计算机科学
COMPUTER SCIENCE

基于多尺度特征融合的遥感图像建筑物提取算法研究

陈国军, 岳雪燕, 朱燕宁, 付云鹏

引用本文

陈国军, 岳雪燕, 朱燕宁, 付云鹏. 基于多尺度特征融合的遥感图像建筑物提取算法研究[J]. 计算机科学, 2023, 50(9): 202-209.

CHEN Guojun, YUE Xueyan, ZHU Yanning, FU Yunpeng. Study on Building Extraction Algorithm of Remote Sensing Image Based on Multi-scale Feature Fusion [J]. Computer Science, 2023, 50(9): 202-209.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220800086

基于多尺度特征融合的遥感图像建筑物提取算法研究

陈国军　岳雪燕　朱燕宁　付云鹏
中国石油大学(华东)计算机科学与技术学院　山东 青岛 266580
(88326309@qq.com)

摘　要　由于高分辨率遥感图像中的建筑物尺寸多样,且背景复杂,因此在对遥感图像中的建筑物进行提取时,往往存在细节丢失、边缘模糊等问题,从而影响模型的分割精度。为了解决这些问题,提出了具有空间和语义信息的双分支架构网络 B2Net。首先,在语义信息分支上建立交叉特征融合模块,充分捕获上下文信息,以聚合更多的多尺度语义特征;其次,在空间信息分支上将空洞卷积和深度可分离卷积进行组合,提取图像的多尺度空间特征,并通过优化膨胀率扩大网络的感受野;最后,构建内容感知注意力模块,对图像中的高频和低频内容进行自适应选择,以达到细化建筑物分割边缘的效果。在两个建筑物数据集上对 B2Net 进行训练与测试。在 WHU 数据集上,与基线模型相比,B2Net 在精度、召回率、F1 分数以及交并比上皆达到了最佳效果,分别为 98.60%,99.40%,99.30%,88.50%;在 Massachusetts 建筑物数据集上,4 个指标比 BiSeNet 分别提高了 0.9%,1.9%,1.7%,2.2%。实验结果证明,B2Net 可以更好地捕获空间细节信息和高级语义信息,提高了复杂背景下的建筑物进行分割精度,满足了对建筑物快速提取的需求。

关键词:建筑物提取;特征融合;空洞卷积;深度可分离卷积;内容感知注意力
中图法分类号　TP751

Study on Building Extraction Algorithm of Remote Sensing Image Based on Multi-scale Feature Fusion

CHEN Guojun, YUE Xueyan, ZHU Yanning and FU Yunpeng
College of Computer Science and Technology, China University of Petroleum(East China), Qingdao, Shandong 266580, China

Abstract　Because of the various size of buildings and complicated background in high-resolution remote sensing images, there are some problems such as loss of details and blurring of edges when extracting buildings in remote sensing images, which affect the segmentation accuracy of the model. In order to solve these problems, this paper proposes a two-branch architecture network B2Net with spatial and semantic information branches. Firstly, the cross feature fusion module is provided in the semantic information branch to fully capture the context information to aggregate more multi-scale semantic features. Secondly, in the spatial branch, we combine the atrous convolution and depthwise separable convolution to extract the multi-scale spatial features of the image, and optimize the dilated rate to expand the receptive field. Finally, we use the content aware attention module to adaptively select the high-frequency and low-frequency content in the image to achieve the effect of refining the edges of building segmentation. We train and test the B2Net on two building datasets. On the WHU dataset, compared with the baseline model, the B2Net achieves the best result in precision, recall, F1 score and IoU, which is 98.60%, 99.40%, 99.30%, and 88.50%, respectively. On the Massachusetts building dataset, the four indicators are 0.9%, 1.9%, 1.7% and 2.2% higher than BiSeNet, respectively. Experiments show that B2Net can better capture spatial detail and high-level semantic information, improve the segmentation accuracy of buildings in complicated backgrounds, and meet the needs of rapid extraction of buildings.

Keywords　Building extraction, Feature fusion, Atrous convolution, Depthwise separable convolution, Content aware attention

1　引言

遥感图像的语义分割已得到广泛的应用[1],如道路检测、土地覆盖物分类、建筑物提取等。其中,建筑物提取作为遥感图像语义分割的一个重要应用,目的是识别出环境中属于建筑的像素,在遥感研究中具有重要意义。

随着遥感图像采集技术的不断发展,遥感图像的分辨率不断提升,并且地面采样距离也不断增大,采集到的图像通常包含了丰富的土地覆盖信息和复杂的环境背景,使得遥感影像具有类内方差大、类间差异小的特性,进一步增加了在遥感

到稿日期:2022-08-09　返修日期:2022-12-10
基金项目:山西省交通建设科技项目(2019-2-8)
This work was supported by the Transportation Construction Science and Technology Project in Shanxi Province(2019-2-8).
通信作者:岳雪燕(yue@s.upc.edu.cn)

图像中提取建筑物的难度。

传统的建筑物提取方法主要采用手工特征作为提取建筑物的关键特征，如局部结构[2]（边、线、角）、阴影[3]、纹理特征[4]和遥感影像多光谱特征[5]等。将这些特征与支持向量机[6]和遗传算法[7]相结合，对建筑物进行检测和分类。但是这类方法的性能依赖手工特征的提取，不能应用于大规模数据集，无法满足当前实际应用的需求。

随着深度学习的发展，卷积神经网络在计算机视觉领域表现出了优异的性能。Long 等[8]提出了全卷积网络（Fully Convolutional Network，FCN），去除了全连接层，提升了分割的效率，降低了模型复杂度；U-Net[9]提出了基于跳层连接的编解码结构，较好地改善了目标边界分割效果较差的问题，在小样本数据集中取得了不错的效果。但是其跳层连接和特征融合模块对高层语义信息与低层信息进行了平等处理，不能很好地提取出具有较强语义性的特征，在大型数据集上特征表达能力受限（提取较小建筑和细化建筑边界能力欠缺）。为了更好地解决这个问题，DeepLab[10-13]提出了空洞空间池化金字塔（Atrous Spatial Pyramid Pooling，ASPP），通过多个具有不同膨胀率的空洞卷积并行分支来提取不同尺度的上下文信息，并结合全局上下文融合模块来进一步提高模型的表征能力。但是其在处理大量高分辨率特征图时，会占用大量内存，阻碍了对高分辨率图像的研究。

遥感图像中的建筑物在不同地区差异较大，例如建筑物的大小不一、建筑物群的疏密程度不同等，这就要求模型提取的信息中包含更多的多尺度信息（其中低级特征包含丰富的空间信息，高级特征包含丰富的语义信息）和局部信息。对于多尺度特征融合，MFRN[14]是直接将相邻层级的特征进行拼接；Link-Net[15]则是直接进行特征的相加。虽然这些设计可以从低级特征中聚合空间信息，但也引入了低层冗余信息（次要细节和噪声）。因此，如何进行高效的融合是改善分割模型的关键。

注意力机制[16-18]是一种用于提取局部信息的有效方法，如 BAM[19]，CBAM[20]，scSE[21]和 CoordAttention[22]等。在遥感图像的处理中，Li 等[23]以及 Xu 等[24]借鉴注意力思想使得模型的分割精度有所提升，但模型的训练参数量剧增，整体的训练时间增加。同时大部分注意力机制中的下采样层会加重分割边缘的锯齿现象，而边缘分割的效果则会直接影响建筑物提取的准确性，因此如何更好地分割建筑物边缘也是建筑提取任务的核心问题。

本文的主要贡献如下：

1）引入空间卷积金字塔模块（Spatial Convolution Pyramid，SCP），提出深度空洞可分离卷积块（Depthwise Atrous Separable Convolution Block，DAS），利用具有不同膨胀率的并行 DAS 卷积块，结合深度可分离卷积和空洞卷积的优势，在降低模型参数量的同时扩大了感受野，以捕获丰富的多尺度信息。

2）提出了交叉特征融合模块（Cross Feature Fusion，CFM），采用选择性融合策略，使得重要特征之间相互补充，抑制冗余信息，避免特征间的污染。

3）提出了内容感知注意力模块（Content Aware Attention，CAA），细化每个阶段的特征，并利用自适应的低通滤波优化边缘锯齿。

4）提出了以 BiSeNet 轻量化模型为基础的 B2Net 网络结构，实验在 WHU 数据集和马萨诸塞州建筑物数据集（Massachusetts Building Dataset，Massa）上有优异的表现，展示了模型良好的泛化能力。

2 网络结构

如图 1 所示，本文提出了以 BiSeNet[25]轻量化网络结构为基础的 B2Net，B2Net 基于双分支结构，分为空间信息分支和语义信息分支，前者用于提取空间信息，后者用于提取生成的语义信息，将两者的特征进行融合后得到最终的分割结果。

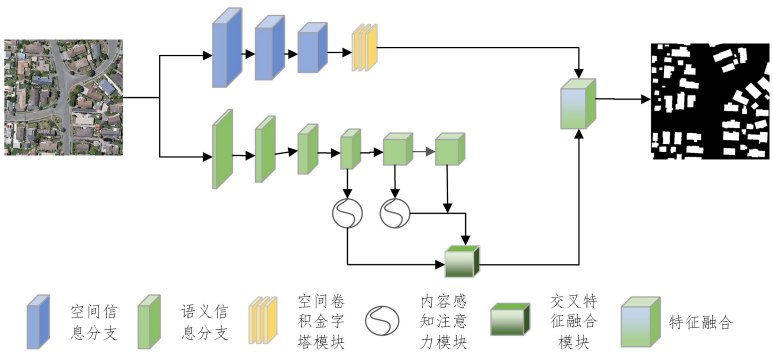


图 1 B2Net 网络的结构图

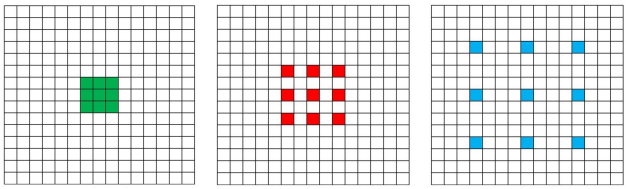
Fig. 1 Network structure diagram of B2Net

3 空间信息分支

空间信息分支由普通卷积层和空间卷积金字塔模块（SCP）组成，通过这个分支可以得到一个粗糙的分割结果。该分支首先通过三层步长为 2 的卷积进行空间细节信息的提取，将特征图尺寸缩小至原始图像的 1/8，然后应用 SCP 模块进行空间上的多尺度特征提取，每个 DAS 具有不同的膨胀率，最终可以得到蕴含丰富空间信息的特征图。

受 ASPP[11]的启发，本文提出了 SCP 模块，用于捕获多尺度信息。空洞卷积[26]可以在不增加参数和不降低分辨率的前提下有效地扩大感受野。如图 2 所示，图 2(a)为 3×3 的普通卷积；图 2(b)为膨胀率为 2 的空洞卷积，其感受野与 5×5 的卷积相同；图 2(c)为膨胀率为 4 的空洞卷积，感受野为9×9 大小，因此可以通过设置不同膨胀率，来得到不同

感受野，以获取多尺度建筑物的空间信息。但是由于空洞卷积的网格效应，进行多次叠加之后部分像素并没有参与计算，这就会丢失信息的连续性。因此，本文采用深度可分离卷积[27]和空洞卷积结合的方式，在提升性能的同时也可以获得较大的感受野。



(a)膨胀率=1 (b)膨胀率=2 (c)膨胀率=4

图 2 空洞卷积在不同膨胀率下的结构图

Fig. 2 Structure diagram of atrous convolution at different expansion rates

如图 3 所示，SCP 模块由 4 个 DAS 卷积块并行组成，其中每个 DAS 卷积块的膨胀率为 2^{i-1}，$i=1,2,\cdots,n$。首先，它将输入特征图的通道数压缩为原来的 1/4，然后将 DAS 卷积块的并行结构用于降维后的特征图。最后，将各个分支的输出进行融合之后与最初的输入特征图进行跳层连接，以突出不同尺度目标的空间细节特征，让模型更好地学习这些特征，提高模型对多尺度建筑物特征提取的能力。

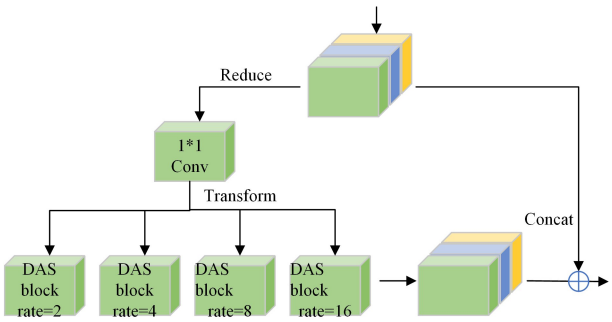


图 3 SCP 模块的结构图

Fig. 3 Structure diagram of SCP module

DAS 卷积块(见图 4)是在空洞卷积之前利用深度卷积进行局部特征的聚合，使得特征图中的每个像素点蕴含其邻域信息。但是简单地将空洞卷积和深度卷积进行组合仍然会忽略大量相邻像素点的信息，对模型的精度也会有所影响，因此 DAS 卷积块使用两个深度上不对称的卷积来聚合更多的局部信息，以减少特征信息的损失。由于深度卷积没有融合通道间的信息，因此 DAS 卷积块在最后使用 1×1 的点卷积[28]来混合不同通道的信息。DAS 卷积块通过聚合局部和全局信息，加强了模型的特征提取能力。

ASPP 模块中空洞卷积的膨胀率为 6,12,18，但在本文的网络结构中，随着空间信息分支对特征的不断提取，特征图的分辨率会逐渐降低，上述组合则不能有效地提取多分辨率特征图的特征。较大的膨胀率会导致模型欠缺分割小建筑物的能力，从而减弱网络分割多尺度建筑物的能力。为了更有效地提取多分辨率特征图的特征，提高网络提取不同大小建筑物的能力，DAS 卷积块的膨胀率设为 2^{i-1}，$i=1,2,\cdots,n$，使不同大小膨胀率的卷积核能捕获多尺度建筑物的信息，不同层级的膨胀率可以有效地提高模型提取多尺度建筑物的能力。

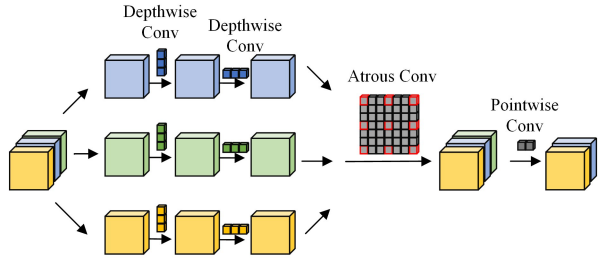


图 4 DAS 卷积块的结构图

Fig. 4 Structure diagram of DAS convolution block

4 语义信息分支

当空间信息分支捕获了丰富的空间信息时，语义信息分支则利用上下文信息，预测出精细的分割边界。首先，为了提取语义上的多尺度特征，本文将图像下采样为原始尺寸的 1/4,1/8,1/16,1/32。对中间两层应用内容感知注意力模块(CAA)，使边缘得到平滑处理，并增强边缘特征。将这两层的输出通过交叉特征融合模块(CFM)进行特征融合，最后将两个分支的输出进行融合，使用双线性插值方法进行 8 倍的上采样，得到与原始图相同尺寸的分割结果。

4.1 交叉特征融合

高分辨率航空遥感图像背景复杂[29]，低层特征虽然含有背景噪声，但具有较多的细节信息和边界信息，这对于生成准确的预测结果非常重要。相反，由于多次下采样，高层特征在边界上比较粗糙，丢失了大量细节信息，但是高层特征具有丰富的语义信息。为了防止引入过多的冗余信息，特征融合的方式尤为重要。因此，本文建立 CFM 模块(见图 5)，用于细化高层特征和低层特征。

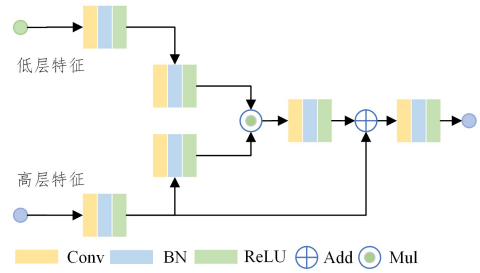


图 5 CFM 模块示意图

Fig. 5 Sketch map of CFM module

CFM 模块包含两个输入分支，一个输入为低层语义特征(F_l)，另一个输入为高层语义特征(F_h)。现有研究中采用的直接相加或拼接的方法会造成特征的冗余，“污染”原始特征，从而对特征图产生影响，影响网络的性能。本文通过逐元素乘法提取 F_l 和 F_h 之间的公共部分，F_h 会逐渐吸收 F_l 的有用信息，即 F_h 的边界会被锐化。如式(1)所示，CFM 模块将两个 3×3 卷积层分别应用于 F_l 和 F_h，输出相同大小的特征图，以适应模块的后续处理，通过元素间的乘法将特征进行变换和融合，融合之后的特征既具有高层语义信息也具有低层

空间信息。最后，将融合的特征添加到高层语义信息 F_h 中，以补充高层特征图所缺失的特征边界。

$$F_h = F_h + M_h(G_l(F_l) * G_h(F_h)) \tag{1}$$

其中，M_h，G_h，G_l 分别为 3×3 卷积层、批标准化（Batch Normalization，BN）和激活函数（Rectified Linear Unit，ReLU）的组合。

4.2 内容感知注意力

锯齿[30]是采样过程中的常见现象，指高频信息被采样后退化成的完全不同的信息。在深度学习领域中，常见于最大池化、跨步卷积等下采样操作中。解决该问题的一种方法是在下采样之前进行低通滤波，如高斯模糊，但这些方法忽略了特征图的频率信息在不同空间位置以及不同通道中的差异。为了解决上述问题，本文在 B2Net 网络的语义信息路径中引入 CAA 模块（见图 6）。由于图像不同位置的频率信息各不相同，不同通道也会从不同角度捕获输入的特征信息，例如某些通道捕获边缘，其他通道捕获颜色等。因此，采用自适应的低通滤波层动态地调整特征信息的权重，避免高频信息混叠，进而细化分割边界。

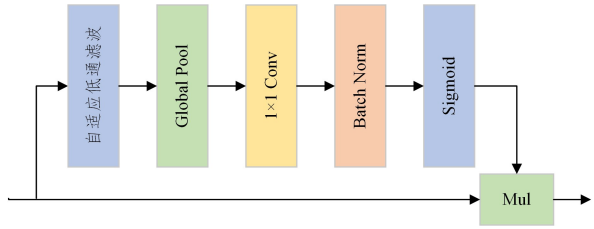


图 6 CAA 模块示意图

Fig. 6 Sketch map of CAA module

如图 7 所示，在自适应低通滤波层中，首先通过压缩激活模块（Squeeze-and-Excitation Block，SE）[31]得到每个空间位置 (i,j) 所需的低通滤波器，然后将其作用于输入特征 X，得到输出特征 Y。为降低计算成本，将输入通道划分为 k 组，每组预测一个低通滤波器，整个过程如式（2）、式（3）所示：

$$Y_{i,j} = \sum_{p,q\in\Omega} w_{i,j}^{p,q} \cdot X_{i+p,j+q} \tag{2}$$

$$Y_{i,j}^{g} = \sum_{p,q\in\Omega} w_{i,j,g}^{p,q} \cdot X_{i+p,j+q}^{c} \tag{3}$$

其中，Y 表示输出特征，Ω 表示应用低通滤波器 (i,j) 周围的位置集合，低通滤波器 w，g 表示分组序号。

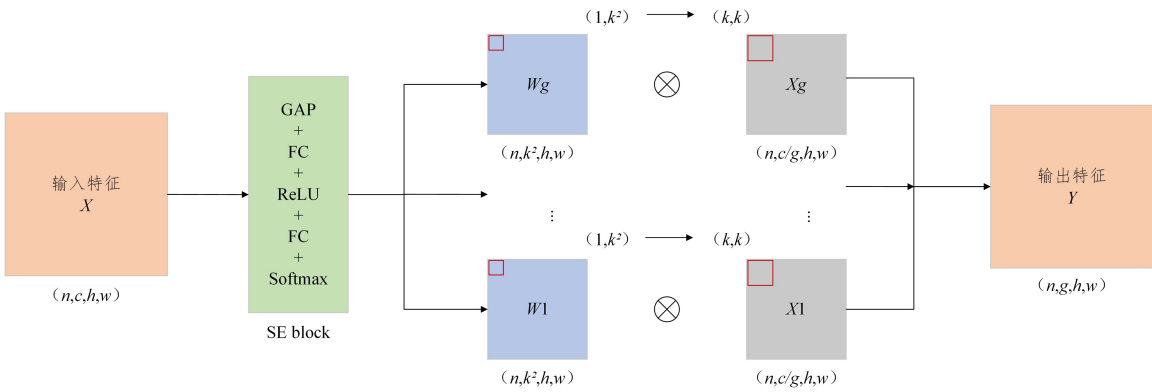


图 7 自适应低通滤波示意图

Fig. 7 Sketch map of adaptive low-pass filter

CAA 模块通过自适应低通滤波对输入特征的不同位置、不同通道预测滤波权值，再使用全局平均池化来聚合全局上下文信息，计算特征权重来指导模型学习。这种设计不仅可以优化语义信息分支中相邻阶段的输出特征，还可以减少锯齿现象，同时突出强相关性的特征。

5 实验与结果

5.1 实验数据与训练设置

本文在 WHU 数据集[32]和 Massa 建筑物数据集[33]上进行训练与评估。

WHU 数据集由从遥感图像中提取的 22 000 多座独立建筑组成，覆盖面积为 450 km^2，拍摄于新西兰基督城。航空图像的大部分被降采样到 0.3m 的空间分辨率，并被裁剪成 8189 个 512×512 像素的非重叠图块，这些图块构成了整个数据集。然后将被裁减后的图像分为 3 部分，训练数据集包含 4736 张，验证数据集包含 1036 张，测试数据集包含 2416 张。

Massa 建筑物数据集由 151 张波士顿地区的航拍图片组成，每张图片的空间分辨率为 1m，像素为 1500×1500。训练集包含 141 张图片，测试集包含 10 张图片。为了与 WHU 数据集的图像保持相同大小，本文将图像有重叠地裁剪成 512×512 像素大小。两个数据集的示例如图 8 所示。

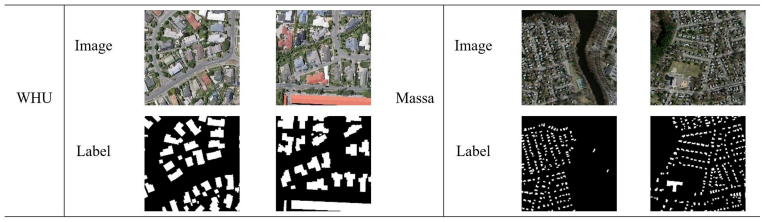


图 8 数据集示例图

Fig. 8 Example diagram of datasets

实验环境为 Ubuntu 16.04 LTS，GPU 配置为 TSLA P100；网络模型均使用 python 3.6，pytorch 1.2 进行训练与测试。

5.2 实验评价指标

为了进行公平比较，本文使用与其他文献中相同的指标。对于 WHU 数据集和 Massa 建筑物数据集，使用准确率、召回率、F1 分数和交并比(IoU)进行定量性能评估；使用浮点运算数(FLOPs)和参数量(Parameters)来衡量模型的计算复杂度和规模。

精度指正确分类的正像素(建筑物)占分类器预测的所有正像素(建筑物)的比例；而召回率(也称完整性)指正确分类的正像素在所有真实的正像素中的比例；F1 分数是准确度和召回率的加权平均值，它同时考虑了 FP 和 FN；IoU 是预测的正像素区域和真实的正像素区域交集与并集的比值。所有性能指标的计算式如式(4)－式(7)所示：

$$Precision=\frac{TP}{TP+FP} \tag{4}$$

$$Recall=\frac{TP}{TP+FN} \tag{5}$$

$$F1=\frac{2\times Precision\times Recall}{Precision+Recall} \tag{6}$$

$$IoU=\frac{TP}{TP+FP+FN} \tag{7}$$

其中，TP 表示正确分类的目标像素数，TN 表示正确分类的非目标像素的数量，FP 是分类为目标的非目标像素数，FN 是分类为非目标的目标像素数。

FLOPs 是浮点运算数，Parameters 是模型的参数数量，计算式如式(8)、式(9)所示：

$$FLOPs=2HW(C_{in}K^2+1)C_{out} \tag{8}$$

$$Parameters=K^2\times C_{in}\times C_{out} \tag{9}$$

其中，H 和 W 是输入特征图的长度和宽度，C_{in} 是输入通道数，C_{out} 是输出通道数，K 是卷积核大小。

5.3 对比实验

本文在 WHU 数据集和 Massa 建筑物数据集上将 B2Net 与一些语义分割模型进行比较，包括 U-Net，SegNet，DeepLabv3＋，EU-Net[34]，BiSeNet，以评估其有效性。

在对比实验中，模型均使用 512×512 像素的图像进行训练。不同的是，U-Net，EU-Net 的优化方法为 RMSprop，损失函数为 BCE WithLogitsLoss；SegNet 和 DeepLabv3＋使用 Adam 优化器，损失函数为 BCELoss。4 个模型的初始学习率均为 0.00001，BiSeNet 和 B2Net 的损失函数为 Dice，优化器为 SGD，初始学习率为 0.0025。表 1 和表 2 分别列出了 6 种模型在 WHU 数据集和 Massa 建筑物数据集上的对比结果。

表 1 6 种模型在 WHU 数据集上的对比结果

Table 1 Comparison results of six models on WHU dataset

(单位：%)

Method	*Precision*	*Recall*	*F1-score*	*IoU*
U-Net	92.28	85.02	88.50	79.38
SegNet	92.11	89.93	91.01	85.56
DeepLabv3＋	93.11	92.99	93.05	85.66
BiSeNet	98.30	98.10	97.20	86.00
EU-Net	94.98	95.10	95.04	87.56
Ours	**98.60**	**99.40**	**99.30**	**88.50**

表 2 6 种模型在 Massa 建筑物数据集上的对比结果

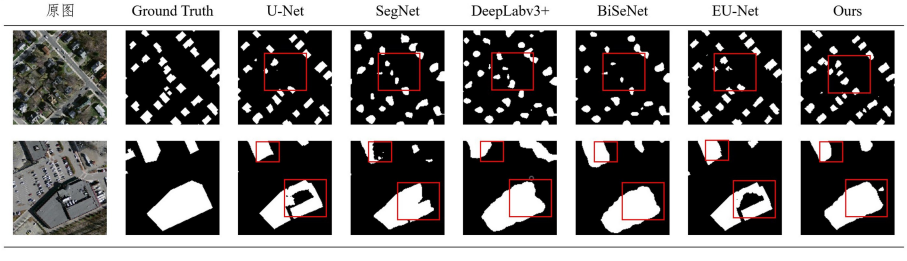
Table 2 Comparison results of six models on Massa building dataset

(单位：%)

Method	*Precision*	*Recall*	*F1-score*	*IoU*
U-Net	85.99	79.61	82.48	70.69
SegNet	89.68	82.64	83.84	71.87
DeepLabv3＋	90.32	87.66	84.43	74.58
BiSeNet	93.10	95.40	94.50	81.10
EU-Net	86.70	83.40	85.01	73.93
Ours	**94.00**	**97.30**	**96.20**	**83.30**

首先分析基于像素的度量。如表 1 所列，在 WHU 数据集上，U-Net 的精度高于 SegNet，但召回率、IoU 等却低于 SegNet。本文提出的 B2Net 网络相比 BiSeNet IoU 提高了 2.5%。如表 2 所列，在 Massa 建筑物数据集上，相比 BiSeNet，B2Net 的 IoU 提升了 2.2%，而 DeepLabv3＋和 EU-Net 等的评价指标值都较低。定量评估表明，不管在 WHU 数据集上还是在 Massa 建筑物数据集上，B2Net 在 IoU、F1 分数、召回率和精度方面都优于所有对比方法，这意味着其可以增强图像整体特征和信息。而且 B2Net 没有任何后处理步骤，节省了分割任务的时间。

其次，通过可视化结果和标签来展示模型的预测效果，并用红色方框标明区别较大的区域。图 9 给出了 Massa 建筑物数据集上的可视化结果。对于图 9 第 1 行的小型建筑物而言，EU-Net 相比 U-Net 漏检现象有所缓解，SegNet 和 DeepLabv3＋只能识别建筑物的大体位置，而 B2Net 网络准确提取了建筑物细节。第 2 行的建筑物有阴影和凹陷，这种情况下，U-Net 和 EU-Net 将阴影凹陷误分为背景，只有 B2Net 提取了建筑物完整的轮廓特征。但图 9 第 2 行的结果也说明了本文模型对于约束建筑物形状的能力有所欠缺。



注：红色框为差异较大的区域。

图 9 Massa 建筑物数据集上的建筑物提取结果(电子版为彩图)

Fig. 9 Results of building extraction on Massa building dataset

图 10 给出了 WHU 数据集上不同场景下的分割结果。第一种情况是超大建筑：U-Net，SegNet，DeepLabv3＋和 EU-Net 将遮挡物误分为建筑物。第二种情况是分布密集的中型建筑：从整体来说，本文方法对相邻建筑物边缘的提取精度较高，保证了建筑物的主要结构，而剩余 5 个模型都没完整地预测出这种情况。第三种情况是大型建筑的孔洞现象：从图 10 的第 3 行可以明显地看出，本文方法分割的建筑物的孔洞有较大的改善。第四种情况（图 10 中的第 4—6 行）是密集和稀疏分布的小型建筑：小建筑物被误分的现象得到改善，但 6 个模型在极其微小的建筑上精度略有下降。

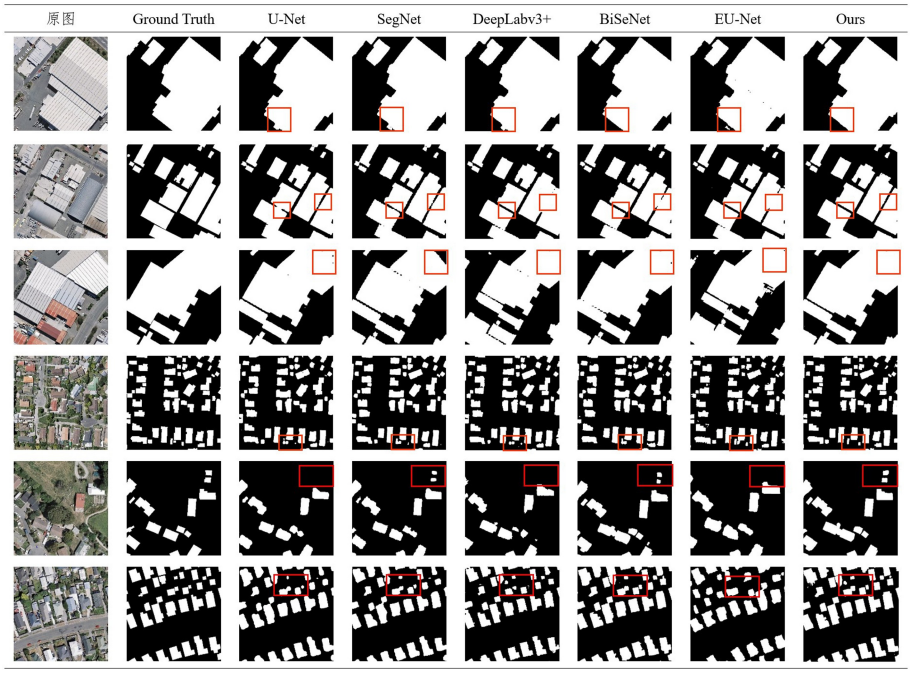


图 10 WHU 数据集上的建筑物提取结果

Fig. 10 Results of building extraction on WHU dataset

最后，由于多数用于分割任务的网络都是编码-解码结构，它们通过深层网络来实现较高精度，因此本文在参数量和浮点计算上将 B2Net 与经典编码-解码结构（U-Net，SegNet）、现有方法（HRNet[35]，LinkNet）以及实时分割网络（ICNet[36]）进行对比。表 3 列出了 6 种模型的复杂度对比结果。由表 3 可知，B2Net 在浮点运算和参数量上相比编码-解码网络有大幅度的下降，使得网络更加轻量化，提升了网络的运行效率。

表 3 模型复杂度的对比结果

Table 3 Comparison results of model complexity

Model	*Parameters*	*GFLOPs*
SegNet	29.44×10^6	160.32
U-Net	26.78×10^6	247.53
HRNet	28.85×10^6	57.64
ICNet	26.50×10^6	28.30
LinkNet	25.65×10^6	27.92
Ours	$\mathbf{23.34\times10^6}$	**16.66**

综上，不论是在模型复杂度，还是在提取建筑物形状的完整性方面，B2Net 都取得了最佳的结果。

5.4 消融实验

为了验证所提模块的有效性，本文在 WHU 数据集上对每个模块进行了消融实验，即 SCP 模块、CFM 模块和 CAA 模块。所有结果都是在 WHU 数据集的训练集上进行训练并在测试集上进行评估而获得的。如表 4 所列，BiSeNet 达到了 86.0% 的 IoU。通过添加 CFM 模块，获得了 86.5% 的 IoU，提高了 0.5%；添加 SCP 模块带来 1.6% 的 IoU 收益；CFM 模块和 SCP 模块是空间多尺度和语义多尺度，两者结合可将 IoU 提高到 88.3%，从图 11 可以看出，增加了多尺度操作之后，模型提取不同大小建筑物的能力有所提高。

表 4 消融实验的定量结果

Table 4 Quantitative results of ablation experiment

（单位：%）

Method	*Precision*	*Recall*	*F1-score*	*IoU*
CFM	98.40	95.50	96.00	86.50
SCP	98.50	98.90	98.10	87.60
CAA	98.40	**99.80**	**99.70**	86.60
CFM＋SCP	97.80	97.70	97.60	88.30
CAA＋SCP	98.40	98.70	98.80	86.90
CAA＋CFM	98.50	98.80	98.60	87.60
CAA＋CFM＋SCP	**98.60**	99.40	99.30	**88.50**

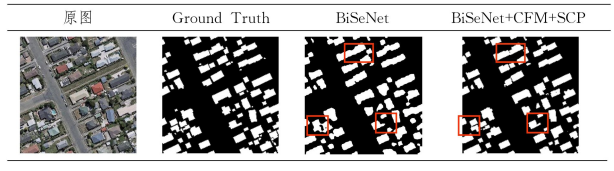


图 11 消融实验结果

Fig. 11 Results of ablation experiment

3 个模块结合后的网络 IoU 可以达到 88.50%，从图 12 可看出，分割的建筑物锯齿现象得到了较大的改善。实验结果表明，本文方法的每个模块都很重要，有助于最终的准确性，并为建筑物提取的语义分割任务带来了巨大的好处。

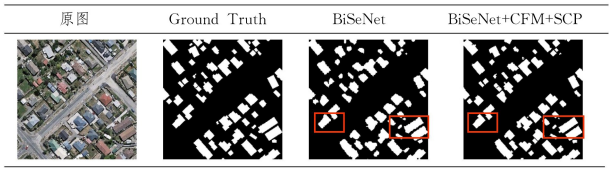


图 12 边缘对比图

Fig. 12 Diagram of edge comparison

结束语 针对建筑物分割方法中存在的空间和语义信息提取不足的问题，本文提出了一个有效的基于双分支的B2Net模型，用于从遥感图像中提取建筑物。该网络由空间信息分支和语义信息分支组成。引入内容感知注意力模块使得边界得到平滑处理；在空间信息分支上，SCP模块对不同尺度建筑物采用不同大小的膨胀率进行特征提取，这不仅扩大了模型的有效接受域，而且捕获了更多的多尺度信息；在语义信息分支上，CFM模块用于多尺度的特征融合；模型从空间和语义两个角度来加强模型提取多尺度建筑物的能力。大量消融实验证明了所提方法的有效性。与现有的语义分割方法的对比实验表明，B2Net在WHU数据集和Massa建筑物数据集上的所有评价指标都有显著提高。虽然本文模型取得了令人满意的结果，但是从可视化结果来看，建筑物的形状未得到很好的约束。在未来的研究中，尝试在本文模型的基础上增加形状约束，修改损失函数或调整网络结构，以改善建筑物的分割效果。

参 考 文 献

[1] ZOU W, JING W, CHEN G, et al. A survey of big data analytics for smart forestry[J]. IEEE Access, 2019, 7: 46621-46636.

[2] HUERTAS A, NEVATIA R. Detecting buildings in aerial images[J]. Computer Vision, Graphics, and Image Processing, 1988, 41(2): 131-152.

[3] PENG J, LIU Y C. Model and context-driven building extraction in dense urban aerial images[J]. International Journal of Remote Sensing, 2005, 26(7): 1289-1307.

[4] LEVITT S, AGHDASI F. An investigation into the use of wavelets and scaling for the extraction of buildings in aerial images [C]// Proceedings of the 1998 South African Symposium on Communications and Signal Processing-COMSIG'98 (Cat. No. 98EX214). IEEE, 1998: 133-138.

[5] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[6] TURLAPATY A, GOKARAJU B, DU Q, et al. A hybrid approach for building extraction from spaceborne multi-angular optical imagery[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2012, 5(1): 89-100.

[7] SUMER E, TURKER M. An adaptive fuzzy-genetic algorithm approach for building detection using high-resolution satellite images[J]. Computers, Environment and Urban Systems, 2013, 39: 48-62.

[8] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3431-3440.

[9] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// International Conference on Medical Image Computing and Computer-assisted Intervention. Cham: Springer, 2015: 234-241.

[10] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. arXiv: 1412.7062, 2014.

[11] 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, 2017, 40(4): 834-848.

[12] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv: 1706.05587, 2017.

[13] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018: 801-818.

[14] LI L, LIANG J, WENG M, et al. A multiple-feature reuse network to extract buildings from remote sensing imagery[J]. Remote Sensing, 2018, 10(9): 1350-1367.

[15] CHAURASIA A, CULURCIELLO E. Linknet: Exploiting encoder representations for efficient semantic segmentation[C]// 2017 IEEE Visual Communications and Image Processing (VCIP). IEEE, 2017: 1-4.

[16] ZHONG Z, LIN Z Q, BIDART R, et al. Squeeze-and-attention networks for semantic segmentation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 13065-13074.

[17] CAO J, CHEN Q, GUO J, et al. Attention-guided context feature pyramid network for object detection[J]. arXiv: 2005.11475, 2020.

[18] DAI Y, GIESEKE F, OEHMCKE S, et al. Attentional feature fusion[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021: 3560-3569.

[19] PARK J, WOO S, LEE J Y, et al. Bam: Bottleneck attention module[J]. arXiv: 1807.06514, 2018.

[20] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018: 3-19.

[21] ROY A G, NAVAB N, WACHINGER C. Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks [C]// International Conference on Medical Image Computing and Computer-assisted Intervention. Cham: Springer, 2018: 421-429.

[22] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 13713-13722.

[23] LI C L, HUANG F H, HU W, et al. Building Extraction from High-Resolution Remote Sensing Image based on Res_Attention Unet[J]. Journal of Geo-Information Science, 2021, 23(12):

2232-2243.

[24] XU C Y，FAN S S，ZHU H. Semantic Segmentation of Remote Sensing Images Using The Channel Domain Attention Mechanism Deeplabv3＋ Algorithm [J]. Control Engineering，2023，30(2)：368-375.

[25] YU C，WANG J，PENG C，et al. Bisenet：Bilateral segmentation network for real-time semantic segmentation[C]// Proceedings of the European Conference on Computer Vision(ECCV). 2018：325-341.

[26] YU F，KOLTUN V. Multi-scale context aggregation by dilated convolutions[J]. arXiv：1511. 07122，2015.

[27] CHOLLET F. Xception：Deep learning with depthwise separable convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017：1251-1258.

[28] DU S J，DU S H，LIU B，et al. Incorporating DeepLabv3＋ and object-based image analysis for semantic segmentation of very high resolution remote sensing images[J]. International Journal of Digital Earth，2021，14(3)：357-378.

[29] ZHANG L，DONG R，YUAN S，et al. Making low-resolution satellite images reborn：a deep learning approach for super-resolution building extraction[J]. Remote Sensing，2021，13(15)：2872.

[30] ZHANG R. Making convolutional networks shift-invariant again [C]// International Conference on Machine Learning. PMLR，2019：7324-7334.

[31] HU J，SHEN L，SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018：7132-7141.

[32] JI S P，WEI S Q. Building extraction via convolutional neural networks from an open remote sensing building dataset [J]. Journal of Geomatics，2019，48(4)：448-459.

[33] MNIH V. Machine learning for aerial image labeling[D]. University of Toronto(Canada)，2013.

[34] KANG W，XIANG Y，WANG F，et al. EU-Net：An efficient fully convolutional network for building extraction from optical remote sensing images[J]. Remote Sensing，2019，11(23)：2813.

[35] WANG J，SUN K，CHENG T，et al. Deep high-resolution representation learning for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2020，43(10)：3349-3364.

[36] ZHAO H，QI X，SHEN X，et al. Icnet for real-time semantic segmentation on high-resolution images[C]// Proceedings of the European Conference on Computer Vision(ECCV). 2018：405-420.

CHEN Guojun，born in 1968，associate professor，is a member of China Computer Federation. His main research interests include graphics and image processing，virtual reality，and BIM technology.

YUE Xueyan，born in 1998，postgraduate. Her main research interest is computer vision.

（责任编辑：喻藜）