

基于锯齿空洞空间卷积池化结构的高分辨率遥感影像建筑物提取

陈小雨^{1*}, 石彦坡^{2*}, 吴少璇^{3#}, 刘冰雨³, 贾永红³

¹自然资源部地图技术审查中心, 北京

²西安光脉汇嘉科技责任有限公司, 陕西 西安

³武汉大学遥感信息工程学院, 湖北 武汉

收稿日期: 2025年4月6日; 录用日期: 2025年5月8日; 发布日期: 2025年5月14日

摘要

尽管高分辨率遥感影像地物纹理清晰并且光谱异质性高, 但影像中的房屋建筑存在复杂多样性问题, 这导致应用卷积神经网络AASPP模型、DenseASPP模型分别进行提取建筑物时, 较易产生误提取和遗漏提取。因此, 本文提出了一种基于空洞金字塔卷积结构DentateASPP的高分高分辨率遥感影像建筑物提取方法, 采用现有建筑物数据集和重庆高分2号影像进行试验, 与AASPP和DenseASPP模型建筑物提取的结果相比, 实验表明, DentateASPP模型提取建筑物误提取和漏提取最少, 效果最优。

关键词

建筑物提取, 高分辨率遥感影像, 卷积神经网络, 空洞卷积池化结构

High-Resolution Remote Sensing Image Building Extraction Based on Dentate Atrous Spatial Pyramid Pooling Structure

Xiaoyu Chen^{1*}, Yanpo Shi^{2*}, Shaoxuan Wu^{3#}, Bingyu Liu³, Yonghong Jia³

¹Map Technology Review Center of the Ministry of Natural Resources, Beijing

²Xi'an Guangmai Huijia Technology Co., Ltd., Xi'an Shaanxi

³School of Remote Sensing Information Engineering, Wuhan University, Wuhan Hubei

Received: Apr. 6th, 2025; accepted: May 8th, 2025; published: May 14th, 2025

*共同第一作者。

#通讯作者。

文章引用: 陈小雨, 石彦坡, 吴少璇, 刘冰雨, 贾永红. 基于锯齿空洞空间卷积池化结构的高分辨率遥感影像建筑物提取[J]. 计算机科学与应用, 2025, 15(5): 558-563. DOI: 10.12677/csa.2025.155128

Abstract

Despite the clear texture and high spectral heterogeneity of ground objects in high-resolution remote sensing images, there is a complex and diverse problem with the buildings in the images, which leads to errors and omissions when using convolutional neural network AASPP model and DenseASPP model to extract buildings separately. Therefore, a high-resolution remote sensing image building extraction method based on DentateASPP (Dentate Atrous Spatial pyramid pooling) are proposed. The existing building dataset and high-resolution GF-2 images were used for experiments. Compared with the results of building extraction using AASPP and DenseASPP models, the experiments show that the DentateASPP model has the least number of errors and omissions in extracting buildings, and the best performance.

Keywords

Building Extraction, High Resolution Remote Sensing Images, Convolutional Neural Network, Hollow Convolutional Pooling Structure

Copyright © 2025 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

高空分辨率遥感影像建筑物提取是遥感影像信息提取的研究热点之一。传统影像分类方法主要以目视解译、统计模式识别方法为主, 但存在解译精度差、自动化程度不高等缺点, 难以满足日益增长的应用需求。卷积神经网络(Convolutional neural networks, CNN)以其自动提取图像由浅到深层次的特征信息的能力[1]。将特征金字塔网络整合到 U-Net 的网络主干中提出了 SU-Net, 有效提升了 U-Net 对于多尺度信息的泛化能力, 实现对多尺度建筑的高精度提取[2]。在 ResNet 中引入 U-Net 模型的解码块, 提出 Res-U-Net 优化建筑物提取结果[3]; DeepLabV3 以微调后的 ResNet 为基础网络框架, 利用残差块的设计方式, 可有效避免获取深层次的语义信息时网络深度增加造成的网络退化问题[4]; 在基础网络的后端, 连接 ASPP (atrous spatial pyramid pooling)对所得的特征图像进行不同采样率的空洞卷积, 利用并行空洞卷积所捕获多尺度信息和全局平均池化所得的图像级特征融合, 通过上采样实现精确的像素级分类效果[5]。将 Resnet 网络中浅层得到的特征图通过 R-MCN 结构, 进一步提取多尺度和细节信息; 然后采用 DUC (dense upsampling convolution)上采样方式, 恢复特征图的尺寸[6]。应用 FCN 跳跃结构融合不同层次的特征图, 达到提高分类精度的目的[7]。采用增强空洞空间金字塔池化模型 AASPP (Augmented Atrous Spatial Pyramid Pooling, AASPP), 实现水系、植被、建筑物和道路多类别地物的像素分类[8]。虽然 AASPP 多层不同空洞率的卷积层理论上既能学习到大视觉域的环境信息又能学习到细部特征信息, 但各卷积层之间互不相关。虽然卷积神经网络在一定范围内提高了提取建筑物的速度与精度, 但是提取结果仍存在提取结果边界模糊不完整、漏提、误提。

密集空洞空间金字塔池化(Dense Atrous Spatial Pyramid Pooling, DenseASPP)采用捷径连接将各个不同视觉域的空洞卷积层两两连接, 这使不同卷积层的输入既能包含视觉域较大的环境信息, 又包含相对完整的局部信息, 因此更有利于卷积层学习以区别建筑与非建筑。为此提出齿状连接的空洞空间金字塔池化模型 DentateASPP 用于改善建筑物提取效果。

2. 原理与方法

DentateASPP 含有编码结构、解码结构和池化结构。编码结构是残差网络 ResNet_v2_101，解码结构使用 Deeplabv3+，如图 1。其中池化结构位于编码结构和解码结构之间，用来获取不同尺度视觉域特征图，目前多使用空洞卷积层代替池化层来获取。

提出的齿状连接空洞空间金字塔池化模型 DentateASPP，主要有两方面的改进：一是修改空洞卷积池化结构中整体结构，选择更适合建筑物提取的空洞卷积层层数以及在卷积层之间增加捷径连接；另一方面是对卷积层的空洞率的数值的选择和排列方式的设置

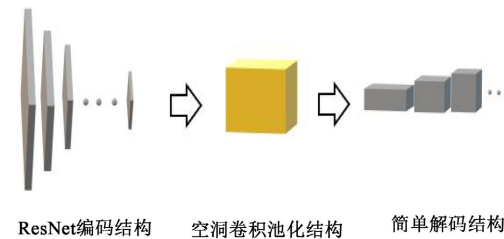


Figure 1. Convolutional neural network structure
图 1. 卷积神经网络结构

2.1. 空洞卷积池化模型结构修改

如图 2，DenseASPP 捷径连接是将某一卷积层输出信息传递某一特定卷积层的方法，在各个层之间添加捷径连接形成密集连接。这样能使卷积层后的特征图与空洞卷积前的特征图都能被不同视觉域的卷积层学习，因此卷积层的输入既包含视觉域较大的背景信息，又包含相对完整的局部信息。

AASPP 每一条路径上有四层空洞卷积层如图 3 所示。针对建筑物提取任务，在空洞率从大到小排列的情况下，一般采用两层空洞卷积池化结构。大尺度空洞卷积池化结构收集图像中大范围信息，小尺度空洞卷积池化结构学习小范围信息。本文采用三层空洞卷积池化结构，在中间增加中尺度空洞卷积池化结构收集中等尺度信息。这样形成的三层空洞卷积池化结构比二层更适合建筑物提取。

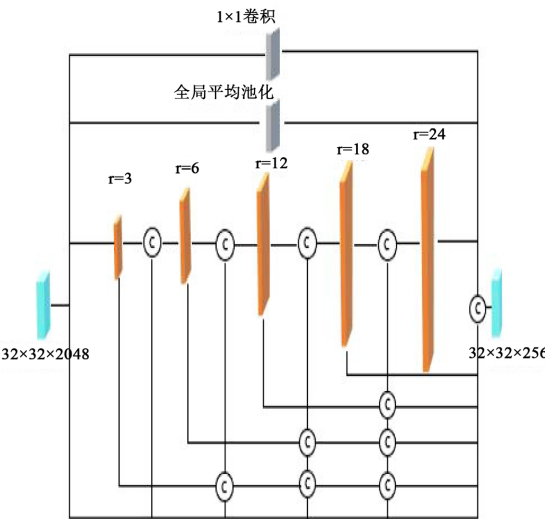


Figure 2. Dense atrous spatial pyramid pooling
图 2. 密集空洞空间金字塔池化

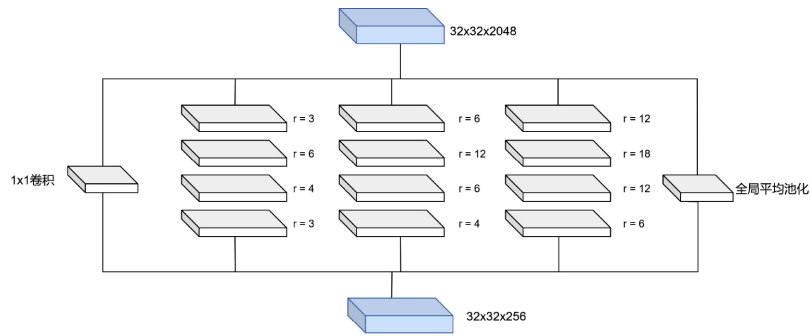


Figure 3. Augmented atrous spatial pyramid pooling
图 3. 增强空洞空间金字塔池化

2.2. 设置空洞卷积池化模型的卷积空洞率

空洞卷积池化模型需考虑空洞率的排列方式和空洞率的设置方式。排列方式是指空洞率不同的卷积层在空洞卷积池化结构中的顺序。空洞率排列一般为空洞率从大到小的方式。而空洞率设置为互质关系，即相互之间没有公约数，形成一种齿状卷积层相互连接的方式。这样互质的空洞率可以提升模型对图像中对象的边缘信息特征的学习能力。

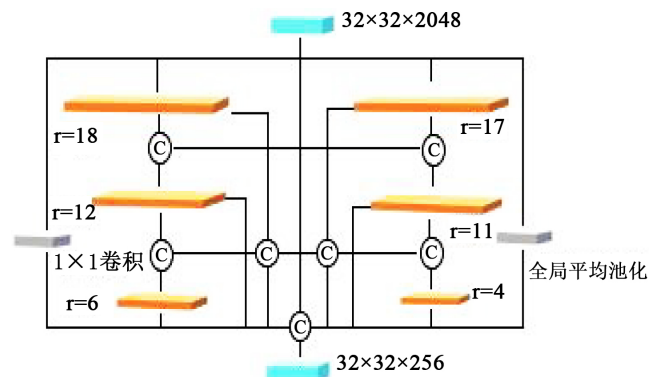


Figure 4. Dentate atrous spatial pyramid pooling
图 4. 齿状连接空洞空间金字塔池化

综合上述，本文提出的齿状连接空洞空间金字塔池化模型结构如图 4 所示。模型中设置了两条相互独立空洞卷积串联合路径，而每条路径上的空洞卷积层数为 3 层，并且各个卷积层之间存在捷径连接。其中一条路径使用了 DenseASPP 中的空洞率，它的空洞卷积空洞率依次为[18,12,6]；另一条路径上则设置对应位置上空洞率大小相近，且同一路径上空洞卷积的空洞率值互为质数，它的空洞率则依次为[17,11,4]。另外，模型中 1×1 卷积层与全局平均池化层是对原空洞空间金字塔池化模型中的继承。这样独立路径的设置方式使模型分别学习三个层次的视觉域下的建筑物及其背景特征，同时交错互质的空洞率增强了模型对复杂建筑物的特征信息的学习能力。

3. 试验与分析

采用 Python 3.6 设计了高分遥感影像提取建筑物的 AASPP、DenseASPP 和 Dentate ASPP 算法。提取建筑物效果评价采用准确率和交并比，如公式(1)和(2)。TP、FP 和 FN 分别为正确提取的建筑像素数、漏提取的建筑像素数和误提取的建筑像素数。

$$\text{prec} = \text{TP}/(\text{TP} + \text{FP}) \tag{1}$$

$$\text{IOU} = \frac{\text{TN}}{\text{FP} + \text{FN} + \text{TN}} \tag{2}$$

为了验证 DentateASPP 提取建筑物的效果，使用 Crowd AI Mapping Challenge 提供的建筑物数据集进行了验证。该数据集划分为训练集、验证集和测试集。其中，训练集包含 280741 张图片，每张图片分辨率为 300×300 像素，包含 RGB 三个通道。对应建筑物标注为 MS-COCO 格式。验证集包含同样格式的图片 60317 张。测试集共 60697 张图片，没有标注数据。

Table 1. Extracting building accuracy using the Crowd AI Mapping Challenge dataset
表 1. 采用 Crowd AI Mapping Challenge 数据集提取建筑物准确度与交并比

模型	准确度	交并比
AASPP	92.85	66.83
DenseASPP	92.71	66.77
DentateASPP	95.19	67.67

表 1 分别是应用 DenseASPP、AASPP 和 DentateASPP 提取数据集建筑物的准确度和交并比，相比 AASPP 与 DenseASPP 模型，DentateASPP 模型提取建筑物精度最高。由此可以得到结论，DentateASPP 模型建筑物提取效果最优。

采用 2019 年重庆地区 GF-2 影像制作样本数据集，2020 年重庆地区 GF-2 影像作为测试数。图 5 给出了 AASPP、DenseASPP 与 DenateASPP 模型在测试区域影像提取建筑的结果。图 5 中第 1 行、第 2 行从左到右分别是学校建筑群、居民建筑群、工业建筑群影像及其标注，第 3 行、第 4 行、第 5 行分别是采用 AASPP、DenseASPP 和 DentateASPP 模型提取建筑物。对比提取结果，可以看到 DentateASPP 模型建筑错误提取和漏提取少。

表 2 分别是 AASPP、DentateASPP 和 DenseASPP 提取 GF-2 影像建筑物的准确度和交并比，相比 AASPP 与 DenseASPP 模型，DentateASPP 模型提取建筑物精度最高。由此可以得到结论，DentateASPP 模型建筑物提取效果最优。

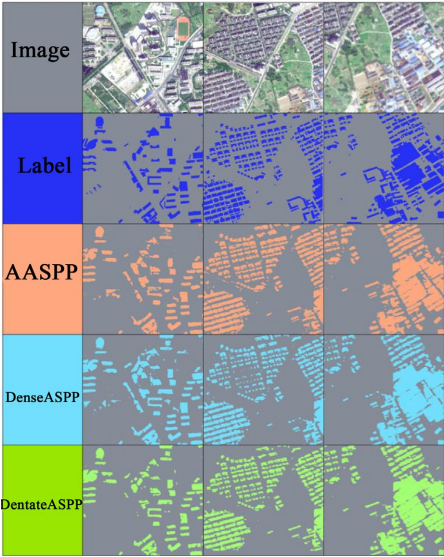


Figure 5. Buildings extracted using AASPP, DenseASPP, and DenateASPP models
图 5. 采用 AASPP、DenseASPP 与 DenateASPP 模型提取的建筑物

Table 2. Accuracy evaluation of extracting buildings from WorldView2 using three models
表 2. 采用三种模型提取 WorldView2 建筑物的精度评价

模型	准确度	交并比度
AASPP	91.2	66.97
DenseASPP	90.2	66.81
DentateASPP	93.1	67.77

4. 结论

通过空洞卷积池化模型结构修改和设置卷积空洞率, 提出了高分遥感影像的齿状连接空洞空间金字塔池化建筑物提取模型 DentateASPP, 相比 AASPP 与 DenseASPP, DentateASPP 模型在现有数据集和 GF-2 上测试, 对建筑物误提取和遗漏提取最少, 准确度最优, 可用于建筑物变化监测。

基金项目

该成果获得国家重点研发计划课题(2022YFC3002702)、武汉大学大学生创新创业训练计划项目资助。

参考文献

- [1] 崔卫红, 熊宝玉, 张丽瑶. 多尺度全卷积神经网络建筑物提取[J]. 测绘学报, 2019, 48(5): 597-608.
- [2] Delassus, R. and Giot, R. (2018) CNNs Fusion for Building Detection in Aerial Images for the Building Detection Challenge. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, 18-22 June 2018, 237-2374. <https://doi.org/10.1109/cvprw.2018.00044>
- [3] 叶沅鑫, 谭鑫, 孙苗苗, 王蒙蒙. 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.
- [4] Zhao, H., Shi, J., Qi, X., Wang, X. and Jia, J. (2017) Pyramid Scene Parsing Network. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 6230-6239. <https://doi.org/10.1109/cvpr.2017.660>
- [5] Yu, B., Yang, L. and Chen, F. (2018) Semantic Segmentation for High Spatial Resolution Remote Sensing Images Based on Convolution Neural Network and Pyramid Pooling Module. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **11**, 3252-3261. <https://doi.org/10.1109/jstars.2018.2860989>
- [6] Chen, L., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L. (2018) Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**, 834-848. <https://doi.org/10.1109/tpami.2017.2699184>
- [7] Chen, L., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H. (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 833-851. https://doi.org/10.1007/978-3-030-01234-2_49
- [8] Chen, G., Li, C., Wei, W., Jing, W., Woźniak, M., Blažauskas, T., et al. (2019) Fully Convolutional Neural Network with Augmented Atrous Spatial Pyramid Pool and Fully Connected Fusion Path for High Resolution Remote Sensing Image Segmentation. *Applied Sciences*, **9**, Article 1816. <https://doi.org/10.3390/app9091816>