

# 改进 SegNet+CRF 高分辨率遥感影像 建筑物提取方法

赵昊罡<sup>1</sup>, 崔红霞<sup>1</sup>, 张芳菲<sup>2</sup>, 顾海燕<sup>3</sup>, 穆潇莹<sup>1</sup>

(1. 渤海大学 信息科学技术学院, 辽宁 锦州 121003;

2. 中电太极(集团)有限公司, 北京 100083; 3. 中国测绘科学研究院, 北京 100086)

**摘要:** 将传统的语义分割 SegNet 网络用于高分辨率遥感影像的建筑物提取时, 分割的建筑物存在边界模糊、精度较低、错检漏检等问题; 为了解决上述问题, 提出一种改进 SegNet 网络+CRF 语义分割方法; 编码阶段的最低分辨率层引入空洞金字塔池化模型 (ASPP, atrous spatial pyramid pooling), 通过并行的空洞卷积操作扩大特征提取的感受野; 解码阶段构建特征金字塔 (FPN, feature pyramid networks) 实现特征多尺度融合, 弥补上采样过程中丢失的特征信息; 最后, 预测图像送入全连接条件随机场模型 (CRF, fully connected/dense CRF) 进行后处理, 优化提取的建筑物边缘; 实验表明, 相较于原 SegNet 网络, 改进方法的建筑物提取像素精度、召回率、平均交并比分别提高了 0.48%、1.29%、2.36%。

**关键词:** 语义分割; 空洞金字塔池化模型; 特征金字塔; 全连接条件随机场; 迁移学习

## Building Extraction from High-resolution Remote Sensing Images Based on Improved SegNet and CRF Method

ZHAO Haogang<sup>1</sup>, CUI Hongxia<sup>1</sup>, ZHANG Fangfei<sup>2</sup>, GU Haiyan<sup>3</sup>, MU Xiaoying<sup>1</sup>

(1. College of Information Science and Technology, Bohai University, Jinzhou 121003, China;

2. CLP Taiji (Group) Co., Ltd., Beijing 100083, China;

3. Chinese Academy of Surveying and Mapping, Beijing 100086, China)

**Abstract:** When traditional semantic segmentation SegNet networks are used to extract buildings from high-resolution remote sensing images, there are the problems of fuzzy boundary, low accuracy, error detection and missing detection. In order to solve above problems, an improved SegNet and CRF semantic segmentation method is proposed. An Atrous Spatial Pyramid Pooling (ASPP) model is introduced to the lowest resolution layer in the encoding stage, parallel multiple hole convolutions are used to extract the feature maps of different receptive fields of an image with the expansion rates. In the decoding stage, Feature Pyramid Networks (FPNs) are built to realize multi-scale feature fusion, and compensate the lost feature information during sampling. Finally, based on the fully connected conditional random field (CRF) mode, the prediction images are reprocessed to optimize the building edges. Experimental results show that the building extraction accuracy of the improved model is higher than that of the original SegNet network, the average pixel accuracy, recall and average cross-over ratio are improved by 0.48%, 1.29% and 2.36% respectively. The improved method is feasible, and can be extended to similar networks.

**Keywords:** semantic segmentation; atrous spatial pyramid pooling; feature pyramid networks; fully connected conditional random field; transfer learning

## 0 引言

高分辨率遥感影像拥有丰富的地物信息, 而建筑物一直都是其主要的地物目标, 从高分辨率遥感影像中提取建筑物对于开展遥感领域各项任务都具有重要的意义<sup>[1]</sup>。由于建筑物几何形状、建筑物周边环境以及高分辨率遥感成像角度、分辨率等多种因素的复杂性, 通过传统方法进行

建筑物的提取不仅效率低下并且精度较低。传统的语义分割方法没有训练过程, 分割效果并不理想<sup>[2]</sup>。计算机视觉领域技术在图像识别、目标检测等方面的快速应用和发展, 深度学习深度卷积神经网络在语义分割方面的快速发展, 都为遥感影像中自动化、高精度提取建筑物带来了机遇, 其效果远远优于传统方法<sup>[3]</sup>。

卷积神经网络 (CNN, convolutional neural networks)

收稿日期: 2022-10-28; 修回日期: 2022-12-08。

基金项目: 自然资源部测绘科学与地球空间信息技术重点实验室项目(2020-2-4); 辽宁省教育厅重点攻关项目(LZ2020004)。

作者简介: 赵昊罡(1995-), 男, 湖北武汉人, 硕士研究生, 在读硕士, 主要从事深度学习图像语义分割方向的研究。

通讯作者: 崔红霞(1969-), 山东聊城人, 女, 博士研究生, 教授, 主要从事低空遥感等遥感影像智能图像处理工作方向的研究。

引用格式: 赵昊罡, 崔红霞, 张芳菲, 等. 改进 SegNet+CRF 高分辨率遥感影像建筑物提取方法[J]. 计算机测量与控制, 2023, 31(7): 177-

在处理图像任务上具有强大的特征提取能力，是传统方法所无法比拟的，为各类图像的语义分割任务提供了自动化、高精度的解决方案，也是本文所选择的解决方案。卷积神经网络具有极其强大的泛化能力，其成功依赖于两个固有的归纳偏置，即平移不变性与局部相关性，通过固定大小的卷积核可以从图片中捕获到许多空域和通道间的信息，并且可以通过共享卷积核来提取图像特征，这大大减少了模型的计算复杂度并提升了效率，近年来卷积神经网络层出不穷，如 AlexNet<sup>[4]</sup>、VGG<sup>[5]</sup>、GoogLeNet<sup>[6]</sup>、ResNet<sup>[7]</sup> 等都具有强有力的捕获丰富空间特征和多尺度信息的能力<sup>[8]</sup>。给语义分割领域带来了无穷的前景，许多基于深度卷积网络的语义分割网络相继发表：Jonathan Long 等人<sup>[9]</sup>提出了全卷积神经网络，通过去掉全连接层适应不同尺寸的图片，通过跳级连接将不同池化层的结果上采样并融合，该网络结构的参数量大，训练效率低并且分割精度低；Ronneberger 等人提出<sup>[10]</sup>U-net 网络，通过使用跳跃连接实现编码部分的特征拼接，上采样的过程中减少了信息的丢失，此网络多用于医学图像分割领域；Badrinarayanan 等人<sup>[11]</sup>提出 SegNet 网络，引入了最大池化索引进行上采样，该网络的性能相较 U-net 没有过多的提升，但是大大的减少了训练时间与内存的消耗；Hengshuang Zhao 等人提出了 PSPnet<sup>[12]</sup>，通过引入金字塔池化模块（PPM，pyramid pooling module）融合多尺度特征细节，但其不足是会丢失目标边界细节；Liang-Chieh Chen 等人提出了 DeepLab 系列网络<sup>[13-16]</sup>，将条件随机场融入网络结构并参与训练，以及以空洞卷积为基础构造空洞金字塔池化模型（ASPP，atrous spatial pyramid pooling），通过提升特征图的感受野获取更多的图像细节，提高了物体边界分割的精准度，可是模型计算成本较高，并且像素位置信息的丢失会影响特征图的局部一致性。

相较于低、中分辨率遥感影像，高分辨率影像的几何结构和纹理特征更加丰富。然而，丰富的信息导致遥感影像“同谱异物”和“同物异谱”现象更为明显。由于卷积神经网络的优越性，近些年来许多学者将深度学习语义分割网络应用于遥感图像的语义分割：卢彻<sup>[17]</sup>等人为解决高分辨率遥感影像中存在同谱异物的问题，利用残差网络思想改进 U-net 网络；薛明<sup>[18]</sup>等人运用残差网络思想对 SegNet 网络进行修改，解决了梯度消失问题；李国清<sup>[19]</sup>等人在 U-net 网络加入多尺度损失函数、跳跃连接和注意力机制以提高网络精度；Ziran Ye<sup>[20]</sup>等人在 FCN 的基础上引入注意力机制，通过加权来融合不同阶段的特征。

建筑物是高分辨率遥感影像中最重要的地物，对该类地物的信息提取必须同时关注提取精度和边缘细节。然而，传统的语义分割网络在地物提取的识别精度以及边缘细节处理方面，还缺少普适性、实用性的方法。本文提出改进的 SegNet 语义分割网络并用于建筑物自动提取，以具有较强迁移性的 VGG16 作为网络的主干部分，引入 ASPP 模型加强编码阶段对建筑物特征的提取，解码阶段构建特征金

字塔网络（FPN，feature pyramid networks）优化上采样结构，最后使用全连接条件随机场（CRF，fully connected/dense crf）进行建筑物分割图像的后处理，对网络分割图像进行进一步的边界优化。

## 1 传统语义分割网络

### 1.1 SegNet 语义分割网络

SegNet 网络结构如图 1 所示，在 FCN 的基础上加入了解码部分，形成了 Encode（编码）-Decode（解码）结构，实现对图像端到端的像素级分类。本文编码部分使用 VGG16 作为编码器，去掉了其中的全连接层，使通过 VGG16 得到的特征图直接和解码器相连接，解码部分对特征图进行上采样得到原始图片大小，最后利用 Softmax 层对逐个像素进行预测。

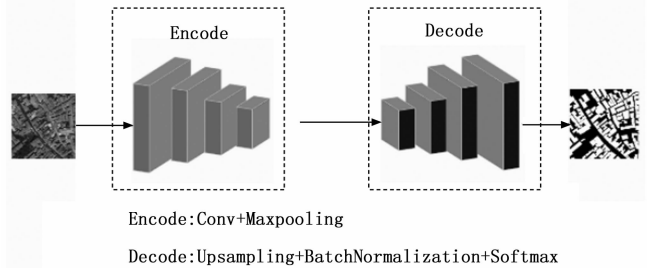


图 1 SegNet 网络结构图

SegNet 的创新之处在于编码部分的最大池化操作会记录像素点的坐标即“最大池化索引”，在解码部分则会使用到这些坐标值来进行上采样，在保留图像特征细节的同时增大图像的尺寸<sup>[21]</sup>，最大池化索引的上采样过程如图 2 所示，在上采样过程中将像素点还原至其记录的索引位置，剩余位置则用 0 进行补充，这一过程无需参数的学习，大大的减少了网络的计算量，提高了训练效率也提高了网络预测的精度。

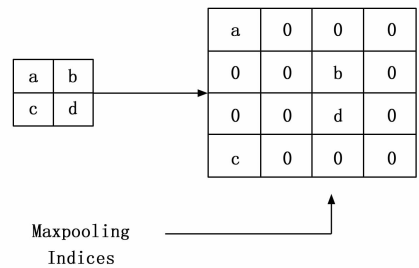


图 2 SegNet 中的上采样操作

### 1.2 空洞金字塔池化模型

如图 3 所示，展示了空洞卷积在不同扩张率下的感受野的大小。空洞卷积相较于普通卷积引入了扩张率（Dilation rate）的参数，通过改变扩张率来改变图像的感受野大小。空洞卷积可以在不改变图像分辨率的情况下，使其图像获得不同的感受野，从而获取更多的图像信息，并且此操作不会增加额外计算量<sup>[22]</sup>。本文使用的 ASPP 模型于 DeepLabv3+ 语义分割网络中提出，在金字塔池化模型的基

基础上引入空洞卷积, ASPP 模型中含有 3 个并行的膨胀率不同的空洞卷积, 还有一个普通卷积以及一个全局平均池化层, 通过 ASPP 模型将得到 5 个相同分辨率但是感受野不同的特征层, 最后将其进行线性融合。

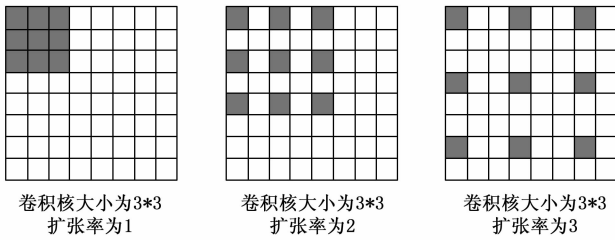


图 3 不同膨胀率下感受野的变化

### 1.3 特征金字塔

FPN 的结构如图 4 所示, 利用了编码阶段卷积神经网络金字塔形状的层级特征, 自顶向下通过对低分辨率特征进行上采样并与高分辨率特征横向融合, 从而创建了形似金字塔的多尺度特征融合模型<sup>[23]</sup>, FPN 通过融合多尺度的特征信息来弥补网络在上采样过程中细节的丢失。

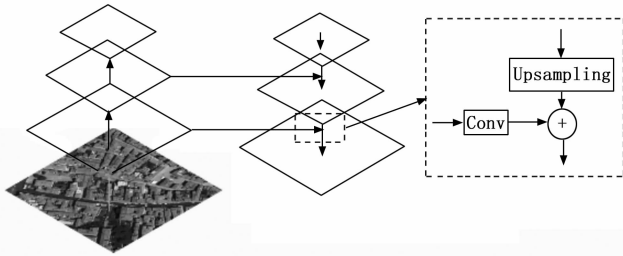


图 4 特征金字塔

## 2 改进的 SegNet+CRF 模型

### 2.1 改进的 SegNet 网络

在 1.1 一节中所述的 SegNet 网络结构是基于 Encode (编码)-Decode (解码) 结构的语义分割网络, 编码部分运用层级结构的卷积神经网络提取不同分辨率大小的图像特征, 解码部分则利用得到的特征图逐步上采样, 恢复到原图分辨率大小, 然后进行逐像素点预测。如图 1 所示, 原始 SegNet 网络的编码部分使用深度卷积神经网络得到进行上采样的最终特征层。由于深度卷积神经网络对图片特征提取的感受野的局限性, 其特征范围根据卷积核大小而定<sup>[24]</sup>, 若想要通过增大卷积核来扩大感受野则会带来巨大的计算量, 并且池化层对特征图分辨率大小的影响, 会丢失过多图像的细节。此外, 在 SegNet 原始结构的解码部分, 虽然使用最大池化索引来进行上采样在一定程度上能够减少图像特征细节的丢失, 但仅仅使用编码部分得到的最后一个特征层进行上采样, 并恢复到原图大小进行逐像素点预测, 这样仍是不够的, 网络编码部分其他特征层的高分辨率特征信息都未进行利用, 这也导致了最终网络性能并不理想。

基于以上问题, 本文提出改进的 SegNet 网络, 如图 5

所示, 分别对编码部分和解码部分进行了改进 (虚线框的“1”和“2”)。如图 5 中改进 1 部分所示, 改进的 SegNet 网络中我们使用 VGG16 作为主干网络进行图像特征的提取, 并引入带有空洞卷积的 ASPP 模型对 VGG16 得到的特征进行进一步的加强, ASPP 是加入了空洞卷积的金字塔池化模型, 其目的也是为了尽可能多的去提取图像特征横。如图 5 中改进 2 部分所示, 改进的 SegNet 网络的解码部分通过构建 FPN 模型优化上采样结构, 改进传统 SegNet 网络的多层特征融合的解码结构。

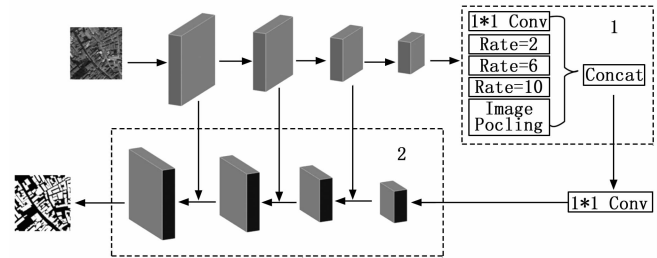


图 5 改进 SegNet 结构

传统的 VGG16 它是由牛津大学的 Visual Geometry Group<sup>[5]</sup>提出的, 是目前常见且成熟的卷积神经网络, 具有良好的迁移性, 被广泛运用于图像领域的各项任务。VGG16 网络的结构包含卷积层、最大池化层、全连接层和 softmax 层, 通过堆叠卷积层、最大池化层来形成层级的特征提取模块, 降低图像的分辨率并且提升其通道数。

在编码部分, 去掉主干特征提取网络 VGG16 的全连接层以及 softmax 层, 使其得到的特征层直接与 ASPP 相连接。首先, VGG16 的结构可以分为 5 个子模块, 每个子模块都是卷积层的堆叠与一个最大池化层, 卷积层用来改变图像的通道数并且提取特征信息, 并且由于池化层的存在, 图像特征的分辨率是逐层递减的, 在 VGG16 中卷积层卷积核的大小均为  $3 \times 3$ , 在此将 5 个子模块得到的特征层其记为  $f_1, f_2, f_3, f_4, f_5$ , 以原图大小  $H \times W \times 3$  为例, 5 个特征层对应的分辨率大小分别为下采样 2、4、8、16、32 倍, 每个特征层的通道数记为  $c_1, c_2, c_3, c_4, c_5$ 。其次, 将 VGG16 与带有空洞卷积的 ASPP 模型相结合, 输入图片通过编码部分的 VGG16 模型得到 5 个特征层, 将大小为  $(H/32) \times (W/32) \times c_5$  的特征层  $f_5$  输入 ASPP 模型, 并行地得到 5 个新的特征图, 这 5 个特征图具有与  $f_5$  相同的大小, 然后将其进行通道数的叠加, 并通过一个卷积层调整特征的通道数, 此时将会得到所需要的最终特征层, 其分辨率大小以及通道数仍保持与特征层  $f_5$  相同。其中, ASPP 中 3 个空洞卷积的扩张率分别调整为 2、6、10, 以适应不同几何尺寸的建筑物检测; 其感受野大小则由式 (1) 导出:

$$K = k + (k - 1)(r - 1) \quad (1)$$

其中:  $K$  表示感受野的大小,  $k$  表示传统卷积核的大小, 一般为  $3 \times 3$ ,  $r$  表示扩张率, 根据公式计算扩张率分别为 2、6、10 的空洞卷积其感受野分别为 5、13、21。

如图 5 所示, 在解码部分, 仅仅使用 SegNet 结构编码

部分得到的最后一个特征层进行上采样, 恢复到原图大小进行逐像素点预测, 虽然使用最大池化索引进行上采样操作减少了参数量优化了内存消耗, 但是不可避免的丢失了图像特征细节, 也因此使得原本网络的预测效果并不理想。改进网络的编码部分通过 VGG16 与 ASPP 的结合, 得到了拥有更多特征细节的最终特征层, 解码部分则通过构建 FPN 模型优化上采样结构。

利用通过 ASPP 获得的最终特征图层进行上采样并用一个  $1 \times 1$  大小的卷积层调整其通道数, 使其与  $f_4$  特征层的分辨率大小以及通道数相同, 即  $(H/16) \times (W/16) \times c_4$ , 此目的是为了将不同的特征层的对应元素相加, 以此来完成特征的融合, 如果通道数以及分辨率大小不相同则无法进行此操作, 通过特征融合后得到的特征层记为  $ff_4$ ,  $ff_4$  由式 (2) 导出:

$$ff_4 = \text{Add}(\text{Conv}(\text{UpSampling}(f_3)), f_4) \quad (2)$$

为了消除两个特征图对应元素相加可能带来的融合不充分的问题, 通过一个  $3 \times 3$  大小的卷积层进行平滑处理。处理后的特征层记为  $ff_3$ , 对  $ff_3$  进行上采样, 使其与  $f_2$  特征层的分辨率以及通道数相同, 再进行融合操作; 重复此步骤, 直到将特征图恢复到原图的分辨率, 实现 FPN 模型的构建。通过解码部分的上采样操作特征图恢复到原图大小, 通过一个卷积层调整其通道数为所需分类的数量, 例如本文使用数据集将图片的背景与建筑物分割, 则分类数量为 2, 最后, 通过 Softmax 层进行逐像素点的预测。本文构建的 FPN 结构融合了不同尺度的特征信息、实现各阶段特征图的特征融合, 有利于不同几何尺寸的建筑物提取且减少了上采样中的信息丢失现象。

### 2.2 基于 CRF 模型的后处理方法

在上文中我们通过引入 ASPP 以及构建 FPN 来提高网络对特征的提取并减少上采样中信息的丢失, 然而卷积神经网络存在一个普遍的问题, 即卷积操作更多的是提取图像中的空间特征信息, 而并未考虑像素点间上下文的关系, 这也是导致最终网络预测目标边界不清晰的重要原因, 为了优化建筑物提取的边缘, 本文引入 CRF 模型进行后处理。CRF 是一种判别式模型, 对初步的网络预测结果进行概率建模, 并优化预测结果<sup>[25]</sup>, 这是全连接条件随机场模型在图像语义分割领域上的运用, 可以充分考虑像素点之间的关系, 从而提高语义分割的精度并且平滑分割区域的边缘。

在此以本文 SegNet 为例, 其网络的分割结果为一组变量  $(X_1, \dots, X_N)$  上的随机场  $X$ , 其每一个变量的范围为标签  $L = \{l_1, l_2, \dots, l_k\}$ ; 另有一定义在变量  $\{I_1, \dots, I_N\}$  上的随机场  $I$ ,  $I$  的范围为可能输入的  $N$  个图片并且  $X$  的范围为像素可能的标签,  $I_j$  为像素  $j$  的颜色信息,  $X_j$  为像素  $j$  的在标签中所属类别。  $(I, X)$  即组成一个随机场, 遵循 Gibbs 分布:

$$P(X | I) = \frac{1}{Z(I)} \exp \left[ - \sum_{c \in C_c} \Phi_c(X_c | I) \right] \quad (3)$$

在式 (2) 中,  $G$  表示  $X$  上的一个图,  $C_c$  为  $G$  中的最大子图集,  $c$  为  $C_c$  中的一个最大子图,  $\Phi_c$  则是这个最大子图的势函数,  $P(X | I)$  则表示在  $I$  的前提下计算所有像素点属于某一类别的条件概率<sup>[26]</sup>。全连接条件随机场的 Gibbs 能量函数为:

$$E(X) = \sum_i \psi_u(x_i) + \sum_{i < j} \psi_p(x_i, x_j) \quad (4)$$

在式 (3) 中,  $\psi_u(x_i)$  为一元势函数, 表示观测场与标签像素独立推断的结果, 在本文中即为网络的预测结果;  $\psi_p$  为二元势函数, 即考虑像素之间成对关系的推断结果, 通过考虑像素点间各项因素, 进一步对像素点所属类别进行判别。由于需要考虑距离、颜色等因素的关系, 因此能够起到了在边界处分割的效果, 从而进一步优化分割物体的边界。

## 3 实验分析

### 3.1 数据集

本文使用公开数据集 Inria Aerial Image Labeling Dataset, 数据集中包含 180 张大小为  $5\,000 \times 5\,000$  的城市航空影像, 进行预处理将每一张图片裁剪为有像素重叠的大小为  $512 \times 512$  的图像, 如图 6 所示, 然后按照 8:1:1 的比列将裁剪后得到的数据集划分为训练集、验证集、测试集, 最终得到训练集 24 336 张, 验证集与测试集均为 3 042 张。网络与模型由 Tensorflow 与 Python 进行构建, 采用迁移学习思想初始化网络权重, 网络初始学习率为  $1 \times 10^{-4}$ , 优化函数选用 Adam, 损失函数选用交叉熵损失函数 (categorical crossentropy loss)。

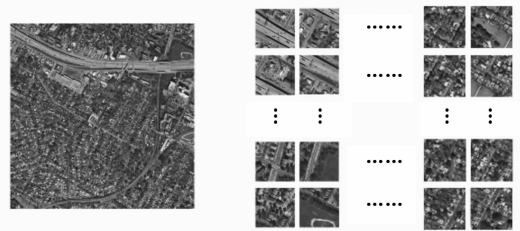


图 6 数据集预处理

在深度学习任务中, 在数据集并不壮大的情况下, 使用迁移学习可以很好地进行网络训练, 由于其强大的泛化能力可以捕捉数据间的关联性, 并且可以减少训练时间、防止过拟合现象的发生<sup>[27]</sup>。本文运用迁移学习的思想 (Transfer learning)<sup>[28]</sup> 在网络的编码阶段对 VGG16 加载 ImageNet 预训练权重。

### 3.2 定量分析

本文将原始 SegNet 模型, 称为 Model1; 改进 SegNet 模型, 称为 Model2; 改进 SegNet+CRF 模型, 称为 Model3, 并与经典语义分割模型 Unet 网络以及 Deeplabv3 网络进行结果对比; 采用像素准确率 (PA, pixel accuracy), 召回率 (Recall), 平均交并比 (MIoU, mean intersection over union) 定量分析不同模型建筑物提取的精度、稳健性。

3 种评价标准的计算参数如表 1 所示, 由表 1 中的参数可得到 3 种评价标准的公式, 像素准确率表示分类正确的像素在总像素个数中的占比, 如式 (5) 所示; 召回率表示预测值为 1、真实值也为 1 的像素在真实值为 1 的像素中的占比, 如式 (6) 所示; 平均交并比表示真实值和预测值两个集合的交集和并集之比, 如式 (7) 所示。

表 1 真值表

真实值 \ 预测值	Positive	Negative
Positive	True Positive(TP)	False Positive(FP)
Negative	False Negative(FN)	True Negative(TN)

$$\text{Pixel Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{MIoU} = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{TP + FP + FN} \quad (7)$$

将不同模型在训练集与验证集上完成训练后, 在测试集上进行图像的预测, 随后将各模型预测结果进行语义分割评价指标的对比, 如表 2 所示。

表 2 不同网络性能量化数据对比

评价指标 \ 模型	PA	Recall	MIoU
Model1	96.17	96.44	85.69
Model2	96.5	97.29	87.14
Model3	96.65	97.73	88.05
Unet	94.53	93.64	84.21
Deeplabv3+	96.48	96.55	86.16

由表 2 可知, 本文提出的模型 2 以及加入 CRF 后处理的模型 3 相较于 Unet 以及 Deeplabv3+ 各项指标均有明显提升, 模型 3 的 PA 指数相较于 Unet、Deeplabv3+ 增加 2.12%、0.17%; Recall 指标增加 4.09%、1.18%; MIoU 指数增加 3.84%、1.89%。模型 2 相比模型 1 的 PA 指数提升 0.33%、Recall 指数提升 0.85%、MIoU 指数提升 1.45%。模型 3 相比模型 1 的 PA 指数提升 0.48%、Recall 指数提升 1.29%、MIoU 指数提升 2.36%, 其中, Recall 指数与 MIoU 指数增长较多。

为检测本文改进方法中各个模块对网络性能提升的有效性, 对网络进行消融实验, 在去除不同模块的条件下对比网络预测结果的 MIoU 指标, 如表 3 所示。从表中不难看出, 移除任意一个模块都可能导致网络性能的下降, 在分别单独使用 ASPP 模块和 FPN 模块后的 MIoU 指数, 相比原始的 SegNet 网络具有明显的增长; 并且在同时使用 3 种模块后 MIoU 指数相比原始网络增长了 2.36%。

在进行网络预测结果的定量分析后, 随机选择三张图片及其预测结果, 从图片上直观地对比不同训练模型的预测差异。由图 7 和图 8 的圆圈标记的区域可视化显示, 可以看出, SegNet 的预测图相比原数据集的标注图, 建筑物分割

表 3 消融实验

ASPP	FPN	CRF	MIoU
×	×	×	85.69
√	×	×	86.85
×	√	×	86.01
√	√	×	87.14
√	√	√	88.05

区域有明显的膨胀, 并且建筑物之间存在粘连现象, 主要存在于尺度较小的并排建筑物之间, 并且建筑物边缘并不平滑且具有锯齿现象; Unet 以及 Deeplabv3+ 的预测结果中并排小建筑间的间隙相较于 SegNet 的预测结果有些许改善, 分割较为清晰, 但仍有很多并排小建筑物间的粘连现象, 并且个别较大建筑识别不完全; 反观本文使用的方法, 总体上较传统方法提取的建筑物边界更加清晰可见、更加平滑, 锯齿现象以及区域膨胀现象明显改善, 更加接近标注图的识别效果; 如图 9 所示, 图中包含几个较大的建筑物, CRF 对边缘的改进效果非常明显, 而原始 SegNet 得到的预测图中建筑物的边缘锯齿状较为严重, 通过本文改进网络以及 CRF 的后处理边缘更加平滑, 大大减少了锯齿现象, 并且识别效果更加接近标注图像; 反观 Unet 以及 Deeplabv3+ 的预测图, 不难看出这两个模型在较大建筑物的识别上效果稍差, 不仅边缘非常的粗糙, 周围还带有识别错误的错分像素, 大建筑更是有较为明显的识别不完全现象。

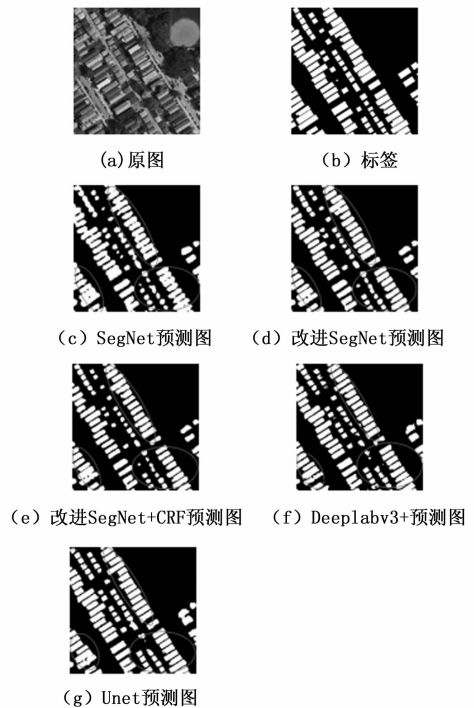


图 7 原图, 标签以及各模型预测结果

### 4 结束语

传统的语义分割网络模型存在上采样阶段丢失过多的特征细节, 忽略了像素点间的上下文关系问题。以传统语

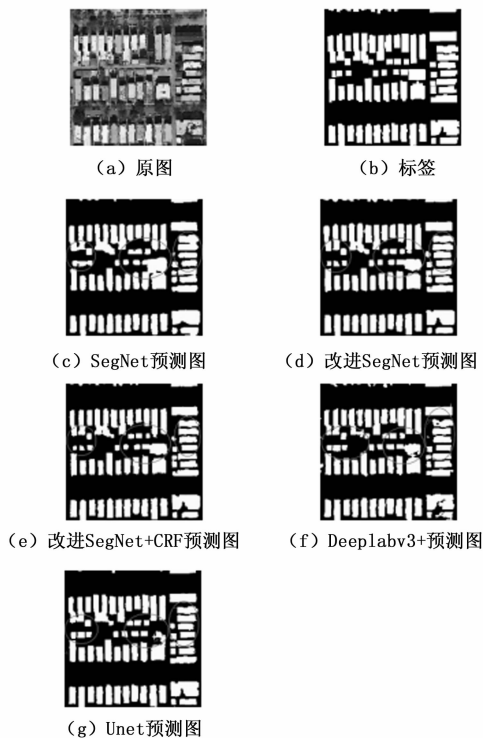


图 8 原图, 标签以及各模型预测结果

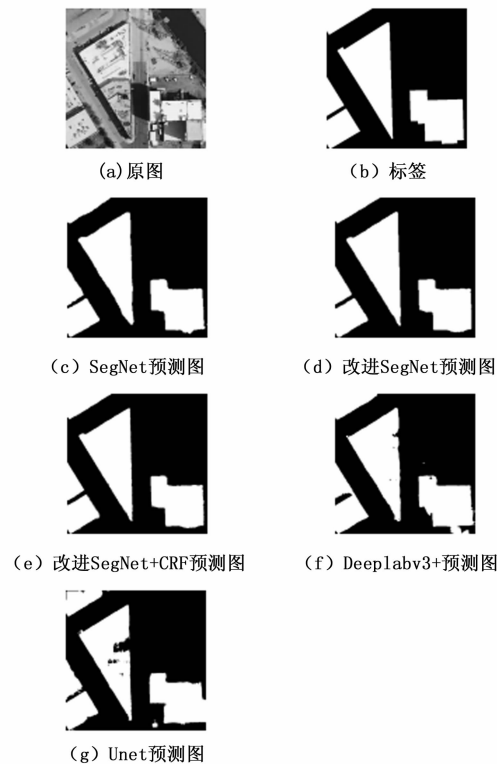


图 9 原图, 标签以及各模型预测结果

义分割网络直接用于高分辨率遥感影像建筑物信息提取, 存在提取精度低、边缘细节丢失的问题。本文提出改进的 SegNet 网络与 CRF 后处理结合的方法来改善以上问题。首先, 在改进的网络中通过引入空洞卷积来获取更多的图像

特征细节, 并构建 FPN 融合多尺度的特征信息, 减少特征细节的丢失; 其次, 引入 CRF 的后处理, 通过构建像素点间的成对关系, 来考量像素间的上下文关系, 从而对分割图像进一步的细化, 纠正错分的像素点, 细化原本粗糙的物体边界。此外, 通过加载预训练权重来训练网络, 一方面减少训练时间, 另一方面弥补数据集有限的问题。实验表明, 本文提出的高分辨率遥感影像建筑物提取方法是可行的, 且可以拓展到其它类似的网络。但是, 对于密集建筑物的分割, 分割精度有待进一步提高; 结合建筑物的几何特点, 进一步优化 SegNet 的编、解码网络是今后需要研究的问题。

参考文献:

[1] 常京新, 王双喜, 杨元维, 等. 高分遥感影像建筑物轮廓的逐级优化方法 [J]. 中国激光, 2020, 47 (10): 241-254.

[2] 罗会兰, 张云. 基于深度网络的图像语义分割综述 [J]. 电子学报, 2019, 47 (10): 1-10.

[3] 王舒洋, 慕晓冬, 杨东方, 等. 融合高阶信息的遥感影像建筑物自动提取 [J]. 光学精密工程, 2019, 27 (11): 2474-2483.

[4] ALEX K, ILYA S, GEOFFREY E. Hinton. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60 (6): 84-90.

[5] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C] // International Conference on Learning Representations (ICLR). 2015: 1-14.

[6] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C] // Proceedings of the IEEE conference on computer vision and pattern recognition, 2015: 1-9.

[7] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 770-778.

[8] 杨潇宇, 汪西莉. 结合多尺度注意力和边缘监督的遥感影像建筑物分割模型 [J]. 激光与光电子学进展, 2021, 59 (22): 1-20.

[9] SHELHAMER E, LONG J, DARRELL T. Fully Convolutional Networks for Semantic Segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017; 39 (6): 640-651.

[10] RONNEBERGER O, FISCHER P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation [J]. Springer International Publishing, 2015.

[11] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017; 1.

[12] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network [C] // Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 2881-2890.

[13] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs [J]. Computer Science, 2014 (4): 357

