

基于改进DeepLabv3+的高分辨率遥感影像屋顶提取方法

丁澎涛, 朱习军

青岛科技大学, 信息科学技术学院, 山东 青岛

收稿日期: 2023年2月24日; 录用日期: 2023年5月5日; 发布日期: 2023年5月16日

摘要

DeepLabv3+网络能够有效地解决高分辨率遥感图像语义分割的挑战。经过ResNet50骨干网络的支持,我们对DeepLabv3+模型进行了深入的研究,利用Adam梯度下降法和RELU激活函数,有效地处理了遥感影像中的建筑屋顶,提高了语义分割的精度和速度,能够更快地收敛到最优解。同时,空洞空间金字塔池化模块(ASPP)与解码器模块(decoder)的普通卷积部分被并行加权的空洞卷积代替,从而减少参数数量,提升模型的性能。IoU的准确度达到89.2%,超过DeepLabv3+算法,降低了特征提取的误差,同时也大大减少了细节信息的丢失,为最终的语义分割带来了显著的改善。

关键词

遥感图像, DeepLabv3+, 语义分割, 特征融合, 通道注意力

Roof Extraction Method of High-Resolution Remote Sensing Image Based on Improved DeepLabv3+

Pengtao Ding, Xijun Zhu

School of Information Technology, Qingdao University of Science and Technology, Qingdao Shandong

Received: Feb. 24th, 2023; accepted: May 5th, 2023; published: May 16th, 2023

Abstract

DeepLabv3+network can effectively solve the challenge of semantic segmentation of high-resolution remote sensing images. With the support of the ResNet50 backbone network, we have conducted

in-depth research on the DeepLabv3+ model. Using the Adam gradient descent method and RELU activation function, we have effectively processed the building roof in remote sensing images, improved the accuracy and speed of semantic segmentation, and can quickly converge to the optimal solution. At the same time, the common convolution part of the hole space pyramid pooling module (ASPP) and the decoder module (decoder) are parallelly weighted to reduce the number of parameters and improve the performance of the model. The accuracy of IoU reaches 89.2%, which exceeds the DeepLabv3+ algorithm, reduces the error of feature extraction, and also greatly reduces the loss of detail information, bringing significant improvement to the final semantic segmentation.

Keywords

Remote Sensing Image, DeepLabv3+, Semantic Segmentation, Feature Fusion, Channel Attention

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

传统的建筑物提取多是将使用手工提取的特征当做判断标准,同时基于边缘检测或者影像特征。在处理复杂的环境中,往往无法获得令人满意的效果。之后有人提出使用概率图模型,同时把卷积操作设置扩张率,这样为后来研究者对模型进行优化提供了很好地思路[1] [2]。通过精确地分割和提取建筑屋顶可以大大改善地理信息的处理效率,从而为城市规划、资源检测以及国家安全等领域带来极大的便利[3] [4] [5] [6] [7]。在本研究中,首先使用遥感图像获取了建筑物屋顶的图像数据,然后使用 deeplabv3+模型对这些数据进行了语义分割。提出的模型在 ASPP 模块中添加注意力机制,并将空洞卷积替换为并行加权空洞卷积,进一步提高模型的分割精度。利用遥感技术可以获得大量的建筑物屋顶信息,这些信息具有丰富的语义和纹理特征,但由于存在大量的冗余数据,所以如何处理海量高质量遥感影像,并且得到有效的信息在遥感领域中显得尤为重要[8] [9] [10] [11]。

2. 研究方法

2.1. DeepLabv3+网络

Google 的 deeplab 已经发布了四个版本,其中 DeepLabv3 基于 DeepLabv2 的基础上,开发出 ASPP 模型[12],但如果 3×3 的空洞卷积无法达到特征图的尺寸,就会导致 1×1 的卷积失效,从而无法完整地捕捉到图像的信息。DeepLabv3 通过增加一个 1×1 的卷积层,大大改善了 aspp 模块的性能,使得图片可以达到更高的分辨率,从而提高了图像处理的效率。DeepLabv3+网络本质上是基于 DeepLabv3 网络进行的改进,在原有的基础上添加了解码器(Decoder)模块,通过结合底层的特征信息来增强分割效果[13] [14] [15] [16]。

DeepLabv3+架构由两个核心部分组成:编码器和解码器。编码器模块负责提取高层特征,它的主要网络结构是 ResNet50,使用空洞卷积来取代传统的标准卷积,可以大大拓宽网络的感知范围。空洞卷积的本质就是带有空洞的卷积,达到在无需大幅度缩小图像尺寸的前提下获得更大的感受野范围的目的。

DCNN (深度卷积神经网络)利用 ASPP 模块提取出的复杂的多尺度上下文数据,可以提供更为全面的多尺度上下文信息。ASPP 模块由 5 个分支组成,包括一个全局平均池化和四个空洞卷积层,其中一个卷

积层是 1×1 的, 其余三个都是 3×3 的空洞卷积, 只不过空洞率分别是 6 的 1、2 和 3 倍数。之后把得到的特征图进行特征融合, 然后用 1×1 的卷积降低通道维度。使用双线性插值后将这些值与从编码器模块提取的低级特征相关联, 以达到空间分辨率的一致性。值得一提的是, 为了提高编码器模块的性能, 我们采用 1×1 卷积层来降低浅层特征的维度, 从而避免高级特征受到损害。经过编码器模块的深度学习和输入学习, 我们可以将其中的特征进行 3×3 卷积处理, 并使用四倍双线性上采样技术来提升预测精度, 从而构建一个完整的网络。如图 1 所示:

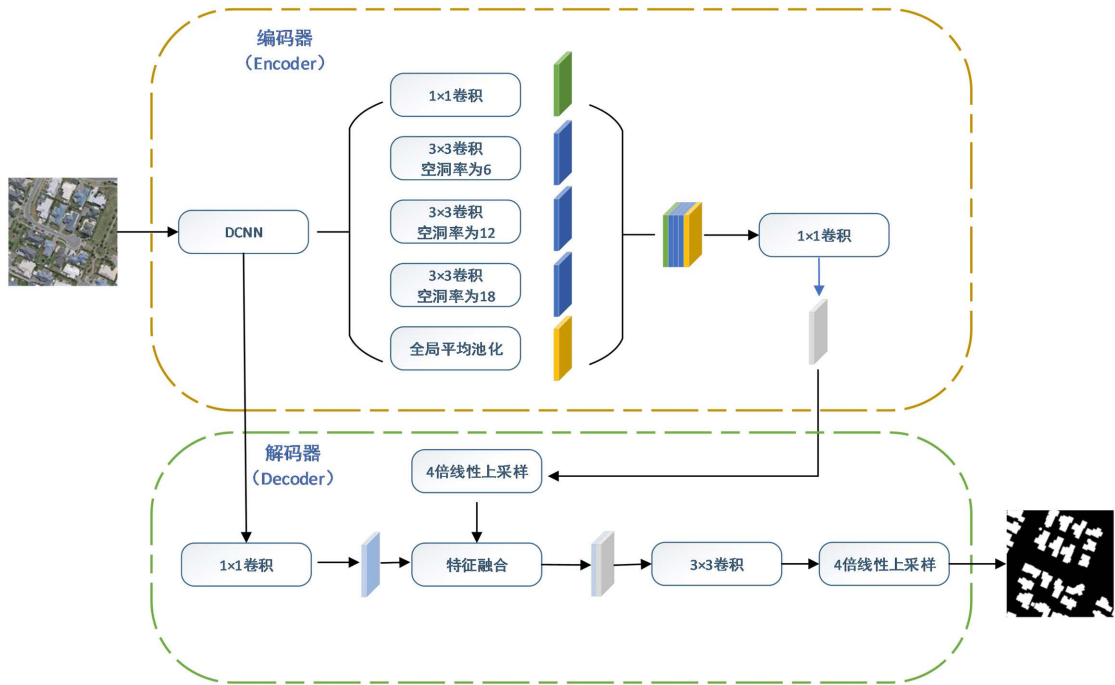


Figure 1. DeepLabv3+ model structure diagram
图 1. DeepLabv3+模型结构图

2.2. 改进的 DeepLabv3+网络模型

2.2.1. 骨干网络

建筑物屋顶多呈面状结构, 分布于高分辨率的遥感图像中, 语义信息非常简单, 但细节信息非常丰富, 这对于语义分割中细节提取能力有很高的要求。传统的特征提取算法主要通过人工手动操作进行, 效率低、准确率不高, 难以满足实际应用需求。原 DeepLabV3+采用骨干网络为 Xception 网络结构, 模型参数量大, 参数数量庞大, 结构复杂, 比较适合于完成多类型语义分割任务, 同时, 造成计算量太大, 不便于培训等等, 显然, 这将非常适合于建筑物屋顶分割的工作。为了解决这个问题, 根据高分辨率遥感影像建筑屋顶特点, 文中把 DeepLabV3+解码器结构的骨干网络设为 ResNet50 进行特征提取。

ResNet50 相比 Xception 网络参数较少, 易于训练, 收敛迅速, 此外在加上空洞卷积后, 感受野也有所增加, 能够较好地融合遥感影像语义信息, 显著提高了运行速度。

2.2.2. 并行加权空洞卷积

DeepLabv3+最初是为了扩大感知范围, 降低运算复杂度而采用空洞卷积, 但是, 这种方法也会导致局部信息的丢失以及长距离信息的不可预测性。通过离散采样, 我们可以在特征图中发现空洞卷积, 它们的相邻像素来自彼此独立的子集, 但是这种卷积方法忽略了连续点间的局部信息, 因此, 在空洞率较

高的情况下, 这种现象尤为明显, 这将严重降低最终的结果准确度。

研究人员经常使用不对称卷积来加速压缩模型, 其中, $1 \times n$ 和 $n \times 1$ 卷积是最常用的, 它们可以有效地分解 $n \times n$ 对称卷积, 从而大大降低模型中的参数数量。如果二维卷积核秩达到 1, 那么就可以通过一系列一维卷积的方式实现对其的等价转换。本文采用并行加权的空洞卷积模块来解决由空洞卷积引起的复杂问题。通过对不同尺寸的图像块采用并行计算的方式, 从而得到一个具有相同阶数的卷积核以达到提升模型精度的目的。通过将 2×2 、 1×3 和 3×1 不对称卷积叠加在一起, 可以取代 3×3 卷积(如图 2 所示), 从而增强卷积核。本模块能够提高模型准确率, 且不会增加模型复杂度。

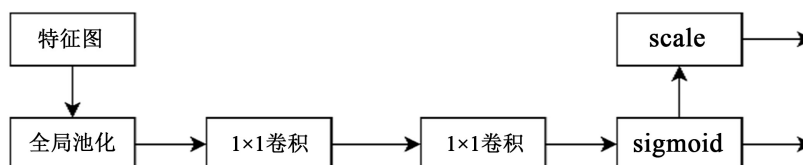


Figure 2. Channel attention structure
图 2. 通道注意力结构

2.2.3. 引入通道注意力机制

DeepLabv3+的注意力机制模块的引入, 得到了刘文祥[10]等学者的积极探索, 他们的成果表明, 将双注意力机制模块融合到 DeepLabv3+中, 可以取得良好的性能。通过位置注意力模块和通道注意力模块, 可以将特征图从一维压缩到更高维度, 从而生成一维的位置权重矩阵, 这样可以更有效地提取出更多的信息。使用双注意力机制能够极大地改善语义分析的精度, 然而, 由于需要计算两种注意力权重矩阵, 因此需要投入大量的计算资源。位置注意力机制的最大目的是将全局信息融合在一起, 但 DeepLabv3+中的空洞金字塔池化模块可以实现这一目标, 从而提高整体的效率和准确性。为了减少网络的复杂性, 本论文仅引入通道注意力机制这个模块。

采用通道注意力机制, 我们能够有效地将主干网络中的数据分析结果转换为多维的、具有多样性的特征图, 从而实现对数据的有效管理。然后, 使用卷积神经网络和 BN 神经网络, 我们能够多个特征向量压缩成一个单独的维度, 其中每个维度都包含了该通道的权重信息, 最后利用 Sigmoid 激活函数, 我们能够大幅提升非线性表达的复杂程度。

2.2.4. 算法流程和网络机构

整个分割算法可分为三个部分, 首先, 对输入的遥感建筑图像实施统一的大小调整, 接下来, 我们将处理过的数据输入 Encoder 模块, 以便从中提取出有用的特征, 并对其进行加权处理, 这样可以使得到的局部和全局特征更加准确。最后通过 Decoder 模块, 我们可以对已获取的特征进行上采样, 以提高图像的分辨率, 并实现语义分割。

通过 ResNet50 网络, 我们可以从图像中提取出多种尺度的特征, 并利用并行加权空洞卷积技术来实现有效的数据处理。接着, 我们将每个尺度的特征赋予不同的权重信息, 并对其进行四倍上采样, 以增加特征的丰富度并减少信息损失。通过对融合的特征进行四倍的增强, 我们可以更准确地预测未来的结果。

3. 实验分析

3.1. 实验数据和环境

通过 WHU 遥感建筑数据集的高分辨率特征, 我们可以有效地评估本文提出的模型的分割性能, 特

别是针对较小的建筑屋顶, 这种特殊的数据集可以更加有效地帮助我们完成分割任务。

首先, 我们对高分辨率遥感影像进行了统一切割, 得到了 512×512 的小分辨率图像。接下来, 我们对这些图像进行了随机旋转和模糊处理, 最终得到了 7252 张遥感建筑物图像。经过 7:3 的比例训练和验证, 我们获得了 5076 张训练集和 1976 张验证集。

在 Ubuntu 操作系统中, 我们使用了 32G 的内存, Nvidia GeForce RTX 2080Ti 显卡, 拥有 12 G 的显存, 并且使用了 PyTorch 1.10 深度学习框架来进行实验。通过将 BatchSize 设置为 4, 并将初始学习率定义为 0.0001, 我们采用了 Adam 算法来不断地更新和迭代参数, 这样就可以有效地控制学习率, 从而加速模型的收敛。

3.2. 评测指标

通过 WHU 遥感建筑数据集的高分辨率特征, 我们可以有效地评估本文提出的模型的分割性能, 特别是针对较小的建筑屋顶, 这种特殊的数据集可以更加有效地帮助我们完成分割任务。

1) 准确率(ACC), 准确率这一指标最为简单, 它只是算出正确分类像素数与全部像素数之比, 计算公式是:

$$ACC = \frac{TP}{FN+TP+TN+FP} \quad (1)$$

这其中 TP 是被正确分割为建筑屋顶的个数, FP 是被错误分割为建筑屋顶的个数。TN 是正确分割为背景的个数, FN 是被错误分割为背景的个数。

2) 召回率(RECALL), 是指在图像中被正确分割的建筑屋顶个数与所有建筑的比值:

$$RECALL = \frac{TP}{FN+TP} \quad (2)$$

3) 精确度(PRECISION), 是指在图像中被正确分割的建筑屋顶个数与所有分割结果为屋顶的比值:

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

4) F1 得分(F1), 是指准确率和精确度的加权平均值:

$$F1 = \frac{2}{\left(\frac{1}{PRECISION}\right) + \left(\frac{1}{RECALL}\right)} \quad (4)$$

5) 交并比(IoU), 指为真实值与预测值两集合之交集与并集之比:

$$IoU = \frac{TP}{FN+TP+FP} \quad (5)$$

4. 结果分析

Table 1. Performance comparison of different models

表 1. 不同模型的性能比较表

模型	ACC	IoU	Precision	Recall	F1
a	0.988	0.904	0.948	0.951	0.950
b	0.949	0.792	0.903	0.916	0.912
c	0.984	0.901	0.946	0.948	0.949
d	0.978	0.892	0.945	0.943	0.947

经过一百次迭代, 我们的模型已经成功地实现了收敛。如下图所示, 训练时 loss 与 IoU 曲线一致, 以 Tensorboard 直观显示, 横坐标为训练迭代数量, 纵坐标是 IoU 与损失值的关系, 当训练迭代次数增加时, 损失值曲线和 IoU 曲线的变化趋势会变得更加平缓, 效果良好。在 WHU 的遥感建筑数据集上 ACC 达到了 0.987, IoU 达到了 0.892。

本文提出的模型与其他算法模型的性能比较如表 1 所示。

其中 a 为 UNet 模型, b 为 Deeplabv3+模型, c 为双注意力机制的 DeepLabv3+模型, d 为本文模型。

本文模型与其他模型在高分辨率遥感图像上的语义分割效果如图 3 所示, 可以看出, 使用本文模型是可以分割出遥感图像中的各个建筑屋顶的, 且根据最终的试验结果可以发现本文改进的 DeepLabV3+模型在建筑屋顶的分割效果是好于原 DeepLabV3+网络的, 也证明了改进后的 DeepLabV3+模型对于高分辨率遥感影像建筑屋顶语义分割问题方面具有一定的先进性。

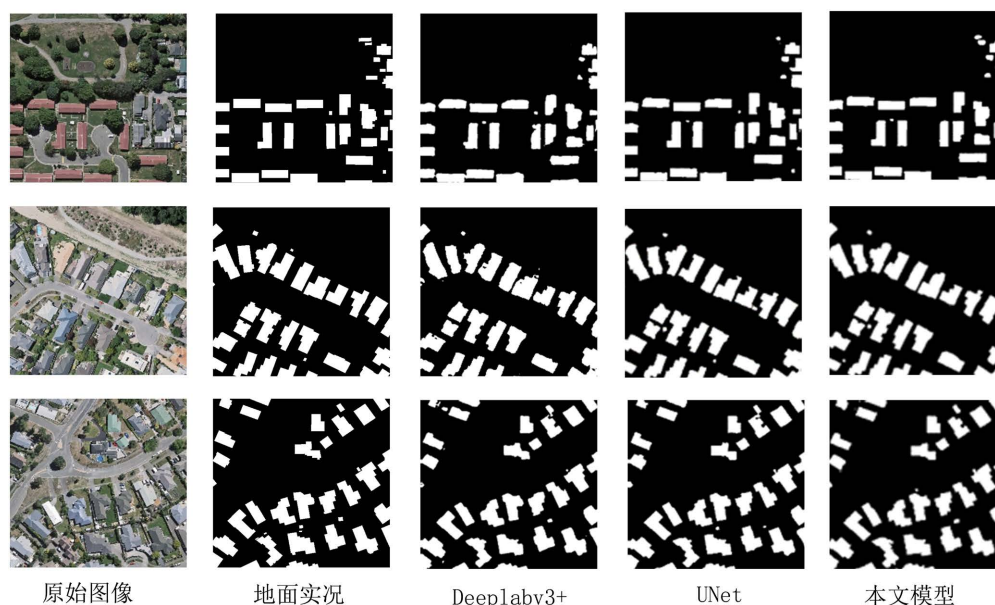


Figure 3. Comparison of segmentation results of various models
图 3. 各模型分割结果比较

5. 结语

本文文提出了一种基于改进 DeepLabv3+网络对高分辨率遥感影像进行屋顶语义分割。尽管 DeepLabv3+网络分割准确, 但是也有一些问题, 如模型的参数数量多, 语义分割效率不够高。鉴于上述挑战, 本文提供了一系列有效的解决方案, 并将它们应用到 DeepLabv3+网络架构上。采用 ResNet50 作为基础架构, 可以显著降低 Xception 模型的参数需求。但是, 这种做法显然会导致一定程度的精确性降低。为了提高精度, 采用了并行加权不对称空洞卷积和通道注意力机制, 以解决分割目标的纠正问题。通过对模型精度、参数量以及其他因素的全面评估, 我们得出了这个结论。尽管模型在各个方面都取得了良好的语义分割效果, 但由于遥感影像中建筑物的尺寸较小, 边界模糊或受到外部障碍物的干扰, 导致边界划分不够准确, 因此未来的研究将重点放在提升边界分割的精确性上。

参考文献

- [1] 韩玲, 杨朝辉, 李良志, 刘志恒, 黄勃学. 利用 Deeplab v3 提取高分辨率遥感影像道路[J]. 遥感信息, 2021, 36(1): 22-28.

- [2] 袁立, 袁吉收, 张德政. 基于 DeepLab-v3+的遥感影像分类[J]. 激光与光电子学进展, 2019, 56(15): 236-243.
- [3] Qiu, C., Schmitt, M., Geiß, C., Chen, T.-H.K. and Zhu, X.X. (2020) A Framework for Large-Scale Mapping of Human Settlement Extent from Sentinel-2 Images via Fully Convolutional Neural Networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, **163**, 152-170. <https://doi.org/10.1016/j.isprsjprs.2020.01.028>
- [4] Xia, L., Zhang, X., Zhang, J., Wu, W. and Gao, X. (2020) Refined Extraction of Buildings With the Semantic Edge-Assisted Approach from Very High-Resolution Remotely Sensed Imagery. *International Journal of Remote Sensing*, **41**, 8352-8365. <https://doi.org/10.1080/01431161.2020.1775322>
- [5] Ahmed, N., Mahbub, R.B. and Rahman, R.M. (2020) Learning to Extract Buildings from Ultra-High-Resolution Drone Images and Noisy Labels. *International Journal of Remote Sensing*, **41**, 8216-8237. <https://doi.org/10.1080/01431161.2020.1763496>
- [6] Hoffmann, E.J., Wang, Y., Werner, M., Kang, J. and Zhu, X.X. (2019) Model Fusion for Building Type Classification from Aerial and Street View Images. *Remote Sensing*, **11**, Article No. 1259. <https://doi.org/10.3390/rs11111259>
- [7] Ji, S., Wei, S. and Lu, M. (2019) Fully Convolutional Networks for Multisource Building Extraction from an Open Aerial and Satellite Imagery Data Set. *IEEE Transactions on Geoscience and Remote Sensing*, **57**, 574-586. <https://doi.org/10.1109/TGRS.2018.2858817>
- [8] 陈丹丹. 基于 Deeplabv3+的高分遥感影像道路损毁信息提取方法研究[D]: [硕士学位论文]. 北京: 中国地震局地震预测研究所, 2020. <https://doi.org/10.27488/d.cnki.ggjfz.2020.000004>
- [9] 林耀辉. 基于 DeepLabv3+的遥感影像语义分割研究[D]: [硕士学位论文]. 福州: 福建师范大学, 2021. <https://doi.org/10.27019/d.cnki.gfjsu.2021.001903>
- [10] 刘文祥, 舒远仲, 唐小敏, 刘金梅. 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. <https://doi.org/10.13284/j.cnki.rddl.003229>
- [11] 元玉梅. 基于深度学习的高分辨率遥感影像建筑区提取方法[D]: [硕士学位论文]. 南京: 南京邮电大学, 2021.
- [12] 高芳, 舒远仲, 朱雯雯. 基于改进 Deeplabv3+的遥感图像语义分割研究[J]. 南昌航空大学学报(自然科学版), 2022, 36(2): 24-31.
- [13] 孙昊堃. 基于深度学习的图像语义分割算法研究[D]: [硕士学位论文]. 贵阳: 贵州大学, 2022.
- [14] 张群. 基于改进的 SegNet 城市遥感图像语义分割算法研究[D]: [硕士学位论文]. 南昌: 南昌大学, 2021. <https://doi.org/10.27232/d.cnki.gnchu.2021.002846>
- [15] 李宇, 肖春姣, 张洪群, 李湘眷, 陈俊. 深度卷积融合条件随机场的遥感图像语义分割[J]. 国土资源遥感, 2020, 32(3): 15-22.
- [16] 赵斐, 张文凯, 闫志远, 于泓峰, 刁文辉. 基于多特征图金字塔融合深度网络的遥感图像语义分割[J]. 电子与信息学报, 2019, 41(10): 2525-2531.