文章编号:1671-4598(2025)10-0072-08 DOI:10.16526/j.cnki.11-4762/tp.2025.10.010 中图分类号:TP274 文献标识码:A

基于改进 YOLOv5 的烟草采摘机器人视觉检测模型研究

郭巍巍，赵 崇，李 翔，许晓敬，唐培培
（河南省烟草公司驻马店市公司，河南 驻马店 463000）

摘要：近年来，智能农业在提高生产效率与减少资源浪费方面展现出广泛前景；针对烟草采摘过程中背景复杂、目标遮挡严重等问题，设计了一种基于改进 YOLOv5 的烟草采摘机器人视觉检测模型；该模型融合双目立体成像与通道注意力机制，引入多尺度特征增强金字塔、Hybrid Backbone 结构、注意力模块、Transformer 融合模块以及动态实例归一化模块，提升了目标检测的准确性与环境适应能力；通过自建烟草图像数据集开展实验测试，结果显示该模型在平均精度上达到 92%，在复杂背景下识别准确率提升至 90%，识别响应时间缩短至 0.2 s；研究成果已应用于实际采摘场景，验证了所提模型在提高检测精度、响应速度与鲁棒性方面的有效性与实用性。

关键词：烟草；机器人；识别检测；YOLOv5；实例归一化

Research on Visual Detection Model of Tobacco Picking Robot Based on Improved YOLOv5

GUO Weiwei, ZHAO Chong, LI Xiang, XU Xiaojing, TANG Peipei
(Zhumadian City Co., Henan Provincial Tobacco Co., Zhumadian 463000, China)

Abstract: In recent years, it is of great potential for intelligent agriculture to improve production efficiency and reduce resource waste. To address complex backgrounds and severe occlusions during the tobacco harvesting process, a visual detection model for tobacco-picking robots based on an improved YOLOv5 was developed. The model integrates binocular stereo imaging and channel attention mechanisms, and incorporates a multi-scale feature enhancement pyramid, hybrid backbone structure, attention module, Transformer fusion module, and dynamic instance normalization module to enhance the detection accuracy and environmental adaptability of targets. Experimental tests is conducted on a self-constructed tobacco dataset, which demonstrates that the model achieves a mean average precision of 92%, improves the recognition accuracy to 90% under complex backgrounds, and reduces the recognition response time to 0.2 s. Research results are applied in actual harvesting scenarios, proving its effectiveness and practicality in detection accuracy, response speed, and robustness.

Keywords: tobacco; robot; identification and detection; YOLOv5; instance normalization

0 引言

随着全球人口的增长与耕地资源的日益紧张，农业生产面临产量提升与成本控制的双重压力，特别是在劳动密集型经济作物的种植与收获环节。烟草作为中国、美国、巴西等国家的重要经济作物，其产业链覆盖种植、加工、流通等多个环节，直接关系到区域经济收入与税收结构。然而，当前我国烟草采摘仍以人工操作为主，普遍存在效率低、人力成本高、作业强度大等问题[1]。据统计，在烟草主产区，人工采摘环节人力投入占总成本的比重超过 40%，且存在采摘不均、烟叶损伤率高、采摘时机控制不准等质量隐患。此外，农村人口老龄化趋势加剧也使得烟草种植劳动力资源日益短缺。当前，尽管部分企业引入了初步的机械化辅助设备，但真正实现对烟叶目标的精准识别与智能采摘仍面临重大技术瓶颈，如环境适应性差、目标检测不稳定、

收稿日期:2025-05-07； **修回日期**:2025-06-23。
作者简介:郭巍巍(1980-),男,大学本科,助理农艺师。
通讯作者:赵 崇(1978-),男,大学专科,助理农艺师。
引用格式:郭巍巍,赵 崇,李 翔,等. 基于改进 YOLOv5 的烟草采摘机器人视觉检测模型研究[J]. 计算机测量与控制,2025,33(10):72-79.

识别速度慢等问题。因此，构建高精度、高鲁棒性、实时响应的视觉检测系统，是推动烟草采摘机器人落地应用的核心环节。在自动化农业技术的发展中，目标检测和识别技术起着关键作用。传统的图像识别技术往往依赖于手工特征提取，效率低下且对复杂环境的适应性差[2]。而近年来，深度学习的快速发展，尤其是卷积神经网络的应用，使得图像识别技术得到了显著提升。闫彬等人[3]为了准确识别果树上的苹果目标并区分枝干遮挡情形，提出了一种基于改进 YOLOv5m（You Only Look Once version 5m）的苹果采摘实时识别方法。研究结果表明，该方法能有效识别不同采摘方式的苹果，可为机器人采摘提供技术支撑，降低采摘损失。Yar 等人[4]为了克服传统火灾探测模型误报率高、时间复杂、模型大等挑战，提出了一种改进的 YOLOv5s 模型，集成了 Stem 模块，替换了 SPP 中的大内核并添加了 P6 模块。研究结果表明，该模型能以较低复杂度和较小尺寸有效检测火灾区域，在自制数据集上实现了更好的检测性能，适用于现实场景。

目前，国外研究主要聚焦于基于深度学习的农作物识别与采摘场景感知，多采用 Faster R-CNN、YOLO 系列、Mask R-CNN 等模型对作物进行目标检测或语义分割，重点放在果蔬类目标（如苹果、草莓）的识别与采摘路径规划上。这类方法在清晰背景条件下性能良好，但在目标密集、重叠遮挡或复杂光照场景下检测精度显著下降，且模型普遍存在结构庞大、响应迟缓等问题，难以直接应用于烟草采摘的实时场景。国内研究则更多关注基于 YOLOv5 等轻量级网络的烟叶识别建模，已有部分方法尝试应用二维图像识别技术完成烟草图像中目标检测任务，在简单环境下获得了一定的检测精度，但仍存在以下问题，缺乏立体视觉建模手段，无法准确判断目标深度与空间位置信息；对烟叶边缘模糊、小目标及复杂光照条件下鲁棒性差；模型泛化能力不足，训练后难以适应多变采摘环境。因此，研究提出了一种基于改进 YOLOv5 的采摘机器人视觉检测模型，该模型采用双目成像技术收集烟草图像数据，结合通道注意力机制（SE，squeeze and excitation）对 YOLOv5 进行改进。研究的创新点在于通过双目成像技术提供了精准的深度信息，有助于提高模型对烟叶的空间定位精度，特别是在复杂遮挡和重叠的场景下。SE 模块通过自适应调整特征图中的通道权重，增强了模型对有效特征的捕捉能力。双路径注意力机制（DPA，dual-path attention）通过同时处理空间和通道信息，进一步提升了特征表达能力。CIoU Loss 函数优化了目标框的拟合度，尤其在目标形状不规则时，提升了定位精度。多尺度特征融合金字塔通过融合不同分辨率的特征图，提高了对多尺度烟叶目标的检测能力。IN-Net 模块通过对每个输入样本进行实例归一化，增强了模型在不同环境下的适应性，尤其在复杂光照变化的情况下减少了误检和漏检。

1　YOLOv5 基础模型框架与改进技术路线设计

1.1　YOLOv5 基础架构分析

YOLOv5 是一种先进的目标检测算法，其设计以速度快、精度高、模型轻便著称，适用于实时和嵌入式应用场景。YOLOv5 主要基于 YOLO 系列架构进行优化，采用了一些新型的计算技术和网络结构，其结构如图 1 所示。

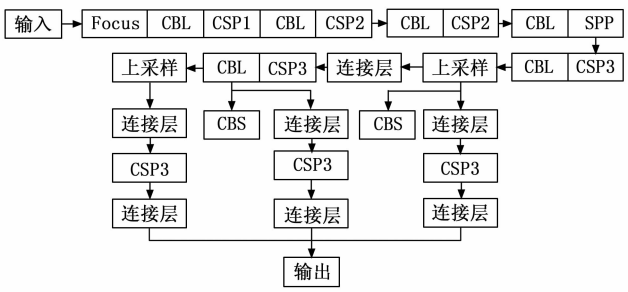


图 1　YOLOv5 模型结构

由图 1 可知，YOLOv5 的模型结构，分为输入端、主干网络、中间层和检测头共 4 部分。首先是输入端，接收输入图像，并对图像进行预处理[5-6]。然后将图像输入到主干网络模块，用于提取图像特征。主干网络模块包括 Focus 层、CBL 模块、跨阶段局部网络（CSP，cross stage partial）模块和空间金字塔（SPP，spatial pyramid pooling）模块。Focus 层在图像中提取重要信息，CBL 模块进行特征的卷积处理，CSP 模块分离和融合特征，减少冗余计算，提高效率。SPP 模块则使用多尺度池化，使模型捕捉多尺度信息[7-8]。中间层负责增强特征的多尺度表达能力。在 YOLOv5 中，中间层包含了 CSP 模块、CBL 模块、上采样层和特征连接层。上采样层用于放大特征图，有助于检测不同尺度的目标，特征连接层将不同层的特征图融合，使高层特征和底层特征共享信息，从而提升检测精度[9-11]。检测头，包含卷积层，将多尺度融合后的特征映射为最终的检测结果。在烟草采摘的实际场景中，目标通常具有小尺寸、高密度和遮挡严重等特征，传统 YOLOv5 尽管具备较快的检测速度和良好的精度表现，但仍存在局限性。如多目标遮挡下，易出现特征混淆与检测误差、在强光或逆光等复杂光照条件下，模型泛化能力不足、面对部分遮挡或细小目标的烟叶，存在漏检等问题，因此需要对传统 YOLOv5 进行改进。

1.2　改进技术路线设计

为解决传统 YOVOv5 存在的问题，研究提出了多策略联合优化的改进模型，构建了一条结合感知增强与

结构轻量化的技术路线，如图 2 所示。

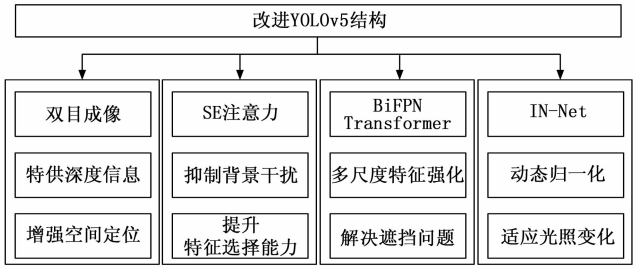


图 2 技术路线设计

由图 2 可知，在该模型中，双目成像模块用于获取图像的深度信息，通过立体视觉提供三维空间的位置信息，从而增强模型的空间定位能力；SE 注意力机制对特征图的通道维度进行加权，通过抑制背景干扰突出关键信息，有效提升特征选择能力；BiFPN 与 Transformer 模块联合构建多尺度特征融合通路，提升模型对复杂目标的表达力，特别是在多目标遮挡场景下能够增强检测精度；IN-Net 模块则通过动态实例归一化方法，自适应调整图像在不同环境下的特征分布，使模型具备较强的光照适应能力。4 个模块各司其职又相互联动，共同构建具备强鲁棒性、适应性和实时性的改进 YOLOv5 结构。

2 双目成像与数据增强模块

2.1 烟草采摘双目立体定位系统构建

为实现对烟叶空间位置信息的精准建模，研究搭建了双目立体视觉系统。该系统由两台工业相机构成，其信息参数如下：型号：Basler ace acA1920，分辨率为 1 920×1 080，基线距离：12 cm。并通过 OpenCV 中的 SGBM 算法实现视差图构建[12-13]。双目成像立体定位数据收集模型的结构如下，首先进行图像采集，双目相机系统由两个相机组成，这两个相机具有相同的参数，并且平行排列，相机之间的距离称为“基线”。双目相机拍摄的两幅图像称为左图像和右图像，通过左右图像进行图像采集。然后进行特征匹配，寻找左图像和右图像中相同的特征点。通过找到这些点之间的对应关系，计算这些点的视差，从而完成特征匹配。其立体视觉方程组如式（1）所示：

$$\begin{cases} d = x_l - x_r \\ Z = \dfrac{f \cdot B}{d} \end{cases} \tag{1}$$

式中，d 为左右视差，Z 为深度信息，即物体到相机的距离。x_l 和 x_r 为左、右图像中的同一物体点的水平坐标。B 为相机的基线距离，d 为视差。在研究中，视差图的构建采用了 OpenCV 库中的半全局块匹配算法（SGBM，semi-global block matching）。该算法兼顾了局部块匹配的效率与全局优化的准确性，适合处理烟叶遮挡多、纹理复杂的图像场景。SGBM 关键参数设置如下：首先，将最小视差设置为 0，代表从无视差起始进行匹配。最大视差范围设为 64，表示左右图像匹配时最多可搜索 64 个像素差值，该值必须为 16 的整数倍，数值越大可覆盖更远距离但计算复杂度更高。匹配块大小设为 9，用于定义计算匹配代价的局部窗口，窗口越大结果越平滑，越小则更敏感于边缘变化。代价平滑项 P_1 和 P_2 用于控制视差结果的连续性，其中 P_1 为低强度平滑项，P_2 为高强度跳变惩罚项，P_2 需大于 P_1，一般根据块大小自动设定。唯一性比率设为 10，用于确保最佳匹配与次佳匹配间的差异明显，避免误匹配。左右一致性检查设置为 1，确保左图与右图匹配结果的一致性，提升可靠性。为滤除背景噪声和误匹配区域，设置了斑点窗口大小为 100，允许的斑点差值变化范围为 32。匹配模式采用三路径优化，在保持较高准确率的同时提升计算效率。

2.2 烟草图像预处理与增强策略

在烟草采集场景中，烟叶的分布复杂且易受遮挡影响。通过双目成像获取精确的深度信息，可以明确烟叶在三维空间中的位置，帮助机器人识别不同层级的烟叶，尤其是靠近地面或重叠的烟叶。但是原始图像常受采集距离、光照强度、背景杂乱度等多种因素影响，易导致模型训练过程中的过拟合与泛化能力不足。为此，研究在训练阶段引入多种图像增强策略，以提升数据多样性与模型鲁棒性。具体增强方式如表 1 所示。

表 1 数据增强策略表格

增强类型	参数设置	主要作用
随机裁剪	缩放比：0.5～1.2	模拟不同距离下烟叶目标的尺寸变化，增强模型的尺度不变性
HSV 调整	色调：±20%，饱和度：±30%	增强模型对不同光照条件下图像色彩变化的适应能力
MixUp 融合	权重：λ=0.2～0.8	将两幅图像按比例混合，增强小目标的检测能力并抑制过拟合

由表 1 可知，随机裁剪可以重构图像的构图比例，使模型能更好地识别距离较近或较远的烟叶；HSV 颜色扰动则模拟不同采摘时间段下的色彩变化，提高对光照敏感区域的容忍度；而 MixUp 融合则通过图像混叠方式引入标签平滑，有效提升模型对边缘目标、遮挡目标的泛化识别能力。

3 主干模型轻量化与注意力机制融合

3.1 Hybrid Backbone 结构设计

在目标检测任务中，主干网络的作用是提取图像的底层、中层和高层特征，对最终的检测精度和速度起决定性作用。传统 YOLOv5 的主干网络采用 CSPDarknet53 结构，该结构主要依赖于 CSP 模块来提高特征提

取能力。然而，CSPDarknet53 在计算量、模型参数量以及特征融合方面仍然存在一定的局限性。因此，研究对 YOLOv5 的主干部分进行创新性改进，提出了一种 Hybrid Backbone 结构，以增强特征表达能力并提高模型推理速度。

传统 YOLOv5 的主干网络在计算量和参数量方面相对较大，导致在嵌入式或实时应用场景中可能存在推理速度不够快的问题。为了解决这一问题，研究在原 CSPDarknet53 的基础上，引入 MobileNetV3 的轻量化特征提取结构。MobileNetV3 采用了深度可分离卷积和硬激活函数，能够在降低计算复杂度的同时提高网络的特征提取能力。其结构如图 3 所示。

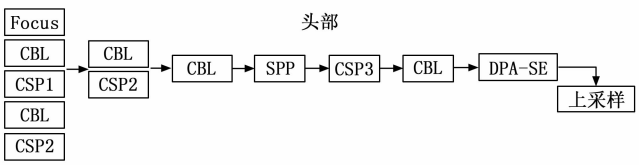


图 3　Hybrid Backbone 结构

由图 3 可知，在改进的 Hybrid Backbone 结构中，输入图像首先通过 Focus 模块进行通道切片，接着依次经过两个 CSP 模块用于提取浅层特征，随后引入两组 ResNet 残差块增强深层语义表达能力，并与后续 MobileNetV3 模块进行串联，其中 MobileNetV3 采用深度可分离卷积（kernel＝3×3，stride＝1）与 Hard-Swish 激活函数构建轻量级中层特征提取单元，同时在每个 MobileNetV3 单元中嵌入 SE 通道注意力模块用于提升重要特征响应，SE 模块采用全局平均池化后接两层全连接网络（隐藏节点数为输入通道数的 1/4 与原通道数），激活函数为 Sigmoid。所有模块输出通过 CSP 结构进行残差融合与统一编码，形成主干输出特征图。该结构在保持模型高效性的同时，增强了对复杂纹理和小目标的多层感知能力。YOLOv5 的 CSP 结构主要通过跨阶段特征融合的方式提高梯度流动性，但在处理高层语义信息时，其特征表达能力仍然有限。因此，研究在 Hybrid Backbone 结构中加入了 ResNet 的残差块，以增强模型对深层语义特征的学习能力。残差块的引入能够有效缓解深度神经网络中的梯度消失问题，同时提升模型对复杂目标的特征提取能力。最后引入 SE 模块，SE 主要分为两个部分，分别是压缩和激励。首先，通过全局平均池化将输入特征图在空间维度上压缩，得到一个维度的向量来表示每个通道的全局信息，其表达式如式（2）所示：

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{i,j,c} \tag{2}$$

式中，H 和 W 分别为空间维度的高度和宽度。X 为输入的特征图，z_c 为压缩后的全局通道特征，c 表示第 c 个通道。压缩后的通道特征输入到两个全连接层，生成通道权重。第一个全连接层用于降维，将全局通道特征映射到一个低维空间。第二个连接层用于升维，将降维后的特征恢复到原通道数并应用 Sigmoid 激活函数，以生成每个通道的权重。最后将得到的权重应用于原特征图的每个通道。

3.2　DPA-SE

本文设计了一种能够同时处理空间和通道两个维度的信息、进一步提升模型表达能力的 DPA-SE 机制，其结构如图 4 所示。

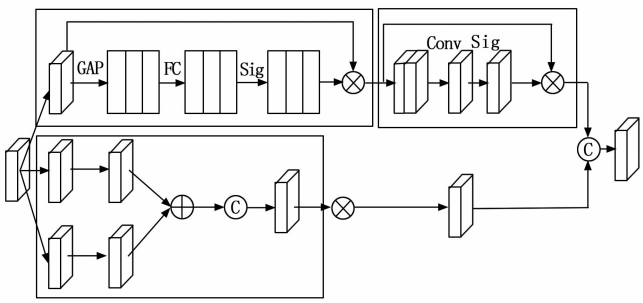


图 4　DPA-SE 结构

由图 4 可知，空间注意力路径通过计算特征图的空间依赖关系，生成空间注意力权重，突出重要的空间区域。其公式如式（3）所示：

$$Spatial\ Attention = \sigma(Conv_1 \times Concat(max(f_c), \setminus avg(f_c))) \tag{3}$$

式中，$Conv_1$ 表示一个卷积操作，max 和 avg 分别为最大池化和平均池化操作。通道注意力路径类似于 SE 模块，通过压缩和激励过程生成通道注意力权重，突出重要的通道。颈部部分通过多次上采样和特征融合操作实现特征金字塔结构，包含多个 CBL 和 CSP 模块，以不同分辨率提取特征[14]。预测部分对 3 种尺度的特征图进行卷积操作，得到目标检测的最终输出，通过输出结果来对烟草进行采摘。

4　颈部网络多尺度特征优化

4.1　BiFPN-Transformer-Dynamic Head 模型

为解决传统 YOLOv5 颈部网络在复杂场景下多尺度特征融合不足的问题，本研究提出一种融合双向特征金字塔网络（BiFPN，bi-directional feature pyramid network）、Transformer 编码器与 Dynamic Head 的模型：

针对传统 PANet 单向特征聚合导致深层与浅层信息交互受限的缺陷，设计 BiFPN 实现多尺度特征的高效融合：通过上行路径将低层细节特征上采样后与高层语义特征融合，增强对小目标的定位能力；借助下行路径将高层语义特征下采样后与底层特征结合，保留全局上下文信息。这一双向融合机制有效打破了传统单向聚合的信息瓶颈，为后续检测头提供了更丰富的特征表示。

在此基础上，为解决卷积神经网络局部感受野受限导致的远距离特征关联缺失问题，本研究在 BiFPN 基础上引入 Transformer 编码器。通过多头自注意力机制，Transformer 能够捕捉全局空间依赖关系，显著提升模型对烟叶边缘模糊或目标交叠场景的适应能力。

为进一步提升特征融合的灵活性与适应性，本研究设计了 Dynamic Head 模块。该模块采用可变形卷积技术，通过动态调整卷积核采样位置自适应匹配目标形变与位移；同时并行执行空间金字塔池化与通道注意力池化，强化关键特征响应。

具体实现中，改进后的颈部网络架构遵循以下流程（见图 5）：首先从主干网络提取 3 组对应小、中、大目标尺度的特征图；随后通过 BiFPN 进行跨尺度双向特征重组，实现多层级特征的互补增强；在此基础上引入 Transformer 编码器捕捉长距离依赖关系，弥补 CNN 局部感受野的不足；最后利用可变形卷积与多尺度池化动态优化特征权重，经卷积归一化后输出至预测分支。该设计通过多模块协同，在保持计算效率的同时实现了特征选择、融合与表征能力的全面提升，为复杂场景下的烟草目标检测提供了强鲁棒性支撑。

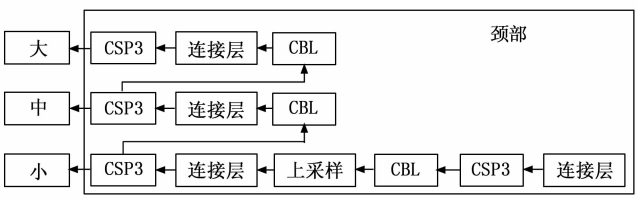


图 5 BiFPN-Transformer-Dynamic Head 模型

4.2 *CIoU* 损失函数改进

在颈部部分，采用 *CIoU Loss* 函数，该损失函数结合了 *IoU* 损失、中心点距离惩罚、长宽比惩罚。*IoU* 损失衡量预测框与真实框的重叠程度，*IoU* 越大表示预测越准确[15]。为了进一步使预测框快速收敛到真实框，*CIoU Loss* 引入了预测框和真实框中心点之间的距离惩罚。通过惩罚预测框和真实框的长宽比差异，使模型能够在不同形状的目标上更精确地拟合。*CIoU* 损失在烟叶采摘任务中的引入，能够显著提高预测框与真实框的拟合程度，尤其是在长宽比差异较大的烟叶目标上[16]。通过对中心点的距离惩罚，采摘机器人可以准确定位烟叶的中心点并优化机械臂的采摘轨迹，避免误采或损伤其他烟叶。长宽比的调节进一步提升了模型对不规则烟叶的适应能力，使得采摘操作更加精准[17]。*CIoU Loss* 的表达式如式（4）所示：

$$CIoU\ Loss = 1 - IoU + \frac{d^2}{c^2} + \alpha v \tag{4}$$

式中，*IoU* 为衡量预测框与真实框的重叠程度，d 为中心点的欧氏距离，c 表示包含预测框和真实框的最小外接矩形的对角线长度，α 根据 *IoU* 大小动态调节，α 为调节参数。为进一步理解其优化机制，研究分析损失函数的梯度计算过程。中心距离项对预测框中心坐标的偏导如式（5）所示：

$$\frac{\partial}{\partial x_p}\left(\frac{d^2}{c^2}\right)=\frac{2(x_p-x_g)}{c^2},\frac{\partial}{\partial y_p}\left(\frac{d^2}{c^2}\right)=\frac{2(y_p-y_g)}{c^2} \tag{5}$$

式（5）有效引导预测框向真实中心快速收敛。长宽比惩罚项则通过对预测框宽高比与真实框宽高比之间的差异进行量化，其定义如式（6）所示：

$$v=\frac{4}{\pi^2}\left(\arctan\frac{w_g}{h_g}-\arctan\frac{w_p}{h_p}\right)^2 \tag{6}$$

式中，w_g、h_g、w_p 和 h_p 分别为真实框和预测框的宽高，调节因子则根据 *IoU* 大小动态调整长宽比项对总损失的影响权重。烟草采集的核心在于快速、精准地识别采摘目标，并结合位置信息优化采摘路径。引入多尺度特征增强金字塔（MFEP，multi scale feature enhancement pyramid）。MFEP 通过构建多层次的特征金字塔结构，整合不同尺度的特征信息，提升模型对复杂场景的建模能力[18]。从主干网络的不同层次提取特征图，这些特征图对应不同的空间分辨率和语义信息。通过上采样和下采样操作，将不同尺度的特征图进行融合，生成多尺度的特征金字塔。利用增强后的多尺度特征进行目标检测，生成最终的预测结果。

5 预测网络自适应归一化模块

5.1 实例归一化动态适配

由于 YOLOv5 网络的每一层输出数据都是由上一层决定的，同时参与影响的还有各层的参数，因此数据分布存在多样性和复杂性。为了解决该问题，研究引入了新型归一化处理方式，实例归一化（IN，instance normalization）。IN 可以对单个输入样本的每个通道进行归一化处理。其在处理相同图像数据分布时，可以规范特征统计量，即均值和方差，因此具有更快的收敛速度。IN 计算流程如图 6 所示。

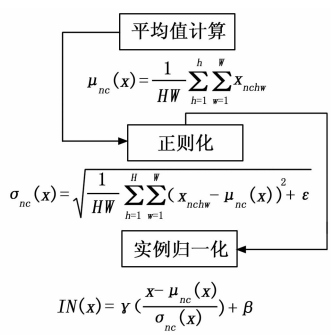


图 6 IN 计算流程图

图 6 中，$\mu_c(x)$ 和 $\sigma_c(x)$ 分别为当前所有图片的每个特征通道计算得到的均值和标准差。γ 和 β 皆表示训练学习过程中的放射参数[19]。W 和 H 分别为图像的宽和高，x_{nchw} 表示在第 n 个样本的第 c 个通道下，张量 x 的

高度为 h，宽度为 w。$\sigma_{nc}(x)$ 表示对单一图像的每个通道所求得的标准差。

5.2　IN-Net 模块部署策略

由于烟草采摘机器人实际应用场景的环境和光线等因素存在多变性，为了进一步提高烟草采摘机器人的目标图像采集效率和检测模型的泛化能力，研究在 YOLOv5 网络的基础上继续引入了 IN-Net 模块[20]。IN-Net 模块是一种能够提高神经网络泛化能力和稳定性的方法。通过在卷积层之前引入实例归一化，IN-Net 能够在处理具有不同环境特征的目标图像时，更加有效地适应环境变化，并减少因光照、背景干扰等因素导致的检测精度下降。

6　改进的 YOLOv5 模型

模型整体结构如图 7 所示。

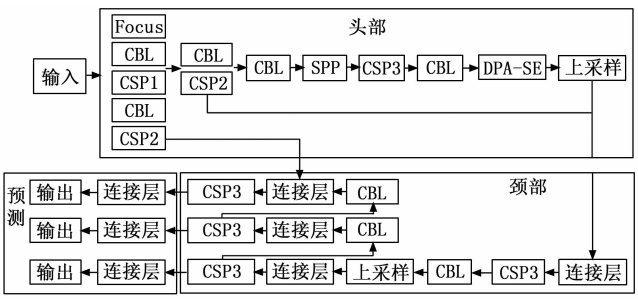


图 7　改进 YOLOv5 的模型结构

由图 7 可知，输入部分，图像首先经过 Focus 模块进行切片处理，随后通过多个卷积块和 CSP 结构提取底层特征。在主干部分，引入了 SE 模块，通过通道注意力机制自适应增强特征表示，该模块嵌入在 CSP 结构内部，直接作用于中高层特征图。经过多层特征提取后，主干输出特征图进入颈部网络。颈部部分通过多次上采样、特征融合与 BiFPN 机制增强跨尺度特征表达，同时结合 Transformer 捕捉全局依赖关系。预测部分则在不同分辨率下分别生成小、中、大目标检测分支，输出最终的目标检测结果。整体结构通过 IN-Net 进一步提升了在复杂环境下的检测鲁棒性。

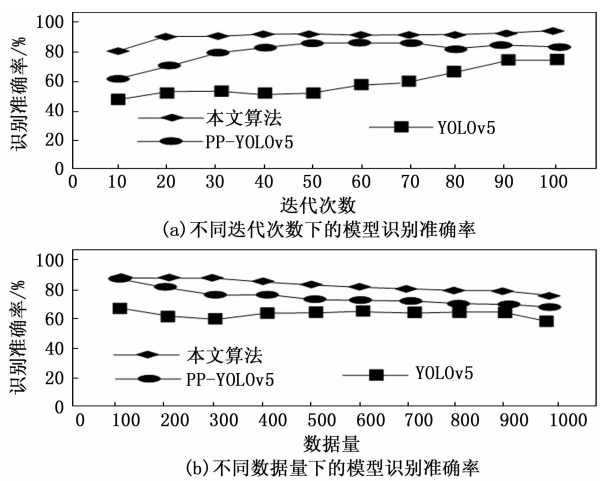
7　基于改进 YOLOV5 的烟草采摘机器人识别性能分析

7.1　自建烟草数据集

研究选择使用配备 Windows 11 操作系统和 32 GB 内存的计算机，并搭载了 Intel Core i5-6300CPU。数据集采用自建数据集，该数据集中包含不同健康状态的烟草植物图像，涵盖健康植物、受病害影响的植物，以及各种常见烟草病害，数据集中的图像分辨率较高，且每张图像附有标签，指明植物的健康状态或病害类型。数据集中共包含 4 000 张烟草图像，其中健康烟叶图像约占 40%，轻微病害、中度病害、重度病害及死亡状态烟叶分别占比 20%、15%、15%与 10%。每张图像均由人工标注烟叶目标的边界框及健康等级标签，采用 VOC 格式进行组织，并划分为训练集（70%）、验证集（20%）与测试集（10%）。为模拟真实采摘环境中的复杂背景与目标重叠情况，数据集中包含大量受遮挡、倾斜、重叠及光照不均的图像样本，显著提升了模型泛化训练的难度。针对小目标检测问题，还特别采集了烟叶远距视角图像，并在标注时保留细粒度边界框，提升模型对微小烟叶的识别能力。为提升模型的鲁棒性与泛化能力，数据集在构建过程中充分考虑了烟草的多样性与采集环境的复杂性。图像采集覆盖常见烟草品种（如黄花烟、中部红花、香料型等）及其典型生长阶段，包括苗期、旺长期、成熟期与采后期，确保训练数据在时序维度上的覆盖广泛。同时，采集任务在晴天、阴天、晨曦、正午及傍晚等不同自然光照条件下进行，部分样本还选取了雨天湿叶、雾霾光照干扰场景，以增强模型对光照与气象变化的适应能力。场景布置覆盖多种典型田间背景，如裸露土壤、杂草干扰、病斑叶片遮挡与叶层重叠等。图像采集采用双工业相机（型号：Basler ace acA1920），固定基线距离与角度控制在±5°，采集角度涵盖俯视、侧视与仰视视角。图像标注由具备农作物病虫识别经验的专家团队人工完成，采用 VOC 格式统一标注烟叶边界框及健康等级标签，确保标签的一致性与可复现性。

7.2　模型性能分析

为研究所提出模型的性能，研究引入 YOLOv5 以及 PP-YOLOv5 模型作为对比模型，采用识别准确率，识别时间等参数作为指标，对模型的性能进行分析，结果如图 8 所示。



(a) 不同迭代次数下的模型识别准确率

(b) 不同数据量下的模型识别准确率

图 8　不同迭代次数和数据量下的模型识别准确率分析

由图 9（a）可知，随着迭代次数的增加，本文算法和 PP-YOLOv5 的准确率均有明显提高，而 YOLOv5 的增长较缓慢。本文算法的准确率在 30 次迭代后已接

近 90%，最终稳定在 95%左右，表现优于其他模型，PP-YOLOv5 在 50 次迭代后达到 85%左右。由图 9（b）可知，随着数据量增加，3 种模型的识别准确率略有波动，但总体较为平稳。SE-YOLOv5 在所有数据量下准确率均维持在 90%以上，而 PP-YOLOv5 略低于本文算法，但仍保持在 80%以上。YOLOv5 的表现最差，在数据量增加后准确率呈下降趋势。这是由于本文算法和 PP-YOLOv5 的模型结构更适合处理大量数据，能够更好地捕捉复杂数据特征，而 YOLOv5 在数据量较大时的适应性不足。实验结果表明，所提出的模型有着更为出色的识别性能。对各个模型的识别时间进行分析，结果如图 9 所示。

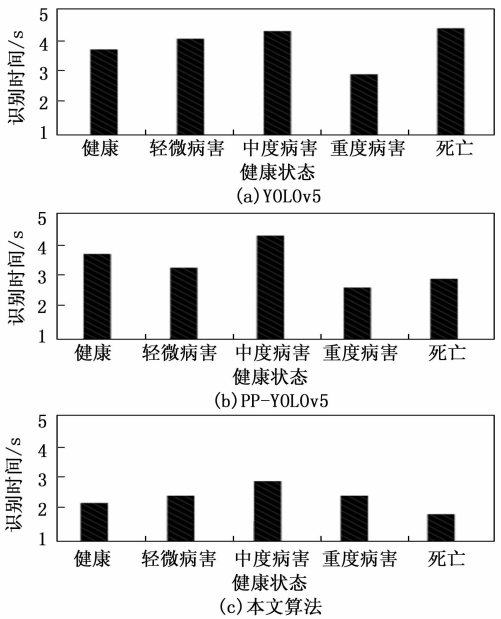


图 9 各个模型识别时间比较

由图 9（a）可知，在 YOLOv5 模型中，对健康状态和轻微病害状态下的识别时间约为 2 s，但对中度和重度病害状态的识别时间显著增加，分别接近于 4 s 和 4.5 s，而对死亡状态的识别时间有所下降，约为 3 s。由图 9（b）可知，在 PP-YOLOv5 模型中，对健康和轻微病害的识别时间也较低，接近于 2 s。对中度和重度病害状态，识别时间分别达到约 3 s 和 3.5 s，对死亡状态的识别时间约为 2.5 s。本文算法模型对健康状态的识别时间最低，约为 1 s 多。对轻微病害状态的识别时间上升至 2 s 多，对中度病害状态的识别时间为 3 s 左右，对死亡状态的识别时间则接近 2 s。实验结果表明，所提出的模型有着更为出色的性能。对各个模型的实际性能进行分析，结果如图 10 所示。

由图 10 可知，在 YOLOv5 的识别结果中，模型成功检测到了一些烟草。然而，在一些区域中，较小的烟草部分或远处的物体没有被识别，这导致了某些区域的边界框较少。在 PP-YOLOv5 的检测结果中，与 YOLOv5

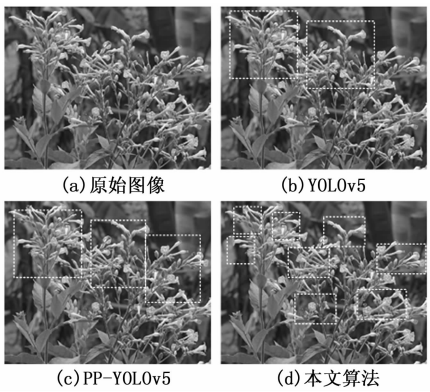


图 10 实际识别效果分析

相比，更多的边界框出现了。这表明 PP-YOLOv5 模型能够在复杂场景中识别更多的物体。在本文算法的识别结果中，检测结果更为精细。与前两种方法相比，模型识别了更多的物体并且定位更为精确，从烟草中可以看到更多的边界框。实验结果显示，所提模型在复杂背景条件下的准确率优于 YOLOv5 与 PP-YOLOv5，主要得益于特征提取模块与结构融合机制的协同优化。一方面，引入的 Hybrid Backbone 结构通过融合 ResNet 与 MobileNetV3 实现深浅特征联合建模，显著提升了模型对细粒度纹理和小目标烟叶的识别能力；另一方面，BiFPN 与 Transformer 联合构建的颈部结构在多尺度上下文建模方面表现优越，能够有效整合局部形态与全局结构信息，增强模型在复杂场景中的泛化表达力。此外，实例归一化模块通过抑制不同光照分布对特征分布的扰动，使模型在光照变化剧烈的田间环境下仍能保持稳定的识别性能。对于错误识别率分析发现，主要误识别样本集中在以下几类：一是严重病害或叶片枯黄区域，其颜色与土壤或杂草背景接近，导致模型边界模糊；二是重叠或半遮挡叶片区域，因深度估计误差与通道压制权重不准而出现目标丢失或误框；三是强反光或阴影区域，造成部分特征图响应异常。相比之下，所提模型在以上场景下表现更稳，反映出多尺度融合结构与注意力机制对边缘区域的定位鲁棒性更强。因此，性能提升不仅源于双目深度信息引入与 SE 模块，还与整体网络感知结构和归一策略高度适配有关。对各个模型的综合性能进行分析，结果如表 2 所示。

表 2 综合性能分析表

模型	平均精度/%	检测速度/FPS	内存占用/MB	错误识别率/%	资源占用率/%	目标检测响应时间/s	复杂背景的准确率/%
YOLOv3	77	22	251	9	78	0.55	76
SE-YOLOv3	82	25	233	7	74	0.47	79
YOLOv5	85	30	200	5	70	0.25	80
PP-YOLOv5	88	28	220	4	75	0.30	85
本文算法	92	32	210	3	72	0.20	90

由表 2 可知，本文算法的平均精度达 92%，优于 PP-YOLOv5 的 88%和 YOLOv5 的 85%。在检测速度方面，本文算法以 32 帧每秒占据优势，适合需要实时响应的采摘任务。在错误识别率方面，本文算法最低，仅为 3%，有助于减少误检，提高采摘效率。且在复杂背景下的准确率达 90%，明显优于 YOLOv5 的 80%，显示出在复杂环境下的稳定性能。目标检测响应时间方面，本文算法表现最佳，仅需 0.2 s，能够有效支持快速的自动采摘需求。实验结果表明，所提出的模型的综合性能表现出色。

8　结束语

针对传统的烟草采摘机器人在复杂环境中往往面临识别精度低和响应速度慢等问题，提出了一种基于改进 YOLOV5 的采摘机器人视觉检测模型，该模型采用双目成像技术收集数据，并通过通道注意力机制对 YOLOv5 进行优化。实验结果表明，本文算法的准确率在 30 次迭代后已接近 90%，最终稳定在 95%左右，表现优于其他模型，随着数据量增加，在所有数据量下准确率均维持在 90%以上。本文算法的平均精度达 92%，优于 PP-YOLOv5 的 88%和 YOLOv5 的 85%。在检测速度方面，本文算法以 32 帧每秒占据优势。在错误识别率方面，本文算法最低，仅为 3%。SE-YOLOv5 在复杂背景下的准确率达 90%，明显优于 YOLOv5 的 80%，目标检测响应时间方面，本文算法表现最佳，仅需 0.2 s。研究结果表明，所提出的方法在各个情况下均表现优秀，但是研究仍然存在不足，尽管 SE-YOLOv5 在大多数条件下表现出色，但在极端条件下，如极度光照变化或极端遮挡的鲁棒性仍需进一步验证。未来研究可以引入更多的改进技术，如集成学习进一步提升模型的鲁棒性和准确性。

参考文献：

[1] 王道累，张世恒，袁斌霞，等. 基于改进 YOLOv5 的轻量化玻璃绝缘子自爆缺陷检测研究 [J]. 高电压技术，2023，49 (10)：4382－4390.

[2] 吕宗喆，徐　慧，杨　骁，等. 动态环境下基于 YOLOv5 的 RGB-D 视觉 SLAM 优化方法 [J]. 制造业自动化，2023，45 (4)：191－195.

[3] 闫　彬，樊　攀，王美茸，等. 基于改进 YOLOv5m 的采摘机器人苹果采摘方式实时识别 [J]. 农业机械学报，2022，53 (9)：28－59.

[4] YAR H，KHAN Z A，ULLAH F，et al. A modified YOLOv5 architecture for efficient fire detection in smart cities [J]. Expert Systems with Applications，2023，231 (10)：120465.1－120465.11.

[5] LI X，LUO R，ISLAM F U. Tracking and detection of basketball movements using multi-feature data fusion and hybrid YOLO-T2LSTM network [J]. Soft Computing，2024，28 (2)：1653－1667.

[6] 何铭亮，王建国，范　斌，等. 基于 yolov5 在线可视铁谱图像磨粒多目标识别方法研究 [J]. 润滑与密封，2023，48 (5)：137－142.

[7] 卢进南，刘　扬，王连捷，等. 基于改进 YOLOX 的电铲铲齿断裂检测方法 [J]. 电子测量与仪器学报，2023，37 (5)：46－57.

[8] HANAFI W，TAMALI M. Implementing distributed collaboration and applying the YOLO algorithm to robots. Studies in Engineering and Exact Sciences，2024，5 (1)：277－296.

[9] 赵金源，贾　迪. 改进 YOLOv5 的多人姿态估计修正算法 [J]. 计算机工程与科学，2024，46 (5)：852－860.

[10] 徐佳鹏，张朝晖，李　智，等. 基于 YOLO 模型的小麦外观分类算法研究 [J]. 自动化仪表，2023，44 (3)：83－87.

[11] LIU S，LIU M，CHAI Y，et al. Recognition and location of pepper picking based on improved Yolov5s and depth camera [J]. Applied Engineering in Agriculture，2023，29 (2)：179－185.

[12] 范先友，过　峰，俞建峰，等. 基于改进 YOLOv7 的液晶面板电极缺陷视觉检测技术研究 [J]. 电子测量与仪器学报，2023，37 (9)：225－233.

[13] DUAN Z，XIE T，WANG Y W J. Microalgae detection based on improved YOLOv5 [J]. IET Image Processing，2024，18 (10)：2602－2613.

[14] 查易艺，王　翀，张明明. 基于知识图谱的变压器匝间短路故障辨识研究 [J]. 自动化仪表，2024，45 (4)：14－18.

[15] WU Q，LI Y，HUANG W，et al. C3TB-YOLOv5：integrated YOLOv5 with transformer for object detection in high-resolution remote sensing images [J]. International Journal of Remote Sensing，2024，45 (8)：2622－2650.

[16] 赵梓杉，桑海峰. 基于改进的 YOLOv5 的交通锥标检测系统 [J]. 电子测量与仪器学报，2023，2 (4)：56－64.

[17] 秦　阳，李欣航. 基于机器学习的土壤锰污染程度预测模型构建 [J]. 中国锰业，2023，41 (6)：80－86.

[18] RUI C，WU Z，LIU C，et al. Surface defect detection of steel strip at low resolution based on SAC-YOLOv5 [J]. Measurement Science & Technology，2025，36 (1)：1－17.

[19] REN H，FAN A，ZHAO J，et al. Lightweight safety helmet detection algorithm using improved YOLOv5 [J]. Journal of Real-Time Image Processing，2024，21 (4)：125.1－125.15.

[20] WANG C，ZHANG Y，ZHOU Y. Automatic detection of indoor occupancy based on improved YOLOv5 model [J]. Neural Computing & Applications，2023，35 (3)：2575－2599.