引文格式：陈洋,黄康江,袁志明,等. 利用改进的U-Net的建筑物提取方法[J]. 地理空间信息, 2025, 23(10):24-28. doi:10. 3969/j.issn. 1672-4623.2025.10.006

利用改进的U-Net的建筑物提取方法

陈　洋[1]，黄康江[1]，袁志明[2*]，蒙友波[1]，熊　伟[1]
（1. 贵州省自然资源勘测规划研究院，贵州 贵阳 550000；2. 江西应用工程职业学院，江西 萍乡 337000）

摘　要：针对传统语义分割方法在进行建筑物的提取存在提取精度不高、错检、漏检和内部空洞的问题，提出一种融合扩张卷积和注意力机制（ARD-UNet）的遥感影像建筑提取方法。该方法以U-Net为基础网络，首先在编码阶段引入残差学习单元，有助于解决深层网络中的梯度消失问题；其次跳跃阶段引入双重注意力机制模块，有助于模型更加聚焦于建筑物的关键特征信息；最后中间层使用扩张卷积单元扩大感受野，进一步整合建筑物的上下文信息。结果表明，该方法在WHU、Massachusetts建筑数据集上进行验证，交并比指标分别提高了2.31%和1.87%，相较于传统的语义分割方法，ARD-UNet融合了扩张卷积、注意力机制和残差学习的优点，能够更加准确地提取遥感图像中的建筑物。

关键词：建筑物提取；U-Net网络；扩张卷积；注意力机制；残差学习单元
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2025）10-0024-05

Building Extraction Method Based on Improved U-Net

CHEN Yang[1], HUANG Kangjiang[1], YUAN Zhiming[2], MENG Youbo[1], XIONG Wei[1]
(1. Guizhou Institute of Natural Resources Survey and Planning, Guiyang 550000, China;
2. Jiangxi Applied Engineering Vocational institute, Pingxiang 337000, China)

Abstract: To address the problems of poor extraction accuracy, error detection, omission and internal voids in traditional semantic segmentation methods for building extraction, we proposed a remote sensing image building extraction method incorporating dilated convolution and dual attention mechanism (ARD-UNet). Taking the U-Net network as the base network, the method introduced residual learning units in the coding phase to enhance shallow feature layers and intermediate layers of contextual information at first. Then, this method introduced a dual attention mechanism module in the jumping phase to enhance the learning of key feature information of the building. Finally, the method used dilated convolutional units to expand the field of perception in the middle layer, and further integrated contextual information about the building. We validated the method on the WHU, Massachusetts construction dataset. The result shows that the cross-ratios are improved by 2.31% and 1.87% on both datasets respectively. ARD-UNet incorporates the advantages of dilated convolution, residual learning and double attention mechanism, which can extract buildings in remote sensing images more accurately than traditional semantic segmentation methods.

Key words: building extraction, U-Net, dilated convolution, attention mechanism, residual learning unit

随着卫星遥感技术的迅速发展，高分辨率卫星遥感图像的分辨率越来越高，使得遥感影像数据被广泛的应用到相关领域[1-3]。作为城市的重要组成部分，建筑物对城市规划、人口估算、地形图更新、灾害预测等具有重要的意义[4-6]。在遥感领域，自动、精确且高效的从图像中提取建筑物一直是研究的重点和挑战[7-11]。

本文提出一种融合扩张卷积（dilated convolutional unit，DCU）和双重注意力机制（convolution block attention module，CBAM）的遥感影像建筑物提取方法ARD-UNet（attention residual dilated convolution u-Net）。在编码阶段并行残差学习单元（residual learning unit，RLU），有助于解决深层网络中的梯度消失问题；引入注意力机制模块，从空间维度、通道维度进行了权重分配，加强建筑物关键特征信息的学习，实现目标提取精度的提升；在中间层阶段引入3个扩张卷积单元，进一步提高建筑物的提取精度。

1　研究方法

1.1　ARD-UNet网络构建

U-Net网络在建筑物的提取过程中易产生内部空洞、过度分割，难以捕获建筑物中有效的特征信息。因此，本文融合扩张卷积和双重注意力机制用于遥感

第一作者简介：陈洋（1982—），本科，高级工程师，主要从事自然资源遥感数据处理工作，E-mail：36242960@qq.com。
通信作者：袁志明（1994—），硕士，工程师，主要从事遥感图像处理工作，E-mail：277649226@qq.com。

影像建筑物的提取，ARD-UNet网络结构由编码阶段、中间层、解码阶段3部分组成。编码阶段主要接收遥感影像数据，其中引入残差学习单元，并采用最大池化操作降低特征的维度。中间层采用了3个连续扩张卷积，在不降低特征空间分辨率的情况下，扩大特征的感受野，进一步整合遥感图像中的建筑物特征信息。解码阶段一方面采用了双重注意力机制，使模型对建筑物特征的学习更具有针对性，从而忽略无关的背景信息；另一方面采用了3×3反卷积的处理方式，恢复了建筑物的细节特征，最后输出的结果图像与输入遥感影像具有相同的空间分辨率。

1.1.1　残差学习单元

随着卷积层数的增加，经典的U-Net网络在训练过程中容易产生计算速度减慢、精度下降、梯度消失以及过拟合的现象。为有效防止此类问题产生，本文在编码器阶段引入残差学习单元（如图1所示），通残差学习单元实际上是将一个 x 恒等映射加入到普通卷积中，通过批量归一化操作和恒等映射机制，大大简化了网络模型的结构，增强了浅层次的特征层和中间层次的上下文信息。残差学习单元结构计算公式如下：

$$F(x)=H(x)-x \tag{1}$$

式中，$H(x)$ 为目标映射函数；x 为残差学习单元的输入；$F(x)$ 为残差函数，即学习后的残差结果。

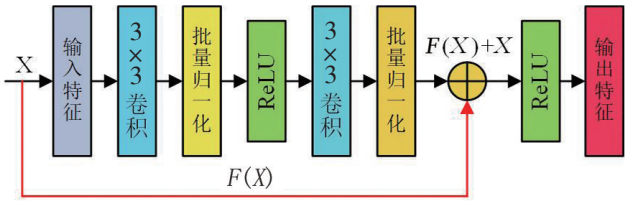


图1　残差学习单元结构

1.1.2　扩张卷积单元

经典的U-Net具备很好的捕捉上下文信息的能力，可以使得特征信息有效的向网络模型更深层次传播。但由于U-Net网络模型池化层的存在，大大降低了遥感影像的分辨率，导致遥感图像语义信息的丢失，直接影响了建筑物提取的准确性。为了在不损失特征空间分辨率的前提下，扩大特征的感受野并整合特征的上下文信息，ARD-UNet网络模型在编码器的第4次降采样后引入了3个扩张卷积，扩张率分别为2、2和3。扩张卷积通过在每个空间维度的连续卷积核中插入 $N-1$ 个零值，实现输入的上采样滤波。扩张卷积核的大小为：

$$K=k+(k-1)(N-1) \tag{2}$$

式中，N 为扩张卷积的扩张率，取值为2、2和3；k 为原始滤波器（卷积核）的大小，通过调整扩张率，可以自适应地扩大感受野，在不增加网络复杂度的前提下，有效地获取遥感图像中不同尺度的局部特征，并更好地保留大部分像素的空间和位置信息。

1.1.3　注意力机制模块

注意力机制的核心思想是聚焦于关键特征的重要信息，忽略无关的背景特征，同时通过反向传播和前向传播来学习获得注意力的权重[12]。在U-Net模型跳跃阶段引入双重注意力机制，如图2所示。注意力机制模块使模型更加关注高分辨率遥感图像中建筑物的特征信息，忽略无关的背景信息，提高网络模型对建筑物的提取精度。注意力机制模块由通道注意力机制、空间注意力机制组成。

通道注意力机制：通过在特征图的空间维度上进行全局池化，将多维的特征图转换成一维的向量，然后利用这个一维向量来调整每个通道的重要性，使模型能够自主学习和强调对当前任务有用的特征通道，同时抑制那些不重要的通道；空间注意力机制：它关注图像中不同区域对于特定任务的贡献程度，并赋予图像中与任务相关的重要区域更高的权重，而忽略那

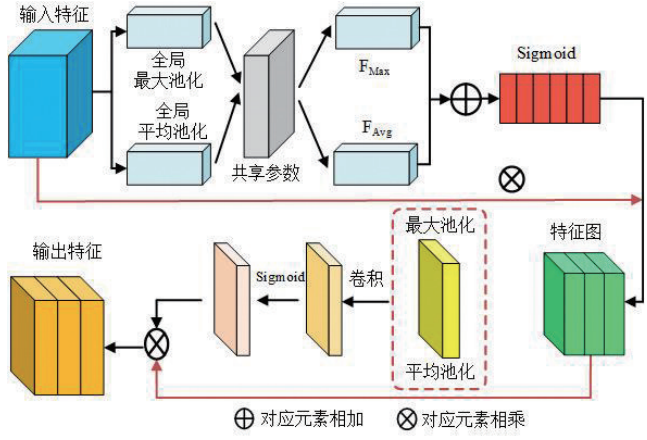


图2　注意力机制结构

些不重要的区域。空间注意力机制的核心在于定位到感兴趣的信息，抑制无用信息。

2 实验结果与分析

2.1 实验数据

Massachusetts建筑物数据集，影像大小为1 500像素×1 500像素，且具有1 m的空间分辨率。为了提高模型训练的效率和灵活性，我们对图像进行预处理，将大尺寸的图像随机裁剪成较小尺寸的子图，获得了4 932幅训练数据集、144幅验证数据集、360幅测试数据集。

WHU建筑物数据集（whu building dataset）[13]，影像的占地面积约为450 km^2，包含22万栋不同的建筑物，具有0.3 m空间分辨率。受GPU的局限性，将图像的尺寸随机裁剪为256像素×256像素的子图，其中包含18 944张训练数据集，4 144张验证数据集，9 664张测试数据集。如图3所示为Massachusetts、WHU建筑物部分影像的示例。



a WHU数据集示例 b Massachusetts数据集示例

图3 WHU与Massachusetts数据集部分示例

2.2 评价指标

本文采用图像语义分割常用的指标来评估模型的分类性能，包括召回率（R）、精确度（P）、总体精度（R_{OA}）、F1-score（F_1）、交并比（R_{IOU}），具体公式如下：

$$P=\frac{TP}{TP+FP} \tag{3}$$

$$R=\frac{TP}{TP+FN} \tag{4}$$

$$F_1=2\times\frac{R\times P}{R+P} \tag{5}$$

$$R_{OA}=\frac{TP+TN}{TP+FP+TN+FN} \tag{6}$$

$$R_{\mathrm{IOU}}=\frac{TP}{FP+TP+FN} \tag{7}$$

式中，TP、FP、TN和FN分别为真阳性、假阳性、真阴性和假阴性。

2.3 评价指标

本文的所有实验都是在Windows 10的64位操作系统上，使用PyTorch 1.3.1深度学习框架进行的。实验配置NVIDIA GeForce RTX 20808 GB显存，并利用CUDA10.2进行加速计算。模型的初始学习率为0.000 1，采用Adam优化器进行优化。迭代次数为50次，每间隔2次保存一次权重参数文件，最终选择最优的权重参数文件对遥感图像中建筑物进行预测。

2.4 实验结果与分析

为了验证本文方法在遥感图像建筑物提取中的有效性。在WHU和Massachusetts建筑物数据集上，将本文的ARD-UNet模型与U-Net、PSPNet和DeepLabV3+进行对比，结果如图如4、5所示。

2.4.1 WHU数据集的结果对比

表1为不同模型在WHU数据集的结果对比。通过对比结果可以发现，ARD-UNet模型在总体精度、精确度、召回率、交并比以及F1-score比其他3种模型有所提高。U-Net模型精确度为78.42%，PSPNet模型精确度为82.47%，DeepLabV3+模型的精确度为82.46%，精确度略低于ARD-UNet模型。U-Net模型提取建筑物的总体精度在3个对比模型里最低。PSPNet和U-Net模型有着较好的分割总体精度、精确度，但召回率、交并比以及F1-score相对较低，不能满足精确提取建筑物的要求。反观ARD-UNet模型，相较于U-Net模型在OA、P、R、IOU、F1-score（F_1）分别提升了2.39、2.41、3.28、2.31和2.41个百分点。

表1 WHU数据集在不同模型对比结果/%

模型	OA	P	R	$F1$	IOU
PSPNet	92.87	82.47	76.54	87.36	78.85
U-Net	91.25	82.38	76.57	86.59	78.63
DeepLabV3+	92.87	82.46	77.75	87.36	79.34
ARD-UNet	93.64	84.79	79.85	89.00	80.94

为了更加直观地对比4种模型在遥感图像建筑物提取效果，分别选取了4幅不同细节的建筑物图像进行了对比，结果如图4所示。从图4可以看出，PSPNet模型的提取建筑物有明显的错误，存在较多的建筑物漏检现象，并且较多的建筑物黏连的现象；DeepLabV3+模型提取建筑物的效果稍微优于PSPNet模型，但在建筑物的边缘分割较为粗糙，部分较小建筑物分割较差；U-Net模型提取建筑物也产生了较为明显的错误，对建筑物边缘划分还不够精确，建筑物内部也出现了漏检、错检的现象；而ARD-UNet模型对遥感图像中建筑物的提取结果，仍存在部分分类错误的现象，但对于较大的建筑物，其他3种对比方法提取建筑的结果边缘模糊、存在内部空洞的现象，

而ARD-UNet模型的边缘提取效果更好；对于较小的建筑物，其他3种对比方法提取建筑的结果出现漏检的现象，并且对建筑物的边缘提取也较模糊，而ARD-UNet模型明显改善了漏检、错检的问题，同时能够较好的保留建筑物的细节特征。

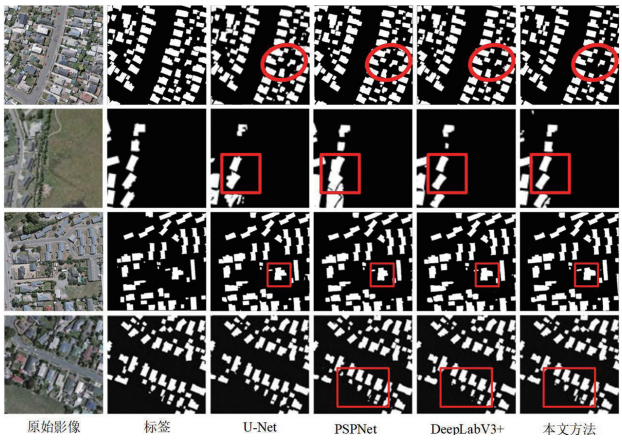


图4　WHU数据集在不同模型的对比结果

2.4.2　Massachusetts数据集的结果对比

为了进一步验证ARD-UNet模型的有效性与可行性，本文在Massachusetts建筑数据集上进行了U-Net、PSPNet和DeepLabV3+模型对比实验，结果如图5所示。从图5可以看出，ARD-UNet模型提取建筑的结果明显优于PSPNet、U-Net、DeepLabV3+模型，尤其在小型建筑物边缘细节方面。PSPNet模型提取建筑物的结果中有着明显的空洞问题，且存在错检、漏检的现象。U-Net和DeepLabV3+建筑物边缘效果较好，但还是存在部分建筑物漏检的问题。相比之下，ARD-UNet模型充分利用了扩张卷积、双重注意力机制以及残差单元的优点，因此，提取建筑物的结果更加接近真实值，尤其是图5中椭圆和矩形标出来的建筑物无论是大型建筑物还是小型建筑物，ARD-UNet模型具有准确的边缘信息。

表2为对比模型在Massachusetts数据集的结果，可以看出，对比模型中PSPNet的召回率明显低于U-Net、DeepLabV3+、ARD-UNet模型，仅为75.33%，这是因为PSPNet提取大型建筑物存在着明显错分的问题。DeepLabV3+模型的各个评价指标仅次于ARD-UNet模型。ARD-UNet的总体精度为94.21%其中较PSPNet、U-Net、DeepLabV3+分别提高了3.18%，1.89%，0.65%；精确度为84.15%其中较PSPNet、U-Net、DeepLabV3+分别提高了0.92%，1.23%，0.08%，交并比为80.83%，较PSPNet、U-Net、DeepLabV3+分别提高了4.46%，1.87%，0.86%。通过这些指标表明，ARD-UNet在总体精度、召回率、精确度、F1-score、交并比均有一定的提升。

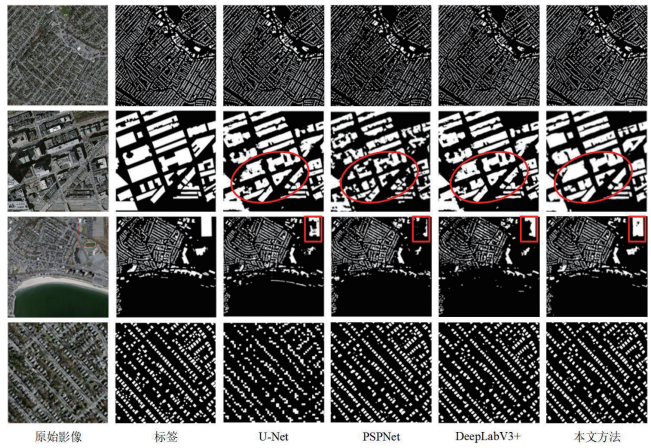


图5　Massachusetts数据集不同模型的对比结果

表2　Massachusetts数据集在不同模型对比结果/%

模型	*OA*	*P*	*R*	*F*1	IOU
PSPNet	91.03	83.23	75.33	86.96	76.37
U-Net	92.32	82.98	77.85	87.40	78.96
DeepLabV3+	93.56	84.07	78.36	88.56	79.97
ARD-UNet	94.21	84.15	79.98	88.90	80.83

2.4.3　消融实验

为了验证ARD-UNet模型中残差学习单元、扩张卷积单元以及双重注意力机制模块的有效性。本文在WHU建筑物数据集上对各个模块在原始的U-Net模型上进行了消融实验。具体精度对比如表3所示。

从表3消融实验结果可以，相较于经典的U-Net模型，加入了残差学习单元IOU提高了0.3%，加入了扩张卷积单元*OA*、IOU分别提高了0.11%和1.26%，引入注意力机制*OA*、IOU分别提高了0.45%和1.36%；加入残差学习单元、扩张卷积单元组合后*OA*、IOU分别提高了0.15%和1.48%；加入残差学习单元、双重注意力机制组合后IOU提高了2.15%；加入扩张卷积单元和注意力机制后*OA*、IOU分别提高了0.66%和2.04%。而同时添加3个模块*OA*、IOU分别提高了0.77%和2.31%。这说明残差学习单元、扩张卷积单元以及双重注意力机制模块的组合在ARD-UNet发挥了积极性的作用，有助于模型提高建筑物提取的精度。

表3　消融实验结果

U-Net	RLU	DCU	CBAM	*OA*/%	IOU/%
√				92.87	78.63
√	√			92.80	78.93
√		√		92.98	79.89
√			√	93.32	79.99
√	√	√		93.02	80.11
√	√		√	92.88	80.78
√		√	√	93.53	80.67
√	√	√	√	93.64	80.94

3 结 语

针对遥感影像建筑物提取精度低、内部产生空洞、错检、漏检等问题，提出一种适用于遥感图像建筑物提取的ARD-UNet模型，在编码阶段保护了信息的完整性；在跳跃阶段使学习更具有针对性。将ARD-UNet、U-Net、PSPNet、DeepLabV3+在WHU、Massachusetts数据集上进行测试，实验结果表明，ARD-UNet模型在2个数据集上都具有较好的检测性能。

参考文献

[1] 申传庆,王凯,王文杰. 基于ResNet-UNet的地表覆盖自动分类技术研究[J]. 地理空间信息,2023,21(6):21-23

[2] 徐佳伟,刘伟,单浩宇,等. 基于PRCUnet的高分遥感影像建筑物提取[J]. 地球信息科学学报,2021,23(10):1 838-1 849

[3] 古煜民,阎福礼. 基于不同骨架UNet++网络的建筑物提取[J]. 中国科学院大学学报,2022,39(4):512-523

[4] 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):2 245-2 254

[5] 罗松强,李浩,陈仁喜. 多尺度特征增强的ResUNet+遥感影像建筑物提取[J]. 激光与光电子学进展,2022,59(8):540-549

[6] 周志诚,李静,史晓明,等. 基于遥感影像的土地利用动态巡查监管应用方法探讨[J]. 地理空间信息,2023,21(3):41-43

[7] 杨栋杰,高贤君,冉树浩,等. 基于多重多尺度融合注意力网络的建筑物提取[J]. 浙江大学学报(工学版),2022,56(10):1 924-1 934

[8] 卢彻,徐胜华,朱军. 改进U-Net的高分影像建筑物提取方法[J]. 测绘科学,2021,46(12):140-146

[9] 许正森,管海燕,彭代锋,等. 高分辨率遥感影像建筑物提取的注意力胶囊网络算法[J]. 遥感学报,2022,26(8):1 636-1 649

[10] 于明洋,陈肖娴,张文焯,等. 融合网格注意力阀门和特征金字塔结构的高分辨率遥感影像建筑物提取[J]. 地球信息科学学报,2022,24(9):785-802

[11] Ding L, Tang H, Liu Y, et al. Adversarial Shape Learning for Building Extraction in VHR Remote Sensing Images[J]. IEEE Transactions on Image Processing,2021,31(10):678-690

[12] Woo S, Park J, Lee J, et al. Cbam:Convolutional Block Attention Module[C]//Proceedings of the European Conference on Computer Vision (ECCV),2018

[13] Ji S, Wei S, 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

（上接第23页）

分类技术研究[J]. 地理空间信息,2023,21(6):21-23

[4] 贺浩,王仕成,杨东方,等. 基于Encoder-Decoder网络的遥感影像道路提取方法[J]. 测绘学报,2019,48(3):330-338

[5] 张永宏,何静,阚希,等. 遥感图像道路提取方法综述[J]. 计算机工程与应用,2018,54(13):1-10

[6] 陈涛,艾廷华. 多边形骨架线与形心自动搜寻算法研究[J]. 武汉大学学报(信息科学版),2004,29(5):443-446

[7] 杜瑞颖,刘镜年. 面状地物名称注记的自动配置研究[J]. 测绘学报,1999,42(4):365-368

[8] 艾廷华,郭仁忠. 基于约束Delaunay结构的街道中轴线提取及网络模型建立[J]. 测绘学报,2000,43(4):348-354

[9] 杨雨雪,王红艳,张玲玲,等. 基于细化和最小生成树的多边形主骨架线提取[J]. 上海师范大学学报(自然科学版),2022,51(2):204-209

[10] 阎逸飞. 基于缓冲区的道路中心线提取方法与应用[J]. 地理空间信息,2023,21(11):36-38

[11] 杨伟,艾廷华. 基于众源轨迹数据的道路中心线提取[J]. 地理与地理信息科学,2016,32(3):1-7

[12] 李功权,蔡祥云. 一种基于约束三角网的道路中心线的提取方法[J]. 长江大学学报(自科版),2013,10(4):47-50

[13] 杜维,梁年,张超. 遥感影像提取数据的后处理技术研究[J]. 地理信息世界,2021,28(3):113-118

[14] 余杰,吕品,郑昌文. Delaunay三角网构建方法比较研究[J]. 中国图象图形学报,2010,15(8):1 158-1 167

[15] Thomson R C, Brooks R. Efficient Generalisation and Abstraction of Network Data Using Perceptual Grouping[C]//Proceedings of the 5th International Conference on GeoComputation,2000:23-25

[16] 刘晶. GIS制图中骨架提取算法的设计与实现[J]. 中南民族大学学报(自然科学版),2007,32(4):78-80