

# 基于 CSS-Cascade Mask R-CNN 的有遮挡多片烟叶部位识别

朱波<sup>1</sup>, 胡朋<sup>1</sup>, 刘宇晨<sup>2</sup>, 张冀武<sup>2\*</sup>

(1. 昆明理工大学机电工程学院, 昆明 650504; 2. 云南烟草质量监督监测站, 昆明 650800)

**摘要:** 烟叶部位信息是进行烟叶分级的重要参考信息, 准确识别烟叶部位对实现烟叶智能分级具有重要意义。在实际的烟叶智能分级应用中, 为了提高分级效率, 需要对多片烟叶等级进行同步识别。受现行上料方式的限制, 同步识别的多片烟叶间往往存在局部遮挡的问题, 给烟叶的目标检测和部位识别带来挑战。该研究提出一种基于改进 Cascade Mask R-CNN, 融合通道、非局部和空间注意力机制, 并引入柔性极大值抑制检测框交并操作与斯库拉交并比损失函数 (SIoU) 的目标检测与识别模型 (CSS-Cascade Mask R-CNN)。该模型对 Cascade Mask R-CNN 进行了三方面的改进: 一是在其骨干网络 Resnet101 上同时引入通道、非局部、空间 3 种注意力机制, 使网络更加关注未被遮挡且部位特征明显区域的显著度; 二是将 Cascade Mask R-CNN 中的损失函数 SmoothL1Loss 替换为 SIoU 损失函数, 将预测框与真实框之间的方向差异引入到模型训练中提高模型检测精度; 三是在筛选候选框时将常规的非极大抑制 (non-max-suppression) 替换为柔性非极大抑制, 以避免删除候选框造成信息丢失。试验结果表明, 利用提出的模型对有遮挡多片烟叶进行检测和部位识别, 检测框平均准确率均值 (bbox\_mAP50) 达到了 80.2%, 与改进前的 Cascade Mask R-CNN 模型相比提高了 7.5 个百分点。提出的模型与多个主流的目标检测模型 (YOLOvX、YOLOv3、YOLOv5、Mask R-CNN、Cascade R-CNN) 相比, 分别高 7.1、10.2、5.8、9.2、8.4 个百分点, 尤其是对较难区分的下部烟叶优势明显, 因此研究结果可以为有遮挡多片烟叶部位的检测识别提供参考。

**关键词:** 机器视觉; 烟叶部位识别;; 局部遮挡;; Cascade Mask R-CNN;; 非局部注意力机制

doi: 10.11975/j.issn.1002-6819.202311090

中图分类号: TP391.41

文献标志码: A

文章编号: 1002-6819(2024)-09-0271-10

朱波, 胡朋, 刘宇晨, 等. 基于 CSS-Cascade Mask R-CNN 的有遮挡多片烟叶部位识别[J]. 农业工程学报, 2024, 40(9): 271-280. doi: 10.11975/j.issn.1002-6819.202311090 <http://www.tcsae.org>

ZHU Bo, HU Peng, LIU Yuchen, et al. Recognition of the position for partially occluded multiple tobacco leaves based on CSS-Cascade Mask R-CNN[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(9): 271-280. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202311090 <http://www.tcsae.org>

## 0 引言

烤烟叶是卷烟工业的主要原料, 其品质直接影响最终烟草制品的感官质量。对烤烟叶进行精确分级对提升烟叶均质化加工水平、合理利用资源和增加烟农收入具有重要意义。近年来, 将计算机视觉技术与机器学习技术相结合取代人工对烤烟叶进行智能分级, 以改变人工操作存在主观偏差和效率低下的问题, 正在成为烟草工业发展的大趋势<sup>[1-2]</sup>。

对烟叶进行智能分级, 首先需要对烟叶在烟株上的着生位置 (即上部叶、中部叶、下部叶) 进行准确划分。从快速成像的多片烟叶图像中分辨出每片烟叶的部位, 需要进行多目标检测与识别。此外, 限于烟叶分级上料设备在技术上还没取得大的突破, 同时进入到成像区域中的多片烟叶间往往还存在局部遮挡, 给其准确检测和部位识别造成更大挑战。

当前, 计算机视觉研究领域已提出了多种基于深度

学习的目标检测算法模型, 包括单阶段和双阶段两大类。单阶段算法如 YOLO 系列<sup>[3-7]</sup> 和 SSD 系列<sup>[8-11]</sup>, 样本标注简单, 检测速度快, 但在复杂场景下准确性难以保证。双阶段算法如 R-CNN 系列<sup>[12-17]</sup>, 先生成目标预选框, 然后再分类, 通常更精确, 但计算量也更大。现有文献中不论是单阶段还是双阶段算法, 绝大多数都只考虑了目标无遮挡的情况, 对有遮挡的情况研究较少, 尤其对多个目标相互遮挡的情况研究更少。薛珊等<sup>[18]</sup> 提出了基于 YOLOX-S 改进的反无人机系统抗遮挡目标检测算法, 通过在 CSPNet 模块中引入 CA 注意力机制突出有用特征抑制无用特征, 各类别平均检测精度的平均值达到 93.4%。王鲁等<sup>[19]</sup> 提出了基于改进 Cascade Mask R-CNN 与协同注意力机制的群猪姿态识别, 通过构建基于协同注意力机制的轻量级识别模型实现猪体姿态识别, 模型准确率达到 98.2%。林耀海等<sup>[20]</sup> 提出结合渐进式中心定位的重叠蜜柚分离方法和果梗定位方法。利用主成分分析方法、渐进式中心定位与区域边缘点, 定位出遮挡程度较小的蜜柚果梗。在自然场景下的图像上进行试验, 结果表明在有阴影、小目标、遮挡和重叠等复杂环境下, 该方法的平均识别率为 94.02%。袁山等<sup>[21]</sup> 提出一种改进 Mask R-CNN 模型对植物叶片进行实例分割。通过引入级联检测模块、注意力机制和一个 2 层 3×3 卷积模块

收稿日期: 2023-11-09 修订日期: 2024-03-22

基金项目: 中国烟草总公司云南省烟草公司重点项目 (2021530000241012)

作者简介: 朱波, 博士, 讲师, 硕士生导师, 研究方向为信息化、机器学习模式识别。Email: [zhubo20110720@163.com](mailto:zhubo20110720@163.com)

\*通信作者: 张冀武, 高级农艺师, 研究方向为烟叶等级智能识别技术与烟叶质量监督检测。Email: [1426857585@qq.com](mailto:1426857585@qq.com)

对模型分割分支进行优化,解决相互遮挡、叶片边缘特征不明显等问题,在 CVPPP 数据集上的对称最佳 Dice 得分 (symmetric best dice, SBD) 为 90.3%。

以上文献都对有遮挡的多个目标进行检测和识别,但是,上述研究只需要关注目标外形轮廓特征即可,而且目标外形本身也较为简单。相比而言,烟叶部位识别需要同时考虑其外形、纹理和颜色等特征,而且烤烟叶还存在卷曲、褶皱甚至破损等情形,其外形轮廓属于高度不规则形体,故其检测和识别要复杂得多,直接运用以上所述算法难以取得较好效果。为此,本文提出一种基于改进 Cascade Mask R-CNN<sup>[22-26]</sup>,融合通道、非局部和空间的注意力机制 (channel-nonlocal-space attention mechanism, CNS-AM),并引入柔性极大值抑制 (soft-NMS) 检测框交并操作与斯库拉交并比损失函数 (Scylla-IoU, SIoU) 的模型 (CSS-Cascade Mask R-CNN),以期解决有遮挡多片烟叶部位的检测与识别问题。

## 1 试验材料与方法

### 1.1 烟叶样本图像采集

为开展模型验证,本文采用自主研发的烟叶智能分级试验设备(如图 1 所示)采集烟叶样本图像。该试验设备配置了 MV-CA050-11UM/UC 型工业相机,其使用 Sony 的 IMX264 芯片,通过 USB3.0 接口实时传输非压缩图像,最高帧率达 35 帧/s,获取图像质量优异,满足试验需求。试验装置为封闭箱体,可以避免不同光源、环境等外部因素的干扰,获得高质量图像。样本图像尺寸为:2 348 像素×1 528 像素,存储为 JPG 格式。



图 1 烟叶智能分级试验设备

Fig.1 Test equipment for intelligent classification of tobacco leaves

烟叶样本取自于 2023 年在云南某地实际收购的烤烟,样本集中包含上部叶、中部叶和下部叶各 1 500 片,不同部位的烟叶外观特征如表 1 所示。

表 1 不同部位的烟叶外观特征

Table 1 Appearance characteristics of different parts of tobacco leaves

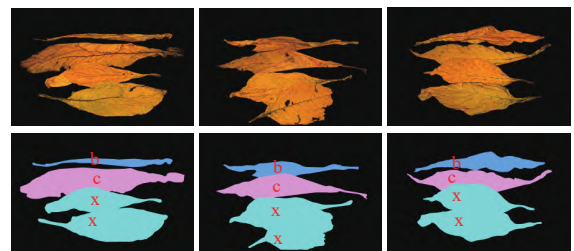
特征 Characteristics	烟叶部位 Tobacco position		
	上部叶 Upper leaf	中部叶 Middle leaf	下部叶 Lower leaf
脉相 Vein characteristics	较粗到粗,较显露至凸起	粗细适中,遮盖至微露,叶尖稍弯	较细,遮盖,叶尖弯曲
叶形 Leaf shape	稍窄至较窄	宽至较宽,叶尖较钝	较宽圆
叶面 Blade characteristics	稍皱至平坦	稍皱缩	皱缩
颜色 Colour	橘黄至红棕	橘黄色	柠檬黄

每次从样本集中随机选取 4 片烟叶散放在托盘上,送入试验装置中进行图像采集。为了模拟实际多片烟叶

进行同步分级时存在的相互遮挡情况,对采集的图像进行人工检测,剔除掉无遮挡或完全遮挡的样本。按照该过程最终采集到有不同程度遮挡的多片烟叶样本图像共计 1 125 张。数据集按照 7:3 的比例随机选择划分,选取 788 幅作为训练集,337 幅作为验证集。

### 1.2 烟叶样本图像标注

为了使采集的有遮挡多片烟叶样本图像能够满足后续研究的需要,采用 Labelme 工具对烟叶对象进行标注。用一个多边形锚框将每个图片中的每片烟叶手工标注出来,确保锚框刚好包络住完整的烟叶轮廓。标注好的样本图像信息存储到 JSON 文件中,其中包含标注标签和标注点坐标信息。标注好的样本烟叶分割掩码图像如图 2 所示。



注:图中第一行是采集后的原始图片;第二行是标注后的掩码图片,其中的 b、c、x 分别表示上部叶、中部叶和下部叶。

Note: The first line in the figure is the original picture after collection; The second line is the labeled mask picture, where b, c, and x represent the upper, middle, and lower leaves respectively.

图 2 标注好的多片烟叶掩码图像

Fig.2 Labeled mask image of multiple tobacco leaves

### 1.3 试验环境及参数设置

开展模型试验的软硬件环境如表 2 所示。模型相关超参数设置如下:初始学习率 lr=0.006,动量 momentum factor=0.9,权重衰减为 0.000 1;输入图像尺寸统一为 1 333 像素×800 像素;模型训练批量大小 Batch\_Size=4,训练迭代次数 Epoch=50;模型训练使用 multi step 学习率调整策略,并设置在 32、42 代进行学习率衰减。

表 2 试验软硬件环境

Table 2 Experimental software and hardware environment

名称 Name	参数 Parameter
处理器 CPU	12 th Gen Intel(R) Core(TM) i7-12 700
图形处理器 GPU	NVIDIA GeForce RTX 3 090
操作系统	Windows 11
编程语言	Python 3.8.16
深度学习框架	Pytorch 2.0.0

## 2 改进的 Cascade Mask R-CNN 模型

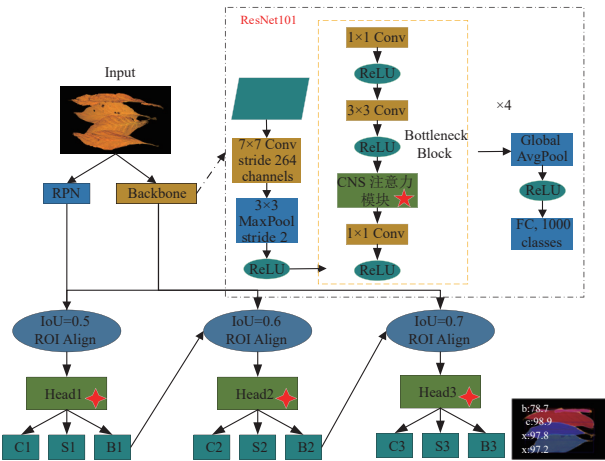
在 R-CNN 模型中引入级联的思想最早在 CVPR2018 上被提出,实现的 Cascade R-CNN 模型<sup>[27-29]</sup>通过级联多个检测器来提高目标检测精度。之后,级联思想又被引入到 Mask R-CNN 模型中进行改进,形成 Cascade Mask R-CNN 模型。级联思想广泛运用在农业和医学图像检测<sup>[30-34]</sup>上且效果显著。



## 2.1 CSS-Cascade Mask R-CNN 模型

本文提出的 CSS-Cascade Mask R-CNN 模型继承了 Cascade Mask R-CNN 模型的基础结构, 整体可分为 4 个部分, 分别为特征提取网络 (Backbone 网络)、区域建议网络 (region proposal network, RPN)<sup>[35]</sup>、感兴趣区域对齐网络 (region of interest align, RoI Align) 以及各个功能输出网络。常用的特征提取网络有 ResNet 系列<sup>[36-38]</sup>与 ResNeXt 系列<sup>[39-41]</sup>, 主要负责对输入图像进行特征提取。区域建议网络用于提取候选框, 在模型训练阶段通过非极大值抑制方法筛选出一批优秀候选框。感兴趣区域对齐网络负责对 RPN 确定的提案框包含的不同大小的特征图进行尺度统一。各功能输出网络包括分类、检测和分割 3 种类型的输出。CSS-Cascade Mask R-CNN 中级联检测头由感兴趣区域对齐网络与各个功能输出网络组成, 在级联检测头中模型将每个候选区域 (RoI) 输入到 RoI Align, 通过对特征进行插值计算确保从各种尺寸的 RoI 中提取的特征图大小一致, 然后各个功能输出网络将 RoI Align 处理后的特征进行分类和边界框回归, 将回归后的候选框送入下一个阶段。在第二阶段, 使用较高的 IoU 阈值 (如 0.6) 来进一步重复上述操作。第三阶段也是如此, 通过三次级联处理得到更加精确的预测框与检测识别结果。

前期试验表明, Cascade Mask R-CNN 模型在目标被部分遮挡使得可视区域减少时, 很容易出现漏检和误检的现象。为此, 提出的 CSS-Cascade Mask R-CNN 模型在 Cascade Mask R-CNN 基础上主要进行了如下三方面的改进: 1) 通过引入 CBAM (convolutional block attention module) 与非局部模块相融合的陈 (channel-nonlocal-space) 注意力机制; 2) 采用 SIoU<sup>[42]</sup> 损失函数取代 SmoothL1loss 作为边界框损失函数; 3) 在 RPN 阶段采用 Soft-NMS<sup>[43]</sup> 算法取代 NMS 算法。CSS-Cascade Mask R-CNN 模型的结构如图 3 所示, 详细的改进工作如 2.2~2.4 节所述。



注: 五角星代表改进注意力机制部分; 四角星代表改进损失和 NMS 部分; C1、S1、B1; C2、S2、B2; C3、S3、B3 分别代表 IOU 阈值为 0.5、0.6 和 0.7 的分类、检测、分割结果。  
Note: The pentagram represents the improved attention mechanism part; the quadrangle represents the improved loss and NMS part; C1, S1, B1; C2, S2, B2; and C3, S3, B3 represent the classification, detection, and segmentation results with IOU thresholds of 0.5, 0.6, and 0.7, respectively.

图 3 CSS-Cascade Mask R-CNN 模型结构

Fig.3 CSS-Cascade Mask R-CNN model structure

## 2.2 CNS 注意力机制

Cascade Mask R-CNN 卷积网络在特征提取过程中关注局部特征, 但在有遮挡的情况下, 这些局部特征可能显示不全, 导致检测精度下降。针对该问题, 提出一种将卷积注意力机制模块 CBAM<sup>[44]</sup> 与非局部模块 Non-local Module<sup>[45]</sup> 相融合的陈 (CNS) 注意力机制, 如图 4 所示。

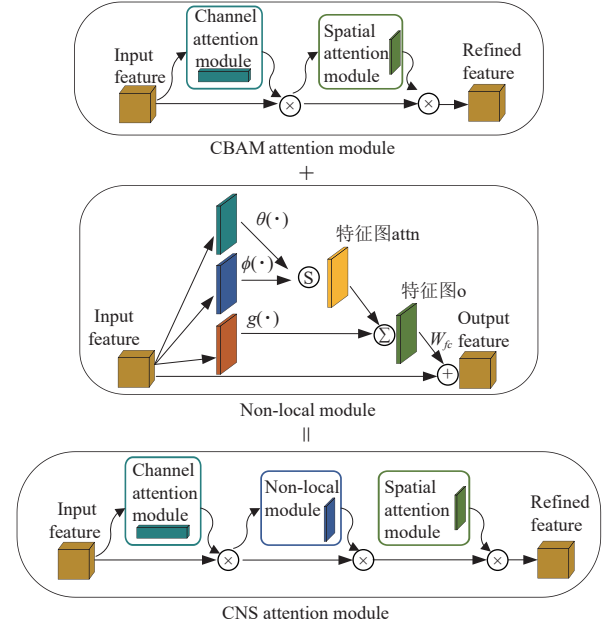


图 4 CBAM、CNS 与 Non-local Module 结构图

Fig.4 CBAM, CNS, and Non-local Module structure diagram

非局部操作关注特征图中任意两位置间的关联性, 超越了传统的局部邻域范围, 具体通过计算两位置间特征向量的点积, 并应用 Softmax 函数进行指数归一化来实现。此操作使得模型能够捕获图像上的长距离依赖关系, 更好理解图像的整体结构, 特别是其具有对遮挡图像中的关联部分进行整合的能力, 可以获得一个完整的特征表示, 因此对处理存在遮挡的图像非常有效。

非局部操作的具体计算过程如下:

步骤 1 特征转换: 假设输入特征图  $X$  的维度为  $(B, C, H, W)$ , 其中  $B$  为批量大小,  $C$  为通道数,  $H$  和  $W$  分别为高度和宽度。特征转换通过使用 3 个  $1 \times 1$  卷积核进行卷积操作实现:

$$\theta(x) = X \otimes \theta \quad (1)$$

$$\phi(x) = X \otimes \phi \quad (2)$$

$$g(x) = X \otimes g \quad (3)$$

式中  $X$  为输入特征图;  $\otimes$  为卷积操作;  $\theta, \phi, g$  为卷积层的权重; 得到的  $\theta(x), \phi(x), g(x)$  维度都变为  $(B, C/2, H, W)$ 。

步骤 2 计算注意力图: 将  $\theta(x), \phi(x)$  通过 reshape 操作转换为  $(B, C/2, H \times W)$  形状, 然后通过矩阵乘法计算它们的关系矩阵  $f$ :

$$f = \theta(x) \times \phi(x)^T \quad (4)$$

式中  $f$  的维度为  $(B, H \times W, H \times W)$ 。再接着用

Softmax 函数对  $f$  的每一行进行归一化, 得到注意力图  $\text{attn}$ 。

步骤 3 特征加权: 将  $g(x)$  通过 reshape 操作转换为  $(B, C/2, H \times W)$  的形状, 并使用注意力图  $\text{attn}$  对  $g(x)$  进行加权, 得到特征图  $o$ :

$$o = \text{attn} \times g(x) \quad (5)$$

然后通过 reshape 操作转换回  $(B, C/2, H, W)$  的形状, 得到加权特征图。

步骤 4 特征融合: 通过  $1 \times 1$  卷积层  $W_{fc}$  调整加权特征图  $o$  的通道数, 使其与输入特征图  $x$  的通道数相匹配, 并将其与原始输入特征图  $x$  相加, 实现特征融合:

$$y = W_{fc}(o) + x \quad (6)$$

式中  $y$  为输出特征图。

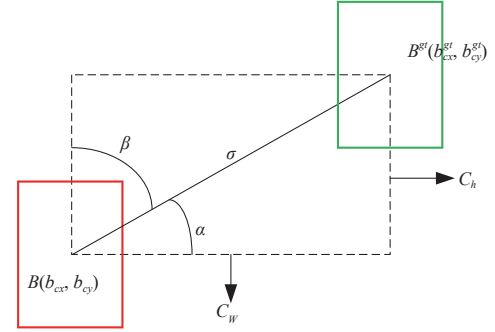
通过上述操作能够在特征图的不同位置之间建立直接的联系, 从而有效增强了网络对全局信息的捕获能力。这种操作对于处理视觉任务中的复杂场景和目标相互遮挡的图像尤为重要。

本文在 Cascade Mask R-CNN 模型中 Backbone 的 Resnet101 中引入 CNS 注意力, 使输入的图像依次进行通道、非局部与空间注意力模块的计算, 使得 Backbone 能够捕获长距离的依赖关系, 进而更为全面地理解被局部遮挡烟叶的整体结构, 而不是仅关注局部特征。从而显著影响模型的性能, 更好地满足实际应用的需求。

### 2.3 边界框回归损失优化

Cascade Mask R-CNN 模型中, 边界框回归损失采用的是 SmoothL1 Loss 损失, 但 SmoothL1 Loss 只关注预测边界框坐标与真实坐标之间的差异。即使预测框与真实框有很好的重叠, 但只要它们的坐标有微小的偏差, 损失就可能很大, 然而在烟叶局部遮挡的情况下, 这种重叠的情况是非常常见的, 这会导致预测的边界框失真。针对以上问题, 本研究采用 SIoU 损失函数替换 SmoothL1 Loss。SIoU 包含 4 个部分: 角度损失 (angle cost)、距离损失 (distance cost)、形状损失 (shape cost) 以及交并比损失 (intersection over union)。其中, 角度损失旨在最小化预测边界框和实际边界框之间的角度偏差。在烟叶识别中, 烟叶可能会以各种角度出现, 尤其是当它们相互遮挡时。通过最小化角度差异, 模型能够更准确地预测烟叶的方向, 这对于正确识别和分类遮挡的烟叶至关重要。距离损失用于最小化预测边界框和真实边界框中心点之间的距离, 这有助于确保预测框准确地定位在目标对象上。在处理遮挡的烟叶时, 准确的定位可以显著提高检测的精度, 特别是当多个烟叶彼此接近或部分重叠时; 形状损失致力于匹配预测边界框和真实边界框的形状。这不仅包括大小, 还包括边界框的长宽比例。由于烟叶的形状和大小可能因种类和遮挡程度而异, 因此精确的形状匹配对于提高识别精度至关重要。交并比损失是衡量预测边界框和真实边界框之间重叠程度的一个标准指标, 它计算两个边界框交集和并集的比例。在多片遮挡的场景中, 高 IoU 值意味着模型能够准确地识

别和分割重叠的烟叶, 这对于准确评估每片烟叶的质量和类别至关重要。如图 5 为 SIoU 参数示意图。



注:  $B$  和  $B^{gt}$  点分别为预测框中心点及真实框中心点;  $C_w$ 、 $C_h$  分别为  $B$  和  $B^{gt}$  点的横坐标与纵坐标之差;  $\alpha$  为两中心点连线与水平线的夹角 ( $^\circ$ ),  $\beta$  为两中心点连线与垂直线的夹角 ( $^\circ$ );  $(b_{cx}, b_{cy})$  为  $B$  坐标,  $(b_{cx}^{gt}, b_{cy}^{gt})$  为  $B^{gt}$  坐标;  $\sigma$  为  $B$  和  $B^{gt}$  距离。

Note:  $B$  and  $B^{gt}$  are the center points of the prediction frame and the real frame, respectively;  $C_w$  and  $C_h$  are the differences between the horizontal and vertical coordinates of points  $B$  and  $B^{gt}$ , respectively;  $\alpha$  is the angle between the line connecting the two center points and the horizontal line;  $\beta$  is the angle between the line connecting the two center points and the vertical line, ( $^\circ$ );  $(b_{cx}, b_{cy})$  and  $(b_{cx}^{gt}, b_{cy}^{gt})$  are coordinates of  $B$  and  $B^{gt}$ ;  $\sigma$  is distance between  $B$  and  $B^{gt}$ .

图 5 SIoU 计算示意图

Fig.5 The scheme for calculation of SIoU

1) 角度损失  $\Lambda$  计算公式如下:

$$\Lambda = 1 - 2\sin^2\left(\arcsin(x) - \frac{\pi}{4}\right) \quad (7)$$

其中  $x = \frac{C_h}{\sigma} = \sin \alpha \cdot C_w$

2) 定义距离损失  $\Delta$  计算公式如下:

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma P_t}) \quad (8)$$

其中:

$$P_x = \left(\frac{b_{cx}^{gt} - b_{cx}}{C_w}\right)^2, \quad P_y = \left(\frac{b_{cy}^{gt} - b_{cy}}{C_h}\right)^2, \quad \gamma = 2 - \Lambda$$

式中  $\gamma$  为角度损失相关的调整因子,  $P_t$  为预测框和真实框中心点的归一化距离, 用于计算两个框中心之间的距离损失。

3) 形状损失  $\Omega$  定义如下:

$$\Omega = \sum_{t=w,h} (1 - e^{\omega_t})^z \quad (9)$$

其中:

$$\omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}, \quad \omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})}$$

式中  $z$  为控制对形状损失的关注程度,  $w_w$ 、 $w_h$  分别为预测框和真实框在宽度与高度上的相对误差。

为避免过于关注形状损失而降低对预测框的移动,  $\theta$  参数范围为 2~6 之间。  $w$  和  $h$  为预测框的宽和高,  $w^{gt}$  和  $h^{gt}$  为真实框的宽和高。

4) SIoU 损失的定义为:

$$L_{siou} = 1 - U + \frac{\Delta + \Omega}{2} \quad (10)$$

式中  $U$  为传统回归损失,  $\Lambda$  为角度损失,  $\Delta$  为距离损失,

$\Omega$  为形状损失。

#### 2.4 Soft-NMS

Cascade Mask R-CNN 模型中, RPN 阶段采用非极大值抑制 (NMS), 它主要就是通过迭代的形式, 不断以最大得分的框去与其他框做 IoU 操作, 并过滤那些 IoU 较大 (即交集较大) 的框。这样做有一个明显的缺点就是它将相邻检测框的分数均强制归零 (即将重叠部分大于重叠阈值的检测框移除)。在这种情况下, 如果一个真实物体在重叠区域出现, 则将导致对该物体的检测失败并降低了算法的平均检测率。针对上述问题, 本研究选择使用了在 NMS 算法基础上进行了改进的 Soft-NMS。与 NMS 算法将交并比大于阈值的框直接剔除不同, soft-NMS 将交并比大于阈值边界框的置信度得分代入式 (11) 计算得出新的置信度得分。

$$s_i = \begin{cases} s_i & \text{IoU}(M, b_i) < N_t \\ s_i(1 - \text{IoU}(M, b_i)), & \text{IoU}(M, b_i) > N_t \end{cases} \quad (11)$$

式中  $s_i$  为边界框置信度得分,  $M$  为当前拥有最高置信度的检测框,  $b_i$  为其他候选检测框;  $\text{IoU}(M, b_i)$  为置信度最高得分边界框与当前边界框的 IoU 计算值;  $N_t$  为阈值。

Soft-NMS 能够降低 IoU 计算结果大于阈值的次优边界的置信度得分, 而不是直接舍弃, 这在一定程度上缓解了因为存在局部遮挡导致烟叶漏检的现象, 提升了烟叶目标检测的效果。

#### 2.5 评价指标

为验证模型的有效性及其优越性, 本文选取交并比作为模型在目标检测和实例分割任务上的性能指标。交并比用于衡量模型预测的边界框与真实边界框之间的一致性, 其计算式如 (12) 所示。

$$\text{IoU} = (B_p, B_{gt}) = \frac{\text{Area}(B_p \cap B_{gt})}{\text{Area}(B_p \cup B_{gt})} \quad (12)$$

式中  $B_p$  为预测的边界框,  $B_{gt}$  为真实边界框。

当预测的边界框与真实边界框的 IoU 大于或等于 0.5 时, 该预测被认为是真正例。精确度 ( $P$ ) 和召回率 ( $R$ ) 计算如式 (13)、(14) 所示。

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (13)$$

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (14)$$

式中 TP 为真正例数, FP 为假正例数, FN 为假反例数。

在  $\text{IoU}=0.5$  的特定类别的平均精度 (AP) 通过计算精确度-召回率曲线下面积得到; 再通过各类别平均精度计算总体平均检测精度均值  $\text{bbox\_mAP50}$ , 如式 (15)、(16):

$$\text{AP}_{\text{IoU}=0.5} = \int_0^1 P(R) dR \quad (15)$$

$$\text{bbox\_mAP50} = \frac{1}{C} \sum_{c=1}^C \text{AP}_{\text{IoU}=0.5,c} \quad (16)$$

式中  $C$  为类别总数。

### 3 试验结果与分析

#### 3.1 CNS 注意力机制各模块组合顺序对检测效果的影响

Cascade Mask R-CNN 选用 Resnet101 作为模型的骨干网络, 对输入图像进行特征提取, 在特征提取过程中在 Resnet101 中引入 CNS 注意力机制, 并将结果作为后续任务的输入。CNS 注意力机制是基于 CBAM 注意力机制改进而来, 其包含 3 个模块: 通道注意力模块、空间注意力模块和非局部模块。

由于不同模块对不同烟叶部位的敏感性不同, 某些烟叶部位可能在空间布局上更为突出, 而其他部位可能在特征通道 (如纹理或颜色) 上更加显著, 再加上不同模块组合的交互效应, 每个注意力模块对输入特征的处理方式不同, 因此它们的组合顺序可能导致不同的交互效应。例如, 空间注意力可能突出某些区域的特征, 而通道注意力随后可能进一步加强这些区域的特定通道。不同的组合顺序可能导致特征在不同层级上的强化或抑制, 影响最终的特征表示。设计网络时, 需要根据任务的具体需求和烟叶的特征来决定最有效的组合顺序。因此, 为了探究哪种组合方式最适用于部分遮挡烟叶部位检测任务, 在 Cascade Mask R-CNN 的 Resnet101 中分别引入 CSN (channel-space-nonlocal)、CNS (channel-nonlocal-space) 和 NCS (nonlocal-channel-space), 并与 CBAM 注意力机制在相同试验条件下进行试验, 试验结果如表 3 所示。

表 3 各模块不同组合顺序检测结果

Table 3 The test results of each module are combined in different sequences

网络 Network	注意力 模块 Attention module	各部位平均检测精度 mAP50 of each part of the leaf/%			平均检测 精度均值 bbox_mAP 50/%	帧率 Frames per second/ (帧·s <sup>-1</sup> )
		上部叶 Upper leaf	中部叶 Middle leaf	下部叶 Lower leaf		
		93.2	68.8	56.7	72.7	51.6
Cascade Mask R-CNN	CBAM	89.0	62.9	73.0	75.0	38.9
	CSN	89.4	70.2	69.9	76.5	34.2
	CNS	93.0	68.9	68.5	76.8	33.6
	NCS	85.4	68.8	69.8	74.6	32.5

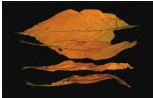
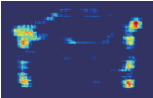
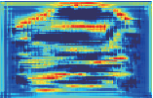
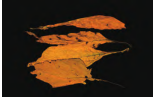
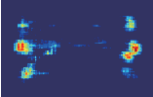
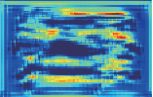
试验结果表明, Cascade Mask R-CNN 在加入 CNS 注意力机制后, 获得最高的平均检测精度, 且从帧率 (frames per second) 值看来, 其效率也并不是所有组合中最低的。由此表明在骨干网络 Resnet101 中引入 CNS 这一组合顺序形成的注意力机制后, 能够更加有效的提取出烟叶的非遮挡区域呈现出的烟叶部位特征, 进而提高对烟叶部位的识别性能。

为进一步探究 Cascade Mask R-CNN 引入 CNS 注意力机制模块后对遮挡多片烟叶部位识别性能提高的机理, 对烟叶样本图片进行热力图可视化分析, 如表 4 所示。其中, 红色区域为模型着重关注的显著区域, 颜色越深代表显著度越高。从表 4 可以看出, Cascade Mask R-CNN 模型只是关注到烟叶图片样本的一些局部特征, 而在烟叶样本之间存在相互遮挡的情况下, 被关注的局部特征可能恰好是被遮挡部分, 这就会导致被遮挡烟叶出现误



检。在骨干网络 Resent101 中引入 CNS 注意力机制可以使检测网络利用整个烟叶的全局特征，更全面地理解烟叶的形态，即使在叶片有局部遮挡情况下也能进行有效检测。

表 4 改进前后 CAM 可视化  
Table 4 Improved before and after CAM visualization

样本 Sample	原始样本图像 Original sample image	Resent101 提取样本特征图 Resent101 extract the sample feature map	引入 CNS 的 Resent101 提取样本特征图 Introduce CNS Resent101 extract the sample feature map
1			
2			

3.2 消融试验

为了探究 CNS 注意力机制、SIoU 损失函数和 Soft-NMS 模块对提出的 CSS-Cascade Mask R-CNN 模型所起的作用，进行了消融试验，结果如表 5 所示。

从表 5 可知，单纯引入 CNS 模块后，Cascade Mask R-CNN 网络的 bbox\_mAP50 达到了 76.8%，比之前的值提升了 4.1 个百分点。这是因为 CNS 注意力机制引入使

得模型能够从仅关注烟叶局部特征的限制中解放出来，转而关注和整合更广泛的区域信息。这种从局部到全局的关注转变，提高了模型对遮挡情况的适应性和检测的准确性，因为它允许模型利用图像中的全局信息来补充或校正局部特征的不足，获取更多烟叶部位信息。

单纯引入 Siou 损失函数的 Cascade Mask R-CNN 网络的 bbox\_mAP50 值达到了 73.9%，相比之前的值提高了 1.2 个百分点，主要是在烟叶局部遮挡的情况下 Siou 损失函数更能关注预测框与真实框重叠的情况，提升了遮挡情况下烟叶检测效果。单纯引入 Soft-NMS 模块的 Cascade Mask R-CNN 网络的 bbox\_mAP50 值达到 74.7%，相比之前的值提高了 2.0 个百分点，验证了 Soft-NMS 的有效性。将 CNS 注意力机制、Siou 损失函数和 Soft-NMS 模块一起引入得到的 CSS-Cascade Mask R-CNN 模型，其 bbox\_mAP50 值达到了 80.2%，相比 Cascade Mask R-CNN 模型提高了 7.5 个百分点。消融试验结果表明，提出的 CSS-Cascade Mask R-CNN 模型在 Cascade Mask R-CNN 模型基础上所做的 3 项改进均发挥了作用，虽然在效率上有所降低，但在检测精度上得到了显著提升，对有遮挡多片烟叶的检测和部位识别更加有效。

表 5 网络模型消融试验  
Table 5 Network model ablation experiment

网络 Network	CNS	Siou	Soft-NMS	各部位平均检测精度 mAP50 of each part of the leaf/%			平均检测精度均值 bbox_mAP50/%	帧率 Frames per second/(帧·s <sup>-1</sup> )
				上部叶 Upper leaf	中部叶 Middle leaf	下部叶 Lower leaf		
Cascade Mask R-CNN	×	×	×	93.2	68.8	56.7	72.7	51.6
	√	×	×	93.0	68.9	68.5	76.8	33.6
	×	√	×	89.5	64.0	68.3	73.9	45.1
	×	×	√	88.0	72.1	64.0	74.7	38.3
	√	√	√	92.0	73.4	75.0	80.2	21.2

3.3 与其他模型的检测精度对比试验

为检验 CSS-Cascade Mask R-CNN 模型相比其他目标检测模型的优势，选择了多个目前流行的目标检测模型进行对比试验，包括单阶段模型 YOLOvX、

YOLOv3 和 YOLOv5，以及双阶段模型 Mask R-CNN、CascadeR-CNN。分别对比了上部叶、中部叶和下部叶的检测精度，以及整体的平均检测精度。试验结果如表 6 所示。

表 6 与其他模型对比结果  
Table 6 Compare results with other models

网络 Network	各部位平均检测精度 mAP50 of each part of the leaf/%			平均检测精度均值 bbox_mAP50/%	平均检测精度均值 bbox_mAP75/%	帧率 Frames per second/(帧·s <sup>-1</sup> )
	上部叶 upper leaf	中部叶 Middle leaf	下部叶 Lower leaf			
YOLOvX	80.9	73.0	65.5	73.1	51.2	84.2
YOLOv3	80.3	63.2	66.4	70.0	43.8	88.6
YOLOv5	86.3	68.7	68.3	74.4	52.3	90.6
Mask R-CNN	93.5	66.4	48.9	71.0	56.7	21.4
Cascade R-CNN	81.0	73.4	61.9	71.8	48.8	51.6
CSS-Cascade Mask R-CNN	92.0	73.4	75.0	80.2	78.6	21.2

从表 6 可知，提出的 CSS-Cascade Mask R-CNN 模型在上部叶、中部叶和下部叶上得到的平均精度 mAP50 值分别为 92.0%、73.4%、75.0%，除了在上部叶上略低于 Mask R-CNN 的 93.5%，在中部叶和下部叶上均达到最高，尤其是在特征最不明显的下部叶上的表现相比几种对比模型优势明显。计算出在 3 种部位的烟叶上得到的平均精度均值均值 bbox\_mAP50，可以看出 CSS-Cascade Mask R-CNN 模型与几种主流的目标检测模型，

即 YOLOvX、YOLOv3、YOLOv5、Mask R-CNN 和 Cascade R-CNN 相比，分别高出 7.1、10.2、5.8、9.2、8.4 个百分点。为了考查模型的鲁棒性，进一步提高 IoU 阈值计算出更加严苛的检测精度指标 bbox\_mAP75 值。几种对比模型的 bbox\_mAP75 值与相应的 bbox\_mAP50 值相比均有不同程度的降低，但 CSS-Cascade Mask R-CNN 模型的值只是略微减小，达到了 78.6%，大幅度领先几种对比模型，相比最好的单阶段模型 YOLOv5

(52.3%) 高出了 26.3 个百分点，相比 Mask R-CNN (56.7%) 也高出了 21.9 个百分点。

以上结果证明了 CSS-Cascade Mask R-CNN 模型在检测精度和鲁棒性上均具有显著的优越性。

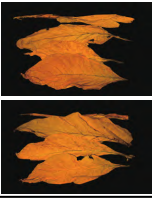
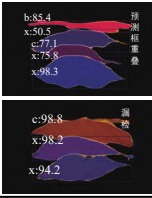
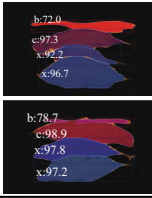

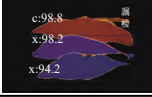

为进行效率对比，依次计算出各模型的 FPS 值。本文提出的 CSS-Cascade Mask R-CNN 模型 FPS 值相比 YOLO 系列模型低，相比 Mask R-CNN 和 Cascade R-CNN 也有不同程度的差距。但是，考虑到目前智能分级设备受限于机械化上料方式造碎率高的问题只能采用人工上料方式，其效率瓶颈并不在识别计算环节，而在人工上料环节，CSS-Cascade Mask R-CNN 模型的识别计算时间相比人工上料时间依然要短得多，因此认为目前追求更高识别精度而不是识别计算效率是值得的，也即使使用该模型比效率明显高出的 YOLO 系列模型更加合理。

3.4 试验结果可视化分析

为展示 CSS-Cascade Mask R-CNN 模型对有遮挡多片烟叶部位的检测效果，从测试集中随机抽取样本进行测试，并与原 Cascade Mask R-CNN 模型的检测结果进行对比。其中较有代表性的 2 个样本对比情况如表 7 所示。

表 7 改进前后检测结果对比

Table 7 Comparison of inspection results before and after improvements

样本 Sample	原始样本图像 Original sample image	Cascade Mask R-CNN	CSS Cascade Mask R-CNN
1			
2			

由表 7 可知，Cascade Mask R-CNN 模型进行检测时，样本 1 中出现了一个目标多个结果的情况，其中一个结果将中部叶误判为下部叶 (x: 50.5)；样本 2 中则出现了上部叶漏检的情况。误判和漏检的情况在 Cascade Mask R-CNN 模型检测时出现较多，而在 CSS-Cascade Mask R-CNN 模型检测时则没有出现。

4 结 论

针对烟叶智能分级中因为多片烟叶间存在局部遮挡导致对烟叶部位识别精度不高的问题，本文提出了一种改进的 Cascade Mask R-CNN 模型。

1) 通过在其骨干网络 Resnet101 上引入通道、非局部、空间融合注意力机制，其目的是通过学习一组权重值来表示每个特征图以及每个空间位置的重要程度，从而使网络可以学习更多有用的特征，增强有遮挡多片烟叶部位特征的显著度，进而实现目标检测网络前期特征的有效提取，改进后的模型平均检测精度均值 bbox\_mAP50 值达到了 76.8%，相比未改进之前提高 4.1 个百分点。

2) 将 Cascade Mask R-CNN 中常用的边界框回归损

失函数 SmoothL1 Loss 替换为 SIoU 损失函数，将预测框与真实框之间的方向差异引入到模型训练中提高模型检测精度，改进后的模型平均检测精度均值 bbox\_mAP50 值达到 73.9%。

3) 在筛选候选框时将常规的非极大抑制 (NMS) 算法替换为柔性非极大抑制 (Soft-NMS) 算法，以避免删除候选框造成信息丢失，改进后的模型平均检测精度均值 bbox\_mAP50 值达到 74.7%，相比之前提高 2.0 个百分点。

试验结果表明，本文提出的目标检测模型 CSS-Cascade Mask R-CNN 应用于有遮挡多片烟叶部位识别，与几种主流的目标检测模型，即 YOLOvX、YOLOv3、YOLOv5、Mask R-CNN 和 Cascade R-CNN 相比，平均检测精度均值 bbox\_mAP50 值分别高出 7.1、10.2、5.8、9.2、8.4 个百分点。下一步工作将针对遮挡率更高和有更多片烟叶的情形进行研究，并就烟叶部位识别的准确率对遮挡比例的敏感性进行定量分析。

[参 考 文 献]

[1] ZENG N, QIU H, WANG Z, et al. A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease[J]. Neurocomputing, 2018, 320: 195-202.

[2] 汪强, 席磊, 任艳娜, 等. 基于计算机视觉技术的烟叶成熟度判定方法[J]. 农业工程学报, 2012, 28(4): 175-179.

WANG Qiang, XI Lei, REN Yanna, et al. Determination of tobacco leaf maturity degree based on computer vision technology[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2012, 28(4): 175-179. (in Chinese with English abstract)

[3] REDMON J, FARHAD A. YOLOv3: An Incremental Improvement[EB/OL]. [2023-11-09] arXiv. 1804.02767. <https://doi.org/10.48550/arXiv.1804.02767>.

[4] 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. 2016: 779-788.

[5] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada. 2021: 2778-2788.

[6] 陈锋军, 陈闯, 朱学岩, 等. 基于改进 YOLOv7 的油茶果实成熟度检测[J]. 农业工程学报, 2024, 40(5): 177-186.

CHEN Fengjun, CHEN Chuang, ZHU Xueyan, et al. Detection of camellia oleifera fruit maturity based on improved YOLOv7[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(5): 177-186. (in Chinese with English abstract)

[7] HU M Z, LI Z Y, YU T, et al. Efficient Lightweight YOLO: Improving small object detection in YOLO for aerial images[J]. Sensors (Basel, Switzerland), 2023, 23(14): 6423

- [8] LIU W, ANGUELOV D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, Proceedings, Part I 14, Springer International Publishing, 2016: 21-37.
- [9] 刘阳, 高国琴. 基于轻量级 SSD 模型的夜间金蝉若虫检测[J]. 农业工程学报, 2022, 38(10): 166-175.
- LIU Yang, GAO Guoqin. Detecting nocturnal golden cicada nymphs using lightweight SSD model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(10): 166-175. (in Chinese with English abstract)
- [10] 彭红星, 李荆, 徐慧明, 等. 基于多重特征增强与特征融合 SSD 的荔枝检测[J]. 农业工程学报, 2022, 38(4): 169-177.
- PENG Hongxing, LI Jing, XU Huiming, et al. Litchi detection based on multiple feature enhancement and feature fusion SSD[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(4): 169-177. (in Chinese with English abstract)
- [11] SUN X, GU J, HUANG R. A modified SSD method for electronic components fast recognition[J]. Optik, 2020, 205: 163767.
- [12] 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, 2017, 39(6): 1137-1149.
- [13] HE K, GKIOXARI G, Dollár P, et al. Mask R-CNN[C] // Proceedings of the IEEE International Conference on Computer Vision.Venice, Italy, 2017: 2961-2969.
- [14] CAI Z, VASCONCELOS N, INTELLIGENCE M. Cascade R-CNN: High quality object detection and instance segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(5): 1483-1498.
- [15] 谭鹤群, 李玉祥, 朱明, 等. 通过图像增强与改进 Faster-RCNN 网络的重叠鱼群尾数检测[J]. 农业工程学报, 2022, 38(13): 167-176.
- TAN Hequn, LI Yuxiang, ZHU Ming, et al. Detecting overlapping fish population using image enhancement and improved Faster-RCNN networks[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(13): 167-176. (in Chinese with English abstract)
- [16] 张海燕, 徐心语, 马雪芬, 等. 超声图像中复合材料褶皱形态的 Mask-RCNN 识别方法[J]. 物理学报, 2022, 71(7): 209-216.
- ZHANG Haiyan, XU Xinyu, MA Xuefen, et al. Mask-RCNN recognition method of composite fold shape in ultrasound images[J]. Acta Physica Sinica, 2022, 71(7): 209-216. (in Chinese with English abstract)
- [17] 汤鹏, 刘毅, 魏宏光, 等. 基于 Mask-RCNN 海上升压站数
- 字式仪表读数的自动识别算法[J]. 红外与激光工程, 2021, 50(S2): 163-170.
- TANG Peng, LIU Yi, WEI Hongguang, et al. Automatic recognition algorithm of digital instrument reading in offshore booster station based on Mask-RCNN[J]. Infrared and Laser Engineering, 2021, 50(S2): 163-170. (in Chinese with English abstract)
- [18] 薛珊, 王亚博, 吕琼莹, 等. 基于 YOLOX-drone 的反无人机系统抗遮挡目标检测算法[J]. 工程科学学报, 2023, 45(9): 1539-1549.
- XUE Shan, WANG Yabo, LÜ Qiongying, et al. Anti-occlusion target detection algorithm for anti-UAV system based on YOLOX-drone[J]. Chinese Journal of Engineering, 2023, 45(9): 1539-1549. (in Chinese with English abstract)
- [19] 王鲁, 刘晴, 曹月, 等. 基于改进 Cascade Mask R-CNN 与协同注意力机制的群猪姿态识别[J]. 农业工程学报, 2023, 39(4): 144-153.
- WANG Lu, LIU Qing, CAO Yue, et al. Posture recognition of group-housed pigs using improved Cascade Mask R-CNN and cooperative attention mechanism[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(4): 144-153. (in Chinese with English abstract)
- [20] 林耀海, 吕钟亮, 杨长才, 等. 自然场景图像中的重叠蜜柚识别及试验[J]. 农业工程学报, 2021, 37(24): 158-167.
- LIN Yaohai, LYU Zhongliang, YANG Changcai, et al. Recognition of the overlapped honey pomelo images in natural scene and experiment[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(24): 158-167. (in Chinese with English abstract)
- [21] 袁山, 汤浩, 郭亚. 基于改进 Mask R-CNN 模型的植物叶片分割方法[J]. 农业工程学报, 2022, 38(1): 212-220.
- YUAN Sha, TANG Hao, GUO Ya. Segmentation method for plant leaves using an improved Mask R-CNN model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(1): 212-220. (in Chinese with English abstract)
- [22] ZHENG Y, QIN L, QIU T, et al. Automated detection and recognition of thyroid nodules in ultrasound images using Improve Cascade Mask R-CNN[J]. Multimedia Tools and Applications, 2021, 81(10): 1-21.
- [23] WEIFENG H, CAIZHI L, XIANGFAN N, et al. Recognition and detection of aero-engine blade damage based on Improved Cascade Mask R-CNN. [J]. Applied optics, 2021, 60(17): 5124-5133.
- [24] YEOL H O, SARFRAZ M K, BAE S J, et al. Automated detection of greenhouse structures using cascade Mask R-CNN[J]. Applied Sciences, 2022, 12(11): 5553-5553.
- [25] DIAZ P M, TITTUS P. Fast detection of wind turbine blade damage using Cascade Mask R-DSCNN-aided drone inspection analysis[J]. Signal, Image and Video Processing,



- 2023, 17(5): 2333-2341.
- [26] ZHANG Y, MA Y, LI Y, et al. Intelligent analysis method of dam material gradation for asphalt-core rock-fill dam based on enhanced Cascade Mask R-CNN and GCNet[J]. *Advanced Engineering Informatics*, 2023, 56: 102001.
- [27] CAI Zhaowei, NUNO Vasconcelos. Cascade R-CNN: Delving into high quality object detection[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA, 2017: 6154-6162.
- [28] 张娜, 包梓群, 罗源, 等. 改进的 Cascade R-CNN 算法在目标检测上的应用[J]. *电子学报*, 2023, 51(4): 896-906.
- ZHANG Na, BAO Ziqun, LUO Yuan, et al. Application of improved Cascade R-CNN algorithm in target detection[J]. *Acta Electronica Sinica*, 2023, 51(4): 896-906. (in Chinese with English abstract)
- [29] 赵振兵, 熊静, 李冰, 等. 基于改进 Cascade R-CNN 的典型金具及其部分缺陷检测方法[J]. *高电压技术*, 2022, 48(3): 1060-1067.
- ZHAO Zhenbing, XIONG Jing, LI Bing, et al. Typical Fittings and its Partial Defect Detection Method Based on Improved Cascade R-CNN[J]. *High Voltage Engineering*, 2022, 48(3): 1060-1067. (in Chinese with English abstract)
- [30] 薛卫, 程润华, 康亚龙, 等. 基于 GC-Cascade R-CNN 的梨叶病斑计数方法[J]. *农业机械学报*, 2022, 53(5): 237-245.
- XUE Wei, CHENG Runhua, KANG Yalong, et al. Pear leaf disease spot counting method based on GC-Cascade R-CNN[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(5): 237-245. (in Chinese with English abstract)
- [31] 谢元澄, 何超, 于增源, 等. 复杂大田场景中麦穗检测级联网络优化方法[J]. *农业机械学报*, 2020, 51(12): 212-219.
- XIE Yuancheng, HE Chao, YU Zengyuan, et al. Optimization method for cascade network of wheat ear detection in complex filed scene[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(12): 212-219. (in Chinese with English abstract)
- [32] JIA Y, WANG H, CHEN W, et al. An attention-based cascade R-CNN model for sternum fracture detection in X-ray images[J]. *CAAI Transactions on Intelligence Technology*, 2022, 7(4): 658-670.
- [33] LIU R, ZHAO Y, LIU Y, et al. Automated detection of lesion in computer tomography images based on Cascade R-CNN[C]//2020 International Conference on Image, Video Processing and Artificial Intelligence. SPIE, Shanghai, China, 2020, 11584: 270-275.
- [34] LI M, LEE S. A study on small pest detection based on a CascadeR-CNN-Swin model[J]. *Computers, Materials Continua*, 2022, 72(3): 6155-6165.
- [35] RAO Y, CHENG Y, XUE J, et al. FPSiamRPN: Feature pyramid Siamese network with region proposal network for target tracking[J]. *IEEE Access*, 2020, 8: 176158-176169.
- [36] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 770-778.
- [37] 谢州益, 冯亚枝, 胡彦蓉, 等. 基于 ResNet18 特征编码器的小稻病虫害图像描述生成[J]. *农业工程学报*, 2022, 38(12): 197-206.
- XIE Zhouyi, FENG Yazhi, HU Yanrong, et al. Generating image description of rice pests and diseases using a ResNet18 feature encoder[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(12): 197-206. (in Chinese with English abstract)
- [38] 万鹏, 赵竣威, 朱明, 等. 基于改进 ResNet50 模型的大宗淡水鱼种类识别方法[J]. *农业工程学报*, 2021, 37(12): 159-168.
- WAN Peng, ZHAO Junwei, ZHU Ming, et al. Freshwater fish species identification method based on improved ResNet50 model[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2021, 37(12): 159-168. (in Chinese with English abstract)
- [39] XIE S, GIRSHICK R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 1492-1500.
- [40] 宋磊, 李嵘, 焦义涛, 等. 基于 ResNeXt 单目深度估计的幼苗植株高度测量方法[J]. *农业工程学报*, 2022, 38(3): 155-163.
- SONG Lei, LI Rong, JIAO Yitao, et al. Method for measuring seedling height based on ResNeXt monocular depth estimation[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(3): 155-163. (in Chinese with English abstract)
- [41] 王军龙, 宣魁, 熊海涛, 等. 基于改进 ResNeXt50 残差网络的锦鲤选美方法[J]. *农业机械学报*, 2023, 54(S1): 330-337.
- WANG Junlong, XUAN Kui, XIONG Haitao, et al. Beauty pageant of koi method based on improved ResNeXt50 residual network[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(S1): 330-337. (in Chinese with English abstract)
- [42] GEVORGYAN Z. Siou loss: More powerful learning for bounding box regression[EB/OL]. [2023-11-09] arxiv: 2205.12740. <http://doi.org/10.48550/arXiv.2205.12740>.
- [43] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS improving object detection with one line of code[C]// Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 5561-5569.
- [44] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany, 2018: 3-19.
- [45] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7794-7803.

# Recognition of the position for partially occluded multiple tobacco leaves based on CSS-Cascade Mask R-CNN

ZHU Bo<sup>1</sup>, HU Peng<sup>1</sup>, LIU Yuchen<sup>2</sup>, ZHANG Jiwu<sup>2\*</sup>

(1. Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming 650504, China;

2. Yunnan Tobacco Quality Supervision and Monitoring Station, Kunming 650800, China )

**Abstract:** Object detection technologies have played a pivotal role in agricultural informatics at present. Among them, the precision and efficiency of grading systems have been enhanced for tobacco leaves. These technologies can also be integrated to realize the accurate identification and classification of tobacco leaf features. Particularly, the partial occlusion has commonly occurred in the images of multiple tobacco leaves. In this study, an innovative recognition model was proposed using an improved Cascade Mask Region-based Convolutional Neural Network (Cascade Mask R-CNN). Three types of attention mechanisms were also introduced, including channel, nonlocal, and spatial attention. Furthermore, a refined detection framework was then incorporated to employ the SIOU loss function, thereby improving the detection accuracy. The automatic grading of tobacco was effectively realized using visual features. Specifically, Cascade Mask R-CNN was then chosen to treat the complex tasks of image recognition. Then, a series of enhancements were employed to improve the overall performance. Firstly, Channel-Nonlocal-Space (CNS) attention mechanism was introduced into the ResNet101 backbone. The improved model was then focused mainly on the unoccluded regions of the image with distinct and salient features. By doing so, this improved model was used to better discern the nuanced characteristics of tobacco leaves for the classification. Secondly, the conventional SmoothL1Loss function was replaced with the more sophisticated SIOU loss function. The former was typically employed in the Cascade Mask R-CNN for the loss calculation, but without considering the directional difference between the predicted and actual bounding box. As a result, the improved model benefited from the enhanced detection accuracy and expedited convergence, thus facilitating more precise identification of tobacco leaf features. Thirdly, the traditional Non-Max Suppression (NMS) was replaced with the Soft Non-Max Suppression (Soft NMS), in order to avoid the information loss caused by simple and rough deletion of candidate boxes in the standard NMS, thereby preserving more available information for the final analysis. A series of experiments was conducted to verify the effectiveness of the improved model. 1125 images of partially occluded multiple tobacco leaves were acquired from the actual tobacco leaf grading system. An average accuracy of 80.2% (bbox\_mAP50) was achieved 7.5 percentage points higher, compared with the Cascade Mask R-CNN model. The better performance was also achieved by 7.1, 10.2, 5.8, 9.2, and 8.4 percentage points, respectively, compared with the multiple mainstream object detection models, namely the YOLOvX, the YOLOv3, the YOLOv5, the Mask R-CNN, and the Cascade R-CNN. Especially, the outstanding performance was also obtained for the lower tobacco leaves difficult to identify. The findings can provide an effective approach for the position recognition of partially occluded multiple tobacco leaves.

**Keywords:** computer vision; recognition of tobacco leaf position; partially occluded; Cascade Mask R-CNN; CNS attention mechanism