

基于拓扑损失的遥感图像植被信息提取

高 昊¹ 阴桂梅¹

GAO Hao YIN Guimei

摘 要

绿色植被在城市的生态系统中扮演着非常重要的角色，非常有必要对城市植被进行合理的规划与检测。近年来，深度学习技术，尤其是深度卷积神经网络（DCNNs）在高分辨率遥感图像的解析和处理方面取得了显著进展。因为高分辨率遥感图像成本低、周期短和精度高等特点，为城市植被信息提取提供了可能性。但是，高分辨率的城市遥感图像中伴随着比较复杂的背景信息，增加了图像预测的难度。将拓扑数据分析方法（持续同调）通过拓扑损失函数和拓扑先验的方法引入到 UNet 模型中，提高模型的结构感知能力。实验结果表明，不同分割模型中进行的实验在一定程度上均取得了良好的表现。

关键词

植被；遥感；持续同调；拓扑数据分析；深度学习

doi: 10.3969/j.issn.1672-9528.2024.02.024

0 引言

城市植被指的是覆盖在城市地表上的所有植物种类，包括自然种类和人工栽培种类，是城市生态系统中唯一的初级生产者，具有重要的自净、美化城市景观，丰富城市建筑轮廓线、提供休闲娱乐和防灾减灾等作用^[1]。所以，对城市的植被信息进行有效检测，对城市植被规划和城市绿化率的监测起着至关重要的作用。目前，主要是通过人工调查和基于遥感的调查这两种方式来获取城市植被信息。但其成本高、周期长且可靠性较差是人工调查的主要缺点，而利用高空遥感图像可以较好地解决以上问题。

事实上，从计算机图像处理的角度来看，高分辨率遥感图像的城市植被信息提取这一类问题的本质是用深度学习网络模型处理 RGB 图片的语义分割问题。计算机视觉中的重要任务是分割出图像物体的位置和形状，本质是哪个像素属于哪个物体。这项任务是通过图像分割来完成的，其中包括为所有输入图像像素分配标签。然而，在涉及精细结构的应用中，如神经元膜和血管中的薄连接，精度是不够令人满意的，因为它可能导致非常严重的错误。为了解决这一问题，在 UNet 模型中引入了拓扑损失函数。在本文中，研究了引入拓扑的两种方法。第一种是在分割模型中常用的交叉熵损失的基础上加入持续同调，构成拓扑损失函数。拓扑损失度量真实值和预测值之间的差异。第二种方法是在训练神经网络之前使用拓扑处理数据。这两种策略都能够在相对低像素的情况下保持被分割物体的拓扑结构。

1 相关工作

1.1 基于深度学习的植被提取研究方法

卷积神经网络（CNN）作为很多深度学习模型的一个基础算子，在提取城市高分辨率遥感图像中的植被信息方面已经被广泛研究并应用。Kattenborn 等人^[2]通过大量的研究和实验发现，与传统的机器学习方法相比，深度学习中的卷积神经网络如果用在遥感图像中的植被信息提取效果更好。为了解决模糊和无定形形状的植被提取问题，Watanabe 等人^[3]采取了在当时看来是一种创新的方法，把要处理的高分辨率遥感图像剪裁成很小的方块，使得用来训练的数据集数量有了量的提升，然后使用深度学习中的卷积神经网络较好地从高空拍摄的城镇遥感图像中提取了部分有效绿色植被信息。王修信^[4]在配合实际情况的同时使用了多源头、多特征的方法，然后将支持向量机与卷积神经网络相结合，提取了森林植被的相关信息，提高了喀斯特地貌中森林植被信息的提取能力。

在 2015 年，U-Net 网络由 Ronneberger 等人^[5]提出，它通过跳接的网络结构更充分地融合了浅层特征与深层特征。这种网络结构兼具轻量化与高性能，它的出生是为了解决医疗影像语义分割问题，即使在数据集不充足的情况下，也能够达到较好的图像分割效果同时避免了过拟合。然而，由于其网络结构简单，很多人将其用于高分辨率遥感图像中植被、建筑物的提取。Alshehhi 等人^[6]提出了一种新型的多尺度对抗生成网络的深度学习模型，这个模型首先在 U-Net 架构中使用了 EfficientNet-B3 作为生成器，然后使用 ResNet18 作为输出结果的判别器，最后分别使用 Potsdam 和 Vaihingen 数据集对高分辨率遥感图像中的城市地表物体

1. 太原师范学院计算机科学与技术学院 山西晋中 030619

分成不同类别,这种网络模型在提取植被信息方面相比传统方法效果更好。蔡淑颖等人^[7]使用改进的 U-Net 模型和面向对象多尺度分割技术,提取了杭州部分市区高分辨率遥感图像的城市森林信息。徐知宇等人^[8]为了解决小样本训练中容易出现过拟合的问题,在 U-Net 网络模型中加入批标准化和 Dropout 层,使得城市高分辨率遥感图像中植被的自动分类精度得到有效提升。

虽然现在基于深度学习的图像分割模型能够提高高分辨率遥感图像植被提取的正确率,但是城市高分辨率遥感图像随着分辨率的提升,也使得图像中出现了更多的背景噪音,比如建筑物和停车场里分布不均匀、形状不规则的植被,这类问题会使植被提取任务存在破碎和不均匀的结果,比如靠近城市建筑物和城市中的大面积公共设施时,部分植物容易被阴影遮挡。所以,提高植被提取区域的孔洞及边界准确性,保证提取区域拓扑结构的准确性是本文研究的主要内容。

1.2 拓扑数据分析

拓扑学在其他领域中已经被广泛应用,例如计算机视觉、医学影像分析、地理信息科学等。在图像分割领域中,拓扑学可以用于约束分割结果的连通性和孔洞结构,提高分割结果的准确性和鲁棒性。已有的将拓扑用于图像分割的相关工作有好几种。例如可以使用基于图的分割方法,并且在目标函数添加基于拓扑约束的正则化项。

总之,拓扑学在遥感图像分割领域中的应用具有广阔的前景和潜力,本文进一步探索和发展拓扑约束的有效性和适用性,以提高遥感图像分割的精度和鲁棒性。

持续同调(persistent homology, PH)是研究空间距离的拓扑特征,是拓扑数据分析中的主要工具之一。最常用的方法就是构建单纯复形(simplicial complexes)并且对单纯复型进行过滤(filtration),然后记录这个计算过程中持续同调性的变化规律。但是在分析图像数据时立方复形(cubical complexes)拥有更高的计算效率,这是一个新的研究领域,被称为拓扑数据分析,其中拓扑的思想被用来从嘈杂和高维数据集中提取信息。

2 模型及方法

2.1 网络主体模型

U-Net 是一种基于编码和解码结构的语义分割模型,因其网络结构像字母 u 而得名,基本网络架构如图 1 所示。其中编码器由多个模块堆叠而成,每个模块内部由卷积层组成,将特征图传递到对应的解码器模块,为解码器提供额外的细节特征。图像数据由编码器输入,在模块之间通过最大池化层连接实现下采样,经过层次结构得到图像压缩后的语义特征。受到 DenseNet 的启发,解码器的每个模块不仅输入前一个模块上采样后的特征,还接收编码器对应模块输出的特征。最后,输出分割后的标签图。

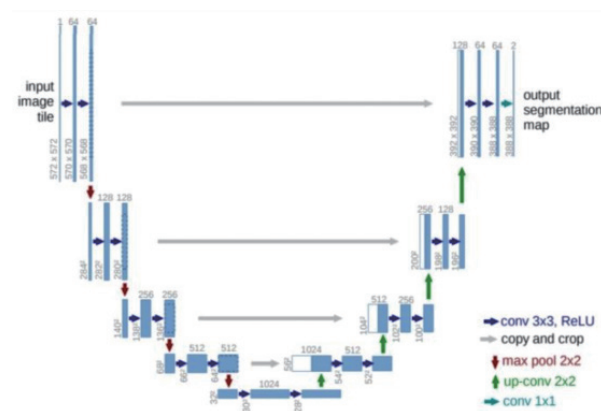


图 1 UNet 模型

2.2 拓扑损失函数

由于 UNet 模型采用的是卷积和池化操作,对于物体的边界信息处理不够充分,导致在物体边界模糊的情况下,分割效果不佳,而拓扑损失函数有望解决这些问题。拓扑损失函数可以考虑预测结果和真实结果之间的拓扑结构,从而处理形状变化和缩放等问题,具有更好的鲁棒性。此外,拓扑损失函数使用的距离度量不仅考虑像素级别的距离,还考虑像素之间的拓扑关系,可以确保在不同区域的像素数量不平衡的情况下,各个区域的预测结果都能够得到合理的评估。拓扑损失函数还可以保证预测结果的连通性,即预测结果中的同一区域内的像素都应该属于同一个连通组件,使得预测结果更加合理。

对于每个模型输出的预测值与真实值,引入拓扑数据分析中持续同调方法(persistence homology, PH)^[9]对预测与真实实例提取拓扑特征,然后将预测值的拓扑特征与真实值之间的差值作为损失。持续同调对图像中的拓扑特征进行提取,然后通过在不同尺度下构建简单复形,统计不同尺度的复形中持续存在的单纯同调群(拓扑结构),获取持续拓扑特征。本文使用 Dice 系数来衡量预测实例的持续图与真实实例的持续图之间的相似性,其值越大意味着这两个样本越相似,Dice 系数的数学表达式为:

$$Dice = \frac{2|X \cap Y|}{|X| + |Y|} \quad (1)$$

对于图像语义分割中的二分类问题,现在主流的方法是通过各种方式减少置信度的预测,置信度高的预测标签像素会使 Dice 系数提高,从而得到较低的 TopoLoss。表达式为:

$$T = 1 - \frac{2 \sum_{i=1}^N y_i \hat{y}_i}{\sum_{i=1}^N y_i + \sum_{i=1}^N \hat{y}_i} \quad (2)$$

3 实验结果和分析

3.1 数据预处理

本文中使用的公开数据集是德国 Potsdam (波茨坦) 地区的数据集。这个数据集将城市高分辨率遥感图像中的地物信息

整体分为6个类别。因为本文的研究重点是城市植被的提取，所以将数据集进行了二值化处理，把树木与低矮植物统一综合为植被类并标注为白色，将城市中的各种建筑物、不透水的地面、车辆还有其他背景合并为非植被类并标注为黑色。

因为深度学习的网络模型在训练过程中需要一定量的训练集，虽然 UNet 模型能够做到在少量数据集的情况下避免过拟合，但是较多的数据集可以增加模型的泛化能力。因此在本研究中，对图像进行了剪裁处理，将波茨坦城市数据集扩容成 3800 张 600×600 的图像，然后以 7:3 的比例将剪裁后的数据集划分为训练集和验证集，用于后续的网络模型训练和训练后的指标验证。

3.2 评价指标

本研究中的城市植被提取的深度学习网络模型的性能评估方法会从以下几种评价指标来进行：准确率 (accuracy, A)、 F_1 值、交并比 (intersection of union, IoU) 和 Kappa 系数，准确率指的是在所有样本中预测正确的概率，一般来说，准确率和召回率反映了分类器性能的两个方面，单一依靠某个指标并不能较为全面地评价一个分类器的性能，像素的错分率越低其值越高。精确率 (precision, P) 是预测的正样本中有多少是真的正确的概率，在本研究中指的是在预测中正样本的植被占实际中正样本植被像素的比例。召回率 (recall, R) 是正样本中有多少是被找了出来，在城市植被信息的提取中指的是分类正确的图像像素个数占总图像像素个数的比例。 F_1 值指的是精准率和召回率的平均值。交并比指的是该网络模型预测的城市植被像素和真实像素的交集和并集的比值，效果越好它的值越高。

$$A = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

$$P = \frac{TP}{TP + FP} \quad (4)$$

$$R = \frac{TP}{TP + FN} \quad (5)$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \quad (6)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (7)$$

式中： T_p (true positive) 指的是正值被预测为正值； T_N (true negative) 指的是负值被预测为负值； F_p (false positive) 指的是负值被预测为正值； F_N (false negative) 表示正值被预测为负类。

Kappa 系数是用在深度学习模型中评估一致性的一种方法，可以用它来进行多分类模型准确度的评估。这个系数的值越高，则代表模型实现的分类准确度越高。

$$Kappa = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \cdot x_{+i})} \quad (8)$$

式中： N 表示像素总元， x_{ii} 表示混淆矩阵对角线元素， r 表示分类数， x_{i+} 表示 i 类的预测样本数， x_{+i} 表示 i 类的真实样本数。

3.3 实验设置

模型实验基于 PyTorch 深度学习框架在一张 NVIDIA RTX3060 显卡上训练。使用了 adamW^[10] 优化器对模型进行优化，初始学习率设置为 0.001。训练的 batch size 设置为 8，总共轮次为 200 轮。随着训练轮数的增加，本研究中的拓扑损失函数的数值会随着时间逐渐变小，当损失函数的值逐渐变小然后稳定在一个值附近时，表示该网络模型收敛，然后程序将自动保存训练过程中最优模型参数。

3.4 实验结果

为验证本文网络模型改进后的有效性，现在将 TP-UNet、U-Net^[5] 和 DeepLabV3+^[11] 三种不同的网络进行了对比分析，模型参数和模型大小进行比较的结果如表 1 所示。

表 1 研究模型参数量 (空间复杂度)

网络模型	参数量 (空间复杂度)
U-Net	6.9
Deep LabV3+	49.80
TP-UNet	8.76

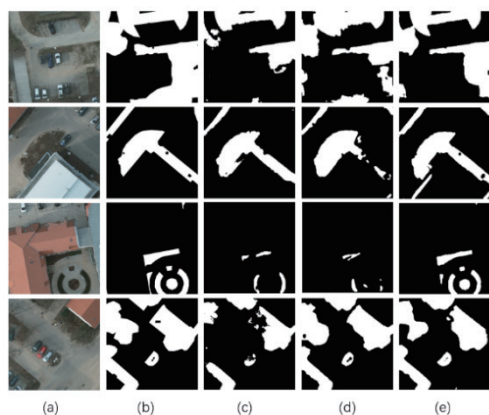
可以很明显的看到，DeepLabV3+ 网络结构最为复杂，参数量最大，U-Net 网络是表中网络结构最简单的，参数量也是最少的。本文提出的 TP-UNet 网络参数量与 U-Net 网络差距较小，更是远小于 DeepLabV3+ 网络。

表 2 列出 TP-UNet 等 3 个模型在波茨坦城市数据集上的训练结果，准确率、 F_1 值、交并比与 Kappa 系数值。从网络模型的准确率能够清晰地看到，三种网络模型都能够有效地将本数据集的城市高分辨率遥感图像中的植被基本形状分割出来。U-Net 网络的各项指标与其他网络模型相比都处于较低水平，表现较差。表现最好的是本文提出的 TP-UNet 模型，相较于其他网络模型而言，表 2 的四种评价指标值都有了不同程度的提高，与 U-Net 基础模型相比，改进后的网络模型的准确率提高了大概 0.5% 左右、 F_1 值提高了大概 0.61% 左右、交并比提高了大概 1.14% 左右、Kappa 系数提高了大概 0.02% 左右，证明了本研究提出的改进后的网络模型的有效性。

表 2 波茨坦城市数据集提取结果指标

模型	$A/\%$	$F_1/\%$	$IoU/\%$	Kappa
U-Net	90.20	85.81	86.65	0.777 7
DeepLabV3+	94.11	89.43	82.30	0.798 7
TP-UNet	90.70	86.42	87.79	0.797 7

图 2 展示了本文改进的网络模型与其他两种模型的结果对比图，这些示例图分别从不同方面展示出这三种的图像分割能力。



(a) 原始图像; (b) 真值标签图; (c) UNet;
(d) DeepLabV3+; (e) 本研究模型

图2 分割图像的示例

第一幅图展示的是建筑物停车场旁灌木植被的情景,由于地表物体距离紧凑,边界模糊,使得模型对这类场景的分割变得比较困难,可以看到 TP-UNet 模型能对其进行比较清晰的区分,说明了 TP-UNet 模型能够很好地保留区域内植被的拓扑结构。第二幅图展示的是建筑物旁出现规则草坪的情景,因为其颜色与建筑物的阴影相近,给模型的图像分割带来了很大的难度。只有 UNet 网络模型和 TP-UNet 网络模型能够将植被信息从大片的建筑物中分割出来,和 DeepLabV3+ 相比有着比较明显的差距。第三幅图展示了回字楼中圆形并带有孔洞的草坪场景,通常小公园草坪上有连通周围道路的小而细的通道,这种小路比较密集而且通道狭窄使得边界模糊比较难区分。可以看到只有 TP-UNet 模型能够在这种困难情况下能够使用研究目标保留较好的拓扑结构同时保持其连通性和孔洞结构,DeepLabV3+ 模型则在此类困难的条件下表现地差强人意。第四张图中有小块建筑物、汽车,还有大片道路,同时汽车还停在草坪上,汽车之间存在的小片植被可能会导致阴影相近的植被的区域产生一样的像素特征,使与汽车阴影相连的植被很容易被当成非植被类。三种模型中,DeepLabV3+ 网络模型虽然可以把植被区域的大体轮廓分割出来,但效果还是差强人意,相比之下 TP-UNet 模型优势更为明显。通过在不同的网络模型下进行的对比实验可以看出,TP-UNet 网络模型可以在伴随复杂噪声的情况下可以较好地保持植被的拓扑结构。

4 结论

综上所述,本研究提出了一种改进的 UNet 模型,结合拓扑损失函数和拓扑先验来提高图像分割的效果和鲁棒性。通过实验结果可以看到,该模型在各项指标上均取得了较好的结果,证明了该模型相比基准模型在拓扑结构方面的有效性。

拓扑损失函数可以考虑预测结果和真实结果之间的拓扑结构,从而处理形状变化和缩放等问题,具有更好的鲁棒性。

同时,拓扑先验可以提供先验知识,指导模型更加准确地预测图像中的物体边界和形状信息。通过将这两种方法结合到 UNet 模型中,成功提高模型的分割效果和鲁棒性。

参考文献:

- [1] 梁志林,张立燕,曾现灵,等.高光谱遥感城市植被识别方法研究[J].地理空间信息,2017,15(2):72-85.
- [2] KATTENBORN T, LEITLOFF J, SCHIEFER F, et al. Review on convolutional neural networks (cnn) in v egetation remote sensing[J]. ISPRS journal of photogrammetry and remote sensing, 2021(173):24-49.
- [3] WATANABE S, SUMI K, ISE T. Identifying the vegetation type in google earth images using a convolutional neural network: a case study for japanese bamboo forests[J]. BMC ecology, 2020,20(1):1-14.
- [4] 王修信,杨路路,汤谷云,等.利用卷积神经网络提取高分辨率遥感图像喀斯特森林信息[J].科学技术与工程,2020,20(17):6773-6777.
- [5] RONNEBERGER O, FISCHER P, BROX T. UNet : convolutional networks for biomedical image segmentation [C]//In: International Conference on Medical Image Computing and Computer-Assisted Intervention.Berlin:Springer, 2015:234-241.
- [6] ALSHEHHI R, MARPU P R. Extraction of urban multi-class from high-resolution images using pyramid generative adversarial networks[J]. International journal of applied earth observation and geoinformation, 2021,102:102-119.
- [7] 蔡淑颖,何少柏,韩凝,等. U-net 深度学习神经网络结合面向对象的森林高分影像信息提取[J]. 园林, 2020, 343(11): 41-50.
- [8] 徐知宇,周艺,王世新,等.面向 GF-2 遥感影像的 U-Net 城市绿地分类[J].中国图象图形学报,2021,26(3):700-713.
- [9] EDELSBRUNNER H, HARER J. Persistent homology-a survey[J]. Contemporary mathematics, 2008,453(26): 257-282.
- [10] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization[EB/OL].(2017-11-14)[2023-09-22].<https://arxiv.org/abs/1711.05101>.
- [11] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//In: Proceedings of the European conference on computer vision (ECCV). Berlin:Springer,2018:801-818.

【作者简介】

高昊(1995—),男,山西太原人,硕士研究生,研究方向:图像分割。

阴桂梅(1975—),通信作者(email: yinguimeicn@126.com),女,山西太原人,教授,研究方向:智能信息处理。

(收稿日期:2023-12-04)