

新型语义分割D-UNet的建筑物提取

龙丽红¹, 朱宇霆², 闫敬文¹, 刘敬瑾¹, 王宗跃³

1. 汕头大学 工学院电子系, 汕头 515063;

2. 中山大学 电子与信息工程学院, 广州 510006;

3. 集美大学 计算机工程学院, 厦门 361021

摘要: 高分辨率遥感图像语义分割在航空图像分析领域具有重要的理论价值和应用价值。但由于高分辨率遥感图像中建筑物语义的丰富性和图像背景的复杂性, 以往的分割方法往往容易产生边缘模糊、细节信息丢失和分辨率低等缺点。为了解决高分辨率卫星图像语义分割边界模糊和信息丢失的问题, 本文提出一种端到端的卷积神经网络 Dilated-UNet (D-UNet)。首先, 通过改进U-Net网络结构, 采用Dilation技术拓展四通道的多尺度空洞卷积模块, 每个通道采用不同的卷积扩张率来识别多尺度语义信息, 从而提取更丰富的细节信息。其次, 设计了一种交叉熵和Dice系数的联合损失函数, 更好的训练模型以达到预期分割效果。最后, 在Inria航空图像数据集上进行综合评估与检验。实验结果表明, 本文提出的遥感图像分割方法能够有效地从高分辨率遥感图像中进行像素级城市建筑物的分割, 与其他方法相比, 分割精度更高, 具有较高的实际应用价值。

关键词: 遥感图像, 语义分割, 多尺度, 空洞卷积, 图像处理

中图分类号: P2

引用格式: 龙丽红, 朱宇霆, 闫敬文, 刘敬瑾, 王宗跃. 2023. 新型语义分割D-UNet的建筑物提取. 遥感学报, 27(11): 2593-2602

Long L H, Zhu Y T, Yan J W, Liu J J and Wang Z Y. 2023. New building extraction method based on semantic segmentation. National Remote Sensing Bulletin, 27(11): 2593-2602 [DOI: 10.11834/jrs.20211029]

1 引言

随着地球观测遥感卫星技术发展取得了卓越的成就, 大量的遥感数据为各种应用提供丰富的可挖掘的信息(孙伟伟等, 2020)。建筑物信息作为地理信息的重要组成部分, 广泛应用于道路规划、土地规划、城市管理等领域, 在生活中有着越来越重要的应用, 本文主要研究遥感图像的楼房分割问题。

在传统分割方法中, 通常根据灰度、颜色、纹理和形状等特征将图像划分成若干互不相交的区域。典型的分割方法有基于阈值、边缘、区域、图论等分割方法。基于阈值的分割方法, 其基本思想是基于图像灰度特征计算灰度阈值, 并将图像的每个像素灰度值与阈值相比较, 得出其类别。如李丽等(2013)根据建筑群的分布和纹理特点, 利用小波变换和自适应全局阈值法提取建筑群标

记信息实现分割; 吴诗嫻等(2018)提出了一种基于直线截距直方图的多阈值分割方法等。基于边缘的分割方法(Hu等, 2013; Wang等, 2015), 主要根据图像边缘灰度、颜色、纹理等特征的突变来进行边缘检测, 比如使用基于Roberts(康牧等, 2008)、Sobel、Prewitt(邹柏贤等, 2013)、Laplacian(桂预风和吴建平, 2011)等微分算子对图像进行边缘检测, 识别出图像的边缘信息, 从而完成分割。基于图论的分割方法, 其基本思想是将图像的分割问题与图的最小分割问题相关联, 最终实现分割效果。如Felzenszwalb和Huttenlocher(2004)介绍一种基于图表示的图像分割方法, 基于贪心聚类算法提出可变型部件模型算法, 奠定了基于图论的分割算法。但由于遥感图像包含丰富的光谱信息, 使用传统的特征提取方法对于要求高的遥感图像分割应用场景, 仍然存在很大的局限性。

收稿日期: 2021-01-22; 预印本: 2021-06-10

第一作者简介: 龙丽红, 研究方向为图像目标检测、图像分割的理论和应用。E-mail: 19lhlong@stu.edu.cn

通信作者简介: 闫敬文, 研究方向为遥感图像分析与识别、深度学习、小波分析。E-mail: jwyan@stu.edu.cn

近年来,深度学习理论及应用取得重要进展,并在图像语义分割中取得良好的效果,特别是对高级语义信息提取,解决了传统图像分割方法中语义信息缺失的问题。2014年,全卷积神经网络FCN (Fully Convolutional Networks) 首次将深度学习应用于图像分割,实现像素级分割,奠定了深度学习神经网络用来解决图像分割问题的基础(Shelhamer等, 2017)。然而FCN采用双线性插值上采样恢复特征图,丢失许多细节信息,导致分割结果比较模糊和平滑。如图1所示,在航空遥感图像数据集INRIA (Inria Aerial Image Labeling Dataset) 上使用全卷积神经网络对遥感影像进行分割,其中图1(a)为原图像,图1(b)为图像的真实标签,图1(c)是使用全卷积神经网络的分割效果图。可以看出,与真实标签相比(如红框所示),测试结果边界出现明显模糊,小尺寸的目标未能精确识别,整体分割效果较为粗糙。2015年,U-Net (Ronneberger等, 2015) 作为FCN的改进与发展,通过捕获上下文信息的收缩路径来实现更精准的像素边界定位。采用网络结构完全对称的典型编码解码结构,但由于其网络结构仅在单一尺度上预测,不能很好地处理多尺度问

题,且速度较慢,冗余较多。然而,遥感影像具有语义丰富,分辨率低,清晰度低等特点,且建筑物图像具有多尺度、大跨度等特点,且上下文联系紧密,与自然图像相比具有极其复杂的特性,简单地使用全卷积神经网络或U-Net不能很好地满足分割的应用需求。为了解决这个问题,Bischke等(2019)采用了一种新型的多任务损失函数,他们的方法可以很好的保存建筑物的分割边缘并且提高了分割准确率,但缺点是网络复杂且不易优化。类似的,Li等(2019)在网络模型中收集多尺度特征信息来提高遥感影像分割的精度,韩彬彬等(2020)在残差卷积网络中加入了带孔卷积和稠密连接机制以构建稠密空间金字塔结构来提取多尺度特征。而Pan等(2019)采用对抗神经网络并且在其中引入注意力机制来提升分割的效果,他们的方法在INRIA数据集上取得了很好的结果。与上述方法不同,本文提出了一种基于U-Net的新模型Dilated-UNet (D-UNet),即在U-Net中间加入精心设计的空洞卷积模块来解决这个问题,并设计了一种交叉熵和Dice-coefficient的联合损失函数,更好的训练模型以达到预期分割效果。

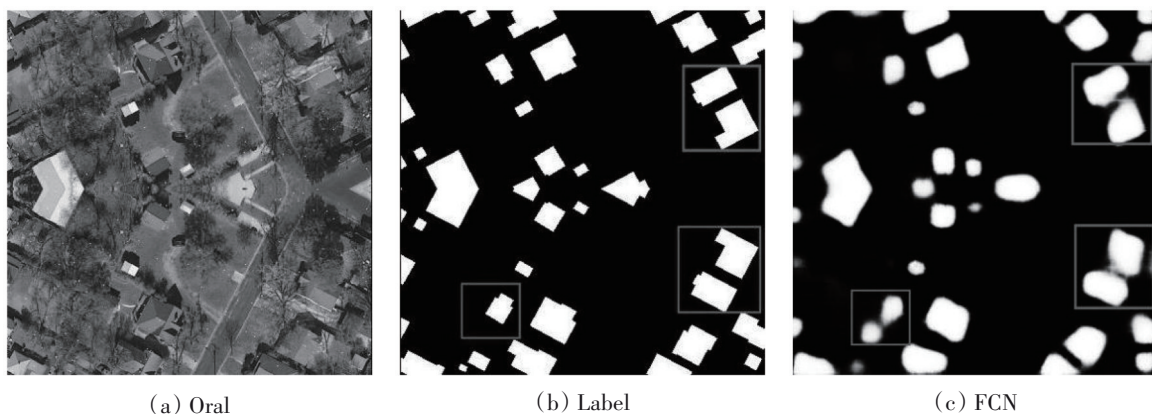


图1 FCN分割效果对比图

Fig. 1 Comparison of FCN segmentation effect

2 基于U-Net改进的卷积网络模型

本文提出一种基于U-Net改进的全新卷积神经网络分割模型D-UNet (Dilated-UNet),接下来将详细介绍D-UNet网络结构,主要包括U-Net、空洞卷积模块、联合损失函数和实现细节等,具体网络结构如图2所示。

2.1 模型结构

(1) U-Net。U-Net,首次是应用于医学图像分割,其采用典型的编码—解码结构。在编码结构上,每经过两个卷积核为 3×3 的卷积层后进入 2×2 的最大池化层,便得到新尺度的特征图,共有4个卷积池化模块。在解码结构上,每进行一次 2×2 的上采样,与上一次卷积操作得到的特征图

在裁剪后进行多尺度的融合, 丰富细节信息, 提高分割精度, 共有4个上采样模块。在编码和解码的结构上具有完全对称的特点。但U-Net网络卷积

层次较少, 对特征的提取与表达不够准确, 边缘模糊现象没有得到很好解决。

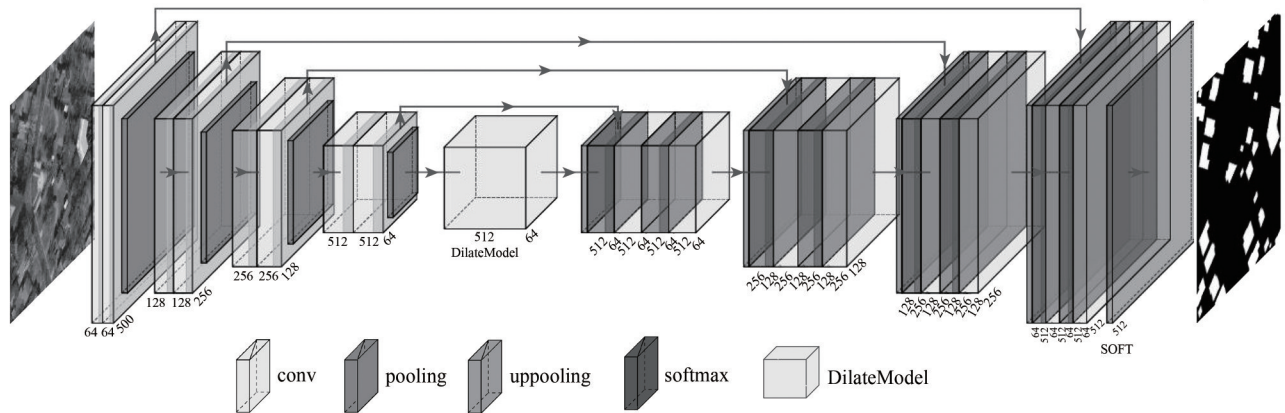


图2 D-UNet网络结构图

Fig. 2 D-UNet network structure diagram

(2) 空洞卷积模块。图像输入到CNN中, FCN或U-Net和传统CNN一样对图像进行卷积、池化等操作, 以降低图像尺寸, 增大感受野。图像分割是像素级输出, 因此要进行上采样恢复到原始图像尺寸大小, 但过程降低了空间分辨率并严重丢失图像的细节信息。空洞卷积极好地解决了这个问题(Yu和Koltun, 2016)。空洞卷积可类似看成是在卷积核内部插入(扩张率)个0来扩大卷积核大小, 从而在扩大感受野的同时捕获多尺度的上下文信息。

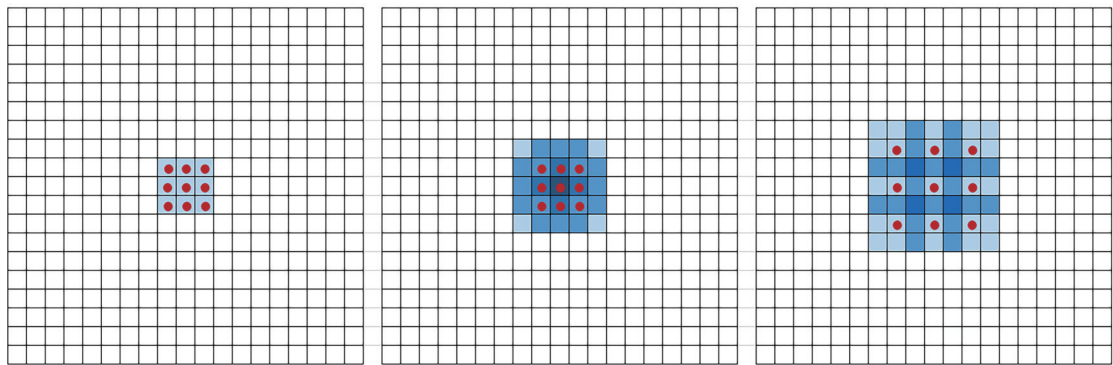
空洞卷积可扩大感受野, 例如3个卷积核大小为 3×3 的普通卷积核叠加, 3层的感受野大小分别是 3×3 , 5×5 , 7×7 。而3个大小为 3×3 的卷积核叠加, 扩张率为1时, 其感受野为 3×3 ; 扩张率为2时, 其感受野为 7×7 ; 扩张率为5时, 其感受野为 17×17 。显然, 上述卷积核参数始终为9, 在保持参数个数不变的情况下, 使用空洞卷积可以大幅度增大卷积核的感受野, 感受野示意图如图3所示。其中红色点表示感受野的中心, 蓝色的深浅表示参与卷积运算的次数, 越深表示参与卷积运算的次数越多。

本文设计的空洞卷积模块如图4所示, 共4条数据信息流通通道, 使用的卷积核大小均为 3×3 。第一通道使用4个 3×3 的卷积核, 且空洞卷积的扩张率分别为1、2、5、8, 其感受野分别 3×3 、 7×7 、 17×17 、 33×33 。随后每条通道相对应的空洞卷积扩张率不变, 逐渐减少卷积层层数, 直

到扩张率为1的卷积核, 即普通卷积。即提取到的特征进入四条通道, 分别生成不同的信息流, 然后进行信息连接。参考空间金字塔模型(Zhao等, 2017)和ResNet的残差结构(He等, 2016), 本文设计的空洞卷积的扩张率多样化, 可提取多尺度信息, 在扩大感受野的同时成功避免空洞卷积理论问题, 即“网格效应”(Wang等, 2018), 且扩张率越大, 能提取到更多边缘信息细节, 提高模型效果。通过多通道进行并行计算, 将经过不同空洞卷积扩张率的特征图进行相加融合, 克服单通道的单一性不足, 且不同通道对不同大小的目标分辨率不同, 从而提高整体的分辨效果。

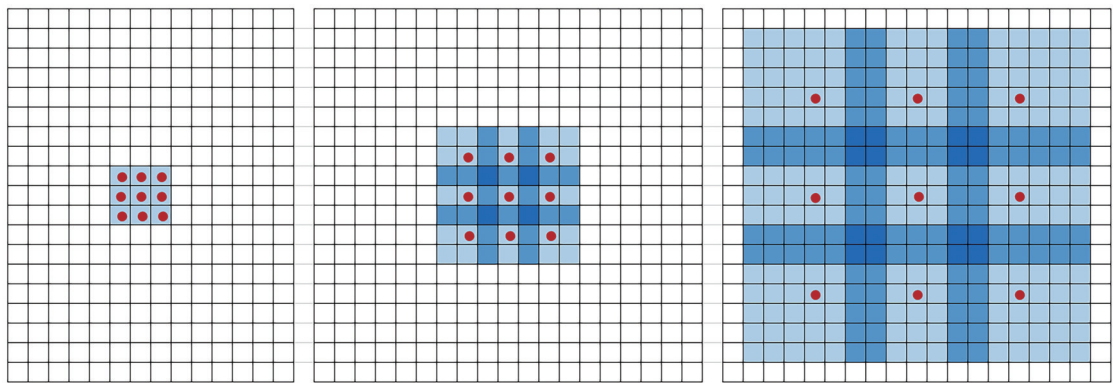
其中, $16 \times 16 \times 512$ 表示特征图的长和宽是均16, 通道数是512; 采用的卷积核均是 3×3 , 箭头上的数字表示所对应的空洞卷积扩张率。

(3) D-UNet。本文提出的D-UNet模型采用典型的编码—解码结构。编码结构由4个卷积模块, 共8个卷积层, 4个最大池化层组成, 每层卷积前都进行归一化操作, 使用ReLU作为激活函数, 每两层卷积后接一个最大池化层。下采样后进入空洞卷积模块, 增大感受野, 提取多尺度信息, 进一步提高边缘分辨率。解码结构包含4个上采样块, 每个上采样块包含转置卷积和上池化操作, 使特征图恢复到与输入图像相同的尺寸, 保留原始图像的空间信息, 最后使用softmax函数进行逐像素分类, 最终实现分割效果。



(a) 3个卷积核的扩张率分别为(1,1,1)卷积核感受野(即普通卷积)

(a) The receptive field of the three convolution kernels whose expansion rates are (1,1,1) (Ordinary convolution)



(b) 3个卷积核的扩张率分别为(1,2,5)的卷积核感受野

(b) The receptive field of the three convolution kernels whose expansion rates are (1,2,5)

● 感受野中心 □ □ □ 可表示卷积运算次数

图3 3个3×3卷积核感受野对比图

Fig. 3 Comparison of three 3×3 convolution kernel receptive fields

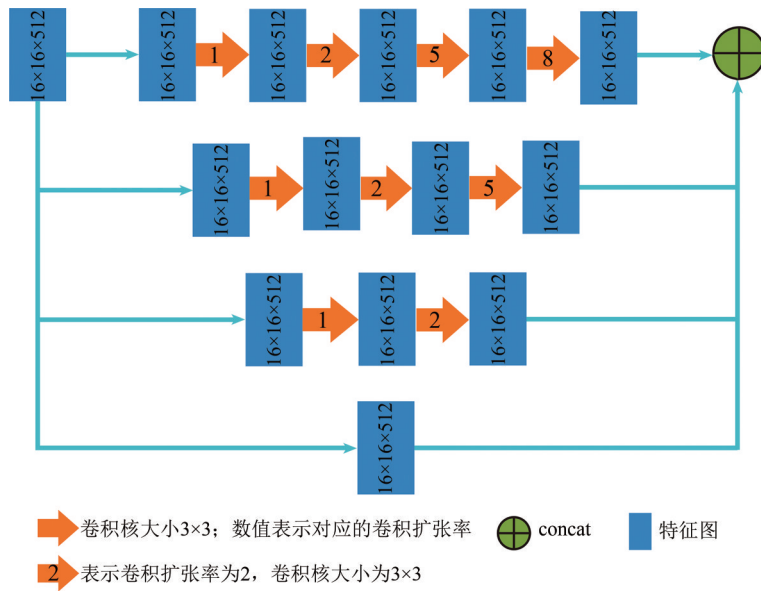


图4 空洞卷积模块图

Fig. 4 Dilated convolution network module

2.2 联合损失函数

本文设计的损失函数采用联合策略，将交叉

熵和Dice-coefficient的进行联合，联合损失函数表达式为

$$L = \lambda L_{\text{BCE}} + (1 - \lambda) L_{\text{Dice}} \quad (1)$$

式中, L_{BCE} 为二值交叉熵损失函数 (Binary Cross Entropy Loss), L_{Dice} 为 Dice 系数损失函数, L 为联合损失函数, λ 为权重调节参数。

假设训练数据 $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$, 其中 $x \in R^n$ 为训练样本, y 为标签, 且 $y \in \{0, 1\}$ 。二值交叉熵损失函数表达式如下:

$$L_{\text{BCE}} = -\sum_{i=1}^n y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \quad (2)$$

式中, i 为第 i 个样本; n 为总样本数; y_i 为第 i 个样本的真实标签; \hat{y}_i 为第 i 个样本的预测标签。然而, 二值交叉熵损失函数衡量标准只是对正确分类的预测概率, 具有单一性, 因此本论文采用联合损失函数。

Dice 系数损失函数, 可度量集合相似度, 可用于计算两个样本的相似度 (取值范围为 $[0, 1]$)。然而, 当样本极度不均的情况下, 二值交叉熵损失函数对样本不平衡的问题很敏感, 训练时会偏向于样本多的一方, 导致训练效果下降, 但 Dice 系数损失函数表现效果较好。假设 X 和 Y 为两个样本, Dice 系数为 (3), Dice 损失函数可表示为 (4), 可微形式可表示为 (5), 其中 y 为真实标签, \hat{y} 为预测标签, 则梯度为 (6)。当样本差异越大时, s 越小, 则梯度越大, 训练越稳定。

$$s = \frac{2|X \cap Y|}{|X| + |Y|} \quad (3)$$

$$D = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (4)$$

$$D = 1 - \frac{2y\hat{y}}{y + \hat{y}} \quad (5)$$

$$\frac{dD}{dy} = \frac{2y^2}{(y + \hat{y})^2} \quad (6)$$

因此, 考虑到样本之间的相似程度, 本文采用 Dice 系数损失函数和交叉熵损失函数进行联合, Dice 系数损失函数表达式为

$$L_{\text{Dice}} = 1 - \frac{2 \times \sum_{i=1}^n y_i \hat{y}_i + \varepsilon}{\sum_{i=1}^n y_i + \sum_{i=1}^n \hat{y}_i + \varepsilon} \quad (7)$$

式中, i 为第 i 个样本; n 为总样本数; y_i 为第 i 个样本的真实标签; \hat{y}_i 为第 i 个样本的预测标签; ε 为调节参数。

权重调节参数 λ , 初始化 $\lambda=1$, 即 $L = L_{\text{BCE}}$, 只有交叉熵损失函数。实验表明, 当 $\lambda = 0.7$ 时, 即

$L = 0.7L_{\text{BCE}} + 0.3L_{\text{Dice}}$, 分割效果最好。

2.3 网络训练细节

实验平台采用 Intel-i7-7700 四核心八线程 CPU 处理器、使用 Pytorch 1.5.1 版本的深度学习框架, 并使用 NVIDIA 公司的 CUDA10.1 GPU 平台进行计算加速。实验统计, 本文模型训练一个 epoch 的时间为 1.1 个小时, 总共训练了 50 个 epoch, 共用时 55 个小时。

(1) 预处理。由于 GPU 内存的限制, 大尺度的遥感图像不能直接作为网络结构的输入, 因此从遥感图像中随机裁剪出图片大小为 512×512 的图像, 并将图像进行归一化处理, 使图像的像素值在 0 到 1 之间, 更加便于计算, 以提高计算速度。

(2) 训练。由于预训练在深卷积神经网络中起着重要作用 (Hinton 等, 2006), 因此本实验使用 VGG13 在 ImageNet 上的预训练模型, 然后以端到端的方式对整个网络进行微调。具体训练过程如下, 训练中使用三通道遥感影像作为输入和输出的分割结果, 以联合损失函数 (1) 作为整体模型的损失函数, 在 GPU 上进行 300 次迭代, 学习率设为 0.001, 每 50 次迭代学习率缩小 10 倍, 权重衰减设为 0.0001, 动量设为 0.9, 使用 Adam 算法优化所有网络参数。

3 实验过程及结果分析

3.1 数据集

本文使用的数据集是 IAIL 航空图像数据集 (Inria Aerial Image Labeling Dataset) (Maggiori 等, 2017)。IAIL 航空图像数据集是一个城市建筑物检测的遥感图像数据集, 包括高度密集的大都市金融区和高山度假村的各种城市景观, 标记只有建筑和非建筑两种。数据集由 360 张彩色 (3 波段 RGB) 正射影像组成, 包括美国和奥地利的城市住区, 其空间分辨率为 0.3 m^2 , 训练面积为 337.5 km^2 , 验证面积为 67.5 km^2 , 测试面积为 405 km^2 , 总覆盖面积 810 km^2 (每幅影像大小为 50002)。选择该数据集, 可使用交叉验证方法, 随机设置 5 个城市作为训练集, 5 个城市作为测试集, 避免训练容易产生过拟合现象, 有效验证模型的可行性。其次, 该数据集是官方公开的数据集, 测试集没有标签,

因此将模型统一提交到官方网站 (<https://project.inria.fr/aerialimagelabeling/leaderboard/>[2021-01-22]) 上进行验证。

3.2 验证指标

为了定量评估网络模型效果, 本文采用像素精度 Acc (pixel accuracy) 和交并比 IoU (Intersection over Union) 这两个度量指标。像素精度是标记正确的像素占总像素的比例, 交并比是真实值和预测值集合的交集与并集之比, 表达式如下所示。

$$Acc = \frac{\sum_{i=0}^k p_{ii}}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}} \tag{8}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{9}$$

式中, 假设有 $k+1$ 个类, p_{ii} 表示本属于 i 类且被预测为 i 类的像素数量, 即被正确预测的像素数量, p_{ij} 表示本属于 i 类却被预测为 j 类的像素数量。TP 为真正值 (True Positive), 即判定为正样本, 实际上是正样本, FP 为假正值 (False Positive), 即判定为负样本, 实际上是正样本, FN 为假负值 (False Negative), 即判定为负样本, 实际上是负

样本。

3.3 主要结果

为了验证本文方法的有效性, 我们将 D-UNet 与航空图像标记数据集上的典型模型和先进方法进行比较, 表 1 为在测试集上的数值评估结果, 有 SegNet, VGG11 (Simonyan 和 Zisserman, 2014), PSPNet (Zhao 等, 2017), LinkNet (Chaurasia 和 Culurciello, 2017), 还包括 2018 年 IAIL 竞赛的获胜者 AMML (Huang 等, 2018), 以及 U-Net 和 ED-Net 的融合模型 (余威和龙慧云, 2019) 和 AMUNet (Guo 等, 2020)。由表 1 可知, 本文方法在测试集的 5 个城市实验中普遍比经典方法好, 交并比 IoU 和准确性 Acc 都有很大的提高。比 2019 年提出的 U-Net 和 ED-Net 融合模型在 IoU 和准确性上分别提高了 4.74% 和 0.64%。同时, 比惠健等 (2019) 在 IoU 和 Acc 分别提高了 8.08%、1.13%, 比 AMUNet 的 IoU 提高了 7.13%, Acc 提高了 0.64%。由表 1 可知 D-UNet 在测试集上有两个地区的结果不如其他方法, 因为每个地区的建筑物具有不同的分布特点, 每个方法在不同的地区中效果较难达到完全的最优, 但 D-UNet 在整体 IoU 和准确性上均超过了其他方法。

表 1 测试集数值评估

Table 1 Test set numerical evaluation

方法\数据集	/%											
	BellingHam		Bloomington		Innsbruck		San Francisco		EastTyrol		Overall	
	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc
SegNet	70.26	97.03	64.67	96.58	72.33	96.65	69.36	89.55	76.70	97.90	70.29	95.54
U-Net	70.14	96.91	69.31	97.06	74.98	96.99	74.55	91.52	78.71	98.09	73.83	96.12
PSPNet	70.26	97.03	66.49	96.83	75.48	97.10	71.17	89.86	78.47	98.08	72.00	95.78
LinkNet	69.59	96.97	66.51	96.82	75.25	97.07	73.27	91.10	78.39	98.08	72.83	96.01
AMLL	67.14	96.64	65.43	96.73	72.27	96.66	75.72	91.80	75.67	97.70	72.55	95.91
AMUNet	64.46	96.33	54.59	95.69	68.75	96.20	70.78	90.29	71.38	97.39	67.69	95.18
U-Net & ED-Net	71.59	97.14	61.03	96.33	70.39	96.53	70.41	90.21	76.97	97.94	70.08	95.63
D-UNet(ours)	72.13	97.20	68.51	97.02	77.10	97.31	75.23	91.59	80.01	98.23	74.82	96.27

注: 加粗数值表示该地区下最优方法的实验结果。

在验证集上对比了近一两年在 IAIL 航空图像数据集上提出的新分割方法, 如表 2 所示。本文方法 D-UNet 比 Li 等 (2019) 提出多尺度的 U-Net (Multi-scale UNet) 的 IoU 和 Acc 分别提高了 4.74% 和 0.65%; 比 Pan 等 (2019) 使用空间和通道双注

意力机制的生成式对抗网络的 IoU 和 Acc 分别提高了 1.46%、0.17%; 比 Han 等 (2020) 提出的稠密连接机制改进的残差卷积网络和空间金字塔结构的 RDASP-Net 模型 IoU 提高了 3.05%, Acc 提高了 0.51%。Sebastian 等 (2020) 提出的 SEResNeXt101-

FPN-CPA 方法，即加入了上下文金字塔模型，通过捕获上下文特征依赖项来改进对不同大小的建筑物进行提取，本文方法 D-UNet 的 IoU 比其高了 1.69%，Acc 高了 0.29%。另外，D-UNet 比 Guo 等 (2020) 提出的加入注意力机制块的改进的 U-Net 方法 AMUNet 的 IoU 提高了 2.02%，Acc 提高了 0.04%。表 2 充分说明了本方法在验证集上的有效性，D-UNet 在整体 IoU 和准确性上优于其他最新方法。

表 2 验证集数值评估
Table 2 Validation set numerical evaluation

方法\数据集	/%											
	Austin		Chicago		KITSAP Co		West Tyool		Vienna		Overall	
	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc
Multi-scale UNet	73.09	96.43	70.38	92.92	72.45	99.43	76.40	98.12	78.88	93.98	74.24	96.12
GAN-SCA	81.01	97.24	71.37	93.32	68.67	99.31	78.68	98.33	81.62	94.80	77.52	96.60
AMUNet	84.43	97.29	81.22	96.45	54.13	93.83	79.97	98.83	85.05	97.28	76.96	96.73
RDASP-Net	78.52	96.85	69.87	92.71	66.73	99.27	80.28	94.33	76.46	98.13	75.93	96.26
SEResNeXt101-FPN-CPA	80.15	97.18	69.54	92.78	70.36	99.32	80.83	98.46	81.43	94.67	77.29	96.48
D-UNet(ours)	82.05	95.02	78.96	98.55	77.76	98.05	72.76	93.45	79.76	98.66	78.98	96.77

注：加粗数值表示该地区下最优方法的实验结果。

实验可视化图如图 5 所示。图 5 (a) 为数据集原图，图 5 (b) 为原图标签，D-UNet 的结果如图 5 (g) 所示，SegNet、VGG11、PSPNet 和 LinkNet 分割的视觉效果分别如图 5 (c)、5 (d)、5 (e)、5 (f) 所示。可以看到，与其他方法的分割结果相比，本文方法 D-UNet 可以更好区分建筑物之间的边界，预测结果较少有边缘信息的丢失，能捕获更多“尖锐”的细节信息。

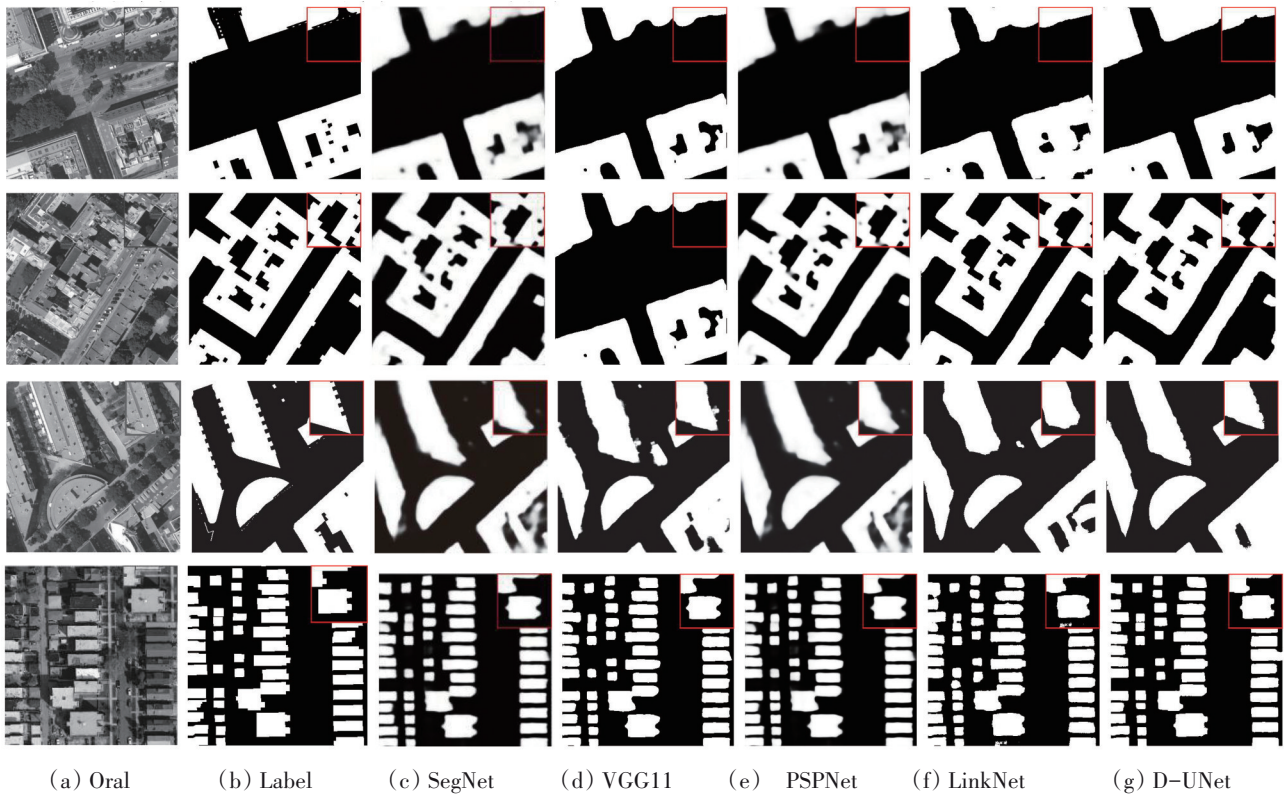


图 5 各方法实验可视化图
Fig. 5 Experimental visualization of each method

图6展示了D-UNet模型在图像尺寸大小为5000×5000的INRIA遥感影像测试集上的分割效果，D-UNet在测试时每次的输入尺寸为1024×1024，最后将其拼接为原图大小。

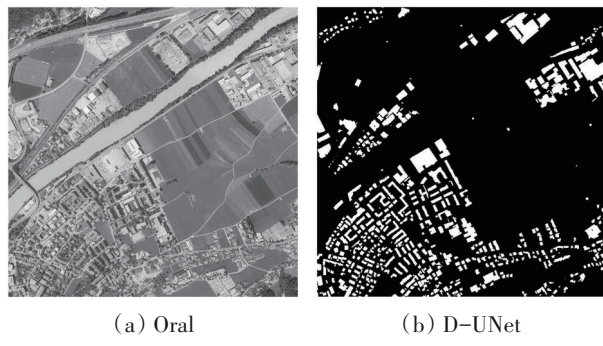


图6 D-UNet分割效果可视化图
Fig. 6 Visualization of D-UNet segmentation effect

3.4 消融实验

为进一步验证本文方法的有效性，在验证集上对网络结构进行有无空洞卷积模块的消融对比实验结果如表3所示。在消融实验中，以本文提出的联合损失函数作为模型的损失函数，对比四通道的空洞卷积模块对模型的整体效果，实验表明，有空洞卷积模块的D-UNet在5个城市的交并比IoU和精确度Acc均有提升，其中，IoU最大提升幅度为5.41%，Acc最大提升幅度为2.18%，模型整体提高分别为：IoU：4.61%，Acc：0.74%，即四通道的空洞卷积模块能够提取大小不同的多尺度特征信息对模型的分割效果具有较好的提升作用。

表3 验证集消融实验评估

Table 3 Validation set ablation experimental evaluation

												%
方法\数据集	Austin		Chicago		KITSAP Co		West Tyool		Vienna		Overall	
	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc
D-UNet(No dilated conv)	78.34	94.32	73.55	96.37	72.35	98.32	68.33	93.02	75.53	97.53	74.37	96.03
D-UNet(dilated conv)	82.05	95.02	78.96	98.55	77.76	98.05	72.76	93.45	79.76	98.66	78.98	96.77

4 结 论

从遥感图像中准确自动地分割建筑物对于城市规划 and 灾害管理等应用领域至关重要。本文提出了一种新的建筑物提取方法，称为D-UNet。考虑到现有的基于全卷积神经网络的方法有很多局限性，如产生模糊边缘和细节信息丢失等问题。D-UNet通过以下3个优势来解决上述问题：（1）D-UNet是端到端结构的像素级分割网络。（2）D-UNet通过融合不同尺度的空洞卷积模块，在不增大模型计算量的同时提高了分割的精确度。（3）提出了一种新的联合损失函数，使模型能更快更稳定的进行参数更新。

在IAIL航空图像数据集上的实验结果充分验证了D-UNet在高分辨率遥感图像语义分割的有效性和优势，其分割精度更高，优于其他方法，具有较高的实际应用价值。但是，D-UNet仍然具有可以提升的空间。我们将在之后的研究中继续探究如何降低D-UNet的训练时长以及如何继续提升分割精度。

参考文献(References)

Bischke B, Helber P, Folz J, Borth D and Dengel A. 2019. Multi-task learning for segmentation of building footprints with deep neural networks//Proceedings of 2019 IEEE International Conference on Image Processing. Taipei, China: IEEE: 1480-1484 [DOI: 10.1109/ICIP.2019.8803050]

Chaurasia A and Culurciello E. 2017. LinkNet: exploiting encoder representations for efficient semantic segmentation//Proceedings of 2017 IEEE Visual Communications and Image Processing. St. Petersburg: IEEE [DOI: 10.1109/vcip.2017.8305148]

Felzenszwalb P F and Huttenlocher D P. 2004. Efficient graph-based image segmentation. International Journal of Computer Vision, 59 (2): 167-181 [DOI: 10.1023/B:VISI.0000022288.19776.77]

Gui Y F and Wu J P. 2011. An image edge detection method base on Laplacian operator and degree of gray relation. Journal of Shantou University (Natural Science), 26(2): 69-73 (桂预凤, 吴建平. 2011. 基于Laplacian算子和灰色关联度的图像边缘检测方法. 汕头大学学报(自然科学版), 26(2): 69-73) [DOI: 10.3969/j.issn.1001-4217.2011.02.011]

Guo M Q, Liu H, Xu Y Y and Huang Y. 2020. Building extraction based on U-net with an attention block and multiple losses. Remote Sensing, 12 (9): 1400 [DOI: 10.3390/rs12091400]

Han B B, Zhang Y T, Pan Z X, Tai X Q and Li F F. 2020. Building ex-

- traction in remote sensing image based on residual dense pyramid network. *Foreign Electronic Measurement Technology*, 39(5): 5-10 (韩彬彬, 张月婷, 潘宗序, 台宪青, 李芳芳. 2020. 基于稠密金字塔网络的遥感影像建筑物提取. *国外电子测量技术*, 39(5): 5-10) [DOI: 10.19652/j.cnki.femt.1901954]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE [DOI: 10.1109/CVPR.2016.90]
- Hinton G E, Osindero S and Teh Y W. 2006. A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7): 1527-1554 [DOI: 10.1162/neco.2006.18.7.1527]
- Hu X Y, Shen J J, Shan J and Pan L. 2013. Local edge distributions for detection of salient structure textures and objects. *IEEE Geoscience and Remote Sensing Letters*, 10(3): 466-470 [DOI: 10.1109/LGRS.2012.2210188]
- Huang B H, Lu K K, Audebert N, Khalel A, Tarabalka Y, Malof J, Boulch A, Le Saux B, Collins L, Bradbury K, Lefèvre S and El-Saban M. 2018. Large-scale semantic classification: outcome of the first year of Inria aerial image labeling benchmark//*Proceedings of 2018 IEEE International Geoscience and Remote Sensing Symposium*. Valencia: IEEE [DOI: 10.1109/IGARSS.2018.8518525]
- Hui J, Qin Q M, Xu W and Sui J. 2019. Instance segmentation of buildings from high-resolution remote sensing images with multitask learning. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 55(6): 1067-1077 (惠健, 秦其明, 许伟, 隋娟. 2019. 基于多任务学习的高分辨率遥感影像建筑实例分割. *北京大学学报(自然科学版)*, 55(6): 1067-1077) [DOI: 10.13209/j.0479-8023.2019.106]
- Kang M, Xu Q G and Wang B S. 2008. A roberts' adaptive edge detection method. *Journal of Xi'an Jiaotong University*, 42(10): 1240-1244 (康牧, 许庆功, 王宝树. 2008. 一种 Roberts 自适应边缘检测方法. *西安交通大学学报*, 42(10): 1240-1244) [DOI: 10.7652/xjtub200810010]
- Li L, Chai W T and Mei S L. 2013. Segmentation of remote sensing images based on adaptive global threshold and fused markers. *Transactions of the Chinese Society for Agricultural Machinery*, 44(7): 222-228 (李丽, 柴文婷, 梅树立. 2013. 基于自适应全局阈值融合标记的遥感图像建筑群分割. *农业机械学报*, 44(7): 222-228) [DOI: 10.6041/j.issn.1000-1298.2013.07.039]
- Li X, Jiang Y C, Peng H and Yin S. 2019. An aerial image segmentation approach based on enhanced multi-scale convolutional neural network//*Proceedings of 2019 IEEE International Conference on Industrial Cyber Physical Systems*. Taipei, China: IEEE: 47-52 [DOI: 10.1109/ICPHYS.2019.8780187]
- Maggiori E, Tarabalka Y, Charpiat G and Alliez P. 2017. Can semantic labeling methods generalize to any city? The inria aerial image labeling benchmark//*Proceedings of 2017 IEEE International Geoscience and Remote Sensing Symposium*. Fort Worth: IEEE [DOI: 10.1109/IGARSS.2017.8127684]
- Pan X R, Yang F, Gao L R, Chen Z C, Zhang B, Fan H R and Ren J C. 2019. Building extraction from high-resolution aerial imagery using a generative adversarial network with spatial and channel attention mechanisms. *Remote Sensing*, 11(8): 917 [DOI: 10.3390/rs11080917]
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//*Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich: Springer [DOI: 10.1007/978-3-319-24574-4_28]
- Sebastian C, Imbriaco R, Bondarev E and de With P H N. 2020. Contextual pyramid attention network for building segmentation in aerial imagery. *arXiv*: 2004.07018
- Shelhamer E, Long J and Darrell T. 2017. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4): 640-651 [DOI: 10.1109/TPAMI.2016.2572683]
- Simonyan K and Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv*: 1409.1556
- Sun W W, Yang G, Chen C, Chang M H, Huang K, Meng X Z and Liu L Y. 2020. Development status and literature analysis of China's earth observation remote sensing satellites. *Journal of Remote Sensing*, 24(5): 479-510 (孙伟伟, 杨刚, 陈超, 常明会, 黄可, 孟祥珍, 刘良云. 2020. 中国地球观测遥感卫星发展现状及文献分析. *遥感学报*, 24(5): 479-510) [DOI: 10.11834/jrs.20209464]
- Wang J, Yang X C, Qin X B, Ye X and Qin Q M. 2015. An efficient approach for automatic rectangular building extraction from very high resolution optical satellite imagery. *IEEE Geoscience and Remote Sensing Letters*, 12(3): 487-491 [DOI: 10.1109/LGRS.2014.2347332]
- Wang P Q, Chen P F, Yuan Y, Liu D, Huang Z H, Hou X D and Cottrell G. 2018. Understanding convolution for semantic segmentation//*Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision*. Lake Tahoe: IEEE [DOI: 10.1109/WACV.2018.00163]
- Wei Y and Long H Y. 2019. A building segmentation method based on deep convolution networks for remote sensing imagery. *Computer Technology and Development*, 29(6): 57-61 (余威, 龙慧云. 2019. 基于深度卷积网络的遥感影像建筑物分割方法. *计算机技术与发展*, 29(6): 57-61) [DOI: 10.3969/j.issn.1673-629X.2019.06.012]
- Wu S H, Wu Y Q and Zhou J J. 2018. Multi-level thresholding for remote sensing image of urban area based on line intercept histogram. *CAAI Transactions on Intelligent Systems*, 13(2): 227-235 (吴诗嫻, 吴一全, 周建江. 2018. 直线截距直方图城区遥感图像多阈值分割. *智能系统学报*, 13(2): 227-235) [DOI: 10.11992/tis.201609012]
- Yu F and Koltun V. 2016. Multi-scale context aggregation by dilated convolutions. *arXiv*: 1511.07122
- Zhao H S, Shi J P, Qi X J, Wang X G and Jia J Y. 2017. Pyramid scene parsing network//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE: 2881-2890 [DOI: 10.1109/CVPR.2017.660]
- Zou B X, Zhang R and Miao J. 2013. Improved Prewitt method for im-

age edge detection. *Microelectronics and Computer*, 30(5): 23-26
(邹柏贤, 张然, 苗军. 2013. Prewitt图像边缘检测方法的改进.)

微电子学与计算机, 30(5): 23-26 [DOI: 10.19304/j.cnki.
issn1000-7180.2013.05.005]

New building extraction method based on semantic segmentation

LONG Lihong¹, ZHU Yuting², YAN Jingwen¹, LIU Jingjin¹, WANG Zongyue³

1. *Department of Electronic Engineering, Shantou University, Shantou 515063, China;*

2. *School of Electronic and Information Technology, Sun Yat-Sen University, Guangzhou 510006, China;*

3. *School of Computer Engineering, Jimei University, Xiamen 361021, China*

Abstract: Semantic segmentation of high-resolution remote sensing image has important theoretical and practical value in the field of aerial image analysis. However, the traditional segmentation methods are prone to edge blur, loss of detail information, and low resolution due to the richness of building semantics and the complexity of image background in high-resolution remote sensing images.

An end-to-end convolutional neural network called Dilated-UNet (D-UNet) is proposed to solve the problem of fuzzy boundary and information loss in high-resolution satellite image semantic segmentation. First, the U-Net network structure is improved and the multiscale dilated convolution module of four channels is expanded using the division technology. Each channel uses different convolution expansion rates to identify the multiscale semantic information for extracting richer detailed information. Second, a joint loss function of cross entropy and Dice coefficient is designed to achieve the desired segmentation effect.

The model is comprehensively evaluated and tested on the Inria aerial image dataset. Experimental results show that the proposed remote sensing image segmentation method can effectively segment urban buildings at pixel level from high-resolution remote sensing images, and the segmentation accuracy is higher and is therefore better than those of other methods.

Our proposed D-UNet can deliver automatic building segmentation from high-resolution remote sensing images with high accuracy. Thus, it is a useful tool for practical application scenarios.

Key words: remote sensing images, semantic segmentation, multiscale, dilated convolution, image processing