 进行处理。 B 通过变形、转换之后形状为 $K \times C (K=H \times W)$ ，再与变形后的 O 矩阵进行相乘，经过 Softmax 处理后得到形状为 $K \times K (K=H \times W)$ 的权重矩阵 M ；再对矩阵 D 进行变形，得到形状为 $C \times K$ 的矩阵，将该矩阵与 M 进行相乘得到结果，再将变形后的结果与原始输入 A 进行相加得到最终输出 $F=C \times H \times W$ 。

1.4 解码器

解码器的主要输入为图像经过 DCNN 处理得到浅层特征以及经过双重注意力机制处理后的深层特征。浅层特征通过 1×1 的卷积将通道数缩减为预期值，深层特征通过上采样调整大小后将得到的两种特征进行拼接；再使用一个 3×3 的卷积对拼接结果进行处理，通过一个上采样把特征图恢复到原始尺寸大小进行最终输出。

2 实验分析

2.1 数据处理

本研究采用 WHU 建筑数据集 (WHU Building Dataset)，该数据集采集于新西兰克赖斯特彻奇，包括 8 189 幅大小为 512 像素 \times 512 像素的遥感影像，其中训练集、验证集和测试集的样本数量分别为 4 910、1 400 和 1 879。验证集影像空间分辨率为 0.3 m；同时本研究对原始影像进行数据增强，数据增强主要通过通过对样本进行变形处理来增加训练样本，避免模型出现过拟合现象^[20]。如图 4 所示，本文通过对样本进行垂直、水平镜像翻转以及不同角度的旋转来实现图像数据增强。

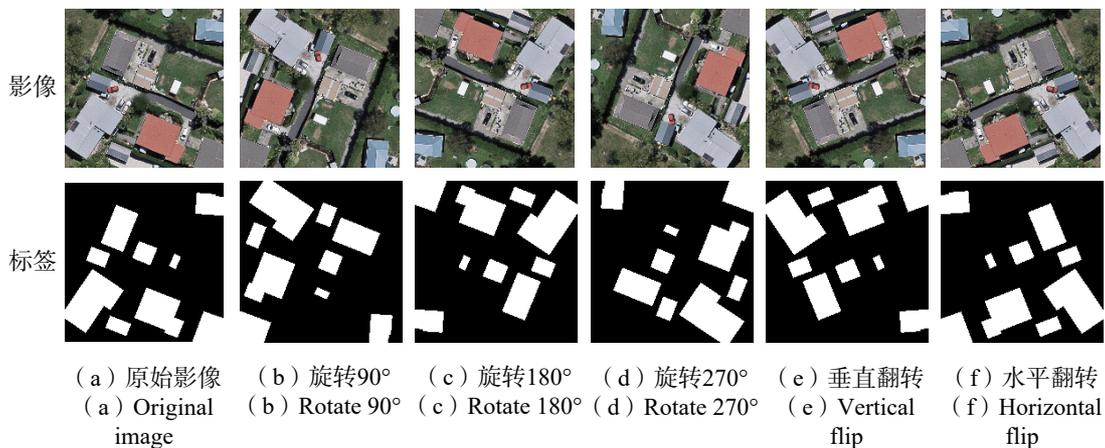


图4 经过数据增强的影像与标签
Fig.4 Data enhanced images and labels

2.2 实验设置

本研究基于 PyTorch 深度学习框架,使用 TorchVision、Scikit-Image、Matplotlib 等开源 python 库进行影像处理,搭配 NVIDIA GeForce GTX 3070 Ti 显卡进行模型训练,显存为 8 GB,使用 CUDA11.0 加速运算,同时对比实验了五种经典的语义分割模型。实验过程中,选用二元交叉熵作为损失函数,优化器的初始学习率设置为 1×10^{-4} 。为了避免过拟合,在所有卷积中引入正则化操作,权重衰减为 1×10^{-4} ,模型训练次数设置为 150;为了克服 GPU 内存的限制,批量大小设为 8。

2.3 精度评价

本研究采用整体精度 OA、召回率 Recall、准确率 Precision、F1 评分、交并比 IoU 和卡帕系数 Kappa 六个指标完成精度评价。部分指标的定义见式(2)~式(5):

$$OA = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$IoU = \frac{TP}{TP+FP+FN} \quad (3)$$

$$Kappa = \frac{P_o - P_e}{1 - P_e} \quad (4)$$

$$P_e = \frac{a_1 \times b_1 + a_2 \times b_2 + \dots + a_t \times b_t}{n \times n} \quad (5)$$

式中 TP 表示被正确分类的正类像素数目; TN 表示被正确分类的负类像素数目; FP 表示被错分为正类的像素数目; FN 表示被错分为负类的像素数目; P_o 表示总体分类精度; P_e 表示偶然一致性误差; a_1, a_2, \dots, a_t 分别表示每一类的真实样本总个数; b_1, b_2, \dots, b_t 分别表示每一类的预测样本总个数; t 为样本的类型数; n 为样本总数。

3 结果讨论

3.1 结果分析

本文对比实验了五种经典的语义分割模型,包括 DANet^[21]、FCN8s^[22]、SegNet^[23]、Unet^[24] 和 DeepLabv3+^[25]。如图 5 所示,红色区域表示各方法相对真实标签多提取的部分,蓝色区域表示各方法相对真实标签未提取的部分。从图中可以看出,除了本文方法外,其他方法都存在较多错误识别的区域。总体来说, DANet 表现最差,检测到的建筑物边界不清,存在目标预测不完整以及多分的问题。FCN8s、SegNet、Unet、DeepLabv3+ 和 ASCP-Net 在测试数据上建筑物边缘的平滑性较好,但相比其他几种方法, ASCP-Net 划分的区域中错分点更少。同时, ASCP-Net 模型提取出的建筑物具有准确的轮廓以及较为完整的内部结构,与实验的其他模型相比提取效果更好。

受计算机内存资源和硬件的影响,无法将获取的高分辨率遥感影像直接输入到神经网络模型中,需要将高分影像裁剪成需要的尺寸大小,但同一目标可能会被裁剪到不同图像的边缘位置,这会破坏边缘目标的纹理特征,为准确提取边缘地物带来挑战。如图 6 所示,黄色线框表示不同模型对边缘建筑的提取效果,对于 FCN8s、Unet、DeepLabv3+ 等网络,所提取的边缘建筑均有漏分、错分等现象,主要原因是边缘地物的部分特征被破坏,网络无法进行准确提取。而本文所提出的方法能够较准确提取出图像边缘被分割的地物,原因是该方法通过加入 ASPP 有效的利用了图像的高级语义信息,同时加入双注意力机制来精确捕获边缘地物像素之间的位置依赖关系和通道依赖关系,增强利用了边缘地物的特征语义信息,使得我们的方法能够推断出边缘不完整地物的位置,同时实现更准确的边缘地物提取,有效改善了其他方法对边缘地物提取不够准确的缺陷。

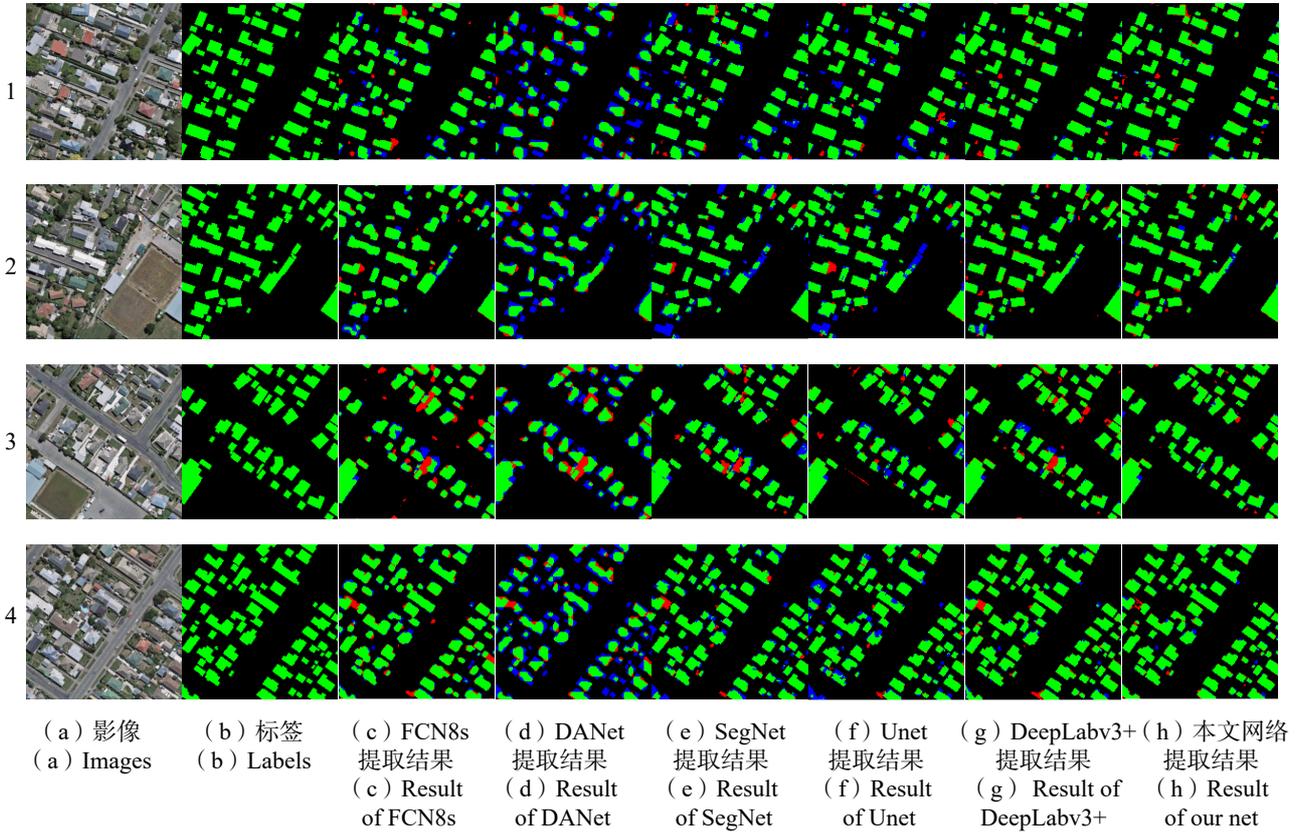


图5 不同模型的实验结果对比

Fig.5 Comparison of experimental results of different models

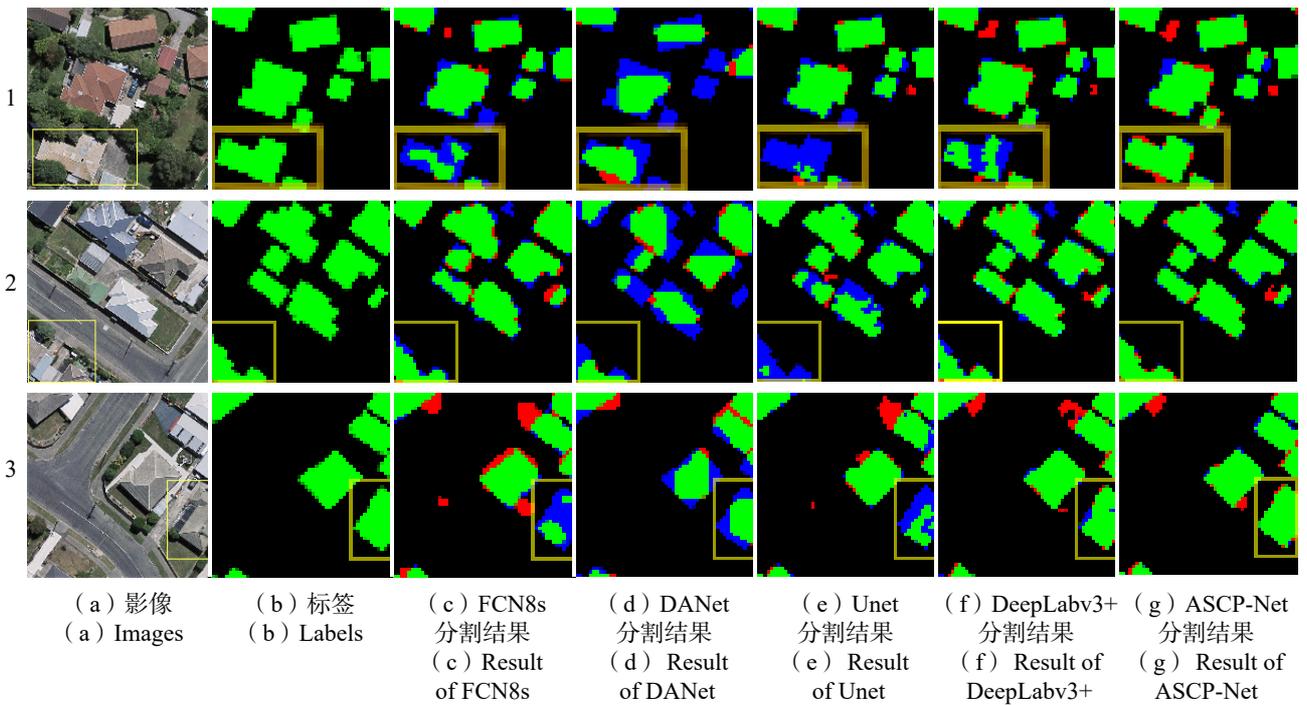


图6 不同模型对图像边缘建筑的分割效果对比

Fig.6 Comparison of segmentation effect of different models on image edge buildings

对于一些建筑区域的提取,本文方法实现了较好的效果,而 FCN8s、Unet、DeepLabv3+等模型均出现了较多的漏分区域。如图 7 所示,黄色线框表示一定范围内各方法漏提的区域分布。从图像上来看,该区域的建筑表面与多种地物相邻,包括绿化植物、硬质地面等,这为精确提取该类建筑带来了挑战。经典语义分割网络 FCN8s 和 Unet 均出现了一些漏分区域,主要原因是网络模型未充分利用图像的高级语义特征,并且所划分的像素之间并不具备多种依赖关系。DANet 虽然加入了双注意力机制,但是对于图像高级特征的挖掘还不够充分。DeepLabv3+未加入注意力机制,无法充分建立相关像素点之间的各种依赖,最终导致无法有效利用地物的特征语义信息。本文提出的方法从多个尺度利用图像的高级特征,并且通过双通道注意力机制进一步加强了高级特征维度上像素之间的依赖,相比其他方法,在该区域实现了更好的提取效果。

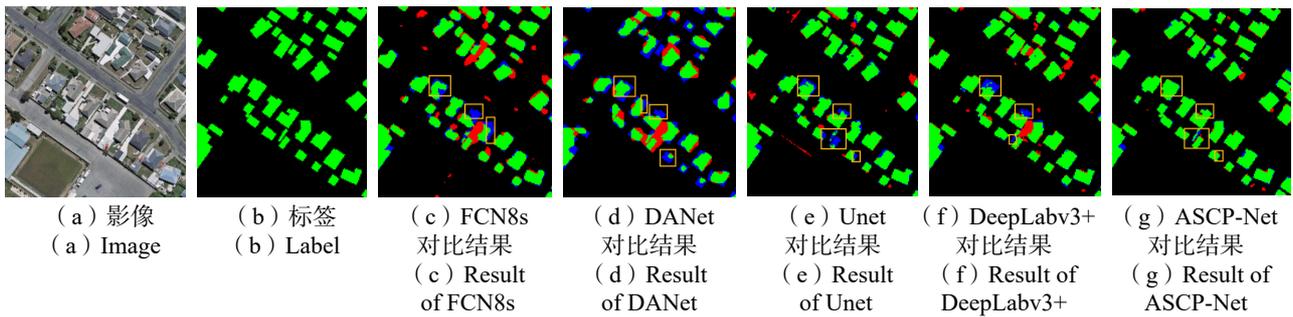


图 7 不同模型漏提的区域分布对比示意图
 Fig.7 Comparison figure of regional distribution omitted from different models

CNN 对图像进行卷积操作时,所处理的图像特征范围都是在一定区域内的,或者说是局部的。通过确定局部区域内每个像素的类别,网络实现对不同目标地物的提取和分类。但当提取大尺度地物时,对局部特征的提取并无法代表整个目标,提取的不连贯的细部特征会造成大尺度地物内部的不一致性,使得整个区域的分割不完整。本文提出的方法通过加入 ASPP 模块和双注意力机制,能够有效聚合图像相关区域的上下文依赖信息,将图像位置不同的局部特征信息联系起来,保证了大尺度区域提取的内部一致性。如图 8 所示,黄色线框表示一些方法所提区域内部存在的空洞点,红色线框表示未能进行有效识别的区域。本文方法加强了不同细节特征之间的联系,可保证所提取目标内部的完整性。虽然 DeepLabv3+模型也避免了该区域的内部空洞现象,但是对于一些稍大的边缘地物的提取,其内部也容易出现不连续的空洞点。此外,对于原始影像的红色线框内区域,所有方法都未能实现准确识别,主要原因是这种类似硬质地面的纹理可能会被识别为非建筑物特征,而实际上它们也代表一些低层建筑的屋顶部分或者与主体建筑相邻的微小建筑。

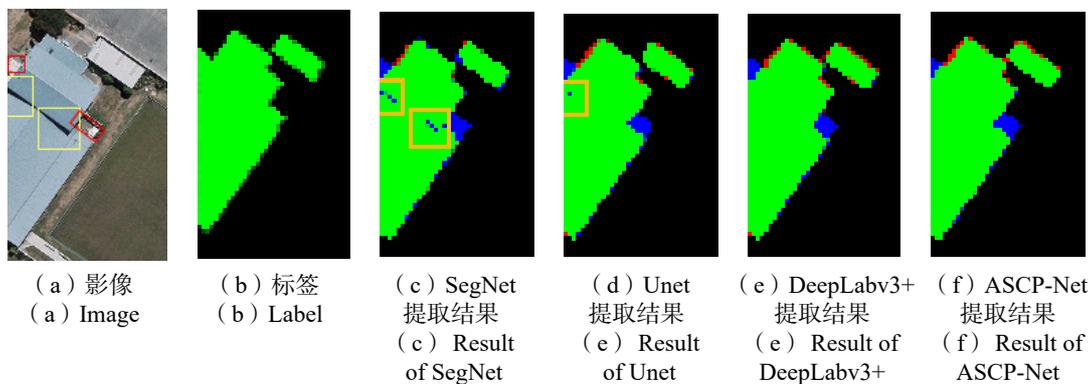


图 8 大尺度建筑的提取效果对比
 Fig.8 Comparison of the extraction effect of large-scale buildings

3.2 精度分析

不同模型对建筑物提取的定量结果如表 1 所示。

表 1 不同模型的精度评估
Tab.1 Accuracy evaluation of different models

模型	OA/%	Precision/%	Recall/%	F1/%	IoU/%	Kappa
FCN8s	94.6	90.8	87.9	89.2	80.7	0.737
DANet	89.1	91.1	72.2	80.5	67.3	0.638
SegNet	94.9	93.7	83.8	88.4	79.2	0.745
UNet	95.3	94.5	86.6	90.7	82.9	0.763
DeepLabv3+	96.5	92.5	93.2	92.8	86.7	0.771
ASCP-Net	97.4	94.9	94.3	94.6	89.8	0.787

DANet 几乎在所有指标上的表现最差, FCN8s 模型在 Precision 参数指标上的表现最差, 本文的模型在所有指标上都优于其他模型。交并比指标比 FCN8s 模型提升了 9.1 个百分点, 比 DANet 模型提升了 22.5 个百分点, 比 SegNet 模型提升了 10.6 个百分点, 比 UNet 模型提升了 6.9 个百分点, 比 DeepLabv3+ 模型提升了 3.1 个百分点; Kappa 值比 DANet 提升了 14.9 个百分点, 比 Unet 提升了 2.4 个百分点, 比 DeepLabv3+ 提升了 1.6 个百分点, 总体提升效果显著。定性分析和定量评价结果表明, ASCP-Net 在建筑物提取任务中具有更为优异的综合性能。

3.3 消融实验

为有效验证各个模块对最终实验结果的影响, 本文在 WHU Building Dataset 数据集上进行了消融实验。具体结果如表 2 所示。

表 2 各个模块对实验结果的影响
Tab.2 Influence of each module on the experimental results

方案	模块				IoU/%	Kappa
	Xception	ASPP	通道注意力	位置注意力		
1	√				84.6	0.713
2	√	√			87.1	0.751
3	√	√	√		88.2	0.768
4	√	√		√	89.1	0.776
5	√	√	√	√	89.8	0.787

注: √表示采用相应模块。

消融研究以 Xception 网络为基础, 依次增加 ASPP 模块、通道注意力模块和位置注意力模块进行实验。ASPP 模块可以提升网络对多尺度特征的提取能力, 同时一定程度上提高了网络的泛化能力; 位置注意力模块选择关注图像关键信息更丰富的空间区域, 通道注意力模块可以重点关注信息更为丰富的影像通道, 它们都可以进一步提升网络对高级特征的代表能力。由表 2 可以得出, 通过加入 ASPP 模块, IoU 指标和 Kappa 系数分别提升了 2.5 个百分点和 3.8 个百分点; 同时, 加入通道注意力模块和位置注意力模块会进一步提升网络的精度, 值得注意的是, 只加入位置注意力模块(方案 4)相比只加入通道注意力模块(方案 3)带来的精度提升更明显。因此, 在建筑物提取任务中, 位置信息要比通道信息更为重要。以上说明了各个模块的组合在 ASCP-Net 中均产生了积极作用, 有助于模型实现更精确的建筑物提取任务。

4 结束语

针对现有相关研究很少基于影像的高级语义特征对模型进行优化改进, 本文提出一种新的深度学习

模型架构 ASCP-Net, 可用于高分辨率遥感影像建筑物较精准的自动提取, 研究选用 WHU Building Dataset 为数据源, 采用融合了 ASPP 与 SCA 的非对称编解码结构模型进行建筑物提取, 并与 FCN8s、UNet、DeepLabv3+等经典语义分割模型方法进行实验对比, 结果表明:

1) 本文所提出的 ASCP-Net 模型对 WHU 数据集的总体准确率(97.4%)、精确度(94.9%)、召回率(94.3%)、F1 评分(94.6%)、IoU(89.8%)、Kappa 值(0.787)比其他方法要高, IoU 指标相比其他方法有显著提高。ASCP-Net 模型具有相对良好的性能, 在建筑物提取中具有潜在的应用前景;

2) 相比其他模型, 本文的方法在对图像边缘不完整建筑的分割中更具优势, 提取的边缘建筑更为完整;

3) 对于一些小地物的提取, 本文方法还存在一些不足。部分独立的微小建筑或者与主体建筑相邻的微小部分容易被漏分。

地物背景信息的复杂多样性为建筑物提取工作带来困难, 小地物的精确提取一直是语义分割任务中需要解决的难题。今后研究中将考虑在模型的解码器部分增加更多的低层特征来源, 并进一步与其他算法相结合, 以提高小地物的提取精度。

参考文献 (References)

- [1] XU P, TANG H, GE J, et al. ESPC_NASUnet: An End-to-end Super-resolution Semantic Segmentation Network for Mapping Buildings from Remote Sensing Images[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 5421-5435.
- [2] LIU Y, GROSS L, LI Z, et al. Automatic Building Extraction on High-resolution Remote Sensing Imagery Using Deep Convolutional Encoder-decoder with Spatial Pyramid Pooling[J]. *IEEE Access*, 2019, 7: 128774-128786.
- [3] LIU Y, ZHOU J, QI W, et al. Arc-net: An Efficient Network for Building Extraction from High-resolution Aerial Images[J]. *IEEE Access*, 2020, 8: 154997-155010.
- [4] GAVANKAR N, GHOSH S. Automatic Building Footprint Extraction from High-resolution Satellite Image Using Mathematical Morphology[J]. *European Journal of Remote Sensing*, 2018, 51(1): 182-193.
- [5] AVUDAIAMMAL R, ELAVENI P, SELVAN S, et al. Extraction of Buildings in Urban Area for Surface Area Assessment from Satellite Imagery Based on Morphological Building Index using SVM Classifier[J]. *Journal of the Indian Society of Remote Sensing*, 2020, 48(9): 1325-1344.
- [6] DONG Y, DU B, ZHANG L. Target Detection Based on Random Forest Metric Learning[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, 8(4): 1830-1838.
- [7] 赵翠晓, 陈曦, 杨辽, 等. 建筑物边界约束的航空影像DSM优化[J]. *遥感信息*, 2016, 31(4): 16-21.
ZHAO Cuixiao, CHEN Xi, YANG Liao, et al. DSM Optimisation of Aerial Imagery with Building Boundary Constraints[J]. *Remote Sensing Information*, 2016, 31(4): 16-21. (in Chinese)
- [8] 王宏涛, 王变利. 融合LiDAR和多光谱影像的建筑物自动检测[J]. *测绘科学*, 2015, 40(8): 64-67.
WANG Hongtao, WANG Bianli. Automatic Building Detection Based on LiDAR and Multispectral Images[J]. *Science of Surveying and Mapping*, 2015, 40(8): 64-67. (in Chinese)
- [9] AYTEKIN Ö, ZÖNGÜR U, HALICI U. Texture-based Airport Runway Detection[J]. *IEEE Geoscience and Remote Sensing Letters*, 2012, 10(3): 471-475.
- [10] DÉ SIR C, BERNARD S, PETITJEAN C, et al. One Class Random Forests[J]. *Pattern Recognition*, 2013, 46(12): 3490-3506.
- [11] 王振庆, 周艺, 王世新, 等. IEU-Net高分辨率遥感影像房屋建筑物提取[J]. *遥感学报*, 2021, 25(11): 2245-2254.
WANG Zhenqing, ZHOU Yi, WANG Shixin, et al. House Building Extraction from High-resolution Remote Sensing Images Based on IEU-Net[J]. *National Remote Sensing Bulletin*, 2021, 25(11): 2245-2254. (in Chinese)
- [12] YE Z, FU Y, GAN M, et al. Building Extraction from Very High Resolution Aerial Imagery Using Joint Attention Deep Neural Network[J]. *Remote Sensing*, 2019, 11(24): 2970.
- [13] 赵元昊, 赵莹莹, 刘东升, 等. 遥感影像建筑物提取多尺度特征深度学习网络[J]. *航天返回与遥感*, 2022, 43(4): 25-35.
ZHAO Yuanhao, ZHAO Yingying, LIU Dongsheng, et al. Building Extraction Deep Neural Network for Remote Sensing

- Images Based on Multi-scale Features[J]. *Spacecraft Recovery & Remote Sensing*, 2022, 43(4): 25-35. (in Chinese)
- [14] 乔文凡, 慎利, 戴延帅, 等. 联合膨胀卷积残差网络和金字塔池化表达的高分影像建筑物自动识别[J]. *地理与地理信息科学*, 2018, 34(5): 7.
QIAO Wenfan, SHEN Li, DAI Yanshuai, et al. Automatic Recognition of Buildings in High-resolution Image Based on Expansive Convolutional Residual Network and Pyramid Pooling Representation[J]. *Geography and Geo-Information Science*, 2018, 34(5): 7. (in Chinese)
- [15] 刘文祥, 舒远仲, 唐小敏, 等. 采用双注意力机制Deeplabv3+算法的遥感影像语义分割[J]. *热带地理*, 2020, 40(2): 303-313.
LIU Wenxiang, SHU Yuanzhong, TANG Xiaomin, et al. Remote Sensing Image Segmentation Using Dual Attention Mechanism Deeplabv3+ Algorithm[J]. *Tropical Geography*, 2020, 40(2): 303-313. (in Chinese)
- [16] CHOLLET F. Xception: Deep Learning with Depthwise Separable Convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 1251-1258.
- [17] HE K, ZHANG X, REN S, 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. IEEE, 2016: 770-778.
- [18] HE K, ZHANG X, REN S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916.
- [19] CHEN L, ZHANG H, XIAO J, et al. Sca-cnn: Spatial and Channel-wise Attention in Convolutional Networks for Image Captioning[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 5659-5667.
- [20] ZHANG Z, WANG Y. JointNet: A Common Neural Network for Road and Building Extraction[J]. *Remote Sensing*, 2019, 11(6): 696.
- [21] FU J, LIU J, TIAN H, et al. Dual Attention Network for Scene Segmentation[C]//2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 16-20, 2019, Long Beach, CA, USA. IEEE, 2019: 3146-3154.
- [22] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 8-10, 2015, Boston, MA, USA. IEEE, 2015: 3431-3440.
- [23] 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.
- [24] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional Networks for Biomedical Image Segmentation[C]//Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, October 5-9, 2015, Munich, Germany. Springer International Publishing, 2015: 234-241.
- [25] CHEN L, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with Atrous Separable Convolution for Semantic Image Segmentation[C]//2018 European Conference on Computer Vision (ECCV), September 8-14, 2018, Munich, Germany. ECCV, 2018: 801-818.

作者简介

于明洋, 男, 1978年生, 2005年获山东师范大学自然地理学专业硕士学位, 副教授。主要研究方向为地理信息工程研发、深度学习。E-mail: ymy@sdjzu.edu.cn。

通讯作者

徐海青, 男, 1997年生, 2019年毕业于安徽师范大学城市规划专业, 现在山东建筑大学测绘科学与技术专业攻读硕士学位。研究方向为遥感图像处理与智能解译。E-mail: 2021160105@stu.sdjzu.edu.cn。

(编辑: 庞冰)