

文章编号: 1008 - 2786 - (2018)6 - 953 - 11

DOI: 10. 16089/j. cnki. 1008 - 2786. 000390

深度语义分割的无人机图像植被识别

林志玮^{1,2,3}, 涂伟豪¹, 黄嘉航², 丁启禄¹, 刘金福^{1,4}

(1. 福建农林大学 计算机与信息学院, 福建 福州 350002; 2. 福建农林大学 林学院, 福建 福州 350002;

3. 福建农林大学 林学博士后流动站, 福建 福州 350002; 4. 福建省高校生态与资源统计重点实验室, 福建 福州 350002)

摘 要: 为有效实施植被信息获取及监测, 亟需分类准确及易于推广的植被信息识别技术。本文利用无人机航拍获取植被光学影像, 利用深度语义分割技术建构植被种类识别模型, 为植被变化动态监测提供准确的植被类别信息。首先, 基于安溪县龙门镇崩岗区的采样点, 获取 20 m 航拍高度的无人机影像, 构建 FCN-VGG19 植被识别模型, 探讨不同特征融合结构对 FCN-VGG19 识别性能的影响, 测算出各植被的覆盖面积; 其次, 取安溪县另一取样点的无人机影像作为验证集, 分析 FCN-VGG19 的迁移学习能力, 验证模型稳健性。结果表明: (1) 基于 20 m 高度的无人机影像建立的 FCN-VGG19-8s 模型识别正确率最高, 为 86.30%; (2) FCN-VGG19-8s 识别精度高于 FCN-VGG19-32s; 并从测试集中随机抽取一张图, 测算该测试图的马尾松覆盖面积为 78.38 m², 芒其覆盖面积为 12.77 m², 柠檬桉覆盖面积为 0.89 m²; (3) 在模型的迁移学习能力试验分析中, 当 A 数据集占训练集的比例下降时, 对模型识别 B 数据集的影响不大; 当 B 数据集的数据量减少时, 其识别精度稍有下降, 仍有 84.5%。本文基于无人机光学影像, 结合深度语义分割模型进行植被识别, 以福建安溪县为例验证模型稳健性, 分析模型在测算植被覆盖面积的适用性, 旨在为植被识别研究提供新思路。

关键词: 语义分割; 全卷积网络; 无人机影像; 植被识别

中图分类号: S719

文献标志码: A

植被作为陆地生态系统的重要组成部分, 是生态系统物质循环与能量流动的中枢, 在水土保持、水源涵养及固碳过程中都起着重要作用。2017 年世界植被现状年度报告^[1]指出, 17 个科的植被可提供人类所需 80% 的食物, 植被对人类具有至关重要的作用。但随着社会经济的发展和工业活动的增加, 大气污染与水土流失加重, 植被遭到大面积破坏, 覆盖面积大幅减小, 植被保护与恢复成为研究热点。在此过程中, 植被类型与分布资讯为植被保护与恢复提供基础决策信息^[2]。精确识别植被类型决定了植被分布信息获取的有效性, 进而影响植被保护与恢复进程, 故设计准确、有效的植被识别手段, 迫

在眉睫。

传统植被识别主要采用实地调查, 此方法可保证植被识别的精度, 但需要人工全程参与, 存在效率低及成本高的不足^[3]。随着计算机和遥感技术的发展与更新, 基于遥感影像的植被识别成为新的识别手段, 按其分类的最小单元可分为像素级分类和面向对象分类。像素级分类是将遥感影像的每个像素作为最小分类单元, 采用分类算法, 如支持向量机、决策树等建立分类模型, 识别植被类型, 取得较高的分类精度^[4-6]。基于像素级分类精度较高, 但遥感影像存在“同物异谱, 同谱异物”的现象, 以像素为分类单位仅能依据其光谱信息作为分类特征,

收稿日期 (Received date): 2018 - 04 - 24; **改回日期** (Accepted date): 2018 - 11 - 13

基金项目 (Foundation item): 海峡博士后交流资助计划; 中国博士后科学基金面上项目 (2018M632565); 福建省自然科学基金项目 (2016J01718)。[Channel Postdoctoral Exchange Funding Scheme; China Postdoctoral Science Foundation (2018M632565); Natural Science Foundation of Fujian (2016J01718)]

作者简介 (Corresponding author): 林志玮 (1981 -), 男, 博士, 讲师, 主要研究方向: 图像处理, 图形识别, 机器学习。[LIN Zhiwei (1981 -), male, Ph. D., lecturer, research on image processing, pattern recognition and machine learning] E-mail: cwlin@fafu.edu.cn

难以有效克服该问题的影响。面向对象分类是以一定面积的规则或者不规则区域作为分类单元,可综合影像的光谱、形状、纹理等信息,在一定程度上能克服“同物异谱,同谱异物”的问题^[7]。常见的基于遥感影像的面向对象方法,是结合多尺度分割技术,将遥感影像分为多个不规则斑块,以斑块为分类单元,提取其光谱、纹理以及形状分类特征,采用分类算法建立植被识别模型^[8-9]。利用面向对象的分类方法对植被识别,具有精度高、识别面积广等优势,但其分类精度受分割尺度影响,对于不同的地区,其最优分割尺度一般不同,造成分割尺度参数难以确定。综上所述,采用遥感影像进行植被识别,可获取较高分类精度,但不管是基于像素级分类或是面向对象分类均存在一定的局限性。考虑到遥感技术易受天气等因素影响,而无人机技术具有数据获取方便、影像清晰以及受天气影响小等优势;相比于遥感技术,其更容易大范围的普及,已广泛应用于森林病虫害监测^[10]、灾害应急救援^[11]等领域。基于无人机搭载光学相机,结合机器学习提取植被 RGB 信息和分布特征进行植被识别,能够改善基于遥感影像植被分类的不足。

深度神经网络作为新兴机器学习方法,可自适应抽取分类特征,特征拟合能力强、分类特征具有鲁棒性,能有效处理图像分类^[12-13]、图像定位^[14-15]、语义分割^[16]等任务。VGG19 模型^[17]是图像识别的深度神经网络,在 ImageNet 大规模视觉识别挑战 (ILSVRC) 有着高精度的识别效果,其具有 16 层卷积层和 3 层全连接层,多层的卷积层能够更好的提取图像的细节信息,由于植被分布错综复杂,细节信息难以提取,将 VGG19 模型应用植被识别可以提取植被的细节信息,提高植被分类精度。考虑到 VGG19 模型无法测算植被的覆盖面积,因此,基于 VGG19 建立全卷积网络模型 (Fully Convolutional Networks, FCN)^[17]进行植被覆盖面积测算。全卷积网络模型是深度语义分割中的典型代表,具有分类精度高、分类流程简易的技术特点,将其应用于植被识别不仅可以精确识别植被的类型,还能测算其覆盖面积及空间分布。回顾国内外文献,目前尚未有研究使用无人机航拍影像进行多数种的植被识别。故本文基于无人机光学影像,首次结合深度语义分割的全卷积网络模型,利用卷积变换自动提取图像局部光谱、纹理特征,建立植被识别模型。实验结果

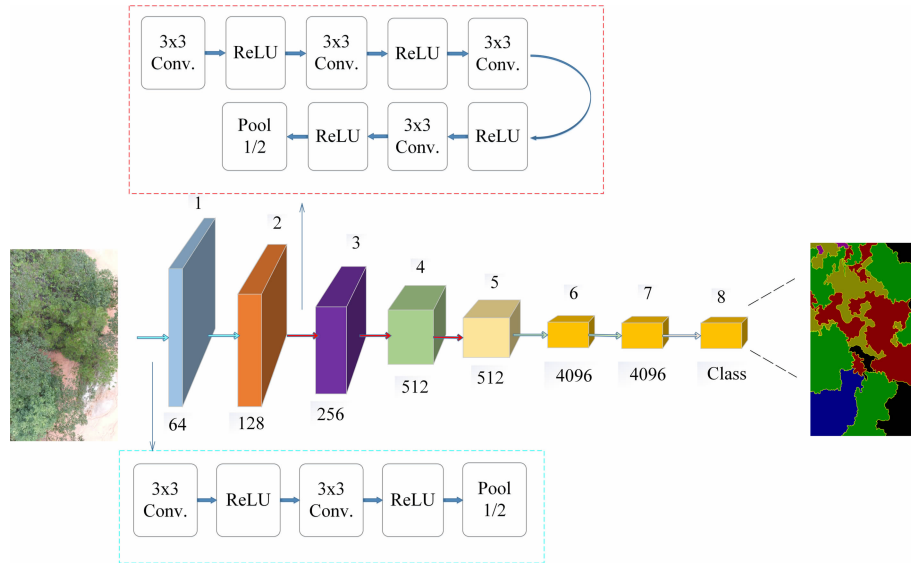
表明,全卷积网络模型不仅能够准确识别植被类别,还能估算出测试图中各植被的覆盖面积,为植被分布信息的收集提供简便、有效手段,为制定植被保护与恢复方案提供合理基础信息。

1 研究区域与数据

安溪县位于福建省东南沿海,隶属泉州市,县域范围为东经 $117^{\circ}36' \sim 118^{\circ}17'$, 北纬 $24^{\circ}50' \sim 25^{\circ}26'$, 东西宽 74 km, 南北长 63 km, 分为外安溪与内安溪, 全县总面积达 3057.28 km^2 。气候为南、中亚热带海洋性季风气候, 外安溪年平均温度 $19^{\circ}\text{C} \sim 21^{\circ}\text{C}$, 年降雨量 1600 mm; 内安溪年平均温度在 $16^{\circ}\text{C} \sim 18^{\circ}\text{C}$, 年降雨量 1800 mm。地势自西北向东南倾斜, 其中西北部山峦起伏, 坡度大; 东南部地势相对较平缓。区内分布南亚热带季雨林和中亚热带常绿阔叶林, 西北部主要分布常绿阔叶林和针阔叶混交林, 东部是马尾松、相思树、木荷混交林。其地处闽东南花岗岩区域, 植被退化和崩岗侵蚀现象严重^[18-19]。调查该地区的植被类型及植被覆盖面积有助于分析崩岗侵蚀现象, 为植被恢复提供指导依据。因此, 将通过无人机搭载光学相机, 分别对安溪县山区的 A、B 采样点进行航拍, 并基于深度学习方法进行植被识别。采用的无人机为大疆 inspire 1 raw, 型号为 T600, 四轴飞行器, 旋转角速度俯仰轴为 $300^{\circ}/\text{s}$, 航向轴为 $150^{\circ}/\text{s}$; 光学相机型号为 zenmuseX5R, 1600 万像素, 并设定不同的航拍高度, 航拍高度为无人机距离地面起飞点的垂直距离。无人机飞行取样时, 预先设定飞行航线, 拍摄路线采用“己”字形进行取样。为降低数据的重复性, 取样完成后进行数据筛选, 剔除重复冗余于图像, 确保图像间不重叠。在实验阶段, 将收集的数据分为训练集和测试集, 数据集间数据无区域重叠。

2 方法

全卷积神经网络模型 (Fully Convolutional Networks, FCN) 可以接受任意尺寸的输入图像, 并产生相应尺寸的输出结果, 是在卷积神经网络的基础上, 实现对图像像素级分类的网络架构。基于 VGG19 的全卷积网络架构 FCN-VGG19 如图 1 所示; FCN-VGG19 模型可简化为 8 个阶段, 第 1 阶段与第 2 阶段的运算均先进行卷积运算 (Conv.) 和激



(从原图到预测图的每一箭头表示为一阶段,相同颜色的箭头表示相同的操作,每一阶段后的立体块为运算后的特征图像,立体块上的数字表示为第 N 次阶段的结果,立体块下的数字表示为特征图像的张数。)

图1 网络结构图

Fig. 1 The network architecture

(Each arrow from the original image to the prediction graph is represented as a stage, the arrows of the same color indicate the same operation, the cubic after each phase are the processed feature images, and the numbers on the cubic are represented as the result of N^{th} stage, the number under the cubic is expressed as the number of features of the feature image.)

活运算 (ReLU), 重复运算两次后进行池化 (pool) 运算, 第 3、4、5 阶段均包含 4 次卷积运算和激活运算, 运算后进行 pool 运算, 第 6、7 阶段则都包含卷积运算和激活运算, 产生 4096 张特征图像, 送入第 8 阶段进行 1×1 卷积运算, 产生的预测图进行反卷积运算, 得到与输入图像大小一致的预测图, 实现像素级的语义分割。

2.1 卷积层

卷积层是利用卷积核对输入图像进行卷积运算的网络层, 在卷积层中, 每一个神经元会接受上一层的局部信息, 并通过卷积核进行权值运算, 使得输入信号的有效信息得到增强, 降低噪声的影响, 提取出输入图像的局部特征。其计算方式可表示为:

$$x_j^l = \sum_{i \in M_j} x_j^{l-1} * k_{ij}^l + b_j^l \quad (1)$$

其中: l 是卷积层所在的层数, x_j 第 j 张特征图的像素值, $*$ 为卷积运算, k_{ij}^l 为卷积核对第 j 张特征图像的第 i 个区域进行卷积运算, b_j^l 是第 j 张特征图的偏差, M_j 为 $l-1$ 层的一个输入特征图。为筛选出有效信息, 输入信息每经过一次卷积层均进行激活函数运算。激活函数是非线性函数, 用于为神经网络模型加入非线性因素, 将输入信息进行非线性的函

数映射, 使其能够处理复杂的非线性分类问题。ReLU 是其中一种激活函数, 可以避免梯度消失问题, 且计算量更低^[20], 其计算公式为:

$$f(z) = \max(0, z) \quad (2)$$

其中, z 表示为非线性激活函数的输入信息。当输入值小于或等于 0 时, 输出为 0; 输入值大于 0 时, 输出原值。若保持原图大小进行卷积运算和激活函数运算, 则网络计算参数量过大, 导致网络计算难度大。为解决参数量过大问题, 同时增强卷积层提取的特征信息, 可以通过 pool 运算降低图像大小。本文采用的 pool 函数是 maximum pooling 函数, 即以图像一个区域上的某个特定特征的最大值来代表这个区域的特征, 减少了图像的维度。在网络的后三层中, 卷积神经网络常在最后网络层添加全连接层, 即将卷积层中提取的不变性特征进行分类^[21]。但全连接层在转换过程中会丢失空间信息, 不同于卷积神经网络, 全卷积网络的后三层均采用卷积核为 1×1 的卷积层替换全连接层, 在避免空间信息丢失的同时, 获取每一个区域的输出结果。

2.2 反卷积

全卷积网络由于将全连接层转化为卷积层, 使

得网络中卷积层的输出信息至少为二维,可以输出一个类别热图。输出结果由于卷积层中的 pool 运算,其维数也随之降低,输出结果的维度无法与输入图像的大小匹配,需要通过反卷积将网络的输出结果转化为与输入图像大小一致。反卷积是通过卷积的转置实现卷积逆过程的运算,其计算公式^[22]为:

$$\hat{z}_j^l = \sum_{i=1}^{n_j} z_{ij}^l * f_{ij}^l \quad (3)$$

其中,*号为卷积运算, l 是卷积层所在的层数, \hat{z}_j 为第 j 张特征图像的重构图像, z_{ij} 为第 j 张特征图像的第 i 个区域的像素, f_{ij} 是对第 k 张特征图像的第 i 个区域的像素进行反卷积运算的卷积核, n_j 为第 j 张特征图的像素点总数。在 FCN-VGG19 中,输入图像由于池化(pool)运算降低了图像的分辨率,因此,最后一层卷积形成的特征图分辨率较低,细节特征体现不明显。FCN-VGG19-32s 将最后一层卷积层输出的热图直接进行反卷积扩大 32 倍,输出为与输入图像大小一致的预测图,直接反卷积的预测图精度不高。为使模型细节处理能力更强,long^[18]提出了多尺度融合的 FCN 结构,结合抽象高层信息与具体低层信息,丰富输出结果的信息量。融合结构如图 2 所示。

多尺度融合结构主要分为 FCN-VGG19-16s 和 FCN-VGG19-8s 两种融合结构:(1) FCN-VGG19-16s 把最后一层卷积层输出的热图进行反卷积,图像扩大 2 倍再与 pool4 的特征图融合成新的特征图,最后对新的特征图进行反卷积还原至原图大小;(2) FCN-VGG19-8s 则把 FCN-16s 融合成的特征图与 pool3 的特征图进行融合,并进行反卷积扩大至原图大小。

2.3 覆盖面积测算

由于 FCN-VGG19 模型是像素级分类的网络架

构,通过 FCN-VGG19 模型统计测试图各植被像素点数,可以测算出每一张测试图对应的覆盖面积,为减少计算量,对每一张图像进行缩放。由于图像经过缩放,像素点数有所改变,应通过计算图像缩小比例分析原有像素点数,图像缩小比例计算公式为:

$$g = s_i / s_o \quad (4)$$

g 为图像缩小比例, s_i 为原图像的大小, s_o 为缩小后图像的大小。

仅测算出各植被的像素点数是无法测算其实际覆盖面积,需通过影像空间解析度计算各像素点所表示的实际大小,影像空间解析度由拍摄的高度所决定,需通过比较参照物来获得影像空间解析度,其计算公式为:

$$r = a / n \quad (5)$$

其中, r 为空间解析度($\text{m}^2/\text{像素点数}$), a 为参照物面积(m^2), n 为参数物像素点数。

通过统计各植被类别的像素点数、图像缩小比例及影像空间解析度可以测算出各植被的覆盖面积,覆盖面积计算公式为:

$$S_p = grN_p \quad (6)$$

S_p 为第 p 类植被的覆盖面积, N_p 为第 p 类植被的像素点数。

2.4 分类精度指标

采用 4 种度量方式验证语义分割模型的分类效果,分别为像素点正确率(pixel accuracy)、平均正确率(mean accuracy)、平均 IU 指数(mean intersection over union)、加权 IU 指数(frequency weighted intersection over union),其计算公式为:

像素点正确率:

$$\sum_p n_{pp} / \sum_p t_p \quad (7)$$

平均正确率:

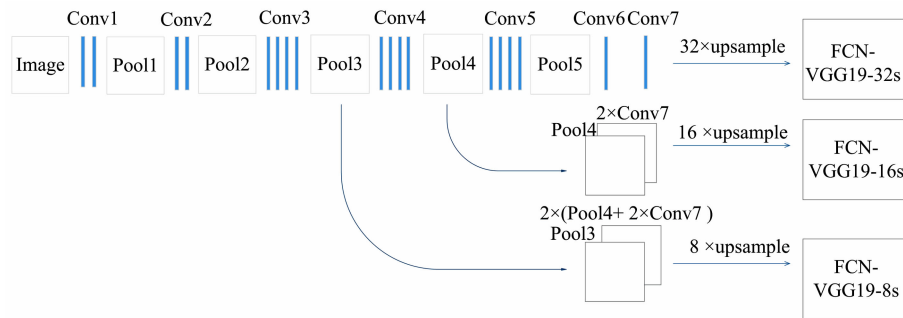


图 2 多尺度融合结构图

Fig. 2 The multi-scale fusion structure

$$(1/n_c) \sum_p n_{pp}/t_p \quad (8)$$

平均 IU 指数:

$$(1/n_c) \sum_p [n_{pp}/(t_p + \sum_q n_{qp} - n_{pp})] \quad (9)$$

加权 IU 指数:

$$(\sum_{m=1}^{n_c} t_m)^{-1} \sum_p [t_p n_{pp}/(t_p + \sum_q n_{qp} - n_{pp})] \quad (10)$$

其中, $t_p = \sum_q n_{pq}$, n_{pq} 表示为属于 p 类的像素点被分类为 q 类的个数, n_{pp} 表示为属于 p 类的像素点被分类为 p 类的个数, n_c 表示为类别的种数。像素点正确率表示为所有像素点识别的精度;平均正确率是计算每一类别像素点判断正确的比例,再把所有类别进行平均;平均 IU 指数为各类别预测的像素点集合和真实的像素点集合的交集与两个集合的并集之比;加权 IU 指数则在平均 IU 指数上考虑各类别出现的频率,以每一类出现频率作为每一类别的权重进行计算平均 IU 指数。

3 结果分析

本节通过三部分实验探讨深度语义分割模型在植被识别任务的适用性:(1)基于不同高度的安溪县 A 采样点无人机光学图像,分析 FCN 模型的植被分类效果;(2)基于安溪县 A 采样点无人机光学图像,研究不同融合结构,即 FCN-VGG19-32s, FCN-VGG19-8s 的分割精度,确定最优结构,并测算出测试图植被的覆盖面积;(3)以安溪县 B 采样点光学图像为训练集,安溪县另一取样点的光学图像为测试集,分析基于全卷积网络模型植被识别的迁移学习稳健性。据此分析深度语义分割模型在测算植被覆盖面积的适用性。

实验主要硬件配置如下:CPU, Inter core i7-7700K 4.2GHZ; GPU, GeForce GTX1070; 内存,

32G。实验损失函数优化采用 Adam (Adaptive Moment Estimation) 算法,学习率设为 10^{-4} 。batch size 设为 1,即训练一个样本(图片)则进行一次反向传播,epoch 设定 100。同时,预载 imagenet-VGG19 权重初始化网络,减少预训练时间,避免过拟合。

3.1 不同高度的航拍影像下 FCN 模型植被分类结果

不同航拍高度的无人机影像解析度不同,影像的覆盖范围也不同,故分别对同一地区 20~40 m 航拍高度的无人机影像建立 FCN 模型,分析高度变化对模型的影响。不同高度下数据集 A (20~40 m) 均由安溪县无人机影像通过镜像和图像旋转进行数据增强,每一高度数据集共计 6336 张光学影像,并将图像大小缩小为 224×224 ,其中,训练集 4608 张,测试集 1728 张。设定不同尺度下的分类模型均为 FCN-8s,参数保持不变,其分类结果见表 1、图 3。

预测图上每一像素点均属于其特定的种类,在图 3 中每一种颜色代表一种树种,颜色所在的区域表示为物种的分布。其中,黑色类别为裸露地面,黄色类别为芒萁,蓝色类别为石楠,绿色类别为柠檬桉,后三种类别均为安溪县山区的优势树种。从图 3 可知,基于 20~40 m 无人机影像的 FCN 模型预测图与标准图整体基本一致,有着较高的植被识别精度,但不同高度下 FCN-8s 模型分类的效果有着明显差异,基于 20 m、30 m 及 40 m 航拍影像的分类精度分别为 86.30%、78.24% 及 76.22%。随着无人机影像高度的上升,FCN 模型的分割精度有所降低,由此可见航拍高度对模型识别精度的影响较大。随着航拍高度的增加,影像中树种的解析度下降,难以准确识别树种间的空间结构,加大了植被识别难度。在图 3 中,随着航拍高度的增加,各植被间的边界界定更加不准确,如 30 m 预测图像上的芒萁以及 40 m 预测图像上的柠檬桉,在空间分布均发生改变。

表 1 不同航拍高度影像的 FCN-VGG19 模型植被识别结果

Tab. 1 The recognition result of FCN-VGG19 model at various aerial altitude images

无人机航拍高度/m Sampling attitude	像素点正确率/% Pixel accuracy	平均正确率/% Mean accuracy	平均 IU 指数/% Mean intersection over union	加权 IU 指数/% Frequency weighted intersection over union
20	86.30	67.13	44.43	79.64
30	78.24	59.47	28.72	69.99
40	76.22	61.51	23.99	67.48

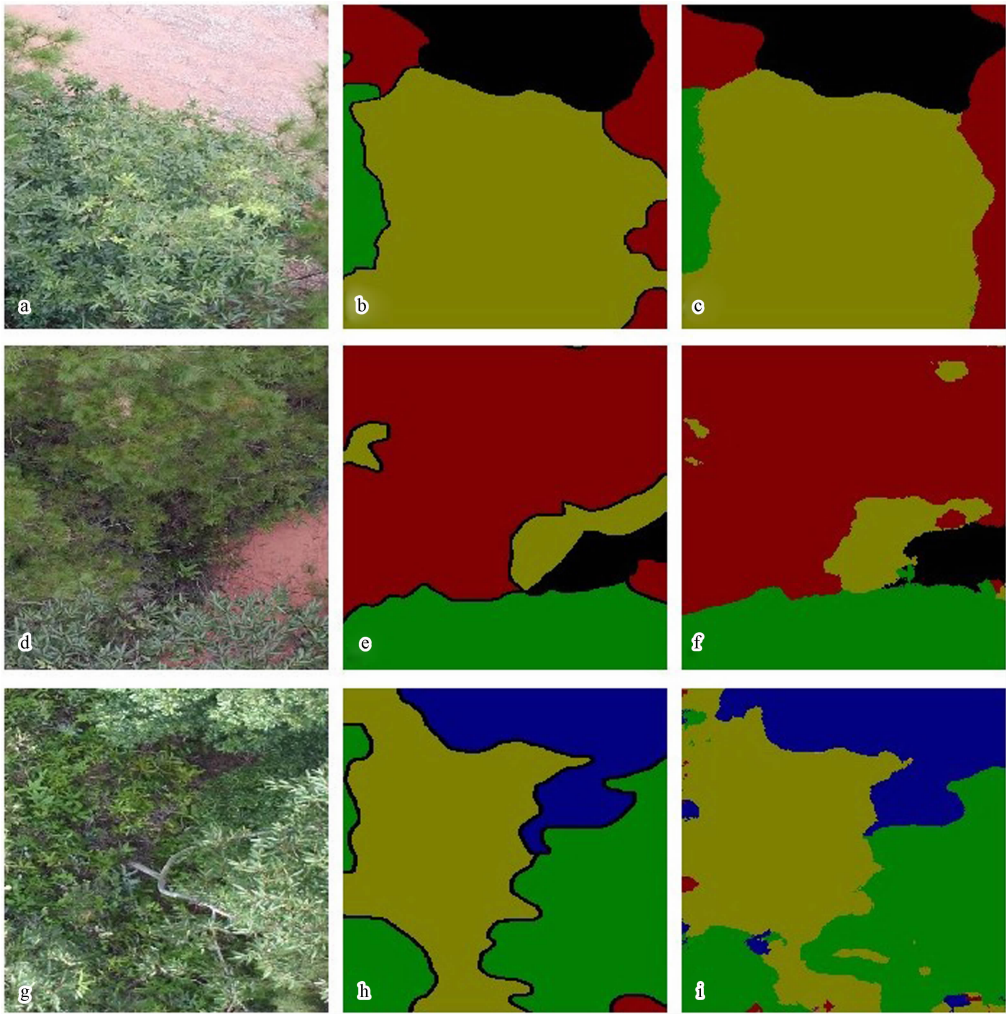


图3 不同航拍高度影像的 FCN-VGG19 模型可视化识别结果
(a: 20 m 原始图像;b: 20 m 标准图;c: 20 m 预测图;d: 30 m 原始图像;e: 30 m 标准图;f: 30 m 预测图;
g: 40 m 原始图像;h: 40 m 标准图;i: 40 m 预测图)
Fig. 3 The recognition results of FCN-VGG 19 model at various aerial altitude images by visualization
(a: 20 m original image; b: 20 m label image; c: 20 m predication image; d: 30 m original image; e: 30 m label image;
f: 30 m predication image; g: 40 m original image; h: 40 m label image; i: 40 m predication image)

3.2 不同融合结构下 FCN 模型植被分类结果

不同尺度的融合结构对空间层次信息结合存在差异,基于安溪县 A 采样点无人机航拍影像,分析融合结构对模型分类效果的影响,数据集 A-20m 的植被类别总计 13 类,包括裸露地面,马尾松 (*Pinus massoniana* Lamb.)、枯萎马尾松 (Withered *Pinus massoniana* Lamb.)、柠檬桉 (*Eucalyptus citriodora* Hook. f.)、幼年柠檬桉 (Young *Eucalyptus citriodora* Hook. f.)、芒萁 (*Dicranopteris pedata* (Houtt.) Nakaike)、枯萎芒萁 (Withered *Dicranopteris pedata* (Houtt.) Nakaike)、石楠 (*Photinia serratifolia* (Desf.) Kalkman)、孔雀豆

(*Adenanthera microsperma* Teijsm. & Binn)、东方乌毛蕨 (*Blechnum orientale* L.)、金银花 (*Lonicera japonica* Thunb.)、芒 (*Miscanthus sinensis* Anderss.)、苔藓 (Bryophyte)。对 FCN-VGG19-8s、FCN-VGG19-32s 实验结果进行对比分析,其识别结果如表 2 及图 4 所示。
由表 2 可知,FCN-VGG19-32s 和 FCN-VGG19-8s 的全卷积网络模型其像素点正确率均在 80% 以上。其中,FCN-VGG19-8s 像素点精度达到 86.30%,高于 FCN-VGG19-32s 模型的分类效果。这是由于不同网络层的感受野不一样,影像上的植被分布面积不一致,不同的感受野能够侦测到不同

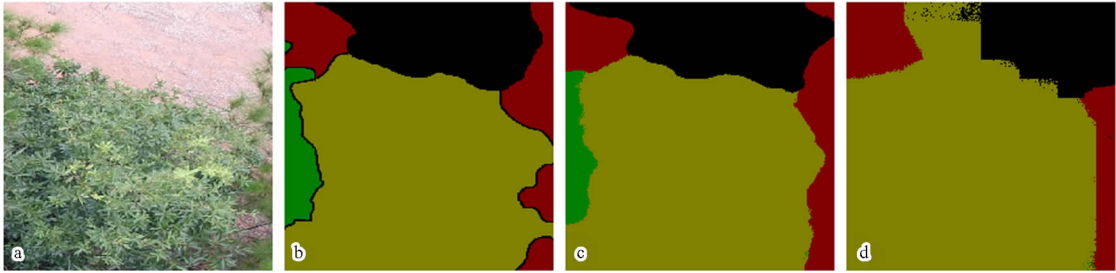


图 4 不同融合结构的 FCN-VGG19 模型可视化识别结果

(a: 原始图像;b: 标准图;c: FCN-VGG19-8s 模型预测图;d: FCN-VGG19-32s 模型预测图)

Fig.4 The recognition results of FCN-VGG19 model with different fusion structures by visualization

(a: Original image; b: Label image; c: Predication image of FCN-VGG19-8s; d: Predication image of FCN-VGG19-32s)

分布面积的植被信息,FCN-VGG19-8s 结合不同层特征信息,获得更精细的预测图。图 4 中,红色部分为马尾松的分布区域,绿色为柠檬桉的分布区域,黄色部分为芒萁的分布区域,FCN-VGG19-8s 模型的预测图和 FCN-VGG19-32s 模型的预测图基本与标识图一致,均能准确识别各个植被的类别,但 FCN-VGG19-8s 模型预测图上每一植被边界划分更为清晰,其识别效果更好。

表 2 不同融合结构的 FCN-VGG19 模型植被识别结果

Tab.2 The recognition result of FCN-VGG19 model with different fusion structures

评估指标 Evaluation index	FCN-VGG19-8s	FCN-VGG19-32s
像素点正确率/% Pixel accuracy	86.30	81.63
平均正确率/% Mean accuracy	67.13	58.32
平均 IU 指数/% Mean intersection over union	44.43	43.11
加权 IU 指数/% Frequency weighted intersection over union	79.64	72.95

FCN-VGG19 模型能够统计每一张无人机航拍图像其植被覆盖的像素点数、图像的放缩比和空间解析率,进一步测算出在该图像中各植被的覆盖面积。为分析无人机影像的空间解析率,采用 A4 纸为参照物进行测量,测试影像及空间解析率见图 5 及表 3。由于无人机航拍存在较多的航向,无人机航拍图像经过筛选后不足以覆盖整个地区,无法测算整个地区各植被覆盖面积。因此,仅从测试图像中随机抽取一张测试图进行测算,其覆盖面积测算

结果见表 4 所示。

表 3 无人机航拍图像分辨率

Tab.3 The resolution of UAV aerial image

高度/m Height	20
A4 白纸像素点数 Pixel number of A4 paper	1273
A4 纸实际面积/cm ² Actual area of A4 paper	623.70
1 像素点所占面积/cm ² The area of 1 pixels	0.49



图 5 无人机航拍图像

Fig.5 The UAV aerial image

3.3 不同地区无人机影像分类结果

在安溪县 B 采样点采用无人机航拍,无人机影像通过图像增强,建立 B 数据集,B 数据集共分为 3 份,均分为 B1、B2 和 B3 数据集,共计 1920 张。考虑到不同地区数据的训练集比例影响着模型的识别精度,将 A-20 m 数据集均分为 A1 数据集和 A2 数据集,分别有 2304 张无人机影像,并设计以下方案。方案一:为探究 A 数据集在训练集的比例变化对模型影响,维持 B 数据集的训练张数,改变 A 数据集

表 4 基于 FCN-VGG19-8s 模型的单张测试图植被面积测算结果
Tab. 4 The coverage of vegetation for single test image based on FCN-VGG19-8s model

类别 Classification	像素点数 Pixel number	放缩比 Scaling ratio	空间解析率 Spatial resolution	覆盖面积/cm ² Coverage area
裸露地面 Bare ground	10 295	8.36	0.49	42 200
马尾松 <i>Pinus massoniana</i> Lamb.	6535	8.36	0.49	783 800
柠檬桉 <i>Eucalyptus citriodora</i> Hook. f.	2178	8.36	0.49	8900
芒萁 <i>Dicranopteris pedata</i> (Houtt.) Nakaike	31 168	8.36	0.49	127 700

表 5 不同图像数据集的 FCN-VGG19-8s 模型识别效果
Tab. 5 The recognition results of FCN-VGG19-8s model with different image datasets

编号 Number	训练集 Training set	数据量 Data size	像素点正确率/% Pixel accuracy	平均正确率/% Mean accuracy	平均 IU 指数/% Mean intersection over union	加权 IU 指数/% Frequency weighted intersection over union
I	A1 + B1 + B2	3584	86.50	71.56	44.35	79.58
II、IV	A1 + A2 + B1 + B2	5888	86.21	71.57	42.53	79.24
III	A1 + A2 + B1	5248	84.50	68.84	40.32	77.21

在训练集的比例:(I)以 A1、B1 和 B2 数据为训练集,以 B3 数据集为验证集建立 FCN-VGG19-8s 模型;(II)以 A1、A2、B1 和 B2 数据为训练集,以 B3 数据集为验证集建立 FCN-VGG19-8s 模型。方案二:为探究 B 数据集在训练集的比例变化对模型影响,维持 A 数据集的训练张数,改变 B 数据集在训练集的比例:(III)以 A1、A2 及 B1 数据集为训练集,以 B3 数据集为验证集建立 FCN-VGG19-8s 模型;(IV)以 A1、A2、B1 和 B2 数据集为训练集,以 B3 数据集为验证集建立 FCN-VGG19-8s 模型。在方案设计中,实验 II 和 IV 一致,因此仅需进行实验 I、II 和 III,其识别效果见表 5。

针对不同取样点不同训练集比例的植被识别,模型分类效果有所差异。在方案一中,当训练集为 A1、B1、B2 数据集时,模型的植被像素点识别正确率为 86.50%,训练集为 A1、A2、B1、B2 数据集时,植被的像素点识别正确率为 86.21%,两者的像素点识别正确率基本一致。这是由于 A1 数据集的影像张数足够多,能够提取到各植被的纹理特征,加入 A2 数据集的影像后,对模型的识别的影响不大。在方案二中,当 B 地区数据集的训练张数下降时,即训练集为 A1、A2 和 B1,正确率有所下降,为 84.50%,略低于训练集为 A1、A2、B1 和 B2 数据集的正确率。这是由于不同地区的生态环境不同,如光照强度、海拔、降水量等,相同植被为适应环境在

形态上表现出不同的适应性,同时,由于拍摄条件很难保持一致,如光照强度,相同植被在不同光照强度下的无人机影像上,其 RGB 信息也会发生改变。当加入 B 数据集的数量不足时,模型难以从 A 数据集中提取到另一取样点植被的形态特征和光谱特征。即使如此,FCN-VGG19-8s 模型基于 A 数据集进行训练,只需加入少量的其他地区的数据集,能够提高模型的泛化能力,在其他地区仍具有很好的识别效果。在应用于大范围的树种识别时,可针对某一特定地区,并基于该地区的模型进行迁移学习,减少对其他地区的数据量需求,降低数据收集成本。

4 结论和讨论

本文基于 FCN-VGG19 对植被进行识别,实现端对端的语义分割。实验结果表明全卷积网络模型能够基于可见光航拍影像实现植被的准确识别:(1)在多尺度融合结构中,以 FCN-VGG19-8s 模型为最优,分类精度最高;(2)针对不同地区的数据,模型无法提取到未知的植被纹理特征和光谱特征,模型的分类效果有所下降,但仍具有很好的识别效果。全卷积网络模型不仅能够识别植被的类别,还能测算出植被的覆盖面积,在保证精度的情况下,解决了人工调查耗时耗力的问题,同时也避免了由于人为主观因素所造成的误差。

全卷积网络可以广泛的应用于多项林业应用,如植被生物多样性的计算、植被覆盖面积的测算及入侵物种的监测等。但仍存在改进之处:(1)无人机搭载光学相机进行拍摄便于图像采集,不需要很大的成本,但数据的人工标识是一个费力耗时的过程,后期将通过半监督学习或无监督学习解决数据处理成本大的问题;(2)由于人工标识的成本较大,数据量较少,本文通过加载训练 imagenet 数据库的 VGG19 权重来克服数据量不足的问题,未来将增加更多植被的影像资料,丰富实验数据;(3)不同地区的无人机影像,其所包含的信息量不一样,每类植被的特征为适应环境均有所变化,模型无法学习到未知的植被特征,在推广本模型应用于大范围的植被识别时,应尽可能采集到各个区域的影像信息。(4)由于无人机影像存在较多的航向,且数据经过筛选,无人机影像不足以覆盖整个地区,无法测算整个地区各植被覆盖面积,因此模型在测算面积时无法测算出整个地区各植被的覆盖面积,未来将收集更为完整的图像数据,透过图像拼贴技术,制作出完整的地区航拍图像,并结合 FCN-VGG19 测算出整个地区各植被的覆盖面积。

参考文献 (References)

- [1] WILLIS K J. The state of the world's plants report-2017 [R]. London, Royal Botanic Gardens, Kew, 2017
- [2] 李钰微,贾坤,魏香琴,等. 中国北方地区植被覆盖度遥感估算及其变化分析[J]. 国土资源遥感, 2015, **27**(2): 112 - 117 [LI Yuwei, JIA Ku, WEI Xiangqing, et al. Fractional vegetation cover estimation in northern China and its change analysis[J]. Remote Sensing for Land and Resources, 2015, **27**(2): 112 - 117]
- [3] 丁雷龙,李强子,杜鑫,等. 基于无人机图像颜色指数的植被识别[J]. 国土资源遥感, 2016, **28**(1): 78 - 86 [DING Leilong, LI Qiangzi, DU Xin, et al. Vegetation extraction method based on color indices from UAV images[J]. Remote Sensing for Land and Resources, 2016, **28**(1): 78 - 86]
- [4] 宋仁飞,林辉,严恩萍,等. 洞庭湖湿地植被分类研究[J]. 中南林业科技大学学报, 2015, **35**(11): 27 - 31 [SONG Renfei, LIN Hui, YAN Enping, et al. Dongting lake wetland vegetation classification[J]. Journal of Central South University of Forestry and Technology, 2015, **35**(11): 27 - 31]
- [5] 李梦莹,胡勇,王征禹. 基于 C5.0 决策树和时序 HJ-1A/B CCD 数据的神农架林区植被分类[J]. 长江流域资源与环境, 2016, **25**(7): 1070 - 1077 [LI Mengying, HU Yong, WANG Zhengyu. Study on vegetation classification in Shennongjia forest district based on C5.0 decision tree and HJ-1A/B data [J]. Resources and Environment in the Yangtze Basin, 2016, **25**(7): 1070 - 1077]
- [6] 付伟,徐涵秋,王美雅,等. 南方红壤典型水土流失区植被分类及植被类型变化的遥感评估——以福建省长汀县河田地区为例[J]. 遥感技术与应用, 2017, **32**(3): 546 - 555 [FU Wei, XU Hanqiu, WANG Meiya, et al. Vegetation classification and variation assessment in a typical red soil erosion area in southern China: Hetian, Changting County of Fujian Province [J]. Remote Sensing Technology and Application, 2017, **32**(3): 546 - 555]
- [7] 贾明明,任春颖,刘殿伟,等. 基于环境星与 MODIS 时序数据的面向对象森林植被分类[J]. 生态学报, 2014, **34**(24): 7167 - 7174 [JIA Mingming, REN Chunying, LIU Dianwei, et al. Object-oriented forest classification based on combination of HJ-1 CDD and MODIS-NDVI data [J]. Acta Ecologica Sinica, 2014, **34**(24): 7167 - 7174]
- [8] 刘伟,王聪华,赵尔平,等. 基于分层分级的遥感图像植被分类方法[J]. 中国矿业大学学报, 2016, **45**(4): 828 - 835 [LIU Wei, WANG Conghua, ZHAO Erping, et al. Vegetation classification of remote sensing image based on hierarchical clustering analysis [J]. Journal of China University of Mining and Technology, 2016, **45**(4): 828 - 835]
- [9] 张华,张改改,吴睿. 基于 GF-1 卫星数据的面向对象的民勤绿洲植被分类研究[J]. 干旱区地理, 2017, **40**(4): 831 - 838 [ZHANG Hua, ZHANG Gaigai, WU Rui. Object-based vegetable classification based on GF-1 imagery in Minqin Oasis [J]. Arid Land Geography, 2017, **40**(4): 831 - 838]
- [10] 费运巧,刘文萍,骆有庆,等. 森林病虫害监测中的无人机图像分割算法比较[J]. 计算机工程与应用, 2017, **53**(8): 216 - 223 [FEI Yunqiao, LIU Wenping, LUO Youqing, et al. Comparison of algorithms for unmanned aerial vehicle image segmentation in monitoring forest diseases and insect pests [J]. Computer Engineering and Applications, 2017, **53**(8): 216 - 223]
- [11] 雷添杰,李长春,何孝莹. 无人机航空遥感系统在灾害应急救援中的应用[J]. 自然灾害学报, 2011, **20**(1): 178 - 183 [LEI Tianjie, LI Changchun, HE Xiaoying. Application of aerial remote sensing of pilotless aircraft to disaster emergency rescue [J]. Journal of Natural Disasters, 2011, **20**(1): 178 - 183]
- [12] KRIZHEVSKY A, SUTSKEVERI, HINTON G E. Imagenet classification with deep convolutional neural networks [C]// Proceeding of the Advances in Neural Information Processing Systems. Harrahs and Harveys, Lake Tahoe, USA: NIPS, 2012: 1097 - 1105
- [13] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015: 1 - 9
- [14] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014:

- 580 – 587
- [15] GIRSHICK R. Fast R-CNN [C]// Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile; IEEE, 2015; 1440 – 1448
- [16] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. Computer Science, 2014
- [17] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA; IEEE, 2015; 3431 – 3440
- [18] 林升钰, 查轩. 基于 GIS 的安溪县土壤侵蚀潜在危险度评价[J]. 水土保持研究, 2010, 17(06): 60 – 63 [LIN Shengyu, CHA Xuan. Study on potential danger of soil erosion based on GIS in Anxi County[J]. Research of Soil and Water Conservation, 2010, 17(06): 60 – 63]
- [19] 张学俭, 沈雪建. 治理与开发相结合实现生态与经济效益双赢[J]. 中国水土保持, 2004(09): 5 – 6 [ZHANG Xuejian, SHEN Xuejian. Realizing win-win of ecological and economic returns through combination of management and development[J]. Soil and Water Conservation in China, 2004(09): 5 – 6]
- [20] 易盟, 隋立春. 基于改进全卷积神经网络的航拍图像语义分类方法[J]. 计算机工程, 2017, 43(10): 216 – 221 [YI Meng, SUI Lichun. Aerial image semantic classification method based on improved full convolution neural network [J]. Computer Engineering, 2017, 43(10): 216 – 221]
- [21] 曹林林, 李海涛, 韩颜顺, 等. 卷积神经网络在高分遥感影像分类中的应用[J]. 测绘科学, 2016, 41(9): 170 – 175 [CAO Linlin, LI Haitao, HAN Yanshun, et al. Application of convolution neural networks in classification of high resolution remote sensing imagery[J]. Science of Surveying and Mapping, 2016, 41(9): 170 – 175]
- [22] ZEILER M D, TAYLOR G W, FERGUS R. Adaptive deconvolutional networks for mid and high level feature learning [C]// Proceeding of the IEEE International Conference on Computer Vision. Barcelona, Spain; IEEE, 2011: 2018 – 2025

Unmanned Aerial Vehicle Vegetation Image Recognition using Deep Semantic Segmentation

LIN Zhiwei^{1,2,3*}, TU Weihao¹, HUANG Jiahang¹, DING Qilu¹, LIU Jinfu^{1,4}

(1. College of Computer and Information Science, Fujian Agriculture and Forestry University, Fujian Fuzhou 350002, China;

2. College of Forestry, Fujian Agriculture and Forestry University, Fujian Fuzhou 350002, China;

3. Forestry Post-doctoral station of Fujian Agriculture and Forestry University, Fujian Fuzhou 350002, China;

4. Key Laboratory for Ecology and Resource Statistics of Fujian Province, Fujian Fuzhou 350002, China)

Abstract: To efficiently monitor and obtain information on vegetation, a vegetation identification technique is required that yields high classification accuracy and can easily be promoted. To provide precise information on vegetation types to dynamically monitor changes in vegetation, an unmanned aerial vehicle (UAV) was equipped with an optical camera to obtain optical images of vegetation, and the deep semantic segmentation technique was used to construct a model for identifying types of vegetation. First, the FCN-VGG19 vegetation recognition model was constructed on the basis of optical images of a permanent gully captured in Anxi County in China's Fujian Province, to assess the effects of feature fusion with various structures on their performance and calculate the coverage of each vegetation type. Subsequently, optical images captured at another sampling site were used as a validation set to analyze the performance of the FCN-VGG19 model with transfer learning and verify its stability. The experimental results indicated that: (1) the FCN-VGG19-8s model constructed using UAV images at an aerial height of 20 m achieved the highest accuracy of 86.30%. (2) The recognition accuracy of the FCN-VGG19-8s model was higher than that of the FCN-VGG19-32s; an image was randomly extracted from the test dataset, and the calculated coverage of *Pinus massoniana* Lamb., *Miscanthus sinensis* Anderss., and *Eucalyptus citriodora* Hook. f. was 78.38 m², 12.77 m², and 0.89 m², respectively. (3) In the validation of transfer learning, the performance

of the model for identifying dataset B was only slightly affected because the proportion of the training data from dataset A was decreased. Although the identification accuracy was slightly decreased when the training data from dataset B were increased, the identification accuracy still reached 84.5%. The optical images obtained from the UAV were associated with a deep semantic segmentation architecture to identify types of vegetation. In addition, Anxi County was used as the sampling area to demonstrate the robustness of the architecture and evaluate its feasibility by calculating the vegetation coverage. Overall, this research aimed to provide an innovative method for vegetation recognition.

Key words: semantic segmentation; fully convolutional networks; UAV image; vegetation recognition