

doi: 10.6046/gtzyyg.2020230

引用格式: 郭文,张莽. 基于注意力增强全卷积神经网络的高分卫星影像建筑物提取[J]. 国土资源遥感,2021,33(2):100 – 107. (Guo W,Zhang Q. Building extraction using high – resolution satellite imagery based on an attention enhanced full convolution neural network[J]. Remote Sensing for Land and Resources ,2021 ,33(2):100 – 107.)

基于注意力增强全卷积神经网络的 高分卫星影像建筑物提取

郭 文¹, 张 莽²

(1. 自然资源部第三航测遥感院,成都 610100; 2. 西南石油大学地球科学与技术学院,成都 610500)

摘要:从卫星遥感影像中自动提取建筑物在经济社会发展中具有广泛的应用价值,由于卫星遥感影像存在地物遮挡、光照、背景环境等因素的影响,传统方法难以实现高精度建筑物提取。采用一种基于注意力增强的特征金字塔神经网络方法(FPN – SENet),利用多源高分辨率卫星影像和矢量成果数据快速构建大规模的像素级建筑物数据集(SCRS 数据集),实现多源卫星影像的建筑物自动提取,并与常用的全卷积神经网络进行对比。研究结果表明: SCRS 数据集的提取精度接近国际领先的卫星影像开源数据集,且假彩色数据精度高于真彩色数据; FPN – SENet 的建筑物提取精度优于其他常用的全卷积神经网络; 采用交叉熵和 dice 系数之和为损失函数能够提升建筑物提取精度,最好的分类模型在测试数据上的分类总体精度为 95.2% ,Kappa 系数为 79.0% ,F1 分值和 IoU 分别达到了 81.7% 和 69.1% 。该研究可为高分辨率卫星影像建筑物自动提取提供参考。

关键词: 国产高分卫星影像; 建筑物; 语义分割; 注意力增强

中图法分类号: TP 751; P 237 **文献标志码:** A **文章编号:** 1001 – 070X(2021)02 – 0100 – 08

0 引言

基于高分辨率遥感影像的建筑物提取对地图制作、城市乡村规划、违建监测、数字城市建立、人口估计等应用具有重要意义,随着越来越多的高分辨遥感卫星投入使用,在提高地物光谱特征,突出地物结构、纹理和细节等信息的同时,也因为卫星观测角度问题造成地物遮挡,尺度的增大带来了严重的异物同谱现象,以及影像噪声增加等问题,限制了遥感影像建筑物自动提取精度。

传统方法主要是基于影像的光谱和空间特征信息,通过图像分割和特征提取技术,如尺度不变特征变换(scale – invariant feature transform, SIFT)、方向梯度直方图(histogram of oriented gradient, HOG)、面向对象^[1–2]等,获得遥感影像中的特征或分割对象,再选取支持向量机(support vector machine, SVM)、决策树、随机森林、条件随机场等分类器进行类别划分^[3–4]。然而,基于光谱和空间特

征信息的分类方法在处理背景复杂、地物繁多的影像时都显得不足,其中最大缺陷是泛化能力较差,即训练后的模型或规则难以适用不同地区、不同数据源的遥感影像。

2006 年 Hinton 等^[5]提出了深度学习的理论,深度学习是一种通过构建神经网络来模拟学习和人脑思维方式的自主特征学习方法,通过非线性表达来获取数据中的高级抽象特征,并构建数学模型以提高分类精度和检测准确率。近年来各种深度学习网络,如卷积神经网络(convolutional neural network, CNN)^[6]、递归神经网络(recurrent neural network, RNN)^[7]完成了各种高性能的计算机视觉任务,如图像分类^[8]、自然语言处理^[9]、语音识别^[10]等。

由于 CNN 在图像分类方面取得了较好的效果,许多学者基于 CNN 发展了遥感影像语义分割算法。2015 年,从 CNN 发展而来的全卷积神经网络(fully convolutional networks, FCN)^[11]语义分割方法,采用全卷积层代替传统 CNN 中的全连接层,可以直接获得像素级别的分类结果,是第一个真正的端到端的

收稿日期: 2020 – 07 – 21; 修订日期: 2021 – 03 – 29

基金项目: 四川省自然资源科研项目“基于深度注意网络的多云多雨地区土地利用精准提取方法”(编号: KJ – 2020 – 4)和国家基础测绘科技与标准计划“信息化测绘基地建设方案设计与论证”(编号: 2018KJ0304)共同资助。

第一作者: 郭 文(1964 –),男,高级工程师,主要从事摄影测量与遥感研究。Email: 451362006@qq.com。

通信作者: 张 莽(1985 –),男,博士,高级工程师,主要从事遥感大数据应用研究。Email: scrs_qiaozh@163.com。

像素分类模型^[12-13], 目前已经有多种语义分割网络, 如 Segnet^[14], U-net^[15], PSPNet^[16] 和 DeepLab^[17] 等, 这些语义分割网络的提出为遥感影像建筑物自动提取提供了可能性。但是将 FCN 应用到遥感影像建筑物提取时, 需要通过逐像素标记类别构建样本数据集, 这也阻碍了 FCN 在遥感影像分类中的深入应用。

目前, 国际上已经有少量开源的遥感影像建筑物样本数据集, 如 ISPRS 的 Vaihingen 和 Potsdam 数据集 (<http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html>), Massachusetts 数据集^[18]、Inria 数据集^[19]、WHU 数据集^[20] 等, 以上数据集主要是以航空影像为主, 仅 WHU 数据集包含一部分卫星影像, 但在一些重大项目中, 如第三次全国国土调查、地理国情监测、全球地理信息资源建设等, 仍需要构建大量的样本以满足重大项目建筑物自动提取的需求。

针对建筑物自动提取的难题, 本文提出基于高分卫星遥感影像和地理国情监测矢量数据快速构建

大规模的建筑物样本数据集, 并采用一种注意力增强的全卷积神经网络, 实现从卫星遥感影像中自动提取建筑物。

1 研究方法

1.1 基于注意力增强的特征金字塔建筑物提取网络

针对遥感影像的多尺度问题, 传统的方式是通过多尺度输入图像的方式来构建多尺度特征, 特征金字塔网络 (feature pyramid network, FPN)^[21] 借鉴了这一思想, 但是在网络内部构建特征金字塔, 实现多尺度特征融合, 将包含更多语义信息的高层网络特征与包含更多几何细节信息的浅层网络特征串联融合。本文针对高分卫星遥感影像的建筑物自动提取, 采用一种基于 FPN 的语义分割网络 FPN-SENet, 提升遥感影像建筑物分类精度, FPN-SENet 主体网络框架如图 1 所示, 而在特征图提取部分, 采用了 SE (squeeze-and-excitation) 模块^[22], 如图 2 所示。

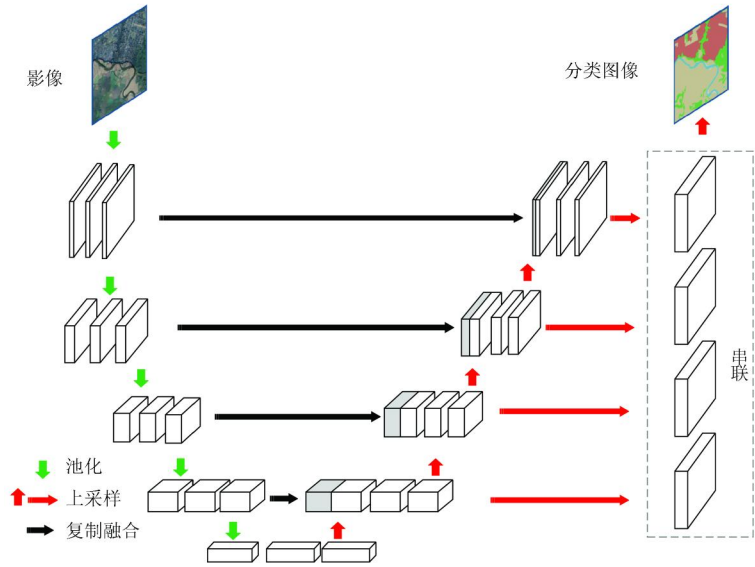


图 1 FPN-SENet 网络框架
Fig. 1 Structure diagram of FPN-SENet

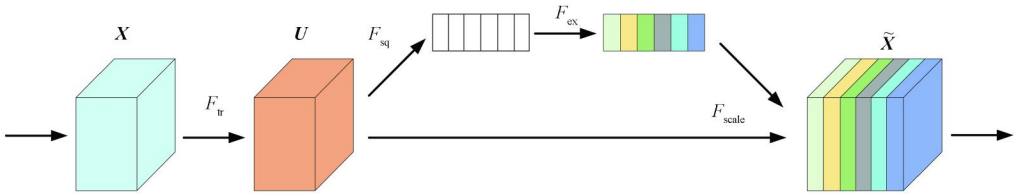


图 2 SE 模块单元
Fig. 2 Squeeze-and-excitation block

一个 SE 模块单元中, F_{tr} 为传统的卷积结构, X 和 U 分别为输入 ($H' \times W' \times C'$) 和输出 ($H \times W \times C$), $U = [u_1, u_2, \dots, u_c]$, 公式为:

$$u_c = v_c \times X = \sum_{s=1}^{C'} v_c^s \cdot x^s, \quad (1)$$

式中: x^s 为第 c 个通道的输入特征图; u_c 为第 c 个

通道的输出特征图; $\mathbf{v}_c$ 为第 c 个卷积核的参数; C' 为输入通道数量。

增加 SE 模块后,首先对卷积层输出的特征图 $\mathbf{U}$,采用全局平均池化(global average pooling, GPA)进行压缩操作 F_{sq} ,公式为:

$$z_c = F_{sq}(\mathbf{u}_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \mathbf{u}_c(i, j) \quad (2)$$

然后对压缩后的 z ,使用 2 个全连接层进行激励操作 F_{ex} ,第一个全连接把 C 个通道压缩成了 C/r 个通道来降低计算量,其中 r 是压缩的比例,后面跟了 ReLU 函数,第二个全连接再恢复回 C 个通道,后面跟接 Sigmoid 函数,公式为:

$$s = F_{ex}(z, W) = \sigma(g(z, W) = \sigma(W_2 \delta(W_1 z))) \quad (3)$$

式中: δ 为 ReLU 函数; σ 为 Sigmoid 函数; $\mathbf{W}_1$ 为 $C/r \times C$ 的降维矩阵; $\mathbf{W}_2$ 为 $C \times C/r$ 的升维矩阵。

通过压缩和激励操作,学习到各个通道间的关系,得到不同通道的权重,然后将 SE 模块计算出的各通道权重值分别与原特征图对应通道的二维矩阵相乘,得出的结果输出,公式为:

$$\tilde{\mathbf{X}}_c = F_{scale}(\mathbf{u}_c, s_c) = s_c \cdot \mathbf{u}_c \quad (4)$$

SE 模块为网络模型提供了一种通道域注意力机制,让模型更加关注信息量较大的特征图,而抑制那些不重要特征图。

1.2 交叉熵与 dice 系数结合的损失函数

交叉熵损失函数是遥感影像分类任务中最常用的一种损失函数(L_{bce}),可从预测值(pr)与真值(gt)之间的交叉熵计算得到,公式为:

$$L_{bce} = -gt \cdot \ln(pr) \quad (5)$$

本文引入 dice 系数,将交叉熵与 dice 系数之和作为新的损失函数,对 FPN-SENet 网络模型进行训练,

定义为:

$$L_{bce_dice} = L_{bce} + L_{dice} \quad (6)$$

$$L_{dice} = 1 - F \quad (7)$$

$$F = (1 + \beta^2) \frac{precision \cdot recall}{\beta^2 \cdot precision + recall} \quad (8)$$

式中: $precision$ 为准确率; $recall$ 为召回率; β 为准确率和召回率之间的平衡系数,本文 β 取值为 1,表示准确率和召回率同等重要。

1.3 基于窗函数平滑预测的分类后处理方法

由于受计算机硬件设备的限制,采用 FCN 对遥感影像进行训练和预测时,需将遥感影像和样本图像裁剪成较小尺寸的图像进行处理,但是在分块进行遥感影像预测分类时,2 块之间的边界附近可能存在明显拼接缝的现象,常见的方法是通过重叠切割减少分块拼接痕迹^[23],提升分类效果,而本文引入一种二次样条窗函数,进行平滑拼接预测分类,该窗函数由一维二次样条窗函数扩展而来,公式为:

$$w(n) = \frac{w_1(n) + w_2(n)}{w(n)}, 1 \leq n \leq M \quad (9)$$

$$w_1(n) = 2t(n)^2, 1 \leq n \leq 0.25M, 0.75M \leq n \leq M \quad (10)$$

$$w_2(n) = 1 - 2(t(n) - 1)^2, 0.25M < n < 0.75M \quad (11)$$

$$t(n) = 0.5 \left[1 - \cos\left(\frac{2\pi n}{M+1}\right) \right], 1 \leq n \leq M \quad (12)$$

式中: $\overline{w(n)}$ 为 $w(n)$ 的平均值; M 为窗口大小; $t(n)$ 为三角窗函数。

该窗函数的一维剖面 and 二维形状如图 3 所示,将该窗函数的权重系数与模型预测分类的二维图斑矩阵相乘,可以突出分块图斑中心区域,而抑制分块图斑边缘区域,并以图像块边长的 1/2 作为滑动步长,在整幅待预测影像上进行滑动预测,获得平滑预测的分类图斑。

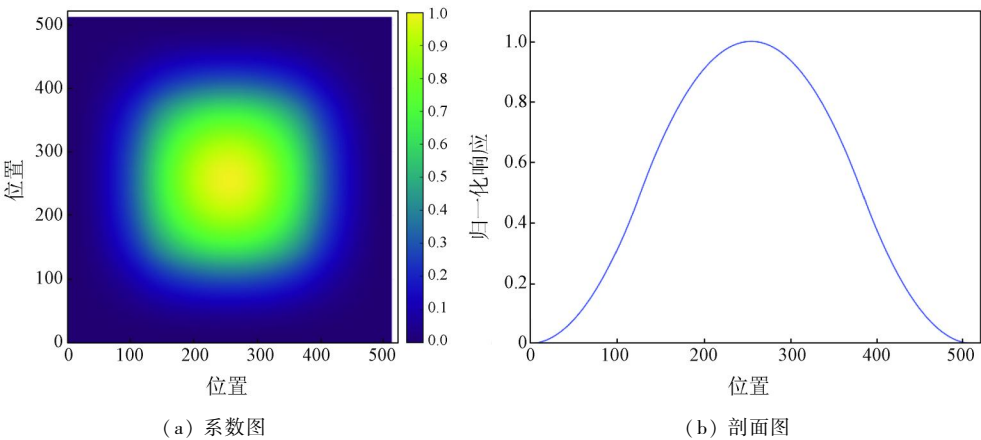


图 3 二次样条窗函数
Fig. 3 Spline window function

2 基于多源卫星影像和矢量成果数据的建筑物样本集制作

深度学习模型训练需要大量的样本库,本文制作了大规模的逐像素标注样本数据集:首先,收集整理地理国情监测成果数据中的地表覆盖矢量数据和对应的高分辨率卫星遥感影像,该地表覆盖数据是基于高分辨率遥感影像进行人工目视解译,并结合一定的指标和规则制作完成,其主要内容和采集指标参考《地理国情普查内容与指标》^[24];然后,选取地表覆盖矢量数据中房屋建筑较多的地区作为样本制作区域,同时去除卫星影像中有云覆盖的区域,裁剪出房屋矢量数据和对应区域的卫星影像,再将房屋矢量数据转化为栅格格式,形成与卫星影像逐像素对应的标签数据,完成建筑物的样本标注;最终,选取了 200 景卫星遥感影像完成多源卫星遥感

建筑物样本数据制作,其中高分二号影像约占 1/2,北京二号和高景一号影像各占 1/4。

本文将制作的建筑物样本数据集简称为 SCRS Buildings Dataset,样本影像的覆盖范围约 1 万 km²,总像素约 10 亿个,其中,建筑物占比为 8%。初始的样本数据尺寸大小并不一致,在模型训练之前,可以根据神经网络模型的输入尺寸将初始样本裁剪成固定大小的小样本,若按照 512 像素 × 512 像素裁剪,可以制作约 50 000 个小样本。由于样本制作采用了融合后的卫星遥感影像,每个样本影像包含 4 个波段,样本影像的空间分辨率为 0.5 ~ 0.8 m,样本示例如图 4 所示,分别是 3 种卫星的真彩色影像、假彩色影像、标签数据。由于样本标签从地理国情监测成果数据而来,许多建筑物标签有图斑综合现象,即房屋建筑区内面积小于 1 600 m²的绿化林地、绿化草地、硬化平地等综合成建筑物类别,如图 5 所示。

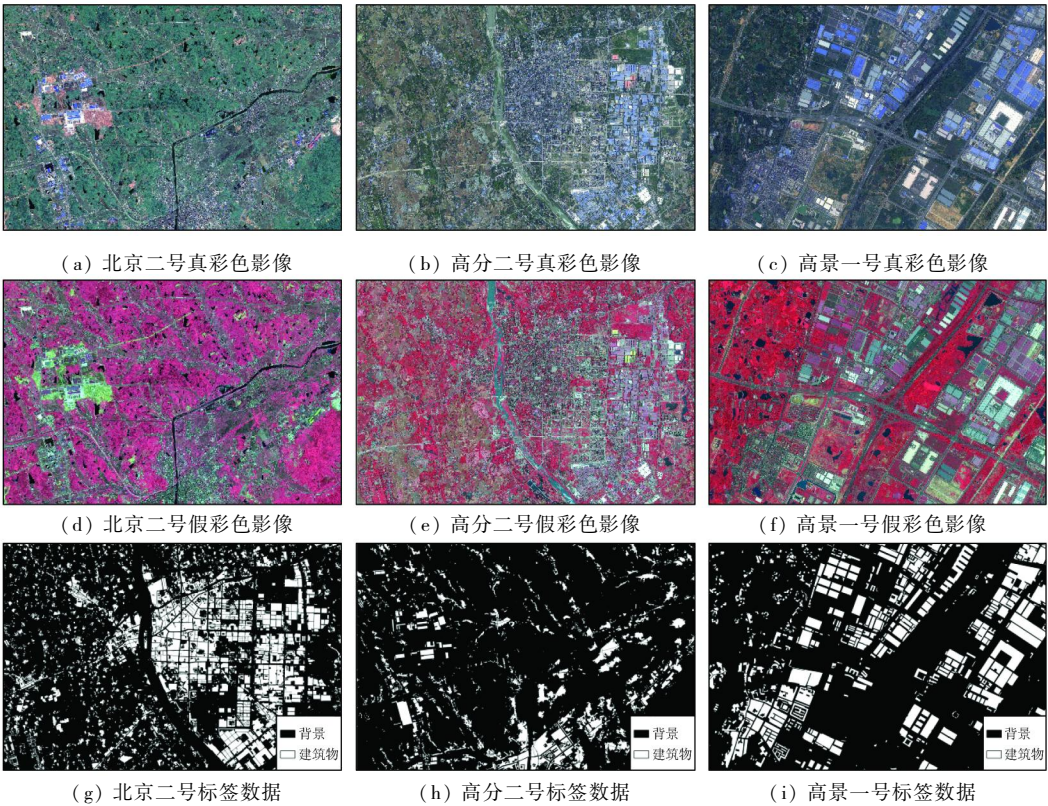


图 4 卫星遥感影像建筑物样本
Fig. 4 Examples of dataset with satellite image and label

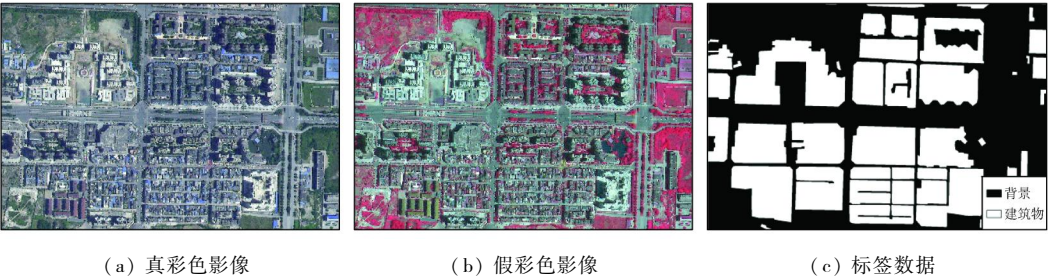


图 5 图斑综合样例
Fig. 5 Examples of path generalization

3 试验与分析

试验基于 TensorFlow 和 Keras 深度学习开源框架,采用 python 语言编程实现,试验操作系统为 Ubuntu 16.04, GPU 显卡选用 NVIDIA GTX1080TI。模型训练时,将训练样本分为训练集和验证集,验证集占总样本的 1/3, Batch_size 设置为 6, 样本图像输入尺寸为 512 像素 × 512 像素, 利用 Adam 算法进行网络优化, 学习率设置为 0.000 1。

精度评估采用基于误差矩阵为核心的分类精度评价方法^[25], 评价指标包括总体精度 (OA) 和 Kappa 系数、F1 分值和 IoU (交并比)^[26], 公式为:

$$F1 = \frac{2precision \cdot recall}{precision + recall}, \tag{13}$$

$$precision = \frac{TP}{TP + FP}, recall = \frac{TP}{TP + FN}, \tag{14}$$

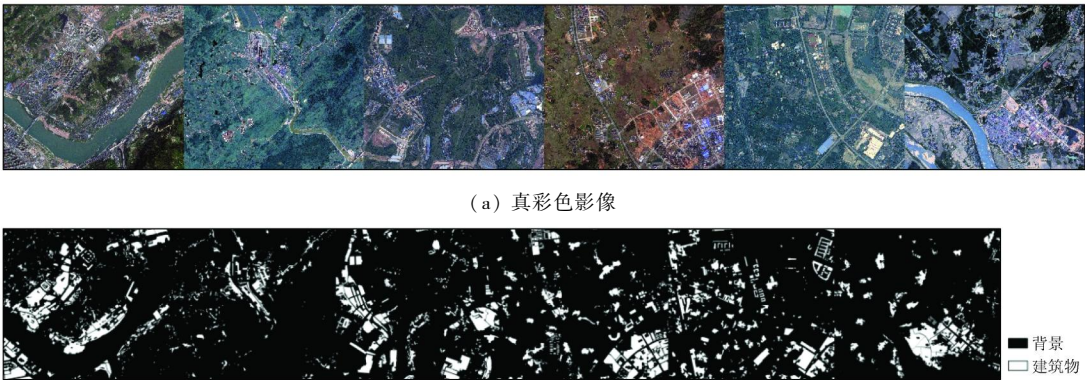
$$IoU = \frac{TP}{TP + FP + FN}, \tag{15}$$

式中: TP 为属于类别 C 被正确分到类别 C 的样本;

FN 为不属于类别 C 的样本被错误分类到类别 C; FP 为不属于类别 C 的样本被正确分类到了类别 C 的其他类。F1 和 IoU 取值为 0 ~ 1 之间, 值越大, 分类精度越好。

3.1 与 WHU 卫星数据集比较

为了分析 SCRS 数据集的性能, 采用 FPN - SENet 分别在 WHU 数据集和 SCRS 数据集上进行训练、测试: ①WHU 卫星数据 II (Satellite dataset II (East Asia)), 包含 2 幅训练数据和 1 幅测试数据, 将训练数据裁切成 512 像素 × 512 像素大小, 2/3 为训练集, 1/3 为验证集, 利用训练后的模型对测试数据进行分类和精度评估; ②SCRS 数据集, 将训练数据裁切成 512 像素 × 512 像素大小的图像, 2/3 为训练集, 1/3 为验证集, 利用训练后的模型对测试数据进行分类和精度评估。其中, SCRS 的测试数据, 分别在不同区域选取高分二号 2 景、北京二号 2 景、高景一号 2 景影像, 从每一景中裁剪出高宽为 5 000 像素 × 5 500 像素建筑物丰富的影像, 拼接成一幅高宽为 5 000 像素 × 33 000 像素的测试数据, 如图 6 所示。



(a) 真彩色影像

(b) 建筑物参考分类结果

图 6 建筑物测试数据

Fig. 6 Test images and ground - truth of building

表 1 给出了 FPN - SENet 分别在 WHU 数据集和 SCRS 数据集上的测试精度, 表明 SCRS 建筑物提取的精度接近目前国际上精度最高的开源卫星影像建筑物数据集 (WHU 数据集), 但是 WHU 数据集的

影像只有 3 个波段, 而 SCRS 数据集的影像有 4 个波段, 可以选择用真彩色影像或假彩色影像, 从表 1 可以看出, 假彩色影像的 F1 值和 IoU 比真彩色影像高 2.9 和 4.0 百分点。

3.2 常用 FCN 模型比较

目前遥感影像的像素级分类几乎都是基于 FCN, 于是本文以 SCRS 数据集的真彩色影像为输入, 对目前较为流行的几种 FCN 模型进行评估, 表 2 给出了 FCN - 8s, Segnet, U - net, PSPNet 和本文方法 (FPN - SENet) 的分类精度, 测试数据仍选择图 5 的影像。从测试结果来看, FPN - SENet 取得了最优的结果, 相对于表现次之的 U - net, 本文方法的 F1 值和 IoU 高出 2.4 和 3.1 百分点, 比 PSPNet

表 1 SCRS 数据集与 WHU 数据集的比较

Tab. 1 Comparison of the SCRS dataset with the WHU dataset

数据集	OA	Kappa	recall	precision	F1/%	IoU/%
WHU	0.995	0.804	0.772	0.845	80.7	67.6
SCRS(真彩色)	0.946	0.751	0.742	0.827	78.2	64.2
SCRS(假彩色)	0.952	0.784	0.778	0.847	81.1	68.2

的结果高出 5.7 和 7.4 百分点。从文献中得知,传统方法的准确率难以超过 50%^[20],而表 2 中语义分割模型的 F1 - score 均超过了 70%,说明深度学习方法在卫星影像建筑物提取中具有很好的应用前景。

表 2 不同网络模型的比较
Tab.2 Comparison of the FPN - SENET with other networks

方法	OA	Kappa	recall	precision	F1/%	IoU/%
FCN - 8s	0.932	0.672	0.639	0.800	71.0	55.1
Segnet	0.934	0.681	0.645	0.809	71.8	56.0
U - net	0.941	0.735	0.707	0.818	75.8	61.1
PSPNet	0.936	0.689	0.645	0.827	72.5	56.8
FPN - SENet	0.946	0.751	0.742	0.827	78.2	64.2

3.3 不同损失函数的比较

为了提升 FPN - SENet 在 SCRS 数据集上提取建筑物的性能,本文尝试用多种损失函数对 SCRS 数据集进行训练和测试,表 3 显示了不同损失函数模型在测试数据的精度,影像输入波段为假彩色 3 波段,L_bce 模型的建筑物提取准确率最高,达到了 84.7%,L_dice 模型的召回率最高,达到了 83.3%,

而 L_bce - dice 能够在准确率和召回率之间取得平衡,保证提取精度最优,F1 分值和 IoU 分别达到了 81.7% 和 69.1%。

表 3 不同损失函数的比较
Tab.3 Comparison of results from models trained with different loss functions

损失函数	OA	Kappa	recall	precision	F1/%	IoU/%
L_bce	0.952	0.784	0.778	0.847	81.1	68.2
L_dice	0.948	0.779	0.833	0.786	80.9	67.9
L_bce - dice	0.952	0.790	0.82	0.814	81.7	69.1

3.4 基于窗函数平滑拼接的效果

本文在整幅测试影像进行分类时,采用了平滑预测的后处理方式,图 7 显示了基于窗函数平滑拼接的效果,可以看出,采用分块直接拼接时,图像块之间存在明显的拼接痕迹,分块边缘处的地物不连续;而基于窗函数平滑拼接的结果中,地物在块与块之间过渡更自然,图斑连续性更好,边界更整齐。采用精度最优的模型,对测试影像的自动提取效果如图 8 所示,可以看出本文方法能够很好地从高分分辨率卫星影像中提取建筑物。

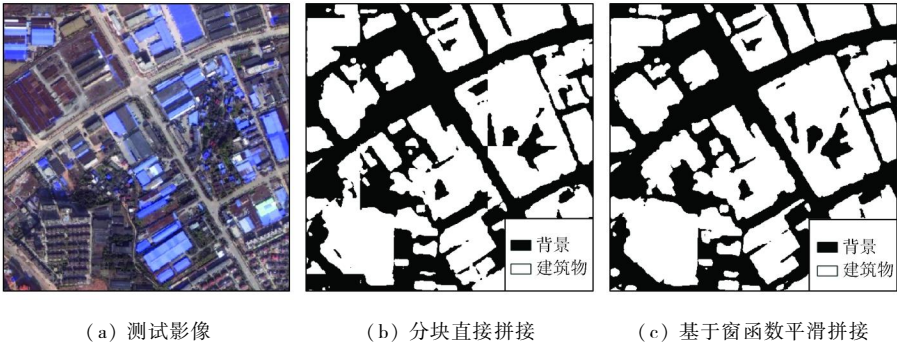


图 7 平滑预测的效果
Fig.7 Results with smooth prediction

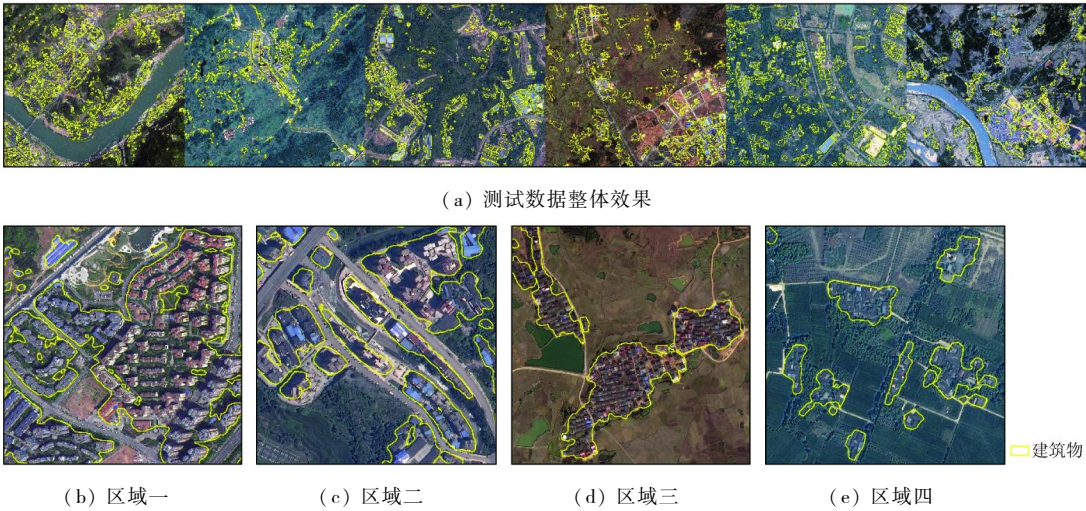


图 8 测试数据提取效果
Fig.8 The classified result of test images

4 结 论

本文基于国产高分卫星影像和地理国情监测成果数据,制作卫星影像建筑物数据集,该方法能够快速制作大规模的像素级标注数据集,且采用一种基于注意力增强的特征金字塔建筑物提取网络对数据集进行测试,主要结论如下:

- 1) 本文制作的卫星影像数据集的建筑物提取精度接近国际先进的开源卫星影像建筑物数据集,且数据集影像为 4 波段。
- 2) 本文提出的网络模型,在建筑物提取上的精度优于其他常用的 FCN 网络。
- 3) 通过改进损失函数、基于窗函数平滑拼接的技术手段,能够进一步提高建筑物提取精度和效果。本文结合国家重大项目需要提出的卫星影像建筑物自动提取方法,能够为深度学习方法在遥感解译领域的深入应用提供借鉴意义。未来将继续研究深度学习方法针对不同数据源、不同时相卫星影像的适用性,并制作更多地物类型、更大规模的遥感影像样本数据,推广深度学习方法在遥感自动解译领域的应用。

参考文献(References):

[1] Blaschke T. Object based image analysis for remote sensing[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2010, 65 (1): 2 - 16.

[2] Yang Y, Newsam S. Geographic image retrieval using local invariant features[J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 51(2): 818 - 832.

[3] Li E, Femiani J, Xu S B, et al. Robust rooftop extraction from visible band images using higher order CRF[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(8): 4483 - 4495.

[4] Melgani F, Bruzzone L. Classification of hyperspectral remote sensing images with support vector machines[J]. IEEE Transactions on Geoscience and Remote Sensing, 2004, 42(8): 1778 - 1790.

[5] Hinton G E, Osindero S, Teh Y. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18: 1527 - 1554.

[6] Lecun Y, Bottou L, Bengio Y, et al. Gradient - based learning applied to document recognition[C]. Proceedings of the IEEE, 1998, 86(11): 2278 - 2324.

[7] Graves A, Liwicki M, Fernandez S, et al. A novel connectionist system for unconstrained handwriting recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31: 855 - 868.

[8] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//International Conference on Neural Information Processing Systems, 2012.

[9] Mikolov T, Deoras A, Kombrink S, et al. Empirical evaluation and combination of advanced language modeling techniques[C]//Flor-

ence: Conference of the International Speech Communication Association, 2011.

[10] Dahl G E, Yu D, Deng L, et al. Context - dependent pre - trained deep neural networks for large - vocabulary speech recognition[J]. IEEE Transactions on Audio, Speech and Language Processing, 2012, 20(1): 30 - 42.

[11] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 2015.

[12] Wu G M, Shao X W, Guo Z L, et al. Automatic building segmentation of aerial imagery using multi - constraint fully convolutional networks[J]. Remote Sensing, 2018, 10(3): 407 - 424.

[13] Zhang W K, Huang H, Schmitz M, et al. Effective fusion of multi - modal remote sensing data in a fully convolutional network for semantic labeling[J]. Remote Sensing, 2018, 10(52): 1 - 14.

[14] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder - decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(12): 2481 - 2495.

[15] Ronneberger O, Fischer P, Brox T. U - Net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer - Assisted Intervention, 2015.

[16] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[17] 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 & Machine Intelligence, 2018, 40: 834 - 848.

[18] Mnih V. Machine learning for aerial image labeling[M]. Toronto: Toronto University of Toronto, 2013.

[19] Maggiori E, Tarabalka Y, Charpiat G, et al. Can semantic labeling methods generalize to any city? The inria aerial image labeling benchmark[C]//IEEE International Geoscience and Remote Sensing Symposium(IGARSS), Fort Worth, United States, 2017.

[20] 季顺平, 魏世清. 遥感影像建筑物提取的卷积神经网络与开源数据集方法[J]. 测绘学报, 2019, 48(4): 448 - 459.

Ji S P, Wei S Q. Building extraction via convolutional neural networks from an open remote sensing building dataset[J]. Acta Geodaetica et Cartographica Sinica, 2019, 48(4): 448 - 459.

[21] Lin S Y, Doll'ar P, Girshick R, et al. Feature pyramid networks for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[22] Hu J, Shen L, Sun G. Squeeze - and - excitation networks[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[23] 杨建宇, 周振旭, 杜贞容, 等. 基于 segnet 语义模型的高分辨率遥感影像农村建设用地提取[J]. 农业工程学报, 2019, 35(5): 251 - 258.

Yang J Y, Zhou Z X, Du Z R, et al. Rural construction land extraction from high spatial resolution remote sensing image based on segnet semantic segmentation model[J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(5): 251 - 258.

[24] 国务院第一次全国地理国情普查领导小组办公室. 地理国情普查内容与指标[M]. 北京: 测绘出版社, 2013.

The First National Geographic National Conditions Census Leading Group Office of State Council. Geographic national conditions census content and indicator [M]. Beijing: Surveying and Mapping Press, 2013.

[25] Congalton R G. A review of assessing the accuracy of classifications of remotely sensed data [J]. Remote Sensing of Environment, 1991, 37 (1) : 35 - 46.

[26] Maggiori E, Tarabalka Y, Charpiat G, et al. Convolutional neural networks for large - scale remote - sensing image classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55: 645 - 657.

Building extraction using high - resolution satellite imagery based on an attention enhanced full convolution neural network

GUO Wen¹, ZHANG Qiao²

- (1. The Third Institute of Photogrammetry and Remote Sensing, Ministry of Natural Resources, Chengdu 610100, China;
2. School of Geoscience and Technology, Southwest Petroleum University, Chengdu 610500, China)

Abstract: Automatic extraction of buildings from satellite remote sensing images has a wide range of applications in the development of economy and society. Due to the influence of mutual occlusion, illumination, background environment and other factors in satellite remote sensing images, it is difficult for traditional methods to achieve high - precision building extraction. This paper proposes an attention enhanced feature pyramid network (FPN - SENet) and constructs a large - scale pixel - wise building dataset (SCRS dataset) by using multi - source high - resolution satellite images and vector data to realize the automatic extraction of buildings from multi - source satellite images, and compares it with the other full convolution neural networks. The results show that the accuracy of building extracted from SCRS dataset is close to the world's leading open source satellite image dataset, and the accuracy of Pseudo color data is higher than that of true color data. The accuracy of FPN - SENet is better than that of other full convolution neural networks. The extraction of building can also be improved by using the sum of cross entropy and Dice coefficient as the loss function. The overall accuracy of the best classification model is 95.2%, Kappa coefficient is 79.0%, and F1 - score and IoU are 81.7% and 69.1% respectively. This study can provide a reference for building automatic extraction from high - resolution satellite images.

Keywords: Chinese high - resolution satellite imagery; buildings; semantic segmentation; attention enhancement (责任编辑: 李 瑜)