基于改进 YOLOv5s 的苗圃内树苗及障碍物目标检测方法

刘 慧，郑鑫澎，沈 跃※，王思远，沈卓凡，开锦茹
（江苏大学电气信息工程学院，镇江 212013）

摘 要：为解决电动喷雾机器人在苗圃作业时对树苗、行人和栽培盆识别准确率低等问题，该研究提出一种基于改进 YOLOv5s（you only look once version 5 small）的多目标检测方法。首先对骨干网络进行改进，将部分卷积（partial convolution，PConv）引入综合卷积模块（comprehensive convolution block，C3）中以减少网络模型计算量；在骨干网络最高维特征后添加坐标注意力机制（coordinate attention，CA）以提升位置感知精度；优化网络模型颈部结构并使用双线性插值进行上采样操作，增强模型特征提取能力；最后对原耦合检测头进行更换并优化检测部分的结构，使用改进的轻量化解耦头 LD 检测头（light decouple detection head）以进一步提升检测精度。试验结果表明，改进后模型的平均精度均值 $mAP_{0.5}$、$mAP_{0.5:0.95}$、精确率、召回率分别达到 88.2%、54.7%、86.0% 和 82.4%。与 YOLOv5s 模型相比，$mAP_{0.5}$、$mAP_{0.5:0.95}$、精确率、召回率分别提高 4.6、5.9、1.8、3.4 个百分点。改进模型在移动端部署后，识别需要避障的行人、盆以及需要作业的树苗的准确率较原模型分别提升了 15.4、4.8 和 7.0 个百分点。研究结果可为电动喷雾机器人在苗圃中的作业提供技术支撑。

关键词：深度学习；目标检测；图像识别；苗圃；YOLOv5s；部分卷积
doi：10.11975/j.issn.1002-6819.202406161
中图分类号：S147.2 **文献标志码：**A **文章编号：**1002-6819(2024)-22-0136-09

刘慧，郑鑫澎，沈跃，等. 基于改进 YOLOv5s 的苗圃内树苗及障碍物目标检测方法[J]. 农业工程学报，2024，40(22)：136-144. doi: 10.11975/j.issn.1002-6819.202406161 http://www.tcsae.org
LIU Hui, ZHENG Xinpeng, SHEN Yue, et al. Method for the target detection of seedlings and obstacles in nurseries using improved YOLOv5s[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(22): 136-144. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202406161 http://www.tcsae.org

0 引 言

近年来，随着果树和观赏类树木需求的增加，苗圃的种植面积相应扩大。广阔的种植范围意味着需要投入更多人力成本。喷雾机器人的出现减轻了相关工作人员的负担[1]。尽管喷雾机器人能够提高作业效率，但在喷洒过程中准确识别作业对象并及时避让周围物体和人员，仍是亟需解决的问题。

当前农业领域的目标检测方法，主要包括基于三维激光雷达点云数据[2-3]以及基于二维数据图像[3-4]的神经网络模型。三维激光雷达采集到的点云能够获取物体的位置信息且不受光照条件的影响，但包含信息量相对较少且设备价格昂贵。相比之下，视觉传感器成本较低，获取的图像数据包含丰富的像素信息，并且可以通过数据增强等技术降低光照对检测结果的影响。

基于图像的目标检测模型主要分为单阶段（one stage）模型和二阶段（two stage）模型[3]。二阶段模型使用了较为复杂的检测流程，因此具有较高的检测精度，但需要更长的检测时间，无法满足农业机器人的实时性需求。相比之下，单阶段模型仅需一次提取特征即可完成目标检测，因此具有更快的检测速度[5]，更适合需要实时检测的应用场景， YOLO（you only look once）系列[6-10]是单阶段模型中较为经典的代表，国内外学者也在此基础上提出了多种改进方法。蔡舒平等[11]用深度可分离卷积（depthwise separable convolution）代替 YOLOv4 中原有的标准卷积，并将网络中的残差组件（residual unit）替换为逆残差组件（inverted residual unit），使用软性非极大值抑制（soft DIoU-non-maximum suppression，Soft-DIoU-NMS）算法，在不损失模型精度的情况下实现了模型轻量化。龚惟新等[12]在 YOLOv5s 中引入 C3HB 模块和交叉注意力（criss-cross attention，CCA）模块，结合样本切分并加入负样本处理方法提升模型精度。王文瑾等[13]在 YOLOv5 中添加选择性核特征纹理迁移模块，并使用前景背景平衡损失函数抑制背景噪声干扰，从而提高了网络识别小目标和纹理模糊目标的能力。孙俊等[14]基于 YOLOv5s，用轻量级网络 MobileNetv3 作为特征提取网络，并在颈部网络中引入 RepVGG 模块，融合多分支特征提升模型的检测精度，最后使用基于动态非单调聚焦机制的损失（wise intersection over union loss，WIoU Loss）作为边界框回归损失函数，加速网络收敛并提高模型的检测准确率。YE 等[15]基于 YOLOv5 引入二阶通道注意机制，并将网络原有的快速空间金字塔池化（spatial pyramid pooling-fast，SPPF）简化为 SimSPPF，同时与目标追踪的 StrongSORT 算法协同使用，

收稿日期：2024-06-21 修订日期：2024-10-10
基金项目：国家自然科学基金项目（32171908）
作者简介：刘慧，教授，博生导师；研究方向为农业电气化与自动化、智能控制与信号处理等。Email：amity@ujs.edu.cn
※通信作者：沈跃，教授，博士导师，研究方向为无人农机与智能控制和农业机器人。Email：shen@ujs.edu.cn

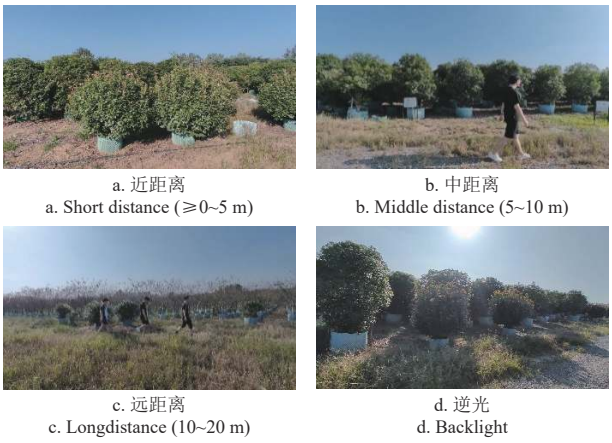
使网络能够以动态的视觉方式跟踪和计数感染松材萎蔫病的树木。ZHANG 等[16]基于 YOLOv5，用轻量级 GhostNet V2 代替 C3 模块中的瓶颈网络以减少网络参数，并将损失函数中的 CIoU 改为 SIoU 使检测框的回归更加准确，同时采用加权框融合代替非极大值抑制对检测框的输出进行处理，最后采用具有噪声的基于密度聚类算法进行主干聚类，实现果园内的目标检测。陈青等[17]基于 YOLOv7 对多尺度特征融合网络进行改进，在骨干网络中增加 160×160 的特征尺度层以增强模型对微小局部特征的识别敏感度，引入注意力机制 CBAM（convolutional block attention module）改善网络对输入图片的感兴趣目标区域的关注度，最后采用软性非极大值抑制（soft NMS）避免高密度重叠目标被一次抑制而造成漏检。

以上研究都致力于提高目标检测的速度，然而苗圃内树木密集，同时存在行人和栽培盆需要及时识别并进行避障处理。对喷雾作业而言，快速、准确地识别作业靶标至关重要，同时精准检测行人、栽培盆等并执行避障操作，是保证喷雾机器人自主可靠作业的关键。因此本研究选取行人、树苗、栽培盆作为主要检测对象，并有针对性地标注相关数据。在数据集制作过程中，根据实际作业需求，仅标注第一排作业范围内的树木和盆，由于行人均出现在第一排或第一排树木前方，故对所有行人都进行标注。YOLOv5s 网络具有较小的深度和特征图，非常适合嵌入式设备的部署[18]。为满足实时性需求，本研究选用 YOLOv5s 作为基础架构，并针对 YOLOv5 网络检测精度较低、模型运算量较大、检测速度较慢等问题进行优化。使用部分卷积代替 C3 模块中的瓶颈模块，以降低训练时的总参数量。同时使用双线性插值优化上采样过程并优化颈部网络结构，在 SPPF 模块前加入坐标注意力模块以弥补模型在位置感知方面的不足，使用解耦检测头替换原耦合检测头，在无锚框（anchor free）的基础上对结构进一步精简，以提高网络的检测精度，为电动喷雾机器人在苗圃内的喷雾作业提供可靠、实时的感知信息。

1　材料与方法

1.1　图像采集与数据集构建

本文数据集为实景拍摄自制数据集，图像采集地点为江苏省镇江市京口区城市有机质协同处理中心苗圃（32°13'N，119°30'E），图像获取设备为索尼（SONY）IMX890，共采集 2 000 张原始图像，为 JPG 格式，分辨率为 1 920×1 080 像素，包含 3 种代表性对象：树，行人，栽培盆。从远、中、近不同距离以及逆光与正常光照条件下进行图像采集，以保证数据的多样性，提升网络的目标检测能力。数据集部分图像示例如图 1，考虑到喷雾机器人的作业范围，定义≥0～5 m 为近距离，5～10 m 为中距离，10～20 m 为远距离，超过 20 m 的目标除行人外不予考虑。



a. 近距离
a. Short distance (≥0~5 m)

b. 中距离
b. Middle distance (5~10 m)

c. 远距离
c. Longdistance (10~20 m)

d. 逆光
d. Backlight

图 1　数据集图像示例
Fig.1　Example of dataset image

使用 Labelme 软件对每张图片中目标物体所在的最小外接矩形框进行人工标注，标注图像中靶标与非靶标类别与位置信息，标注结果以 JSON 格式保存。将 2 000 张图像划分为 1 500 张训练集，200 张验证集和 300 张测试集。训练集用于训练网络模型参数，验证集用于训练时调整网络超参数，防止网络过拟合，测试集不参与训练，用于最终评估模型检测效果。使用代码对标记后的类别总标签数进行统计，具体数量如表 1。

表 1　数据集不同目标数量
Table 1　Number of different targets in the dataset

种类 Class	训练集 Training set	验证集 Validation set	测试集 Test set	总计 Total
树	4 270	604	895	5 769
行人	1 388	138	162	1 688
栽培盆	4 626	616	936	6 178

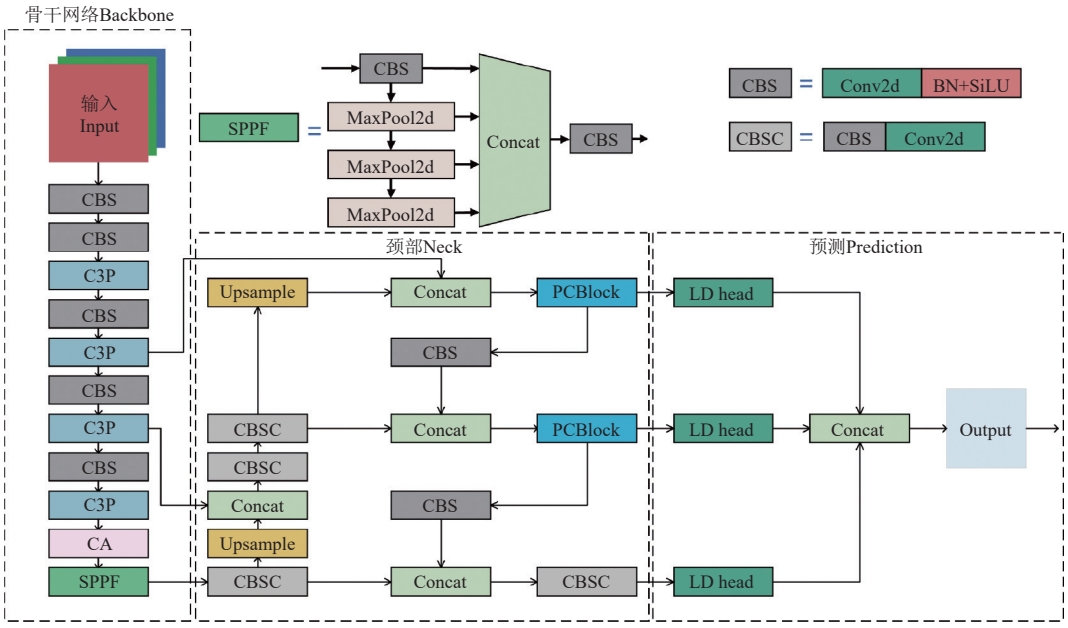
训练时对数据集进行图像增强以进一步提升数据集的多样性，包括对输入图像的随机裁剪、比例变换以及曝光调整等。

1.2　苗圃目标检测方法

YOLOv5s 的网络结构分为输入端 Input、骨干网络 Backbone、颈部网络 Neck 和预测 Prediction。主干网络 Backbone 用于提取特征。颈部网络 Neck 用于更好地融合并提取骨干网络给出的特征，从而提高网络的性能。预测端检测头分大、中、小三部分，用以分别检测不同大小的物体。YOLOv5 模型根据网络深度差异由浅至深分为 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x，

本研究选用网络深度最浅的 YOLOv5s，模型权重文件小，更易于在移动设别上部署。本研究对 YOLO 网络进行以下改进：1）将网络模型内 C3 模块的 BottleNeck 模块使用 PConv 替换用以减少计算量。2）优化颈部网络结构，改进 C3 部分和 CBS 模块，同时使用双线性插值优化上采样过程。3）使用改进的轻量化解耦检测头替换原网络的耦合检测头。4）加入 CA 注意力机制弥补网络在位置感知方面的不足。

改进后的网络结构如图 2 所示。



注：CBS 为卷积+批量归一化+ SiLU 激活函数；C3P 为引入部分卷积的 C3 模块；CA 为坐标注意力；SPPF 为空间金字塔池化模块；MaxPool2d 为最大池化操作 CBSC 为 CBS+卷积; Concat 为特征连接模块; PCBlock 为改进的特征提取模块；Upsample 为上采样模块； LDHead 为轻量化解耦检测头。
Note：CBS is convolution + batch normalization(BN) + sigmoid linear unit (SiLU) activation function; C3P is the C3（comprehensive convolution block）module introducing partial convolution; CA is coordinate attention; SPPF is the spatial pyramid pooling-fast module; MaxPool2d is the maximum pooling, CBSC is CBS+ convolution; Concat is the feature connection module; PCBlock is the improved feature extraction module; Upsample is the up-sampling module; LDHead is the light decouple detection head.

图 2 改进的 YOLOv5s 结构图
Fig.2 Structure of improved YOLOv5s

1.2.1 C3P 模块

C3 模块主要用于增加网络深度和感受野，但计算量的增长也较大。主流的轻量化网络 MobileNet[19-21]、ShuffleNet[22-23]、GhostNet[24-25] 等利用参数量较少的深度可分离卷积或组卷积提取空间特征，减少了网络模型的总计算量（floating point of operations，FLOPs），但会带来精度损失。部分卷积（partial convolution，PConv）是 FasterNet 中的一个轻量化卷积，与常规卷积相比，该卷积方式只在部分通道上执行规则卷积，提取空间特征的同时保持其他通道不变，减少了计算冗余和内存的访问。与部分卷积相比，深度可分离卷积或组卷积虽然也可以减少网络计算量，但运算符经常遭受内存访问增加的副作用。运算符进行频繁的内存访问，算子经常会受到内存访问增加的影响而导致运行效率低下[26]。因此本研究将部分卷积（partial convolution，PConv）[26] 引入 C3 模块中，取代常规卷积的瓶颈模块（BottleNeck），瓶颈模块模块结构如图 3a。将改进后的模块命名为 C3P 模块，改进后的具体结构如图 3 所示。

图 3b 为部分卷积的结构，首先对输入特征信息的一部分输入通道应用常规卷积进行特征提取，保持其余通道不变。将前c_p个或者最后c_p个连续的通道进行特征映射运算，最后直接进行 Concat 操作并进行卷积操作，在不失一般性的情况下认为输入和输出特征具有相同数量的通道。

假设输入特征图的长为h，宽为w，输入通道数为c，参与卷积的通道为c_p，卷积核尺寸为k，计算量F_{PConv}为

$$F_{PConv} = h \cdot w \cdot k^2 \cdot {c_p}^2 \quad (1)$$

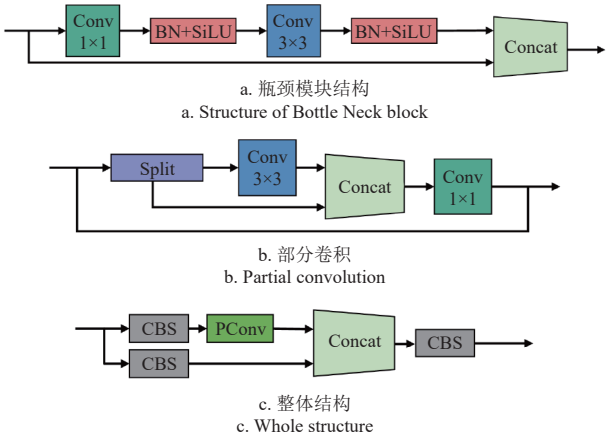


图 3 C3P 结构图
Fig.3 Structure diagram of comprehensive convolution block using partial convolution(C3P)

1.2.2 颈部网络优化

在 CBS 模块后再次加入卷积操作，并命名为 CBSC（CBS+convolution）模块。PC 模块（图 4）在数据输入前进行卷积操作，在不改变特征图空间尺寸的情况下，改变特征图的深度并引入非线性映射，有助于提高模型的表达能力。在输出时加入 2 个 CBSC 模块，在高效特征提取的基础上，加入的 BN（batch normalization）和 SiLU（sigmoid linear unit）能及时对梯度的传播进行优化。改进后的模块命名为 PCBlock，并取代用于提取大目标和中目标特征的 C3 模块。

将 YOLOv5 原上采样方法的最近邻插值替换为双线性插值，其原理如图 5 所示，尽管会增加计算量，但在特征图传递过程中能够将更多的有用信息保留，保证了

特征图的灰度连续性，有利于提高目标检测精度。

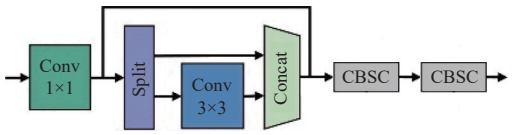
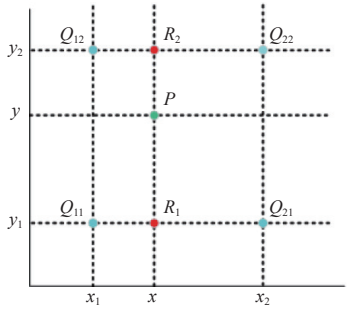


图 4　PC 模块

Fig.4　Structure of PC (partical convolution) block



注：Q_{11}、Q_{12}、Q_{21}、Q_{22} 为原像素点，x、x_1、x_2，y、y_1、y_2 为相应点的横纵坐标；P 为假定上采样后的像素点，R_1、R_2 为上采样过程的中间点。

Note: Q_{11}、Q_{12}、Q_{21}、Q_{22} represent the original pixel points with coordinates of x, x_1, x_2, y, y_1, y_2, P denotes the presumed upsampled pixel location, R_1、R_2 are intermediate points in the upsampling process.

图 5　双线性插值原理图

Fig.5　Schematic diagram of bilinear interpolation

在上采样过程中，当确定原图像素点Q_{11}、Q_{12}、Q_{21}、Q_{22}位置时，假定上采样后存在像素点P，若要确定P在函数f上的数值，首先在x方向进行线性插值，得到$f(R_1)$与$f(R_2)$，如式（2）和式（3）所示。

$$f(R_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x_2 - x}{x_2 - x_1} f(Q_{21}) \tag{2}$$

$$f(R_2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x_2 - x}{x_2 - x_1} f(Q_{22}) \tag{3}$$

随后在y方向上进行线性插值，得到$f(P)$，如式（4）所示。

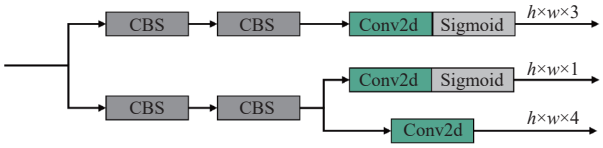
$$f(P) \approx \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y_2 - y}{y_2 - y_1} f(R_2) \tag{4}$$

将式（2）～（3）带入式（4）得到双线性插值后 P 点的坐标，即$f(P)$，如式（5）。

$$\begin{aligned} f(P) \approx & \frac{f(Q_{11})}{(x_2 - x_1)(y_2 - y_1)}(x_2 - x)(y_2 - y) + \\ & \frac{f(Q_{21})}{(x_2 - x_1)(y_2 - y_1)}(x - x_1)(y_2 - y) + \\ & \frac{f(Q_{12})}{(x_2 - x_1)(y_2 - y_1)}(x_2 - x)(y - y_1) + \\ & \frac{f(Q_{22})}{(x_2 - x_1)(y_2 - y_1)}(x - x_1)(y - y_1) \end{aligned} \tag{5}$$

1.2.3　轻量化解耦头

YOLOv5 网络使用的耦合检测头可能会导致模型在特定任务上的表现下降[27]，而解耦头可以提高网络训练的收敛速度，且在 YOLO 网络的整个目标检测流程中至关重要[27]。在分离特征提取和像素预测方面，解耦头可以将两者分离开，使网络更灵活地处理不同尺度和语义信息。而且由于其将像素级的预测作为独立的任务进行处理，也可以更好地保留细节和边缘信息，有利于预测目标种类。改进后的检测头命名为 LD（light decouple）检测头，结构如图 6 所示。



注：3、1、4 为图像通道数，h 为图像高度，w 为图像宽度，像素。

Note: 3、1、4 is the number of image channels, h is the height of the image, w is the width of the image, pixel.

图 6　轻量化解耦检测头

Fig.6　Light decouple(LD) detection head

1.2.4　CA 注意力机制

注意机制可以视为一个动态的选择过程，通过根据输入内容的重要性自适应地加权特征来实现[28]。由于移动设备计算能力有限，使用最为广泛的仍然为 SE（squeeze-and-excitation）注意力[29]，在二维全局池化的帮助下计算通道注意力，可以以较低的计算成本提供显著的性能提升。然而，SE 注意力只考虑了通道间信息的编码，而忽略了位置信息的重要性，位置信息对于视觉任务中捕获目标结构至关重要[30]。CBAM（convolutional block attention module）[31] 注意力试图降低输入张量的通道维数以利用位置信息，然后使用大尺寸卷积来计算空间注意力。然而，卷积只能捕获局部关系，无法对视觉任务中必需的长距离依赖关系进行建模，而 CA 注意力很好地解决了这个问题。

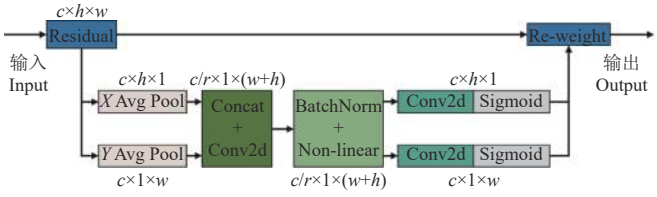
本研究在骨干网络中插入 CA 注意力机制[30]，对图像的最高维特征进行处理，以精确捕获苗圃图像中的跨通道信息，同时有效提升方向感知和位置感知能力，有助于模型准确识别和定位目标区域。CA 注意力机制的结构如图 7 所示。

CA 注意力首先对骨干网络输入的最高维度特征从水平和垂直方向一维全局池化得到特征图，获得不同方向的权重特征，如式 (6)～(7) 所示。

$$z_c^h(h) = \frac{1}{w} \sum_{0 \leqslant i \leqslant w} x_c(h, i) \tag{6}$$

$$z_c^h(w) = \frac{1}{h} \sum_{0 \leqslant j \leqslant h} x_c(j, w) \tag{7}$$

式中$z_c^h(h)$、$z_c^h(w)$分别为第c个通道输出的高度特征图和宽度特征图，$x_c(h,i)$和$x_c(j,w)$分别表示输入特征图的第c个通道沿h方向和w方向的输入。



注：c 为图像通道数，r 为下采样缩减比。

Note: c is the number of image channels, r is the downsampling reduction ratio.

图 7　CA 注意力结构图

Fig.7　Structure of coordinate attention (CA)

然后拼接获得的全局感受野特征图，并将通道数缩小r倍，随后进行批量归一化处理得到中间特征图f。其编码操作如式（8）所示。

$$f=\delta(F_1(z_c^h \oplus z_c^w)) \tag{8}$$

式中δ为非线性激活函数，F_1为卷积变换函数，$\oplus$ 表示沿空间维度的拼接操作。

最后将特征图沿原来方向再次拆分为f^h和f^w，并使用 1×1 卷积将特征图通道数扩张r倍恢复为原来的c，得到与输入特征图通道数相同的特征图F_h和F_w，并使用 Sigmoid 激活函数得到特征图垂直、水平方向的注意力权重g^h，g^w，具体表示为

$$g^h=\sigma(F_h(f^h)) \tag{9}$$

$$g^w=\sigma(F_w(f^w)) \tag{10}$$

式中 σ为 Sigmoid 激活函数，g^h和g^w为特征图垂直、水平方向的注意力权重。

最后使用注意力权重g^h和g^w与输入图像进行乘法加权计算，得到在宽度方向和高度方向带有注意力权重的特征图，计算式如下：

$$y_c(i,j)=x_c(i,j)\cdot g_c^h(i)\cdot g_c^w(j) \tag{11}$$

式中$x_c(i,j)$和$y_c(i,j)$分别为第c个通道的输入和输出。

1.3 试验平台及训练方法

硬件平台为一台配置有 Intel(R) Core(TM)i9-10900kCPU、NVIDIAQuadro RTX4000 GPU 和 64G 运行内存的计算机，使用 Python 编程语言和 PyTorch 深度学习框架训练模型。

本文将随机梯度下降法设置为优化器对网络进行优化，初始学习率设置为 0.015，权重衰退系数为 0.001，每次训练抓取样本批次大小为 16，训练轮次设置为 300。

1.4 评价指标

使用精确率、召回率、均值平均精度[32]表征模型精度；使用计算量、模型权重大小表征模型的复杂程度；使用每秒检测出的图像帧数表征模型实时检测速度。

2 结果与分析

2.1 消融试验

以 YOLOv5s 为基础分析 C3P 模块、CA 注意力机制、轻量化解耦头以及新的上采样方式和颈部结构的有效性。具体试验结果如表 2 所示。

表 2 消融试验性能对比

Table 2 Ablation experiment performance comparison

模型 Model	精确率 Precision/%	召回率 Recall/%	$mAP_{0.5}$/%	$mAP_{0.5:0.95}$/%	FLOPs/G	模型大小 Model size/MB	检测速度 Detection speed/（帧·s^{-1}）
YOLOv5s	84.2	79.0	83.6	48.8	15.9	13.6	62.1
YOLOv5s+C3P	82.0	81.8	84.8	48.7	12.6	11.0	57.8
YOLOv5s+C3P+CA	85.2	81.2	86.4	49.5	12.7	11.3	54.6
YOLOv5s+C3P+CA+LD	83.5	83.0	87.0	52.2	19.7	14.3	51.3
YOLOv5s+C3P+CA+LD+bilinear	86.4	81.0	86.7	53.3	19.7	14.3	50.9
本文	86.0	82.4	88.2	54.7	18.9	14.1	51.3

注：$mAP_{0.5}$ 指 IoU 设为 0.5 时的平均精度，$mAP_{0.5:0.95}$ 表示不同 IoU 阈值（从 0.5 到 0.95，步长 0.05）上的平均精度。1 G FLOP = 10^9 FLOPs，即每秒 10 亿次浮点运算。下同。

Note：$mAP_{0.5}$ refers the mean average precision of the intersection over union (IoU) threshold is 0.5. $mAP_{0.5:0.95}$ indicates the mean average precisio nof IoU thresholds ranging from 0.5 to 0.95 with step size of 0.05. 1 GFLOP = 10^9 FLOPs, which means 1 billion floating-point operations per second.The same below.

根据表 2 可知，与原网络相比，使用 C3P 模块替换原 C3 模块后，虽然精确率略微降低，但召回率和 $mAP_{0.5}$ 分别提高了 2.8 和 1.2 个百分点，模型计算量减少了 20.3%，权重文件大小减少了 19.1%。由于单个部分卷积的计算量远低于常规卷积，因此使用 C3P 替换的部分计算量显著减少。然而，消融试验显示，尽管轻量化改进可以降低网络复杂度，但不可避免地导致精确度下降，因为减弱了模型对部分重要特征的提取能力。在骨干网络的最高维度后引入 CA 注意力模块明显提高了精确率，且增加的计算量和权重文件大小较少。因此 CA 注意力模块有效弥补了使用 C3P 模块带来的精确率下降。为进一步提升网络性能，将原耦合头更换为改进后的轻量化解耦头，虽然部分计算量有所增加，但检测性能得到了很大的提升，在精确率略微降低的情况下，召回率、$mAP_{0.5}$、$mAP_{0.5:0.95}$ 分别提高了 2.8，0.6，2.7 个百分点。为弥补精确率的损失，采用更有利于提高目标检测精度的双线性插值法进行上采样，并将颈部网络中结构进一步优化。最终改进后网络与原网络相比，浮点运算量仅增加 3 G，但精确率、召回率、$mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别提升了 1.8，3.4，4.6 和 5.9 个百分点。单幅图像检测速度较原网络慢 3.4 ms，但整体网络检测速度仍达每秒 51.3 帧，满足实际应用中的实时性要求。

2.2 不同目标检测模型性能对比

为客观展示本文苗圃目标检测模型的可靠性和优势，将改进后的模型与原 YOLOv5s 和当前的主流单阶段目标检测算法 SSD（single shot multibox detector）[33]，Faster-RCNN[34]，YOLOv3-tiny，YOLOv7-tiny 以及 YOLOv8s 进行对比试验。试验中所涉及的模型均使用相同的数据集，各模型检测效果如表 3所示。

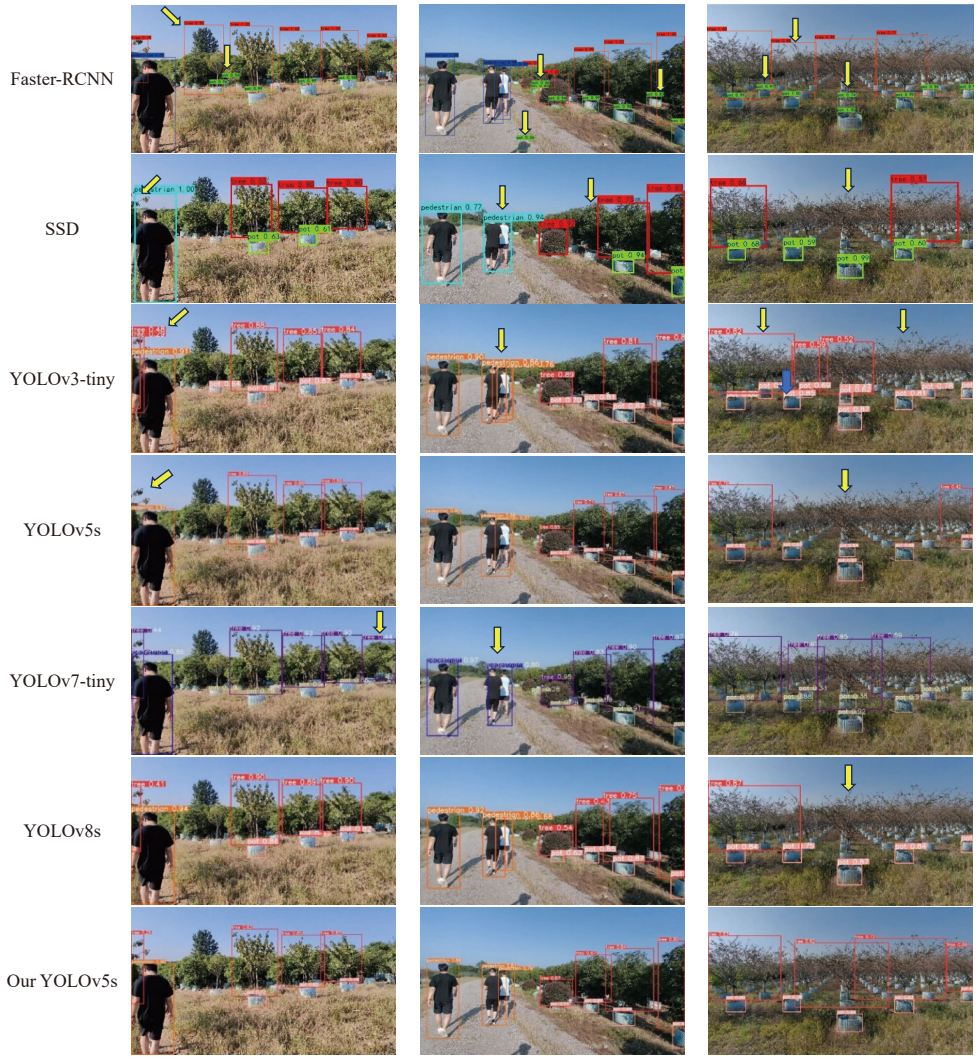
表 3 不同目标检测模型的性能对比

Table 3 Performance comparison of different target detection models

模型 Model	精确率 Precision/%	召回率 Recall/%	$mAP_{0.5}$/%	$mAP_{0.5:0.95}$/%	FLOPs/G	模型大小 Model size /MB
Faster-RCNN	79.2	47.5	70.7	37.1	134.4	108.0
SSD	85.6	54.9	75.3	32.8	34.9	91.6
YOLOv3-tiny	78.0	82.6	84.1	51.4	18.9	23.2
YOLOv5s	84.2	79.0	83.6	48.8	15.8	13.6
YOLOv7-tiny	82.4	83.3	85.7	49.8	13.0	11.6
YOLOv8s	82.4	78.9	84.9	52.6	28.4	21.4
本文	86.0	82.4	88.2	54.7	18.9	14.1

从表 3 中可以看出，YOLOv7-tiny 和 YOLOv5s 的模型尺寸最小，本文 YOLOv5s 模型大小仅比原网络增加 0.5MB，且与所有网络模型相比，改进后的网络拥有最高的精确率、$mAP_{0.5}$ 和 $mAP_{0.5:0.95}$，仅召回率略低于 YOLOv7-tiny。与 Faster-RCNN 相比，本网络仅用其 14% 的计算量和 13% 的模型大小，精确率、召回率、$mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别高 6.8，34.9，17.5 和 17.6 个百分点。与 YOLOv8s 相比，本文 YOLOv5s 精确率、召回率、$mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别高 3.6，3.5，3.3 和 2.1 个百分点，且总计算量和模型大小仅为 YOLOv8s 的 66%。为了进一步说明本文模型的有效性，部分检测结果如图 8 所示。

Faster-RCNN 没有产生漏检，但检测出过多无需识别的目标且产生了误检（图 8b），将地面影子检测为栽培盆。从图 8a 可以看出，对最左侧的不完整树木，YOLOv3-tiny 产生了误检，出现了重复的检测框，而 SSD 与原 YOLOv5s 则产生了漏检。对图片右侧远处不需要进行喷雾作业的树木，Faster-RCNN 与 YOLOv7-tiny 将其检出，并将 2 棵树误检为 1 棵树，其余网络检测效果都较为理想。从图 8b 可看出，对连续行走的 3 个行人，YOLOv3-tiny 出现了误检，将 2 个行人识别为 3 人，而 YOLOv7-tiny 出现了漏检，将 2 个行人识别为 1 人。图 8c 中本文 YOLOv5s 准确识别出了第一排可喷雾树木和第一排栽培盆，而其他网络则出现了漏检误检。



a. 场景1检测结果
a. Detection results of scene 1

b. 场景2检测结果
b. Detection results of scene 2

c. 场景3检测结果
c. Detection results of scene 3

注：图中矩形框为检测结果，黄色箭头所指为漏检以及误检处。
Note：In the figure, rectangular boxes denote detection results, yellow arrows indicate missed detections and false positives.

图 8　部分验证集检测结果对比
Fig.8　Comparison of detection results on partial validation set

2.3　移动终端模型部署与验证

为了进一步验证模型在移动端的有效性，在喷雾机器人平台上部署本文改进模型和原始 YOLOv5s 模型。该平台主要由 AI 边缘计算机 T600、导航处理器、喷雾执行末端、差速履带底盘和 Real Sense 双目相机等核心组件。在试验过程中，利用 Real Sense 双目相机捕获视觉区域信息，传输至模型进行目标类别判断并生成锚框等相关数据。将需要避障的行人与盆识别，自主判断需要作业的树苗并进行喷雾作业，硬件平台如图 9 所示。

机器人行进最大速度 1.5 m/s，双目相机采样视频分

辨率为 1 080 像素，帧率为 30 Hz，检测效果如表 4 所示。

1.差分信号接受天线 2.风送式喷雾结构 3.导航处理器 4. AI 边缘计算机 5.双目相机 6.水箱 7.差速履带底盘
1. Differential signal receiving antenna 2. Air-assisted spray structure 3. Navigation processor 4. AI edge computing device5. Binocular camera 6. Water tank 7. Differential track chassis

图 9 硬件平台
Fig.9 Hardware platform

表 4 模型在移动端的检测效果对比
Table 4 Comparison of detection performance on mobile device

模型 Model	精确率 Precision/%			检测速度 Detection speed/（帧·s^{-1}）
	树木 Tree	栽培盆 Cultivation pot	行人 Pedestrian	
YOLOv3-tiny	75.2	82.8	56.5	34
YOLOv5s	76.2	82.7	62.1	31
YOLOv7-tiny	79.7	82.4	72.1	36
YOLOv8s	66.0	78.9	72.5	25
本文	83.2	87.5	77.5	24

从表 4 中数据可以看出，在实际部署后，由于行驶中设备晃动和采集设备的性能限制，所有网络检测准确率均有所下降，但本文模型对树木、栽培盆与行人的检测准确率相较原模型分别提高了 7.0、4.8 和 15.4 个百分点，证明了本文模型在移动端部署具有较好的检测效果，能够对需要避障的行人、栽培盆以及需要作业的树冠进行准确识别。虽然检测速度并非最佳，但仍满足电动喷雾机器人在苗圃场景的作业需求。

3 结 论

为实现电动喷雾机器人在苗圃环境作业时对目标的准确识别，按照实际应用需求对 YOLOv5s 模型进行改进，主要研究结论如下：

1）使用部分卷积改进 C3（comprehensive convolution block）模块，将模型计算量降低了 20.8%，权重文件减小 19.1%；在此基础上，在骨干网络加入 CA（coordinate attention）注意力机制，并进一步改进网络颈部结构，对 CBS（ convolution + batch normalization + sigmoid linear unit activation function）模块添加卷积并使用 PC（partical convoluution）模块替换 C3 模块，同时使用双线性插值法优化训练过程中的上采样方式，最后使用改进后的轻量化解耦头替换原网络的耦合头。改进后的 YOLOv5s 相较于原网络，精确率、召回率、$mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别提高 1.8、3.4、4.6 和 5.9 个百分点。

2）与 YOLOv3-Tiny、YOLOv7-Tiny 和 YOLOv8s 模型相比，均值平均精度 $mAP_{0.5}$ 分别提高 4.1、2.5 和 3.3 个百分点，$mAP_{0.5:0.95}$ 分别提高 3.3、4.9 和 2.1 个百分点。

3）本文改进后的 YOLOv5s 模型均值平均精度 $mAP_{0.5}$ 为 88.2%，$mAP_{0.5:0.95}$ 为 54.7%，精确率为 86.0%，召回率为 82.4%，在移动端对网络部署后，本文改进模型对树木的检测精确率为 83.2%，对栽培盆的检测准确率为 87.5%，对行人的检测精确率为 77.5%，检测速度满足实时检测要求，可以为电动喷雾机器人在苗圃环境的作业提供更可靠的技术支撑。

[参 考 文 献]

[1] MARINOUDI V, SØRENSEN C G, Pearson S, et al. Robotics and labour in agriculture. A context consideration[J]. Biosystems Engineering, 2019, 184: 111-121.

[2] XU, LIU, SHEN, et al, Individual nursery trees classification and segmentation using a point cloud-based neural network with dense connection pattern[J]. Scientia Horticulturae, 2024, 328: 112945.

[3] ZOU Z, CHEN K, SHI Z, et al. Object detection in 20 years: A survey[J]. Proceedings of the IEEE, 2023, 111(3): 257-276.

[4] 刘慧，姜建滨，沈跃，等. 基于改进 DeepLab V3+ 的果园场景多类别分割方法[J]. 农业机械学报，2022，53(11)：255-261. LIU Hui, JIANG Jianbin, SHEN Yue, et al. Multi-category Segmentation of Orchard Scene Based on Improved DeepLab V3+[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(11): 255-261.

[5] DIWAN T, ANIRUDH G, TEMBHURNE J V. Object detection using YOLO: Challenges, architectural successors, datasets and applications[J]. Multimedia Tools and Applications, 2023, 82(6): 9243-9275.

[6] 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. Las Vegas, USA: IEEE, 2016: 779-788.

[7] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Piscataway, IEEE Computer Society: 2017: 7263-7271.

[8] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. (2018-04-18)[2023-08-14]. https://doi.org/10.48550/arXiv.1804.02767.

[9] BOCHKOVSKIY A, WANG C, HONG Y. YOLOv4: Optimal speed and accuracy of object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020.

[10] WAND C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE. 2023: 7464-7475.

[11] 蔡舒平，孙仲鸣，刘慧，等. 基于改进型 YOLOv4 的果园障碍物实时检测方法[J]. 农业工程学报，2021，37(2)：36-43.

CAI Shuping, SUN Zhongming, LIU Hui, et al. Real-time detection methodology for obstacles in orchards using improved YOLOv4[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(2): 36-43. (in Chinese with English abstract)

[12] 龚惟新，杨珍，李凯，等. 基于改进 YOLOv5s 的自然环境下猕猴桃花朵检测方法[J]. 农业工程学报，2023，39(6)：177-185.
GONG Weixin, YANG Zhen, LI Kai, et al. Detecting kiwi flowers in natural environments using an improved YOLOv5s[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(6): 177-185. (in Chinese with English abstract)

[13] 王文瑾，游子绎，邵历江，等. 融合超分辨率重建的 YOLOv5 松枯死木识别模型[J]. 农业工程学报，2023，39(5)：137-145.
WANG Wenjin, YOU Ziyi, SHAO Lijiang, et al. Recognition of dead pine trees using YOLOv5 by super-resolution reconstruction[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(5): 137-145. (in Chinese with English abstract)

[14] 孙俊，吴兆祺，贾忆琳，等. 基于改进 YOLOv5s 的果园环境葡萄检测[J]. 农业工程学报，2023，39(18)：192-200.
SUN Jun, WU Zhaoqi, JIA Yilin, et al. Detecting grape in an orchard using improved YOLOv5s[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(18): 192-200. (in Chinese with English abstract)

[15] YE X, PAN J, SHAO F, et al. Exploring the potential of visual tracking and counting for trees infected with pine wilt disease based on improved YOLOv5 and StrongSORT algorithm[J]. Computers and Electronics in Agriculture, 2024, 218: 108671.

[16] ZHANG J, TIAN M, YANG Z, et al. An improved target detection method based on YOLOv5 in natural orchard environments[J]. Computers and Electronics in Agriculture, 2024, 219: 108780.

[17] 陈青，殷程凯，郭自良，等. 基于改进 YOLOv7 的苹果生长状态及姿态识别[J]. 农业工程学报，2024，40(6)：258-266.
CHEN Qing, YIN Chengkai, GUO Ziliang, et al. Apple growth status and posture recognition using improved YOLOv7[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(6): 258-266. (in Chinesewith English abstract) (in Chinese with English abstract)

[18] 兰玉彬，孙斌书，张乐春，等. 基于改进 YOLOv5s 的自然场景下生姜叶片病虫害识别[J]. 农业工程学报，2024，40(1)：210-216.
LAN Yubin, SUN Binshu, ZHANG Lechun, et al. Identifying diseases and pests in ginger leaf under natural scenes using improved YOLOv5s[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(1): 210-216. (in Chinese with English abstract)

[19] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [EB/OL]. (2017-04-17) [2023-01-12]. https://arxiv.org/abs/1704.04861.

[20] SANDLER M, HOWARDA, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.. Salt Lake City: IEEE, 2018: 4510-4520.

[21] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 1314-1324.

[22] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 6848-6856.

[23] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C] //Proceedings of the European Conference on Computer Vision (ECCV). Munich,2018: 116-131.

[24] HAN K, WANG Y, TIAN Q, et al. Ghostnet: More features from cheap operations[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle, WA, USA: IEEE, 2020: 1580-1589.

[25] TANG Y, HAN K, GUO J, et al. GhostNetv2: Enhance cheap operation with long-range attention[J]. Advances in Neural Information Processing Systems, 2022, 35: 9969-9982.

[26] CHEN J, KAO S, HE H, et al. Run, Don't walk: Chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Visionand Pattern Recognition. Vancouver, BC, Canada: IEEE, 2023: 12021-12031.

[27] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021[EB/OL]. (2021-08-06) https://doi.org/10.48550/arXiv.2107.08430.

[28] GUO M H, XU T X, LIU J J, et al. Attention mechanisms in computer vision: A survey[J]. Computational Visual Media, 2022, 8(3): 331-368.

[29] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Salt Lake City: IEEE, 2018: 7132-7141.

[30] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE, 2021: 13713-13722.

[31] 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,2018: 3-19.

[32] ZAIDI S S A, ANSARI M S, ASLAM A, et al. A survey of modern deep learning based object detection models[J]. Digital Signal Processing, 2022, 126: 103514.

[33] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2015, 28:91-99.

[34] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]//Proceedings of the 14th European Conference on Computer Vision. Netherlands: ECCV, 2016: 21-37.

Method for the target detection of seedlings and obstacles in nurseries using improved YOLOv5s

LIU Hui , ZHENG Xinpeng , SHEN Yue※ , WANG Siyuan , SHEN Zhuofan , KAI Jinru

(*School of Electrical and Information Engineering*, *Jiangsu University*, *Zhenjiang* 212013, *China*)

Abstract: Nursery planting is ever expanding at present, particularly with the increasing demand for fruit and ornamental trees. Simple tools cannot fully meet the efficient work of large-scale production, due to the labor intensity and pesticide demand. Alternatively, spray robots can be expected to reduce the large number of tasks in modern agriculture. However, it is still lacking in the recognition accuracy and target identification of spray robots in nurseries. In this study, a nursery target detection was proposed using an improved YOLOv5s. Firstly, partial convolution (PConv) was introduced into the comprehensive convolution block (C3) to reduce the computational complexity, in order to improve the backbone network. A coordinate attention mechanism was added at the highest dimensional feature to enhance location awareness. Secondly, the neck structure of the network model was optimized to enhance the feature extraction of the model. At the same time, the bilinear interpolation was used for up-sampling operations during training. Finally, the original coupling detection head was replaced with the improved light decouple one, in order to further improve the detection accuracy. A nursery dataset was constructed with a total of 2 000 pictures, including three representative objects: trees, pedestrians, and cultivation pots. The images of the objects were collected from pedestrians with various postures at varying distances (far, medium, and near). The dataset included a total of 5 769 trees, 1688 pedestrians, and 6 178 pots. Only the trees and pots were marked in the first row during labeling, according to the operation requirements. Once the pedestrians are detected, they must be identified to consider the safety of pedestrians. Therefore, all pedestrians were labeled regardless of distance. The experimental results show that the detection of the model was improved differently via the C3 module and coordinate attention mechanism, as well as the optimized neck network structure and up-sampling mode using a lightweight decoupling head. Finally, the average $mAP_{0.5}$, $mAP_{0.5:0.95}$, accuracy, and recall reached 88.2%, 54.7%, 86.0%, and 82.4%, respectively. The size of the improved network model was 14.1 MB, the average detection speed of a single image was 19.5 ms, and the average frame rate was 51.3 frames. The $mAP_{0.5}$, $mAP_{0.5:0.95}$, accuracy, and recall increased by 4.6, 5.9, 1.8 and 3.4 percentage points, respectively, compared with the original YOLOv5s. The improved network had the highest accuracy, $mAP_{0.5}$ and $mAP_{0.5:0.95}$, compared with the current mainstream single-stage target detection YOLOv3-tiny, YOLOv7-tiny, and the latest YOLOv8s model. The experimental results show that the improved model performed better in accurately and rapidly identifying the pedestrians, pots, and trees in a complex environment. The research findings can also provide technical support to the operational activities of electric spraying robots in nurseries.

Keywords: deep learning; object detection; image recognition; nursery; YOLOv5s; partial convolution