doi:10.6041/j.issn.1000-1298.2026.03.019

基于改进 YOLO v5n 的移栽机栽植部件辣椒苗识别方法

张斯源[1]　苑严伟[2,3]　崔　巍[2,3]　朱凤武[1]　吕程序[2,3]　张学东[2,3]

（1. 吉林农业大学工程技术学院，长春 130118；2. 中国农业机械化科学研究院集团有限公司，北京 100083；
3. 农业装备技术全国重点实验室，北京 100083）

摘要：全自动移栽机在实际工作过程中经常遇到堵苗、漏苗和幼苗栽植状态异常等问题，实时监测移栽机栽植部件中的幼苗情况，是提高移栽机工作效率与移栽质量的关键。因此，本文提出了一种基于 YOLO v5n 的轻量化识别方法，用于对栽植部件中的辣椒苗进行精确目标检测。首先在顺光、逆光光照情况下，使用摄像头采集了单株和多株的移栽机栽植部件辣椒苗图像，构建辣椒苗数据集；其次在 YOLO v5n 神经网络架构的基础上，使用 Ghost 卷积替换普通卷积，并插入改进后的 FastGhost 模块和 SimAMGhost 模块，有效降低模型的运算量和计算延迟，提高检测速度；引入 EMA 注意力机制，提高对重要细节信息的注意程度，改善模型对高度重叠的多株辣椒苗图像识别效果，解决了辣椒苗的部分多检和漏检问题；最终使用 Shape－IoU 损失函数替换 CIoU 损失函数，消除边界框自身形状对边界框回归的影响，提高边界框回归准确度。实验结果表明，与 YOLO v5n 相比，改进后的 YOLO v5n－GE 模型的检测平均精度均值为 95.3%，比原模型提高 0.3 个百分点，模型参数量和计算量分别缩小 52.5% 和 51.2%，检测速度提升 12.2%。与当前 YOLO 系列主流模型相比，YOLO v5n－GE 能够在大幅度减少参数量和运算量的情况下，保持较高的检测精度，证明了改进算法的有效性，可为硬件资源有限的移栽机栽植部件中的辣椒苗识别工作提供技术支持。

关键词：全自动移栽机；辣椒苗识别；YOLO v5n；SimAMGhost 模块；EMA 注意力机制

中图分类号：S24；TP389.1　　**文献标识码**：A　　**文章编号**：1000-1298(2026)03-0196-10　　**OSID**：

Pepper Seedling Recognition in Transplanting Machine Components Based on Improved YOLO v5n

ZHANG Siyuan[1]　YUAN Yanwei[2,3]　CUI Wei[2,3]　ZHU Fengwu[1]　LÜ Chengxu[2,3]　ZHANG Xuedong[2,3]

(1. *School of Engineering and Technology*, *Jilin Agricultural University*, *Changchun* 130118, *China*
2. *Chinese Academy of Agricultural Mechanization Sciences Group Co.*, *Ltd.*, *Beijing* 100083, *China*
3. *State Key Laboratory of Agricultural Equipment Technology*, *Beijing* 100083, *China*)

Abstract: Aiming to address key operational issues in automatic transplanting machines, such as missed planting, seedling blockages and abnormal seedling planting state, an optimized lightweight detection model, YOLO v5n－GE, was proposed for real-time monitoring of seedling conditions within transplanting equipment. The research began by collecting images of both single and multiple pepper seedlings under varying lighting conditions (front and backlighting) using a camera. To reduce computational load and latency, traditional convolutions were replaced with Ghost convolutions on the basis of YOLO v5n model, and the main feature extraction modules were substituted with improved FastGhost and SimAMGhost modules. EMA attention mechanism was applied to enhance the network's focus on important detail information, effectively improving the model's recognition results for highly overlapping pepper seedlings, reducing sensitivity to occlusions, and increasing recognition accuracy. Additionally, Shape－IoU loss was used to replace CIoU loss, addressing the influence of the bounding box shape on bounding box regression and improving bounding box regression accuracy. Experimental results on the self-built dataset demonstrated that the improved YOLO v5n－GE model achieved an mAP of 95.3%, representing a 0.3

收稿日期：2024－10－29　修回日期：2024－12－12

基金项目：国家重点研发计划项目(2023YFD2001203)

作者简介：张斯源(2001—)，男，硕士生，主要从事农业电气化与自动化研究，E-mail：zsy01227@163.com

通信作者：苑严伟(1980—)，男，研究员，博士，主要从事农业智能化技术研究，E-mail：yyw215@163.com

percentage points improvement over the original model. The model's parameter count and computational load were reduced by 52.5% and 51.2%, respectively, detection speed was increased by 12.2%. These enhancements enabled efficient, real-time detection of pepper seedlings while maintaining high accuracy, demonstrating the improved algorithm's effectiveness. The research result can provide technical support for seedling recognition in transplanting machine components with limited hardware resources.

Key words: automatic transplanting machine; pepper seedling recognition; YOLO v5n; SimAMGhost module; EMA attention mechanism

0　引言

全自动移栽机是未来移栽种植机械的发展方向[1]，但在全自动移栽机的工作过程中经常出现漏苗、堵苗和栽植状态异常等问题，显著影响了移栽机的工作效率和移栽质量，传统的光电检测方法在识别移栽机作业中的错误栽植状态方面存在局限性。因此，采用图像识别技术对移栽机栽植部件的作业状态进行监测和精确识别，具有提升全自动移栽机作业效率的潜力，逐步成为推动农业装备智能化、高效化的重要方向。

随着计算机技术和深度学习技术的深入研究，深度卷积神经网络在农业领域体现出较高的优越性，这种方法可以在背景复杂的图像信息中提取目标特征，在处理复杂任务时有更好的应用效果[2-4]。目前主流的目标检测识别算法有：Faster R-CNN 系列[5-7]、YOLO 系列[8-11]、SSD[12] 等。Faster R-CNN 是两阶段检测的目标检测算法。它用一个区域提示器生成候选的物体区域，再通过神经网络对其进行分类，这种算法虽然精确度很高，但生成候选区域步骤需占用大量资源，检测时间较长。SSD 和 YOLO 算法是单阶段检测的目标检测算法，SSD 的核心思想是将图像分为多个区域，并在每个区域内进行物体检测，但是在复杂背景条件下，SSD 算法的检测精度和定位不够精准。YOLO 算法通过单一的神经网络来检测多个物体，并且仅在 1 次前向传播中完成目标检测、分类和定位 3 个任务。相比之下 YOLO 系列算法在速度和简单性上具有较高的优越性，便于部署在全自动移栽机这类移动设备上。

由于农业环境下对作物检测的精确度和实时性通常有较高需求，国内外学者对 YOLO 系列算法进行了多种方法改进。马宏兴等[13] 使用带有 Transformer 模块的主干网络 BoTNet 和坐标注意力机制对 YOLO v5s 进行轻量化改进，在保持对植物识别精度的同时提高了模型检测速度；冀汶莉等[14] 结合了 PP-LCNet 和 Ghost 卷积降低 YOLO v5s 模型的参数量和计算量，实现对复杂农田环境中的杂草识别；YAN 等[15] 设计了更为轻量的 BottleNeckCSP-2 模块，大幅提升了对被遮挡的苹果目标的实时检测速度，但模型参数量和权重大小依然较大；ZHANG 等[16] 引入 GhostBottleneck 模块和双注意力机制，针对自然背景环境下的黄花菜目标检测，将网络模型参数量和运算量分别压缩 63.58% 和 68.95%，但模型的识别精度仍有优化空间。

为实现全自动移栽机的高速移栽与作业状态实时检测，需要在保持识别精度的情况下对模型进行有效轻量化，减少对计算资源的占用。针对以上问题，本文提出一种改进的 YOLO v5n 算法，以辣椒苗为主要识别对象，在网络结构中引入轻量化 Ghost 卷积模块，在不影响准确率的情况下，降低模型运算量，加快识别速度；引入 EMA 注意力机制，帮助网络模型检测边界模糊的重叠遮挡目标；将损失函数 CIoU Loss 替换为 Shape-IoU Loss，以提高识别精度。

1　材料与方法

1.1　数据集制作

本文所用的移栽机栽植部件为鸭嘴式栽植机构，其主要功能是将辣椒苗从育苗盘转移至田地中，为验证所提识别方法的有效性，设计并制作了针对该机构的实验装置，如图 1 所示。装置由可调速直流电机作为主动力源，通过齿轮减速箱驱动鸭嘴完成圆周运动，减速比为 4∶1，鸭嘴的开关动作则通过凸轮结构实现。

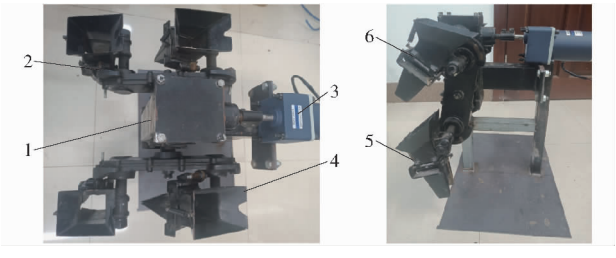

图 1　鸭嘴式栽植机构实验装置

Fig. 1　Experimental apparatus of duck-billed transplanting mechanism

1. 齿轮减速箱　2. 凸轮结构　3. 直流电机　4. 鸭嘴　5. 栽植位　6. 接苗位

辣椒苗数据集的采集地点位于吉林农业大学农业机械化实验室的田间实验地，采集环境为温度适

宜的露天环境。为模拟移栽机的实际工作状态，鸭嘴式栽植机构由可调速电机驱动并处于运动状态，摄像头根据机构的转速以固定频率进行拍摄，捕捉辣椒苗进入鸭嘴区域的瞬时图像。拍摄设备为 RealmeX7Pro 型手机，搭载索尼 IMX686 型传感器。为确保采集图像质量，摄像头固定在距离栽植部件鸭嘴顶部垂直向上约 20 cm 的位置，该安装高度下的相机视角能够完整覆盖鸭嘴部件的工作区域，同时避免遮挡及运动模糊影响，采集场景如图 2 所示。



图 2 数据集采集场景

Fig. 2 Dataset collection scenarios

1. 实验装置 2. 摄像头位置

考虑到移栽机在实际工作中可能遇到的复杂环境，提升数据样本的多样性，在顺光、逆光 2 种光照条件下，拍摄了鸭嘴式栽植机构中包含单株、双株和多株辣椒苗的图像。共计拍摄原始图像 1 763 幅，经过人工筛选剔除冗余图像后，最终构建了包含 1 534 幅有效图像的数据集。采集图像原始分辨率为 4 608 像素 ×3 456 像素，在输入深度学习模型前进行了适当缩放，以适配模型的输入尺寸，不同作业情况下的辣椒苗图像如图 3 所示。

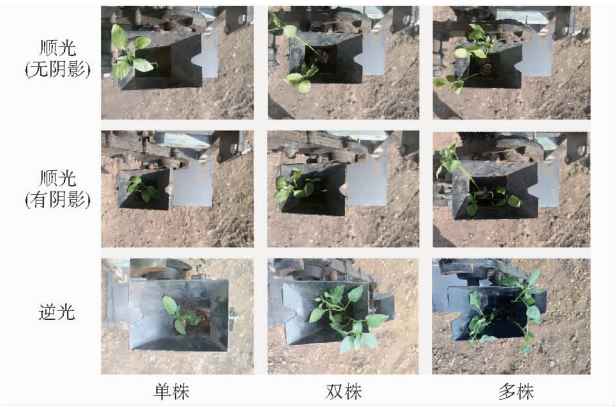

图 3 采集的图像数据集例图

Fig. 3 Acquired image dataset

1.2 数据增强

为了丰富数据集多样性，保持数据均衡，并增强模型的鲁棒性，对辣椒苗数据集进行数据增强，扩充样本数量，使得目标检测模型的识别效果更加显著。

在 Pytorch 框架下结合 OpenCV，在保留原标注信息的情况下，对辣椒苗图像进行随机平移、随机旋转、亮度调整、部分裁剪、随机遮挡和添加高斯噪声等不同的图像变换。最终得到辣椒苗图像数据集共 6 350 幅，数据增强的具体方式如图 4 所示。

使用 LabelImg 工具软件对数据集中的辣椒苗进行精确矩形框标注，并保存为 YOLO 系列模型训练所需要的 YOLO 格式。按照比例 7∶3对数据集进行划分，分为训练集 4 340 幅，验证集 2 010 幅。

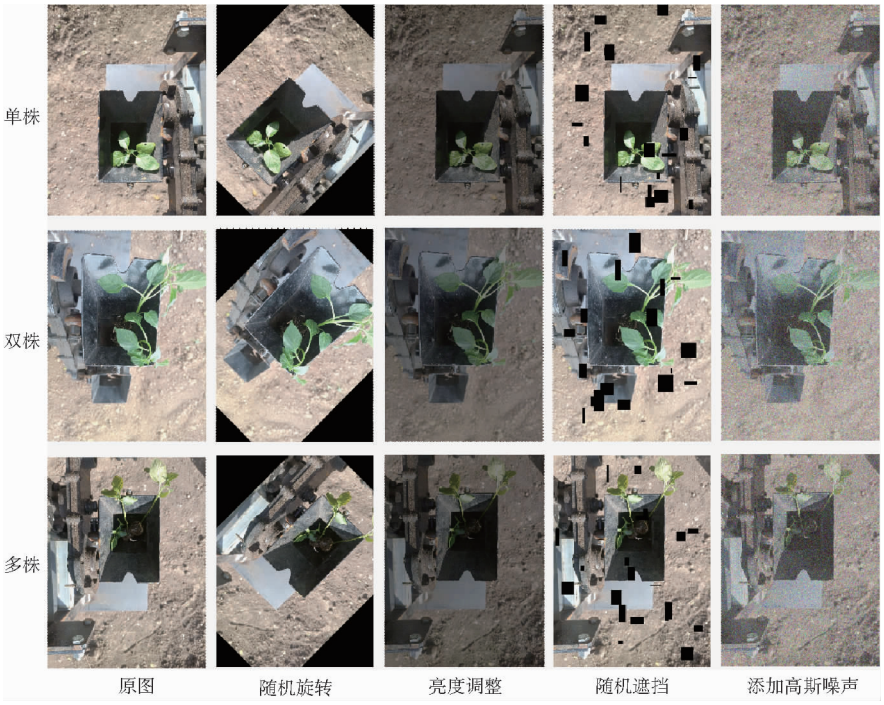

图 4 处理后的图像数据集例图

Fig. 4 Processed image dataset

2　YOLO v5n 改进模型设计

2.1　YOLO v5 网络模型

考虑到算法在移动端的部署需求，选择了 YOLO v5n 作为基础目标检测模型。与 YOLO 系列的最新模型 YOLO v9、YOLO v10 和 YOLO 11 相比，YOLO v5n 的优势在于其更小的模型参数和权重，从而实现更快的检测速度。这一特性使 YOLO v5n 特别适合于资源受限的移动端环境，确保在保证检测精度的同时，满足实时处理的需求。

YOLO v5 网络模型主要由骨干网络（Backbone）、颈部网络（Neck）和目标检测头（Head）3 部分组成，骨干网络部分由 4 个 C3 模块和 1 个空间金字塔池化结构 SPPF 构成，其中 C3 模块贯穿整个网络模型，主要作用是增加网络的深度和感受野，提高特征提取的能力；颈部网络采用的是 FPN + PAN 的特征结构，FPN（Feature pyramid network）通过上采样和特征图融合操作将高层特征图的语义信息传递下来[17]，PAN（Path aggregation network）通过卷积和特征图融合操作将低层特征图的定位信息传递上去，这样形成的特征金字塔便结合了不同层次的丰富信息[18]；目标检测头主要包括边界框生成、分类和回归 3 部分。YOLO v5n 的损失函数包括分类损失函数、定位损失函数和置信度损失函数，主要使用 CIoU 作为损失函数，实现边界框的位置回归。

2.2　Ghost 轻量化网络结构

移栽机在工作过程中的速度较快，且对辣椒苗检测实时性需求较高，因此需要目标检测模型具有更快的检测速度和更低的延迟。针对以上问题，本文选择用 Ghost 卷积模块代替普通卷积作为模型网络结构的主干，帮助目标检测模型获得更低的运算量和更高的帧速率，从而降低模型的计算延迟，提高检测速度[19]。

Ghost 卷积是 GhostNet 网络中提出的轻量化卷积模块，它的工作原理如图 5a 所示，图中 H 和 W 分别表示特征矩阵的高和宽，C 表示特征矩阵的通道数。首先从输入特征图中使用标准卷积生成一组基础特征图；然后通过简单的线性变换即不同大小的卷积核操作从基础特征图中生成多个 Ghost 特征图，这些操作具有较低的计算复杂度；最后，将生成的 Ghost 特征图和基础特征图组合在一起，组合后的特征图不仅有效地保留了输入信息，还保存了一定数量的 Ghost 特征。Ghost 卷积模块的网络结构如图 5b 所示。

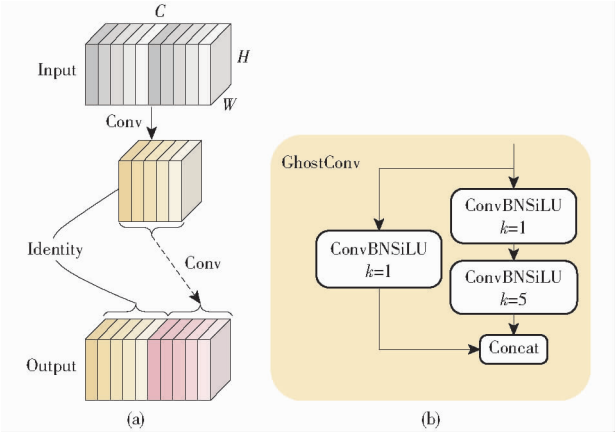

图 5　Ghost 卷积模块工作原理和结构图

Fig. 5　GhostConv module mechanism and structure

在 YOLO v5 模型结构中，C3 模块是主要的特征提取层，如图 6a 所示，从 C3 模块的结构中可以看

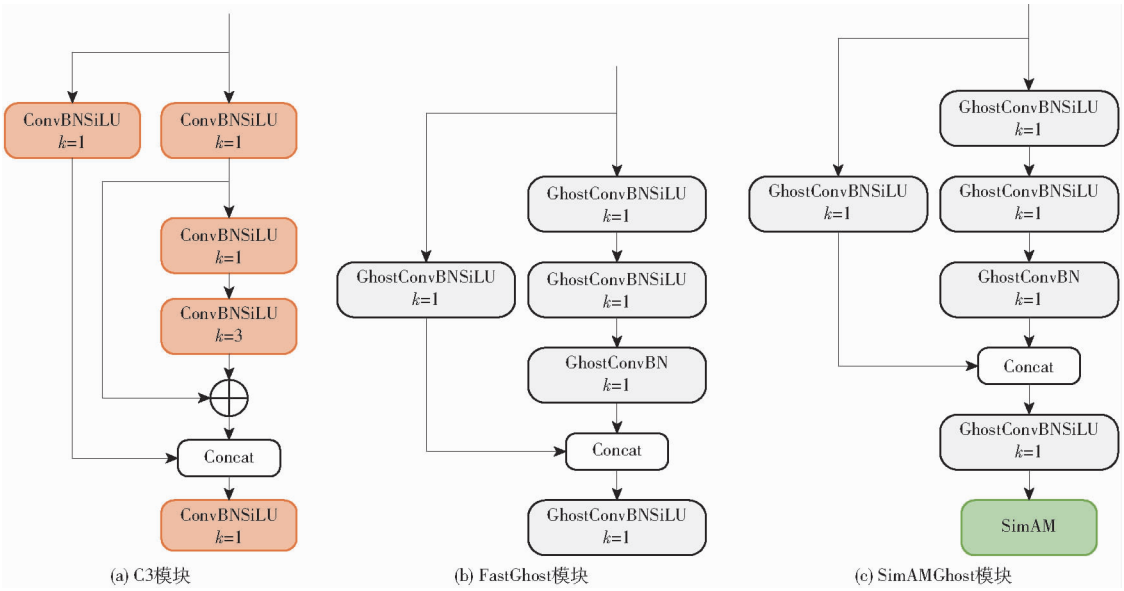

图 6　C3 模块、FastGhost 模块和改进的 SimAMGhost 模块

Fig. 6　C3 module, FastGhost module and improved SimAMGhost module

出,该模块与 Ghost 卷积模块的工作原理相同,都是通过部分保留和部分卷积的方式得到多个尺度的特征输出。基于这个思想,对 C3 模块做了更深层次的轻量化改进,将模块中全部普通卷积都替换为 Ghost 卷积,构成新的 FastGhost 模块,如图 6b 所示。

假设输入特征图与输出特征图的高、宽和通道数全部一致,且只考虑卷积的乘法,输入特征矩阵 $\boldsymbol{X}\in\mathbf{R}^{H\times W\times c_{in}}$,$c_{in}$ 为特征矩阵的输入通道数,普通卷积计算量 $FLOPs$ 计算式为

$$FLOPs = k^2 c_{in} c_{out} HW \tag{1}$$

式中 k——卷积核大小

c_{out}——特征矩阵的输出通道数

如果忽略输入特征矩阵的高和宽,Ghost 卷积模块的计算量 $FLOPs_{ghost}$ 计算式为

$$FLOPs_{ghost} = k^2 c_{in}\frac{c_{out}}{r} + s^2\frac{c_{out}}{r}(r-1) \tag{2}$$

式中 r——输入特征图中使用标准卷积生成基础特征图的比例

s——生成 Ghost 特征的线性操作卷积核大小

两者的计算量比值 r_s 为

$$r_s = \frac{k^2 c_{in} c_{out}}{k^2 c_{in}\frac{c_{out}}{r} + s^2\frac{c_{out}}{r}(r-1)} = \frac{k^2 c_{in} r}{k^2 c_{in} + s^2(r-1)} \approx \frac{c_{in} r}{c_{in} + r - 1} \approx r \tag{3}$$

在本文提出的模型中,s 与 k 相同,$r \ll c_{in}$,且 Ghost 卷积中进行线性操作的通道数是总通道数的 1/2,根据计算量比值,最终 Ghost 卷积的计算量约为普通卷积的 1/2。

2.3 注意力机制的引入

2.3.1 SimAMGhost 模块

由于 Ghost 轻量化网络结构会导致一定程度的精度下降。为了在不影响轻量化的情况下进一步提升模型的识别精度,将 SimAM(Simplified attention module)注意力机制[20]集成到了 FastGhost 模块中,构成新的 SimAMGhost 模块,如图 6c 所示。

SimAM 是一种轻量且高效的模块,它提取了特征图的三维权重,让每个像素都拥有一个唯一的权重,能够捕捉特征图中重要的空间信息,增强模型对目标的辨识能力,如图 7 所示。

在赋予像素特征权重后,利用能量函数计算最小化能量,提升有用特征的激活值,同时抑制无关特征。这种思想是根据 WEBB 等[21]在视觉神经科学领域提出的空间抑制现象。潜在能量函数 e_t 的计算公式为

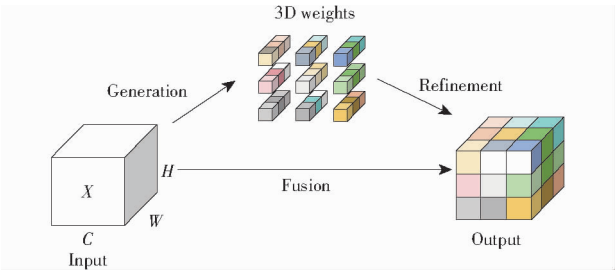

图 7 SimAM 注意力模块工作原理

Fig. 7 SimAM attention module mechanism

$$e_t = (y_t - \hat{t})^2 + \frac{1}{M-1}\sum_{i=1}^{M-1}(y_o - \hat{x}_i)^2 \tag{4}$$

式中 M——该通道中神经元的总数量

$(y_t - \hat{t})^2$——计算目标神经元 t 的输出 $\hat{t}$ 与其标签 y_t 之间的误差

$\frac{1}{M-1}\sum_{i=1}^{M-1}(y_o - \hat{x}_i)^2$——计算其他神经元的输出 $\hat{x}_i$ 与标签 y_o 之间的平均误差

通过对该能量函数引入正则化并化简,SimAM 模块可以根据神经元的最小能量计算出每个神经元的重要性,最终使用缩放系数对原始特征图进行权重重构,实现特征细化,提升其重要信息的表达。

2.3.2 EMA 注意力机制

OUYANG 等[22]提出的 EMA(Efficient multi-scale attention)注意力机制是基于 CA 注意力机制的改进方法,它是一种跨空间的多尺度并行子网络,通过将部分通道维度重新整形为批次维度,避免了降维操作,融合了空间和位置上的有效信息,且所需的参数量并不高。

因此,本文在 YOLO v5n 网络模型的骨干网络上,引入了 EMA 注意力机制,使得模型能够更加关注有用的特征信息,提高模型的性能和精度。EMA 的注意力机制结构如图 8 所示,图中 g 表示通道分组数。

并行处理结构(Parallel structure)。首先,将输入特征矩阵 $\boldsymbol{X}$ 划分为跨通道维度方向的 g 个子特征,用于学习不同的语义,然后将张量输入到 3 个并行的分支中,其中 2 个分支分别通过水平方向和竖直方向的一维全局平均池化将 $\boldsymbol{X}$ 分解为 2 个并行的一维特征编码向量,得到关于宽度和高度的特征表示,保留位置信息,最后将 2 个一维特征编码向量拼接起来,经过 1×1 卷积核捕捉跨通道交互信息,提高特征表示的质量。

跨空间学习(Cross spatial learning)。在 1×1 和 3×3 分支中,将之前得到的跨空间交互信息的输出向量,通过二维全局平均池化编码,得到全局空间信息,对全局平均池化的结果应用二维高斯分布的

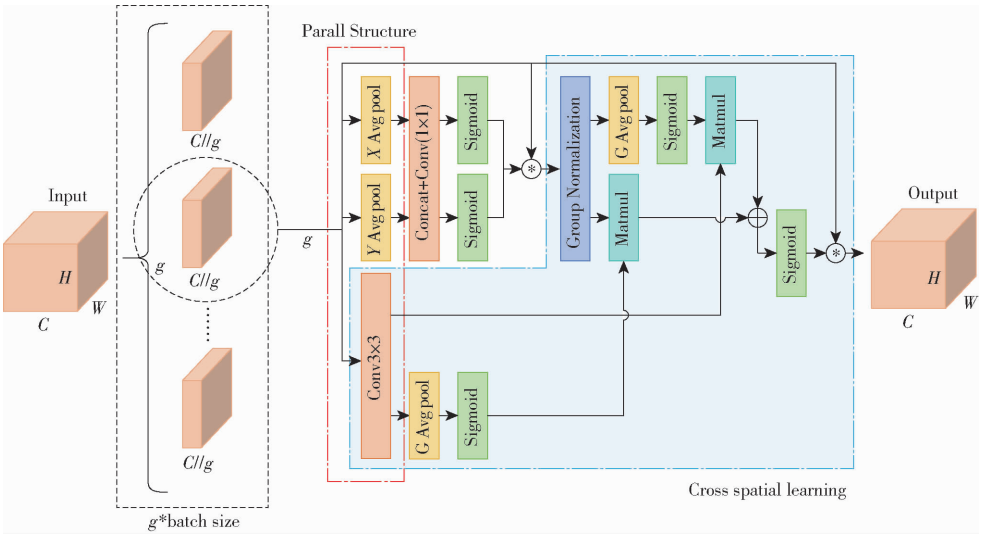

图 8　EMA 注意力机制

Fig. 8　Efficient multi-scale attention module

Softmax 激活函数，突出在特征表示中的通道权重，之后将输出矩阵与另一分支的并行处理输出矩阵进行点积运算，得到空间注意力特征矩阵。

最后将 2 个分支的空间注意力特征矩阵聚合到一起，经过 Sigmoid 函数生成空间注意力权重值，对特征图进行重新加权，得到具有像素级全局上下文关系的输出特征图。

2.3.3　Shape－IoU 损失函数

作为目标检测中的标准性能指标，交并比（Intersection over union，IoU）可以有效衡量预测边界框和真实边界框之间的对齐程度，从而判断模型预测的准确性。

YOLO v5n 网络模型使用了 CIoU 作为损失函数，与其他 IoU 损失函数相比，CIoU 考虑了重叠面积、中心点距离、纵横比，并基于 DIoU 提升回归精确度[23]。但是 CIoU 等损失函数计算相对复杂，且只考虑了利用预测框的相对位置和形状来计算损失，忽略了边界框的形状和尺度等固有属性对预测框回归的影响。ZHANG 等[24]针对这一问题，提出了 Shape－IoU 损失函数。Shape－IoU 的原理如图 9 所示，其计算公式为

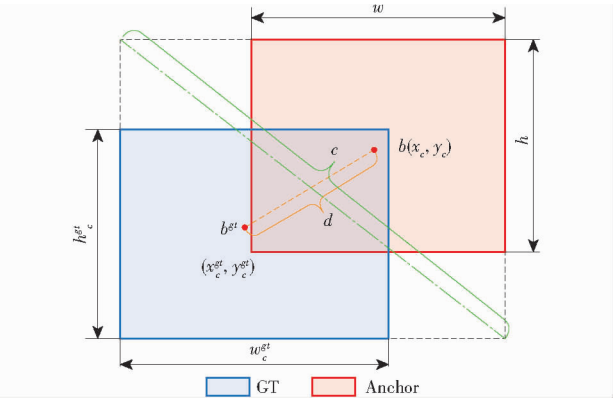

图 9　Shape－IoU 损失函数

Fig. 9　Shape－IoU loss

$$L_{\text{Shape-IoU}} = 1 - IoU + d^{\text{shape}} + 0.5\Omega^{\text{shape}} \tag{5}$$

其中

$$d^{\text{shape}} = h_h \frac{(x_c - x_c^{gt})^2}{c^2} + w_w \frac{(y_c - y_c^{gt})^2}{c^2} \tag{6}$$

$$\Omega^{\text{shape}} = \sum_{t=w,h} (1 - \mathrm{e}^{-\omega_t})^{\theta} \tag{7}$$

$$w_w = \frac{2 (w^{gt})^u}{(w^{gt})^u + (h^{gt})^u} \tag{8}$$

$$h_h = \frac{2 (h^{gt})^u}{(w^{gt})^u + (h^{gt})^u} \tag{9}$$

$$\begin{cases} \omega_w = h_h \dfrac{|w - w^{gt}|}{\max(w, w^{gt})} \\ \omega_h = w_w \dfrac{|h - h^{gt}|}{\max(h, h^{gt})} \end{cases} \tag{10}$$

式中　$L_{\text{Shape-IoU}}$——Shape－IoU 损失函数

d^{shape}——距离损失

Ω^{shape}——形状损失

θ——形状损失的关注程度系数，取 4

u——数据集中的目标比例因子

w_w、h_h——水平、垂直方向权重系数

w^{gt}、h^{gt}——真实框的宽度和长度

x_c^{gt}、y_c^{gt}——真实框中心点的位置

x_c、y_c——预测框中心点的位置

ω_w、ω_h——宽度、高度损失权重

Shape－IoU 损失函数的使用可以帮助模型对辣椒苗的预测框回归更加精确，同样有助于对栽植部件中高重叠度情况下的辣椒苗进行识别。

2.4　改进后的 YOLO v5n－GE 网络模型

为实现复杂环境下移栽机栽植部件的辣椒苗识别，对 YOLO v5n 网络模型进行以下改进：在主干网络中使用 Ghost 卷积替换 3×3 普通卷积，用改进后的 FastGhost 模块和 SimAMGhost 模块替换普通 C3

模块,降低计算延迟,提高检测速度,实现网络轻量化;同时为了弥补精度不足,引入 EMA 注意力机制,使得模型更加关注有用的细节信息,提高识别精度;使用 Shape - IoU 损失函数替换 CIoU 损失函数,提高预测框回归的准确率。改进后的模型称为 YOLO v5n - GE 模型,如图 10 所示。

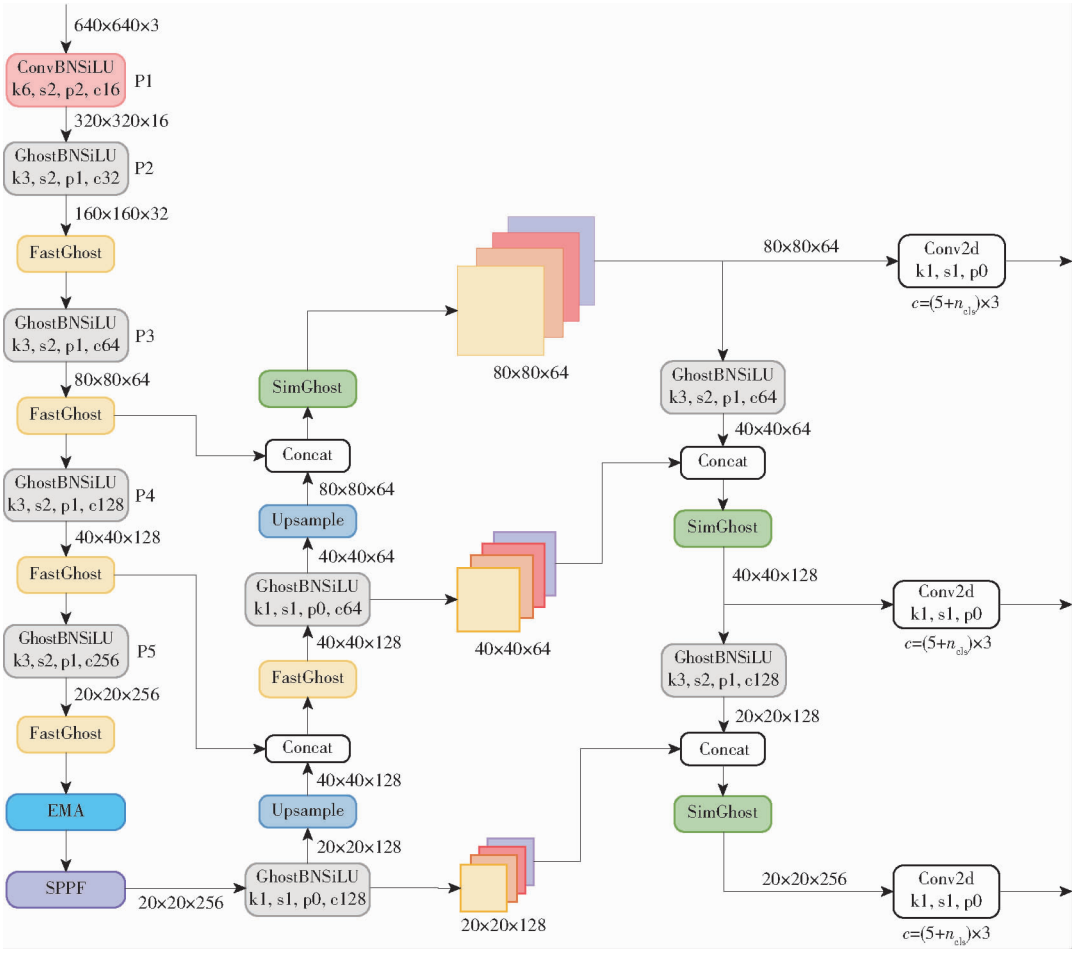

图 10 YOLO v5n - GE 网络模型

Fig. 10 YOLO v5n - GE network model

3 实验结果与分析

3.1 实验环境与参数设置

本实验使用的操作系统是 Windows 11 版本,搭载 Intel Core i7 - 13700KF CPU,主频 3.40 GHz,GPU 采用 NVIDIA RTX4070,CUDA 版本为 12.2,深度学习框架为 Pytorch 2.0.0,编译环境为 Python 3.10.9。

模型训练时的批量尺寸(Batch size)为 16,训练 200 个迭代周期,输入图像尺寸为 640 像素 ×640 像素,使用自适应学习率(Adaptive moment estimation, Adam)优化器来优化模型。

3.2 实验评价指标

为验证改进方法有效性,主要从识别准确度和轻量化效果进行评估。对于识别精确度主要采用精确率(Precision)、召回率(Recall)、平均精度均值(mAP)来评估模型性能;轻量化效果方面选取指标为:模型参数量(Parameters)、浮点运算量(GFLOPS)和台式机上的运行帧速率(FPS)。

3.3 实验结果分析

3.3.1 Ghost 轻量化网络结构改进效果验证方法

为了验证 Ghost 卷积模块替换普通卷积的改进有效性,在辣椒苗数据集上进行了对比实验,实验结果如表 1 所示,表 1 中模型 2 是在原 YOLO v5 模型的基础上替换了主干网络的普通卷积,模型 3、4 是在模型 2 的基础上替换 C3 模块。

表 1 不同 Ghost 模块的实验结果

Tab. 1 Comparision experiment results of different Ghost modules

序号	模型	平均精度均值/%	参数量	浮点运算量	帧速率/(f·s^{-1})
1	YOLO v5n	95.0	1 672 470	4.1×10^9	279.87
2	YOLO v5n Ghost	94.5	1 374 790	3.2×10^9	294.96
3	+ FastGhost	94.3	759 994	1.9×10^9	312.60
4	+ SimAM Ghost	94.6	742 234	1.9×10^9	322.63

由表 1 可知,与原 YOLO v5n 模型相比,模型 2

的参数量下降 17.7%，运算量下降 21.9%，计算速度提高 5%，但是平均精度均值下降 0.5 个百分点，这主要是因为 3×3 普通卷积具有一定的感受野，能够捕捉局部的空间关系和上下文信息，而 Ghost 卷积由于减少了独立的卷积核数量，其感受野受到一定的限制。

模型 3 为用 FastGhost 模块替换 C3 模块，相比模型 2，参数量会进一步下降 44.7%，运算量下降 40.6%，计算速度提高 5.9%，平均精度均值只降低了 0.2 个百分点。当整个 C3 模块中的所有普通卷积都被替换为 Ghost 卷积时，尽管单个卷积的表达能力较弱，但整个模块通过多个 Ghost 卷积层的叠加，逐渐恢复部分丢失的特征信息，尤其是深层的 Ghost 卷积可以从前面的层中补充丢失的信息，从而减小了精度的下降幅度。

从模型 4 的结果得出，设计的 SimAMGhost 模块在不影响模型的参数量和计算量的情况下，将模型的平均精度均值在模型 3 的基础上进一步提高了 0.3 个百分点。

综上所述，以 Ghost 卷积模块为核心的轻量化网络结构十分有效，它可以大量降低模型的复杂程度，并削减冗余的复杂计算，提高计算速度，适用于对识别速度有较高需求的移栽机辣椒苗检测任务当中。

3.3.2　EMA 注意力机制改进效果验证方法

为了验证 EMA 注意力机制的有效性，在模型的相同位置用其他不同的注意力机制替换 EMA，包含 SE、CA 和 CBAM 等，进行了对比实验。实验结果表明，在辣椒苗检测数据集中，EMA 注意力机制的作用相比其他的注意力机制取得了更好的识别结果，并且参数量增加极少。

为了更直观地观察注意力机制的效果，在算法中加入了热力图工具，将模型的注意力可视化，以展示模型在处理输入图像时，各个像素区域的注意力权重分布。注意力可视化的效果如图 11 所示。

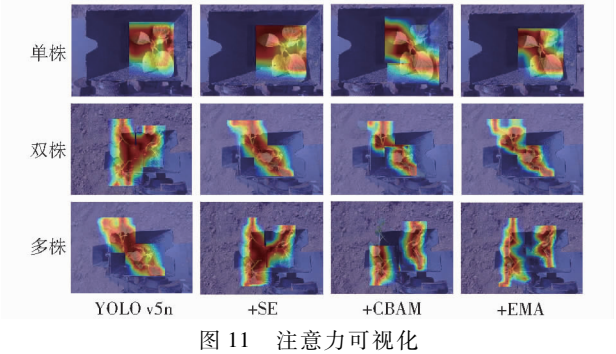

图 11　注意力可视化

Fig. 11　Attention visualization

通过热力图发现原 YOLO v5n 模型的关注点较为分散，而 EMA 注意力机制的加入能够有效地将更多的关注点集中在比较细节的特征区域，能够帮助模型在对重叠度较高的辣椒苗识别任务中取得更好的成效，提升了模型的识别能力和准确性。

3.3.3　损失函数改进效果验证方法

为证明损失函数改进的有效性，根据评价指标绘制了曲线图，记录模型训练和验证过程中的平均精度均值和损失值变化情况。

由于 Shape－IoU 损失函数的设计更侧重于优化边界框形状的准确性，而非单纯的数值降低，在计算的过程中增加了距离损失和形状损失，因此损失值无法成为评价 Shape－IoU 损失函数效果的唯一标准，还需要平均精度均值曲线作为对比，如图 12 所示。

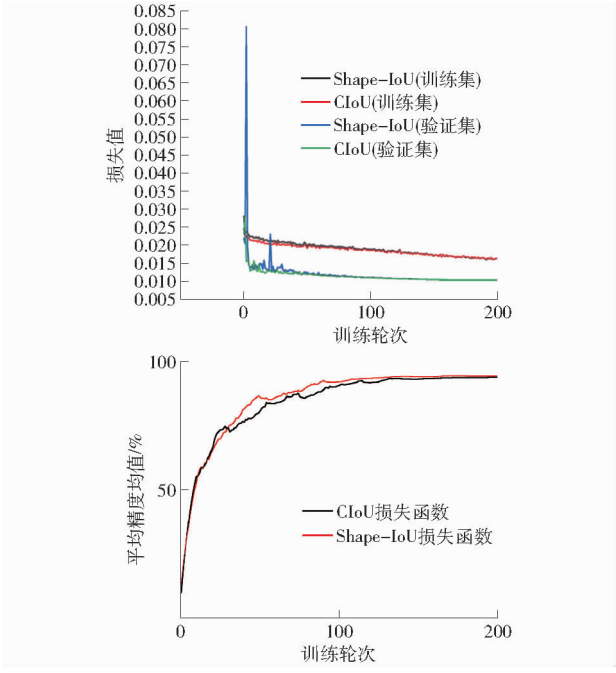

图 12　平均精度均值曲线和损失函数曲线

Fig. 12　Average accuracy mean curve and loss function curve

综上所述，Shape－IoU 损失函数虽然在损失值的表现上与原先 CIoU 损失函数相差不大，而在平均精度均值取得了更优的结果，说明 Shape－IoU 损失函数提升了模型的稳定性和预测准确性，在优化边界框形状和位置上具有更好的表现。

3.3.4　消融实验结果分析

为验证改进方法的鲁棒性和可靠性，设计了消融实验，初始学习率设置为 0.01，动量为 0.937，权重衰减为 0.000 5。默认情况下，所有比较方法的网络都经过了 200 个迭代周期（epoch）的训练，实验结果如表 2 所示。

从表 2 可得出，Ghost 相关模块的引入使得模型的参数量下降 55.6%，浮点运算量下降 53.6%，帧

表 2 消融实验结果

Tab. 2 Ablation experiments results

序号	Ghost	EMA	Shape－IoU	精确率/%	召回率/%	平均精度均值/%	参数量	浮点运算量	帧速率/(f·s^{-1})
1	×	×	×	91.1	92.9	95.0	1 672 470	4.1×10^9	279.87
2	√	×	×	90.9	92.5	94.6	742 234	1.9×10^9	322.63
3	×	√	×	91.8	93.1	95.4	1 723 798	4.2×10^9	271.15
4	×	×	√	91.6	92.4	95.3	1 672 470	4.1×10^9	278.46
5	×	√	√	91.5	93.4	95.7	1 723 798	4.0×10^9	270.34
6	√	×	√	92.6	91.1	94.9	742 234	2.0×10^9	320.45
7	√	√	×	92.1	91.9	95.0	793 562	2.1×10^9	315.54
8	√	√	√	92.2	92.2	95.3	793 562	2.1×10^9	314.12

注:√表示包含此模块,×表示不包含此模块。

速率提高 15.2%,而平均精度均值只下降 0.4 个百分点,这说明轻量化效果十分显著;EMA 的引入使精确率、召回率和平均精度均值分别提高 0.7、0.2、0.4 个百分点,而参数量只增加 3%,浮点运算量只提高 2.4%,说明该注意力机制的加入在对模型轻量化结果影响不大的同时,有效提升了精度;Shape－IoU 的效果主要体现在平均精度均值提升 0.3 个百分点。综合来看,模型经过改进后不仅参数量大小和浮点运算量分别减小 52.5% 和 51.2%,帧速率提升 12.2%,同时平均精度均值上提升 0.3 个百分点,说明改进让模型具有更好的识别精度和实时性。

3.3.5 不同模型实验结果分析

为了验证 YOLO v5n－GE 模型的优越性,使用 YOLO v5n、YOLO v6n、YOLO v7－tiny、YOLO v8n、YOLO v9－T、YOLO v10n、YOLO 11n 共 7 种主流模型在同一实验环境下,在辣椒苗数据集上进行训练和验证,得到实验结果如表 3 所示。

表 3 不同模型对比实验结果

Tab. 3 Comparison experiment results of different models

模型	平均精度均值/%	参数量	浮点运算量	帧速率/(f·s^{-1})
YOLO v5n	95.0	1 672 470	4.1×10^9	279.87
YOLO v6n	93.5	4 531 982	1.14×10^{10}	247.31
YOLO v7－tiny	94.6	6 014 988	1.32×10^{10}	231.25
YOLO v8n	94.7	3 005 843	8.1×10^9	258.13
YOLO v9－T	95.5	2 616 950	1.07×10^{10}	265.42
YOLO v10n	94.6	2 707 414	8.4×10^9	260.51
YOLO 11n	95.2	2 590 035	6.4×10^9	271.67
YOLO v5n－GE	95.3	793 562	2.0×10^9	314.12

从表 3 的实验结果中得出,YOLO v6n 在平均精度均值表现最差;YOLO v7－tiny 虽然平均精度均值处于中等偏上水平,但参数量和运算量过大,需要能承受较高计算资源的设备,不便于部署在移动终端上;YOLO v8n、YOLO v10n 各项指标处于中等水平;YOLO 11n 各项指标虽然处于优等水平,但不及 YOLO v5n－GE;YOLO v9－T 和 YOLO v5n－GE 的平均精度均值表现最优,分别达到 95.5% 和 95.3%,但 YOLO v9－T 的浮点运算量较大。

综上所述,YOLO v5n－GE 模型参数量和运算量达到最低,帧速率最高,表明其处理速度最快,适合对实时性要求较高的应用场景,便于部署在计算资源有限的设备中,在精度上也有较优的表现。

3.4 识别效果

在全自动移栽机的实际工作环境中,由于移栽机栽植部件经常会出现漏苗和多株辣椒苗堵苗的情况,因此需要目标检测模型能够在辣椒苗重叠程度高的干扰下,准确地识别出辣椒苗的数量。为验证改进方法的有效性,对不同改进情况下的辣椒苗进行了识别,检测效果如图 13 所示。



图 13 模型检测结果

Fig. 13 Results of module test

从图 13 可以看出,YOLO v5n 模型在面对重叠度高的辣椒苗识别任务中容易出现多检和漏检的情况,而 EMA 注意力机制的加入可以有效解决该问题,且 Shape－IoU 损失函数的加入帮助模型在框选目标时更加精确。

4 结论

(1)提出了一种改进 YOLO v5n 模型的目标检测算法 YOLO v5n－GE,该算法可以部署在全自动移栽机上,用于对移栽机栽植部件中的辣椒苗进行

识别，以监测栽植部件作业状态。

（2）为满足移栽机实时性要求，引入 Ghost 卷积模块实现网络轻量化，降低了模型的计算延迟；针对真实场景中工作环境复杂、辣椒苗目标重叠程度高的问题，在模型中加入了 EMA 注意力机制，有效改善了多检、漏检的问题；用 Shape－IoU 替换 CIoU 损失函数，使模型对目标的框选更加精确，提高容错率。改进后的 YOLO v5n－GE 模型平均精度均值达到 95.3%，参数量和浮点运算量比 YOLO v5n 模型降低 52.5% 和 51.2%，帧速率提升 12.2%，在实现轻量化的同时保持了较高的识别精度。

（3）与主流的 YOLO v9、YOLO v10、YOLO 11 等模型相比，YOLO v5n－GE 模型的精确率和平均精度均值表现优异，轻量化程度达到最优水平。此外，模型在光照多变、重叠度高的移栽机栽植部件辣椒苗识别任务当中，展现出了更强的泛化能力和鲁棒性，检测速度可以满足实时检测的需求，便于部署在全自动移栽机上。

参　考　文　献

[1] 刘蒙滋，姜凯，王秀，等. 蔬菜自动移栽机研究现状与展望[J]. 农机化研究，2024，46(10)：1－8，20.
LIU Mengzi，JIANG Kai，WANG Xiu，et al. Research status and prospect of automatic vegetable transplanter[J]. Journal of Agricultural Mechanization Research，2024，46(10)：1－8，20.（in Chinese）

[2] RAHNEMOONFAR M. SHEPPARD C. Deep Count：fruit counting based on deep simulated learning[J]. Sensors，2017，17(4)：905.

[3] WU J，YANG G，YANG X，et al. Automatic counting of in situ rice seedlings from UAV images based on a deep fully convolutional neural network[J]. Remote Sensing，2019，11(6)：691.

[4] JIANG H，ZHANG C，QIAO Y，et al. CNN feature based graph convolutional network for weed and crop recognition in smart farming[J]. Computers and Electronics in Agriculture，2020，174：105450.

[5] GIRSHICK R，DONAHUE J，DARRELL T，et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2014：580－587.

[6] GIRSHICK R. Fast R－CNN[C]// Proceedings of the IEEE International Conference on Computer Vision，2015：1440－1448.

[7] 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，2016，39(6)：1137－1149.

[8] REDMON J，FARHADI A. YOLOv3：an incremental improvement[J]. arXiv Preprint，arXiv：1804.02767，2018.

[9] BOCHKOVSKIY A，WANG C Y，LIAO H Y M. YOLOv4：optimal speed and accuracy of object detection[J]. arXiv Preprint，arXiv：2004.10934，2020.

[10] LI C，LI L，JIANG H，et al. YOLOv6：a single-stage object detection framework for industrial applications[J]. arXiv Preprint，arXiv：2209.02976，2022.

[11] WANG 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，2023：7464－7475.

[12] LIU W，ANGUELOV D，ERHAN D，et al. SSD：single shot multibox detector[C]// Computer Vision-ECCV 2016：14th European Conference，2016：21－37.

[13] 马宏兴，董凯兵，王英菲，等. 基于改进 YOLO v5s 的轻量化植物识别模型研究[J]. 农业机械学报，2023，54(8)：267－276.
MA Hongxing，DONG Kaibing，WANG Yingfei，et al. Lightweight plant recognition model based on improved YOLO v5s[J]. Transactions of the Chinese Society for Agricultural Machinery，2023，54(8)：267－276.（in Chinese）

[14] 冀汶莉，刘洲，邢海花. 基于 YOLO v5 的农田杂草识别轻量化方法研究[J]. 农业机械学报，2024，55(1)：212－222，293.
JI Wenli，LIU Zhou，XING Haihua. Lightweight method for identifying farmland weeds based on YOLO v5[J]. Transactions of the Chinese Society for Agricultural Machinery，2024，55(1)：212－222，293.（in Chinese）

[15] YAN B，FAN P，LEI X，et al. A real-time apple targets detection method for picking robot based on improved YOLOv5[J]. Remote Sensing，2021，13(9)：1619.

[16] ZHANG L，WU L，LIU Y. *Hemerocallis citrina* Baroni maturity detection method integrating lightweight neural network and dual attention mechanism[J]. Electronics，2022，11(17)：2743.

[17] LIN T Y，DOLLÁR P，GIRSHICK R，et al. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2017：2117－2125.

[18] LI H，XIONG P，AN J，et al. Pyramid attention network for semantic segmentation[J]. arXiv Preprint，arXiv：1805.10180，2018.

[19] 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，2020：1580－1589.

[20] YANG L，ZHANG R Y，LI L，et al. Simam：a simple，parameter-free attention module for convolutional neural networks[C]// International Conference on Machine Learning. PMLR，2021：11863－11874.

[21] WEBB B S，DHRUV N T，SOLOMON S G，et al. Early and late mechanisms of surround suppression in striate cortex of macaque[J]. Journal of Neuroscience，2005，25(50)：11666－11675.

[22] OUYANG D，HE S，ZHANG G，et al. Efficient multi-scale attention module with cross-spatial learning[C]// ICASSP 2023—2023 IEEE International Conference on Acoustics，Speech and Signal Processing（ICASSP）. IEEE，2023：1－5.

[23] ZHENG Z，WANG P，LIU W，et al. Distance－IoU loss：faster and better learning for bounding box regression[C]// Proceedings of the AAAI Conference on Artificial Intelligence，2020，34(7)：12993－13000.

[24] ZHANG H，ZHANG S. Shape-IoU：more accurate metric considering bounding box shape and scale[J]. arXiv Preprint，arXiv：2312.17663，2023.