doi: 10.6046/zrzyyg.2021219
引用格式: 王华俊,葛小三. 一种轻量级的 DeepLabv3+ 遥感影像建筑物提取方法[J]. 自然资源遥感,2022,34(2):128 - 135. (Wang H J, Ge X S. Lightweight DeepLabv3+ building extraction method from remote sensing images[J]. Remote Sensing for Natural Resources, 2022, 34(2): 128 - 135.)

一种轻量级的 DeepLabv3+ 遥感影像建筑物提取方法

王华俊, 葛小三
(河南理工大学测绘与国土信息工程学院,焦作 454003)

摘要: 快速从遥感影像中提取出具有较高精度的建筑物是遥感智能化应用服务的重要研究内容之一。针对 DeepLab 模型对遥感影像建筑物边缘分割不精确、分割大尺度目标存在孔洞现象、网络参数量大等问题,提出一种轻量级 DeepLabv3+ 模型的遥感影像建筑物提取方法。该方法使用轻量级网络 MobileNetv2 替换 DeepLabv3+ 的主干网络 Xception,从而减少参数量、提高训练速度; 对空洞空间金字塔池化(atrous spatial pyramid pooling, ASPP)的空洞率进行优化组合,提高多尺度语义特征提取效果。改进的模型在 WHU 和 Massachusetts 数据集上进行验证实验,结果表明,在 WHU 数据集中得到的交并比和 F1 分数分别为 82.37% 和 92.89%,比 DeepLabv3+ 分别提高 2.71 百分点和 2.14 百分点,在 Massachusetts 数据集中的交并比和 F1 分数比 DeepLabv3+ 分别提高 2.04 百分点和 2.32 百分点,训练参数量和训练时间减少,建筑物提取精度得到有效提高,能够满足快速提取高精度建筑物的要求。

关键词: 深度学习; 语义分割; 改进 ASPP; DeepLabv3+; MobileNetv2
中图法分类号: TP 79 **文献标志码**: A **文章编号**: 2097 - 034X(2022)02 - 0128 - 08

0 引言

遥感技术的发展使遥感影像空间分辨率提高,地物细节信息更丰富、几何结构和纹理特征等更明显[1],这导致了噪声相应增加。如何从高空间分辨率遥感影像上准确提取建筑物成为研究的热点。

随着深度学习的发展,建筑物提取成为遥感数据智能化应用处理的研究重点。众多学者提出了各种基于深度学习的建筑物提取方法,包括基于 U - Net 网络的方法[2]、结合模糊度和形态学指数的深度学习建筑物提取方法[3],基于特征增强和 ELU 神经网络的建筑物提取方法[4]、结合深度残差网络结构和金字塔式层级连接的建筑物提取方法[5]、基于编解码结构的卷积神经网络的方法[6]、基于 R - MCN 模型的方法[7]等等。但这些方法普遍存在网络参数量大、训练时间长、算法速度难以得到提升的问题。

近年来不断涌现出大量深度学习模型,主要集中在以下几方面: ①数据处理速度提升与训练参数量减少的模型,如 LeNet[8], AlexNet[9], VGGNet[10] 和 ResNet[11]等; ②能减少模型训练参数量的轻量级网络 SqueezeNet[12], MnasNet[13] 和 MobileNet[14]等; ③能提高准确率与能进行多尺度特征提取的 DeepLab 系列模型等。在语义分割中,DeepLab 系列是常用模型之一,主要用于逐像素分类。Chen 等[15]提出 DeepLabv1,结合深度卷积神经网络(deep convolution neural network, DCNN)和概率图模型,提出空洞卷积(atrous convolution),算法速度和准确率较高; Chen 等[16]提出 DeepLabv2,使用空洞空间金字塔池化(atrous spatial pyramid pooling, ASPP)扩大感受野、降低计算量; Chen 等[17]提出 DeepLabv3+,引入编码 - 解码(Encoder - Decoder)形式进行多尺度信息融合,使物体边界分割效果更好; 王俊强等[18]提出 DeepLabv3+ 语义分割与全连接条件随机场相结合的提取方法,能从高分辨率遥感影像中获得典型要素边界信息; 刘文祥等[19]在网络中引入双注意力机制模块(dual attention mechanism module, DAMM),提出将 DAMM 结构与 ASPP 层串联和并联 2 种不同连接方式的网络模型,能有效改善 DeepLabv3+ 的不足。但是利用 DeepLabv3+ 提取遥感影像建筑物仍存在边界信息较粗糙、拟合速度慢、小尺度目标分割模糊和大尺度目标分割有孔洞等问

收稿日期: 2021 - 07 - 14; **修订日期**: 2021 - 11 - 02
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)资助。
第一作者: 王华俊(1997 -),男,硕士研究生,主要从事遥感、地理信息服务技术方面的研究。Email: 1029803406@qq.com。
通信作者: 葛小三(1971 -),男,教授,主要从事地理信息服务技术、空间数据处理与分析方面的研究。Email: gexiaosan@163.com。

题。

本文针对 DeepLabv3+ 提取建筑物存在边界信息粗糙和训练量大等问题,提出一种轻量级 DeepLabv3+ 模型的遥感影像建筑物提取方法,使用 MobileNetv2[20] 替换 DeepLabv3+ 的主干网络,并将 ASPP 模块中空洞卷积的空洞率组合改为 4,8,12,16,以期提高 DeepLabv3+ 的训练速度和目标分割精确度,使模型能达到更好的建筑物提取效果。

1　方法与原理

本文提出一种轻量级 DeepLabv3+ 模型的遥感建筑物提取方法:使用轻量级网络 MobileNetv2 替换原模型的主干网络 Xception;在此基础上,将 ASPP 中空洞卷积的空洞率进行优化组合,提出一种新的 ASPP 模块结构,通过调整模型中的学习率和卷积核等参数,使模型达到更优的建筑物提取效果。

1.1　DeepLabv3+ 模型

DeepLabv3+ 引入 Encoder – Decoder 结构,主要分为编码(Encoder)部分和解码(Decoder)部分。在此结构中,引入可任意控制 Encoder 提取特征的分辨率,主干网络将原始 Xception[21] 进行改进,并将深度可分离卷积应用到 ASPP 和 Decoder 模块中。

1)Encoder 部分。在主干 DCNN 里使用串行空洞卷积,在图像经过主干 DCNN 后,得到的结果分别传入 Decoder 和并行的空洞卷积用不同空洞率(rate)的空洞卷积进行特征提取,提取后合并,用 1 × 1 卷积压缩特征,进入 Decoder 部分,并使用双线性插值方法进行 4 倍上采样。

2)Decoder 部分。一部分是 DCNN 经过 4 倍上采样输出的特征,另一部分是 DCNN 输出以后,经过并行空洞卷积后的结果。为防止 Encoder 得到的高级特征被弱化,用 1 × 1 卷积对低级特征降维,2 个特征融合后,用 3 × 3 卷积进一步融合特征,使用双线性插值方法进行 4 倍上采样,得到与原始图像相同大小的分割预测结果。

DeepLabv3+ 在语义分割任务中将 Xception 模型进行改进,与改进之前的 Xception 相比,DeepLabv3+ 的输入流保持不变,但中间流更多;所有的最大池化被深度可分离卷积(depthwise separable convolution)替代;在每个 3 × 3 深度卷积之后,增加批标准化(batch norm, BN)和整流线性单元(rectified linear units, ReLU)。Xception 结构如图 1 所示。

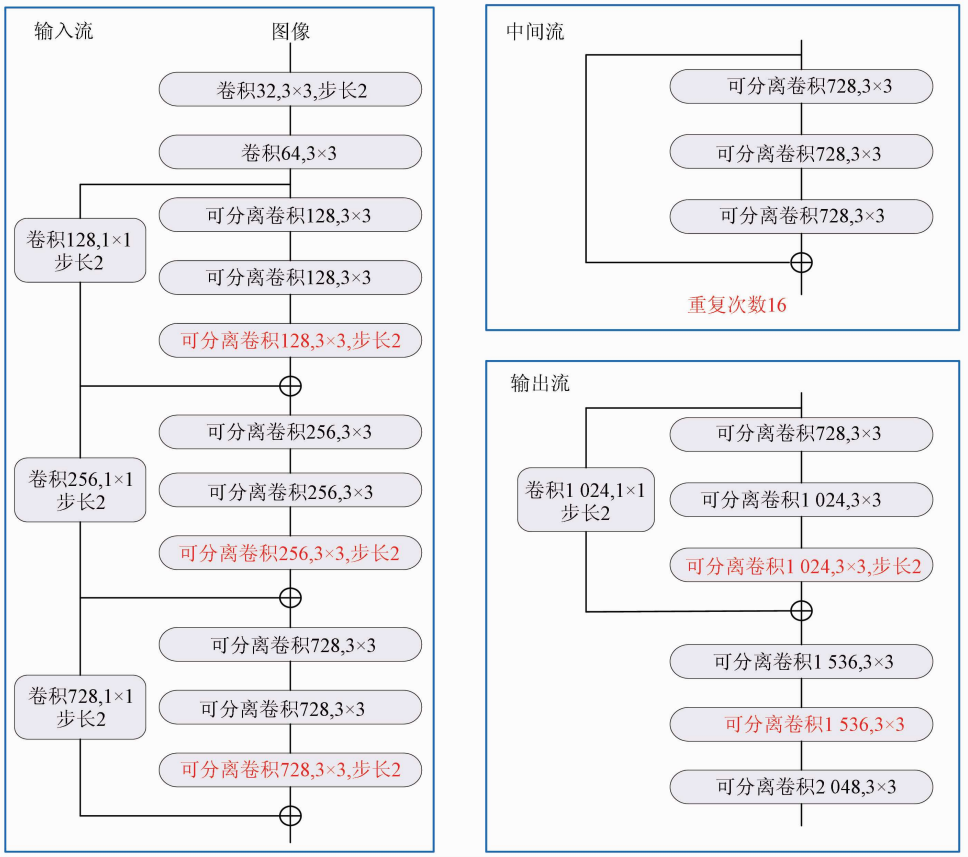


图 1　DeepLabv3+ 中的 Xception 网络结构

Fig. 1　Xception network structure in DeepLabv3+

1.2　改进的 DeepLabv3+ 网络

为有效提取遥感影像建筑物在不同尺度下的语义信息,本文使用 Encoder – Decoder 结构,并将深度可分离卷积应用在 ASPP 和 Decoder 模块中减少运

算量,提高 Encoder – Decoder 网络对遥感影像建筑物提取的运行速率和健壮性。为增大感受野而不损失信息,使用空洞卷积增加每个卷积的输出信息量,并通过空洞卷积平衡精度和耗时。为减少训练参数量和训练时间,将 DeepLabv3+ 的主干网络 Xception 替换为轻量级网络 MobileNetv2; 为增强网络对不同大小目标的分割能力,将 ASPP 模块的空洞卷积中的 6,12,18 组合的空洞率改为 4,8,12,16 的组合。经过主干网络后的处理结果传到 Decoder 层,Decoder 层的主要结构不发生改变。在上述工作完成后,对整个 Encoder 层进行优化,从而使替换后的网络能准确从遥感影像中提取出更高精度的建筑物。改进 DeepLabv3+ 模型结构如图 2 所示(其中包含与原模型对比)。

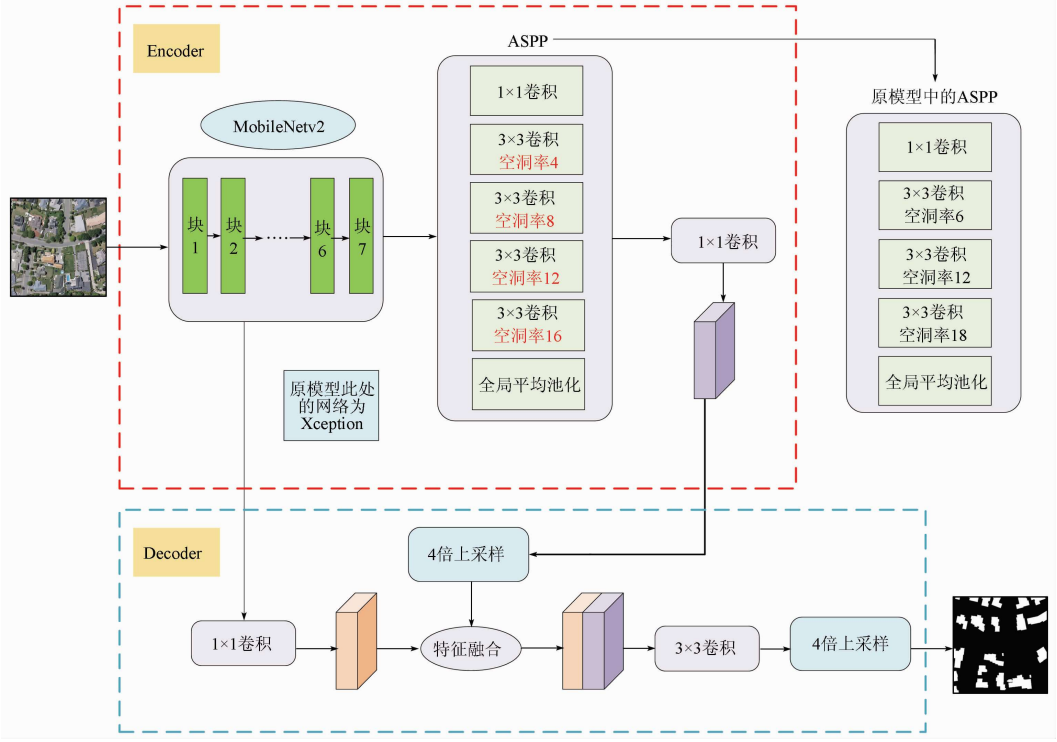


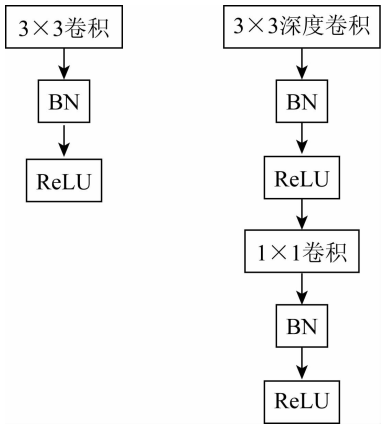
图 2 本文方法网络结构

Fig. 2 The network structure of this method

1.2.1 深度可分离卷积

本文主干网络核心内容为深度可分离卷积。在含有大量噪声和信息的遥感影像处理中,深度可分离卷积与常规卷积操作相比,有参数量少、训练时间短等优点,在精度保持不变的情况下,能更好地在遥感影像中快速提取建筑物的特征信息。

卷积核大小代表感受野大小,卷积核越大感受野越大,若卷积核过大,会使计算量增加,对含有大量信息的遥感影像进行处理时,随着网络深度的增加,计算能力和训练速度会逐渐降低,所以在本文网络中,使用 1 × 1 卷积和 3 × 3 卷积进行卷积操作。标准卷积是将过滤和输入合并为一组新的输出,而深度可分离卷积是由深度卷积和逐点卷积 2 部分相结合,一个用于过滤,另一个用于合并,以此用来提取特征。深度卷积是一个卷积核对应一个输入通道,独立对每个输入通道做空间卷积; 逐点卷积用于结合深度卷积的输出,即每个通道单独做卷积,通道数不变,然后将第一步的卷积结果用 1 × 1 卷积跨通道进行组合。卷积操作如图 3 所示。



(a) 标准卷积 (b) 深度可分离卷积

图 3 标准卷积和深度可分离卷积

Fig. 3 Standard convolution and depth separable convolution

在本文深度可分离卷积中,首先采用深度卷积对不同输入通道分别进行卷积,然后采用逐点卷积将上面的输出进行结合,整体效果和一个标准卷积相同,但是会大大减少计算量和模型参数量,更适合提取建筑物特征。

1.2.2　主干网络

DeepLabv3+ 的主干网络 Xception 对种类多的提取任务有较好效果，但其网络复杂度高、参数量大，而遥感影像复杂、信息量大，随着训练的进行，参数量会逐渐加大，故 Xception 不适合提取遥感影像建筑物信息，因此使用轻量级网络 MobileNetv2 将 DeepLabv3+ 的主干网络 Xception 替换，其网络体积小、参数量少，可以更快速、精准地从大量遥感影像信息中提取建筑物。

MobileNetv2 网络（图 4 和表 1）有更小的体积、更少的计算量、更高的准确率、更快的速度和多种应用场景等优点，在遥感影像建筑物提取中具有极大优势。MobileNetv2 引入线性瓶颈结构（linear bottlenecks）和反向残差结构（inverted residuals），构成线性瓶颈倒残差结构，使遥感影像建筑物提取的参数量和计算量减少、训练速度和提取精度更高。在此结构中，反向残差结构将输入的低维通过 1 ×1 卷积进行升维，使用轻量级深度卷积进行过滤并提取特征图，并利用 1 ×1 卷积进行降维。为避免降维后 ReLU 损失建筑物提取精度和破坏建筑物特征，在深度卷积处理后使用线性瓶颈结构替换 ReLU 进行降维，并使用限制最大输出值为 6 的 ReLU6 替换普通 ReLU。MobileNetv2 中添加扩张倍数控制网络大小，虽然使网络结构更深，但计算量更少，能节省训练时间和资源，对遥感影像中建筑物提取有很大优势。

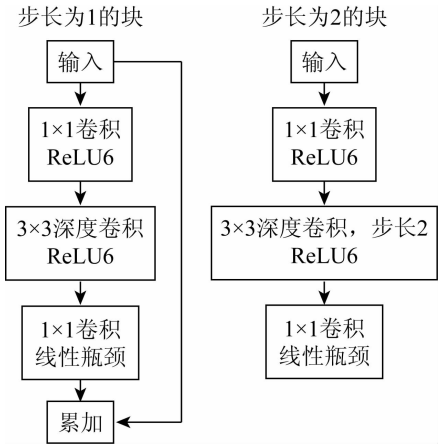


图 4　MobileNetv2 网络结构

Fig. 4　MobileNetv2 network structure

表 1　MobileNetv2 网络结构及参数

Tab. 1　Network structure and parameters of MobileNetv2

输入	操作	扩张倍数	输出通道数	重复次数	步长
$224^2 \times 3$	卷积层	—	32	1	2
$112^2 \times 32$	瓶颈层	1	16	1	1
$112^2 \times 16$	瓶颈层	6	24	2	2
$56^2 \times 24$	瓶颈层	6	32	3	2
$28^2 \times 32$	瓶颈层	6	64	4	2
$28^2 \times 64$	瓶颈层	6	96	3	1
$14^2 \times 96$	瓶颈层	6	160	3	2
$7^2 \times 160$	瓶颈层	6	320	1	1
$7^2 \times 320$	1 ×1 卷积层	—	1 280	1	1
$7^2 \times 1\,280$	7 ×7 平均池化层	—	—	1	—
$1 \times 1 \times k$	1 ×1 卷积层	—	k	—	—

1.2.3　空洞空间金字塔池化

ASPP 是由空洞卷积与空间金字塔池化（spatial pyramid pooling，SPP）[22] 融合形成，能有效提取遥感影像中多尺度语义特征，从而在遥感影像建筑物提取中被广泛使用。

DeepLabv3+ 中 ASPP 模块空洞卷积的空洞率组合为 6，12，18，随着主干网络对特征提取的进行，特征图分辨率会逐渐减小，6，12，18 的组合不能更有效地提取多分辨率特征图特征，没有设置较小的空洞率，导致分割小目标的能力欠缺，从而使网络对不同大小分割目标的分割能力较弱。为更有效地提取多分辨率特征图特征，提高不同大小分割目标的分割能力，本文将空洞卷积的空洞率组合改为 4，8，12，16，使较大的分割目标能被较大空洞率的卷积核分割，相反，较小的目标可以被较小空洞率的卷积核分割，较小的空洞率可使特征提取更有效。经过主干网络 MobileNetv2 得到的特征图输入到本文 ASPP 模块中，经过 1 ×1 卷积操作、不同空洞率 3 ×3 卷积操作和最后的池化操作后，不同大小的分割目标依次被卷积提取出特征图，将输出的 6 张特征图进行融合，得到由本文 ASPP 产生的特征图。本文改进的 ASPP 结构如图 5 所示。

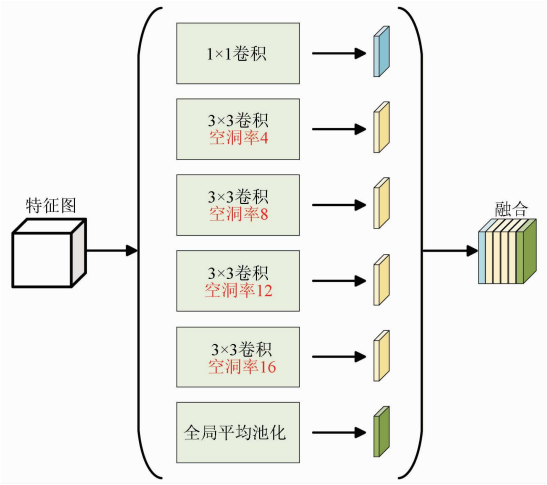


图 5　本文的 ASPP 结构

Fig. 5　ASPP structure of in this paper

2　实验与分析

2.1　实验数据

实验所用训练数据集为武汉大学季顺平团队制作的 WHU 建筑数据集[23] 和 Volodymyr 制作的 Massachusetts 建筑数据集[24]。

WHU 数据集主要包含航空图像、覆盖 1 000 m^2 卫星图像、栅格标签和矢量地图，航空数据集由 22 万多个独立建筑物组成，这些建筑物由空间分辨率为 0.075 m、覆盖范围为 450 m^2 的新西兰克赖斯特彻奇航空图像中提取，此地区包含多种地物种类，各种不同颜色、大小和用途的建筑类型。数据集将大部分航空图像（包含 187 000 栋建筑物）降至 0.3 m 空间分辨率，并将其无缝裁剪为 512 像素 ×512 像素的 8 188 个无重叠图块，同时将数据集分为训练集、验证集和测试集，其中用于训练的图像有 4 736 张、用于验证的图像有 1 036 张、用于测试的图像有 2 416 张。

Massachusetts 建筑物数据集由波士顿地区的 151 张航拍图像组成，每幅图像为 1 500 像素 × 1 500 像素、空间分辨率为 1 m、单张覆盖面积为 2.25 km^2，整个数据集覆盖约 340 km^2。数据集预先划分为含有 137 张图像的训练集、10 张图像的验证集和 4 张图像的测试集。为使 Massachusetts 数据集与 WHU 数据集中的图像大小保持相同，将 Massachusetts 数据集的每张图像分别裁剪为 9 张 512 像素 × 512 像素大小的图像，裁剪后的图像数量为 1 359 张，并将其进行旋转，旋转后的图像数量为 2 718 张。

WHU 与 Massachusetts 数据集实例图像如图 6 所示。WHU 数据集影像的空间分辨率高于 Massachusetts 数据集影像的空间分辨率，并且 Massachusetts 数据集中的建筑物密度高，建筑物大小相对更小，更能体现出深度学习网络提取小型建筑物的能力。

（a）WHU 数据集示例 1　（b）WHU 数据集示例 2　（c）Massachusetts 数据集示例 1　（d）Massachusetts 数据集示例 2

图 6　WHU 与 Massachusetts 数据集实例图像

Fig. 6　Example images of WHU and Massachusetts datasets

2.2　实验设置

实验所用机器主要软硬件配置见表 2。

表 2　计算机配置

Tab. 2　Computer configuration

项目	配置
CPU	Intel(R) Core(TM) i5 – 10200H CPU @ 2.40 GHz
GPU	NVIDIA GeForce GTX 1650 Ti
RAM	16 G
操作系统	64 位 Windows10
开发语言	Python 3.7
深度学习框架	Tensorflow – GPU 1.13.1

实验主要设置：定义输入图片的高和宽及需要分割的种类数量，读取输入的图像和标签，进行归一化和大小调整，使用迁移学习思想获取主干特征提取网络的权重，并将数据集图像随机打乱，送入网络进行训练。初始学习率设置为 0.000 3；每次送入网络训练的图像批次为 4，迭代次数为 100 次，每迭代 2 次保存一次训练细节；使用交叉熵作为损失函数；主干网络使用 MobileNetv2 网络；优化器选择 Adam 优化器，该优化器能动态调整每个参数的学习率；激活函数使用 ReLU 激活函数；膨胀系数 α 设为 1。评价指标为交并比（intersection over union，IoU）和 F1 分数，交叉熵损失函数、IoU 和 F1 分数的公式分别为：

$$Loss = \frac{1}{N}\sum_{i=1}^{N} L_i = \frac{1}{N}\sum_{i=1}^{N} -[y_i \cdot \ln(p_i) + (1 - y_i) \cdot \ln(1 - p_i)] , \tag{1}$$

$$IoU = \frac{TP}{TP + FP + FN} , \tag{2}$$

$$F1 = \frac{2TP}{2TP + FP + FN} , \tag{3}$$

式中：y_i 为样本 i 的标签，建筑物为 1，背景为 0；p_i 为样本 i 预测为建筑物的概率；N 为样本数量；TP 为正确提取建筑物的样本数量；FP 为把背景像素错误提取为建筑物像素的样本数量；FN 为把建筑物像素错误提取为背景像素的样本数量。

2.3　结果与分析

DeepLabv3+ 中 ASPP 模块空洞卷积的空洞率为 6，12，18，本文 ASPP 模块的空洞率组合为 4，8，12，

16。在本文模型基础上,依次使用 6,12,18 和 4,8,12,16 的空洞率组合进行实验,验证 ASPP 模块的改进在网络模型中的效果。在使用原空洞率组合的情况下,测试结果的 IoU 值为 80.54%,使用改进的 ASPP 空洞率组合的情况下,测试结果的 IoU 值为 82.37%,比原组合提高 1.83 百分点,所以使用 4,8,12,16 的空洞率组合对遥感影像建筑物有更优的提取效果。

表 3 为不同模型分别在 2 个数据集中的评价指标值,其评价指标主要为 IoU 与 F1 分数。较其他经典模型相比,本文方法在 2 个数据集中实验结果的交并比均较高,遥感影像建筑物提取精度得到进一步提高。相比于 DeepLabv3+ 模型,本文方法在 WHU 数据集中的 IoU 提升 2.71 百分点、F1 分数提高 2.14 百分点,在 Massachusetts 数据集中的 IoU 提升 2.04 百分点、F1 分数提高 2.32 百分点,U – Net 与 SegNet 的评价指标值较低。总体上,本文提出的方法与其他模型相比均有所提升,对建筑物提取具有较高的有效性。由于本文模型使用的是轻量级网络,与 DeepLabv3+ 模型的主干网络 Xception 相比,本文方法的主干网络参数量少,所以训练时间更短,能有效提升模型训练速度。

表 3　建筑物提取评价结果

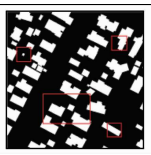
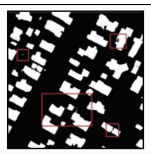
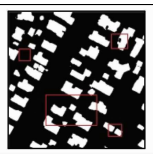
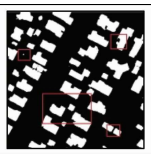
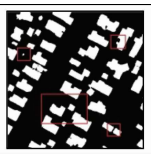
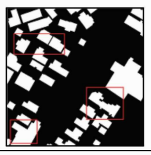
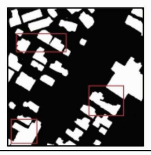
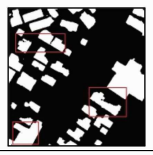
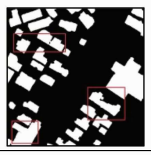
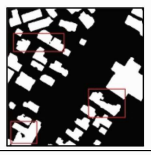
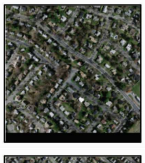
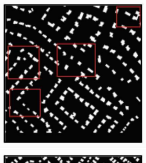
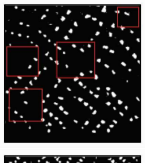
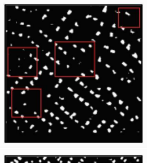
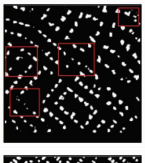
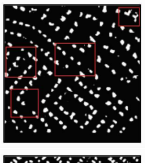
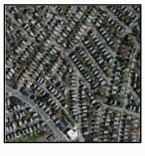
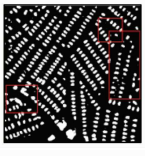
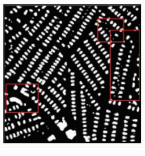
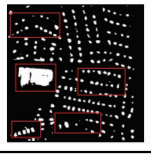
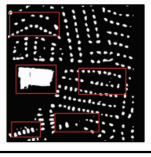
Tab. 3　Building extraction evaluation results

数据集	模型	*IoU*/%	F1 分数/%	训练时间/h
WHU	U – Net	77.28	88.23	10.53
	SegNet	77.15	89.14	15.64
	PSPNet	78.23	89.53	11.32
	DeepLabv3+	79.66	90.75	10.65
	本文方法	82.37	92.89	6.15
Massachusetts	U – Net	71.25	82.11	12.88
	SegNet	71.87	83.84	17.32
	PSPNet	72.15	82.26	13.55
	DeepLabv3+	74.58	84.43	12.72
	本文方法	76.62	86.75	7.35

建筑物提取结果如表 4 所示,在预测结果中随机选取图像作为本次实验结果的对比分析。在 WHU 数据集中 U – Net 和 SegNet 的提取结果相似,整体效果不佳,对小型建筑物的提取有时会失效或提取面积极小,在 Massachusetts 数据集中也可看出其对小型建筑物有提取效果不佳、提取面积极小的情况,并出现多处建筑漏提现象。较其他经典模型相比,DeepLabv3+ 模型的提取效果较好,对大型建筑物边界的提取精度更高,与本文方法相比,对小型建筑物的提取效果不佳、提取面积小、数量少。由于

表 4　WHU 和 Massachusetts 数据集建筑物提取结果

Tab. 4　Building extraction results of WHU and Massachusetts dataset

数据集	序号	原始图像	标签	U – Net	SegNet	DeepLabv3+	本文方法
WHU	1						
	2						
	3						
Massa – chusetts	1						
	2						
	3						

本文对 ASPP 中的空洞卷积设置较小的空洞率,因此本文方法对小型建筑物的提取面积有所增大,比 DeepLabv3+模型提取的建筑物数量更多,改善了 DeepLabv3+漏提和少提现象,边界信息进一步提高,优于 DeepLabv3+模型提取效果; 对建筑物错误提取率较低,提取出的建筑物完整度较高,总体上提取效果较好。但是对小型建筑物的提取精度仍然有待提高,小型建筑物的边界信息提取不够完善,对具有复杂边界的建筑物提取时,其边界细节信息提取不够精细,对大型建筑物提取时偶尔会出现一些孔洞或提取模糊现象。

3 结论和展望

针对 DeepLabv3+网络参数量大的问题,本文对 DeepLabv3+中的主干网络进行替换,利用 MobileNetv2 网络轻便的特点,减少网络参数量、优化网络结构,实验结果也证明了本文方法的有效性,训练速度和精度得到有效提升; 对网络中 ASPP 模块进行调整,将原 ASPP 模块中 6,12,18 的空洞率调整为 4,8,12,16 的组合,经过对 ASPP 模块的实验结果可得出,本文改进空洞率的 ASPP 模块对建筑物的提取效果优于原空洞率组合的提取效果。本文方法总体上对遥感影像建筑物的提取精度较高、参数量少、训练成本更低,能更有效提取遥感影像建筑物。由实验结果可看出,本文网络对建筑物提取仍然存在不足,在后续研究中,继续对 ASPP 中的空洞率组合和主干网络参数进行调整实验,使其对小型建筑物能够达到更好的提取效果; 根据边界损失等思想进一步思考,提出能够提高边界提取精度的新方法; 考虑在本文模型基础上加入其他结构和机制,以此加强网络的健壮性和提取效果、能更有效地改善提取模糊和孔洞现象。

参考文献(References):

[1] 范荣双,陈 洋,徐启恒,等. 基于深度学习的高分辨率遥感影像建筑物提取方法[J]. 测绘学报,2019,48(1):34 -41.
Fan R S,Chen Y,Xu Q H,et al. A high - resolution remote sensing image building extraction method based on deep learning[J]. Acta Geodaetica et Cartographica Sinica,2019,48(1):34 -41.

[2] 张浩然,赵江洪,张晓光. 利用 U - net 网络的高分遥感影像建筑提取方法[J]. 遥感信息,2020,35(3):143 -150.
Zhang H R,Zhao J H,Zhang X G. High - resolution image building extraction using U - net neural network[J]. Remote Sensing Information,2020,35(3):143 -150.

[3] 许泽宇,沈占锋,李 杨,等. 结合模糊度和形态学指数约束的深度学习建筑物提取[J]. 地球信息科学学报,2021,23(5):918 -927.
Xu Z Y,Shen Z F,Li Y,et al. Building extraction by deep learning method combined with ambiguity and morphological index constraints[J]. Journal of Geo - Information Science,2021,23(5):918 -927.

[4] 唐 璎,刘正军,杨 懿,等. 基于特征增强和 ELU 的神经网络建筑物提取研究[J]. 地球信息科学学报,2021,23(4):692 -709.
Tang Y,Liu Z J,Yang Y,et al. Research on building extraction based on neural network with feature enhancement and ELU activation function[J]. Journal of Geo - Information Science,2021,23(4):692 -709.

[5] 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.
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.

[6] 陈凯强,高 鑫,闫梦龙,等. 基于编解码网络的航空影像像素级建筑物提取[J]. 遥感学报,2020,24(9):1134 -1142.
Chen K Q,Gao X,Yan M L,et al. Building extraction in pixel level from aerial imagery with a deep encoder - decoder network[J]. Journal of Remote Sensing,2020,24(9):1134 -1142.

[7] 叶沅鑫,谭 鑫,孙苗苗,等. 基于增强 DeepLabV3 网络的高分辨率遥感影像分类[J]. 测绘通报,2021(4):40 -44.
Ye Y X,Tan X,Sun M M,et al. High - resolution remote sensing image classification based on improved DeepLabV3 network[J]. Bulletin of Surveying and Mapping,2021(4):40 -44.

[8] Lecun Y,Bottou L. Gradient - based learning applied to document recognition[J]. Proceedings of the IEEE,1998,86(11):2278 -2324.

[9] Krizhevsky A,Sutskever I,Hinton G. ImageNet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems,2012,25(2):1097 -1105.

[10] Simonyan K,Zisserman A. Very deep convolutional networks for large - scale image recognition[J]. Computer Science,2014:1409 -1556.

[11] He K M,Zhang X Y,Ren S Q,et al. Deep residual learning for image recognition[EB/OL]. (2015 -12 -10)[2021 -05 -12]. https://arxiv.org/abs/1512.03385.

[12] Iandola F N,Han S,Moskewicz M W,et al. SqueezeNet:AlexNet - level accuracy with 50x fewer parameters and <0.5 MB model size[EB/OL]. (2016 -11 -04)[2021 -05 -12]. https://arxiv.org/abs/1602.07360.

[13] Tan M,Chen B,Pang R,et al. MnasNet:Platform - aware neural architecture search for mobile[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition,IEEE,2019:2820 -2828.

[14] Howard A G,Zhu M,Chen B,et al. MobileNets:Efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017 -04 -17)[2021 -05 -12]. https://arxiv.org/abs/1704.04861.

[15] Chen L C,Papandreou G,Kokkinos I,et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[J]. Computer Science,2014(4):357 -361.

[16] 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.

[17] Chen L C,Zhu Y,Papandreou G,et al. Encoder - decoder with atrous separable convolution for semantic image segmentation[C]// Conference on Computer Vision(ECCV),2018:801 - 818.

[18] 王俊强,李建胜,周华春,等. 基于 DeepLabv3+ 与 CRF 的遥感影像典型要素提取方法[J]. 计算机工程,2019,45(10):260 - 265,271.
Wang J Q,Li J S,Zhou H C,et al. Typical element extraction method of remote sensing image based on DeepLabv3+ and CRF [J]. Computer Engineering,2019,45(10):260 - 265,271.

[19] 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 W X,Shu Y Z,Tang X M,et al. Remote sensing image segmentation using dual attention mechanism DeepLabv3+ algorithm[J]. Tropical Geography,2020,40(2):303 - 313.

[20] Sandler M,Howard A,Zhu M,et al. Mobilenetv2:Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE,2018:4510 - 4520.

[21] Chollet F. Xception:Deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition,2017:1800 - 1807.

[22] He K,Zhang X,Ren S,et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2014,37(9):1904 - 1916.

[23] 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, 2018,57(1):574 - 586.

[24] Mnih V. Machine learning for aerial image labeling[D]. Toronto: University of Toronto,2013.

Lightweight DeepLabv3+ building extraction method from remote sensing images

WANG Huajun, GE Xiaosan

(*School of Surveying and Mapping and Land Information Engineering*, *Henan Polytechnic University*, *Jiaozuo* 454003, *China*)

Abstract: Fast extraction of buildings with high accuracy from remote sensing images is an important research of remote sensing intelligent application services. To address the problems of imprecise segmentation of building edge in remote sensing images, holes in large - scale target segmentation, and a large amount of network parameters in the DeepLab model, a lightweight DeepLabv3 + model for building extraction from remote sensing images is proposed. In this method, the lightweight network MobileNetv2 is used to replace Xception, the backbone network of DeepLabv3 +, so as to reduce the number of parameters and improve the training speed; The hole rate of hole convolution in ASPP is optimized to improve the effect of multi - scale semantic feature extraction. The improved model has been tested on WHU and Massachusetts data sets. The results show that the IOU and F1 score in WHU dataset are 82.37% and 92.89%, respectively, 2.71 percentage points and 2.14 percentage points higher than those of DeepLabv3+, 2.04 percentage points, and 2.32 percentage points higher than those of DeepLabv3+ in Massachusetts data set. The number of training parameters and training time is reduced, and the accuracy of the building extraction is effectively improved, which can meet the requirements of fast extraction of high - precision buildings.

Keywords: deep learning; semantic segmentation; improved ASPP; DeepLabv3+; MobileNetv2

(责任编辑: 张 仙)