引文格式：乔志勇. 基于改进DeepLabV3+的高分辨率遥感影像建筑物提取[J]. 地理空间信息, 2024, 22(12): 69-73. doi:10. 3969/j.issn. 1672-4623.2024.12.016

基于改进DeepLabV3+的高分辨率遥感影像建筑物提取

乔志勇[1]
（1. 厦门精图信息技术有限公司，福建 厦门 361008）

摘　要：针对管道传统的高后果区识别方式不但工作效率低，而且很难进行定量判断的问题。提出了一种基于改进DeepLabV3+的遥感影像建筑物提取算法，采用基于全卷积神经网络识别算法，用于高分辨率遥感影像语义分割，以提取出兴趣区域（AOI）的像素点，并采用形态学方法进行后处理，以获得最终的提取结果并应用于高后果区识别。实验结果表明：该文所提算法对高分辨率遥感影像中建筑物识别效果具有较大增强，且该算法也适用于油气管道高后果区的识别工作。此外还可结合其他数据以更精确地实现高后果区的识别。

关键词：遥感影像识别；全卷积神经网络；建筑物识别；高后果区
中图分类号：P237　　文献标志码：B　　文章编号：1672-4623（2024）12-0069-05

Building Extraction from High-resolution Remote Sensing Images Based on Improved DeepLabV3+

QIAO Zhiyong[1]
(1. Xiamen Kingtop Information Technology Co., Ltd., Xiamen 361008, China)

Abstract: According to the issue of low efficiency and difficulty in quantitative judgment in traditional high consequence area identification methods for pipelines, we proposed a remote sensing image building extraction algorithm based on improved DeepLabV3+. In order to extract pixels of the area of interest (AOI), we used recognition algorithm based on a fully convolutional neural network to conduct semantic segmentation of high-resolution remote sensing images. Then, we employed morphological method to conduct post-processing, and obtained the final extraction results, which were applied to high consequence area identification. Experimental results demonstrate that the algorithm proposed in this paper significantly improves the recognition of buildings in high-resolution remote sensing images. Additionally, this algorithm is suitable for identifying high consequence areas in oil and gas pipelines. Furthermore, it can be combined with other data for a more accurate identification of high consequence areas.

Key words: remote sensing image recognition, fully convolutional neural network, building recognition, high consequence area

高空间分辨率遥感影像已成为地理信息重要的来源之一[1]。高分影像具有丰富的地物细节信息、清晰的几何结构和纹理特征[2-3]。建筑物遥感图像提取一直是遥感信息提取的热点之一。传统的基于边缘检测[4]、建筑物指数[5]、辅助信息[6]的提取方法特征表示能力有限，又受困于分类器性能，难以在复杂场景进行应用，而且存在精度低的问题。

为了提高高分辨率遥感影像中建筑物提取的精确度和稳健性[7-10]，本文提出了一种基于改进DeepLabV3+的算法，用于对高分辨率遥感影像进行语义分割，以提取出兴趣区域（AOI）的像素点，并采用形态学方法进行后处理，以获得最终的提取结果并应用于高后果区识别。

1　方　法

1.1　方法流程

本文采用编码器-解码器架构的全卷积神经网络（FCN）对高分辨率遥感影像进行语义分割。具体的方法流程详见图1。

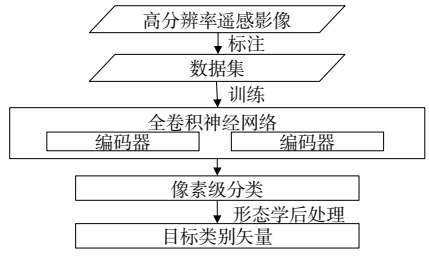


图1　方法流程

收稿日期：2023-09-20。
项目来源：厦门市科技计划资助项目（3502Z20231038）。
作者简介：乔志勇（1981—），本科，高级工程师，主要研究方向为地理信息系统应用和软件设计开发，E-mail：qzy@kingtopinfo.com。

1.2 预处理

本文采用了以下数据增强策略以提升模型的泛化能力：形态学变换，包括翻转、旋转和缩放操作；像素值调整，涉及亮度变换和对比度增强。此外，为了有效缓解类别不平衡问题，本研究还实施了复制粘贴增强技术。

1.3 Reg-DeepLabV3+网络总体结构

模型整体结构采用DeepLabV3+，DeepLabv3+采用了空洞卷积和ASPP（空间金字塔池化）模块，能够更好地捕捉物体的上下文信息和多尺度特征，DeepLaby3+还具有多尺度特征融合和空洞卷积，能够更好地处理图像中的细节和复杂情况，从而提高了语义分割的精度[11-12]。

1.3.1 编码器网络结构

RegNet系列模型的结构如图2。

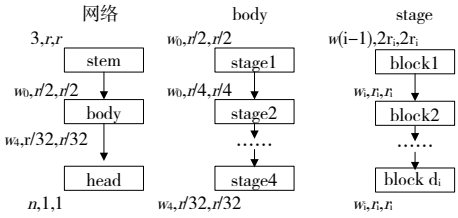


图2 模型结构

对于输入x，Batch Normalization可以将其标准化为：

$$\frac{x-E[x]}{\mathrm{sqrt}\left(\mathrm{Var}[x]+\varepsilon\right)} \tag{1}$$

式中，$E[x]$为x的均值；$\mathrm{Var}[x]$为x的方差；ε为一个非常小的数，用于防止分母为0的情况。然后，对上述标准化后的值进行缩放和平移，得到新的激活值y：

$$y=\gamma\times\frac{x-E[x]}{\mathrm{sqrt}\left(\mathrm{Var}[x]+\varepsilon\right)}+\beta \tag{2}$$

式中，γ和β为可学习的参数，用于缩放和平移。最后，将y作为下一层的输入，继续进行正向传播和反向传播。

如图3所示，网络的基本构建块（block）设计包含一个主分支和一个捷径分支（shortcut branch）。在主分支中，结构通常由3个卷积层依次排列而成：首先是1×1卷积层，随后是3×3的组卷积层，再次是1×1卷积层。为了优化网络性能，每一卷积层之间均嵌入批量标准化（Batch Normalization）和激活函数（例如ReLU）。捷径分支的处理策略取决于卷积的步幅（stride）值。具体而言，当stride等于1时，捷径分支不执行任何操作，直接将输入数据传递至下一层；而当stride等于2时，捷径分支通过一个1×1卷积层进行下采样，以减少数据的空间维度。

在此，r表示特征矩阵的分辨率，即其高度和宽度。

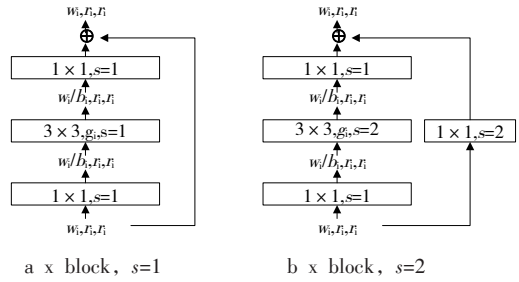


图3 Block结构

当步幅s为1时，输入和输出的分辨率保持不变；而当s为2时，输出的分辨率是输入的一半。w代表特征矩阵的通道数（channel）。值得注意的是，当$s=2$时，输入的通道数为w_{i-1}，而输出的通道数为w_i，因此通道数会发生变化。G为组卷积中每个组的宽度，而b为瓶颈比例（bottleneck ratio），即输出特征矩阵的通道数将缩减至输入特征矩阵通道数的$1/b$。

head是网络的顶部部分，它接受来自body的特征，并将其映射到最终的输出。此处head由一个全局平均池化层和全连接层构成。

在RegNet系列模型中存在2种结构：RegNet-X与RegNet-Y，它们的不同之处是后者的SE（Squeeze-and-Excitation）模块连接在block的Group Conv之后。

SE（squeeze-and-excitation）是一种用于增强深度神经网络性能的注意力机制。它通过学习每个通道的权重来动态调整特征图中的特征表示，以提高模型对重要特征的关注，从而改善模型在各种视觉任务中的性能。SE模块通常由2个主要步骤组成：Squeeze（压缩）：在这一步骤中，SE模块通过全局平均池化（Global Average Pooling）操作，将每个通道内的特征图转化为一个单一的数值。这一数值可以被看作是该通道内特征的全局统计信息，用于表示该通道的重要性。全局平均池化操作将特征图的空间维度降低为一个标量。Excitation（激发）：在这一步骤中，SE模块通过一系列全连接层（通常包括一个或多个全连接层）来学习每个通道的权重或缩放因子。这些权重用于重新加权特征图中的每个通道，以强调重要的通道并抑制不重要的通道。这样，模型可以自适应地关注对特定任务更有信息量的特征。

如图4所示，在本文中提出的SE（Squeeze-and-Excitation）模块构成包括3个主要部分：全局平均池化层、第一个全连接层（FC1）以及第二个全连接层（FC2）。在RegNet架构中，FC1的节点数量被设定为输入特征矩阵通道数的1/4，这与组卷积输出特征矩阵通道数的1/4不同。此外，FC1层采用了ReLU作为

激活函数。至于FC2，其节点数量与组卷积输出特征矩阵的通道数保持一致，并且采用Sigmoid激活函数以实现特征的重新校准。

本算法选择RegNet-Y-040作为编码器，其结构如图4所示：其4个Stage中分别包含2、6、12、2个block。

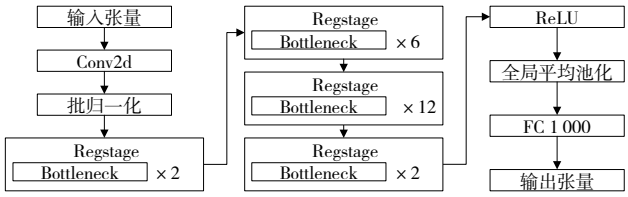


图4　RegNet-Y-040结构

在网络的每个stage中，首个构建块（block）的主分支所采用的33卷积核的步幅（stride）设定为2，以实现特征图的空间降维。相应地，其捷径分支包含一个11卷积核，其步幅同样为2，以便与主分支的输出维度匹配。在其余卷积层中，步幅均维持为1，以保持特征图的空间尺寸。对于同一阶段内的其余构建块，所有卷积层的步幅均为1，并且捷径分支直接将输入数据传递至下一构建块，不进行额外的处理。

在算法中RegNet-Y-040作为编码器，输出特征图（feature map），故不需要全局平均池化层与全连接层（FC 1000），因此本算法中删去这两层。

1.3.2　解码器网络结构

解码器部分采用了一种基于空间金字塔池化（ASPP）的结构，用于将编码器的特征图上采样到输入图像大小，并生成像素级别的预测结果。ASPP结构包括多个并行的卷积分支，每个分支采用不同的空洞卷积核大小，以便捕捉不同大小的上下文信息。另外，ASPP结构还包括一个全局池化分支，用于处理整个图像的信息。最后，ASPP结构将所有分支的输出进行级联和处理，生成最终的分割结果。在标准的DeepLabV3+网络架构中，编码器部分采用Xception模型以生成特征图，随后这些特征图被送入解码器中的ASPP（atrous spatial pyramid pooling）模块进行处理。得益于DeepLabV3+采用的编码-解码结构，若需更换编码器，仅需保留ASPP模块及其后续结构，即可实现网络架构的灵活调整。

1.3.3　基于RegNet-Y-040修改DeepLabV3+

本文对RegNet-Y-040模型进行了修改，移除了其全局平均池化层和全连接层，并将其输出直接连接至DeepLabV3+模型的ASPP模块。鉴于在DeepLabV3+框架中，ASPP模块的输出需经过二倍上采样以与Xception模型初始块（init block）的特征图进行拼接（concatenation），因此，本文采取了以下策略：将RegNet-Y-040模型的第一个构建块（block）的输出传递至一个1×1卷积层，随后通过一个批归一化层和一个ReLU激活函数进行处理。处理后的特征图再与经过二倍上采样操作的ASPP模块输出进行拼接。

1.4　损失函数

本文所采用的损失函数为加权交叉熵损失与Dice损失的联合形式。选择Dice损失的原因在于，训练数据集中可能存在类别不平衡的现象，其中目标类别所占的像素比例相对较小，这可能导致模型学习过程中偏向于背景类别。Dice损失函数的应用旨在缓解这一问题，通过优化目标类别的预测性能，从而提高模型对于小面积目标的识别能力。

$$\mathrm{Dice}=\frac{2|X\cap Y|}{|X|+|Y|} \tag{3}$$

Dice损失公式如式（4），其中X为预测影像，Y为真实影像。

$$\mathrm{loss}=1-\frac{2|X\cap Y|}{|X|+|Y|} \tag{4}$$

鉴于仅采用Dice损失可能导致训练过程的稳定性降低，故引入带权重交叉熵损失。

$$L=-\sum_{C=1}^{M}w_c\boldsymbol{y}_c\log(p_c) \tag{5}$$

式中，w_c为权重；M为类别，$\boldsymbol{y}_C$为向量；p_c为概率。

$$w_c=\frac{N-N_C}{N} \tag{6}$$

式中，N为像素总数；N_C为C所占像素总数。

在实验的具体操作中，本文对目标对象的边缘像素赋予了较高的权重，以此引导训练过程更加关注边缘像素，进而提升边缘区域的分割精度。

因此完整的损失函数为：

$$L=1-\frac{2|X\cap Y|}{|X|+|Y|}-\sum_{C=1}^{M}w_c\boldsymbol{y}_c\log(p_c) \tag{7}$$

1.5　训练过程

优化器选择带momentum的SGD，设置学习率为$1*10^{-3}$，权值衰减（weight decay）为$1*10^{-3}$。

学习率的调整策略采用了结合预热阶段（warm-up）的余弦退火（cosine annealing）方法。此举旨在增强模型训练初期的稳定性，并防止模型在整个训练周期中陷入局部最小值。余弦退火学习率调整的公式表述如下：

$$\eta=\eta_{\min}+\frac{1}{2}(\eta_{\max}-\eta_{\min})\left(1+\cos\left(\frac{T_{\mathrm{cur}}}{T_i}\pi\right)\right) \tag{8}$$

式中，η为当前的学习率；$\eta_{\min}$为最小学习率；$\eta_{\max}$为初始学习率；T_{cur}为上一次重启后所经历的训练轮

数（epoch），T_i为2次重启间的训练轮数（epoch）。当$T_{cur}=T_i$，则$\eta=\eta_{min}$，当重启后$T_{cur}=0$，$\eta=\eta_{max}$。

设置进行5 epochs的预热之后采用余弦退火学习率调整，余弦退火学习率调整的初始学习率为10^{-5}，$T_0=7$，$T_{mult}=2$，其中T_0为第一次重启时的epoch，T_{mult}为每次重启后的增加系数T_i。

1.6 实验与结果

为了对网络性能进行定量评估，使用6个指标来评估分类结果，它们分别是准确率（Accuracy）、召回率（Recall）、Kappa系数、重叠度（Intersection over Union，IoU）、精确率（Precision）和F1分数（F1-score）。表示正确分类的像素数量与总像素数量之间的比例。具体而言，对于每个像素，模型的分割结果与真实标签一一对应，并且该像素的类别标签完全相同，即被视为正确分类的像素。

在WHU-Building数据集上进行测试。数据来自新西兰土地信息服务网站。大约包含22 000栋独立建筑物。图像原始分辨率为0.075 m，被降采样到0.3 m的分辨率，并将它们裁剪成8 189张512×512像素图片。样本分为3个部分：训练集（4 736张、130 500栋建筑）、验证集（1 036张、14 500栋建筑）和测试集（2 416张、42 000栋建筑）。

在WHU数据集上进行建筑物检测实验，训练所用PC配置为CPU：I5-12600 K，显卡：Nvidia RTX3060、32 G RAM，模型代码使用pytorch编写。使用了DeepLabV3+和Reg-DeepLabV3+ 2种模型。并在图7中展示了实验结果。在这2种模型中，Reg-DeepLabV3+表现最为稳定，且能够更精准地保留细节信息。相比之下，DeepLabV3+存在一些明显的不足，例如较难提取出影像上尺寸差别较大的建筑物，边缘角点提取结果模糊等。这些问题Reg-DeepLabV3+网络中得到了一定的缓解，如图5所示，该模型能够精确提取建筑物的细节信息。这些结果表明Reg-DeepLabV3+是一种很有潜力的建筑物检测模型，并且可以进一步发展以解决现有模型的局限性。

1.7 参数量对比

两种模型参数量见表1，通过参数量对比可知，Reg-DeepLabV3+与DeepLabV3+网络参数量相似。

2 应用实例

本文基于地理信息系统开发平台KingMap实现建筑物识别算法，并将方法应用于高后果区识别中，以验证其对遥感影像中目标地物的准确度和效率。其

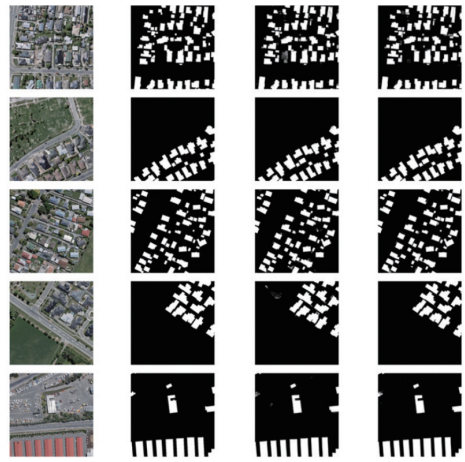


图5 实验结果

表1 模型对比

模型名称	DeepLabV3+	Reg-DeepLabV3+
参数量	22 437 457	21 466 345

方法包括通过建筑物识别算法从遥感影像图中获取建筑图斑；利用缓冲空间分析生成基于管道中心线的输油管道综合影响区域；通过缓冲区分析，获取影响区域内的建筑图斑；根据获取的建筑图斑，通过算法计算实现高后果区的调识别。实验以GB 32167—2015的规范为标准，创建缓冲区并进行分析识别：沿管道中心线两侧各200 m范围和潜在影响半径分别创建缓冲区；使用建筑物识别算法获取建筑图斑矢量；结合高后果区和潜在影响半径要求进行缓冲分析，获取高后果区与潜在影响区域内的建筑图斑；通过算法计算实现高后果区的识别；将识别结果输出为矢量数据。

实验以贵州省管线中某段长度为21 245.9 m，管径为700 mm，材质为管线钢的管线数据为对象进行高后果区识别，底图采用0.8 m空间分辨率的遥感影像图。通过原始影像（图6）和缓冲分析提取高后果区建筑物图斑矢量（图7）。



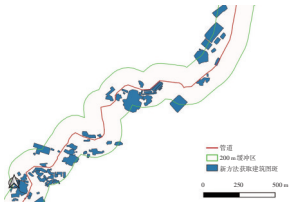


图6 贵州某管段原始影像 图7 建筑物图斑提取结果图

生成的建筑物图斑矢量与传统采集的建筑图斑进行对比（图8和图9）。

据实验结果，传统采集的建筑图斑数量为198个，耗时1~2 h，而使用建筑物识别算法只需1 min左右即可生成建筑物图斑矢量，图斑数量为190个，准确率达到96%。其中不同的原因在于部分建筑过小且在山

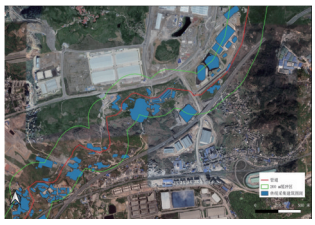


图8　传统采集的建筑图斑

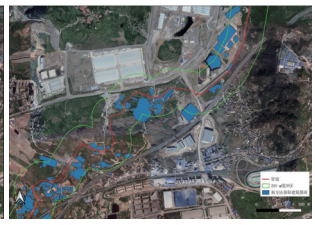


图9　新方法获取的建筑图斑

区中周围有树木遮挡。实验结果表明，建筑物识别算法能够有效地基于高分辨率遥感影像生成建筑图斑矢量，从而提高识别效率。

通过对识别出的建筑图斑矢量结合影响数据进行叠加分析，实现对高后果区的识别。程序将建筑图斑作为带有户数属性的点来处理。这些点在管线周围的分布没有特定形状，只是散布在一定范围内。通过沿管线计算确定建筑（户数）密集区域，可以实现高后果区的识别。实验采用了基于密度的DBSCAN聚类算法，该算法考虑密度而不是距离来计算簇和簇边界，能够有效处理噪声数据。其基本思路如下：以建筑密集区域中的某一建筑为圆心，选择合适半径画圆，即规定了圆的半径及圆内最少包含的建筑密度。若圆内建筑密度大于等于预先指定的值，那么这个圆圈的圆心就转移至该圆圈内的其他建筑，继续画圆去计算，直至圆内所包含的建筑密度少于预先指定的值，聚类终止。基于密度的DBSCAN聚类算法适用于在管线周围建筑分布区域中，寻找高密度分布的区域，并不受建筑分布的形状影响。

实验根据量化后的管道高后果区识别准则，将基于密度的DBSCAN聚类算法应用到高后果区识别中。以管线为基本单位，调用识别出的建筑图斑，半径设置成100 m，数量设置为10，计算出建筑图斑簇的区域，满足识别标准时将该区域划分为高后果区。利用上述方法进行识别，共识别出高后果区4段，其总长为13 886 m。比较程序识别和人工识别的差别，程序基于遥感影像数据高后果区识别的结果于与人工识别基本符合。

相较于人工划分，程序划分在一定程度上降低了人为划分的误差因素，有效地减少了工作人员在高后果区识别过程中产生的各种人为误差，包括但不限于识别误差、打分误差及等级划分误差等，为高度准确且迅速的高后果区划分提供了可靠的参考依据。实验证明，通过建筑物识别算法从遥感影像图中获取建筑图斑能够有效识别高后果区，并能大幅度降低人工成本和管理成本。

3　结　论

本文提出了一种基于改进DeepLabV3+的高分辨率遥感影像建筑物提取方法的算法，实验结果数据证明了该算法有效提升了高分辨率遥感影像中建筑物识别效果。通过结合管道数据进行实验，可以推断出可能受到高后果区影响的区域。此外，该算法还有进一步推广应用的潜力。

参考文献

[1] Cheriyadat A M. Unsupervised Feature Learning for Aerial Scene Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2013,52(1):439-451

[2] 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, 2016,55(2):645-657

[3] Luo B, Zhang L. Robust Autodual Morphological Profiles for the Classification of High-resolution Satellite Images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2013,52(2):1 451-1 462

[4] Canny J. A Computational Approach to Edge Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986,8(6):679-698

[5] Rüther H, Martine H M, Mtalo E G. Application of Snakes and Dynamic Programming Optimisation Technique in Modeling of Buildings in Informal Settlement Areas[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2002, 56(4): 269-282

[6] 徐涵秋. 基于压缩数据维的城市建筑用地遥感信息提取[J]. 中国图象图形学报,2005,10(2):223-229

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

[8] 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

[9] Xu Y, Wu L, Xie Z, et al. Building Extraction in very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters[J]. Remote Sensing,2018,10(1):144

[10] Shao Z, Tang P, Wang Z, et al. BRRNet: A Fully Convolutional Neural Network for Automatic Building Extraction from High-resolution Remote Sensing Images[J]. Remote Sensing, 2020,12(6):1 050

[11] Guo M, Liu H, Xu Y, et al. Building Extraction Based on U-Net with an Attention Block and Multiple Losses[J]. Remote Sensing,2020,12(9):1 400

[12] 李传林,黄风华,胡威,等. 基于Res_AttentionUnet的高分辨率遥感影像建筑物提取方法[J]. 地球信息科学学报, 2021,23(12):2 232-2 243