

引用格式:

徐森, 朱波, 刘宇晨, 张冀武. 基于改进 Mask R-CNN 的多片烟叶部位的同步识别[J]. 湖南农业大学学报(自然科学版), 2023, 49(2): 170–175.

XU M, ZHU B, LIU Y C, ZHANG J W. Synchronous recognition method of multiple tobacco leaves position based on improved Mask R-CNN[J]. Journal of Hunan Agricultural University(Natural Sciences), 2023, 49(2): 170–175.

投稿网址: <http://xb.hunau.edu.cn>



基于改进 Mask R-CNN 的多片烟叶部位的同步识别

徐森, 朱波, 刘宇晨, 张冀武*

(昆明理工大学机电工程学院, 云南 昆明 650504)

摘要:为解决烟叶智能分级识别中需对多片散放烟叶同步进行部位识别的问题,提出一种基于改进 Mask R-CNN 的多片烟叶的部位同步识别方法:在 Mask R-CNN 区域建议网络中引入 K-means 聚类算法,对已标注目标检测框进行聚类,实现对预设的 5 种尺度的锚点尺寸和 3 种比例的锚点长宽比的优化,使其更加符合烟叶图像数据的分布特性,达到提高生成建议框的精确性、缩短识别时间的目的。基于采集的烟叶图像数据集,验证改进 Mask R-CNN 方法的有效性。结果表明,当 IoU 为 0.5 时,改进 Mask R-CNN 单样本耗时 313 ms,比 Mask R-CNN 的 326 ms 快,在测试集上的均值平均精度(mAP)提高了 3.56%。与 Faster R-CNN 和 SSD 目标检测算法相比,在准确率和召回率上也表现出优势。

关键词: 烟叶部位识别; Mask R-CNN; 区域建议网络; K-means 聚类

中图分类号: TP391.41

文献标志码: A

文章编号: 1007-1032(2023)02-0170-06

Synchronous recognition method of multiple tobacco leaves position based on improved Mask R-CNN

XU Miao, ZHU Bo, LIU Yuchen, ZHANG Jiwu*

(Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming, Yunnan 650504, China)

Abstract: In order to solve the problem of the synchronously identification on multiple scattered tobacco leaves in the intelligent classification and identification of tobacco leaves, a method for simultaneous recognition of multiple tobacco leaves position was proposed based on improved Mask R-CNN. The K-means clustering algorithm was introduced into the region proposal network of Mask R-CNN, to cluster the marked target detection frames. It realized the optimization of 5 preset anchor point sizes of sizes and 3 aspect ratio of the anchor point, and made it more in line with the distribution characteristics of tobacco leaf image data, so as to improve the accuracy of the generated suggestion box and shorten the recognition time. Based on the collected tobacco leaf image dataset, the effectiveness of the proposed method was verified. The experimental results show that when the IoU is 0.5, the single-sample time of this improved Mask R-CNN is 313 ms with the improved mAP value of 3.56% on the test set, which is faster than the 326 ms of the original Mask R-CNN. Comparing the target detection algorithms of Faster R-CNN and SSD, it also shows advantages in precision and recall rate.

Keywords: position recognition of tobacco leaves; Mask R-CNN; region proposal network; K-means clustering

收稿日期: 2022-02-10

修回日期: 2022-11-10

基金项目: 国家自然科学基金项目(52065033); 中国烟草总公司云南省烟草公司重点项目(2020530000241003、2021530000241012)

作者简介: 徐森(1998—),女,湖南岳阳人,硕士研究生,主要从事图像识别、计算机视觉技术研究, xm_someone@foxmail.com; *通信作者,张冀武,高级农艺师,主要从事烟叶质量等级质量监督检测研究, 1426857585@qq.com

烟叶分级是烟草加工工艺流程中的首要环节,对合理利用烟叶原料、充分发挥烟叶的最大价值意义重大^[1]。烟叶的智能分级已成为行业的研究热点。赵世民等^[2]利用计算机视觉技术提取烟叶颜色、形状特征进行分级,准确率达 90% 以上。LIU 等^[3]采用数字图像技术提取叶片特征,并采用广义回归神经网络确定其等级,最终检测精度为 93.5%。HE 等^[4]提出了一种基于机器视觉的烟叶等级识别方法,基于模糊模式识别算法,通过提取烟叶的外观特征,对预测集的识别准确率达到 80.23%。赖燕华等^[5]采集烟叶的近红外光谱,利用随机森林算法建立了不同霉变烟叶的识别模型,模型的识别和预测精度分别达 93.82% 和 94.84%。雒慧心^[6]通过图像处理技术对烟叶的现场数据进行预处理,基于深度卷积网络和迁移学习研究烟叶分组和分级。李亚召等^[7]基于霉杂烟叶图像分类问题,建立了一种基于多分支堆叠式的卷积神经网络霉杂烟叶图像分类模型。这些烟叶智能分级技术都局限于对单片烟叶的分级识别,在实际应用中存在效率瓶颈。为解决这一问题,业界提出了对多片散放烟叶同步进行分级识别的需求,其中包括对烟叶部位的识别。

HE 等^[8]提出的 Mask R-CNN 模型可以用于目标检测、关键点检测等计算机视觉任务,试验证明,该模型各方面均优于 Faster RCNN。权龙哲等^[9]提出一种基于改进 Mask R-CNN 的植物叶龄获取方法,通过构建不同天气和采集角度数据集,并将 NMS 更换为 Soft-NMS 算法、RoIAlign 代替 RoI Pooling 方法,其模型精度为分割模型,精度为 73%。YU 等^[10]为了提高机器视觉在草莓采摘机器人的检测性能,引入 Mask R-CNN,对于重叠、遮挡及不同光照下的水果检测都表现出较高的通用性和鲁

棒性。YANG 等^[11]为了实现堆叠电子元件的快速高效的自动检测,通过优化特征提取网络提出改进 Mask R-CNN,在缩减模型大小的同时提高了 2.1% 的精度。钟伟镇等^[12]通过引入 Mask R-CNN 模型,实现了复杂背景下多目标叶片的分割和识别,分割准确率达 97.51%。

鉴于 Mask R-CNN 的优越性能,笔者以 Mask R-CNN 为基础框架,构建了多片烟叶部位同步识别模型,并针对烟叶图像分布特性,对 Mask R-CNN 框架的区域建议网络通过 K-means 聚类算法进行了改进。为验证模型的有效性,基于采集的烟叶数据集进行了验证,结果表明,改进的 Mask R-CNN 模型在识别效率和精度上相比直接使用 Mask R-CNN 的方法都得到显著提升,并且在准确率和召回率上与目前主流目标检测算法 Faster R-CNN 和 SSD 相比有一定优势。现将结果报道如下。

1 烟叶图像数据的采集

运用云南佳叶公司开发的智能分级设备,拍摄烤烟上部叶、中部叶和下部叶的散放烟叶图像 3677 张,构建数据集。采用水平翻转、垂直翻转^[13]、椒盐噪声、高斯模糊^[14] 4 种数据增强方法获得 3677 幅图像(图 1)。按照 8 : 1 : 1 的比例,随机选择划分 2974 幅作为训练集,331 幅用作验证集,其余 372 幅作为测试集。

为了实现对烟叶检测与分割模型的训练,使用 Labelme 对烟叶图像进行标记。在标注过程中,对每幅图像中不同部位的烟叶分别进行标注,生成相应的 json 文件。该文件包含标注标签、标注点坐标信息。将标注后的烟叶分割掩码图像可视化,如图 2 所示。



1 不同部位烟叶混合图像; 2 垂直翻转; 3 水平翻转; 4 椒盐噪声; 5 高斯模糊。

图 1 烟叶图像的数据增强

Fig. 1 Data enhancement of images for tobacco leaves

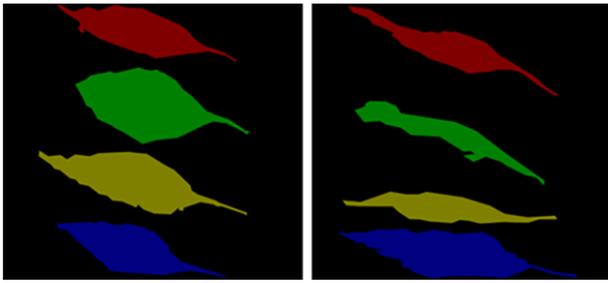


图 2 烟叶分割掩码

Fig.2 Examples of segmentation mask for tobacco leaves

2 基于改进 Mask R-CNN 的烟叶部位识别模型的构建

2.1 构建流程

提出基于改进 Mask R-CNN 的多片烟叶分割与烟叶部位识别方法，实现对目标叶片准确的定位、识别与分割。改进 Mask R-CNN 烟叶部位模型构建流程如图 3 所示。在改进 Mask R-CNN 中，将图片输入主干特征提取网络，提取的共享特征层被输入到区域建议网络(RPN)中，结合 K-means 聚类构建尺寸、比例不一的建议框；将包含感兴趣的目标的建议框传递到 RoIAlign 层，将其池化为大小固定的特征图；最终将 RoIAlign 层池化后的对象特征输入到全连接层和全卷积网络，进行对象分类和实例分割，获取类别置信度、边界框和分割掩膜。

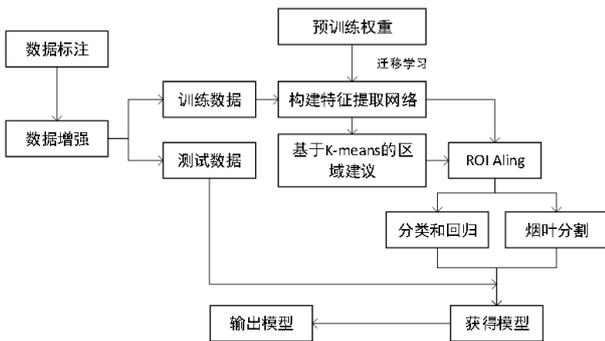


图 3 改进 Mask R-CNN 模型构建流程

Fig.3 Building process of improved Mask R-CNN model

2.2 基于迁移学习的模型初始化

为了有效地学习烟叶部位的特征信息，提升模型的泛化性能，基于迁移学习的思想，载入 Microsoft COCO 数据集^[15]预训练权重对模型进行微调。

2.3 烟叶特征提取网络结构的设计

为兼顾效率与网络的特征提取能力，选取

Resnet-50 作为特征提取网络。由于 ResNet-50 包含 50 层网络，如果只利用 ResNet-50 最后一层提取到的高层特征图进行预测，会影响目标检测的准确性；因此，将数据集输入残差网络后提取出图像数据压缩了 2 次(C2)、3 次(C3)、4 次(C4)、5 次(C5)的结果，构造特征金字塔网络(FPN)结构，保留不同阶段生成的特征图。采用 5 个卷积层模块的 ResNet-50 网络融合 FPN 进行烟叶样本图像的多维特征提取(图 4)，其中，P2 至 P5 为融合特征层。

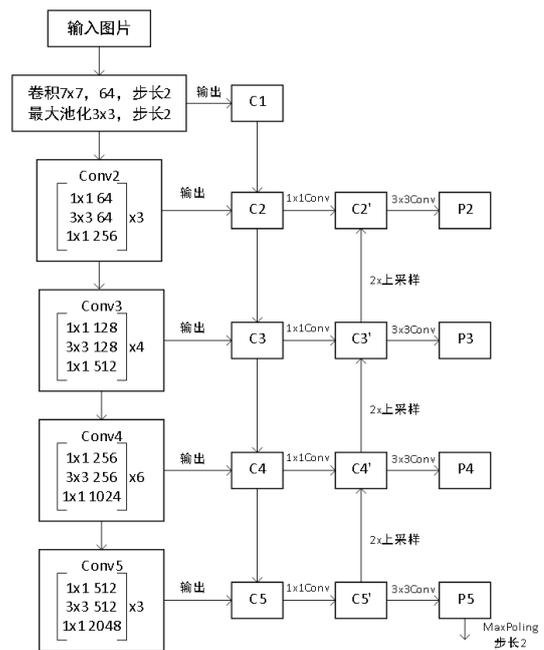


图 4 烟叶特征提取的 ResNet+FPN 结构

Fig.4 ResNet+FPN structure for characteristic extraction of tobacco leaves

针对 MS COCO、VOC^[16]数据集的目标类别，Mask R-CNN 预设了 5 种尺度大小的锚点和 3 种比例的锚点长宽比，但不一定适用于特定的烟叶检测目标。为了更好地适用于烟叶这一类检测目标的应用场景，采用 K-means 聚类^[17]方法重新生成锚点的尺寸及比例，列于表 1。

表 1 改进前后锚点的尺寸及比例

Table 1 Comparison of anchor size and scale between the improved model and the original model

网络	各层锚点尺寸	长宽比
Mask R-CNN	(32, 64, 128, 256, 512)	(0.5, 1, 2)
改进 Mask R-CNN	(146, 246, 309, 364, 441)	(1.5, 2.5, 3.5)

3 模型验证与分析

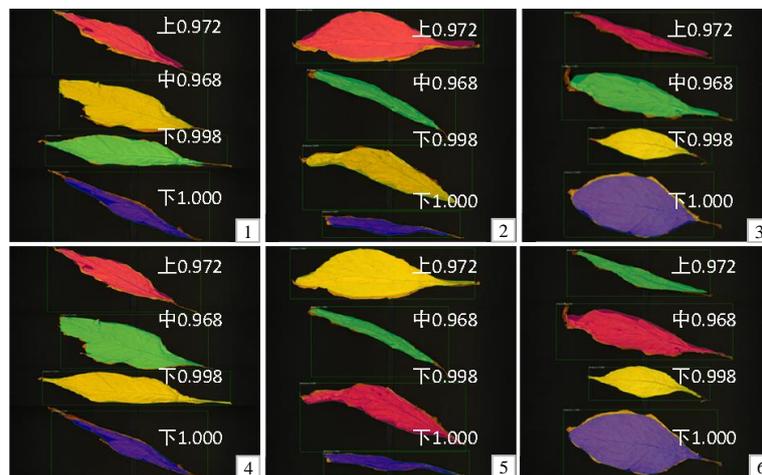
基于迁移学习的方法，提取 Microsoft COCO 数据集预训练模型权重作为初始化设置进行训练。

COCO 数据集中的图像目标通过精确的分割来进行其位置的标定, 主要从日常场景中截取。使用 Tensorflow-gpu 1.10.0 和 Keras 2.0.8 深度模型框架构建基于 Mask R-CNN 的烟叶实例分割与部位识别模型。模型基于 Windows 10 64 位版本操作系统, NVIDIA RTX2070(Super) GPU 进行训练。在调整参数的过程中, 将初始学习率设为 0.000 1, 动量因子为 0.9, 权重衰减系数为 0.000 1。

为了测试模型对于烟叶分割和部位识别的能力, 使用精确率、召回率、F1 值^[18]、平均精度 AP、均值平均精度 mAP、IoU^[19]作为评价指标。

在训练过程中, 记录目标检测过程中各个类别的平均精度值, 采用 mAP 作为目标检测预测位置和类别的性能指标。在实例分割部分, 采用各类 IoU 取平均值后的 mIoU 进行评估。

测试基于改进 Mask R-CNN 模型的多片烟叶实例分割与部位识别性能, 在测试集上对 2 种模型进行比较分析。为了检验模型的鲁棒性, 测试集中新增了对比度和亮度的增强和降低的图像数据。通过对改进前后 Mask R-CNN 的试验结果设置曲线平滑度并绘制 loss 曲线图, 如图 5 所示。由图 5 可知, Mask R-CNN 模型在第 40 次迭代接近收敛, 而改进 Mask R-CNN 在第 35 次迭代时已接近拟合。



1、2、3 Mask R-CNN 分割结果; 4、5、6 改进 Mask R-CNN 分割结果。

图 6 改进前后 Mask R-CNN 烟叶的分割结果

Fig. 6 Tobacco leaves segmentation results for the improved Mask R-CNN and the original Mask R-CNN

在识别精度上, 图 6-2 测试图片的第 1 片烟叶应为下部叶, 而 Mask R-CNN 将其误分为中部叶; 第 3 片烟叶应为上部叶, Mask R-CNN 将其误分为中部叶。在图 6-4、图 6-5、图 6-6 中, 改进 Mask R-CNN 不仅对烟叶部位的识别准确率较高, 而且

除此之外, 改进 Mask R-CNN 相比于 Mask R-CNN 有着更低的损失值。

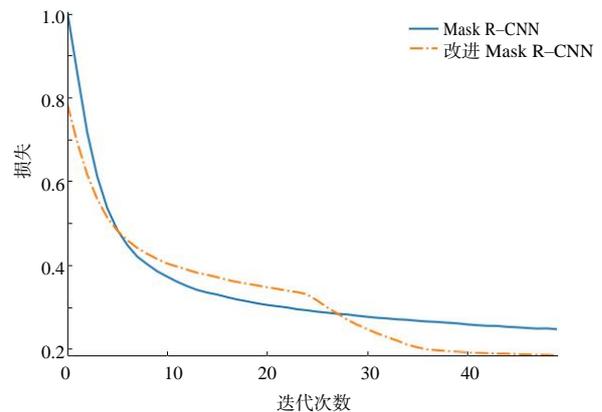


图 5 改进前后损失函数变化曲线

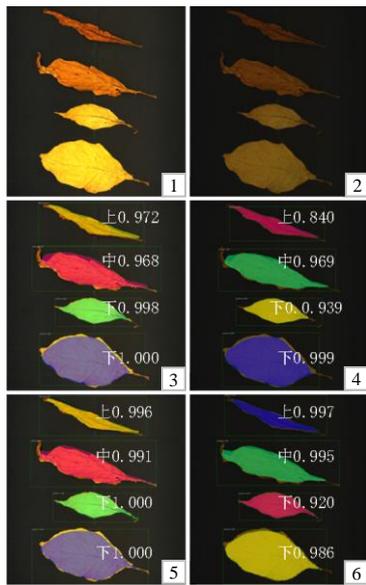
Fig.5 Change curve of loss function for the improved model and the original model

为了更好地验证基于改进 Mask R-CNN 的多片烟叶实例分割与烟叶部位识别模型的性能, 对测试集中部分分割结果可视化(图 6)。由图 6 可知, 虽然 Mask R-CNN 和改进 Mask R-CNN 都能取得不错的检测和分割效果, 但 Mask R-CNN 对烟梗部分几乎没有分割出来(图 6-1), 而对于烟梗的判定是烟叶部位识别中的重要指标, 因而稍显逊色。

不会出现过多的欠分割、过分割情况, 因此, 基于改进 Mask R-CNN 的模型更适用于多片烟叶的实例分割与部位识别。

对改变烟叶图像对比度和亮度的测试结果可视化分析如图 7 所示。在对比度和亮度增强时, 改

进前后的 Mask R-CNN 模型都能准确地识别和分割目标烟叶,但改进 Mask R-CNN 的识别精度更



1、3、5 对比度和亮度增强;2、4、6 对比度和亮度降低;3、4 Mask R-CNN;5、6 改进 Mask R-CNN。

图7 改变图像对比度和亮度的检测与分割结果

Fig.7 Detection and segmentation results by changing contrast and brightness of the images

高。在对比度和亮度降低时,尽管改进 Mask R-CNN 模型对于下部叶识别的精度略低于 Mask R-CNN 模型,但仍然表现出较高的识别和分割效果。在这种情况下,改进模型对于上部叶的识别要优于 Mask R-CNN,如图7中对于对比度和亮度降低的测试数据,Mask R-CNN 模型对于该上部叶的锚框置信度为0.840,但改进 Mask R-CNN 模型的锚框置信度达到0.997,具有更好的鲁棒性。

改进 Mask R-CNN 在测试集上 mAP 值可达到 92.76%,比 Mask R-CNN 的 89.20%提高 3.99%,且模型对于上部叶的识别效果最好。在 IoU 为 0.5 时,2 种模型的单样本耗时、精确率和召回率如表 2 所示。在 Resnet-50 作为特征提取网络的条件下,相比 Mask R-CNN,改进 Mask R-CNN 单样本耗时由 326 ms 缩短到 313 ms,提高了模型的检测速度。对比 2 种模型的 F1 值,改进 Mask R-CNN 和 Mask R-CNN 的 F1 值分别为 93.69%和 89.99%。可见改进 Mask R-CNN 在多片烟叶的部位识别上能表现出更好的检测性能。

表2 改进前后 Mask R-CNN 单样本耗时与精确率和召回率

网络模型	单样本耗时/ms	mAP/%	mIoU/%	精确率/%	召回率/%	F1/%
Mask R-CNN	326.0	89.20	77.39	87.13	90.94	89.99
改进 Mask R-CNN	313.0	92.76	82.34	92.26	95.16	93.69

在实例分割阶段,使用 mIoU 评估改进前后 Mask R-CNN 模型对烟叶的分割精度。改进 Mask R-CNN 的 mIoU 为 82.34%,而改进前的仅为 77.39%,可见改进 Mask R-CNN 模型在烟叶分割上依旧表现出良好的性能。

将改进 Mask R-CNN 与目标检测算法 Faster R-CNN 和 SSD 进行对比,Faster R-CNN 使用 VGG16 作为主干特征提取网络,并首次提出使用 RPN 网络生成感兴趣区域。SSD 为单阶段目标检测算法,去除了全连接层,经卷积神经网络提取特征后一次完成目标分类与定位。在 IoU 为 0.75 时检验模型性能,精确率和召回率列于表 3。改进 Mask R-CNN 多片烟叶部位识别模型精确率为 97.32%,相比于 Faster R-CNN、SSD 分别提高了 41.54%、13.93%;召回率相比 Faster R-CNN 和 SSD 分别提高了 2.20%、4.58%。可见改进 Mask R-CNN 烟叶识别模型检测性能最优。

表3 改进 Mask R-CNN 与其他模型精确率和召回率对比

网络模型	精确率/%	召回率/%
Faster R-CNN	68.76	87.62
SSD	85.42	85.63
改进 Mask R-CNN	97.32	89.55

4 结论

为解决烟叶智能分级中在烟叶部位识别上效率和精度不高的问题,提出了一种基于改进 Mask R-CNN 的多片烟叶实例分割与部位识别方法:首先,使用 K-means 聚类生成锚框的方法替换 Mask R-CNN 锚框设计,结合迁移学习的方法,加快了模型的拟合速度;在 IoU 为 0.5 时,将模型的 mAP 提高了 3.56%。改进 Mask R-CNN 与 Mask R-CNN 相比具有更好的分割性能,且单样本耗时缩短了 13 ms。

将改进 Mask R-CNN 方法与 2 种主流的目标检测算法(Faster R-CNN 和 SSD)相比,在准确率和召回率上也有优势。后续将研究使数据集背景和数据分配多样化,克服光照、角度因素的影响,进一步提高模型的分割性能。

参考文献:

- [1] ZHANG J, SOKHANSANJ S, WU S, et al. A trainable grading system for tobacco leaves[J]. *Computers and Electronics in Agriculture*, 1997, 16(3): 231–244.
- [2] 赵世民, 宋正雄, 贺智涛, 等. 基于图像特征的烟叶分级方法研究[J]. *安徽农业科学*, 2018, 46(28): 191–193.
- [3] LIU J J, SHEN J Y, SHEN Z Y, et al. Grading tobacco leaves based on image processing and generalized regression neural network[C]// 2012 IEEE International Conference on Intelligent Control, Automatic Detection and High-End Equipment. Beijing, China: IEEE, 2012: 89–93.
- [4] HE Y, WANG H J, ZHU S P, et al. Method for grade identification of tobacco based on machine vision[J]. *Transactions of the ASABE*, 2018, 61(5): 1487–1495.
- [5] 赖燕华, 林云, 陶红, 等. 烟叶霉变的快速识别: 基于近红外光谱与随机森林算法[J]. *中国烟草学报*, 2020, 26(2): 36–43.
- [6] 雒慧心. 基于深度学习的烤烟分选算法研究[D]. 北京: 北京交通大学, 2019.
- [7] 李亚召, 云利军, 叶志霞, 等. 基于卷积神经网络的霉变烟叶图像识别方法研究[J]. *计算机工程与科学*, 2021, 43(3): 473–479.
- [8] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 386–397.
- [9] 权龙哲, 吴冰, 毛首人. 基于 Mask R-CNN 农田杂草实例分割与叶龄识别方法[J]. *东北农业大学学报*, 2021, 52(4): 65–76.
- [10] YU Y, ZHANG K L, YANG L, et al. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN[J]. *Computers and Electronics in Agriculture*, 2019, 163: 104846.
- [11] YANG Z X, DONG R X, XU H, et al. Instance segmentation method based on improved mask R-CNN for the stacked electronic components[J]. *Electronics*, 2020, 9(6): 886.
- [12] 钟伟镇, 刘鑫磊, 杨坤龙, 等. 基于 Mask-RCNN 的复杂背景下多目标叶片的分割和识别[J]. *浙江农业学报*, 2020, 32(11): 2059–2066.
- [13] 项荣, 张茂琛. 基于改进 Mask R-CNN 的番茄茎秆分类方法[J]. *湖南大学学报(自然科学版)*, 2023, 50(2): 31–39.
- [14] 刘坤, 杨怀卿, 杨华等. 基于循环残差注意力的群养生猪实例分割[J]. *华南农业大学学报*, 2020, 41(6): 169–178.
- [15] YU W J, CHEN Z D, LUO X, et al. DELTA: a deep dual-stream network for multi-label image classification[J]. *Pattern Recognition*, 2019, 91: 322–331.
- [16] EVERINGHAM M, ZISSERMAN A, WILLIAMS C K I, et al. The 2005 PASCAL visual object classes challenge [M]// *Machine Learning Challenges Workshop*. Berlin, Germany: Springer, 2006: 117–176.
- [17] MACQUEEN J B. Some methods for classification and analysis of multivariate observations[C]// *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Los Angeles, USA: University of California Press, 1967: 281–297.
- [18] HUANG H, XU H H, WANG X H, et al. Maximum F1-score discriminative training criterion for automatic mispronunciation detection[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2015, 23(4): 787–797.
- [19] REZATOFI G H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2020: 658–666.

责任编辑: 罗慧敏
英文编辑: 吴志立