

基于多源图像融合的自然环境下番茄果实识别

王文杰 贡亮 汪韬 杨智宇 张伟 刘成良
(上海交通大学机械与动力工程学院, 上海 200240)

摘要: 蔬果采摘机器人面对的自然场景复杂多变,为准确识别和分割目标果实,实现高成功率采收,提出基于多源图像融合的识别方法。首先,针对在不同自然场景下单图像通道信息不充分问题,提出融合 RGB 图像、深度图像和红外图像的多源信息融合方法,实现了机器人能够适应自然环境中不同光线条件的番茄果实。其次,针对传统机器学习训练样本标注低效问题,提出聚类方法对样本进行辅助快速标注,完成模型训练;最终,建立扩展 Mask R-CNN 深度学习算法模型,进行采摘机器人在线果实识别。实验结果表明,扩展 Mask R-CNN 算法模型在测试集中的检测准确率为 98.3%、交并比为 0.916,可以满足番茄果实识别要求;在不同光线条件下,与 Otsu 阈值分割算法相比,扩展 Mask R-CNN 算法模型能够区分粘连果实,分割结果清晰完整,具有更强的抗干扰能力。

关键词: 番茄果实; 机器视觉; 目标识别; 深度学习; Mask R-CNN 算法

中图分类号: TP391.41 文献标识码: A 文章编号: 1000-1298(2021)09-0156-09 OSID: 

Tomato Fruit Recognition Based on Multi-source Fusion Image Segmentation Algorithm in Open Environment

WANG Wenjie GONG Liang WANG Tao YANG Zhiyu ZHANG Wei LIU Chengliang
(School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: The natural scenes faced by fruit and vegetable picking robots are complex and changeable. Accurate identification and segmentation of the target fruit are crucial for high success rate harvesting. The instance segmentation is an effective method to solve the problem. However, existing instance segmentation algorithms have some drawbacks, such as the limited effect of edge segmentation accuracy for single-source images, the workload and time spent on image labeling. Therefore, a tomato fruit recognition algorithm based on multi-source fusion image and extended Mask R-CNN model was proposed. Firstly, aiming at the problem of insufficient information in different natural scenes with a single image channel, a multi-source information fusion method combining RGB images, depth images and infrared images was proposed, which enabled the robot to adapt to different lighting and fruits at different maturity stages. Secondly, aiming at the problem of inefficiency of traditional machine learning training sample standards, a clustering method was proposed to assist the rapid labeling of samples to complete the model training. Thirdly, an extended Mask R-CNN deep learning algorithm model was established for online fruit recognition by picking robots. The experimental results showed that the extended Mask R-CNN algorithm model achieved 98.3% detection accuracy and 0.916 detection IoU in the test set, which can well meet the requirements of tomato fruit recognition; under different lighting conditions, compared with the Otsu threshold segmentation algorithm, the extended Mask R-CNN algorithm model was able to distinguish the adherent fruits with clear and complete segmentation results and stronger anti-interference ability.

Key words: tomato fruit; machine vision; target recognition; deep learning; Mask R-CNN algorithm

0 引言

我国是番茄种植面积最大,生产总量最多的国家^[1]。目前,国内番茄采摘主要依靠人工进行。一方面,受人口老龄化影响,劳动力资源紧缺导致劳动力成本不断升高^[2];另一方面,人工采摘效率较低,而且质量无法保证。因此,研究番茄果实识别算法,提高采摘设备的自动化、智能化水平对缓解人力资源紧缺,提高生产效率具有重要意义。

随着农业自动化、智能化技术的不断发展,机器视觉技术已经成为农业采摘机器人的核心技术之一^[3-5]。图像分割是机器视觉技术的前提和关键^[6]。由于番茄果实表皮柔软,而末端执行器一般为金属结构,且动作简单,前期的图像分割精度较低会导致果实采摘损伤率高。因此,提高图像分割的准确率和精度是目前国内外研究的热点问题。传统图像分割算法一般使用 RGB 图像,以颜色阈值作为分割依据^[7-9],对图像中的几何形状和空间位置等深层特征的提取较为困难,算法的适应性差。特定的阈值参数只针对特定的场合,当自然环境变化,如光照变化时,设定的阈值参数失效进而导致识别失败。

近年来,随着计算机硬件性能不断提升,深度神经网络(DNN)方法凭借其强大的特征提取能力和自主学习能力被广泛应用于农作物目标检测^[10-11]。与传统算法相比,神经网络模型能够提取出图像的隐性特征(无明显实际含义的特征),并通过自我学习和迭代得出划分的依据^[12-16]。增加图像的信息量能够提高分割的准确率和精度,使用多源图像是一种有效方法。通过加噪声、改变图像亮度、对比度等方法对样本集进行数据增强,基于深度学习的果实识别算法在场景变化的自然环境具有更好的检测效果,但存在一定的局限性,例如需要大量训练样本,样本标注工作量大、训练时间长等问题。

自然环境下,环境复杂多变,干扰因素多,信息量大,特征提取困难;番茄果实柔软易损,图像分割精度要求高;单源 RGB 图像对重叠、粘连果实识别效果不佳;传统样本集标注方法重复性高,工作量大。为此本文在 Mask R-CNN 的基础上,提出一种基于多源图像融合的图像分割算法,借助聚类方法完成样本集的快速标注,以提高标注效率和检测精度,实现对番茄果实图像精准分割。

1 实例分割算法模型

1.1 图像配准及多源图像融合

番茄在温室种植环境下,光线变化、重叠粘连等复杂情况很难用单纯的 RGB 图像经过深度学习模

型进行准确预测,而加入深度信息与红外信息将有利于这些样本的预测,提高分割精度^[17]。如图 1 所示,深度图像不受光照影响,能通过深度信息区分不同番茄个体,红外图像中心亮周围暗的特点有利于提取边缘特征。

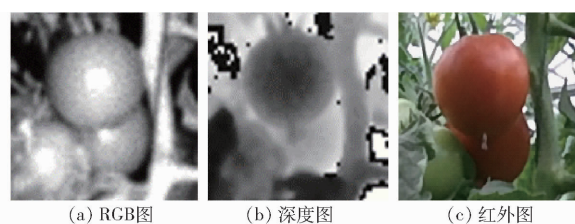


图 1 重叠粘连番茄果实示意图

Fig. 1 Diagrams of overlapping adherent tomato fruits

本研究使用不同传感器采集到的多源图像,其分辨率存在差异。普通彩色图像的分辨率为 1 920 像素 × 1 080 像素,而深度图和红外图的分辨率为 512 像素 × 424 像素,需要使同一样本下的图像具有相同的分辨率。由于不同传感器视野范围不同,简单的图像缩放不能实现图像配准,通过转换矩阵可以实现像素点的精准匹配。采集到的深度图和红外图的图像分辨率均为 512 像素 × 424 像素,下面以红外图向 RGB 彩色图匹配为例说明图像配准的过程。一般地,红外图像像素坐标 (X, Y) 和彩色 RGB 图像的像素坐标 (x, y) 关系式为

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \mathbf{T} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (1)$$

式中 $\mathbf{T}$ ——坐标变换矩阵

$a_1 \sim a_8$ ——待求解参数

可以利用棋盘标注图像获得多组(远大于 4 组)对应的 (x, y) 和 (X, Y) 坐标,然后通过最佳平方逼近法求得超定方程组的最小二乘解,得到矩阵 $\mathbf{T}$,进而得到分辨率为 1 920 像素 × 1 080 像素的红外图像。

特别地,Microsoft 官方提供了可以获得红外图和彩色图对应映射关系的内置函数 MapDepthFrameToColorSpace,生成坐标映射关系表,然后通过坐标关系映射表得到配准后的图像。图 2 描绘了图像配准的过程,深度图和红外图通过矩阵 $\mathbf{T}$ 或者坐标关系映射表生成与彩色图匹配的红外图和深度图。

多源图像融合将针对同一目标采集到的多源信道图像综合成高质量图像,能够最大限度地挖掘各个信道的有用信息,提高图像信息的利用率,便于后续计算机模型的读取和处理^[18]。图像可以视为一个矩阵,配准后的彩色 RGB 图像(1 920 像素 ×

1 080 像素×3)、深度图像(1 920 像素×1 080 像素×1)、红外图像(1 920 像素×1 080 像素×1)具有相同的分辨率(1 920 像素×1 080 像素),可以融合成一幅图像(1 920 像素×1 080 像素×5)。如图 3 所示,融合后图像具有 r (红色)、 g (绿色)、 b (蓝色)、 d (深度)、 i (红外)5 个通道的信息, r 、 g 、 b 、 d 、 i 分别为各通道对应数值。图像融合公式为

$$F = C \cup D \cup I = (x, y, r, g, b, d, i) \tag{2}$$

其中 $C = (x, y, r, g, b)$
 $D = (x, y, d)$
 $I = (x, y, i)$

式中 C ——彩色图像 D ——深度图像
 I ——红外图像 F ——融合后图像

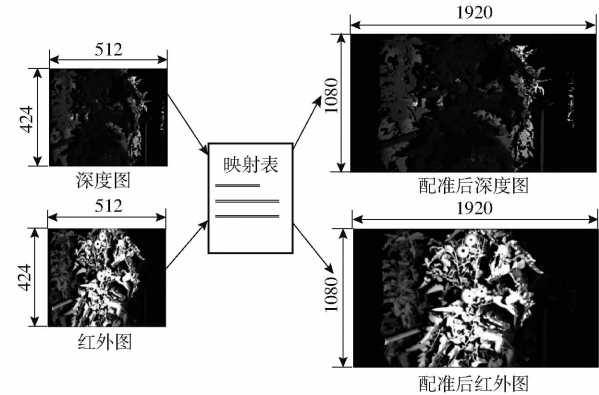


图 2 图像配准过程

Fig. 2 Image registration process

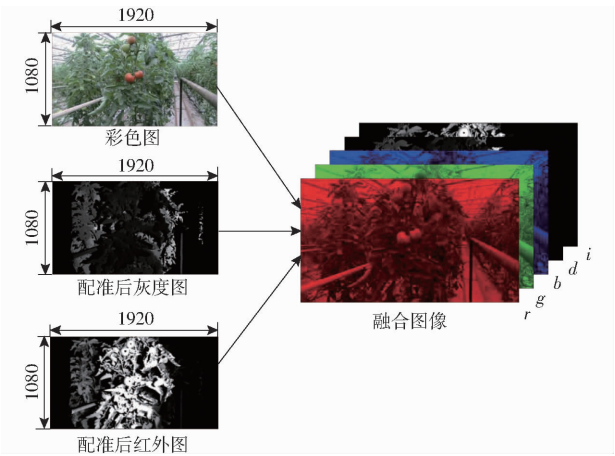


图 3 图像融合过程

Fig. 3 Image fusion process

1.2 基于聚类算法的数据集快速标注方法

完整的训练数据不仅要包括清晰有效的图像信息,还需要有充分有效的标注信息。已知番茄果实图像区域准确对应的语义标签,能够帮助神经网络进行有效地自我学习和迭代。对于大量的开放环境中番茄果实图像样本数据集,其标注难点在于,语义分割任务需要对图像中的每一个具体像素点分配语义标签。一方面,数据集标注工作任务量大,需要投

入大量的时间和精力;另一方面,人工长时间标注出错率高,标注过程会受标注者的主观因素影响。因此,一种能够提高标注效率、缩短标注时间的辅助标注工具或算法是必要的。

目前,有许多公开的标注工具如 VIA (VGG Image Annotator,图 4a)、Labelme(图 4b)等能够辅助研究人员进行标注工作。然而,这些标注工具只能对标注对象使用矩形框框选或者折线包络等粗略的标记,无法满足对图像中每个像素分配语义标签的像素级标注需求。针对这一问题,本文提出一种基于聚类算法的半自动快速标注方法,利用融合后图像信息广泛且综合的特点,将像素点的颜色、深度和红外信息作为聚类特征,沿用 K-means 算法的思想对场景进行初步分割,在此基础上完成样本集标注工作。

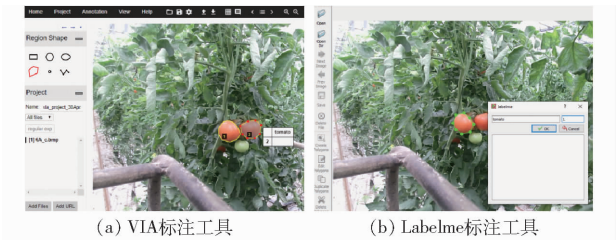


图 4 传统数据集标注工具

Fig. 4 Dataset annotation tools

K-means 是一种常用的聚类算法^[19-20],根据数据样本内部特征的相似性,将一个数据集划分为固定数量的不相交子数据集^[21],算法一般过程为:计算所有样本到所有聚类中心的距离并分配给最近的聚类中心,循环迭代直到全局样本到各自聚类中心的距离之和最小,无法继续优化时算法结束。

基于 K-means 聚类的样本集快速标注算法流程如图 5 所示,所使用的特征包括:像素点横坐标 x 、像素点纵坐标 y 、像素点红色分量 $F_r(x, y)$ 、像素点绿色分量 $F_g(x, y)$ 、像素点蓝色分量 $F_b(x, y)$ 、像素点深度值 $F_d(x, y)$ 、像素点红外强度 $F_i(x, y)$ 。输入为经过图像配准和融合后的样本图像 F ,迭代次数设置为 2 000。聚类中心数量 k 和分类特征权重 w_j 为人工超参数需要多次调试后选择合适值输入。分类特征权重 w_j 是相对值,表示相对于其它特征的重要程度,默认为 1。标注过程如图 6 所示:①根据 K-means 聚类算法,图像根据其特征不同被分割为许多子区域。②根据分割图像的效果通过改变各个特征权重 w_j 进行细微调整,接着选择成熟番茄果实所在色块。③通过图像二值化处理,生成最终标注图像。根据国家标准 GB 8852—88,本文选择红熟中后期,即着红面 4/10 以上的番茄果实作为研究对象,标注过程中红熟中后期番茄的选择由人工完成。

该算法完成了像素级的精确标记,极大地减轻

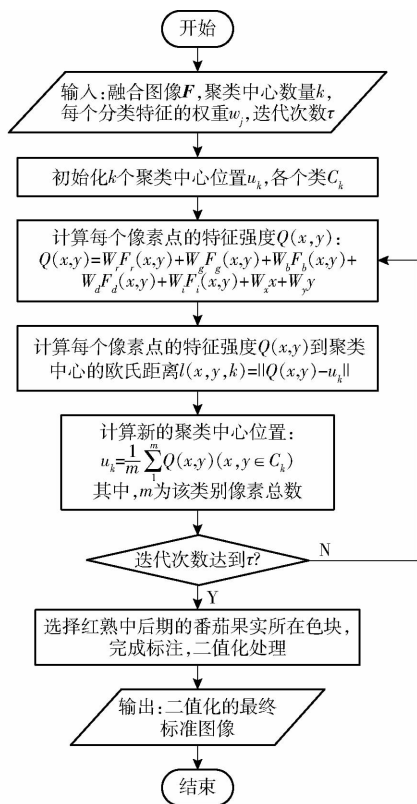


图 5 基于 K-means 聚类的样本集快速标注算法流程图
Fig. 5 Flowchart of fast labeling algorithm for sample sets based on K-means clustering

了研究人员的工作负担,一幅图像的标注时间由原来的 60 s 降低到 20 s 左右。对于个别样本图像,使用 K-means 聚类算法初步分割效果不能满足标注精度要求,需要手动进行调整。

1.3 基于扩展 Mask R - CNN 的图像实例分割算法

Mask R - CNN 综合了研究成果,沿用了 Faster R - CNN 的思想,采用 ResNet - FPN 架构进行特征

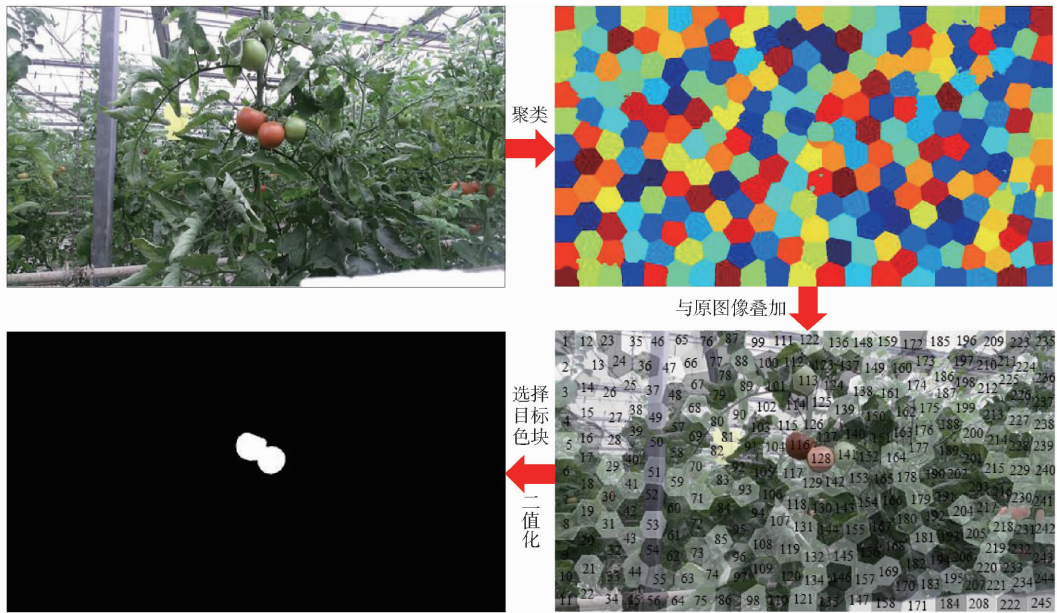


图 6 基于 K-means 聚类算法数据集标注过程
Fig. 6 Data labeling process based on K-means clustering algorithm

提取,与 Faster R - CNN 相比添加了 Mask 预测分支,不仅能够实现物体分类、定位,还能进行实例分割。与语义分割相比,实例分割对同类物体更精确。Mask R - CNN 的优点在于:具有较高的精度,较快的运行速度,较强的泛化性能,被广泛应用于各种场景。但是在农业采摘场景中,Mask R - CNN 模型存在一定的不足:农业场景中番茄成熟果实相互遮挡,由于卷积网络实例分割依赖于纹理特征和颜色差异,而相互覆盖成熟番茄果实表面平滑,颜色均匀,果实的边界难以被准确分割。为适应多源融合图像的特点,发挥多源图像的优势,本文在 Mask R - CNN 的框架上进行扩展,其模型框架示意图如图 7 所示,网络输入由原来的 RGB 图像更改为多源融合图像,网络输出为目标物体的类别、最小包围框和目标物体像素级掩膜结果。由于输入图像的维数由传统的 3 维更改为 5 维融合图像,因此本研究对 ResNet^[22-23] 浅层网络结构进行修改,并将修改过的网络部分参数进行高斯随机初始化。因为输入后为全卷积神经网络,所以只需修改第 1 层参数。加载权重参数排除 conv1,调整 heads 包含的网络层将 conv1 添加至训练层。迁移学习是指预先使用大量通用数据对模型参数进行粗调,然后使用本地数据进行微调^[24-25]。利用迁移学习的优势,先使用 ImageNet 数据集对特征提取网络 ResNet50 预训练,然后使用本地制作的番茄果实数据集训练。

数据增强能够提高模型的精度和泛化性能,针对自然环境下的任务特点,通过改变图像亮度、对比度、添加高斯噪声等方式(图 8)对原始数据集进行数据增强操作,强化模型对光线变化的适应能力。

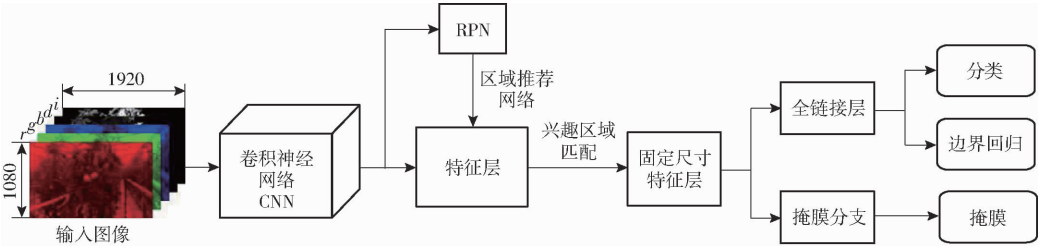


图 7 扩展 Mask R - CNN 框架示意图

Fig. 7 Diagram of extended Mask R - CNN framework

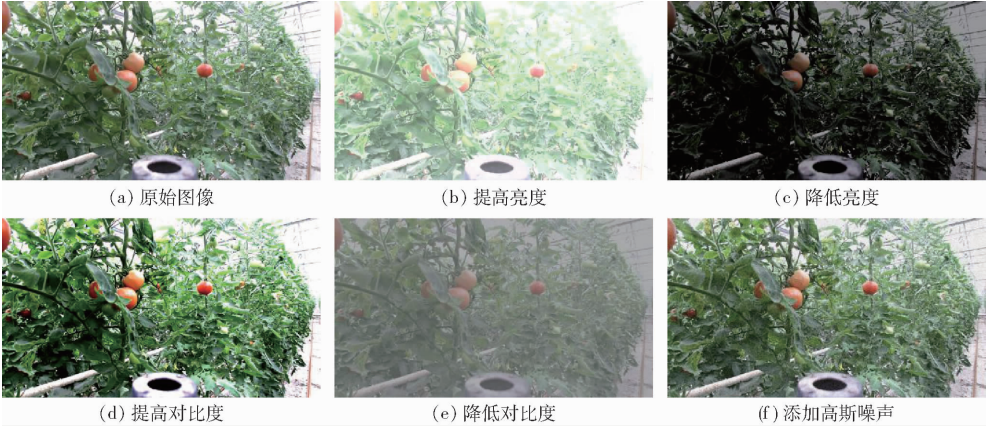


图 8 数据增强示意图

Fig. 8 Diagrams of data augmentation

2 实验验证和比较

2.1 样本采集及模型训练

Microsoft 的 Kinect 相机是一种被广泛应用的多模态视觉传感器,可以同时采集 RGB 图、红外图和深度图^[26]。近年来大规模数据集(如 ImageNet、COCO 等)为计算机视觉研究工作提供了便捷,研究人员在缺乏实验条件和样本数据集的条件下,能够借助现有数据集资源实现一些基础、普适的计算机视觉任务^[27-28]。但对于具体的数据驱动的计算机视觉识别模型,还需要提供大规模标注良好的样本数据集进行针对性训练。

本文样本采集于上海市金山区某现代温室示范基地番茄设施温室(图 9),在计算机和 Kinect V2 相机平台上,编写控制程序采集开放农业环境下的番茄果实图像。共采集了 2 000 份(约 30 GB)样本作



图 9 现代温室示范基地番茄设施温室样本采集场景

Fig. 9 Sample collection scene in modern tomato greenhouse

为训练集原始材料,其中每份样本包含一幅 RGB 彩色图像(1 920 像素 × 1 080 像素)、一幅红外图像(512 像素 × 424 像素)、一幅深度图像(512 像素 × 424 像素)和一张坐标映射关系表。

将采集到的图像进行预处理,并通过配准、融合等图像处理方法生成融合图像。通过数据增强操作将 2 000 份样本扩充到 12 000 份,随机选取 9 600 幅作为训练集,1 200 幅作为验证集,1 200 幅作为测试集用于模型训练。模型训练时采用 0.001 的初始学习率训练头部网络层,然后以 0.000 1 的学习率训练整个网络层,根据模型在验证集的效果调整模型超参数。

服务器配置为: Intel Xeon (R) CPU E5 - 2620 v4 @ 2.10 GHz × 32, 16 GB 的 NVIDIA TESLA P100 GPU, 运行内存 128 GB。基于上述训练策略在服务器上经过 8 h 迭代,训练误差逐渐下降并趋于稳定,最终得到网络参数文件,模型训练完成。

2.2 基于多源图像融合的扩展 Mask R - CNN 算法实验验证

为了评估模型的性能,验证模型对成熟番茄果实的检测效果,使用测试集中没有参与模型训练的番茄果实图像进行模型预测,其实现过程为:首先,将最终得到的模型参数文件加载到神经网络模型中,然后输入待检测图像,经过约 0.45 s 得到模型预测结果。预测结果包括 3 部分:目标物体的类别、目

标物体的边界框以及目标物体的掩膜。

由于原 Mask R - CNN 对于果实检测已经可以取得很好的效果,因此与原 Mask R - CNN 算法模型相比,基于多源图像融合的扩展 Mask R - CNN 算法模型性能及检测效果的变化难以直接在检测结果图像中体现出来,因此需要引入量化指标对模型进行评价。

假设一幅图像中判断为成熟番茄的成熟番茄数目为 T_p ,判断为环境背景的成熟番茄数目为 F_p (漏检数目),将未成熟番茄或枝叶判断为成熟番茄的数目为 T_N (误检数目)。

检测准确率 (Accuracy) 可以表示为

$$A = \frac{T_p}{T_p + T_N + F_p} \times 100\%$$

(3)

实例分割的精度可以用交并比 p_{IoU} (Intersection over union) 表示^[29],计算式为

$$p_{\text{IoU}} = \frac{f_{\text{area}}(P \cap T)}{f_{\text{area}}(P \cup T)}$$

(4)

式中 P ——模型预测的成熟番茄果实区域

T ——实际的番茄果实所在区域

f_{area} ——计算某区域像素数的函数

消融实验是研究神经网络系统的一种有效方法^[30]。分别使用 RGB 通道数据、RGB + D 通道数据、RGB + I 通道数据、RGB + D + I 通道数据进行消融实验训练,并在相同的 200 份测试集上进行测试,得到模型的检测准确率和 IoU 结果如表 1 所示。与基于 RGB 彩色图像传统 Mask R - CNN 检测模型 (检测准确率为 95.4%、交并比为 0.851) 相比,基于多源融合图像的检测模型性能都有一定提升,RGBD 模型检测准确率为 97.4%、交并比为 0.897,RGBI 模型检测准确率为 96.0%、交并比为 0.891,而基于 RGBDI 融合图像的 Mask R - CNN 模型的检测准确率为 98.3%、交并比为 0.916。

表 1 消融实验模型性能评价结果

Tab.1 Performance evaluation results of ablation experimental model

模型	性能指标	RGB	RGBD	RGBI	RGBDI
Mask R - CNN	检测准确率/%	95.4	97.4	96.0	98.3
	交并比	0.851	0.897	0.891	0.916
YOLO v3	检测准确率/%	88.2	89.6	88.9	90.3
	交并比	0.848	0.863	0.861	0.872

2.3 算法效果对比

文献[31]提出的 YOLO v3 算法是目前应用最广泛的目标检测算法之一,使用基于残差神经网络的 Darknet - 53 作为特征提取器,采用多尺度融合的

方法进行类别和位置预测。

分别使用基于多源图像融合的扩展 Mask R - CNN 检测模型、基于 YOLO v3 算法的检测模型和传统 Otsu 阈值分割算法对测试集中的番茄图像进行测试,图 10 为 3 种算法不同光线条件的预测结果对比,图 11 为 Otsu 算法模型与扩展 Mask R - CNN 算法模型预测结果局部细节对比。Otsu 算法受环境因素干扰较大,模型预测结果噪声明显,检测准确率较低。例如,当亮度提高时,番茄果实高亮部分会检测失败。当亮度降低时,番茄的底部阴影部分由于太暗而无法识别 (图 10c),在实际采摘作业过程中,枝叶摆动、光线变化都会导致识别效果无法满足需求。此外,对于粘连果实,即当两个番茄距离很近时,Otsu 算法无法进行单独区分。YOLO v3 算法能够区分独立的番茄果实个体,受光线变化影响小,但算法输出结果为目标边界框,只能表示大致区域而无法实现对目标的精确定位。基于多源图像融合的扩展 Mask R - CNN 图像分割算法综合了深度图、红外图和彩色图的信息,通过对图像的颜色、形状、纹理、空间位置关系等特征进行深度挖掘,其检测效果明显好于 Otsu 算法和 YOLO v3 算法。其优点在于:①检测准确率更高,图像分割效果更好。Otsu 算法以主要颜色信息作为阈值,识别结果中存在许多白色噪点,而扩展 Mask R - CNN 识别结果有完整的番茄轮廓。②抗干扰能力更强,光线变化对检测效果几乎没有影响。③能够进一步区分不同番茄果实个体。

2.4 扩展 Mask R - CNN 算法的不足

在实际识别过程中,扩展 Mask R - CNN 模型也存在许多不足之处:Mask R - CNN 使用滑动窗格的方法寻找窗格中的最佳目标,因此部分番茄果实识别结果的边缘存在小部分缺失 (图 12a,检测边缘是直线而不是弧线);当番茄枝叶遮挡住番茄果实的 1/5 以下时,有时会被误识别为番茄的一部分 (图 12b);当枝叶遮挡区域超过 2/3 时,可能出现漏检现象 (图 12c)。番茄果实的枝叶遮挡问题工况复杂,涉及要素多,可作为未来研究和探索的重点。

3 结论

- (1)提出了一种基于多源图像融合的扩展 Mask R - CNN 图像实例分割模型,其特点在于使用了彩色、深度、红外融合图像作为神经网络的输入,有效地扩充了图像的信息密度。
- (2)提出了使用 K-means 聚类算法对数据集进行快速标注的方法,提高了图像标注效率,单幅图像

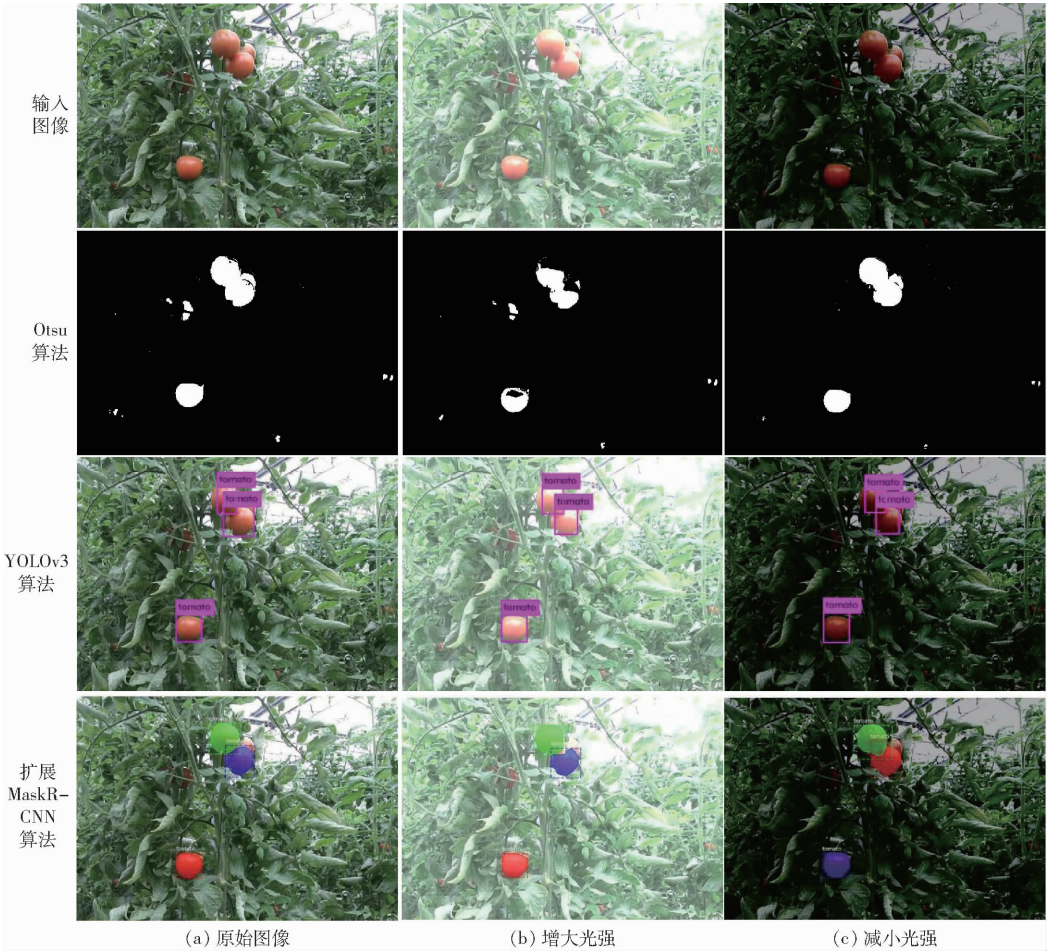


图 10 Otsu 算法、YOLO 算法和本文算法番茄识别结果

Fig. 10 Tomato recognition results of Otsu algorithm, YOLO algorithm and proposed algorithm

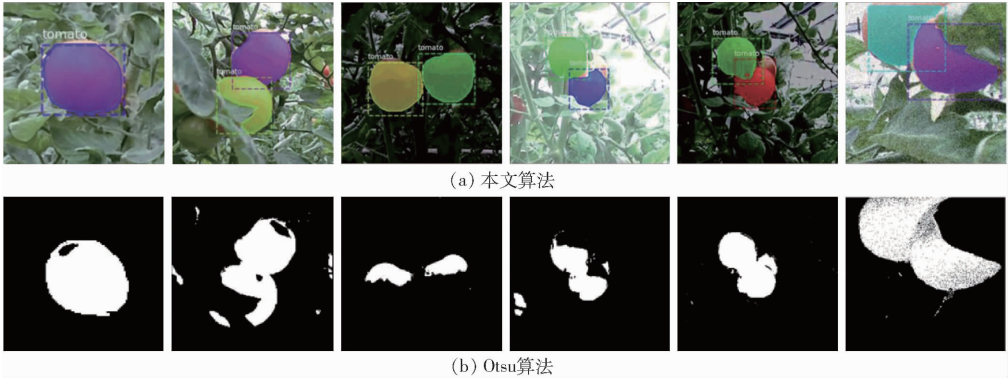


图 11 本文算法和 Otsu 算法预测结果局部对比

Fig. 11 Results of Otsu algorithm and proposed method

标注时间由原来 60 s 左右减少到 20 s 左右。

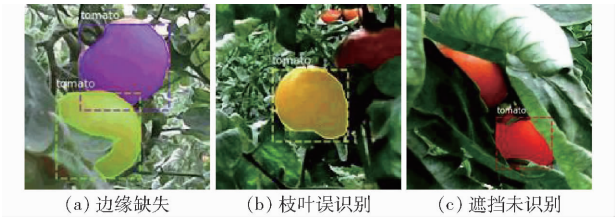


图 12 扩展 Mask R-CNN 算法识别失败案例示意图

Fig. 12 Failure cases with extended Mask R-CNN algorithm

(3)与以输入单一 RGB 图像的 Mask R-CNN 模型相比,基于多源图像的扩展 Mask R-CNN 算法其检测精度由 95.4% 提升至 98.3%, 交并比由 0.851 提升至 0.916。

(4)与传统 Otsu 检测算法和 YOLO v3 算法相比,该算法能够区分不同番茄果实个体,图像分割结果清晰,边缘轮廓完整,且受光线变化影响小,适合自然环境下的作业需求。

参 考 文 献

- [1] 熊露,朱孟帅. 番茄2016年市场分析及2017年市场预测[EB/OL]. [2017-01-22]. http://jiuban.moa.gov.cn/zwllm/jcyj/201701/t20170122_5461550.htm.
- [2] 董坦坦. 成熟番茄的图像识别及其位姿的获取研究[D]. 南京:南京农业大学, 2009.
DONG Tantan. Image recognition and posture detection of ripe tomato based on machine vision[D]. Nanjing: Nanjing Agricultural University, 2009. (in Chinese)
- [3] 杨长辉,刘艳平,王毅,等. 自然环境下柑橘采摘机器人识别定位系统研究[J/OL]. 农业机械学报, 2019, 50(12): 14-22.
YANG Changhui, LIU Yanping, WANG Yi, et al. Research and experiment on recognition and location system for citrus picking robot in natural environment[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(12): 14-22. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20191202&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2019.12.002. (in Chinese)
- [4] HENTEN E J V, TUIJL J V, HEMMING J, et al. Field test of an autonomous cucumber picking robot[J]. Biosystems Engineering, 2003, 86(3): 305-313.
- [5] 毕松,高峰,陈俊文,等. 基于深度卷积神经网络的柑橘目标识别方法[J/OL]. 农业机械学报, 2019, 50(5): 181-186.
BI Song, GAO Feng, CHEN Junwen, et al. Detection method of citrus based on deep convolution neural network[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(5): 181-186. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190521&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2019.05.021. (in Chinese)
- [6] 崔红光,任文涛,杨懿,等. 图像处理技术在农业机器人中的应用[J]. 农机化研究, 2008, 30(1): 168-170.
CUI Hongguang, REN Wentao, YANG Yi, et al. Application of graphic treatment technique in agriculture[J]. Journal of Agricultural Mechanization Research, 2008, 30(1): 168-170. (in Chinese)
- [7] 熊俊涛,郑镇辉,梁嘉恩,等. 基于改进YOLO v3网络的夜间环境柑橘识别方法[J/OL]. 农业机械学报, 2020, 51(4): 199-206.
XIONG Juntao, ZHENG Zhenhui, LIANG Jiaen, et al. Citrus detection method in night environment based on improved YOLO v3 network[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(4): 199-206. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20200423&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2020.04.023. (in Chinese)
- [8] WANG D, SONG H, YU X, et al. An improved contour symmetry axes extraction algorithm and its application in the location of picking points of apples[J]. Spanish Journal of Agricultural Research, 2015, 13(1): 0205.
- [9] LINKER R, COHEN O, NAOR A. Determination of the number of green apples in RGB images recorded in orchards[J]. Computers and Electronics in Agriculture, 2012, 81: 45-57.
- [10] DIAS P A, TABB A, MEDEIROS H. Apple flower detection using deep convolutional networks[J]. Computers in Industry, 2018, 99: 17-28.
- [11] ZHOU Y, XU T, ZHENG W, et al. Classification and recognition approaches of tomato main organs based on DCNN[J]. Transactions of the CSAE, 2017, 33(15): 219-226.
- [12] SA I, GE Z, DAYOUB F, et al. Deepfruits: a fruit detection system using deep neural networks[J]. Sensors, 2016, 16(8): 1222.
- [13] 孙哲,张春龙,葛鲁镇,等. 基于Faster R-CNN的田间西兰花幼苗图像检测方法[J/OL]. 农业机械学报, 2019, 50(7): 216-221.
SUN Zhe, ZHANG Chunlong, GE Luzhen, et al. Image detection method for broccoli seedlings in field based on Faster R-CNN[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(7): 216-221. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190723&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2019.07.023. (in Chinese)
- [14] 岳有军,田博凯,王红君,等. 基于改进Mask RCNN的复杂环境下苹果检测研究[J]. 中国农机化学报, 2019, 40(10): 128-134.
YUE Youjun, TIAN Bokai, WANG Hongjun, et al. Research on apple detection in complex environment based on improved Mask RCNN[J]. Journal of Chinese Agricultural Mechanization, 2019, 40(10): 128-134. (in Chinese)
- [15] 欧攀,路奎,张正,等. 基于Mask RCNN的目标识别与空间定位[J]. 计算机测量与控制, 2019, 27(6): 172-176.
OU Pan, LU Kui, ZHANG Zheng, et al. Target recognition and spatial location based on Mask RCNN[J]. Computer Measurement & Control, 2019, 27(6): 172-176. (in Chinese)
- [16] 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=20200601&journal_id=jcsam.

- 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)
- [17] GUPTA S, GIRSHICK R, ARBELÁEZ P, et al. Learning rich features from RGB - D images for object detection and segmentation[C]//European Conference on Computer Vision. Springer, Cham., 2014: 345 - 360.
- [18] 李俐,陈琦琦,张超,等. 基于 RNMU 的多源星载 SAR 影像融合与土地覆盖分类[J/OL]. 农业机械学报,2020,51(3): 191 - 200.
LI Li, CHEN Qiqi, ZHANG Chao, et al. Multi-source spaceborne SAR image fusion based on RNMU and land cover classification[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020,51(3):191 - 200. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag = 1&file_no = 20200322&journal_id = jcsam. DOI:10. 6041/j. issn. 1000-1298. 2020. 03. 022. (in Chinese)
- [19] KHAN S S, AHMAD A. Cluster center initialization algorithm for K-means clustering[J]. Pattern Recognition Letters, 2004, 25(11): 1293 - 1302.
- [20] MIRKIN B. Clustering: a data recovery approach[M]. Chapman and Hall/CRC, 2016.
- [21] MADHULATHA T S. An overview on clustering methods[J]. arXiv preprint arXiv:1205.1117, 2012.
- [22] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2016.
- [23] 姜红花,张传银,张昭,等. 基于 Mask R - CNN 的玉米田间杂草检测方法[J/OL]. 农业机械学报,2020,51(6):220 - 228,247.
JIANG Honghua, ZHANG Chuanyin, ZHANG Zhao, et al. Detection method of corn weed based on Mask R - CNN[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020,51(6):220 - 228,247. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag = 1&file_no = 20200623&journal_id = jcsam DOI: 10. 6041/j. issn. 1000-1298. 2020. 06. 023. (in Chinese)
- [24] TORREY L, SHAVLIK J. Transfer learning[M]//Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques. IGI Global, 2010: 242 - 264.
- [25] ERHAN D, BENGIO Y, COURVILLE A, et al. Why does unsupervised pre-training help deep learning? [J]. Journal of Machine Learning Research, 2010, 11(2): 625 - 660.
- [26] 刘振,杨玮,李民赞,等. 基于 Kinect 相机的土壤表面三维点云配准方法[J/OL]. 农业机械学报,2019,50(增刊):144 - 149.
LIU Zhen, YANG Wei, LI Minzan, et al. Three-dimensional point cloud registration method for soil surface based on Kinect camera[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(Supp.): 144 - 149. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag = 1&file_no = 2019s023&journal_id = jcsam. DOI:10. 6041/j. issn. 1000-1298. 2019. S0. 023. (in Chinese)
- [27] DENG J, DONG W, SOCHER R, et al. Imagenet: a large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 248 - 255.
- [28] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211 - 252.
- [29] DESELAERS T, ALEXE B, FERRARI V. Weakly supervised localization and learning with generic knowledge[J]. International Journal of Computer Vision, 2012, 100(3):275 - 293.
- [30] REN S, HE K, GIRSHICK R, et al. Faster R - CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(6): 1137 - 1149.
- [31] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.