

引文格式:贺浩,王仕成,杨东方,等.基于 Encoder-Decoder 网络的遥感影像道路提取方法[J].测绘学报,2019,48(3):330-338. DOI:10.11947/j.AGCS.2019.20180005.
HE Hao,WANG Shicheng,YANG Dongfang,et al.An road extraction method for remote sensing image based on Encoder-Decoder network[J]. Acta Geodaetica et Cartographica Sinica,2019,48(3):330-338. DOI:10.11947/j.AGCS.2019.20180005.

基于 Encoder-Decoder 网络的遥感影像道路提取方法

贺 浩¹,王仕成¹,杨东方¹,王舒洋²,刘 星¹

1. 火箭军工程大学控制工程系,陕西 西安 710025; 2. 火箭军工程大学信息工程系,陕西 西安 710025

An road extraction method for remote sensing image based on Encoder-Decoder network

HE Hao¹,WANG Shicheng¹,YANG Dongfang¹,WANG Shuyang²,LIU Xing¹

1. The Rocket Force University of Engineering, The Department of Control Engineering, Xi'an 710025, China; 2. The Rocket Force University of Engineering, The Department of Information Engineering, Xi'an 710025, China

Abstract: According to the characteristics of the road features, an Encoder-Decoder deep semantic segmentation network is designed for road extraction of remote sensing images. Firstly, as the features of the road target are rich in local details and simple in semantic features, an Encoder-Decoder network with shallow layers and high resolution is designed to improve the ability of representing detail information. Secondly, as the road area is small proportion in remote sensing images, the cross-entropy loss function is improved, which solves the imbalance between positive and negative samples in training process. Experiments on large road extraction dataset show that, the proposed method gets the recall rate 83.9%, precision 82.5% and F1-score 82.9%, which can extract the road targets in remote sensing images completely and accurately. The Encoder-Decoder network designed in this paper performs well in road extraction task and needs less artificial participation, so it has a good application prospect.

Key words: remote sensing; road extraction; deep learning; semantic segmentation;Encoder-Decoder network

Foundation support: The National Nature Science Foundation of China (Nos. 61403398; 61673017); The General Project of Shaanxi Nature Science Foundation (No. 2017JM6077)

摘 要:针对道路目标特点,设计实现了用于遥感影像道路提取的 Encoder-Decoder 深度语义分割网络。首先,针对道路目标局部特征丰富、语义特征较为简单的特点,设计了较浅深度、分辨率较高的 Encoder-Decoder 网络结构,提高了分割网络的细节表示能力。其次,针对遥感影像中道路目标所占像素比例较小的特点,改进了二分类交叉熵损失函数,解决了网络训练中正负样本严重失衡的问题。在大型道路提取数据集上的试验表明,所提方法召回率、精度和 F1-score 指标分别达到了 83.9%、82.5%及 82.9%,能够完整准确地提取遥感影像中的道路目标。所设计的 Encoder-Decoder 网络性能优良,且不需人工设计提取特征,因而具有良好的应用前景。

关键词:遥感;道路提取;深度学习;语义分割;编解码网路

中图分类号: P237 **文献标识码:** A **文章编号:** 1001-1595(2019)03-0330-09

基金项目:国家自然科学基金(61403398;61673017);陕西省自然科学基金面上项目(2017JM6077)

道路作为交通的主要组成部分,在人类各项活动中发挥着不可替代的作用。在现代社会中,道路也是地图和地理信息系统中重要的标识对象。随着交通地理信息系统的建设,道路的自动提取技术随之出现并不断发展^[1]。及时而完备的

道路交通信息系统,可在交通导航、城市规划、农林及飞行器自动驾驶等诸多领域发挥重要作用。人类观测手段的更新和交通建设的日新月异对道路自动提取技术不断地提出更高的要求。一方面,当今人类建设活动快速地改变着交通和地貌

信息,与此同时,人们对数字地图、导航等需求也十分旺盛。另一方面,人类观测地球的手段也更加丰富,高分辨率卫星、航拍无人机、成像雷达等设备的应用使得可用数据出现海量增长。在此情况下,如何利用海量的数据,快速且完备地提取道路信息以更新地理信息系统,成为该领域的研究热点。

在过去的几十年里,人们对遥感图像的道路自动提取技术进行了大量的研究。这些研究绝大多数都采用“特征匹配+道路连接”的思路^[2-10],即,先通过设计道路的一般特征,如灰度特征、平行双边缘、道路交叉口等,来确定道路种子的位置,并通过连接算法,如 Snake 算法等完成道路的提取。如文献[10]通过形状、距离、语义等设计度量特征,并使用支持向量机(SVM)训练道路网的回归模型以实现道路提取。这些方法在一定的应用场合下发挥了积极作用,但是其特征设计和连接算法中有诸多阈值参数需要人工调节,才能在某类型图片下取得较好效果,这就限制了其在大规模数据上的应用,使其难以实现通用^[11]。

随着深度学习技术在计算机视觉领域的开疆拓土,图像分类、物体识别和语义分割任务等都获得了重大突破。遥感影像的道路提取任务,可作为图像的语义分割任务的子集进行研究,即设计分割算法,将遥感影像中的道路与背景分割开来。随着深度语义分割研究的进展^[12-18],遥感影像道路提取技术的研究也不断深入。文献[11,19—20]通过深度神经网络对 Massachusetts 城市道路和建筑的标注及提取进行了大量研究,并建立了相应的道路提取数据集。文献[21]基于经典深度神经网络和滑动窗口来识别道路方向,并用有限状态机来连接道路,该方法采用了滑动窗口操作实现对像素区域的分类,因而其提取速度较为缓慢。文献[22]基于 FCN^[15] 网络架构,将道路结构信息约束到损失函数中,改进了深度分割网络在道路提取任务中的表现效果。文献[23]基于 SegNet^[17] 网络结构和 ELU 激活单元,进一步提升了道路提取效果,在 Massachusetts Roads 数据集^[20] 测试中其 F1-score 达到了 81.2%。但是这两种方法均未考虑道路特征自身特点,而使用通用的分割网络结构完成道路提取任务。

本文针对道路目标的特点,针对性地设计了深度卷积编解码网络结构(deep convolutional Encoder-Decoder network, DCED): 为保留道路

网络丰富的低层细节特征,增加了低层特征向高层的跳连,并减少 Encoder 过程中下采样次数,保证中间特征图的分辨率;针对道路语义信息较为简单的特点,压缩了 DCED 网络层数,减少了需要训练的参数数量,降低了网络训练难度。同时,针对遥感影像中道路目标所占像素较少的特点,针对性地改进了损失函数,改善了网络训练中正负样本严重失衡的问题。最后,通过试验验证了本文方法性能,并与其他算法进行了比较分析。

1 Encoder-Decoder 网络及训练

1.1 网络结构

道路是线形的,具有网状分布的特殊结构,其细节信息丰富,但是语义信息较为简单。这样的特点对分割网络的细节特征提取能力提出了较高要求。经典的语义分割网络面对的图像复杂多样,对于语义信息的提取要求更高,其 Encoder 部分大多采用了经典分类网络(如 VGG16)的预训练模型。在这些网络架构中,输入图片经过多次下采样,网络得到的中间特征图尺寸被多倍压缩,如文献[22]采用 FCN 架构,最小的中间特征图被压缩了 32 倍,因而丢失了目标的部分细节信息。随后这些网络虽然采用了“上采样+跳连”的方式来解码图片细节信息,但是对于道路目标,仍然不足以有效地还原道路网的细节。同时,道路目标语义信息简单的特点,也决定了道路提取任务并非必须采用大规模图片分类网络的预训练模型来进行特征提取。针对道路提取问题的特殊性,本文设计了深度相对较浅的 DCED 网络,其结构如图 1 所示。

将输入图片规范到 512×512 像素大小,经过若干层卷积和两次最大池化对图片完成编码部分,此时特征图 6 大小压缩为原图大小的 $1/4$,而后将特征图送入解码网络。在完成第 1 次上采样后,将原图 $1/2$ 大小的特征图 5 跳连至上采样层后,与特征图 10 并联送入后续网络。在完成第 2 次上采样后,将原图大小的特征图 2 跳连至本次上采样层后,与上采样输出的特征图 14 并联送入后续网络。而后,通过包含两个卷积核的卷积层将特征图映射到 $512 \times 512 \times 2$ 大小的特征图 17,实现输入的 RGB 图到输出二分类分割图的映射。而后通过包含一个卷积核,卷积核大小为 1×1 的卷积层,将输出映射成为像素分类的概率图,其大小为 $512 \times 512 \times 1$ 。

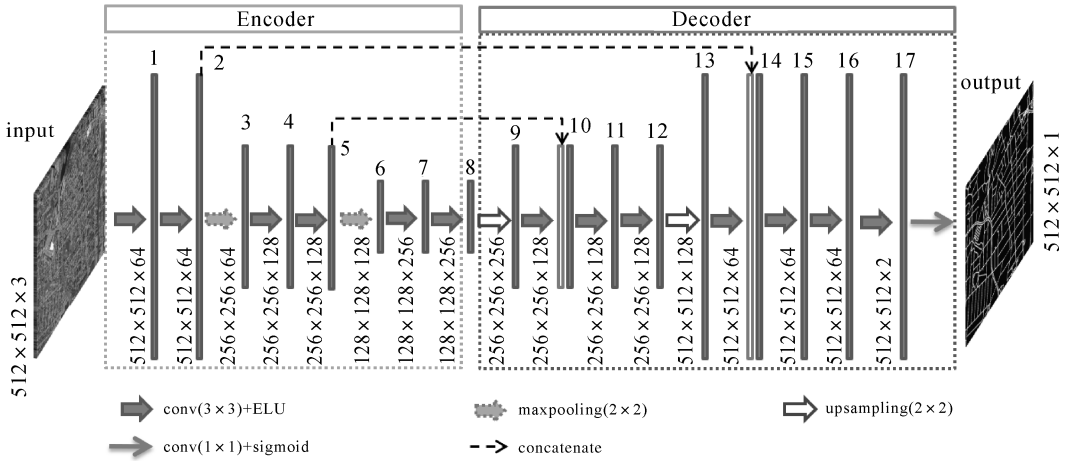


图 1 针对道路提取任务的 Encoder-Decoder 网络结构

Fig.1 The architecture of the designed Encoder-Decoder network for road extraction

本文所设计的网络结构只经过两次下采样处理,其最小特征图仍然具有原图 1/4 的尺寸,因而能够更多地保留道路的局部细节特征。同时,在较浅的网络结构下,通过跳连操作将低层局部特征与高层语义特征融合起来,使得最终输出的特征图对道路网的语义信息及细节信息均具有良好的表示能力。网络中使用到的方法具体情况如下。

1.1.1 卷积

除去用于输出的 1×1 卷积层,其他卷积层卷积核大小均为 3×3 ,并设置 padding 为 1,使得卷积层输入特征图与输出特征图的宽、高一一致。

1.1.2 ELU 激活

在过去一段时间里,修正线性单元(rectified linear unit,ReLU)激活在深度卷积神经网络中被广泛采用。文献[24]提出了指数线性单元(exponential linear unit,ELU),可以加速网络收敛速度,有效克服梯度消失等问题。假设某节点输出为 x_i ,则经过 ELU 层后的输出 r_i 可表示为

$$r_i = \begin{cases} a(e^{x_i} - 1) & x_i < 0 \\ x_i & x_i \geq 0 \end{cases} \quad (1)$$

1.1.3 最大池化

池化操作可以汇合低层特征信息,缩减计算数据量,扩大高层滤波器的感受野。本网络的池化操作均采用最大池化,尺寸设置为 2×2 ,步长也为 2×2 。经过最大池化层后的特征图,其输出的深度不变,高和宽变为原来的 1/2。

1.1.4 跳层连接

在所设计的网络中,特征图经过卷积层后其宽和高是不变的。因此,经过若干卷积层,两次下

采样与两次上采样的特征图 14 与特征图 2 的尺寸是一致的。同理,特征图 10 与特征图 5 的尺寸也是一致的。所谓跳连操作,即将低层特征图与相应高层特征图整合为新的张量,并进行后续的计算和处理。其具体的实现方式为将特征图 2 ($512 \times 512 \times 64$)与特征图 14 ($512 \times 512 \times 64$)在深度维度上串联,构成新的组合特征图 ($512 \times 512 \times 128$)并送入后续卷积层进行运算。同理,对特征图 5 与特征图 10 也进行相似的操作。跳连操作能够融合高层的语义信息和低层的局部特征信息,从而实现准确而又精细的道路特征提取。

1.1.5 上采样

与池化操作相反,对较小的特征图进行上采样,其输出的特征图深度不变,高和宽变为原来的 2 倍。本文采用了双线性插值实现上采样。通过如下方式计算上采样后某一点 (x, y) 的值 P :假设输入的特征图在相邻 4 个像素 (x_1, y_1) 、 (x_1, y_2) 、 (x_2, y_1) 和 (x_2, y_2) 处的值分别为 Q_{11} 、 Q_{12} 、 Q_{21} 和 Q_{22} 。首先在 x 方向进行线性插值

$$R_1 \approx \frac{x_2 - x}{x_2 - x_1} Q_{11} + \frac{x - x_1}{x_2 - x_1} Q_{21} \quad (2)$$

$$R_2 \approx \frac{x_2 - x}{x_2 - x_1} Q_{12} + \frac{x - x_1}{x_2 - x_1} Q_{22} \quad (3)$$

式中, R_1 和 R_2 分别代表输出图在 (x, y_1) 和 (x, y_2) 位置的值。然后在 y 方向进行线性插值,可得

$$P \approx \frac{y_2 - y}{y_2 - y_1} R_1 + \frac{y - y_1}{y_2 - y_1} R_2 \quad (4)$$

联立式(2)、式(3)、式(4)可得

$$P \approx \frac{(x_2 - x)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} Q_{11} + \frac{(x - x_1)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} Q_{21} +$$

$$\frac{(x_2-x)(y-y_1)}{(x_2-x_1)(y_2-y_1)}Q_{12} + \frac{(x-x_1)(y-y_1)}{(x_2-x_1)(y_2-y_1)}Q_{22} \quad (5)$$

本文将特征图上采样为原来大小的2倍,则式(5)可以简化为

$$P = \frac{1}{4}(Q_{11} + Q_{12} + Q_{21} + Q_{22}) \quad (6)$$

1.1.6 Sigmoid

经过该操作,输出为

$$a_i = \frac{1}{1 + e^{-z_i}} \quad (7)$$

式中, z_i 为上层节点输出值。输出值介于0与1之间,可作为二分类预测的概率描述。

1.2 道路结构约束的损失函数改进

道路提取问题可看作是对像素的二分类问题。对于二分类问题,TP(true-positive)代表标签为正,预测也为正;TN(true-negative)代表标签为负,预测为负;FP(false-positive)代表标签为负,预测为正;FN(false-negative)代表标签为正,预测为负。

通常此类问题采用二分类交叉熵损失函数:

$$L = -\frac{1}{n} \sum_i (y_i \log a_i + (1 - y_i) \log(1 - a_i)) \quad (8)$$

式中, y_i 为第 i 个像素的真实值, $y_i = 0$ 代表该像素属于背景区域, $y_i = 1$ 则代表该像素为道路区域; a_i 为经过 sigmoid 函数第 i 个像素的预测值, a_i 取值落于(0,1)内, a_i 越趋近于1,代表其属于道路的概率越高。训练的过程即通过不断调整网络的权值参数,使得损失函数 L 达到最小的过程。

可以看出,式(8)无差别对待输出 y_i , 给予了所有像素相同的权重,因而忽略了 y_i 的位置与其分类之间的联系。基于此观点,文献[22]将损失函数改进为

$$L = -\frac{1}{n} \sum_i (y_i \log a_i + e^{-f(d_i)} (1 - y_i) \log(1 - a_i)) \quad (9)$$

式中, $f(d_i)$ 的定义为

$$f(d_i) = \begin{cases} 0 & d_i = 0 \\ \frac{d_i}{\max_{j \in I} \{d_j\}} & 0 < d_i < T \\ \frac{T}{\max_{j \in I} \{d_j\}} & d_i > T \end{cases} \quad (10)$$

d_i 表示第 i 个像素距离道路区域的最小欧氏距

离,需要在训练前从样本标签中提取得到。 T 是判断某像素距离道路是否足够远的阈值,文献[22]将其设置为 $0.3 \max_{j \in I} \{d_j\}$ 。式(9)的设计对道路附近的像素施加了更大的负样本惩罚,使得道路结构的连续性能对损失函数产生影响。分析式(9)及式(10)可以发现,该损失函数对负样本作了动态的加权处理,其权值落于 $[e^{-0.3}, 1]$ 内。这样的处理是有效的,但是在一次计算损失函数中,背景像素占据了画幅的绝大部分,式(10)的加权对损失函数的影响较为微弱,改进提取效果不够明显。遥感影像中的道路目标为网状结构,其像素在图片中所占比例较小,如图2所示。统计 Massachusetts Roads 数据集 1108 张训练样本,其正样本平均比例为 4.77%,其中最小的只有 1.04%,最大的也不足 20%。在这样正负样本严重不均衡的情况下,在损失函数中给予正负样本相同的权重是不合理的;正样本所占比例较小,其带来的有效信息在损失函数中被平均后变得不显著,因而使得训练效率下降。因此,本文重新设计加权交叉熵损失函数为

$$L = -\frac{1}{n} \sum_{i=1}^n (\lambda_1 y_i \log a_i + \lambda_2 e^{-f(d_i)} (1 - y_i) \cdot \log(1 - a_i)) \quad (11)$$

式中,比例系数 λ_1 和 λ_2 均为正数且有 $\lambda_1 > \lambda_2$ 。对于二分类问题而言,正负样本的合理比例约为 1:1。



(a) 遥感影像



(b) 道路标注

图2 遥感影像与其道路标注图

Fig.2 The remote sensing images and the ground truth of the road

针对每一份训练样本,统计其正样本数目 N_p 和负样本数目 N_n , 设置加权系数使得其满足 $\lambda_1 + \lambda_2 = 2$ 且满足

$$\frac{\lambda_1 \times N_p}{\lambda_2 \times N_n} \approx P \quad (12)$$

可得

$$\left. \begin{aligned} \lambda_1 &= \frac{2N_n}{PN_p + N_n} \\ \lambda_2 &= \frac{2N_p}{N_p + PN_n} \end{aligned} \right\} \quad (13)$$

一般而言,二分类问题中,正负样本所占比例应大致相等,因而设置常数 $P \approx 1$,但是 P 值的设置还与正负样本特征的多样性、复杂程度等多种因素相关,在应用中应进一步深入研究。

1.3 网络的训练

1.3.1 数据扩充

深度学习是一种数据驱动技术,要训练性能

优异的深度网络,数据量是其基础。同时,目前规范的遥感影像道路数据集容量还不能绝对满足网络对数据的需求。在目标样本量不足的情况下,一种解决方案是利用迁移学习,使用大规模图片识别预训练模型的强大表示能力来提取图片特征,另一种思路则是通过对样本图像的处理和变换来扩充训练数据容量。

本文通过对道路影像的多角度旋转和镜像映射来扩充训练数据集。如图 3 所示,通过旋转以及上下、左右的镜像映射,可用的训练数据量扩充为原来的 6 倍。

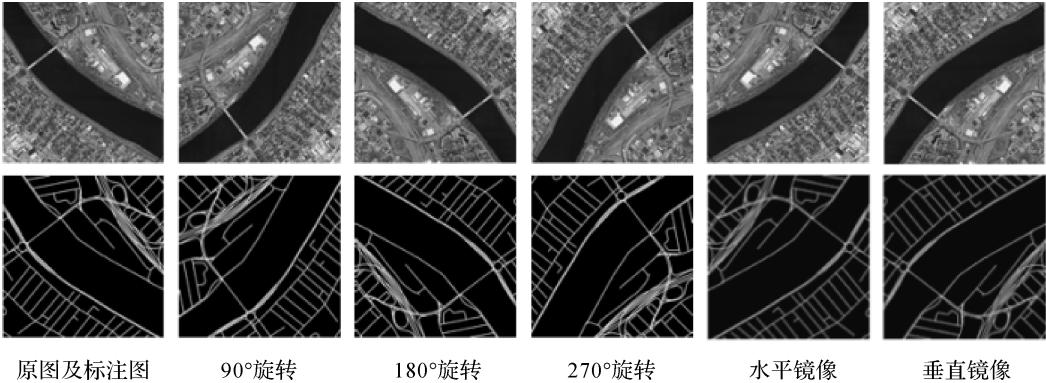


图 3 扩充训练数据

Fig.3 The augmentation of the training data

1.3.2 样本预处理

在训练深度网络之前,对输入图片进行规范化是必要的,本文对输入图片进行了缩放、归一化和去均值处理。其中,归一化操作是将输入的 RGB 图 I 从 $[0,255]$ 规范到 $[0,1]$ 区间,对于输入的任一像素 $x_{i \in I}$,该操作可表示为 $\bar{x}_i = x_i / 255$,其中 \bar{x}_i 为归一化后相应的输入值。去均值处理即将归一化后的每一幅输入图片减去其平均值,处理后的输入分布于 $[-0.5,0.5]$ 。记归一化后图片为 \bar{I} ,去均值化后输入 $\tilde{x}_i = \bar{x}_i - \text{mean}(\bar{I})$, \tilde{x}_i 为去均值化后对应像素的输入值。0 均值归一化处理使得网络输入限制在一定范围内,避免了训练样本差异造成的训练过程不收敛,加快了梯度下降优化过程。

1.3.3 优化算法配置

深度学习网络的训练,一方面需要足够的计算资源和大量训练数据,另一方面也需要配置合理的超参数,采用有效的优化算法。本文采用了 Adam 优化算法^[26]来训练网络参数。该算法可利用梯度的一阶矩估计及二阶矩估计动态地调整

每个参数的学习率。

Adam 算法:

```
require:  $\alpha$ , 步长
require:  $\beta_1, \beta_2 \in [0,1)$  # 矩估计的指数衰减速率
require:  $f(\theta)$  # 基于参数  $\theta$  的损失函数
require:  $\theta_0$  # 初始参数向量
require:  $\epsilon$  # 较小的常数
 $m_0 = 0$  # (初始化一阶矩向量)
 $v_0 = 0$  # (初始化二阶矩向量)
 $t = 0$  # (计数器归 0)
while  $\theta_t$  未收敛 do
     $t = t + 1$ 
     $g_t = \nabla_{\theta} f_t(\theta_{t-1})$  # 从损失函数获取梯度
     $m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$  # 更新有偏一阶矩估计
     $v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$  # 更新有偏二阶矩估计
     $\hat{m}_t = m_t / (1 - \beta_1^t)$  # 修正一阶矩估计偏差
     $\hat{v}_t = v_t / (1 - \beta_2^t)$  # 修正二阶矩估计偏差
     $\theta_t = \theta_{t-1} - \alpha \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon)$  # 更新参数
end while
return  $\theta_t$ 
```

这里,根据实际情况,配置基础学习率 $\alpha =$

10^{-4} , $\beta_1=0.9$, $\beta_2=0.999$, $\epsilon=10^{-8}$ 。另外,受限于计算资源,本文设置批处理大小(batch size)为 6。

2 试验及结果分析

试验数据集采用目前最大的遥感影像道路数据集,Massachusetts Roads。该数据集采用覆盖了美国马萨诸塞州超过 2600 km² 面积的卫星遥感图像,图片为 1500×1500 像素的 RGB 图,地面分辨率约 1 m/像素。数据集包含训练集 1108 张卫星照片及其对应标注图,测试集 49 张卫星照片及其对应标注图,验证集 14 张卫星照片及其对应标注图,每一幅卫星照片中都包含有道路目标。在训练集中,为了增强数据的容错性,数据集的制作者故意对一些照片添加了遮挡。网络的训练采用了数据集中训练集及增强数据共 6648 张照片及对应标注图,并将原始图片从 1500×1500 像素缩放为 512×512 像素,使用其测试集共 49 张照片及其标注图对 DCED 网络性能进行评估。

2.1 提取结果的评价方法

当作语义分割任务来看,道路提取所得的结果也通常采用语义分割的评价方法,即准确率(accuracy),召回率(recall),F1-score 和精度(precision)等。其中,准确率定义为 $accuracy = (TP+TN)/(TP+TN+FP+FN)$,召回率的定义为 $recall = TP/(TP+FN)$,F1-score 定义为 $F1 = 2TP/(2TP+FN+FP)$,精度定义为 $precision = TP/(TP+FP)$ 。

从经典的道路提取方法来看,文献[27—28]采用道路的完整度(completeness)、正确度(correctness)和质量(quality)等作为指标。完整度的定义为 $completeness = TP/(TP+FN)$,正确度的定义为 $correctness = TP/(TP+FP)$,质量的定义为 $quality = TP/(TP+FP+FN)$ 。

可以看出这两种评价体系基本观点的类似,本文采用召回率(完整度)和精度(正确度)及 F1-score 作为评价指标。

2.2 试验结果及分析

2.2.1 对比试验结果及分析

在 Massachusetts Roads 数据集下,通过与文献[18—23]所提方法进行对比,证明了本文所提方法的良好性能。U-net^[18]主要运用于医疗影像分割,其应用对象,如血管阴影等,与道路目标结构相似,因而也可用于遥感影像的分割。与一般的深度分割网络类似,U-net 也使用了包含跳

连、上采样和全卷积的 Encoder-Decoder 架构。虽然 U-net 并未使用通用的预训练模型,但是其基本结构与一般分割模型并无大的区别,其下采样缩小倍率为 32 倍,最小特征图为 28×28,对于遥感道路网络的细节分辨率不足。RSRCNN^[22]的基本网络架构是基于 FCN 实现的,其 Encoder 部分采用了 VGG16 预训练模型,对输入进行了 4 次下采样,缩小 32 倍,最小特征图仅为 7×7,因而其细节表达存在缺陷。ELU-SegNet^[23]引入了更加先进的分割网络结构 SegNet 和激活函数 ELU,使得道路分割结果大为改善,在对训练数据进行了扩充后,ELU-SegNet-R 取得了更好的效果。但是 ELU-SegNet 仍然基于 SegNet 网络基本架构,其 Encoder 部分使用了 VGG16 预训练模型,与 RSRCNN 存在同样的问题。

使用 Massachusetts Roads 数据集中 49 张测试图片进行了验证,所得结果指标如表 1 所示。

U-net 网络难以获得令人满意的效果,可见其模型结构设计并不能完全适应遥感影像的道路提取任务。在使用 FCN 架构情况下,RSRCNN 通过对损失函数进行改进,使 F1-score 达到了 66.2%,证明了其对损失函数的改进是有效的。ELU-SegNet-R 则由于采用了更加先进的网络结构和激活函数,并扩充了训练数据,将 F1-score 提到了 81.2%;但是 ELU-SegNet-R 的改进并没有考虑道路本身的特征特点,而直接利用了适合于街景分割的 SegNet 基本架构,因而限制了其提取道路网络的性能。

表 1 不同方法的道路提取结果对比

Tab.1 The experimental results of road extraction in different methods

方法	recall	precision	F1-score
U-net ^[18]	0.632	0.541	0.545
RSRCNN ^[22]	0.729	0.606	0.662
ELU-SegNet-R ^[23]	0.847	0.780	0.812
本文方法(网络结构)	0.831	0.814	0.822
本文方法(网络结构+改进损失函数)	0.839	0.825	0.829

由于本文方法的网络结构适应了遥感影像的特征特点,在不引入损失函数改进的情况下,本文对网络结构的改进将 F1-socre 提高到了 82.2%,并且在细节辨识能力上明显优于其他方法。在改进的网络结构基础上,引入所提出的加权交叉熵损失函数,进一步提高了对道路网络的提取能力,

使得 F1-score 进一步达到了 82.9%。除了结果指标有所提升,道路边缘区域预测的置信度也显著高于未改进之前,提高了约 10%,证明了对损失函数的改进可以改善数据样本不均衡带来的影响。

本文所提方法除了在以上数据指标上表现更好,其优势更在于能够保留道路细节信息,实现更加精确清晰的道路提取。对预测输出进行可视化来比较本文与文献[22—23]的道路提取效果,如图 4 所示。

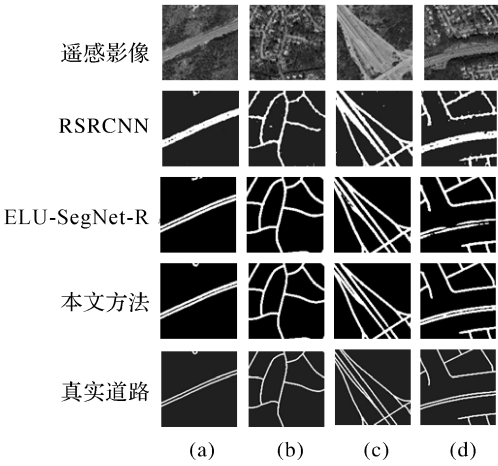


图 4 道路提取结果的可视化对比

Fig.4 Subjective results of road extraction

2.2.1.1 在细节表示能力方面

在图 4(a)、(c)、(d)中,RSRCNN 方法受限于网络结构不足,其细节提取能力较差,不能有效地分离双车道,而 ELU-SegNet-R 及本文方法能够结合低层特征,有效地提取了双车道信息;在图 4(a)中,由于网络结构过深,对细节特征分辨率不足,RSRCNN 及 ELU-SegNet-R 均不能准确地提取图片顶端的圆形道路区域,其结果形状和结构均存在缺陷。本文方法可以完整准确地提取该目标。

可以看出,本文方法适应了道路特征特点,因而能够准确有效地提取道路局部特征信息,在细节表示能力方面明显优于对照方法。

2.2.1.2 在提取结果准确性方面

在图 4(b)、(c)、(d)中,RSRCNN 方法存在较多误检测区域,野值点明显多于 ELU-SegNet-R 及本文方法,同时,在所有图例中,RSRCNN 方法所提取的道路结构都存在着边缘较为粗糙的问题。在图 4(c)、(d)中,ELU-SegNet-R 出现了道

路结构缺损的情况。相较于对照方法,本文方法所得结果野值点较少,且道路结构清晰完整。

相较于对照方法,本文方法分辨率高,较好地保存了道路结构细节信息,因而能够有效避免周围背景环境干扰,准确地提取道路结构。可以看出,本文方法在结果的准确性方面优于其他方法。

2.2.2 试验结果的适应性分析

本文方法是一种有监督的机器学习方法,因此其适应性能依赖于训练数据的多样性。对于道路提取任务而言,训练样本能够涵盖同类型图片中大部分道路,因此本文方法在同类数据中具有较好的适应性能。例如,遥感影像或航拍照片中的道路存在有周围建筑物、行道树阴影遮挡的问题,这对传统的道路提取方法造成了严重的干扰,而本文方法则不受影响。

试验所采用的测试数据中包含了部分遮挡较为明显的样本,某些样本只是道路的边缘及颜色特征发生了改变,而极端的例子中道路被树木所阻断,如图 5 中遥感影像所示。

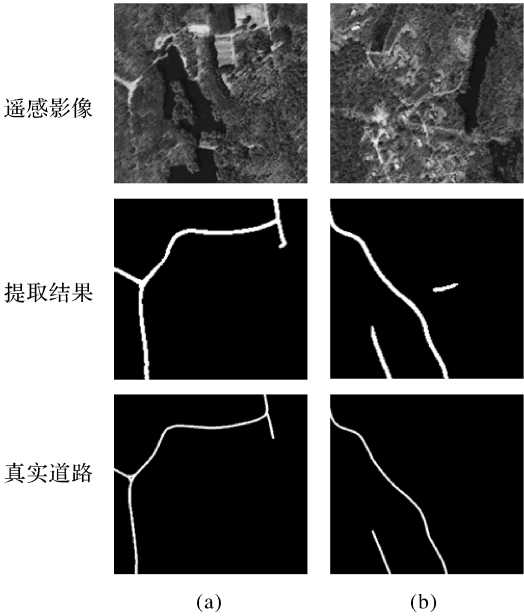


图 5 被遮挡道路区域提取效果

Fig.5 The extraction of the obscured road area

通过图 5 的提取结果可以看出,本文方法对于被树木遮挡的道路也有良好的提取效果,甚至于被树木阻断的道路结构也能有效提取。这是由于训练样本中包含了被树荫遮挡的道路,在训练过程中,网络自动学习了这部分道路的特征,因而能够在测试样本中提取具有相同特征的道路区

域。这证明了本文方法对于复杂环境下的道路提取具有良好的适应性能。

3 结 论

针对遥感影像中,道路目标细节丰富且语义较为简单的特点,本文针对性地设计了层数相对较浅、细节分辨率较高的 Encoder-Decoder 深度分割网络,相较一般方法,更多地保留了道路的局部特征信息,提高了道路提取的细节识别能力;针对目标图像正负样本严重不均衡的情况,改进了二分类交叉熵损失函数,改善了正样本对损失函数影响不够的问题。通过对训练数据进行旋转和镜像映射扩充了样本容量,在此基础上训练并利用试验验证了 DCED 网络性能。

试验表明,本文所提的 Encoder-Decoder 网络在召回率、精度和 F1-score 指标上均表现优异,所提取道路结构完整清晰,且具有良好的适应性能,可以有效地从遥感影像中提取道路网络。相较于传统的“特征提取+道路连接”的方法,深度分割网络在应用阶段不需要人工设计特征和预设阈值参数,并且在特定应用场合下泛化性能好,能够经受大量样本检验,因而具有良好的应用前景。

同时,深度学习技术的性能依赖于训练数据,其特征都是从训练数据中学习得到的,因而其泛化性能受到训练数据集多样性的制约。在单一类背景和视场下训练得到的网络并不能完全适应新的应用场合。对于本文的应用场合而言,同一组卫星或飞行器对一定区域获取的影像在视场和背景上均相似,其特征能够从训练样本中学习得到,因此适应性能良好。如需在新的应用场合使用深度分割网络,一般需将该类型数据加入训练集并对网络权重值进行训练微调。

深度语义分割是从像素域对道路目标进行提取的,相应的结果评价方法也是基于像素域的,因此缺乏对道路拓扑结构完整性和准确性的表示能力。下一步的研究可以基于道路的拓扑结构信息,设计和实现深度网络提取道路的拓扑结构,与本文所提取的分割结果进行融合,以达到更加精确和完整的道路提取成果。

参考文献:

[1] 吴亮,胡云安. 遥感图像自动道路提取方法综述[J]. 自动化学报, 2010, 36(7): 912-922.

WU Liang, HU Yunan. A survey of automatic road extraction from remote sensing image[J]. Acta Automatica Sinica, 2010, 36(7): 912-922.

[2] ZHOU Y T, VENKATESWAR V, CHELLAPPA R. Edge detection and linear feature extraction using a 2-D random field model[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, 11(1): 84-95.

[3] STEGER C, GLOCK C, ECKSTEIN W, et al. Model-based road extraction from images[C]// Proceedings of Automatic Extraction of Man-Made Objects from Aerial and Space Images. Basel, Schweiz: Springer, 1995: 275-284.

[4] STEGER C. An unbiased detector of curvilinear structures[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(2): 113-125.

[5] TUPIN F, MAITRE H, MANGIN J F, et al. Detection of linear features in SAR images: application to road network extraction[J]. IEEE Transactions on Geoscience and Remote Sensing, 1998, 36(2): 434-453.

[6] SHI Wenzhong, ZHU Changqing. The line segment match method for extracting road network from high-resolution satellite images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2002, 40(2): 511-514.

[7] HU Jiuxiang, RAZDAN A, FEMIANI J C, et al. Road network extraction and intersection detection from aerial images by tracking road footprints[J]. IEEE Transactions on Geoscience and Remote Sensing, 2007, 45 (12): 4144-4157.

[8] CHRISTOPHE E, INGLADA J. Robust road extraction for high resolution satellite images[C]// Proceedings of 2007 IEEE International Conference on Image Processing. San Antonio, TX: IEEE, 2007, 5: 437-440.

[9] POULLIS C, YOU Suya. Delineation and geometric modeling of road networks[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2010, 65(2): 165-181.

[10] 付仲良, 杨元维, 高贤君, 等. 道路网多特征匹配优化算法[J]. 测绘学报, 2016, 45(5): 608-615. DOI: 10.11947/j.AGCS.2016.20150388.

FU Zhongliang, YANG Yuanwei, GAO Xianjun, et al. An optimization algorithm for multi-characteristics road network matching[J]. Acta Geodaetica et Cartographica Sinica, 2016, 45(5): 608-615. DOI: 10.11947/j.AGCS.2016.20150388.

[11] MNIH V, HINTON G E. Learning to detect roads in high-resolution aerial images[C]// Proceedings of the 11th European Conference on Computer Vision. Berlin: Springer, 2010: 210-223.

[12] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH: IEEE, 2014: 580-587.

- [13] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//Proceedings of the 3rd International Conference on Learning Representations, San Diego:[s.n.], 2015.
- [14] SZEGEDY C, LIU Wei, JIA Yangqing, et al. Going deeper with convolutions[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA; IEEE, 2015: 1-9.
- [15] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [16] NOH H, HONG S, GAN B. Learning deconvolution network for semantic segmentation[C]//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile; IEEE, 2015: 1520-1528.
- [17] 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.
- [18] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]//Proceedings of Medical Image Computing and Computer-Assisted Intervention, Cham; Springer, 2015, 234-241.
- [19] MNIH V, HINTON G E. Learning to label aerial images from noisy data[C]//Proceedings of the 29th International Conference on Machine Learning, Edinburgh, Scotland; ACM, 2012: 567-574.
- [20] MNIH V. Machine learning for aerial image labeling[D]. Toronto: University of Toronto, 2013.
- [21] WANG Jun, SONG Jingwei, CHEN Mingquan, et al. Road network extraction: a neural-dynamic framework based on deep learning and a finite state machine[J]. International Journal of Remote Sensing, 2015, 36(12): 3144-3169.
- [22] WEI Yanan, WANG Zulin, XU Mai. Road structure refined CNN for road extraction in aerial image [J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(5): 709-713.
- [23] PANBOONYUEN T, VATEEKUL P, JITKAJORNWANICH K, et al. An enhanced deep convolutional Encoder-Decoder network for road segmentation on aerial imagery [C]//Proceedings of Recent Advances in Information and Communication Technology 2017. Bangkok: Springer, 2017: 191-201.
- [24] CLEVERT D A, UNTERTHINER T, HZOCHREITER S. Fast and accurate deep network learning by exponential linear units (ELUs) [C]//Proceedings of International Conference on Learning Representations, San Juan: [s.n.], 2016.
- [25] CHENG Guangliang, WANG Ying, XU Shibiao, et al. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network [J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(6): 3322-3337.
- [26] KINGMA D P, BA J L. ADAM: a method for stochastic optimization[C]//Proceedings of International Conference on Learning Representations, San Diego:[s.n.], 2015.
- [27] WIEDEMANN C. External evaluation of road networks [J]. Proceedings of the ISPRS Workshop on Photogrammetric Image Analysis, Munich, Germany, 2003, 34(3): 93-98.
- [28] WEGNER J D, MONTOYA-ZEGARRA J A, SCHINDLER K. A higher-order CRF model for road network extraction [C]//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR; IEEE, 2013: 1698-1705.
- [29] ZHONG Zilong, LI J, CUI Weihong, et al. Fully convolutional networks for building and road extraction: preliminary results [C]//Proceedings of 2016 IEEE International Geoscience and Remote Sensing Symposium, Beijing, China; IEEE, 2016: 1591-1594.

(责任编辑:陈品馨)

收稿日期: 2018-01-08

修回日期: 2018-11-15

第一作者简介: 贺浩(1991—),男,博士生,研究方向为深度学习和计算机视觉。

First author: HE Hao (1991—), male, PhD candidate, majors in deep learning and computer vision filed.

E-mail: hehao209@126.com

通信作者: 杨东方

Corresponding author: YANG Dongfang

E-mail: yangdf301@163.com