

基于改进 SOLO v2 的番茄叶部病害检测方法

刘文波¹ 叶涛¹ 李 颀²

(1. 陕西科技大学电气与控制工程学院, 西安 710021; 2. 陕西科技大学电子信息与人工智能学院, 西安 710021)

摘要: 为实现对多种番茄叶部病害的精确检测,提出了一种基于改进 SOLO v2 的番茄叶部病害实例分割方法。该方法以 SOLO v2 模型为主体框架,将 ResNet-101 作为骨干网络融合特征金字塔网络 (Feature pyramid networks, FPN),引入可变形卷积对卷积结构进行优化,并将损失因子 δ 融入掩膜损失函数中,在语义分支与掩膜分支上对实例进行检测与分割。通过对模型的改进,实现了对形状复杂多变的番茄叶片的精确检测与分割,并提升了模型的泛化能力与鲁棒性。基于 Plant Village 数据集的试验结果表明,ResNet-101 比 ResNet-50 在 SOLO v2 上的性能表现更好。在相同骨干网络下,SOLO v2 模型的单幅图像处理时间比 Mask R-CNN 减少了 72.0%,平均精度均值 (Mean average precision, mAP) 提升了 3.2 个百分点,改进后的模型在训练过程中收敛效果有所提升,受叶片形状多变的影响较小,最终的平均精度均值达到了 42.3%,单幅图像处理时间仅需 0.083 s,在提升检测精度的同时保证了运行的实时性。该研究较好地解决了番茄病叶识别与分割难的问题,为农业自动化生产中番茄疾病情况与症状分析提供了参考。

关键词: 番茄叶部; 病害; 实例分割; 可变形卷积; 损失因子; SOLO v2

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 1000-1298(2021)08-0213-08

OSID: 

Tomato Leaf Disease Detection Method Based on Improved SOLO v2

LIU Wenbo¹ YE Tao¹ LI Qi²

(1. School of Electrical and Control Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China
2. School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an 710021, China)

Abstract: In order to achieve accurate detection of a wide range of tomato leaf diseases, an instance segmentation method was proposed based on improved SOLO v2 for tomato leaf diseases. The SOLO v2 model was adopted as the main framework, using ResNet-101 as the backbone network to fuse feature pyramid networks (FPN), optimize the convolutional structure by introducing deformable convolution, and integrate the loss factor δ into the mask loss function to detect and segment the instances on the category branch and the mask branches. By improving the model, it achieved accurate detection and segmentation of tomato leaves with complex and variable shapes, and the generalisation and robustness of the model were improved. On the basis of the public dataset of Plant Village, the data were cleaned and synthetic multi-instance images were added. The images were manually annotated to create a training set, a validation set and a test set with nine tomato leaf cases and healthy leaves. After setting the parameters and structure of the models, a performance comparison of SOLO v2 models with different depths of residual networks was carried out in the same experimental environment. Finally, model performance comparison tests of different models and the performance comparison tests of SOLO v2 models before and after optimisation were respectively conducted on the basis of the better performing residual networks. The experimental results showed that ResNet-101 performed better than ResNet-50 on SOLO v2. With the same backbone network, the SOLO v2 model reduced the processing time of a single image by 72.0% compared with Mask R-CNN and improved the mean average precision (mAP) metric by 3.2 percentage points. The enhanced model improved convergence in the training process and was less affected by the variable shape of the blade, with a final mAP of 42.3% and a single image processing time of 0.083 s, ensuring real-time operation while improving detection accuracy. The research solved the

problem of identification and segmentation of diseased tomato leaves, and provided a reference for the analysis of tomato disease conditions and symptoms in automated agricultural production.

Key words: tomato leaf; diseases; instance segmentation; deformable convolution; loss factor; SOLO v2

0 引言

作物病害一直是困扰农业生产的难题,是制约农业可持续发展的主要因素之一^[1]。番茄是我国重要的经济作物之一,同时作为世界上主要的蔬菜作物,种植面积和产量都在不断增加。然而番茄病害会导致产量的显著下降,给农业经济带来巨大损失^[2-3]。因此,番茄病害的早期诊断与防治是番茄高产的重要保障。

随着机器学习在图像检测方面的不断发展,有学者将其用于对植物叶片病害的检测识别。魏丽冉等^[4]提出了基于核函数支持向量机的多分类检测方法对植物叶片的4种疾病进行检测识别,识别率在70%~89.5%之间,精度有限。刘君等^[5]采用卷积神经网络(Convolutional neural network, CNN)与传统的HOG(Histogram of oriented gradient)+SVM(Support vector machines)算法相结合的方法对番茄叶片的疾病进行检测,通过CNN提取浅层特征,融合HOG更好地描述特征,最后输入到分类器SVM中识别病害,识别准确率均值(92.49%)优于传统HOG+SVM算法(71.88%),证明了卷积神经网络在叶片病害检测领域的巨大潜力。杨森等^[6]在Faster R-CNN的基础上,通过构造复合特征字典来训练马铃薯叶片病害识别模型,平均识别准确率为84.16%,但在识别及定位精度上有待加强。熊俊涛等^[7]使用了以VGG16为骨干网络的Mask R-CNN的实例分割算法,实现了对大豆叶片实例的像素级识别与分割,分类准确率达到89.42%,平均分割准确率为88.3%,但运行速度较慢,单幅图像识别平均需要0.8 s。

为实现对番茄叶片疾病高精度与快速检测,本文基于Plant Village开源数据库,提出以单阶段(one-stage)的实例分割算法SOLO v2^[8]为主体框架,并将DCN v2^[9](Deformable convolutional networks version 2)与ResNet-101相融合作为骨干网络来提高对形状各异的番茄叶片的识别精度,为番茄疾病的防治提供信息支持。

1 模型结构设计

实例分割旨在精确地检测和描绘图像中的对象^[10-11],通过掩膜对对象进行定位,将对象定位推向像素级的极限。目前,实例分割方法可分为3类:

自上而下(top-down)的方法^[12-19],首先进行目标检测,然后再分割框内的实例对象,该方法由于性能良好而使用广泛;自底而上(bottom-up)^[20-23]的方法,该方法首先将每个实例看作一个类别,然后进行聚类,一般来说其效果劣于第1类方法;直接得到实例分割结果的方法^[8,24]。本文所选取的SOLO v2网络框架即为直接得到实例分割结果的方法,它以位置信息来动态分割对象,由于网络结构相对简明,模型训练、识别速度都较高。同时,在基本不影响模型运行速度的情况下,本研究将提升模型的识别精度以进一步优化分割性能。

1.1 SOLO v2 模型

SOLO v2 主要由5部分构成:全卷积网络(Fully convolutional networks, FCN)、特征金字塔网络^[25]、卷积核分支(Mask kernel branch)、特征分支(Mask feature branch)以及语义分支(Category branch),如图1所示。

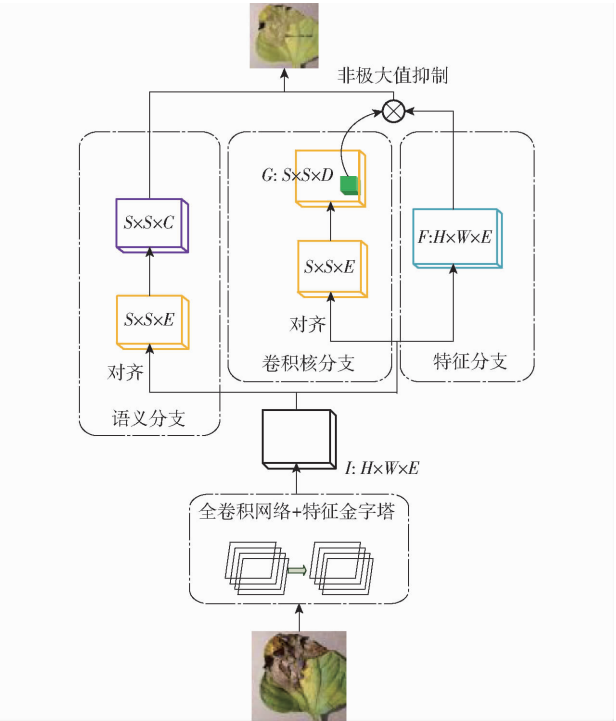


图1 SOLO v2 结构框图

Fig. 1 Framework diagram of SOLO v2

图1中,H、W、E分别为特征金字塔的输出特征图I及掩膜特征图F的高度、宽度及通道数;S表示对齐后特征图的高度或宽度;C为语义总类别数,等效于语义分支输出特征图的通道数;D为卷积核权值,等效于掩膜内核G的通道数。

SOLO v2 与YOLO^[26]类似,对于输入图像,从概

念上将其均分为 $S \times S$ 的网格。通过全卷积网络与特征金字塔网络进行特征提取,判别实例的中心是否落入某个网格中,满足条件的网格将进入语义分支与掩膜分支,分别对对应的实例进行语义以及掩膜大小及位置的判别。

1.1.1 卷积骨干

模型在进行预测实例信息前,必须通过卷积网络进行特征提取。由于残差网络 ResNet^[27] 可以解决优化网络较深带来梯度爆炸及梯度消失的问题,泛化能力较强。因此,本研究使用 ResNet-101 作为特征提取网络,并与层数较少的 ResNet-50 进行对比试验,以比较不同深度的残差网络的性能。

1.1.2 特征金字塔网络

由于番茄叶片在图像中的尺度不一,同时,在训练过程中使用的 Plant Village 数据库中的叶片较大,而实际场景中大多会存在较小尺寸的番茄叶片。因此模型存在实际应用中难以识别小尺寸叶片的问题,从而导致模型精度较低。

目前,特征金字塔网络已广泛与卷积骨干网络融合,以区分不同尺寸的检测对象。在原模型中,如图 2 左侧所示,ResNet 网络根据输出图像的尺寸及维度可分为 5 层,每层的最后 1 个残差块特征作为 1 个特征金字塔网络的输入特征图。考虑到低层的语义信息表达过少,且其输出的特征图进行卷积、上采样的计算量过大,故仅使用上 4 层(C2 ~ C5)的特征图。然后, C5 通过 1×1 卷积核降维得到 M5。下层(M2 ~ M4)通过上一层的上采样与左侧特征图经过 1×1 卷积结果的对应元素相加得到,由此构建自上而下的金字塔结构。金字塔每层分别通过 3×3 卷积运算得到特征图(P2 ~ P5)。最终,语义信息与位置信息各不相同的 4 类特征图(P2 ~ P5)送到头部网络(Head)中进行预测处理。为了获得更准

确的实例信息,对 P5 层使用 CoordConv^[28] 操作得到 P1。

1.1.3 语义分支

由于 SOLO v2 算法不需要边框先对实例进行目标检测,故不需要全连接层,仅采用分类器预测网格内实例的类别。如图 1 所示,1 幅被分为 $S \times S$ 个网格的输入图像,总类别为 C ,那么从特征金字塔送至语义分支前为 $H \times W \times E$ 的输入空间,对齐到 $S \times S \times E$,经过语义分支处理(多个 3×3 卷积)后展开为 $S \times S \times C$ 的输出空间。即最终生成 $S \times S$ 个 C 维输出,对每个网格分别进行语义类别概率的预测。

1.1.4 掩膜分支

在掩膜分支中进行掩膜的尺寸及位置预测,如图 1 所示,取特征金字塔的输出特征图 I 作为输入,经过 2 个分支计算学习后输出掩膜内核 $G \in \mathbf{R}^{S \times S \times D}$ 及掩膜特征图 $F \in \mathbf{R}^{H \times W \times E}$ 。最后通过动态卷积操作完成对掩膜的预测,计算公式为

$$M_{i,j} = G_{i,j} F \quad (1)$$

其中 $M_{i,j} \in \mathbf{R}^{H \times W}$ $G_{i,j} \in \mathbf{R}^{1 \times 1 \times E}$

式中 (i,j) ——原图坐标

$M_{i,j}$ ——位置 (i,j) 单实例最终预测的掩膜

$G_{i,j}$ ——位置 (i,j) 上 1×1 卷积的掩膜内核

1.1.4.1 掩膜内核

特征图 I 进入掩膜卷积核分支时,首先通过对齐操作转换成 $S \times S \times E$ 的空间特征图。然后经过 4 层卷积层及最后一个 $3 \times 3 \times D$ 卷积核的卷积运算后生成最终的掩膜内核 $G \in \mathbf{R}^{S \times S \times D}$,其中进行 1×1 卷积时 D 等于 E , 3×3 卷积时 D 等于 $9E$ 。

1.1.4.2 掩膜特征图

区别于掩膜卷积核分支需要对每层特征金字塔输出图像进行处理后再合并空间得到掩膜内核,掩膜特征分支首先通过 3×3 卷积操作、组归一化后,由激活函数 ReLU 与 2 个双线性插值运算后将 4 层特征图(P1 ~ P4)统一到原图尺寸的 $1/4$ 并合为 1 幅与原图同样尺寸的特征图。最终通过 1×1 卷积计算、组归一化后经激活函数 ReLU 得到最终的掩膜特征图 F 。

1.1.5 非极大值抑制

在经过掩膜分支后会生成最多 $S \times S$ 个掩膜,通过输出张量的第 3 维(通道)对掩膜进行编码,通道值上限为 S^2 ,编码公式为

$$k = iS + j \quad (2)$$

式中 k ——网格 (i,j) 处的掩膜编码

但并非所有的掩膜都表示不同的实例对象,其中存在多个掩膜均表示同一个实例,在网络输出时,应仅保留一个置信度最高的掩膜。非极大值抑制

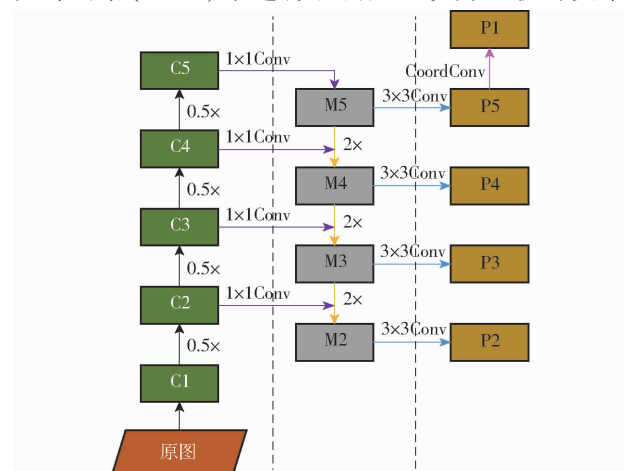


图 2 SOLO v2 中特征金字塔结构图

Fig. 2 Structure diagram of feature pyramid network in SOLO v2

(Non-maximum suppression, NMS) 的思想是抑制非极大值的元素,应用于实例分割即保留某一实例置信度最高的掩膜,去除其他冗余掩膜。

Matrix NMS 考虑使用衰减因子对冗余掩膜的置信度进行降低,因此对于每组相对掩膜 a 置信度较低的 b 的衰减函数 $f(u_{a,b})$ 计算公式为

$$f(u_{a,b}) = 1 - u_{a,b} \tag{3}$$

式中 $u_{a,b}$ ——掩膜 a 与 b 的交并比

由于冗余掩膜的抑制概率与交并比 (Intersection over union, IoU) 正相关, $f(u,a)$ 可表示为

$$f(u,a) = \min_{\forall s_k > s_a} f(u_{k,a}) \tag{4}$$

式中 s_k ——掩膜 k 的置信度

s_a ——掩膜 a 的置信度

$f(u,a)$ ——掩膜 a 关于交并比的正相关函数

$f(u_{k,a})$ ——掩膜 k 与 a 的衰减函数

最后通过惩罚因子单步更新预测得分,从而通过设定阈值保留最佳的预测掩膜,掩膜 b 的惩罚因子计算公式为

$$d_b^* = \min_{\forall s_a > s_b} \frac{f(u_{a,b})}{f(u,a)} \tag{5}$$

$$s_b \leftarrow s_b d_b^* \tag{6}$$

式中 d_b^* ——对掩膜 b 的惩罚因子

s_b ——掩膜 b 的置信度

试验表明^[8], Matrix NMS 对掩膜的预测效果较好,且得益于 GPU 并行计算功能,能够一次完成掩膜 IoU 运算,大幅减少了运行时间。

1.2 改进 SOLO v2

1.2.1 卷积结构优化

在番茄叶片图像中,由于叶片种类繁多,形状大小不一,且某些病症在叶片的表现形式上较为类似,仅采用多变化的训练数据集难以将卷积神经网络 (CNN) 拟合为适应识别多类别叶片病症的理想模型。因为卷积神经网络在构建新模型时的卷积核结构、采样位置、池化比例是不变的,固定的几何结构会导致感受野不适应不同尺度、形状的实例,卷积网络对特征提取能力较差,最终出现网络识别精度较低的情况。DAI 等^[29]提出的第 1 代可变形卷积网络 (DCN) 结构主要解决卷积神经网络的建模问题,引入可变形卷积可提升模型灵活变换的能力。

对于每个输入的特征图,普通卷积在 p_0 位置的像素点处特征值 $y(p_0)$ 计算公式为

$$y(p_0) = \sum_{p_n \in R_f} w(p_n) x(p_0 + p_n) \tag{7}$$

式中 p_n ——以 p_0 为中心卷积核范围内所有位置

$w(p_n)$ —— p_0 位置的采样点权重

$x(p_0 + p_n)$ —— p_0 位置特征值

R_f ——感受野区域

可变形卷积在此基础上增加了一个偏移量 Δp_n 与权重 $\Delta m_n \in [0,1]$, 计算公式为

$$y(p_0) = \sum_{p_n \in R_f} w(p_n) x(p_0 + p_n + \Delta p_n) \Delta m_n \tag{8}$$

经过偏移后的卷积核形状更为灵活,并优化了感受野范围,使卷积骨干可以适应各种类型的番茄叶片,从而提升模型对番茄病叶的检测与分割精度。而引入的权重 Δm_n 则是第 2 代可变形卷积网络 (DCN v2) 所提出的,主要解决加入偏移量后导致感受野对应范围大于目标范围的问题。在 Plant Village 数据集的基础上,本研究将在 ResNet-50 和 ResNet-101 模型中选择性能较优的模型改进卷积结构,并进行对比试验分析。

1.2.2 损失函数优化

在实例分割中,需要用损失函数对语义类别和掩膜输出的期望输出与实际输出的误差进行评价,进而优化模型。为更好地解决图像样本尺寸多、正负样本不平衡的问题,原模型将 Focal^[30] 损失函数和 Dice^[31] 损失函数分别应用于语义类别分类和掩膜预测。对训练损失函数的定义为

$$L = L_{\text{cate}} + \lambda L_{\text{mask}} \tag{9}$$

$$\text{其中 } L_{\text{mask}} = \frac{1}{N_{\text{pos}}} \sum_k l_{|p_{i,j}^* > 0|} d_{\text{mask}}(\mathbf{m}_k, \mathbf{m}_k^*) \tag{10}$$

$$d_{\text{mask}}(\mathbf{p}, \mathbf{q}) = 1 - D(\mathbf{p}, \mathbf{q}) \tag{11}$$

$$D(\mathbf{p}, \mathbf{q}) = \frac{2 \sum_{x,y} (p_{x,y} q_{x,y})}{\sum_{x,y} p_{x,y}^2 + \sum_{x,y} q_{x,y}^2} \tag{12}$$

式中 L ——总损失函数

L_{cate} ——Focal 损失函数

L_{mask} ——掩膜损失函数

λ ——自定义权重参数

N_{pos} ——正样本个数

$p_{i,j}^*$ ——原图位置 (i,j) 的实例真实类别得分

$\mathbf{m}_k$ ——掩膜 k 的预测像素矩阵,等效于 $\mathbf{p}$

$\mathbf{m}_k^*$ ——掩膜 k 的真实像素矩阵,等效于 $\mathbf{q}$

$d_{\text{mask}}(\mathbf{m}_k, \mathbf{m}_k^*)$ —— $\mathbf{m}_k, \mathbf{m}_k^*$ 对应的 Dice 损失函数

$D(\mathbf{p}, \mathbf{q})$ ——矩阵 $\mathbf{p}, \mathbf{q}$ 对应的 Dice 系数

$l_{|p_{i,j}^* > 0|}$ ——关于 $p_{i,j}^*$ 的指示函数 (仅当 $p_{i,j}^* > 0$ 时取 1, 其余情况为 0)

(x,y) ——原图位置 (i,j) 对应的特征图坐标

$p_{x,y}$ ——特征图 (x,y) 处的预测掩膜像素值

$q_{x,y}$ ——特征图 (x,y) 处的真实掩膜像素值

然而,在输入的番茄叶片训练集中,存在像素点

的预测值与真实值都很小的情况 ($p_{x,y}$ 与 $q_{x,y}$ 的值接近 0), Dice 损失函数的梯度则可能会出现非常大的情况, 从而导致训练过程不稳定。因此引入平滑系数 ε , 计算公式为

$$D(p, q) = 2 \frac{\sum_{x,y} (p_{x,y} q_{x,y} + \varepsilon)}{\sum_{x,y} (p_{x,y}^2 + q_{x,y}^2 + \varepsilon)} \quad (13)$$

通过计算发现, 当 $p_{x,y}$ 与 $q_{x,y}$ 的值接近 0 时, 损失函数对 $p_{x,y}$ 的梯度值略大于 0, 消除了可能出现梯度过高导致难以训练的情况。在训练过程中发现平滑系数 ε 取值范围在 0.01 ~ 1 内的模型收敛效果较好, 由于该系数为固定值, 在计算中占比较大, 导致损失函数对不同偏差的预测值的衡量结果较为接近, 会出现误导优化器反向传播时参数更新方向的情况, 进而降低模型的鲁棒性与精度。为了解决此问题, 本研究设计了损失因子 δ 对 Dice 损失函数进行调整, 表达式为

$$D(p, q) = 2 \frac{\sum_{x,y} (p_{x,y} q_{x,y} + \varepsilon \delta)}{\sum_{x,y} (p_{x,y}^2 + q_{x,y}^2 + \varepsilon \delta)} \quad (14)$$

其中
$$\delta = e^{-\left(\sum_{x,y} p_{x,y} - \sum_{x,y} q_{x,y}\right)^2} \quad (15)$$

从理论上可知, 由于损失因子的引入, 使损失函数对预测值与真实值之间偏差的衡量结果与偏差值的联系更为紧密, 并加快了训练收敛的速度, 提升了模型的精度与学习能力。为了探究该损失因子在实际应用中的优化效果, 本研究通过试验进行对比分析及验证。

2 模型评价指标

采用平均精度 (Average precision, AP) 和检测速度作为模型性能评价指标, 从而衡量不同模型对番茄病叶检测与分割的精度、效率。

计算每类番茄叶病的 IoU 阈值对应的 AP, 并求取每类的 mAP。

3 试验设计

3.1 试验平台与参数设置

试验平台选择 GeForce GTX 1660Ti GPU, 显存为 6 GB, 内存为 16 GB, Ubuntu 16.04 运行系统, Python 3.7 + pytorch 1.4.0 下的 mmdetection 1.0 框架作为环境配置。

由于 GPU 性能及数量限制, 故设置 GPU 数量为 1; GPU 每次仅处理 1 幅图像; 迭代轮数为 60; 设置输入模型的图像尺寸为 256 像素 × 256 像素; 优化器采用 SGD, 由 GPU 个数设置初始学习率为 0.00125, 保留原模型中设置的动量因子 0.9 及正

则化权值衰减系数 0.0001; 预热学习率, 选择在第 10 000 步前线性增长学习率, 在第 10 000 步后线性下降学习率, 并在迭代训练至第 27 次与第 33 次时降低学习率, 以加速收敛, 提升收敛后模型的鲁棒性; 设置 Focal 损失函数的权重为 1, Dice 损失函数的权重为 2; 对于各类模型的 IoU 阈值设置为 0.5。

3.2 试验数据

试验数据取自公开数据集 Plant Village 中的番茄叶片集。其中包括早疫病、晚疫病、叶霉菌病、七叶虫斑病、二斑叶螨病、靶点病、黄曲叶病、花叶病、细菌性斑点病以及健康叶共 10 类番茄叶片, 病例如图 3 所示。注意到原数据集有部分图像是错分的, 所以首先需要进行数据清洗保证每类图像在训练前类别正确。并对较少的类别叶片样本进行数据增强, 以解决数据不平衡的问题。同时为了提升模型对单幅图像中多实例的识别效果, 使用 Photoshop 对原样本进行裁剪, 将多个叶片整合到一幅图像中。由于合成图像与原数据集的图像的分辨率不同, 故需要将尺寸归一化到 256 像素 × 256 像素。最后, 从合成图像及经预处理后的 Plant Village 数据集中筛选出 9 000 幅样本, 随机抽取每类的 60% 作为训练集, 20% 作为验证集, 剩余的 20% 作为测试集。



图 3 数据集图像示例
Fig. 3 Data set image example

模型训练的数据集为 coco 标注格式。经清洗后的 Plant Village 番茄数据样本存在分割好的 RGB 图像 (黑色背景), 通过编写程序将其批量转换为黑白二值图像 (白色描述叶片轮廓, 黑色为背景)。对于合成图像则通过 LabelMe 进行实例标注, 得到黑白二值图像后与 Plant Village 数据集的黑白二值样本结合原彩色图像统一使用 pycococreator 工具包转换为 coco 标注文件。

3.3 对比模型训练

模型训练都采用 coco 数据集的预训练模型, 通过迁移学习的方式对各个模型进行训练, 提升训练速度并改善模型的泛化能力。

3.3.1 模型对比试验

为得到在数据集上表现最优的模型, 选用 Mask R-CNN + FPN + ResNet-101、SOLO v2 + FPN + ResNet-50、SOLO v2 + FPN + ResNet-101、SOLO v2 +

FPN + DCN v2 + ResNet - 101 进行训练与验证。其中,对于引入可变形卷积的模型,在卷积骨干 C3 ~ C5 层中所有 3×3 的卷积核替换成可变形卷积核。在训练过程中通过损失函数曲线收敛情况判断模型是否适用于训练样本。最后,将适用的模型在测试集上计算出相应的 AP 和运行速度并进行对比分析。

3.3.2 损失函数对比试验

为了验证本文提出的损失函数优化方法的有效性,试验在 3.3.1 节处理后,选择性能最优的模型为基础,将优化 Dice 损失函数后的模型与原模型进行对比测试。对比 2 个模型训练过程中的损失函数曲线,分析收敛速度与振荡程度。最后在测试集上计算出相应的 AP 并进行对比分析。

4 结果与分析

4.1 不同模型的试验效果

试验将各个训练至收敛的模型应用于测试集中对番茄叶片进行实例检测与分割,并得到最终的性

能评价结果如表 1 所示,其中所有模型都融合了特征金字塔。表 1 中, $AP_{0.5}$ 与 $AP_{0.75}$ 表示 IoU 阈值分别取 0.5 与 0.75 时的 AP 指标; mAP 表示 IoU 阈值以步长 0.05 从 0.5 到 0.95 所对应 AP 的平均值。

在 SOLO v2 模型下, ResNet - 101 在 $AP_{0.5}$ 、 $AP_{0.75}$ 、 mAP 指标上表现较好,分别为 55.9%、44.3%、41.3%,尤其在 IoU 阈值为 0.75 时相对 ResNet - 50 提升了 0.8 个百分点,单幅图像的处理时间基本相同。说明在残差网络中层数较深的网络精度更高,因此优先将 ResNet - 101 作为骨干网络。

选用 ResNet - 101 作为骨干网络,使用 SOLO v2 模型与 Mask R - CNN 模型进行对比发现:相比 Mask R - CNN, SOLO v2 的 $AP_{0.5}$ 、 $AP_{0.75}$ 、 mAP 分别增加了 3.7、2.9、3.2 个百分点。同时, SOLO v2 仅使用 0.059 s 即完成对单幅图像的识别与分割,单幅图像处理时间相比 Mask R - CNN 减少了 72.0%,极大地提升了模型的运行效率。综上所述,无论是从精度还是运行效率上, SOLO v2 模型都具有较好的效果。

表 1 不同网络模型的性能试验结果对比
Tab.1 Comparison of performance test results of different network models

网络模型	骨干网络	$AP_{0.5}/\%$	$AP_{0.75}/\%$	$mAP/\%$	单幅图像处理时间/s
Mask R - CNN	ResNet - 101 - FPN	52.2	41.4	38.1	0.211
	ResNet - 50 - FPN	55.3	43.5	41.1	0.057
SOLO v2	ResNet - 101 - FPN	55.9	44.3	41.3	0.059
	ResNet - 101 - DCN v2 - FPN	57.2	45.2	42.1	0.083

在 SOLO v2 模型及骨干网络 ResNet - 101 基础上,引入可变形卷积后进一步提升了模型的识别与分割精度。改进后的模型在 IoU 阈值为 0.5 时的效果最好,平均精度达到了 57.2%,相比原模型提升了 1.3 个百分点, mAP 提升了 0.8 个百分点。虽然模型运行效率略有下降,但仍具备实时处理图像的能力。所以,本研究使用 DCN v2 对模型进行优化具有一定的可行性。

4.2 不同损失函数试验效果

由分析可知,以 ResNet - 101 - DCN v2 - FPN 为骨干网络的 SOLO v2 模型具有最优的性能,故本研究将该模型作为本次试验的原模型,并在此基础上对损失函数进行相应变化以进行对比试验,试验结果如图 4 所示。

从图 4 中可知,2 个模型训练时的收敛速度较快,证明 2 个模型的泛化性能良好,且在图 4b 中曲线振荡幅度略小于图 4a,说明损失函数的优化进一步提升了模型的泛化能力。而在验证集上,原模型曲线出现了一次较大幅度的振荡(图 4c),而损失函数优化后的模型曲线(图 4d)则较平稳地收敛。这

说明原损失函数的评价误导了优化器反向传播时的参数更新方向,而本研究的优化策略较好地弥补了这一缺陷,从试验中进一步验证了损失因子 δ 的优化效果。

优化后的模型识别分割效果如图 5 所示,无论是对病症表现较为相似的番茄叶片还是对背景颜色干扰大、有噪声的番茄病叶图像,都能够做到正确分类,掩膜的位置信息也较为准确。图 5 中从左到右依次为无背景的番茄叶片、病症表现类似的番茄叶片、背景干扰大且有噪声的番茄叶片。

最终的模型性能评价如表 2 所示,其中 δ -Dice Loss 表示加入损失因子 δ 后的损失函数。

改进后的模型在 $AP_{0.5}$ 、 $AP_{0.75}$ 、 mAP 指标上分别比原模型提高了 0.3、0.1、0.2 个百分点,证明损失函数的优化进一步提升了模型的精度。本文方法通过引入损失因子 δ 来优化模型的掩膜损失函数,提升了整体性。

5 结束语

提出了一种基于 SOLO v2 的番茄病叶检测与

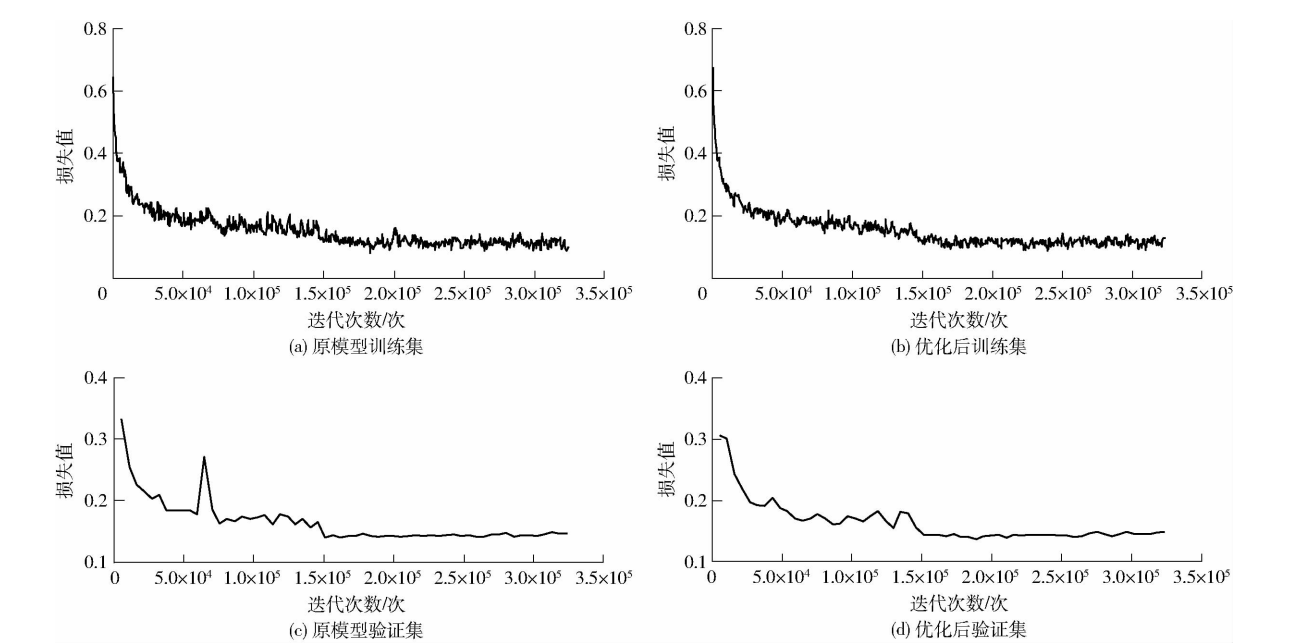


图 4 不同模型在训练集与验证集上的损失值对比

Fig.4 Comparison of loss values of different models on training set and validation set



图 5 番茄叶片检测结果

Fig.5 Results of tomato leaf detection

分割方法，并在卷积结构与损失函数上对模

表 2 加入损失因子后模型性能对比				
Tab.2 Comparison of model performance after adding loss factor				%
		adding loss factor		
骨干网络	掩膜损失函数	AP _{0.5}	AP _{0.75}	mAP
ResNet-101-DCN v2-FPN	Dice Loss	57.2	45.2	42.1
ResNet-101-DCN v2-FPN	δ -Dice Loss	57.5	45.3	42.3

型进行改进。通过对比试验得到,在同样条件下,以 ResNet-101 为骨干网络,SOLO v2 模型比 Mask R-CNN 在 mAP 指标上提高了 3.2 个百分点,单幅图像处理时间减少了 72.0%。而引入 DCN v2 及损失因子 δ 后,最终优化后的模型 mAP 为 42.3%,且有效提高了模型的鲁棒性与泛化能力,能够较精确地实现番茄病叶的识别与分割。

参 考 文 献

[1] 王翔宇,温皓杰,李鑫星,等. 农业主要病害检测与预警技术研究进展分析[J/OL]. 农业机械学报,2016,47(9):266-277. WANG Xiangyu,WEN Haojie,LI Xinxing,et al. Research progress analysis of mainly agricultural diseases detection and early warning technologies[J/OL]. Transactions of the Chinese Society for Agricultural Machinery,2016,47(9):266-277. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20160937&flag=1. DOI:10.6041/j.issn.1000-1298.2016.09.037. (in Chinese)

[2] 赵杨,苗则彦,李颖,等. 番茄灰霉病防治研究进展[J]. 中国植保导刊,2014,34(7):21-29. ZHAO Yang,MIAO Zeyan,LI Ying,et al. Research progress on controlling against tomato gray mold[J]. China Plant Protection,2014,34(7):21-29. (in Chinese)

[3] ALVARO F, SOOK Y, SANG K, et al. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition[J]. Sensors,2017,17(9):2022.

[4] 魏丽冉,岳峻,李振波,等. 基于核函数支持向量机的植物叶部病害多分类检测方法[J/OL]. 农业机械学报,2017,48(增刊):166-171. WEI Liran,YUE Jun,LI Zhenbo,et al. Multi-classification detection method of plant leaf disease based on kernel function SVM[J/OL]. Transactions of the Chinese Society for Agricultural Machinery,2017,48(Supp.):166-171. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=2017s027&flag=1. DOI:10.6041/j.issn.1000-1298.2017.S0.027. (in Chinese)

[5] 刘君,王学伟. 融合 CNN 多卷积特征与 HOG 的番茄叶部病害检测算法[J]. 北方园艺,2020(4):147-152. LIU Jun,WANG Xuewei. A tomato leaf disease detection algorithm based on CNN multi-convolution feature and HOG[J]. Northern Horticulture,2020(4):147-152. (in Chinese)

- [6] 杨森,冯全,张建华,等.基于深度学习与复合字典的马铃薯病害识别方法[J/OL].农业机械学报,2020,51(7):22-29. YANG Sen,FENG Quan,ZHANG Jianhua,et al. Identification method for potato disease based on deep learning and composite dictionary[J/OL]. Transactions of the Chinese Society for Agricultural Machinery,2020,51(7):22-29. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20200703&flag=1. DOI:10.6041/j.issn.1000-1298.2020.07.003. (in Chinese)
- [7] 熊俊涛,戴森鑫,区炯洪,等.基于深度学习的大豆生长期叶片缺素症状检测方法[J/OL].农业机械学报,2020,51(1):195-202. XIONG Juntao,DAI Senxin,OU Jionghong,et al. Leaf deficiency symptoms detection method of soybean based on deep learning[J/OL]. Transactions of the Chinese Society for Agricultural Machinery,2020,51(1):195-202. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20200121&flag=1. DOI:10.6041/j.issn.1000-1298.2020.01.021. (in Chinese)
- [8] WANG X, ZHANG R, KONG T, et al. SOLO v2: dynamic and fast instance segmentation[C]//In Advances in Neural Information Processing Systems,2020:17721-17732.
- [9] ZHU X, HU H, LIN S, et al. Deformable convnets v2: more deformable, better results[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2019:9308-9316.
- [10] THOMIO W, DENIS W. Instance segmentation as image segmentation annotation[C]//Proceedings of the IEEE Intelligent Vehicles Symposium,2019:432-437.
- [11] LIU L, OUYANG W L, WANG X G, et al. Deep learning for generic object detection: a survey[J]. International Journal of Computer Vision,2020,128(2):261-318.
- [12] BOLYA D, ZHOU C, XIAO F, et al. Yolact: real-time instance segmentation[C]//Proceedings of the IEEE International Conference on Computer Vision,2019:9157-9166.
- [13] YI L, HAO Z Q, JI F D, et al. Fully convolutional instance-aware semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:4438-4446.
- [14] HE K M, GEORGIA G, PIOTR D. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,42(2):386-397.
- [15] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:8759-8768.
- [16] HUANG Z, HUANG L, GONG Y, et al. Mask scoring R-CNN[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2019:6409-6418.
- [17] CHEN X, GIRSHICK R, HE K M, et al. TensorMask: a foundation for dense object segmentation[C]//Proceedings of the IEEE International Conference on Computer Vision,2019:2061-2069.
- [18] CHEN H, SUN K, TIAN Z, et al. BlendMask: top-down meets bottom-up for instance segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2020:8570-8578.
- [19] ZHANG R, TIAN Z, SHEN C, et al. Mask encoding for single shot instance segmentation[C]//Proceedings of the IEEE International Conference on Computer Vision,2020:10223-10232.
- [20] ALEJANDRO N, HUANG Z, DENG J. Associative embedding: end-to-end learning for joint detection and grouping[C]//Advances in Neural Information Processing Systems,2017:2277-2287.
- [21] BRABANDERE B D, NEVEN D, GOOL L V. Semantic instance segmentation for autonomous driving[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops,2017:478-480.
- [22] LIU S, JIA J, SANJA F, et al. SGN: sequential grouping networks for instance segmentation[C]//Proceedings of the IEEE International Conference on Computer Vision,2017:3516-3524.
- [23] GAO N, SHAN Y, WANG Y, et al. SSAP: single-shot instance segmentation with affinity pyramid[J]. IEEE Transactions on Circuits and Systems for Video Technology,2021,31(2):661-673.
- [24] WANG X, KONG T, SHEN C, et al. SOLO: segmenting objects by locations[C]//European Conference on Computer Vision. Springer,2020:649-665.
- [25] LIN T, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA,2017:2117-2125.
- [26] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE,2016:779-788.
- [27] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas,NV, USA,2016:770-778.
- [28] ROSANNE L, JOEL L, PIERO M, et al. An intriguing failing of convolutional neural networks and the CoordConv solution[C]//Advances in Neural Information Processing Systems,2018:9628-9639.
- [29] DAI J F, QI H Z, XIONG Y W, et al. Deformable convolutional networks[C]//The IEEE Conference on Computer Vision and Pattern Recognition,2017:764-773.
- [30] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,42(2):318-327.
- [31] MILLETARI F, NAVAB N, AHMADI S A. V-Net: fully convolutional neural networks for volumetric medical image segmentation[C]//2016 Fourth International Conference on 3D Vision (3DV). IEEE,2016:565-571.