

doi: 10.6046/gtzyyg.2020.04.11

引用格式: 刘尚旺,崔智勇,李道义. 基于 Unet 网络多任务学习的遥感图像建筑地物语义分割[J]. 国土资源遥感,2020,32(4):74-83. (Liu S W, Cui Z Y, Li D Y. Multi-task learning for building object semantic segmentation of remote sensing image based on Unet network[J]. Remote Sensing for Land and Resources, 2020,32(4):74-83.)

基于 Unet 网络多任务学习的遥感
图像建筑地物语义分割

刘尚旺^{1,2}, 崔智勇^{1,2}, 李道义^{1,2}

(1. 河南师范大学计算机与信息工程学院, 新乡 453007; 2. “智慧商务
与物联网技术”河南省工程实验室, 新乡 453007)

摘要: 为准确分割出高分辨率遥感图像中的建筑地物,提出一种基于 Unet 网络多任务学习的建筑地物语义分割方法。首先,根据遥感图像建筑地物真值图生成边界距离图,并将该遥感图像及其真值图共同作为 Unet 网络的输入;然后,在基于 ResNet 网络构建的 Unet 网络末端加入建筑地物预测层与边界距离预测层,搭建多任务网络;最后,定义多任务网络的损失函数,并使用 Adam 优化算法训练该网络。在 Inria 航空遥感图像建筑地物标注数据集上进行实验,结果表明,与全卷积网络结合多层感知器方法相比,VGG16 网络、VGG16 + 边界预测、ResNet50 和本文方法的交并比值分别提升 5.15、6.94、6.41 和 7.86 百分点,准确度分别提升至 94.71%、95.39%、95.30% 和 96.10%,可实现高精度的建筑地物提取。

关键词: Unet 网络; 多任务学习; 遥感图像; 语义分割; ResNet 网络

中图法分类号: TP 751.1 **文献标志码:** A **文章编号:** 1001-070X(2020)04-0074-10

0 引言

如今,随着遥感卫星技术的快速发展,大量高分辨率遥感图像随之产生并得到应用^[1-3],其中,从高分辨率遥感图像中自动分割提取建筑地物一直广受关注^[4-5],研究高效的建筑地物提取方法将极大促进城市规划、灾害管理和环境管理等领域的发展^[4]。然而,由于不同地区建筑地物的多样性(如颜色、形状、大小、材料等),以及建筑地物与背景或其他物体的相似性^[5],致使遥感图像建筑地物提取方法的研究极具挑战性^[4]。

目前,从高分辨率遥感图像中提取建筑地物,主要有以下 2 类方法:①基于传统的计算机视觉方法^[6],使用人工特征,如植被指数、纹理和颜色特征;②传统的机器学习分类器(如 AdaBoost、支持向量机(support vector machine, SVM)、随机森林)实现建筑地物提取,并且通常会采用相应的后处理步骤来细化分割结果^[7-8]。然而,此类方法不仅具有

较高的模型复杂度,需大量人机交互,而且往往受制于人工知识与经验的限制^[7-8]。而诸如完全卷积网络(fully convolutional network, FCN)或基于编码器与解码器体系结构的卷积神经网络(convolutional neural network, CNN)已经成功应用于该领域且优于传统的计算机视觉方法^[9]。Yuan^[10]使用 FCN 来预测像素到边界的距离,并对其进行阈值处理以获得最终的分割结果;Zhang 等^[11]在 Google Earth 图像上训练 CNN 并使用最大抑制的后处理步骤移除虚假建筑地物。但是,这些方法由于使用池化层进行下采样,尽管增加卷积核的感受野可以提取图像的全局特征,但同时丢失图像中的高频细节致使分割结果中缺失边界信息,易将建筑地物分割成许多“圆斑”,难以提取建筑地物完整边界^[12]。为此,研究人员使用串联连接重新引入高频细节^[13-14],或扩张卷积^[15-16],以及相应的后处理(如条件随机场(conditional random field, CRF))^[15]来解决这个问题。Maggiori 等^[17-18]在 FCN 之上,训练多层感知器(multi-layer percep-

tron, MLP)以结合网络中不同卷积层提取的局部特征,实现前端细节信息与后端抽象信息的结合,从而界定出建筑地物的位置;Marmanis 等^[19]和 Peng 等^[20]提出将来自不同尺度的多个网络的特征图组合起来,并在这些连接的特征图之上进行最终预测以识别建筑地物;Huang 等^[21]提出不依赖于额外后处理步骤的方法,使用 RGB 和 NRG 波段组合 2 个并行流训练反卷积网络,并融合这 2 个流的预测结果;Li 等^[5]和 Bischke 等^[22]也使用类似的 2 个流网络,并行处理 RGB 和数字高程模型(digital elevation model, DEM),以保留分割结果的边界信息,该方法首先使用 SegNet 网络作为特征提取器并应用边缘检测网络来提取边缘信息,之后将 SegNet 的特征图与边缘预测连接,将边界预测注入网络。尽管上述研究取得较好的建筑地物提取效果,但存在 2 个较为严重的问题^[14]:①加入后处理步骤的建筑地物分割方法,存在模型过于复杂且模块之间整合困难的问题;②通过多种不同的网络提取不同特征并融合这些特征的方法,存在网络复杂、硬件设备要求高且训练时间长的问題。为此,研究人员将应用于医学图像,且可高精度提取目标物边界的具有编码与解码层结构的 Unet 网络^[23]引入到遥感图像建筑地物提取的任务中,Iglovikov 等^[24]利用 VGG11 重新搭建 Unet 网络二值化提取建筑地物,在仅使用单个网络的情况下其提取精度获得巨大提升;Xu 等^[25]结合 ResNet 网络^[26]与 Unet 网络搭建 Res - Unet 网络,既可实现图像的边界提取,又可准确地获取建筑地物位置,且网络参数共享,不仅降低模型的整体复杂性,而且避免边缘检测不准确的问题。

综上所述,基于 ResNet 网络改进的 Unet 网络能够提取较为清晰的边界而分割出准确的目标对象^[25];以及基于多任务学习搭建的深度网络,可在一个网络主体上训练多种不同的任务,从而避免为多种任务搭建不同网络的资源浪费^[27]。故本文在 Unet 网络框架下,搭建基于 ResNet 网络的语义分割网络;为进一步提升建筑地物提取精度,使用多任务学习策略在该网络中加入边界距离预测层来提取完整建筑地物边界,从而实现建筑地物的高精度提取,同时避免计算资源浪费。

1 多任务网络

本文提出的多任务网络,通过使用多任务学习,

不仅将建筑地物分割结果加入到目标损失函数中,而且引入建筑地物分割结果的边界信息,以改进建筑地物的最终分割结果。

1.1 边界距离图

多任务网络的目标是让网络除具有分割建筑地物的语义信息之外,还能在训练中赋予提取建筑地物边界信息的几何属性。从建筑地物真值图中,可方便提取建筑地物边界信息的多种几何属性,如形状、边缘信息。本文采用建筑地物像素到边界的距离作为训练网络产生几何属性的训练数据。使用该训练数据可使网络具备以下优点:①可从现有的建筑地物真值图中,通过距离变换快捷地制作出边界距离图;②使用边界距离图设计的损失函数(如均方差或负对数),计算形式方便,从而使网络学习到图像中每个像素的边界位置信息并隐式地捕获其几何属性。

为此,假设 Q 表示建筑地物边界的像素集合, C 表示属于建筑地物的像素集合,则对于图像中的每个像素 p ,其截断距离 $D(p)$ 可定义为^[28]:

$$D(p) = \delta_p \min[\min_{q \in Q} d(p, q), R],$$
$$\delta_p = \begin{cases} +1 & (p \in C) \\ -1 & (p \notin C) \end{cases}, \quad (1)$$

式中: $d(p, q)$ 为像素 p 与 q 的欧氏距离; R 为截断阈值;符号函数 δ_p 为对像素 p 的距离加权,表示像素是否位于建筑地物的内部或外部。

然后,均匀地量化连续距离值以便于训练。将边界距离图使用 one-hot 编码为 K 维二进制矢量 $b(p)$ ^[29],即

$$D(p) = \sum_{k=1}^K r_k b_k(p), \sum_{k=1}^K b_k(p) = 1, \quad (2)$$

式中 r_k 为对应于 k 的距离值。 k 得到的二进制像素距离图可以被理解为第 k 个边界距离中的每一个像素的边界距离图。

至此,训练多任务网络的数据已生成。表 1 展示了 10 幅训练样本图像及其对应的语义分割和边界距离真值,其中,两两相似的图像用于测试网络的鲁棒性及对小规模建筑地物分割的有效性。其中,第三列表示建筑地物到边界的距离,其距离越大说明该像素越不属于边界像素,值越小说明该值最有可能是边界像素,由此边界距离真值图训练的网络可最大限度保留建筑地物的边界信息。

表 1 训练数据可视化
Tab.1 Visualization of training data

图像编号	训练样本	语义分割真值	边界距离真值	图像编号	训练样本	语义分割真值	边界距离真值
1				6			
2				7			
3				8			
4				9			
5				10			

1.2 多任务网络结构

本文提出的多任务网络是基于 FCN 网络 Unet^[23]而构建。Unet 是具有卷积编码与卷积解码完全对称的网络结构,常用于图像语义分割。然而,Unet 的网络架构简单,尽管将其应用于遥感图像建筑地物提取可准确检测建筑地物的位置,但检测结

果经常为一些尺寸大小不等的圆斑且大多数建筑地物不能检测出来,并且边界信息大量丢失。为此,本文的多任务网络,使用 ResNet 网络^[26]为基础网络,重新构建 Unet 网络;之后,在网络末端,分别加入建筑地物预测层与边界距离预测层,以实现多任务学习,如图 1 所示。

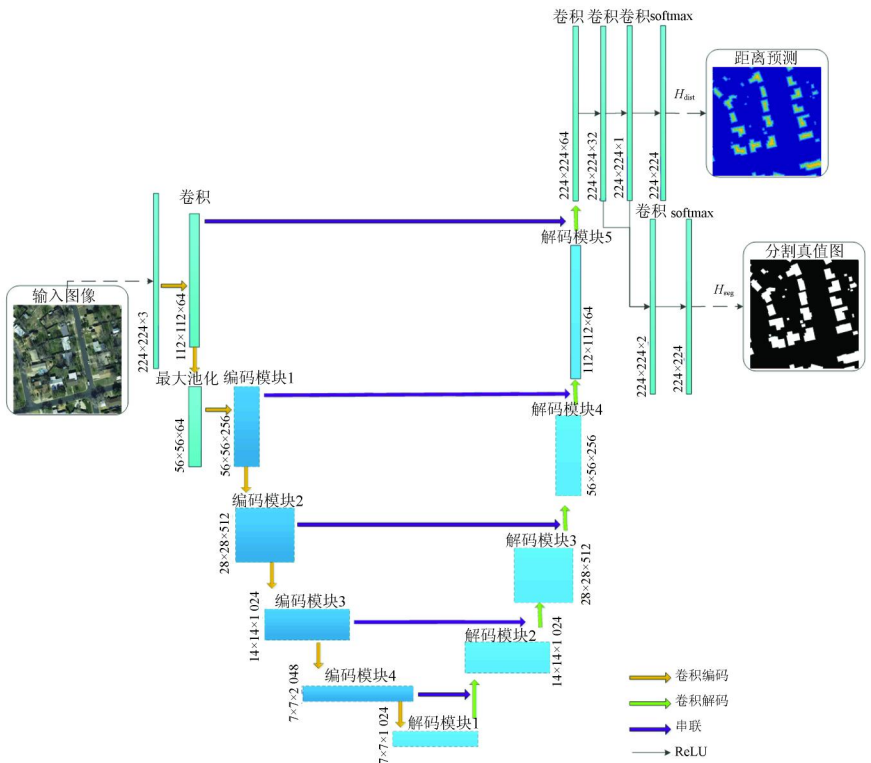


图 1 多任务网络结构
Fig.1 Framework of the multi – tasking network

图 1 中,多任务网络由卷积编码(下采样)和卷积解码(上采样) 2 条路径组成。卷积编码部分是 ResNet 网络(本文选择的是 ResNet50^[26]),用于提取输入图像的特征。首先,输入图像通过卷积、标准化和激活函数操作进入卷积层和最大池化层,以提高网络对失真和迁移的鲁棒性。之后,在特征提取期间,有 4 个编码模块,每个模块包括几个 ResNet 网络的残余块。同一模块中的特征图具有相同的大小,下层模块的特征图是上层模块的一半。不同模块中的特征映射具有不同的缩放特征。而相对应的卷积解码部分旨在使用卷积编码获得的特征图提取建筑地物。参考 Unet 的特征金字塔设置^[23],为获得多尺度的特征,在卷积解码部分中设计与卷积编码部分相应模块的串联。卷积解码部分中的每个模块都包括对应卷积编码和下层模块的输入,以此使卷积解码部分重新获得高频信息。在网络的末端,添加 2 个卷积层,分别用于预测图像中每个像素到建筑地物边界的距离 H_{dist} 和用于预测建筑地物分割结果的距离 H_{seg} ,由距离预测卷积层构建;且加入的 2 个卷积层均伴随着相应的 softmax 层以完成不同的预测任务,从而使多任务网络可充分利用卷积解码特征映射中存在的语义属性和几何属性。因

$$L_t(x, \theta, \sigma_{\text{dist}}, \sigma_{\text{seg}}) = L_{\text{dist}}(x, \theta, \sigma_{\text{dist}}) + L_{\text{seg}}(x, \theta, \sigma_{\text{seg}}), \quad (4)$$

式中: L_{dist} 和 L_{seg} 分别为用于预测建筑地物边界和分割结果的分类损失函数; σ_{dist} 和 σ_{seg} 分别为 λ_i 的相应任务权重。

将每个分类任务模型可能性使用模型输出 $f(x)$ 导出,其中不确定性通过 softmax 函数可简化为:

$$P(C = 1 | x, \theta, \sigma_i) = \frac{\exp[\frac{1}{\sigma_i^2} f_c(x)]}{\sum_{c'=1}^c \exp[\frac{1}{\sigma_i^2} f_{c'}(x)]}, \quad (5)$$

式中: P 为多任务估计; $f_c(x)$ 为期望输出; $f_{c'}(x)$ 为原始的实际输出; σ_i 为标量缩放系数。

对式(5)使用负对数似然,用不确定性表示分类损失,即

$$L_t(x, \theta, \sigma_i) \approx \frac{1}{2} \sum_{c=1}^c -C_c \log_{\text{istic}} P(C_c = 1 | x, \theta) + \log_{\text{istic}}(\sigma_i^2). \quad (8)$$

式中使用近似方程自适应地组合 2 个分类任务边界预测 L_{dist} 和分割结果预测 L_{seg} ,其中,为了数值稳定性,不同任务的损失函数权重使用 $\log_{\text{istic}}(\sigma_i^2)$ 而不是 σ_i^2 。

此,使用 ResNet 网络作为特征提取器,可解决卷积层增加而导致的梯度消失问题,并且在卷积编码部分提取出有效图像特征;卷积解码部分中的串联连接能够学习多个尺度和不同网络层的特征,可增加网络的鲁棒性以提高建筑地物分割的准确性;最后加入多任务预测结构,能够使网络具备提取目标对象的语义和几何属性的能力。

1.3 基于不确定性的多任务损失

在多任务网络中,其损失函数 L_t 为单任务损失函数的加权和,即

$$L_t(x, \theta) = \sum_{i=1}^T \lambda_i L_i(x, \theta), \quad (3)$$

式中: x 为训练图像; T 为任务数; L_i 为相对于网络参数 θ 最小化的相应任务损失函数; λ_i 为权重,以模拟每个任务 L_i 在多损失函数 L_t 中的重要性。其中,多任务损失中的加权项 λ_i 可结合单任务中的超参数,以形成适应多任务网络的总体损失。考虑每个任务中模型预测的不确定性来学习相应任务的权重 λ_i ,其目的是根据单任务预测的置信度来学习相对任务权重^[27]。为此,将多损失函数 L_t 定义为 2 个像素级分类损失的组合,即

$$\begin{aligned} L_t(x, \theta, \sigma_i) &= \sum_{c=1}^c -C_c \log_{\text{istic}} P(C_c = 1 | x, \theta, \sigma_i) \\ &= \sum_{c=1}^c -C_c \log_{\text{istic}} \left\{ \exp \left[\frac{1}{\sigma_i^2} f_c(x) \right] \right\} +, \quad (6) \\ &\log_{\text{istic}} \sum_{c=1}^c \exp \left[\frac{1}{\sigma_i^2} f_{c'}(x) \right] \end{aligned}$$

式中 C_c 为 c 类的像素集。

假设网络的多个输出由连续输出和离散输出组成,分别用高斯似然和 softmax 似然建模^[27],可进一步简化损失函数,即

$$\frac{1}{2} \sum_{c=1}^c \exp \left[\frac{1}{\sigma_i^2} f_{c'}(x) \right] \approx \left\{ \sum_{c=1}^c \exp [f_{c'}(x)] \right\}^{\frac{1}{\sigma_i^2}}. \quad (7)$$

故结合式(6)—(7)的多任务网络的损失函数为:

1.4 训练多任务网络

在完成多任务网络搭建,以及自适应损失函数定义之后,使用动量与 RMSProp 算法结合的 Adam 优化算法^[30]训练该网络,以优化更新网络参数,即

$$\begin{cases} w_t = w_{t-1} - \alpha_t \frac{\dot{m}_{d_w}}{\sqrt{\dot{v}_{d_w}} + \varepsilon} \\ b_t = b_{t-1} - \alpha_t \frac{\dot{m}_{d_b}}{\sqrt{\dot{v}_{d_b}} + \varepsilon} \end{cases}, \quad (9)$$

式中： w_t 、 b_t 和 α_t 分别为第 t 次迭代的网络权重、偏置和学习速率； $\dot{m}_{d_w}$ 和 $\dot{m}_{d_b}$ 分别为网络权重与偏置的偏差纠正动量值； $\sqrt{\dot{v}_{d_w}}$ 和 $\sqrt{\dot{v}_{d_b}}$ 分别为 w_t 和 b_t 的偏差纠正 RMSProp 值； ε 为超参数，一般取 10^{-8} 。

在式(9)中,为加速收敛多任务网络,使用离散指数衰减动态调整当前迭代的学习速率,即

$$\alpha = \alpha_0 g^{f(t/10)}, \quad (10)$$

式中： α 为当前的学习速率； $f(\bullet)$ 为下取整； g 为速率衰减因子； t 为当前的迭代次数； α_0 表示初始学习速率。

2 实验与分析

2.1 数据集及数据增广

本文在大规模 Inria 航空遥感图像建筑地物标注数据集^[17]上进行实验。该数据集中的样本图像是经过正射校正且空间分辨率为 0.3 m 的 RGB 图像,每幅图像尺寸为 5 000 像素 × 5 000 像素,覆盖地表面积为 1 500 × 1 500 m²。所有样本图像从 10 座城市的遥感图像中获取,总面积为 810 km²。这些图像具有不同的城市建筑地物,从人口稠密地区到高山城镇。该数据集仅标注建筑地物与非建筑地物 2 个语义类,其训练集均有标注完整的真值。为了具有可比性,本文按照文献[17]的数据集划分方法,来划分训练集与验证集(每个城市的第 1—5 幅样本图像用于验证,第 6—36 幅用于训练)。

数据增广的目的是生成新的样本实例,并且当训练样本较少时,数据增广对于提高网络的鲁棒性非常有用。对于遥感图像,有许多常用的数据增广方法,如颜色抖动、随机裁剪、水平/垂直翻转、移位、旋转/反射、噪声、切割和切换频带等。由于大多数遥感图像是正射影像,因此变化主要反映在方向和

尺度上。然而,本文使用的数据集中的图像具有相同的空间分辨率,没有大的尺度变化,因此仅使用其中 3 种常见的增广方法:水平/垂直翻转、旋转和随机裁剪。从原始图像中随机提取尺寸为 224 像素 × 224 像素的图像块,对其进行水平和垂直翻转以及不同角度的旋转。经过数据增广后,原数据集可扩大 14 倍。需说明的是,仅对原数据训练集进行增广,并不再为验证集进行数据增广。

2.2 评价指标

为验证提出方法的语义分割性能,采用 2 种评价指标来评估不同方法在数据集上的表现:①准确度(accuracy, Acc),表示正确分类像素的百分比;②交并比(intersection over union, IoU),其被定义为预测结果与真值图均是建筑地物的交集除以其并集,即

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}, \quad (11)$$

式中： A 为不同方法预测的建筑地物； B 为真值图中的建筑地物。

2.3 实验平台

采用高性能服务器实验平台:操作系统为 Ubuntu18.04;CPU 为 Intel(R) Xeon(R) cpu E5-2650 V2 @ 2.6 GHz (×2);内存为 64 GB;GPU 为 Tesla K40m (×3);显存为 23 GB;深度学习平台为 Keras 2.2.4;数据可视化工具为 Matplotlib 2.2.0;编程语言为 Python 3.6.4。

2.4 实验结果与分析

为系统验证本文方法通过加深语义分割网络的编码与解码层,以及加入边界预测层使用级联多任务学习搭建的 Unet 网络的优越性,选取其中 5 个城市的遥感图像,分别将本文方法与 FCN 结合 MLP 方法(FCN + MLP)^[17]、基于 VGG16 搭建的 Unet 网络(VGG16)^[24]、加入本文边界预测的基于 VGG16 搭建的 Unet 网络(VGG16 + 边界预测)和基于 Res-Net50 搭建的 Unet 网络(ResNet50)^[25]的实验结果进行对比分析,实验结果如表 2 所示。

表 2 不同方法的实验结果

Tab. 2 Experimental results of different methods (%)										
城市	FCN + MLP		VGG16		VGG16 + 边界预测		ResNet50		本文方法	
	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc
Austin	61.20	94.20	70.66	95.28	72.81	95.82	72.38	95.79	74.41	96.09
Chicago	61.30	90.43	66.37	91.44	67.38	91.92	66.12	91.50	67.76	92.02
Kitsap Co.	51.50	98.92	57.55	98.19	57.54	98.90	58.68	98.95	60.19	98.63
West Tyrol	57.95	96.66	67.82	95.35	67.18	97.01	67.32	97.07	69.09	97.74
Vienna	72.13	91.87	77.01	93.28	77.19	93.31	76.86	93.21	78.21	93.63
均值	64.67	94.42	69.82	94.71	71.61	95.39	71.08	95.30	72.53	96.10

从表 2 可以看出,本文的多任务网络具有以下优势。

1) 具有更深的编码与解码层搭建的 Unet 网络可取得较优的建筑地物分割结果。表 2 中,FCN + MLP 方法采用简单的 4 层卷积编码层搭建的 FCN,之后使用 MLP 组合不同层的特征映射以输出最终建筑地物预测结果,尽管 MLP 组合来自不同层的特征映射,但因本身的编码与解码层较浅,无法充分提取建筑地物本身的多变特征致使建筑地物提取结果较差。为了验证编码器和解码层的深度对构建 Unet 网络的重要性,本文分别使用 VGG16 与 ResNet50 重新搭建 Unet 网络。新搭建的 Unet 网络中的编码层的网络权重分别使用 VGG16 与 ResNet50 网络在 ImageNet^[28] 上预训练的网络权重初始化,以及使用高斯分布初始解码层的网络权重;之后,使用 Adam 优化算法,设置初始学习速率率为 0.01,采用指数的衰减速率,速率衰减因子为 0.7,使用预测结果与分割标签的负对数似然函数计算损失并传递损失优化网络。从表 2 中所得到的不同网络搭建的 Unet 网络的实验结果可以看出:由于编码层与解码层采用更深层网络架构,相比于 FCN + MLP 方法,VGG16 网络^[24]、VGG16 + 边界预测、ResNet50 网络和本文方法的 IoU 均值分别提升 5.15、6.94、6.41 和 7.86 百分点,Acc 均值分别提升至 94.71%、95.39%、95.30% 和 96.10%,这表明 FCN + MLP 方

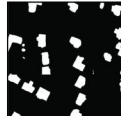
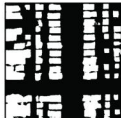
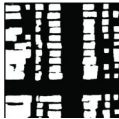
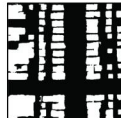
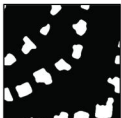
法所提取的特征对于遥感图像建筑地物分割任务来说难以提取图像的深度抽象特征;采用深层网络搭建编码与解码层在语义分割任务中起着至关重要的作用,相对于 VGG16 网络,ResNet50 网络的 IoU 和 Acc 均值分别提高了 1.26 和 0.59 百分点。

2) 距离预测的重要性。为验证本文提出的加入边界距离预测的多任务网络的优势,分别在基于 VGG16 与 ResNet50 搭建的 Unet 网络中加入边界距离预测层,即除了用于分割结果预测层 H_{seg} ,同时附加用于距离预测层 H_{dist} 。经大量实验验证,在制作边界距离图生成训练数据时,当式(1)中截断距离 $R = 20$,式(2)中区间数 $k = 10$ 时,本文方法可取得最高的遥感图像建筑地物分割精度。从表 2 可以看出,VGG16 + 边界预测方法比 VGG16 分割结果的 IoU 和 Acc 均值分别提高 1.79 和 0.62 百分点;本文方法的 IoU 和 Acc 均值分别达到 72.53% 和 96.10%,取得了最高的遥感图像建筑地物分割精度。为此,得出结论:加入边界距离预测层的多任务网络可获得更高的分类精度,究其原因是加入边界距离预测层的 Unet 网络在训练过程中使该多任务网络的主体层(编码与解码层)学习到建筑地物的几何属性,在分割预测层尽可能保留建筑地物的边界信息,从而获得更高的语义分割精度。

为进一步验证本文方法的有效性,不同方法遥感图像建筑地物的分割结果如表 3 所示。

表 3 不同方法遥感图像建筑地物分割结果

Tab. 3 Building object segmentation results of remote sensing image by using different methods

图像 编号	遥感图像	FCN + MLP	VGG16	VGG16 + 边界预测	ResNet50	本文方法	真值
1							
							
3							
							
5							
							

(续表)

图像 编号	遥感图像	FCN + MLP	VGG16	VGG16 + 边界预测	ResNet50	本文方法	真值
7							
8							
9							
10							

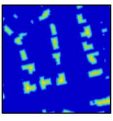
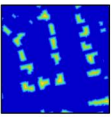
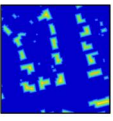
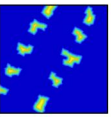
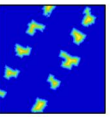
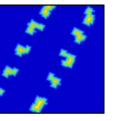
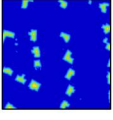
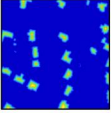
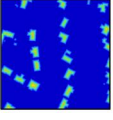
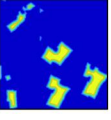
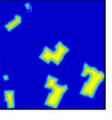
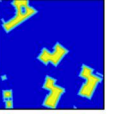
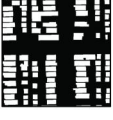
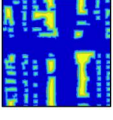
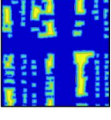
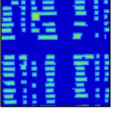
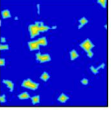
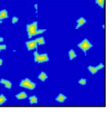
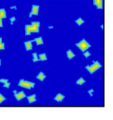
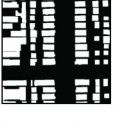
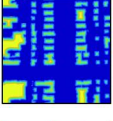
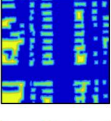
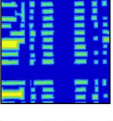
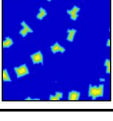
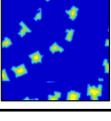
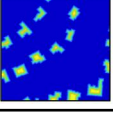
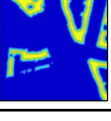
从表 3 可以看出,不同方法提取 10 幅遥感图像建筑地物的结果中,仅 FCN + MLP 方法的分割结果表现出明显的“圆斑”效应,而其他方法的分割结果与真值十分接近。这也可以从表 2 中看出,相比于 FCN + MLP 方法,其他 4 种方法的 IoU 和 Acc 值均有较大提升,因此其分割结果会出现较大差异;而其他 4 种方法的 IoU 和 Acc 值的差异较小,并且表 3 中的 10 幅图像的尺寸为 500 像素 × 500 像素,仅仅是一景遥感图像的万分之一,故视觉效果较差。但细致观察图像 1 与 2 的分割结果,本文方法相比其他 4 种方法可更加准确地分割小规模建筑地物。图

像 4 的分割结果中,由于建筑地物之间间隔较小,很容易造成分割结果中出现不同程度的合并以及边界出现毛边,而 ResNet50 的分割结果合并与分割边界出现毛边的现象较少。图像 5—10 中,VGG16 + 边界预测方法和本文方法的分割结果更接近真值,其边界更准确可辨。

为进一步分析本文提出的多任务网络中的边界距离预测层对上述分割结果的直观视觉作用,可视化 VGG16 + 边界预测方法与本文方法的边界预测层的输出结果如表 4 所示。

表 4 边界距离预测层输出结果可视化

Tab. 4 Visualization of the boundary distance prediction layer output

图像 编号	真值	VGG16 + 边界预测	本文方法	边界距离 真值	图像 编号	真值	VGG16 + 边界预测	本文方法	边界距离 真值
1					6				
2					7				
3					8				
4					9				
5					10				

从表 4 可以看出,VGG16 + 边界预测方法和本文方法与边界距离真值非常接近,从 10 组图像中均可观察到完整的建筑地物边界。因此,这也是在同种框架下,加入边界预测层的多任务网络提取的遥感图像建筑地物优于单任务网络的直接原因;另外,加入边界预测层的多任务网络可充分发挥 Unet 网络提取建筑地物边界的优势,从而为分割结果预测层提供建筑地物边界的更多几何信息。需注意的是,在图像 3,4 和 10 的边界距离图中,本文方法要明显优于 VGG16 + 边界预测的方法,从而侧面验证

了具有更深编码与解码层的 Unet 网络具有较强的提取遥感图像建筑地物细节特征的能力。

2.5 效率分析

影响基于深度学习的语义分割方法时效性的 2 个关键因素为将所有训练样本输入到网络训练一个周期所需的时间和网络的收敛速率。为此,不同方法分别在训练集与验证集上进行测试时,其损失值随训练周期增加的变化趋势如图 2 所示。为了进行公平对比,5 种不同方法训练的优化均采用 Adam 算法。

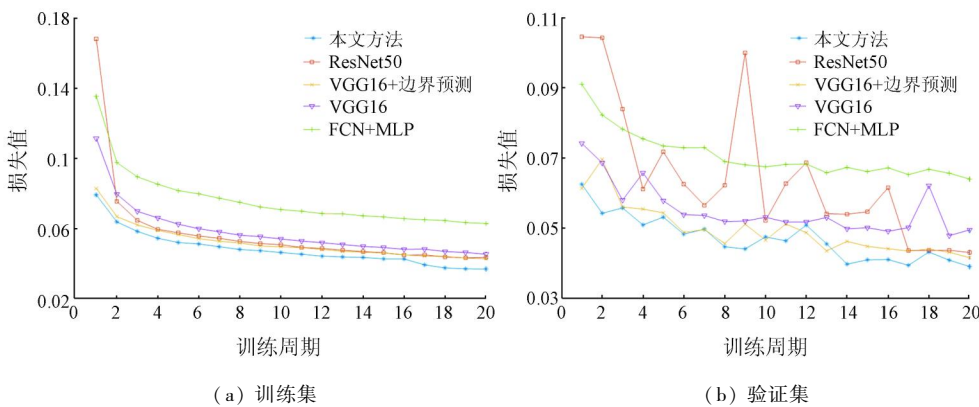


图 2 训练周期与损失值折线图

Fig. 2 Line chart of training period and loss value

从图 2(a)不同方法在训练集上的损失值变化趋势可以看出,FCN + MLP 基准方法的损失值降低速率最慢,而本文方法的损失值降低速率最快,表明本文方法可以在更少的训练周期内获得网络收敛。此外,ResNet50 和 VGG16 + 边界预测的损失值降低速率基本持平,亦表明加入边界预测层的多任务网络可使单一任务的 Unet 网络具有更好的性能。从图 2(b)在验证集上的损失值变化趋势可以看出,本文方法的曲线波动较小,表明本文方法可有效避免过拟合或欠

拟合现象,从而使网络收敛加快,减少网络训练时间。

3 实际应用

河南省新乡市高新区的覆盖范围达 52 km²,包含建筑、道路、植被、裸地和水体等多种地物类型。以我国高分 2 号卫星于 2018 年 4 月 16 日拍摄的该区域实际遥感图像为例,不同方法分割该实际遥感图像建筑地物的实验结果如图 3 所示。

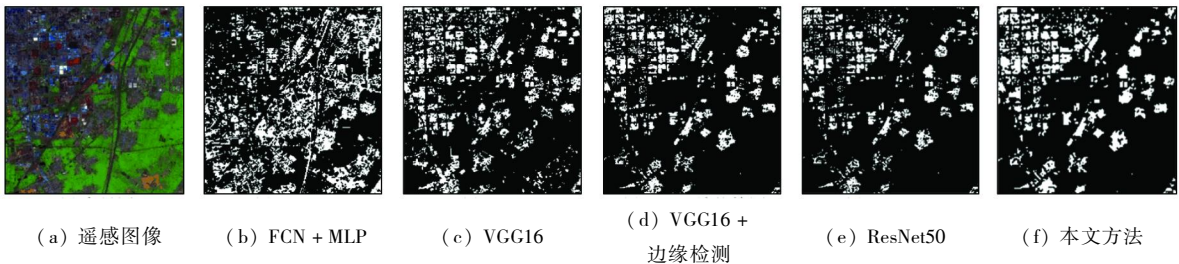


图 3 不同方法的实际遥感图像建筑地物分割结果

Fig. 3 Building object segmentation results of a factual remote sensing image by using different methods

结合图 3(a)遥感图像的目视解译,从图 3 可以看出 FCN - MLP 和 VGG16 方法存在过分割现象,究其原因是它们将部分道路或邻近建筑地物误分割为同一建筑地物,尤其 FCN + MLP 方法的误分割现象较为严重;VGG16 + 边界检测方法,较 VGG16 方

法建筑地物边缘部分更为准确;而 ResNet50 方法对建筑地物则存在欠分割现象,这是由于该方法对边缘检测不够准确的缘故。如上节所述,本文方法擅长准确地分割小规模建筑地物,故其对该覆盖范围较大、包含较多类型地物的实际遥感图像的建筑

地物分割较为准确。通过对图 3(f) 像素数目的计算,并与该研究区域 2017 年的统计年鉴资料对比,本文方法对该实际遥感图像建筑地物的分割准确度 Acc 值为 86.93%。

4 结 论

为实现高精度分割遥感图像中建筑地物,本文提出一种多任务学习的基于 ResNet50 搭建的 Unet 网络。该网络主要从 2 方面提高遥感图像的建筑地物语义分割结果:①采用更深层的 ResNet 网络搭建 Unet 网络;②使用级联多任务学习,使搭建的 Unet 网络可结合建筑地物的几何边界信息,输入到 FCN 进行有效语义分割。

实验结果表明,本文方法可使遥感图像建筑地物语义分割结果的 IoU 均值提高至 72.53%,Acc 均值提高至 96.10%,能够在一定程度上满足实际遥感图像的建筑地物分割的准确性与时效性要求。在实际应用中,对新乡市高新区遥感图像建筑地物的分割准确度达到 86.93%。但是,本文网络的深度仍然有限,其边界距离采用的是简单欧氏距离。为此,接下来拟采用 ResNet101 和 ResNet200 网络继续加深 Unet 网络的编码和解码层,以及采用马氏距离等生成边界距离预测图,来提升遥感图像的建筑地物语义分割精度。

参考文献 (References):

[1] Zhang B, Wang C, Shen Y, et al. Fully connected conditional random fields for high-resolution remote sensing land use/land cover classification with convolutional neural networks[J]. Remote Sensing, 2018, 10(12): 1889-1903.

[2] Li W, Dong R, Fu H. Large-scale oil palm tree detection from high-resolution satellite images using two-stage convolutional neural networks[J]. Remote Sensing, 2019, 11(1): 11-31.

[3] 张永宏, 夏广浩, 阚希, 等. 基于全卷积神经网络的多源高分辨率遥感道路提取[J]. 计算机应用, 2018, 28(7): 2070-2075.

Zhang Y H, Xia G H, Kan X, et al. Road extraction from multi-source high resolution remote sensing image based on fully convolutional neural network[J]. Journal of Computer Applications, 2018, 28(7): 2070-2075.

[4] Demir I, Koperski K, Lindenvaum D, et al. DeepGlobe 2018: A challenge to parse the earth through satellite images[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Salt Lake City: IEEE, 2018: 17201-17209.

[5] 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): 1350-1368.

[6] 施文灶, 刘金清. 基于邻域总变分和势直方图函数的高分辨率遥感影像建筑物提取[J]. 计算机应用, 2017, 37(6): 1787-1792.

Shi W Z, Liu J Q. Building extraction from high-resolution remotely sensed imagery based on neighborhood total variation and potential histogram function[J]. Journal of Computer Applications, 2017, 37(6): 1787-1792.

[7] Sun X, Lin X, Shen S, et al. High-resolution remote sensing data classification over urban areas using random forest ensemble and fully connected conditional random field[J]. ISPRS International Journal of Geo-Information, 2017, 6(8): 245-271.

[8] Jabri S, Zhang Y, Suliman A. Stereo-based building detection in very high resolution satellite imagery using IHS color system[C]//2014 IEEE Geoscience and Remote Sensing Symposium. Quebec City: IEEE, 2014: 2301-2304.

[9] Garcia-Garcia A, Orts-Escolano S, Oprea S, et al. A review on deep learning techniques applied to semantic segmentation[EB/OL]. (2017-04-22) [2019-02-05]. <http://arxiv.org/abs/1704.06857>.

[10] Yuan J. Automatic building extraction in aerial scenes using convolutional networks.[EB/OL]. (2016-02-21) [2019-02-20]. <http://arxiv.org/abs/1602.06564>.

[11] Zhang Q, Wang Y, Liu Q, et al. CNN based suburban building detection using monocular high resolution Google Earth images[C]//2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Beijing: IEEE, 2016: 661-664.

[12] Zhou B, Zhao H, Puig X, et al. Scene parsing through ade20k dataset[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, 2017: 633-641.

[13] Long J, Shelhaner E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, 2015: 3431-3440.

[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, 2017, 39(12): 2481-2495.

[15] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[EB/OL]. (2016-04-30) [2019-03-25]. <http://arxiv.org/abs/1511.07122>.

[16] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-12-05) [2019-03-26]. <http://arxiv.org/abs/1706.05587>.

[17] Maggiori E, Tarabalka Y, Charpiat G, et al. Can semantic labeling methods generalize to any city? The inria aerial image labeling benchmark[C]//2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Fort Worth: IEEE, 2017: 3226-3229.

[18] Maggiori E, Tarabalka Y, Charpiat G, et al. High-resolution semantic labeling with convolutional neural networks[EB/OL]. (2016-11-07) [2019-03-26]. <http://arxiv.org/abs/1611.01962>.

[19] Marmanis D, Schindler K, Wegner J D, et al. Classification with an edge: Improving semantic image segmentation with boundary detection[J]. ISPRS Journal of Photogrammetry and Remote Sensing,

2018,135:158 – 172.

[20] Peng C,Zhang X,Yu G,et al. Large kernel matters: Improve semantic segmentation by global convolutional network [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii,2017:4353 – 4361.

[21] Huang Z,Cheng G,Wang H,et al. Building extraction from multi – source remote sensing images via deep deconvolution neural networks [C]//2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Beijing: IEEE, 2016: 1835 – 1838.

[22] Bischke B,Helber P,Folz J,et al. Multi – task learning for segmentation of building footprints with deep neural networks [EB/OL]. (2017 – 09 – 18) [2019 – 03 – 26]. <http://arxiv.org/abs/1709.05932>.

[23] 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. Springer, Cham,2015:234 – 241.

[24] Iglovikov V,Shvets A. Terausnet: U – net with VGG11 encoder pre – trained on ImageNet for image segmentation [EB/OL]. (2018 – 01 – 17) [2019 – 03 – 29]. <http://arxiv.org/abs/1801.05746>.

[25] 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 – 162.

[26] 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. San Francisco,2016:770 – 778.

[27] Kendall A,Gal Y,Cipolla R. Multi – task learning using uncertainty to weigh losses for scene geometry and semantics [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City,2018:7482 – 7491.

[28] Russakovsky O,Deng J,Su H,et al. Imagenet large scale visual recognition challenge [J]. International Journal of Computer Vision,2015,115(3):211 – 252.

[29] Hayder Z,He X,Salzmann M. Boundary – aware instance segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii,2017:5696 – 5704.

[30] Kingma D P,Ba J. Adam: A method for stochastic optimization [EB/OL]. (2017 – 01 – 30) [2019 – 04 – 11]. <http://arxiv.org/abs/1412.6980>.

Multi – task learning for building object semantic segmentation of remote sensing image based on Unet network

LIU Shangwang^{1,2}, CUI Zhiyong^{1,2}, LI Daoyi^{1,2}

(1. College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China; 2. “Intelligent Business and Internet of Things Technology” Henan Engineering Laboratory, Xinxiang 453007, China)

Abstract: In order to accurately segment the building object of high – resolution remote sensing image, this paper proposes a multi – task learning method based on Unet network. Firstly, boundary distance map is generated from the ground – truth map of the building object remote sensing image; the boundary distance map, original remote sensing image and ground – truth map together are regarded as the input of Unet network. Then, based on the ResNet network, a multi – task network is built by adding the building object prediction layer and the boundary distance prediction layer at the end of the Unet network. Finally, the loss function of the multi – task network is defined, and the network is trained by using Adam optimization algorithm. Experiments on the Inria aerial remote sensing image building object dataset show that, compared with the full convolutional network combined with the multi – layer perceptron method, the intersection – over – unions of VGG16, VGG16 + boundary prediction, ResNet50 and this method have been increased by 5.15, 6.94, 6.41, and 7.86 percentage points, and the accuracy has been increased to 94.71%, 95.39%, 95.30%, and 96.10% respectively, which ensures that the building object of high – resolution remote sensing image can be segmented effectively.

Keywords: Unet network; multi – task learning; remote sensing image; semantic segmentation; ResNet network
(责任编辑: 陈 理)