DOI:10.19651/j.cnki.emt.2212226

基于改进 YOLOv5-s 的火龙果多任务识别与定位*

孔凡国 李志豪 仇展明 王 鑫
(五邑大学智能制造学部 江门 529000)

摘 要:在复杂的农业环境下,水果采摘机器人系统感知端的识别与定位性能是提高水果采摘成功率的重要指标。本文以复杂外形的火龙果作为研究对象,针对采摘机器人的视觉系统提出了一种适用于火龙果图像自主检测的实时多任务卷积神经网络——SegYOLOv5。该网络基于 YOLOv5-s 卷积神经网络的主体架构进行适应性改进,通过提取 3 层加强特征作为改进级联 RFBNet 语义分割网络层的输入,实现图像检测和语义分割的多任务目标识别检测,有效提升了模型的整体性能。改进的 SegYOLOv5 网络结构能够适应对边界敏感的图像语义分割农业场景,测试集的平均精度均值和平均交并比分别为 93.10%和 83.64%,与 YOLOv5-s+原始 RFBNet 和 YOLOv5-s+BaseNet 模型相比,高出了前者 1.23%和 2.74%,高于后者 2.38%和 1.45%。SegYOLOv5 平均检测速度达到 71.94 fps,相比 EfficientDet-D0 提高 40.79 fps,平均精度均值高出 5.8%。通过端到端输出 SegYOLOv5 检测结果并结合图像几何矩算子,能够实时准确定位火龙果质心作为理想采摘点。改进的算法具有较高的鲁棒性和通用性,为基于视觉感知的水果采摘机器人奠定了有效的实践基础。

关键词:YOLOv5;图像处理;火龙果;采摘机器人;采摘点定位
中图分类号:TP391.4 **文献标识码**:A **国家标准学科分类代码**:520.2040

Multitasking recognition and positioning of pitaya based on improved YOLOv5-s

Kong Fanguo Li Zhihao Chou Zhanming Wang Xin
(Intelligent Manufacturing Department, Wuyi University, Jiangmen 529000, China)

Abstract: The recognition and positioning capabilities of the visual perception terminal of the fruit-picking robot system are crucial indicators to increase fruit-picking success rates in the complicated agricultural environment. A real-time multi-task convolutional neural network SegYOLOv5 suited for autonomous Pitaya fruit image detection for the visual system of the picking robot was proposed in this paper using Pitaya fruit with complicated shape as the research object. The network is enhanced based on the primary architecture of YOLOv5′s convolutional neural network. The multitasking target recognition and detection task of image detection and semantic segmentation is realized, and the overall performance of the model is substantially improved, by extracting three-layer enhanced features as the input of the improved cascaded RFBNet semantic segmentation network layer. With a mean Average Precision and mean Intersection Over Union of 93.10% and 83.64%, respectively, for the testing dataset, the enhanced SegYOLOv5 network architecture can adapt to the boundary-sensitive image semantic segmentation agricultural scene, compared with YOLOv5-s+original RFBNet and YOLOv5-s+BaseNet models, it is 1.23% and 2.74% higher than the former, and 2.38% and 1.45% higher than the latter. The average detection speed of SegYOLOv5 can reach 71.94 fps which is 40.79 fps faster than EfficientDet-D0, and the mean Average Precision is 5.8% higher. The center of mass of Pitaya fruit may be precisely positioned in real time as the best picking position using the end-to-end output of SegYOLOv5 detection output and the fusion of image geometric moment operator. The improved algorithm has high robustness and versatility, which lays an effective practical foundation for fruit picking robot based on visual perception.

Keywords: YOLOv5;image processing;pitaya fruit;picking robot;picking point positioning

0 引 言

鉴于水果采摘机器人作业环境的复杂性,现有多数研究并没有综合建立机器人感知端与执行端的联系,出现了采摘机器人系统协调性差且执行效率偏低等问题[1]。实时性是视觉感知型机器人执行视觉闭环控制(“look and

收稿日期:2022-11-28
* 基金项目:广东省普通高校重点领域专项(新一代信息技术)(2021ZDZX1045)资助

move"模式)的重要前提。然而,复杂的算法尽管能够准确定位目标果实,但是高延迟性限制了执行端的规划能力,反之,点到点的开环控制("look then move"模式)很难克服农业环境的外源性干扰(障碍物、阵风等)因素[2];而满足实时性需求的轻量化算法在定位准确度和全局特征的处理能力上往往力不从心,进而限制了感知端的性能[3]。

基于上述矛盾,合理的方案是在机器人感知端设计能满足实时性且具有多尺度特征融合能力的算法。传统视觉检测方法只能提取图像的底层特征,而基于深度学习框架的卷积神经网络能将原始图像直接输入到网络中,通过自主提取图像特征信息,包括图像的低级、中级和高级语义的深层特征[4],实现端到端的图像分类与回归任务。

张勤等[5]通过 YOLOv4 模型确定目标感兴趣区域(region of interest,RoI),融合 RGB-D 图像的深度信息,利用深度分割算法、形态学操作、K-means 聚类算法和细化操作,进而定位番茄串的果梗采摘点,识别成功率为 93.83%,但是单帧图像的平均耗时为 54 ms,大于 30 ms 的实时需求且仍受到光照条件的变化及聚类算法固有的随机性影响,适合于大棚环境下固定品种果实的检测任务[6]。宁政通等[7]结合改进的 Mask R-CNN 网络模型和传统图像处理算法进行特征提取并通过计算葡萄果梗的质心点从而确定采摘点。经试验证明,该方法在不同光照条件下能准确定位果梗采摘点,但是单帧图像平均处理时间是 4.90 s,机器人视觉控制中难以满足闭环反馈需求,作业条件偏理想化。Qi 等[8]通过 YOLOv5 得到荔枝主茎检测框并提取其 RoI 区域输入到 PSPNet 模型进行语义分割,得到主茎分割轮廓并通过二值化和开运算等图像处理方法,确定了荔枝主茎采摘点的位置。该工作得到了 92.5%的查准率,但算法的冗余性和对平台算力庞大的需求可能会限制系统的整体性能。

目前,我国火龙果的年产量已超过 100 万吨,年产值高居世界第 2[9]。实现火龙果机械自动化采收,或将节省高昂的农务采收成本。但是,目前国内外针对火龙果识别与定位的方法等关键技术研究还寥寥无几。商枫楠等[10]提出一种将 YOLOX-Nano 与注意力机制(concentration based attition modiule,CBAM)结合的火龙果检测方法。改进后的模型 $AP_{0.5}$ 指标为 98.9%,检测时间为 21.72 ms。由于单一检测任务实现的果实定位方案不太适用于这类形态各异的火龙果,算法的的鲁棒性恐难以适应果园的复杂环境。邓子青等[11]提出将 Otsu 阈值分割算法融合传统形态学操作,实现了火龙果的图像分割与定位,该方法仅仅区分了成熟个体和果园背景,且算法性能受到自然环境因素影响较大,定位点的鲁棒性低。

因此,本文拟将火龙果这类外观差异较大的热带水果作为识别与定位的研究对象。以 YOLO 系列[12-16]这类广受欢迎的目标检测卷积神经网络为基础,构建一种实时的多任务(图像检测和语义分割)目标检测器——SegYOLOv5。该方法有效克服了自然农业环境下,背景复杂、光照变化等因素对检测算法的影响,解决了现有识别检测方法在高实时性和高精度两方面的矛盾,为实现更通用、有效的水果识别与定位方法提供了有效参考。

1　火龙果多任务识别网络模型

1.1　SegYOLOv5 网络模型构建

SegYOLOv5 检测器改进于 YOLOv5-s 网络模型,基于 Pytorch 深度学习框架。整体网络结构由主干(Backbone)、颈部(Neck)和头部(Head)3 部分组成,如图 1 所示,图中的 Conv2DBNSiLU 表示连续的卷积层、BN 层和 SiLU 激活层;K6 表示卷积核大小为 6×6;S2 表示纵横步长为 2;P2 表示边缘填充数为 2;R2 表示膨胀系数为 2。该模型在原有 YOLOv5-s 的网络架构基础上新增了分割头网络层(SegMask Head),同时对原始结构进行火龙果特征的适应性改进。具体设计原则满足采摘机器人感知端的轻量化及实时性需求,且具备多任务、多尺度目标的识别检测能力。下文将对改进模型的各个主体结构部分进行介绍和理论分析。

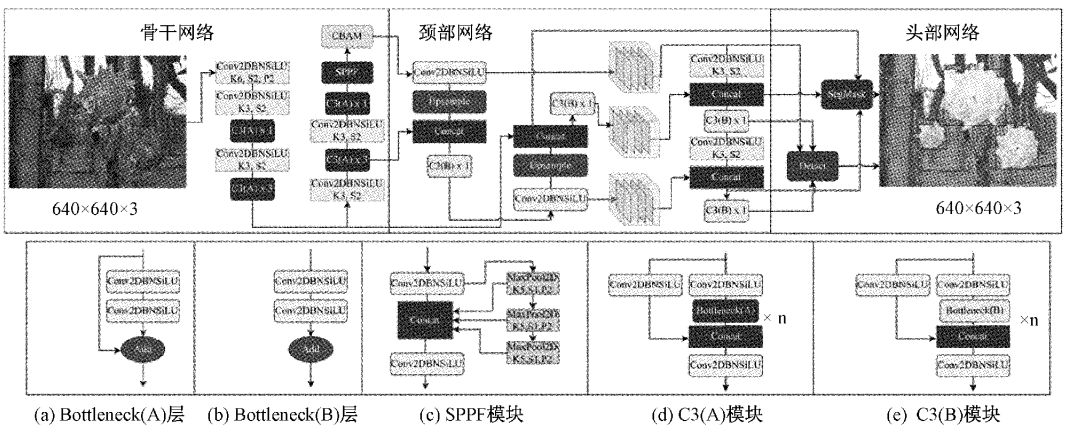

图 1　SegYOLOv5 多任务网络结构

1) Backbone 网络结构

该网络结构继承了 YOLO 系列的多层主干残差结构,有效避免深层网络的梯度消失和梯度爆炸现象,以提高网络训练的表现。当分辨率为 640×640×3 的彩色图(R-G-B 通道)输入改进后的主干网络,分别提取中间层(L2)、中高层(L3)和最高层(L4)3 个初始有效特征层,其维度依次为 80×80×128、40×40×256、20×20×512。对 L4 特征层引入了 CBAM[17] 以改善火龙果个体差异或遮挡现象导致的长范围特征失效问题,抑制无关信息引入的噪声。

2) Neck 网路结构

该网络保留了 YOLOv4 的路径聚合 PaNet 网络结构[18],用于加强特征提取,区别在于将 5 次卷积块替换为包含残差网络的 3 次卷积块,即图 1 中的 C3(B)层。将输入的 3 个有效特征层经过 Down-up-Down 的特征融合、提取操作,加强后的特征层具有如下特点:高层的语义信息保留了更多的浅层细节信息(边缘、形状等),有效提高小尺度目标的检测能力,同时缓解了梯度回传时多次上、下采样造成的特征离散,为后续生成的预测框分类、回归和掩码(Mask)提供更强的语义信息。

3) Head 网络结构

该结构由 SegMask Head 和检测头(Detect Head)两部分组成。检测头网络结构与 YOLOv5-s 相同。扩展的语义分割头网络将 3 层加强特征(16、19、22 层)经过核大小为 1×1 卷积层(没有特殊说明,默认步长为 1)统一隐藏通道数,并上采样为同一尺度大小后融合为 80×80×384 大小的特征层,随后输入改进的级联 RFBNet[19] 分割层网络,以进一步扩大感受野并提高网络非线性,经过 C3SPPF 的残差模块过渡了输出层的瓶颈结构,再通过 1×1 卷积操作调整输出通道数,最后借助 8 倍上采样方法还原成原图大小后,得到语义分割掩码图。

4) 改进 RFBNet 网络模块

像素级图像分类要求对全局和局部特征具有较大的感受野,因此,SegMask Head 的核心结构借鉴了 RFBNet 语义分割网络并加以改进,该网络的基本思想是基于扩张卷积(dilated convolution)的并行分支进行融合,以模拟真实人眼视觉系统的感受野,从而加强网络的特征提取能力。改进后的网络结构如图 2 所示,具体包括如下改进:

(1)原始 RFBNet 网络所使用的非对称卷积块尽管能有效减少参数量,但在低特征尺度(如 3×3)下,其表现并不理想。基于火龙果采摘任务需求的多尺度特性,将原始分支以独立 1×1 卷积整合输出通道。分支 1(Branch1)首层增加 3×3 的卷积层并引入膨胀系数为 2 的扩张卷积,核大小为 3,卷积后进行批标准化和 SiLU 激活函数,进一步提高网络非线性拟合能力;Branch2 舍弃了原始 3×3 卷积层;Branch3 设定隐藏层通道数为 128。

(2)用膨胀系数组 R (1、2、3、5)代替膨胀系数组 R(1、3、5)的扩张卷积层形成的多分支结构以扩大输出层对输

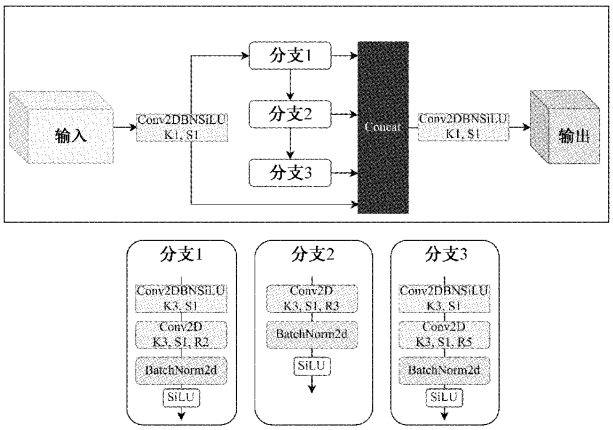

图 2　改进级联 RFBNet 模块

入特征的有效感受野,平衡了局部特征到全局特征的提取能力,避免边缘信息的丢失及网格化效应导致的特征不连续现象。

(3)将 RFBNet 各个改进的分支级联并统一减少隐藏层的输出通道数,实现多尺度特征融合,且不损失深层语义信息,有效地减少了深层网络冗余重复的梯度信息。

综合上述设计原则,网络非线性增强且参数量减少。整体改进后的 SegYOLOv5 网络能鲁棒的分割图像中火龙果个体及茎干目标区域。模型中各个隐藏单元能够融合上下文信息并有效扩大输出层单元的感受野以提供多尺度目标识别的能力。改进后的模型与原始 YOLOv5-s 模型相比参数量仅增加了 7.73%,依旧符合轻量化设计要求,模型权重大小为 15.6 MB,可供后续嵌入式/移动端的部署。

1.2　火龙果数据集制作

1) 数据采集

制作用于网络训练的火龙果数据集,考虑到火龙果采摘机器人末端搭载感知相机的距离对精度的影响[20],设定果实图像采集的间距为 0.5～1.5 m。以 9:00～10:00 和 16:00～18:00 的 2 个时间段的随机光照条件下,在广东佛山、江门等地的火龙果园区进行采集。果实表皮颜色为青色或深绿色的设定为未成熟果实;而红色及紫红色的认为是成熟果实。共采集 600 张用于选择性收获的图像检测数据样本和 450 张用于语义分割的样本。

2) 数据标注及数据增广

在 PC 端上进行 MS COCO 标准格式的数据集制作。借助标注软件“labelimg”制作检测数据集,标注图像中成熟与未成熟果实的真实框(ground truth,GT);利用软件“labelme”标注图像中成熟果实和茎干的轮廓以制作语义分割数据集。

由于人工采集的数据相对单一,在复杂多样的环境下不能兼顾非理想化的对象,包括受遮挡目标和小目标等对象样本分布不均衡问题。为了提高模型的泛化能力和鲁棒性,并有效减少人工标注的耗费,对数据集进行数据增

广。实验中将针对相应的样本标签以一定的概率引用几种常用的数据增广策略，如表 1 所示。其中，各个样本标签的含义为：Pitaya（成熟火龙果）、Unripe_Pitaya（未成熟火龙果）、Stem（火龙果茎干）。增强后的数据集中包含 5 050 张检测数据和 2 760 张语义分割数据。该火龙样本及标签随机打乱后按照 7∶2∶1 的比例分别作为模型的训练、验证和测试数据集。

表 1　数据集标签及数据增强策略

数据集类别	样本标签	数据增强策略	引用概率/%
图像检测数据集	Pitaya	Mosaic	30
		Random affine	40
	Unripe_Pitaya	Mix-up	10
		Random-horizontal-flip	40
语义分割数据集	Pitaya	Copy-paste	100
	Stem	Random rotation	40
		Letterbox	30

1.3　SegYOLOv5 网络训练损失函数

1）图像检测损失函数

图像检测网络的总损失为预测框损失（$Loss_{Rect}$）、置信度损失（$Loss_{Obj}$）、分类损失（$Loss_{Clc}$）三者的加权和。SegYOLOv5 网络的损失函数定义为：

$$Loss_{Det} = a \cdot Loss_{Rect} + b \cdot Loss_{Obj} + c \cdot Loss_{Clc} \tag{1}$$

其中，a，b，c 分别为对应损失的权重系数，通常置信度损失取最大权重，矩形框损失和分类损失的权重次之。

SegYOLOv5 预测框损失函数也即定位损失函数基于交并比（intersection over union，IoU）改进的 CIoU loss。IoU 是目标检测中的一个重要概念，即目标框与真实框交集与并集的比值，其表达式为：

$$IoU(B1,B2) = \frac{|B1 \cap B2|}{|B1 \cup B2|} \tag{2}$$

CIoU 通过更多的维度来考虑预测框与真实框的差异，进一步提升训练的稳定性和收敛速度，得到预测框损失函数表达式为：

$$Loss_{Rect} = 1 - IoU(B,B_{gt}) + \frac{\rho^2(B,B_{gt})}{c^2} + \alpha v \tag{3}$$

$$v = \frac{4}{\pi}\left(\alpha\arctan\frac{\omega^{gt}}{h^{gt}} - \alpha\arctan\frac{\omega}{h}\right)^2 \tag{4}$$

$$\alpha = \frac{v}{1 - IoU(B,B_{gt}) + v} \tag{5}$$

其中，v 为预测框和真实框长宽比例差值的归一化，α 为权衡长宽比例造成的损失和 $IoU(B,B_{gt})$ 部分造成的损失的平衡因子。

SegYOLOv5 分类损失函数和置信度损失函数均采用 BCE Loss 损失函数（binary cross entropy，ECE），如式（6）所示。

$$Loss_{Clc\&Obj} = -\sum_{n=1}^{N} y^* \log(y_i) + (1 - y_i^*)\log(1 - y_i) \tag{6}$$

$$y_i = Sigmoid(x_i) = \frac{1}{1 + e^{-x_i}} \tag{7}$$

当作为类别损失函数时，N 表示类别总数，x_i 表示当前类别的预测值，y_i 为经过激活函数后的当前类别概率值，y_i^* 则为当前类别的真实值（0 或 1）；作为置信度损失函数时，x_i 表示当前置信度的预测分数，y_i 为经过激活函数后的当前置信度概率，y_i^* 则为当前置信度的真实标签值。不同的是，使用 CIoU 作为该预测框的置信度标签参数，其范围取值为 0～1，因此标签值的大小与预测框、目标框的重合度有关，两框的重合度越高则标签值越大。

2）语义分割损失函数

考虑到实际农业果实采样场景中，针对部分果实其成熟与非成熟果实数量差异较大以及难易样本数据集标注困难等问题导致的正负样本数据类别不平衡问题，近年来已陆续提出多种策略以提高训练效果和检测精度。本文采用基于在线困难样本挖掘（online hard example mining，OHEM）算法[21]的 OhemCELoss 作为语义分割的损失函数，即从预测框分类任务的交叉熵扩展到基于像素（pixel）任务计算分类交叉熵，进而根据损失选取难样本（hard example）的策略。

OHEM 算法的核心是选择一些 Hard Example 作为训练的样本从而改善网络参数效果，Hard Example 体现在具有多样性和高损失值。所以 OhemCELoss 损失函数的难样本根据每个像素的交叉熵损失（Cross-entropy loss）来选择，Cross-entropy loss 函数的表达式如下：

$$Loss_{ce}(x,cls) = -\log\left(\frac{\exp(x[cls])}{\sum_j \exp(x[j])}\right) \tag{8}$$

式中：x 为预测值，cls 为像素真实类别。损失的计算原则如下：引入 2 个超参数：N、阈值 T，规定每个批次中至少有 N 个 pixel 参与训练，前 N 个 pixel 经过交叉熵函数得到损失值并经过展平为从大到小排序的向量，随后与预先设定的 T 值比较，若第 N 个损失仍大于阈值 T，则取所有大于该阈值的元素计算交叉熵损失和；否则取前 N_i 个参数计算损失。最后，输出这批次（Batch）样本的损失的均值作为语义分割损失量。

2　实验与分析

2.1　实验平台配置

用于实验的软硬件算力平台的主要配置参数如表 2 所示，编程开发环境为 VScode，利用图像处理单元（graphics processing unit，GPU）进行并行推理运算。

2.2　模型评价指标

实验利用像素精度（pixAcc）和平均交并比（mean intersection over union，mIoU）[22]作为语义分割子任务的

表 2 实验环境的配置参数

软硬件项目	参数/规格
中央处理器型号	Intel ® Core™ i7-10700 CPU@2.90 GHz
GPU 型号	NVIDIA GeForce RTX 2070(8 GB)
操作系统	Ubuntu 20.04
CUDA 版本	11.4.120
cuDNN 版本	8.2.2
高级编程语言	Python 3.9

性能指标。采用平均精度均值(mean average precision,$mAP_{0.5}$)[23]指标评价图像检测子任务的性能,其中,0.5 表示 IoU 阈值为 0.5。检测帧率(frame per second,FPS)描述网络模型的图像推理速度,参数量(parameters)和浮点运算次数(floating point of operations,FLOPs)用于衡量模型的大小及算法复杂度。

2.3 模型训练流程分析

每个迭代周期(epoch)内,每个批次随机载入 m 个批量大小的检测样本和语义分割样本,进行数据预处理调整为输入图片大小,利用官方 YOLOv5-s 预训练权重初始化本研究改进模型的 Backbone 网络参数,每个样本输入网络进行计算并存储各层网络的中间变量,输出层会利用骨干网络(backbone)特征层以加强 Neck 网络,并分别提取 3 个隐藏层的特征张量以输入 Detect Head 和 SegMask Head 进行计算并预测结果。利用输出层的损失函数得到真实标签与预测值的偏差 $\vec{e}$,然后通过微积分链式求导反向传播最终得到最靠近输入层的模型参数梯度。为满足并行计算要求,经过梯度累计后,更新学习率并利用批量随机梯度下降算法朝着最小化损失函数的负梯度方向进行更新,以加速收敛训练误差,如此交替进行反向和前向传播,直到指定的迭代周期结束或训练损失收敛到理想状态。

2.4 SegYOLOv5 模型训练

本研究将通过制作的火龙果数据集,对 YOLOv5-s+级联 RFBNet(即 SegYOLOv5)网络模型进行训练。实验中的主要参数包括:输入彩色图片大小为 480×480×3,批次大小(batch size)为 9,迭代周期设为 190 次。重要的超参数如下:初始学习率为 0.001,随机梯度下降动量为 0.9,权重衰减为 0.000 5,定位、分类及置信度损失的权重系数分别设为 0.05、0.5 和 1.0。

为了客观评估本文提出的 SegYOLOv5 网络在图像检测和语义分割等方面的综合性能,拟在同一实验环境和数据集中增设 3 组对比模型进行训练,分别是:EfficientDet-D0[24]、YOLOv5-s+原始 RFBNet 和 YOLOv5-s+BaseNet。其中,EfficientDet-D0 图像检测网络模型用以对比 mAP 和 FPS 等性能指标;YOLOv5-s+原始 RFBNet 模型用于比对本文改进的 RFBNet 模块性能;BaseNet 语义分割层由 C3 基础网络搭建。各个模型综合训练时长为 3.8 h。

2.5 模型泛化性能实验

为了真实验证上述训练后的各个模型对未知样本的泛化能力,将测试集样本以一个批次大小输入网络,进行泛化实验测试。输入的图片大小为 480×480×3,各个实验指标如表 3 所示。由于 EfficientDet-D0 模型的标准输入为 512×512×3,因此少数指标不作为统计范畴。

表 3 模型泛化测试集实验对比

网络模型	膨胀系数	隐藏层通道数	参数量	FLOPs/10^9	mAP/%	pixAcc/%	mIOU/%	帧率/(fps)
EfficientDet-D0	—	—	—	—	87.30	—	—	31.15
YOLOv5-s+原始 RFBNet	1,3,5	128	8 124 852	28.90	91.87	91.46	80.90	52.91
YOLOv5-s+BaseNet	—	256	7 609 311	23.37	90.72	92.66	82.19	61.35
SegYOLOv5	1,2,3,5	128	7 652 931	22.84	93.10	93.34	83.64	71.94

实验结果表明,YOLOv5-s+BaseNet 网络结构使网络更深,隐藏层通道数为 256,mIoU 为 82.19%,而减少了一半隐藏层通道数的 YOLOv5-s+原始 RFB 网络在不加深网络层数的条件下,mIoU 为 80.90%,相比前者低了 1.29%,但 mAP 同比高出 1.15%。

与其他 3 类检测器相比,本文提出的 SegYOLOv5 网络则进一步减少了算力要求,参数量略高于 YOLOv5-s+BaseNet,但网络复杂度最小,推理速度最快,平均达到 71.94 fps,相比 EfficientDet-D0 高出了 40.79 fps,mAP 值高出 5.8%;与 YOLOv5-s+原始 RFB 网络和 YOLOv5-s+BaseNet 的 mIoU 值和 mAP 值作比较,相比前者分别提高 2.74% 和 1.23%,与后者相比提高了 1.45%和 2.38%。不过由于自制火龙果数据集样本难以避免的平衡性差异,本文的方法在 Pitaya 类别的精度均值高出了 Unripe_Pitaya 类别的 2.8%,但仍满足火龙果自主采摘的识别性能要求。

2.6 模型试验结果对比

为客观评价本研究的 SegYOLOv5 网络模型的识别性能,与具有代表性的 EfficientDet-D0 检测网络和 YOLOv5-s+BaseNet 网络进行火龙果树全视场检测试验。试验场景根据火龙果园不同实景分为远景和近景,结果如图 3 所示。

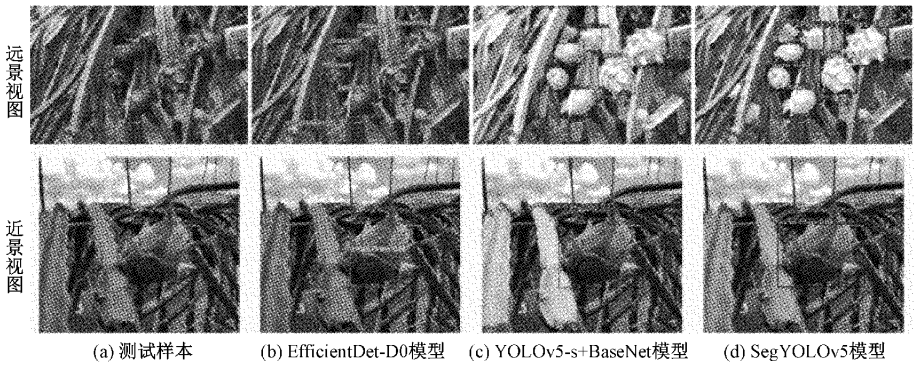

图 3　不同网络模型测试结果对比

试验结果表明，融合了检测与分割 2 种任务的 SegYOLOv5 模型继承了 YOLOv5 和 RFBNet 的特点，以 Neck 网络加强的多尺度特征，综合了底层的细节信息和高层语义信息，分割边界更精细，而 BaseNet 分割层通过重复的卷积操作加深网络层数提高了输出层的感受野，但是低分辨率的特征图丢失了较多的细节，全局目标检测效果优于 EfficientDet-D0，但略低于 SegYOLOv5 模型。整体而言，由于火龙果训练样本的限制，SegYOLOv5 模型语义分割性能还有待提高，对于低分辨率图像的分割效果一般。

3　火龙果采摘点定位

对于不存在遮挡以及规则的类球形目标，诸如苹果、柑橘等，检测器输出果实的预测框中心点作为果实理想的采摘点，然而对于诸如火龙果、麒麟果以及成簇的果实如葡萄、荔枝等不规则外观的对象，上述方法容易误判。因此，该类果实需要定位果实有效质心位置，以提高果实采摘成功率。

3.1　火龙果质心求解分析

图像的矩通常描述了图像形状的全局特征，几何矩作为发展最成熟且最简单的矩描述子，优点是对外部噪声不敏感且运算量小，利用零阶矩、一阶矩可求出轮廓的质心，求火龙果的质心坐标 $P(x_i,y_i)$，其表达式为：

$$\begin{cases} x_i = \dfrac{\sum_{y=1}^{N}\sum_{x=1}^{M} x f(x,y)}{\sum_{y=1}^{N}\sum_{x=1}^{M} f(x,y)} \\ y_i = \dfrac{\sum_{y=1}^{N}\sum_{x=1}^{M} y f(x,y)}{\sum_{y=1}^{N}\sum_{x=1}^{M} f(x,y)} \end{cases}, i = 1,2,\cdots,n \tag{9}$$

式中：x，y 表示像素坐标点 $f(x,y)$ 是火龙果轮廓的像素点处的像素值，N、M 均取值为 480，分别表示图像的像素高度和像素宽度。

3.2　火龙果采摘点定位试验

将上述传统算法与本文识别检测方法结合以确定火龙果图像的二维质心位置。试验将单帧图像输入权重初始化后的 SegYOLOv5 网络模型，推理运算得到检测及分割结果，并单独提取分割掩码图进行后处理，基于 OpenCV 分别进行二值化、形态学腐蚀处理，解算各个火龙果的轮廓并取其最大包围面积的果实，作为首个采摘目标并计算其质心像素坐标，即为理想目标采摘点，最终输出融合结果。

本试验对晴天及阴天光照 2 组光照条件下的图像进行验证，其试验结果如图 4 所示。结果表明，通过融合几何矩图像处理算法，可以动态获取理想的火龙果二维采摘点位置。单帧图像全过程平均耗时小于 0.02 s，仍满足实时性需求。

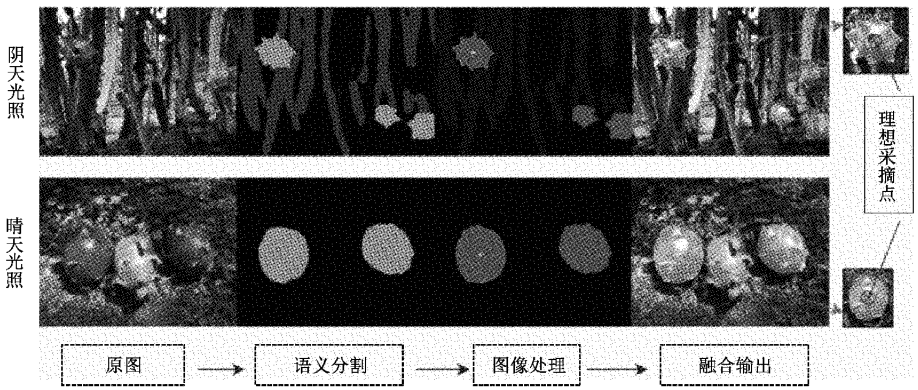

图 4　火龙果采摘点定位试验流程

4 结 论

本文提出了基于 YOLOv5-s 改进的火龙果多任务识别网络模型 SegYOLOv5,具备单输入、多输出的分支结构,能够同步完成火龙果的全视场目标检测和语义分割。分割任务能够辨别目标丰富的细节信息(果实轮廓、茎叶遮挡、果园背景);而检测任务能够区分火龙果类别(成熟或未成熟),预测框可作为先决条件与分割任务交互预测果实间的遮挡或重叠情况。通过端到端的输出结果融合图像几何矩算子,仍可以实时、动态的获取火龙果二维质心采摘点。

实验和测试结果表明,改进后的 SegYOLOv5 网络模型 mAP 值可达到 93.10%,目标像素的 mIOU 达到了 83.64%。与 EfficientDet-D0 和 YOLOv5-s+BaseNet 多任务网络相比,图像检测和语义分割精度更高,单帧火龙果图像的平均推理速度可远超实时。因此本文提出的方法可以为基于视觉控制技术的采摘机器人闭环反馈任务提供实时动态的原始输入,为选择性采摘技术和视觉感知技术奠定了实践基础。本文的方法具备指导水果采摘机器人在复杂农业环境中安全、自主作业的潜力。

参考文献

[1] 金玉成,高杨,刘继展,等. 采摘机器人深度视觉伺服手-眼协调规划研究[J]. 农业机械学报,2021,52(6):18-25,42.

[2] 初广丽,张伟,王延杰,等. 基于机器视觉的水果采摘机器人目标识别方法[J]. 中国农机化学报,2018,39(2):83-88.

[3] 陈燕,王佳盛,曾泽钦,等. 大视场下荔枝采摘机器人的视觉预定位方法[J]. 农业工程学报,2019,35(23):48-54.

[4] CHIAGOZIEM C U, QIN Z G, MD B B H, et al. Recent advancements in fruit detection and classification using deep learning techniques [J]. Mathematical Problems in Engineering, 2022, 2022: 1-29.

[5] 张勤,陈建敏,李彬,等. 基于 RGB-D 信息融合和目标检测的番茄串采摘点识别定位方法[J]. 农业工程学报,2021,37(18):143-152.

[6] 郑太雄,江明哲,冯明驰. 基于视觉的采摘机器人目标识别与定位方法研究综述[J]. 仪器仪表学报,2021,42(9):28-51.

[7] 宁政通,罗陆锋,廖嘉欣,等. 基于深度学习的葡萄果梗识别与最优采摘定位[J]. 农业工程学报,2021,37(9):222-229.

[8] QI X K, DONG J S, LAN Y B, et al. Method for identifying litchi picking position based on YOLOv5 and PSPNet[J]. Remote Sensing, 2022, 14(9): 2004.

[9] 徐磊磊,金琰,侯媛媛,等. 我国火龙果市场与产业调查分析报告[J]. 农产品市场,2021,(8):43-45.

[10] 商枫楠,周学成,梁英凯,等. 基于改进 YOLOX 的自然环境中火龙果检测方法[J]. 智慧农业(中英文),2022,4(3):120-131.

[11] 邓子青,王阳,张兵,等. 基于 Otsu 算法与形态学的火龙果图像分割研究[J]. 智能计算机与应用,2022,12(6):106-109,115.

[12] JOSEPH R, SANTOSH D, ROSS G, et al. You only look once: Unified, real-time object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016: 779-788.

[13] JOSEPH R, ALI F. YOLO9000: Better, faster, stronger [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), IEEE, 2017: 6517-6525.

[14] 段禄成,谭保华,余星雨. 基于改进 YOLOv3 的酒瓶盖瑕疵检测算法[J]. 电子测量技术,2022,45(15):130-137.

[15] 谢国波,郑晓锋,林志毅,等. 基于改进 YOLOv4 算法的高压塔鸟巢检测[J]. 电子测量技术,2022,45(18):145-152.

[16] 颜学坤,楚建安. 基于 YOLOv5 改进算法的印花图案疵点检测[J]. 电子测量技术,2022,45(4):59-65,DOI:10.19651/j.cnki.emt.2108370.

[17] WOO S H Y, PARK J C, LEE J Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018: 3-19.

[18] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.

[19] LIU S T, HUANG D, WANG Y H. Receptive field block net for accurate and fast object detection[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018: 385-400.

[20] FU L S, GAO F F, WU J Z, et al. Application of consumer RGB-D cameras for fruit detection and localization in field: A critical review[J]. Computers and Electronics in Agriculture, 2020, 177: 105687.

[21] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training region-based object detectors with online hard example mining [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 761-769.

[22] GARCIA-GARCIA A, ORTS-ESCOLANO S, OPREA S, et al. A review on deep learning techniques applied to semantic segmentation [J].

ArXiv Preprint,2017, ArXiv:1704.06857.

[23] EVERINGHAM M, VAN G L, WILLIAMS C K I, et al. The pascal visual object classes (voc) challenge [J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[24] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.

作者简介

孔凡国,博士,教授,主要研究方向为人工智能与图像处理、机电控制等。

E-mail: 407901008@qq.com

李志豪,硕士研究生,主要研究方向为智能化检测与自动控制技术。

E-mail:1715519378@qq.com

仇展明,硕士研究生,主要研究方向为智能化检测与自动控制技术。

E-mail: jammingchiu@163.com

王鑫,硕士研究生,主要研究方向为智能化检测与自动控制技术。

E-mail: 915097352@qq.com