

魏超宇, 韩文, 庞程, 等. 基于多尺度特征融合和密集连接网络的疏果期黄花梨植株图像分割[J]. 江苏农业学报, 2021, 37(4): 990-997.

doi: 10.3969/j.issn.1000-4440.2021.04.023

基于多尺度特征融合和密集连接网络的疏果期黄花梨植株图像分割

魏超宇, 韩文, 庞程, 刘辉军

(中国计量大学计量测试工程学院, 浙江 杭州 310018)

摘要: 由于自然环境下果蔬植株的果实、枝干和叶片等目标尺度不一、边缘不规则, 因此造成其准确分割较为困难。针对该问题, 提出 1 种多尺度特征融合和密集连接网络 (Multi-scale feature fusion and dense connection networks, MDNet) 以实现黄花梨疏果期植株图像的准确分割。在研究中借鉴了编码-解码网络, 其中编码网络采用 DenseNet 对多层特征进行复用和融合, 以改善信息传递方式; 解码网络使用转置卷积进行上采样, 结合跳层连接融合浅层细节信息与深层语义信息; 在编码、解码之间加入空洞空间金字塔池化 (Atrous spatial pyramid pooling, ASPP) 用于提取不同感受野的特征图以融合多尺度特征, 聚合上下文信息。结果表明, ASPP 有效提高了模型的分割精度, MDNet 在测试集上的平均局域重合度 (*MIoU*) 为 77.97%, 分别较 SegNet、Deeplabv2 和 DNet 提高了 8.10 个、5.77 个和 2.17 个百分点, 果实、枝干和叶片的像素准确率分别为 93.57%、90.31% 和 95.43%, 实现了黄花梨植株果实、枝干和叶片等目标的准确分割。在翠冠梨植株图像的独立测试中, *MIoU* 为 70.93%, 表明该模型具有较强的泛化能力, 对自然环境下果蔬植株图像的分割有一定的参考价值。

关键词: 黄花梨植株; 多尺度特征融合; 密集连接网络; 图像分割; 空洞空间金字塔池化 (ASPP); 感受野
中图分类号: TP391.4 **文献标识码:** A **文章编号:** 1000-4440(2021)04-0990-08

Image segmentation of Huanghua pear plants at fruit-thinning stage based on multi-scale feature fusion and dense connection network

WEI Chao-yu, HAN Wen, PANG Cheng, LIU Hui-jun

(College of Metrological Technology and Engineering, China Jiliang University, Hangzhou 310018, China)

Abstract: As the fruits, branches and leaves of fruit and vegetable plants vary in scales and margins under natural environment, it is difficult to segment them accurately. To solve the problem, a multi-scale feature fusion and dense connection network (MDNet) was proposed to achieve the accurate segmentation of Huanghua pear images at fruit-thinning stage. The coding-decoding network was adopted in this study and DenseNet was adopted to reuse and fuse multi-layer features in the coding network, so as to improve the transfer mode of information. The transposed convolution was used to carry out up-sampling in the decoding network, and the skip-layer connection was also employed to fuse shallow detail information and deep semantic information.

The atrous spatial pyramid pooling (ASPP) was added between coding and decoding to extract future maps with different receptive fields so as to fuse multi-scale feature and aggregate context information. The results showed that ASPP improved the segmentation accuracy of the MDNet model effectively. The mean intersection over union (*MIoU*) of the MDNet on the test set was 77.97%, improved by 8.10, 5.77 and 2.17 percentage points respectively compared with Seg-

收稿日期: 2020-10-13

基金项目: 国家自然科学基金项目 (51606181); 国家级大学生创新创业训练计划项目 (201910356009)

作者简介: 魏超宇 (1995-), 男, 浙江嘉兴人, 硕士研究生, 研究方向为计算机视觉、深度学习等。(E-mail) P1802085258@cjl.u.edu.cn

通讯作者: 刘辉军, (E-mail) liuhj@cjl.u.edu.cn

Net, Deeplabv2 and DNet. The pixel accuracy for fruits, branches and leaves was 93.57%, 90.31% and 95.43%, respectively. Therefore, the accurate segmentation of fruits, branches and leaves of Huanghua pear plants was realized. The *MIoU* was 70.93% in the independent test of Cuiguan pear plants images, indicating that the model had strong generalization ability and was particularly valuable for the image segmentation of fruit and vegetable plants in natural environment.

Key words: Huanghua pear plants; multi-scale feature fusion; dense connection network; image segmentation; atrous spatial pyramid pooling (ASPP); receptive field

植株分割对于作物表型信息获取、生长状态监测和自动采收等具有重要意义。目前,机器视觉技术已广泛用于小麦^[1]、水稻^[2]和玉米^[3]等大田作物的分割。果蔬多采用矮株密植,冠层结构复杂,在自然环境下,存在果叶近色、光照度不均、枝叶遮挡和尺度不一等问题,使得对其果实、枝干和叶片等部位的准确分割具有挑战性。黄花梨是中国南方地区广泛种植的优质水果,本研究选择标准化商业果园的疏果期黄花梨植株作为研究对象,以期为研究大宗果蔬植株的分割提供参考。

果蔬植株的分割是智能农业领域的研究热点之一。Cheng 等^[4]利用白色幕布简化了苹果植株的背景,在 YCbCr 颜色空间依次设定阈值分割出果实和叶片,并根据果实数量、面积及果叶比等参数建立了估产模型。赵德安等^[5]通过白炽灯照明削弱了图像中的阴影,将色差法分割结果与二次分割得到的果实高亮反光区域进行融合,提出了一种适用于夜间的苹果识别方法。Zhang 等^[6]将简单线性迭代聚类 (Simple linear iterative clustering, SLIC) 与 K-means 聚类方法结合,提出了一种针对黄瓜植株的病害叶片分割方法,其中对炭疽病叶片的像素识别准确率达到了 92.15%。Zemmour 等^[7]提出了一种多颜色空间融合的自适应阈值分割方法,在自然环境下对黄椒分割的 F_1 值为 99.31%,对绿葡萄分割的 F_1 值为 73.52%。此类基于特征工程的方法,在二元分类问题中可以取得较高的精度,但是对多类分割问题的研究较少,其特征设计过程复杂、费时,依赖于目标的颜色、纹理和形态等特征,且多需要特定的环境条件约束,在非结构化环境下,该方法难以保证鲁棒性。

近年来,深度卷积神经网络 (Deep convolutional neural networks, DCNN) 发展迅速,其可直接将原始图像作为输入,通过自学习提取特征,表现出了优异的泛化性,在图像分类、目标检测和语义分割等领域取得了巨大成就^[8]。其中, R-FCN、YOLO 和 SSD 等

主流目标检测网络已经在果蔬领域得到了应用,有效提高了遮挡、阴影和果叶近色等复杂条件下果实的检测和定位精度^[9-10]。然而,目标检测的框选方式不能反映果蔬植株中果、枝、叶目标的轮廓与姿态等信息。语义分割是对图像进行像素级分类,融合了图像分割和目标检测 2 个任务^[11],适用于多类分割问题。典型的语义分割网络多采用编码-解码结构,如 FCN^[12]、SegNet^[13] 和 DeepLab^[14] 等。Lin 等^[15]利用 FCN 对绿色番石榴果实和枝干进行分割,并根据分割结果估计果实的姿态,指导机器人采摘。Majeed 等^[16]为了进行落叶期苹果植株的自动整枝,从点云数据中获取植株的深度信息和红绿蓝 (Red-green-blue, RGB) 图像,在利用深度信息去除背景中的植株后,使用 SegNet 对颜色近似的树干、树枝进行语义分割,像素准确率分别为 92%、93%。Ambrozio 等^[17]提出了 1 种基于 Deeplab 的自然环境下苹果花语义分割的方法,结果显示, IoU (Intersection over union) 为 71.4%,在对梨花、桃花图像的测试中也取得了优异的效果,表明该模型具有较强的泛化能力。Kang 等^[18]设计了 1 种多功能网络 DAS-Net-V2,可在对果实进行检测的同时对枝干进行语义分割,在不同时间段拍摄的图像上的测试结果表明,模型对光照具有较强的鲁棒性,最后通过对 RGB-D 图像进行点云可视化,实现了植株的三维重建。总体看出,基于 DCNN 的语义分割有利于克服人工设计特征的不足,减少果叶近色和环境光照的影响,提高非结构化环境下果蔬植株的分割精度。

植株图像分割过程中果、枝、叶等目标的细节信息有助于农业机器人更好地感知和理解复杂的植株环境,提高目标的定位精度,并减少对植株的损伤。DCNN 通过卷积与池化交替的方式进行连续的下采样,具有空间上的平移不变性,对位置信息不敏感,有利于图像分类,但是对于语义分割是不利的,会丢失目标的细节信息,使得目标边缘过于平滑。并且大多数语义分割网络检测尺度单一,存在大目标过

分割和小目标丢失^[19]的问题。此外,果蔬植株中果、枝、叶等目标边缘不规则、尺度不一,又由于相互间的遮挡,目标形态往往变得更为复杂,使得果蔬植株的语义分割较为困难。

DenseNet^[20]采用密集连接实现了多层特征的复用和融合,有利于改善信息的传递方式。跳层连接可融合浅层细节信息与深层语义信息^[12]。空洞空间金字塔池化(ASPP)^[21]则通过多个并行的空洞卷积得到不同感受野的特征图并将其融合,可提取多个尺度上的特征。因此,本研究利用 ASPP 提取并融合多尺度特征,结合 DenseNet 和跳层连接恢复细节信息,提出了 1 种基于编码-解码结构的密集连接网络(MDNet),并分别在疏果期黄花梨、翠冠梨植株数据集上进行了测试。

1 数据与方法

1.1 图像采集

试验果园位于杭州市钱塘江南岸(地理位置为 120.30°E、30.26°N),为国家级蜜梨标准化示范基地。图像采集时间为 2018 年 6 月上旬,此时黄花梨植株正值疏果期,并已完成初次疏果,其冠层枝叶遮挡严重,果、叶均呈近绿色,果径约为 40 mm,具有一定的代表性。图像采集设备为 FUJIFILM F47fd 佳能数码相机,图像分辨率为 3 488 像素×2 616 像素,图像保存格式为 JPEG。图像采集时,为了模拟机器人作业环境,采用三脚架将相机固定于植株行间,相机至主枝干的水平距离约为 1.2 m,高度随植株冠层的分布而调整。同期采集了翠冠梨植株图像,用于测试模型的泛化能力。

1.2 数据集的制作

共采集 400 幅黄花梨植株图像,随机选取 250 幅组成训练集,50 幅组成验证集,其余 100 幅组成测试集;共采集 160 幅翠冠梨植株图像,全部作为独立测试集。为了减少试验的运行时间,将图像分辨率缩小至 872 像素×654 像素。图 1 为数据集的制作过程,根据植株的特点,利用 Labelme 将图像标注为果实、枝干、叶片和背景(包括草地、天空和远景中的黄花梨植株等)等 4 个类别,并将不同类别的像素值映射为不同颜色,得到标注图像(图 1b)。可以看出,果、枝、叶等目标的边缘不规则,目标之间的尺度变化较大。

针对果园中光照复杂多变和枝叶姿态不定的问

题,分别采用亮度变换、图像旋转处理对训练集进行数据增强。其中亮度变化处理将图像亮度随机调整为原始亮度的 50%~150%,图像旋转处理则对图像进行 $\pm 30^\circ$ 的随机旋转。数据增强后,训练集图像数量增加为原来的 3 倍,共 750 张。由于网络的下采样倍率为 16,输入图像高度、宽度须是 16 的倍数,因此将图像随机裁剪为 512 像素×512 像素。对于旋转后的图像,仅对其中心部分进行裁剪,以避免引入黑色边缘中的无关像素,详见图 1c、图 1d 和图 1e。

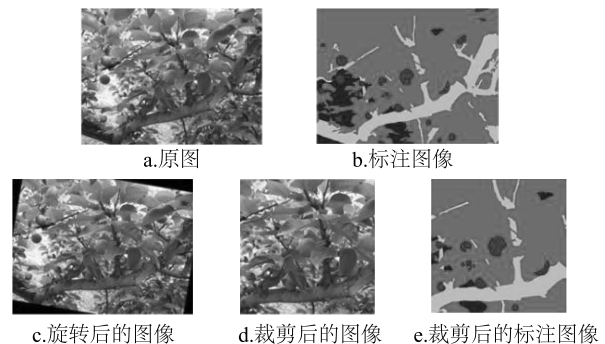


图 1 数据集的制作过程

Fig.1 Manufacturing process of dataset

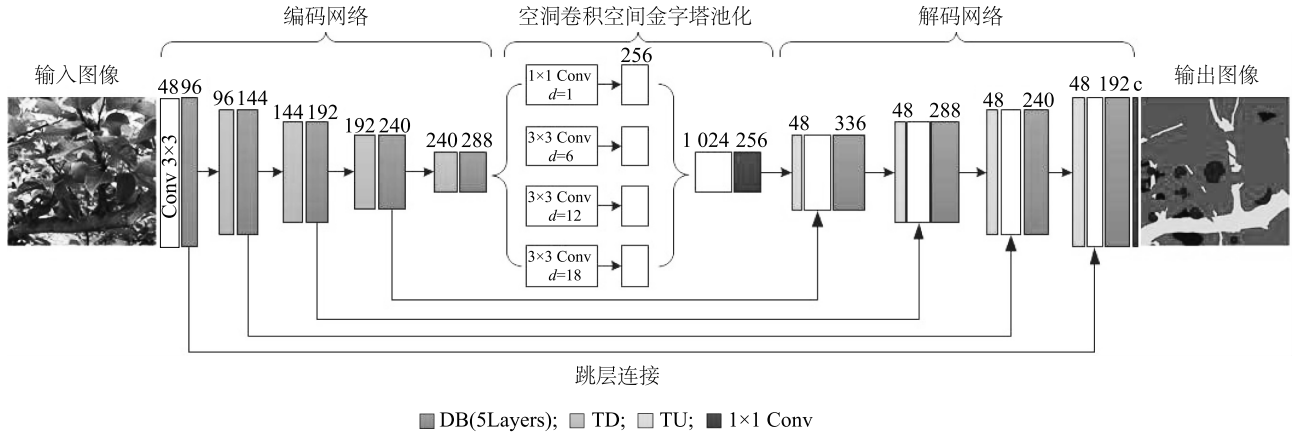
1.3 基于 MDNet 的黄花梨植株分割网络

MDNet 由编码网络、ASPP 和解码网络共 3 个部分组成,网络结构见图 2。首先,编码网络对输入的植株图像进行特征提取。然后,ASPP 通过提取不同感受野的特征图进一步融合多尺度特征。最后,解码网络对特征图进行上采样,实现像素的分类,输出分割结果。

1.3.1 编码网络 目前,AlexNet、VGG 等源于图像分类任务的网络已被用于语义分割,其通过卷积与池化进行特征提取。池化层可以增加输出神经元的感受野大小,从而提取更为抽象的特征,但是过多的池化层也会使特征图的分辨率锐减,导致细节信息丢失,这对语义分割而言是非常不利的。

DenseNet 借鉴残差结构,在卷积层间增加了密集连接,具有特征复用和深层监督的特点^[20],优化了信息传递方式,有利于果蔬植株细节信息的传递。DenseNet 有 5 个池化层,为了减少细节信息的丢失以更适于果蔬植株的语义分割任务,本研究删去了其中最后 1 个池化层。

DenseNet 由密集块(Dense block,DB)与下采样(Transition down,TD)构成。TD 由 1×1 卷积层和 2×



柱子上方的数字表示特征图的数量;conv:卷积(Convolution);1×1、3×3:卷积核的大小; d :空洞率; $c=4$ 。

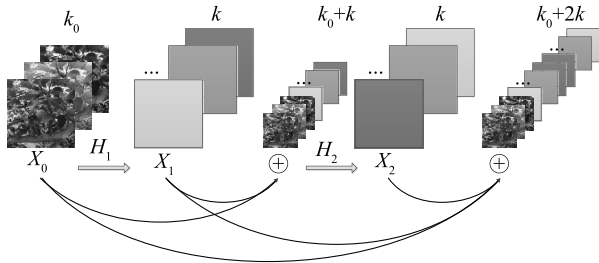
图2 基于MDNet的黄花梨植株语义分割网络结构

Fig.2 Structure of semantic segmentation network of Huanghua pear plants based on multi-scale feature fusion and dense connection network (MDNet)

2 最大池化层组成,DB 中第 i 层的输出 X_i 可表示为如下公式:

$$X_i = H_i([X_0, X_1, \dots, X_{i-1}]) \quad (1)$$

式中, H_i 为非线性变换,由 BN 层、ReLU 激活函数和 3×3 的卷积层组成, $[X_0, X_1, \dots, X_{i-1}]$ 表示第 0 层到第 $i-1$ 层输出特征图的叠加。经 H_i 输出的特征图数量为 k ,称为增长率(本研究中 k 为 12)。令 X_0 的特征图数量为 k_0 ,则 DB 输出的特征图数量为 $k_0 + (i-1) \times k$ 。图 3 是 1 个 3 层 DB 的结构示意,其中“ \oplus ”表示特征图的叠加。



k_0 : X_0 的特征图数量; k : 经 H_i 输出的特征图数量; H_i : 非线性变换。

图3 3层密集块的结构示意

Fig.3 Structure schematic of three-layer dense block

1.3.2 空洞空间金字塔池化(ASPP) ASPP 的主要结构为空洞卷积,空洞卷积可在不降低特征图分辨率的前提下增大感受野,其一维公式如下:

$$y[i] = \sum_{n=1}^N x[i+dn] \cdot w[n] \quad (2)$$

式中, $y[i]$ 为输出; $x[i]$ 为输入; N 为卷积核大小; $w[n]$ 为卷积核中第 n 个参数; d 为空洞率,表示

在卷积核的 2 个连续值之间插入 $d-1$ 个空洞。

大小为 N 、空洞率为 d 的卷积核获得的感受野 (R)^[22] 可表示为如下公式:

$$R = (d-1) \times (N-1) + N \quad (3)$$

如图 4a 所示,当 d 为 1 时(即标准卷积),输出神经元的感受野较小,更关注局部信息,上层信息被利用多次,冗余信息较多^[22]。图 4b 为 $d=2$ 的空洞卷积,空洞卷积允许卷积核与间隔为 $d-1$ 的输入神经元进行运算,可以在不增加参数的情况下增大每个输出神经元的感受野,有利于利用全局信息、聚合上下文信息。

经空洞卷积输出的特征图具有固定的感受野,只能利用单一尺度上的特征^[23]。ASPP 采用多个不同的空洞卷积得到不同感受野的特征图并将其融合,以提取多尺度特征。如图 2 所示,ASPP 结构包含 4 个并行分支,每个分支中卷积的空洞率不同,相应输出特征图的感受野也不同,可以提取不同尺度上的特征。各分支上特征图的通道数均为 256 个,叠加后得到通道数为 1 024 个的特征图,包含多个尺度的信息,最后采用 1×1 卷积进行跨通道信息的融合和降维。

果蔬植株上果、枝、叶等目标尺度不一,经 ASPP 输出的特征图的感受野变得更大,尺度范围也更广,可在多个尺度进行特征提取,以聚合更多的上下文信息,实现果蔬植株的多尺度检测。由于不同分割任务的目标尺度范围不同,因此空洞率的设置需要根据具体情况进行讨论。

1.3.3 解码网络 解码网络采用 DB 与上采样

(Transition up, TU) 结合的方式进行, 将特征图分辨率恢复至原图大小。TU 为步长为 2 的 3×3 转置卷积, 卷积核数量设置为 $(i-1) \times k$, 以避免上采样过程中特征图通道数的快速增长。为了更好地恢复细节信息, 在解码过程中加入了跳层连接, 将编码中的

DB 输出与解码中对应的 TU 输出叠加, 以实现浅层细节信息与深层语义信息的融合, 最后通过 1×1 卷积输出分辨率为原图大小、通道数为类别数 (在本研究中类别数为 4) 的特征图, 之后对每个像素进行分类, 并将交叉熵作为损失函数。

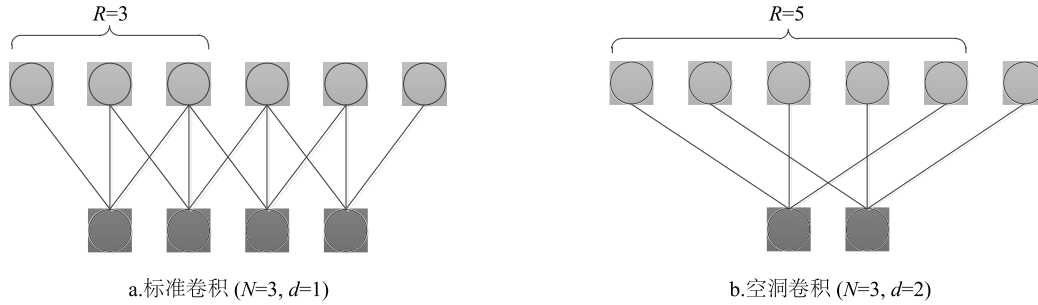


图 4 标准卷积与空洞卷积的一维表示

Fig.4 One dimensional representation of standard convolution and dilated convolution

2 结果与分析

本试验在 Windows 10 系统下运行, 基于 NVIDIA GTX1080Ti GPU 的硬件平台搭建 Tensorflow 深度学习框架, Python 版本为 3.6。设置超参数批尺寸为 2, 学习率为 0.001, 每经过 1 个 epoch, 乘以衰减系数 0.995 以更新学习率, epoch 设置为 150。梯度下降方式采用均方根传递 (Root mean square prop, RMSprop), 以减小梯度下降时的振幅, 加速训练。迁移学习是常用的学习方法, 可有效提高模型的精度。由于 MDNet 模型的特征提取网络是基于 DenseNet 改进的, 目前未有公开的预训练模型可供迁移学习, 因此其权重初始化采用常用的 Xavier 方式。

2.1 评价指标

评价指标^[24]采用类别像素准确率 (w_j)、统计像素准确率 (PA) 和平均区域重合度 (MIoU), 各指标的定义见下式:

$$w_j = \frac{p_{jj}}{\sum_{j=0}^{c-1} p_{jm}} \quad (4)$$

$$PA = \frac{\sum_{j=0}^{c-1} p_{jj}}{\sum_{j=0}^{c-1} \sum_{m=0}^{c-1} p_{jm}} \quad (5)$$

$$MIoU = \frac{1}{c} \sum_{j=0}^{c-1} \frac{p_{jj}}{\sum_{m=0}^{c-1} p_{jm} + \sum_{m=0}^{c-1} p_{mj} - p_{jj}} \quad (6)$$

式中, p_{jm} 为本属于 j 类却被预测为 m 类的像素

点数量, c 为类别数; MIoU 反映了分割结果的完整性和准确性, 常被用作最终的评价指标。

2.2 空洞空间金字塔池化参数对模型的影响

为了考察 ASPP 的有效性并确定最佳空洞率, 本研究对比了模型在不同空洞率参数下的训练损失曲线和分割结果, 分别见图 5 和表 1, 其中 DNNet 模型无 ASPP 结构。

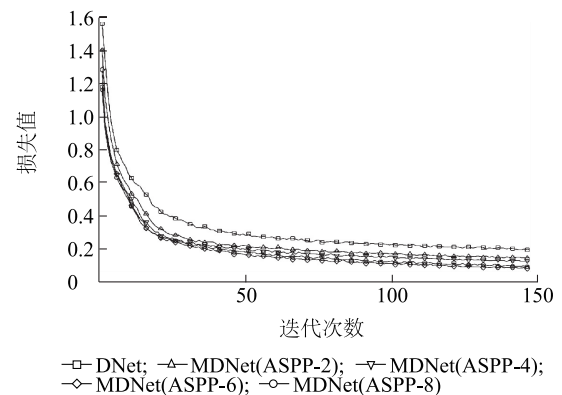


图 5 不同空洞空间金字塔池化 (ASPP) 的 MDNet 模型的训练损失曲线

Fig.5 Loss curve of MDNet models with different atrous spatial pyramid pooling (ASPP)

不同分割任务的目标尺度范围不同, 因此需要寻找最佳空洞率。空洞率过小, 经 ASPP 输出的特征图感受野尺度分布范围也较小, 对于较大目标的分割效果可能较差。空洞率过大, 感受野尺度分布范围也较大, 但是不同感受野之间过于稀疏, 不能充

分提取特征,分割效果也会变差。

表 1 不同 ASPP 对 MDNet 模型分割结果的影响

Table 1 Influence of different atrous spatial pyramid pooling (ASPP) on the segmentation results of MDNet models

模型	空洞率	平均区域重合度 (%)
DNet	—	75.80
MDNet(ASPP-2)	1,2,4,6	77.20
MDNet(ASPP-4)	1,4,8,12	77.36
MDNet(ASPP-6)	1,6,12,18	77.97
MDNet(ASPP-8)	1,8,16,24	77.69

从图 5、表 1 可以看出,ASPP 的加入可使模型的训练损失达到更低的收敛值,有效提高了模型的分割精度,其中 MDNet(ASPP-6)的 $MIoU$ 最高,表明该空洞率较适合黄花梨植株分割任务。由公式(3)可知,ASPP-6 空洞率对应的 4 种感受野分别为 1、13、25 和 37,经 ASPP-6 输出的特征图感受野变得更大、尺度范围更广,可在多个尺度上提取特征。以下关于 MDNet 的讨论均基于 MDNet(ASPP-6)模型。

DNet、MDNet 的分割结果示例分别见图 6e、图 6f。在第 1 行场景中,两者均准确分割了较小的果实,效果较为接近。在第 2 行场景中,上方存在较大

且表面有阴影的果实,其局部特征与枝干特征相似。DNet 的感受野相对较小,仅关注目标的局部信息,导致将该果实的一部分误判为枝干。而 MDNet 通过 ASPP 融合多个尺度上的特征,聚合了上下文信息,利用目标像素与周围像素的联系,实现了该果实的准确分割。在第 3 行场景中,由于 DNet 更关注局部信息,因此难以区分远处植株叶片与当前植株叶片。而 MDNet 的感受野更大、范围更广,能够提取更抽象的特征,因而较好地地区分了当前植株与背景,减少了背景的干扰。

2.3 不同模型的语义分割结果比较

为了进一步验证 MDNet 模型对黄花梨植株分割的有效性,本研究将其与 SegNet、Deeplabv2 和 DNet 等模型在测试集上的分割结果进行对比分析。

从表 2 可以看出,MDNet 模型的精度最高,其 $MIoU$ 、 PA 分别为 77.97% 和 93.31%,果实、枝干和叶片的像素准确率分别为 93.57%、90.31% 和 95.43%。MDNet 模型的 $MIoU$ 分别较 SegNet、Deeplabv2 和 DNet 模型提高了 8.10 个、5.77 个、2.17 个百分点。由于背景包含草地、天空和远景中的植株等,目标多且差异大,不同模型的果实、枝干和叶片的像素准确率均高于背景。

表 2 不同模型的测试结果

Table 2 Test results of different models

模型	类别像素准确率(w_i) (%)				统计像素准确率 (%)	平均区域重合度 (%)
	果实	枝干	叶片	背景		
SegNet	82.38	84.78	85.32	63.57	84.23	69.87
Deeplabv2	92.46	86.44	90.39	70.83	89.79	72.20
DNet	91.55	90.52	93.21	68.18	91.70	75.80
MDNet	93.57	90.31	95.43	75.87	93.31	77.97

从图 6 可以看出,不同模型均克服了果叶近色的问题。图 6c 为 SegNet 的分割结果,可见网络有 5 个池化层,细节信息丢失严重,分割结果中果、枝、叶的边缘较粗糙,第 1 行场景中未分割出上方较小的果实,第 2 行场景中的果实及第 3 行场景中的背景均被错误分割,不同类别像素的准确率均为模型中最低的。

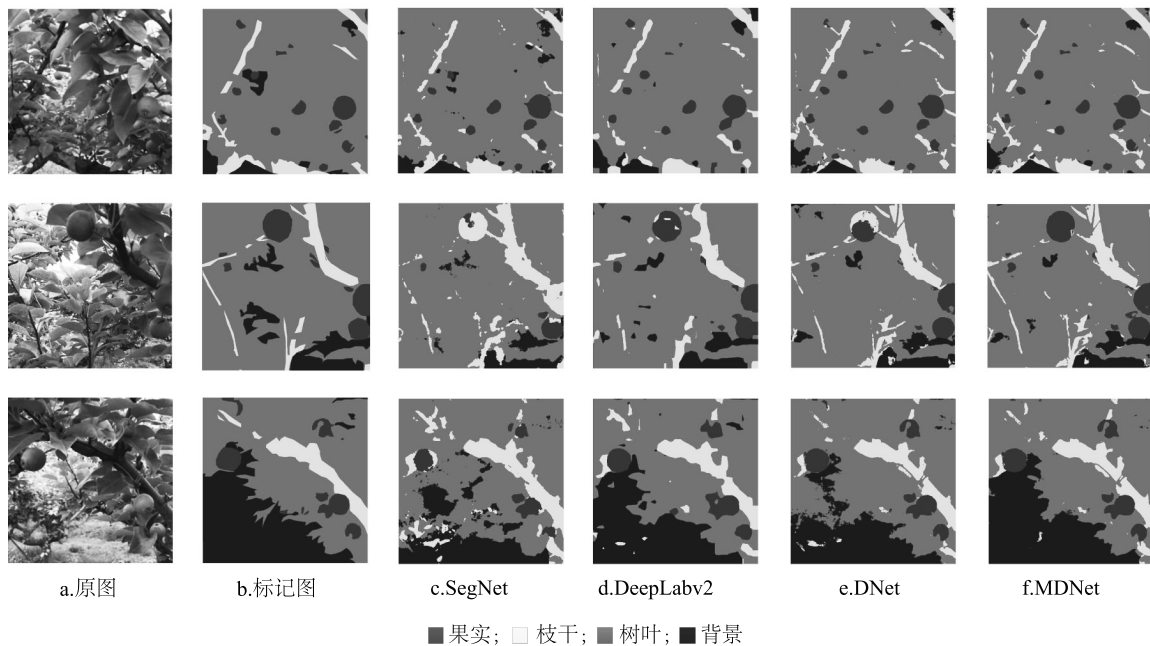
Deeplabv2 与 MDNet 均采用了 ASPP 融合多尺度特征,较准确地分割了不同尺度的果、枝、叶目标,

详见图 6d、图 6f,但 Deeplabv2 采用双线性插值进行上采样,不能较好地恢复目标的细节信息^[23],从分割结果可以看出,果实之间产生了粘连,果、枝、叶边缘多被平滑处理,失去了原有细节。MDNet 采用 DenseNet 作为特征提取网络,对多层特征进行复用和融合,优化了信息传递方式,并通过跳层连接融合浅层细节信息与深层语义信息,可有效地保留并恢复细节信息。MDNet 的分割结果较 Deeplabv2 具有更好的细节性,目标的边缘均得到了准确分割,如图

6 中第 2 行场景中的树干;此外,MDNet 的 $MIoU$ 也超过 Deeplabv2。

综上,本研究提出的 MDNet,通过 ASPP 提取不同感受野的特征图以融合多尺度特征,增强多尺度

目标检测能力,并结合 DenseNet 网络和跳层连接恢复细节信息,优化了边缘分割,有效地提高了非结构化环境下黄花梨植株的语义分割精度。



第 1、2、3 行分别为不同的场景。

图 6 不同模型的语义分割结果示例

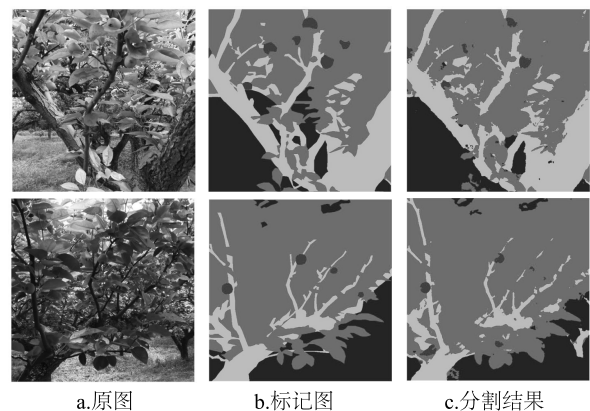
Fig.6 Examples of semantic segmentation results of different models

2.4 对翠冠梨植株的语义分割

为了进一步考察 MDNet 模型的泛化能力,采用 160 幅同期翠冠梨植株图像对模型进行了独立测试, $MIoU$ 达到了 70.93%。图 7a 为输入的原图,可见翠冠梨植株与黄花梨植株的冠层结构较为相似,果实呈扁圆形,略小于同期的黄花梨果实,果、叶颜色更为相近。由图 7b、图 7c 可以看出,在不同光照条件下,翠冠梨枝干与叶片均具有较好的细节。由于翠冠梨颜色、形态的差异,使得翠冠梨果实的边缘信息不同程度地丢失,但多数果实得到了有效检测。草地、天空和远景中的植株也较准确地被分类为背景。由此可见,所建模型对疏果期翠冠梨植株也有较好的分割效果,具有较强的泛化能力。

3 讨论

本研究采用 ASPP 提取并融合多尺度特征,结合 DenseNet 网络与跳层连接,提出了 1 种基于 MDNet 的果蔬植株语义分割方法。在疏果期黄花梨植



第 1、2 行分别为不同场景。

图 7 MDNet 对翠冠梨植株的语义分割

Fig.7 Semantic segmentation of Cuiguang pear plants by MDNet

株数据集上的试验结果表明,ASPP 通过提取不同感受野的特征图融合多尺度特征,聚合了上下文信息,提高了黄花梨植株的语义分割精度,当空洞率为 1、6、12 和 18 时,MDNet 模型的分割精度最高。所

建模型 MDNet 有效克服了自然条件下果叶近色、光照度不均、枝叶遮挡及尺度不一等问题的影响,且较好地保留了植株的细节信息。在黄花梨植株测试集上,该模型的 $MIoU$ 为 77.97%,效果优于 SegNet、DeepLabv2 和 DNet 等模型;在翠冠梨植株独立测试集上,该模型的 $MIoU$ 达到 70.93%,表现出较强的泛化能力。本研究所提方法对自然环境下果蔬植株的分割有一定参考价值。后续将研究在该方法的基础上结合点云信息来实现植株的三维重建,从而提高农业机器人对复杂植株环境的感知和理解水平。

参考文献:

- [1] GUO W, RAGE U K, NINOMIYA S. Illumination invariant segmentation of vegetation for time series wheat images based on decision tree model[J]. Computers and Electronics in Agriculture, 2013, 96: 58-66.
- [2] XIONG X, DUAN L F, LIU L B, et al. Panicle-SEG: a robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization[J]. Plant Methods, 2017, 13(1): 104.
- [3] 苏伟,蒋坤萍,郭浩,等. 地基激光雷达提取大田玉米植株表型信息[J]. 农业工程学报, 2019, 35(10): 125-130.
- [4] CHENG H, DAMEROW L, BLANKE M, et al. ANN model for apple yield estimation based on feature of tree image[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(1): 14-19.
- [5] 赵德安,刘晓洋,陈玉,等. 苹果采摘机器人夜间识别方法[J]. 农业机械学报, 2015, 46(3): 15-22.
- [6] ZHANG S W, WANG H X, HUANG W Z, et al. Plant diseased leaf segmentation and recognition by fusion of superpixel, K-means and PHOG[J]. Optik, 2017, 157: 866-872.
- [7] ZEMMOUR E, KURTSEY P, EDAN Y. Automatic parameter tuning for adaptive thresholding in fruit detection[J]. Sensors, 2019, 19(9): 2130-2151.
- [8] 卢宏涛,张秦川. 深度卷积神经网络在计算机视觉中的应用研究综述[J]. 数据采集与处理, 2016, 31(1): 1-17.
- [9] 王丹丹,何东健. 基于 R-FCN 深度卷积神经网络的机器人疏果前苹果目标的识别[J]. 农业工程学报, 2019, 35(3): 156-163.
- [10] 薛月菊,黄宁,涂淑琴,等. 未成熟芒果的改进 YOLOv2 识别方法[J]. 农业工程学报, 2018, 34(7): 173-179.
- [11] 魏云超,赵耀. 基于 DCNN 的图像语义分割综述[J]. 北京交通大学学报(自然科学版), 2016, 40(4): 82-91.
- [12] LONG J, SHELHAMER, DARRELL T. Fully convolutional networks for semantic segmentation[C]//IEEE. IEEE conference on computer vision and pattern recognition. Seoul, Korea: IEEE, 2015: 3431-3440.
- [13] 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: 2481-2495.
- [14] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [15] LIN G C, TANG Y C, ZOU X J, et al. Guava detection and pose estimation using a low-cost RGB-D sensor in the field[J]. Sensors, 2019, 19(2): 428-443.
- [16] MAJEED Y, ZHANG J, ZHANG X, et al. Apple tree trunk and branch segmentation for automatic trellis training using convolutional neural network based semantic segmentation[J]. International Federation of Automatic Control, 2018, 51(17): 75-80.
- [17] AMBROZIO D P, AMY T, HENRY M. Multi-species fruit flower detection using a refined semantic segmentation network[J]. IEEE Robotics and Automation Letters, 2018, 3(4): 3003-3010.
- [18] KANG H W, CHEN C. Fruit detection, segmentation and 3D visualisation of environments in apple orchards[J]. Computers and Electronics Agriculture, 2020, 171: 1016-1033.
- [19] NOH H, HONG, HAN B. Learning deconvolution network for semantic segmentation[C]//IEEE. IEEE international conference on computer vision. Los Alamitos, USA: IEEE Computer Society Press, 2015: 1520-1528.
- [20] HUANG G, LIU Z, WEINBERGER K Q, et al. Densely connected convolutional networks[C]//IEEE. IEEE conference on computer vision and pattern recognition. Honolulu, USA: IEEE, 2016: 4700-4708.
- [21] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. International Conference on Learning Representations, 2014(4): 357-361.
- [22] 徐子豪,黄伟泉,王胤. 基于深度学习的监控视频中多类别车辆检测[J]. 计算机应用, 2019, 39(3): 700-705.
- [23] YANG M K, YU K, ZHANG C, et al. DenseASPP for semantic segmentation in street scenes[C]//IEEE. IEEE computer society conference on computer vision and pattern recognition. Piscataway, NJ: IEEE Press, 2018: 3684-3692.
- [24] 李云伍,徐俊杰,刘得雄,等. 基于改进空洞卷积神经网络的丘陵山区田间道路场景识别[J]. 农业工程学报, 2019, 35(7): 150-159.

(责任编辑:徐艳)