doi:10.6041/j.issn.1000-1298.2021.06.026

基于深度语义分割网络的荔枝花叶分割与识别

熊俊涛　刘柏林　钟　灼　陈淑绵　郑镇辉
(华南农业大学数学与信息学院, 广州 510642)

摘要: 针对使用目标检测、实例分割方法无法对复杂自然环境下稠密聚集的荔枝花进行识别的问题,提出一种基于深度语义分割网络识别荔枝花、叶像素并实现分割的方法。首先在花期季节于实验果园拍摄荔枝花图像;然后制作标签,并进行数据增强;构建深度为 34 层的 ResNet 主干网络,在此基础上引入稠密特征传递方法和注意力模块,提取荔枝花、叶片的特征;最后通过全卷积网络层对荔枝花、叶片进行分割。结果表明,模型的平均交并比(mIoU)为 0.734,像素识别准确率达到 87%。本文提出的深度语义分割网络能够较好地解决荔枝花的识别与分割问题,在复杂野外环境中具有较强的鲁棒性和较高的识别准确率,可为智能疏花提供视觉支持。

关键词: 荔枝花; 深度网络; 像素分割; 稠密特征传递; 注意力模块
中图分类号: TP391.41　　**文献标识码**: A　　**文章编号**: 1000-1298(2021)06-0252-07　　**OSID**:

Litchi Flower and Leaf Segmentation and Recognition Based on Deep Semantic Segmentation

XIONG Juntao　LIU Bolin　ZHONG Zhuo　CHEN Shumian　ZHENG Zhenhui
(*College of Mathematics and Informatics*, *South China Agricultural University*, *Guangzhou* 510642, *China*)

Abstract: In recent years, deep learning has gradually developed in flower recognition research, which has a positive impact on the growth management and fruit production of orchard fruit trees. In order to tackle the problem that the densely gathered litchi flowers cannot be recognized by instance segmentation method in natural environment, a deep semantic segmentation network was proposed to recognize and segment flowers and leaves pixels. Firstly, pictures of litchi flowers were shoot in the experimental orchard in the flowering stage, which were taken to make pixel-level images, and then were used for data augmentation. Then a backbone network of 34 layers based on ResNet was built, in which dense features were connected layers by layers and in order to exploit the useful information, attention blocks were also added into the networks. A dense features connection method meant each layer was connected to every other layer in a feed-forward fashion, different from that features were only from the last consecutive layer. Attention block was a mechanism of propagating information useful for the specific task and suppress the useless one. Finally, a full convolution layer was added for image pixel prediction. The average intersection union ratio of the proposed model was 0.734, and the pixel recognition accuracy reached 87%. In summary, with good robustness and high recognition accuracy, the proposed deep semantic segmentation model can solve the problem of litchi flower recognition and segmentation, and provide visual support for intelligent flower thinning.

Key words: litchi flower; deep network; pixel segmentation; dense features transmition; attention module

0　引言

在果园管理中,利用疏花疏果技术控制水果植株等农作物的开花时间、开花数量和果实密度,对提高果实产量和质量具有重要作用。随着种植规模的不断增大、人工成本的不断提高,利用机器视觉对花果进行监控,以辅助疏花疏果、精准施肥等管理显得越来越重要。

相关学者利用机器视觉技术在识别花朵果实、实现智能采摘方面进行了许多研究。熊俊涛等[1]

收稿日期: 2020-09-05　修回日期: 2020-09-30
基金项目: 广东省重点研发计划项目(2019130202230002)、国家自然科学基金项目(32071912)、广东省自然科学基金项目(2018A030313330)和广州市科技计划项目(202002030423)
作者简介: 熊俊涛(1981—),男,副教授,主要从事农业机器人和智能设计与制造研究,E-mail: xiongjt2340@163.com

采用 LED 照明系统获取夜间荔枝果实图像，在 YIQ 颜色空间使用 Otsu、模糊 c 均值聚类和 hough 圆检测，实现了荔枝果实识别分割，准确率达到 95.3%，平均识别时间为 0.46 s。武星等[2]简化 YOLO v3 模型的卷积块数量，采用深度可分离卷积、融合损失函数和多阶段优化方法，提出了 Light－YOLO v3 模型，并用于苹果检测，其平均精度达到 94.69%，平均检测速度为 116.96 f/s。刘芳等[3]提出含有残差模块的 YOLO 模型，融合了多尺度特征检测模块，构建了轻便型的复杂环境下番茄快速检测模型，其准确率达到 96.36%，交并比为 0.833，检测时间为 7.719 ms。此外，结合不同传感器进行的葡萄[4]、芒果[5]、梨[6]、草莓[7]等果实识别研究也越来越多。

相对果实识别而言，花朵识别研究相对较少。ACANSEN 等[8]提出一种非物理接触的花朵检测方法，使用 CCD 相机拍摄雷斯克勒花的彩色图像，根据 RGB 颜色空间的像素直方图确定阈值，并根据阈值将花分割出来，有效实现了花朵数量评估，且避免了物理损伤。WANG 等[9]提出夜间环境下修正的芒果花分割方法，该方法将拍摄效果好的图像作为参照图像，使用归一化放缩法将其他图像灰度放缩到参照图像的灰度范围内，从而起到修正作用。在 RGB 颜色空间分割出芒果花、叶片和枝干后，训练识别花的支持向量机，精确分割出芒果花，并计算花密度，其准确率达到 85.8%。WANG 等[10]提出采用颜色修正方法增强样本，在使用 SURF 方法提取花朵特征后，使用支持向量机实现分类，最后计算芒果花的密度，其准确率达到 86.6%。这些传统的手工特征设计方法一般具有多个阶段，每个阶段的设计需要耗费大量的精力，而且只能处理一些背景简单、规模较小的数据集，其泛化能力较差。

随着深度学习的提出和广泛应用，深度学习方法被用于识别花朵。LIU 等[11]设计了一个 8 层的网络结构，包括 5 层“卷积、池化”组合模块和 3 层全连接层，并配合局部归一方法、重叠池化和 Dropout 技术。本文数据集来源于著名的细粒度识别数据集，Oxford 102 Flower Species Dataset，一共包括 102 个不同种类的花朵、8 189 幅图像。FARJON 等[12]提出苹果花识别、密度计算和开花峰值期预估的方法，首先使用 Faster R－CNN 网络识别树上花朵，然后基于线性回归方法根据识别花的数量和分布预测花密度，最后根据开花量衰减时间判定开花峰值时间，从而协助疏花决策。TIAN 等[13]使用以 U－Net 为主干框架的 Mask R－CNN 模型检测和分割不同成长期的苹果花，其预测精度达到 96.43%，平均交并比为 91.55%。深度特征提取的端到端学习，从低层语义特征通过线性卷积核和非线性激活组合传递到高层语义特征，使各个阶段的特征具有较强的依赖性，特征的表示能力显著。

荔枝花朵稠密聚集、形态上难以区分，在繁花期花朵粘连成簇，目标检测、实例分割方法并不适合，而传统的机器学习方法在复杂环境下效果不佳。针对上述问题，本文提出一种基于深度语义分割网络的像素分割方法，对图像中花朵和叶片的像素进行分类，实现整幅图像的花和叶的识别，以期解决花朵稠密、粘连形态问题。

1　材料与方法

1.1　图像采集与数据集构建

使用 Nikon 单反摄像机拍摄荔枝花图像，焦距、曝光和拍摄模式均为自动。拍摄时间为 2019 年 2 月 28 日至 3 月 25 日，拍摄地点在华南农业大学园艺学院实践基地荔枝园。拍摄垂直角度变化范围为 0°～60°，拍摄条件包括顺光、逆光，远距离和近距离，清晰和模糊，如图 1 所示，共拍摄了 675 幅图像。经数据增强后为 3 375 幅，2 700 幅为训练集，350 幅为测试集，325 幅为训练集，比例约为 8∶1∶1。使用 Matlab 2018 的 Image Labeler 软件人工手动标注荔枝花和周围叶片的标签，标签共分 3 类，分别是背景、花朵和叶片。标注后生成 8 位 PNG 灰度图，0 表示背景（黑色），1 表示花朵（红色），2 表示叶片



图 1　荔枝花数据集部分样本

Fig. 1　Samples of litchi flower data set

(绿色)。原图和可视化标注结果如图 1g、1h 所示。

1.2 模型构建

AlexNet[14]具有较好的的特征提取能力,并被应用、改进以解决各种机器视觉问题,如图形分类、图像识别等。随后提出的 VGG 网络[15]、GoogleNet[16]和 ResNet[17],在深度上不断地堆叠卷积层,在宽度上增加并行的卷积核,再加上归一化、激活函数的运用,从而极大地提升了特征提取能力。

Deeplab V3[18]是 Deeplab 系列的第 3 个版本,源于全卷积网络(Fully convolutional networks, FCN)[19]。FCN 使用 VGG16 的主干网络结构提取图像的高层语义特征,在网络中增加了反卷积结构用于恢复特征的空间信息,并将最后的线性全连接层替换为保留空间信息的全卷积层,实现像素预测,解除了固定网络输入的限制。Deeplab V3 借鉴了 FCN 上采样恢复空间信息和全卷积像素预测的思想,使用空洞卷积、空洞空间金字塔池化(ASPP)结构替代了 FCN 的反卷积上采样。空洞卷积增大提取特征的感受野,使空间信息更加丰富,但又不增加内存消耗;ASPP 结构则使尺度多样化,有利于提取多尺寸变化对象的特征。此后许多语义分割研究都在这些思想的基础上发展。本文深度花朵像素分割模型基于 Deeplab V3 网络的搭建思路,构建一个合适深度的稠密特征传递、多尺度特征提取的主干网络,然后使用全卷积层实现花朵像素的预测和分割。整体模型包括主干网络和像素预测两部分,模型结构如图 2 所示。图 2 中 c 表示特征通道上的拼接操作,d 表示空洞卷积率。

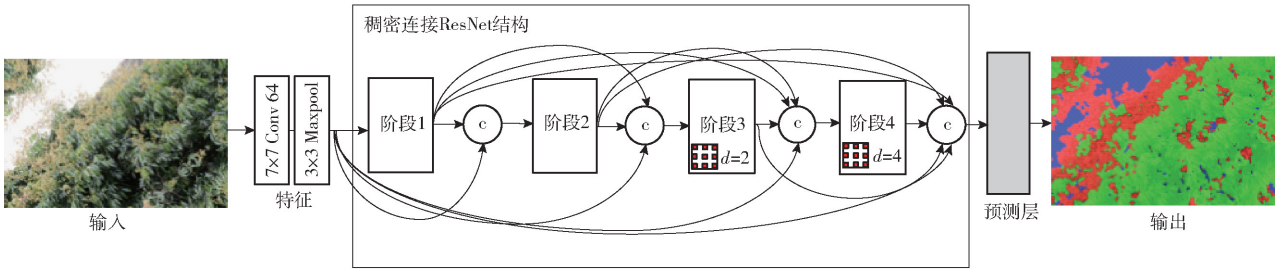


图 2 花朵识别分割网络模型结构图

Fig. 2 Flower recognition and segmentation model structure diagram

1.2.1 主干网络构建

本文基于 Deeplab V3 构建一个适当深度的特征提取主干网络,主干网络一共 34 层,包括两部分。第 1 部分为普通卷积层,卷积核为 7×7,并包括一个尺寸为 3×3 的最大值池化层,粗略提取特征并减小特征尺寸提高语义性。第 2 部分为堆叠残差层,主要有 4 个阶段(Stage),每个阶段包括一定数量的卷积块。在卷积块中使用 3×3 卷积核处理特征,并且特征数量随着卷积块的加深而增加,搭建的网络主干结构如图 3 所示。

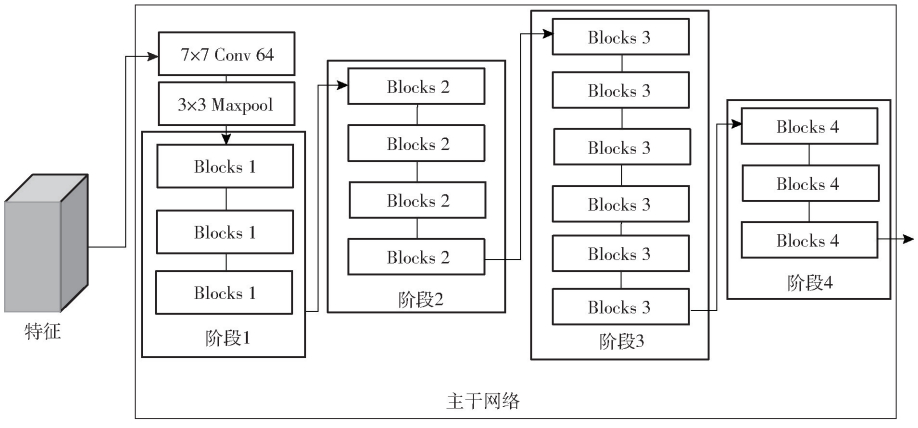


图 3 网络主干结构图

Fig. 3 Backbone structure diagram of network

每个卷积块表示一个残差模块[17],该结构能有效减轻梯度消失和梯度爆炸问题,权重学习更有效。本文卷积块中加入了注意力模块,其结构如图 4 所示。

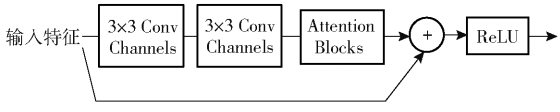


图 4 加入注意力的卷积块结构

Fig. 4 Blocks structure diagram with attention module

在主干网络的所有卷积操作后都有批量归一化层和 ReLU 激活层。Channels 为特征数,阶段 1 和阶段 2 的输入特征数为 64,阶段 3 为 128,阶段 4 为

256，同一个阶段的卷积块之间的特征传递数量不变。Attention Blocks 为注意力模块。

1.2.2　稠密特征传递

DenseNet[20] 提出的深度网络稠密特征传递能够提高特征的重复使用性和减轻梯度消失的问题，DenseNet 的第 $k+1$ 层输入是前 k 层网络的输出，能充分地利用所有浅层特征信息。稠密特征传递的 ASPP 结构构成稠密特征金字塔（Dense ASPP）[21]，通过并行的多尺度空洞卷积核编码特征中不同尺寸物体的信息，有效解决图像中尺寸差异大的物体识别问题，充分提取多尺度物体特征信息。本文借鉴稠密传递和多尺度特征思想，将稠密多尺度特征传递应用在主干残差网络中。主干网络共 4 个阶段，第 i 个阶段的特征输入为前面所有阶段的特征输出，同时为了减少计算量，阶段的输出特征数量不增加；为了增大特征的感受野，在第 3 阶段和第 4 阶段中分别采用空洞卷积率为 2 和 4 的空洞卷积核。稠密特征传递公式为

$$S_i = H_i([S_{i-1}, S_{i-2}, \cdots, S_1]) \tag{1}$$

其中，S_i 表示第 i 个阶段的输出特征，$[S_{i-1}, S_{i-2}, \cdots, S_1]$ 是 $1 \sim i-1$ 个阶段输出特征在通道维度上的拼接，S_0 为第 1 个阶段的输入，即图 3 中卷积块 1 的输入。H_i 表示在第 i 个阶段里的卷积块模块运算。稠密特征传递、多尺度空洞卷积融合的主干结构如图 5 所示。

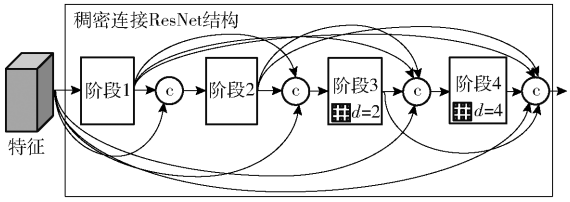


图 5　稠密特征传递、多尺度空洞卷积融合的主干结构图

Fig. 5　Backbone structure diagram of dense-connected features and multiscale astrous convolution

1.2.3　注意力模块

挤压-激励网络（Squeeze-and-excitation networks，SENet）[22] 注意力模块在图像分类中具有显著成效。SENet 是在 ResNet 基础上加入 2 个模块，挤压模块（squeeze bolck）和激励模块（excitation block）。通过挤压，将输入的特征集压缩成一个向量，然后使用可学习的权重实现激励，在损失的反向传播中调整激励权重来筛选前向传递的特征，即“将注意力放在有用的地方”。挤压模块压缩公式为

$$z_c = F_{sq}(u_c) = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(i,j) \tag{2}$$

式中　z_c——输出特征

u_c——输入的第 c 个通道的特征

H、W——特征图的高和宽

F_{sq}——挤压函数

(i,j)——二维特征的像素空间位置

使用均值池化生成该特征的空间描述子，然后输入激励模块公式为

$$s = F_{ex}(z,W) = \mathrm{ReLU}(g(z,W)) = \mathrm{ReLU}(W_2\,\mathrm{sigmoid}(W_1 z)) \tag{3}$$

式中　F_{ex}——激励函数

W_1、W_2——全连接层权值

g——全连接层　　z——输出特征

ReLU、sigmoid——激活函数

$W_1 \in \mathbf{R}^{\frac{C}{r}\times C}$，$W_2 \in \mathbf{R}^{C\times\frac{C}{r}}$，结合两个全连接层和激活函数实现过滤特征功能。$r$ 为压缩率，以减少全连接层的参数量，减轻模型的复杂度。激励模块的最终输出为

$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c u_c \tag{4}$$

式中　$\tilde{x}_c$——输出的第 c 个通道特征

F_{scale}——特征注意力权值 s_c 与输入特征图 u_c 的元素乘积函数

激励权重依赖于输入特征和损失函数，从而实现自适应调整的自注意力功能，其结构如图 6a 所示。

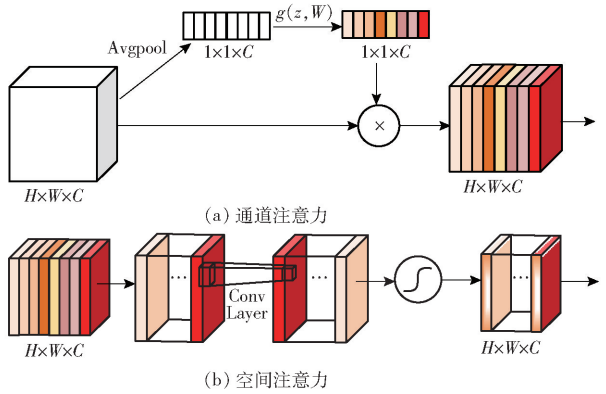


图 6　注意力模块

Fig. 6　Attention module

但在 SENet 中，模型仅关注通道注意力，缺少特征的空间信息注意力[23]。本文在 SENet 基础上，构建关注特征注意力和空间注意力模块，实现特征的筛选，其结构如图 6b 所示。空间特征注意力计算公式为

$$M_s(\tilde{x}_c) = \mathrm{sigmoid}(\mathrm{Conv}^{7\times7}([A_P(\tilde{x}_c); M_P(\tilde{x}_c)])) \tag{5}$$

式中　M_s——正空间注意力计算后的特征

A_P、M_P——对空间注意力计算后特征的均值池化最大值、池化值

$[A_P(); M_P()]$——通道维度上拼接对应池化后的特征图函数

$Conv^{7\times7}$——核为 7×7 的卷积滤波器运算函数

本文使用的注意力模块如图 6 所示。其中 C 为特征通道数。

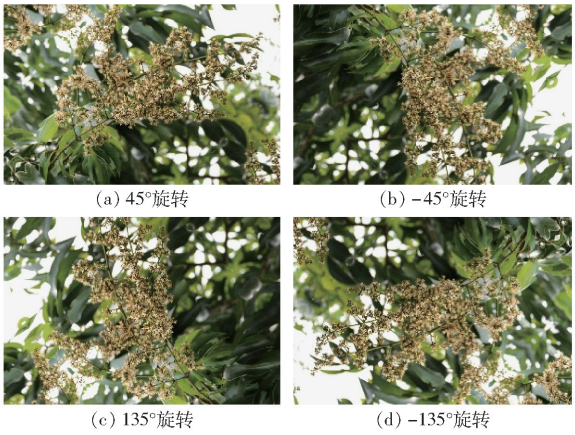
2 实验与结果分析

2.1 实验环境

实验选用 i7-8700 CPU,3.2 GHz 主频,六核十二线程;32 GB 内存;GeForce GTX2080Ti GPU;操作系统为 Ubuntu 16.04;编程语言为 Python;深度学习框架为 Pytorch。

2.2 数据集增强

实验采用线下和线上数据集增强方法对荔枝花图像进行增强。线下增强采用图像旋转方法[24],将采集的 675 幅图像旋转 45°、-45°、135°、-135°后加上原图共 3 375 幅图像,荔枝花图像如图 7 所示。然后在训练时使用线上增强方法,包括翻转和裁剪。旋转图像如图 7 所示。



(a) 45°旋转 (b) -45°旋转

(c) 135°旋转 (d) -135°旋转

图 7 线下数据集图像增强示例

Fig. 7 Samples of offline dataset augmentation

2.3 评价指标

采用像素准确率(Pixel accuracy)、交并比(Intersection over union,IoU)和平均交并比(Mean intersection over union,mIoU)作为评价指标。

2.4 模型训练与对比分析

本文模型采用迁移学习方法,主干网络加载 ImageNet 预训练的 ResNet-34 权重初始化网络。模型迭代次数为 2 500 次,批量训练尺寸为 32;动态调整学习率,初始学习率为 0.01,在迭代 400、1 000 次时迭代学习率为原来的 1/10,动量因子为 0.9,权重衰减为 0.001,优化算法为随机梯度下降(SGD),损失函数为对数交叉熵损失(Cross entropyloss)。本文模型可直接输入荔枝花图像训练,原图尺寸为 1 200 像素×1 800 像素,训练前经过 RGB 的颜色归一化处理,图像随机裁剪为 256 像素×256 像素;根据数据集每个类别标签的比例设置损失函数的类别权重参数,从而解决类别不平衡问题。

模型的主干网络分别选择 18、34、50 层,通过深度对比实验,初步确定适合荔枝花分割的主干网络深度。实验中,除了网络深度变化,其余条件不变,3 个模型的训练损失值如图 8a 所示,验证损失值如图 8b 所示。

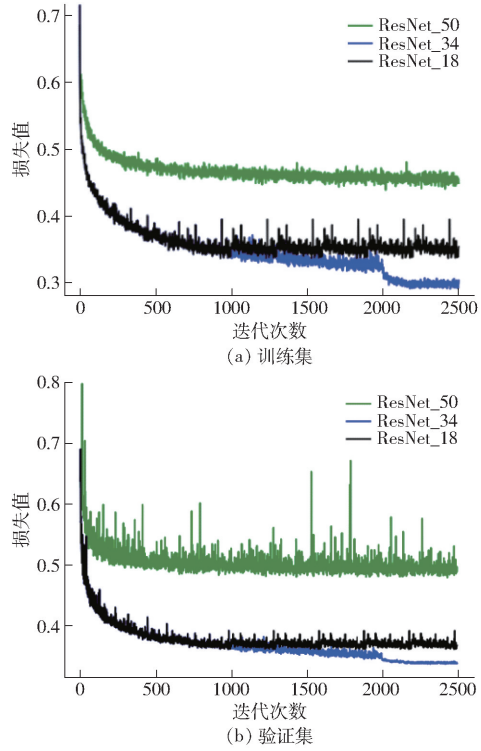


图 8 不同数据集的损失值变化曲线

Fig. 8 Loss value curves of different datasets

从图 8a 可看出,当训练至模型收敛时,ResNet_50 的损失值为 0.47 左右,ResNet_18 损失值为 0.38 左右,ResNet_34 损失值为 0.29 左右,损失值最小。损失值越低,模型在训练集上的拟合能力越强,反映了主干网络的特征提取能力。ResNet_18 的层网络太浅,对花、叶片的分割特征学习不足,表征不充分;而 ResNet_50 拟合能力过强,学习了图像中噪声信息,导致模型损失难以进一步下降。从图 8b 中可看出,ResNet_50 的验证损失波动较大,且损失值也较大;ResNet_34 的验证损失值波动较小,且损失值最小,维持在 0.31 左右,接近训练损失值,说明该模型泛化能力较好。综上所述,模型深度选择 34 层最为合适。

为了进一步验证上述结论,使用测试集进行测试,对比 2.3 节中提及的评价指标,测试集图像共 325 幅,测试结果如表 1 所示。

由表 1 可知,花朵 IoU 在所有类别中最大,叶片比背景稍低。因为拍摄的数据集中,花朵在图像中较为突出,而园林环境的荔枝树比较密集,拍摄图像时,其他荔枝树会形成背景造成干扰,所以导致叶片

表 1　不同深度网络的测试性能指标

Tab. 1　Performance of different depth networks

模型	交并比				像素准确率/%
	花	叶	背景	平均	
ResNet_50	0.635	0.603	0.622	0.620	79.6
ResNet_34	0.698	0.661	0.671	0.677	83.6
ResNet_18	0.693	0.643	0.657	0.664	82.7

的分割难度变大。从平均交并比来看，ResNet_34 达到 0.677，预测像素与真实像素的交集最大；从像素准确率来看，有 83.6%，预测精度最高，进一步验证了 ResNet_34 模型最为合适。

确定模型深度后，需要验证模型增加稠密特征传递和注意力模块的有效性。增加稠密特征传递和注意力模块的模型测试结果如表 2 所示。

表 2　不同模型的性能对比

Tab. 2　Performance comparison of different models

模型	交并比				像素准确率/%
	花朵	叶片	背景	平均	
ResNet_34	0.698	0.661	0.671	0.677	83.6
Dense-ResNet_34	0.703	0.689	0.692	0.695	84.5
ResNet_34 + SE	0.717	0.698	0.694	0.703	84.9
ResNet_34 + CBAM	0.719	0.719	0.707	0.715	85.9
Dense-ResNet_34 + CBAM	0.739	0.740	0.721	0.734	87.0

由表 2 可知，ResNet_34 和 Dense-ResNet_34 的对比验证了稠密特征的作用，ResNet_34 的平均交并比和像素准确率比 Dense-ResNet_34 模型分别低 0.018 和 0.9 个百分点。稠密特征传递使每层的特征提取能够充分利用浅层的所有信息，增强了模型的学习能力。但是丰富的特征可能会造成干扰，所以稠密特征连接未能带来比较明显的性能提升。为了增强特征提取的有效性，在特征提取模块中加入了注意力模块。加入 CBAM 模块的平均交并比和像素准确率比加入 SE 模块的分别高 0.012 和 1 个百分点，CBAM 除了计算通道注意力外，还关注每个通道的空间注意力，比 SE 模块要全面，所以效果更好。综合稠密连接和注意力模块的 Dense-ResNet_34 + CBAM 模型，既充分利用了浅层信息，同时能够筛选有助于目标任务的信息。其平均交并比和像素准确率分别为 0.734 和 87.0%，比 Dense-ResNet_34 分别高 0.039 和 2.5 个百分点，比 ResNet_34 + CBAM 高 0.019 和 1.1 个百分点，有了比较明显的提升。综上所述，在残差主干网络的基础上增加稠密特征传递和注意力模块是有效的。

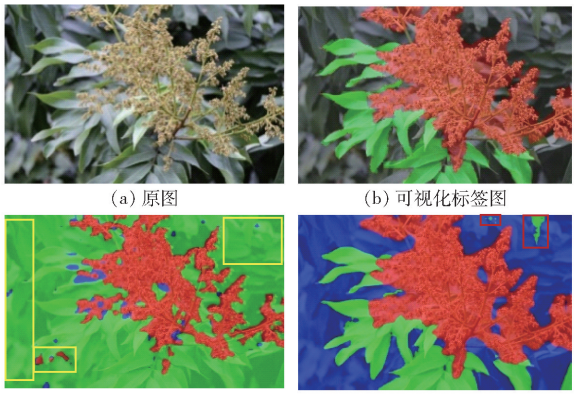
目前花朵识别和分割主要使用传统的机器学习方法，深度卷积网络主要针对易于区分的花朵，使用实例分割方法进行识别。将荔枝花分割、苹果花分割研究作为对比，同时增加一组 FCN 模型实验结果进行对比，验证本文荔枝花识别分割模型的有效性。

从表 3 可看出，MASU R - CNN 预测单朵可分花朵的平均识别准确率最高，为 96.4%，但并不适用于荔枝花、芒果花等对象。而与荔枝花生长形态相近的芒果花识别研究使用传统机器学习方法，平均识别准确率只有 85.8%，比本文模型低 1.2 个百分点。本文模型实现荔枝花语义分割的准确率比 FCN 模型高 6.2 个百分点，效果较好。本文模型与 FCN 模型在测试集的分割效果如图 9 所示。

表 3　不同模型花朵识别效果对比

Tab. 3　Performance comparison of different flower detection models

模型	花的类型	应用类型	平均识别准确率/%
MASU R - CNN[13]	单朵可分	实例分割	96.4
FCN 模型	重叠稠密	语义分割	80.8
芒果花检测模型[9]	重叠稠密	机器学习	85.8
本文模型	重叠稠密	语义分割	87.0



(a) 原图　(b) 可视化标签图　(c) FCN模型分割效果　(d) 本文模型分割效果

图 9　荔枝花分割效果

Fig. 9　Segmentation effect diagrams of litchi flower

对比图 9b、9c 发现，FCN 模型的分割效果较差，模型对背景和叶片的区分度低，某些背景区域会误分割成荔枝花，如黄色方框区域所示。图 9b、9d 分割效果较好，图 9d 的红色方框区域是将背景误分割成叶片，从图 9a 中发现，该区域的背景与叶片十分相似，所以比较难区分。对比图 9c 和图 9d 发现，本文模型的分割效果具有明显优势，能清楚地区分叶片和暗色背景，类别之间的分界线比较清晰。

综上所述，本文提出的方法能够有效识别分割荔枝花，同时也适用于类似荔枝花形态的花朵，能够为智能果园疏花提供视觉支持。

3　结论

(1)为了实现自动化疏花，构建了一个像素级的荔枝花识别分割模型。以 Deeplab V3 为基础，通

过对比实验确定合适深度的特征提取主干网络。加载预训练模型,并进行微调,使之适用于荔枝花数据集,微调后的 ResNet_34 网络能有效地提取野外环境下荔枝花特征,为后续识别和分割奠定基础。

(2)为了提高特征提取的有效性,引入了稠密特征传递方式,本层特征从所有浅层特征中运算得出,浅层特征的高分辨率有助于深层特征的准确定位,提高了定位的准确度。同时,在特征传递过程中加入 CBAM 注意力模块,起到筛选特征的作用,提高了特征的有效性。二者结合提高了模型性能。

参 考 文 献

[1] 熊俊涛, 林睿, 刘振, 等. 夜间自然环境下荔枝采摘机器人识别技术[J/OL]. 农业机械学报, 2017, 48(11): 28 - 34.
XIONG Juntao, LIN Rui, LIU Zhen, et al. Visual technology of picking robot to detect litchi at nighttime under natural environment[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(11): 28 - 34. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=20171104&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2017.11.004. (in Chinese)

[2] 武星,齐泽宇,王龙军,等. 基于轻量化 YOLOv3 卷积神经网络的苹果检测方法[J/OL]. 农业机械学报,2020,51(8):17 - 25.
WU Xing, QI Zeyu, WANG Longjun, et al. Apple detection method based on light-YOLOv3 convolutional neural network[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020,51(8):17 - 25. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=20200802&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2020.08.002. (in Chinese)

[3] 刘芳,刘玉坤,林森,等. 基于改进型 YOLO 的复杂环境下番茄果实快速识别方法[J/OL]. 农业机械学报, 2020, 51(6): 229 - 237.
LIU Fang, LIU Yukun, LIN Sen, et al. Fast recognition method for tomatoes under complex environments based on improved YOLO[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(6): 229 - 237. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=20200624&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2020.06.024. (in Chinese)

[4] XIONG J, LIU Z, LIN R, et al. Green grape detection and picking-point calculation in a night-time natural environment using a charge-coupled device (CCD) vision sensor with artificial illumination[J]. Sensors, 2018, 18(4):969 - 985.

[5] STEIN M, BARGOTI S, UNDERWOOD J, et al. Image based mango fruit detection, localisation and yield estimation using multiple view geometry[J]. Sensors, 2016, 16(11):1915 - 1939.

[6] 闫建伟, 赵源, 张乐伟, 等. 改进 Faster RCNN 自然环境下识别刺梨果实[J]. 农业工程学报, 2019, 35(18): 143 - 150.
YAN Jianwei, ZHAO Yuan, ZHANG Lewei, et al. Recognition of *Rosa roxbunghii* in natural environment based on improved Faster RCNN[J]. Transactions of the CSAE, 2019, 35(18): 143 - 150. (in Chinese)

[7] 刘小刚, 范诚, 李加念, 等. 基于卷积神经网络的草莓识别方法[J/OL]. 农业机械学报, 2020, 51(2):237 - 244.
LIU Xiaogang, FAN Cheng, LI Jianian, et al. Identification method of strawberry based on convolutional neural network[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020,51(2):237 - 244. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=2002026&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2020.02.026. (in Chinese)

[8] ADANSEN F J, COFFELT T A, NELSON J M, et al. Method for using images from a color digital camera to estimate flower number[J]. Crop Science, 2000, 40(3): 704 - 709.

[9] WANG Z, VERMA B, WALSH K B, et al. Automated mango flowering assessment via refinement segmentation[C]// Image and Vision Computing, New Zealand, 2016: 1 - 6.

[10] WANG Z, UNDERWOOD J, WALSH K B, et al. Machine vision assessment of mango orchard flowering[J]. Computers and Electronics in Agriculture, 2018,151(8): 501 - 511.

[11] LIU Y, TANG F, ZHOU D, et al. Flower classification via convolutional neural network[C]// IEEE International Conference on Functional-structural Plant Growth Modeling. IEEE, 2017.

[12] FARJON G, KRIKEB O, HILLEL A B, et al. Detection and counting of flowers on apple trees for better chemical thinning decisions[J]. Precision Agriculture, 2020, 21(3): 1 - 19.

[13] TIAN Y, YANG G, WANG Z, et al. Instance segmentation of apple flowers using the improved Mask R - CNN model[J]. Biosystems Engineering, 2020, 193: 264 - 278.

[14] KRIZHEVSKY A, SUTSKEVER I, HINTON G E, et al. ImageNet classification with deep convolutional neural networks[C]// Neural Information Processing Systems, 2012: 1097 - 1105.

[15] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]// Computer Vision and Pattern Recognition, 2014.

[16] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// Computer Vision and Pattern Recognition, 2015: 1 - 9.

[17] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Computer Vision and Pattern Recognition, 2016: 770 - 778.

[18] CHEN L, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[J]. Computer Vision and Pattern Recognition, arXiv,2017: 1706.05587.

[19] LONG J, SHELHAMER E, DARRELL T, et al. Fully convolutional networks for semantic segmentation[C]// Computer Vision and Pattern Recognition, 2015: 3431 - 3440.

[20] HUANG G, LIU Z, DER MAATEN L V, et al. Densely connected convolutional networks[C]// Computer Vision and Pattern Recognition, 2017: 2261 - 2269.

[21] YANG M, YU K, ZHANG C, et al. DenseASPP for semantic segmentation in street scenes[C]// Computer Vision and Pattern Recognition, 2018: 3684 - 3692.

[22] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018: 7132 - 7141.

[23] WOO S, PARK J, LEE J, et al. CBAM: convolutional block attention module[C]// European Conference on Computer Vision, 2018: 3 - 19.

[24] GIDAIRS S, SINGH P, KOMODAKIS N, et al. Unsupervised representation learning by predicting image rotations[C]// International Conference on Learning Representations, 2018.