

基于改进 YOLOv7 的猕猴桃果实识别优化*

何翔 朱洪前

(中南林业科技大学, 湖南长沙 410004)

摘要 为了解决猕猴桃 *Actinidia chinensis* 果实识别过程中存在果实之间重叠导致的遮挡严重、检测结果易受叶片影响等问题, 建立不同日照条件下的猕猴桃果实图像数据集, 对 YOLOv7 模型做了 3 方面改进: 将 Backbone 部分的卷积模块替换成 GhostConv 模块, 在维持原有精度的程度上降低模型的参数量; 针对猕猴桃果实之间存在大量重叠的情况, 引入非极大值抑制 NMS (Soft-NMS) 策略提高检测框回归精度; 融合 SimAM 注意力机制, 增强模型对于高密度猕猴桃特征的提取能力。通过对比实验表明, 优化后的模型与 Faster RCNN 相比, mAP 值增加了 12.7 个百分点, 检测速度提升 106.8 帧/s, 综合性能较好, 满足机器实时对于猕猴桃果实识别的需求。

关键词 果实识别; YOLOv7; GhostConv; Soft NMS; SimAM

中图分类号: S776 **文献标志码:** A **文章编号:** 2096-2053 (2024) 02-0036-10

Optimization of *Actinidia chinensis* Fruit Recognition based on Improved YOLOv7

HE Xiang ZHU Hongqian

(Central South University of Forestry and Technology, Changsha, Hunan 410004, China)

Abstract In order to solve the problems of severe occlusion caused by overlapping fruits and susceptibility to leaf influence in the recognition process of *Actinidia chinensis* fruit, *A. chinensis* fruit image dataset was established under different sunlight conditions. Three improvements were made to the YOLOv7 model: replacing the convolutional module of the Backbone part with the GhostConv module, reducing the number of model parameters while maintaining the original accuracy; to address the significant overlap between *A. chinensis* fruits, a Non Maximum Suppression NMS (Soft NMS) strategy is introduced to improve the accuracy of detection box regression; integrating SimAM attention mechanism to enhance the model's ability to extract high-density *A. chinensis* fruit features. Through comparative experiments, it was shown that the optimized model increased mAP value by 12.7% and detection speed by 106.8 frames/s compared to Faster RCNN. The overall performance is good and meets the real-time recognition needs of machines for *A. chinensis* fruit.

Key words fruit recognition; YOLOv7; GhostConv; Soft NMS; SimAM

作为猕猴桃 *Actinidia chinensis* 生产大国, 我国无论是种植面积还是年产量均居世界第一^[1], 但由于猕猴桃的种植与管理技术落后以及果园产

量的不断上升, 果农劳动强度大、采摘效率低。收获采摘是猕猴桃生产过程中费力最大、耗时最多的一个环节。为保证产品的质量, 必须做到适

* 第一作者: 何翔 (1999—), 男, 在读硕士, 研究方向为林业机器视觉, E-mail: 1453978711@qq.com.

通信作者: 朱洪前 (1974—), 男, 副教授, 研究方向为林业机器视觉, E-mail: zhuhongqian@163.com.

时采摘，否则会使得猕猴桃果质变软从而影响到果实的品质、运输、存储以及后续的加工，进而影响到市场价格和果农的收入。目前国内大部分猕猴桃采摘作业是手工进行^[2]，随着我国人口老龄化水平的增长，造成农业劳动力的减少，农业生产成本也相应提高，农产品的市场竞争力大大降低，所以急需开发一款自动化猕猴桃采摘机器人投入实际应用。

近年来，国内外有众多学者对自然环境下果实的识别与定位进行了大量的研究，主要有番茄^[3]、苹果^[4]、葡萄^[5]等果实目标特征的提取与定位。猕猴桃作为我国主要的经济作物，相关的研究还比较少：慕军营等^[1]利用 Lab 空间的 a 通道分量进行特征提取，通过图像预处理和图像分割得到含有目标特征的二值图像；为保证算法的实时性，利用简化的正椭圆 Hough 变换提取猕猴桃的特征参量，为猕猴桃果实的采摘提供必要信息，正确提取率为 88.5%，平均提取时间为 3.976 s，鲁棒性和时效性较差；Gao 等^[6]为了让采摘机器人在果园环境中快速准确地检测猕猴桃，提出一种基于轻量级 YOLOv4-GhostNet 网络的猕猴桃检测方法，其将原有的 CSP-Darknet53 骨干网模型替换为 GhostNet，最终得到 93.07% 的平均准确率，92.43% 的召回率，但识别速度只有 53 帧/s，难以满足实际识别需求。综上，前人利用传统图像识别检测技术，借助采摘机器人的摄像头对自然环境下的猕猴桃果实进行识别，但随着采摘效率的不断提高，对识别算法的准确度以及识别速度的要求也在不断提升。

原始 YOLOv7 网络在进行特征提取时容易忽略小目标中的显著性信息，且特征提取网络在训练过程中会产生部分冗余参数不利于模型效率的提升。YOLOv7 的损失函数和 anchors 聚类方法在多尺度目标任务的预测精度方面同样具有提升空间，因此，为解决上述问题，在继承原有 YOLOv7 网络的 E-ELAN 结构、Rep 重参数化和辅助训练策略的基础上，本文提出一种以 YOLOv7 改进的猕猴桃检测模型：将模型主干网络的卷积模块替换为 GhostConv (Ghost Convolution) 卷积模块降低模型参数，同时将 YOLOv7 中原有的 NMS (Non-maximum suppression) 极大值抑制策略替换为对高密度高重叠的目标更为敏感的非极大值抑制 NMS (Soft-NMS)。并为增强模型对于

高密度猕猴桃的特征提取能力融合 SimAM (A Simple, Parameter-Free Attention Module) 注意力机制。通过实验，训练出能快速精确识别各种日照环境下的猕猴桃果实且具备轻量化优点的猕猴桃识别模型。

1 材料与方法

1.1 YOLOv7 算法概述^[7]

作为 2022 年 7 月才推出的一款图像识别算法，在计算效率和精度方面，YOLOv7 相对 YOLOv5 有所提升。YOLOv7 使用更快的卷积操作和更小的模型，因此在相同的计算资源下可以达到更高的检测速度。此外，YOLOv7 还提供较高的精度，能够检测更多的细粒度对象。

YOLOv7 在保持 YOLO 系列算法的实时性特点的同时，其准确度超过所有已知的实时目标检测器，在各大规模图像识别数据集上具有最高的平均准确率。尤其是其中的 E-ELAN (extended ELAN) 模块，采用组卷积扩展计算块的通道和基数，通过控制梯度路径长度，有效地学习图像特征，加速网络收敛。该模型对计算层的所有计算块采用相同的组参数和通道乘数，预测一个包含边界框、置信度和类别的三维张量编码，实现目标定位和分类。

YOLOv7 网络结构主要包括 Input 层、Backbone 层和 Head 层。Input 层为输入层，主要作用是对输入的图片进行预处理后输入到 Backbone 层；Backbone 层又被称作特征提取层，主要功能是提取 3 种不同大小的目标信息特征，由 51 层不同的卷积组合模块构成，其最大的特点是采用高效的 E-ELAN 网络架构，可以提高算法的检测效率，Backbone 层的输出为 3 个不同大小的特征信息，位置分别位于第 24 层、第 37 层以及第 50 层；Head 层主要是融合 Backbone 层给出的特征生成边界框并预测类别，主要包括 SPPCSPC (Spatial Pyramid Pooling Connected Spatial Pyramid Convolution) 层、若干 Conv (Convolution) 层、MPConv (Max-pooling Convolution) 层以及 REP (Re-parameterization) 层组成，Head 层在第 75 层、第 88 层以及第 101 层输出 3 种不同大小的特征图，经过重参数化结构 REP 层输出 3 个不同大小的预测结果，便于算法后续部署。

1.2 网络模型改进

为提高网络检测性能和检测速度，本文对该模型做了以下改进（图 1）：

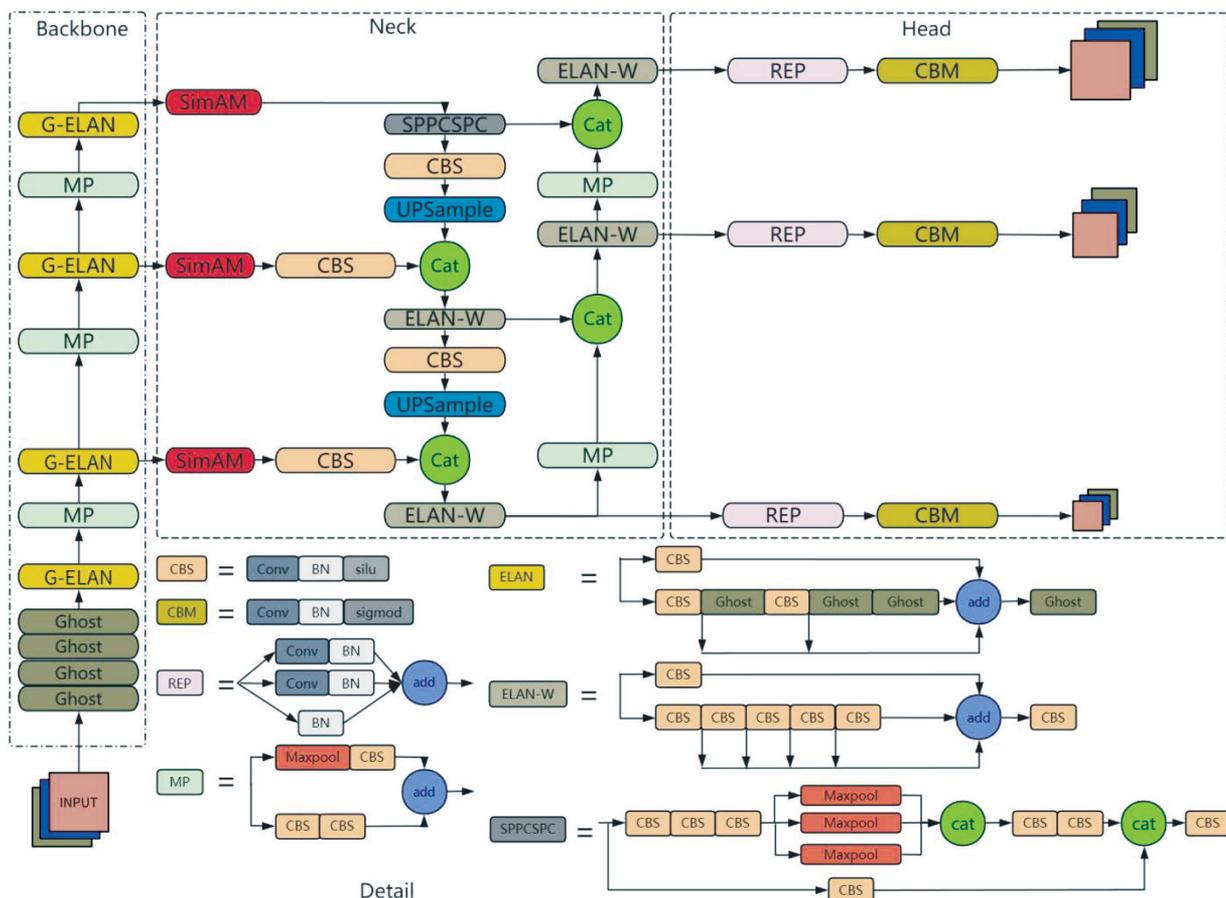
(1) 将主干网络的卷积模块替换成更为轻量化的 GhostConv 模块，在维持原有精度的前提下降低模型的参数量从而提高模型的检测速度。

(2) 对于经过骨干特征提取网络输出的特征采用 SimAM 注意力机制模块进行猕猴桃特征信息的进一步提取，在网络在不添加额外模型计算量的前提下减少噪声信息对于检测结果的影响。使模型能够更加专注于小目标重要特征信息的提取，从而更高效地完成特定的特征提取与重建，提升网络检测精度，降低网络计算量。

(3) 将改建模型中原有的 NMS 极大值抑制策略替换为对高密度高重叠的目标更为敏感的非极

大值抑制 NMS（Soft-NMS），有效减少同类别检测框大量重叠时造成的漏检情况，有效提升检测精确度。

1.2.1 GhostConv 卷积模块^[8] 在实际的模型应用场景中，神经网络模型的计算能力和运算速度常常会受到实际的硬件环境以及设备空间布局设计的限制而受到影响。为适应网络模型在自动化采摘机械嵌入式设备中的应用，引入专门为移动设备设计的 GhostConv 网络结构，其结构核心是利用线性操作来生成丰富的特征图。在原始的网络结构中，提取图片特征的过程中会产生大量冗余，极大占用硬件存储空间，降低模型运算速度，实时性较差，不符合快速识别定位的需求。为实现网络的轻量化，使用深度可分离卷积 GhostConv 模块替代原有的普通卷积层。



注：相较于原 YOLOv7 模型，该改进模型替换了 Backbone 层的卷积模块为 ghostconv 卷积模块，并在 Backbone 层与 Neck 层三处信息交汇处添加 SimAM 注意力机制模块。

Note: Compared with the original YOLOv7 model, the improved model replaces the convolution module of the backbone layer with the ghostconv convolution module, and adds the SimAM attention mechanism module at the intersection of the three information intersections of the backbone layer and the neck layer.

图 1 YOLOv7 改进模型
Fig. 1 YOLOv7 improved model

传统的卷积特征提取中会存在大量的冗余，如图 2-1 所示，GhostConv 之所以能减少参数量和计算复杂度，是因为舍弃了部分常规卷积，采用线性变换进行特征提取，能够更高效地借助特征之间的相关性和冗余性。通过常规卷积 Conv 获取到的特征层内，有一些特征图的特征信息较为相似，因此可以通过消除这些冗余的特征信息来减少获取特征信息的操作的计算量。GhostConv 首先

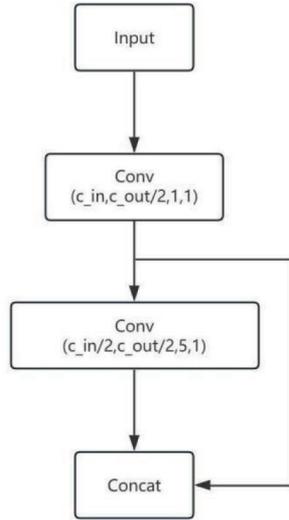


图 2-1 GhostConv 模块结构
Fig.2-1 GhostConv module structure

通过常规卷积提取特征信息，获取到一般特征图，然后将一般特征图中的每一个通道的特征做线性运算，获取到同等通道数的附带特征图，Ghost 特征图就是冗余的特征图数量，通过简单的线性变换得到的 Ghost 特征图，能够在几乎不影响检测精度的同时有效降低原本卷积模块 Conv 的冗余性以及计算复杂度；最后堆叠基础特征图和 Ghost 特征图作为输出。

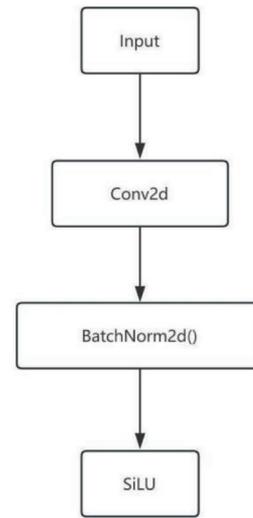


图 2-2 Conv 模块结构
Fig.2-2 Conv module structure

图 2 GhostConv, Conv 卷积模块结构示意图

Fig. 2 GhostConv, Conv convolutional module structure diagram

1.2.2 SimAM 注意力机制模块^[9] SimAM 注意力机制是中山大学在 2021 年提出的一种概念简单且非常有效的注意力机制模块。现有的通道注意力机制例如 SE (Squeeze-and-excitation networks) 注意力机制^[10] 或空间注意力机制还有融合空间和通道的 CBAM (Convolutional Block Attention Module) 注意力机制^[11] 模块。这些注意力机制模块都是从特征 X 中生成一维通道或二维空间权重，并扩展通道或空间注意的权重。而 SimAM 注意力机制的独特优势在于无需增加额外参数，即可直接为特征图推导出三维权重（即考虑空间和通道维度），使网络学习更多有辨别性的神经元以提高网络的特征提取能力。具体来说，作者基于神经科学理论提出了优化能量函数以挖掘神经元的重要性，从而增强重要特征的有效提取，有效抑制非重要特征的干扰。同时，该注意力机制的大部分操作均基于优化能量函数选择，避免过多的结构调整，加速注意力权值的计算，从而在确保 SimAM 注意力机制融入模型后能够保持网络原

本期望的轻量化的同时，更好地发挥注意力机制 SimAM 的有效性和灵活性。

由于本文所使用的数据集为实际田间环境下的猕猴桃数据集，包含环境背景等非重要特征的干扰，在识别过程中会伴随许多噪声信息。这些噪声信息也会在网络模型的学习过程中传递，随着网络层数的增加，噪声信息在特征图的权重也在不断增加，对模型造成负面影响。因此，本文在 YOLOv7 模型中引入注意力机制 SimAM，在特征融合网络在未对获取信息进行融合前添加 SimAM 注意力机制模块。如图 3 所示，它使得网络在不添加额外模型参数的前提下，更加专注于小目标重要特征信息的提取，有利于抑制噪声信息对网络的干扰，从而更高效地完成特定的特征提取与重建，提升网络识别精度，降低网络复杂度。

1.2.3 非极大值抑制 NMS (Soft NMS)^[12] 自然条件下对猕猴桃果实识别在某种意义上属于高密度目标检测，尤其是大规模种植的复杂环境下还可能

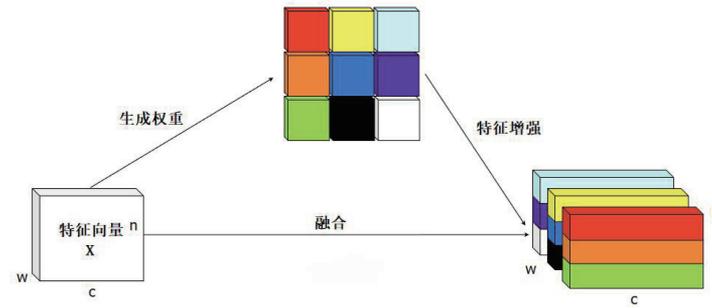


图3 SimAM注意力机制
Fig. 3 SimAM attention mechanism

预测框，通过 IoU（Intersection over Union，交并比）来抑制冗余的预测框，而重叠部分会使筛选存在错误。为保证召回率候选框输出有时会有多个，多余的候选框会影响检测精度，因此需要利用 NMS 过滤掉多余的候选框进而得到最佳的预测输出。非极大值抑制 NMS 方法简单且效果明显，对检测的结果也起到重要作用，因此在检测算法中有着非常广泛的应用。目前几种常见的 NMS 有基本 NMS、Soft NMS、Softer NMS^[13]、Iou-Net^[14] 和 DIoU-NMS^[15]（Distance-IoU-Non-maximum suppression）等多种方法。其中基本 NMS 方法是利用得分高的边框抑制得分低且重叠度高的边框。然而基本 NMS 存在一些缺陷，当两个物体较近重叠时，简单地过滤掉得分低且重叠度高的边框可能会导致漏检。

针对以上问题陆续产生一系列改进方法，Soft NMS 方法相比较基本 NMS 做了如下改进：Soft NMS 中对于 IoU 交并比大于阈值的边框并没有和基本 NMS 那样将其得分直接置为 0，而是降低边框的得分，具体公式如下：^[11]

$$S_i = \begin{cases} S_i, & \text{IoU}(M, b_i) < N_i \\ S_i(1 - \text{IoU}(M, b_i)), & \text{IoU}(M, b_i) \geq N_i \end{cases} \quad (1)$$

从公式中可以看出通过利用边框得分，与 IoU 交并比共同来确定新的边框得分，如果当前边框与边框 B 的 IoU 交并比超过阈值 T，边框得分线性衰减。但是上述公式是非连续函数，即当某个边框与边框 B 的 IoU 交并比超过阈值时会产生跳跃，进而对检测结果造成影响，所以进一步对 Soft NMS 进行改进，最终 Soft NMS 计算公式如下：

$$S_i = S_i e^{-\frac{\text{IoU}(M, b_i)^2}{\sigma}}, \quad \forall b_i \in D \quad (2)$$

采用这种高斯加权方式对于某些得分很高的边框在后续计算中可能被作为正确检测框。NMS 与 Soft NMS 都使用预测分类置信度作为衡量指标，即置信度越高位置也更准确。本文用 Soft NMS 替代原有 NMS，大大减少同类检测框重叠时的漏检情况，有效提升检测精确度。

1.3 数据集建立

1.3.1 图像采集 本文使用的数据集为不同日照环境下采集的猕猴桃图像，采集于湖南省长沙市开福区昱涵生态农业发展有限公司下的猕猴桃种植基地，采集时间为 2023 年 8 月 19 日（阴天）上午 8:00—11:30 时和 8 月 19 日（晴天），14:30—17:30 时，猕猴桃品种为中华红心猕猴桃（成熟）。采集设备为 Canon EOS 6D Mark II 6D2 全幅单反相机，配备 Canon RF24-70 mm 微单镜头，拍摄图像大小为 3 648×2 432 像素，图像采集过程中模拟机器人实际采摘路径，在果树单边位置，从外围、下部距离猕猴桃果实 0.4~0.6 m 处进行拍摄。为增强图像的多样性，采集果实未遮挡和相互遮挡状态以及顺光和逆光下的果实。通过人工筛选去掉图像效果较差的样本后，共采集 1 251 幅图像，包含未遮挡果实 601 幅，果实重叠 650 幅，其中逆光 556 幅，顺光 695 幅。采集的图像如图 4 所示。

1.3.2 数据集图像预处理 猕猴桃数据集中包含采集到的 1 251 张自然条件下不同光照强度不同遮挡程度下的猕猴桃图像，为方便模型的训练，将图片压缩至 640×640 像素，并保存为 JPG 格式。通过随机翻转旋转扩充到 2 502 张（图 5），其中 2 000 张作为训练集，252 张为验证集，250 张为测试集。并随机给 600 张图片添加高斯噪声和椒盐噪声模拟实际的恶劣拍摄环境（图 5）。在该数据集上进行模型的训练并验证算法的



图 4-1 未遮挡果实图像
Fig.4-1 Unobstructed fruit image

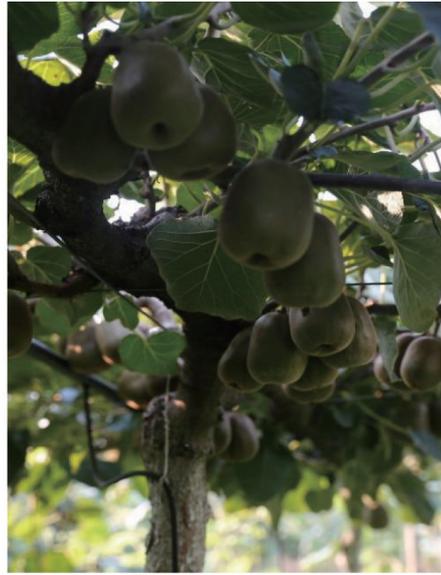


图 4-2 有遮挡果实图像
Fig.4-2 Images of obstructed fruits



图 4-3 顺光图像
Fig.4-3 Smooth image

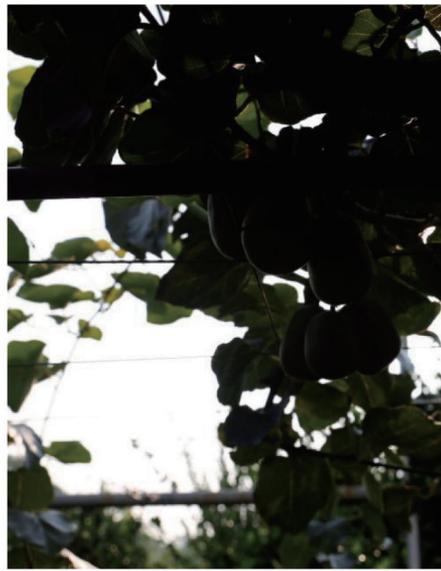


图 4-4 逆光图像
Fig.4-4 Backlit images

图 4 自然条件下猕猴桃图像
Fig. 4 *Actinidia chinensis* fruit images under natural conditions

通用性。采用 Labelimg 标注工具对样本图像进行分类标注，标注格式为 PASCAL VOC 2007，并对标注结果进行校正，减少因标注错误引入的无关标注。

1.4 试验方法

1.4.1 试验环境 本论文采取 Pytorch 框架作为深度学习框架进行训练。开源深度学习框架 Pytorch 是一款轻型深度学习框架，该框架基于 Python 与 CUDA 开发，相较于其他框架更容易安装，可移植性高，同时支持 CPU 和 GPU 两种计算方

式，被广泛应用于计算机深度学习目标检测领域。与普通的算法流程相比，深度学习具有更复杂的网络结构，对网络的学习、参数的调整和权重的更新需要海量的数据进行学习，因此深度学习模型对硬件的要求更加严格。通常而言，研究人员多使用具有高带宽、多核心且能够进行多线程并行计算等优势的 GPU 显卡来完成深度学习的相关计算。本文采用显存 8 G 的 NVIDIA GeForce 3060 TI 显卡作为运算核心的 PC 端实验平台，具体实验环境配置如表 1。



图 5-1 原图
Fig.5-1 Original image



图 5-2 旋转
Fig.5-2 Rotation



图 5-3 翻转
Fig.5-3 Flipping



图 5-4 添加噪声
Fig.5-4 Addnoise

图 5 数据增强
Fig. 5 Data augmentation

表 1 实验环境配置

Table 1 Experimental environment configuration

配置 Allocation	版本 Version
CPU	I5-12400F
GPU	GeForce 3060TI
系统	Window 10
语言	Python 3.9
训练框架	Pytorch1.12.1

1.4.2 评价指标 使用不同的评价指标可以客观准确地评估网络的检测性能, 采摘识别要求识别速度快、准确率高, 抗环境干扰能力强, 部署模型灵活, 本研究从不同角度对模型识别效果进行评价, 分别是以下 4 种关键指标, 分别介绍如下:

(1) mAP (mean Average Precision): 反映多个类别 AP (Average Precision) 的平均准确率均值, AP 是 Precision-recall 曲线下方的面积, 通常

情况下, 该值越大, 模型就更好。计算公式为:

$$AP = \int_0^1 p(r) dr \quad (3)$$

(2) Precision: 反映模型检测准确程度, 计算公式为:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

其中 TP (True Positive) 表示待检目标被正确的检测, FP (False Positive) 表示非待检目标被视为正确的检测。

(3) Recall: 反映模型检测准确程度, 计算公式为:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

其中 TP 表示待检目标被正确的检测到的数据, FN (False Negative) 表示被误认为是负例但实际上是正例的数据。

(4) Fps (Frames Per Second): 反映模型推

理速度的刷新率，计算公式为：

$$Fps = \frac{\text{Frameum}}{\text{Elapsed Time}} \quad (6)$$

其中 Frameum 表示检测图片总数，Elapsed Time 表示检测所花费的总时间，每秒帧率指每秒内可以处理的图片数量，用于评估目标检测模型的预测速度。

2 结果与分析

2.1 模型改进消融实验

由表 2 可以看出，各部分改进的采用均对模型有积极影响。相比于原始 YOLOv7 模型，引入 SimAM 注意力机制对检测精度的提升最为明显，平均精度均值提升 1.1 个百分点，但对于检测速度有不好影响，检测速度降低 7.7 FPS。引入 GhostConv 结构对检测速度的提升最为明显的同时检测精度不降反而有提升。检测速度提高 38.9 Fps，平均精度均值也提升 0.8 个百分点。最后使

用 Soft NMS 算法提高模型对高密度检测目标的识别度，模型精度提升，整体的训练效率更高，模型性能发挥更好。

2.2 引入不同注意力机制对模型检测结果的影响

为了研究注意力模块的引入对猕猴桃识别精度的提升是否有效，本文基于相同数据集进行 4 组对比实验，即改进后的 YOLOv7 模型与分别引入通道注意力模块 SE、混合注意力模块 CBAM 以及注意力模块 SimAM 模型进行比较。从表 3 可见，相比于原改进模型，分别引入混合注意力模块 CBAM 以及注意力模块 SimAM 模型的网络模型识别准确率提高 0.2、0.7 个百分点，引入 SE 注意力模块反而降低 0.9 个百分点；实验结果表明注意力模块 SE 和 CBAM 都增加网络模型的参数导致运算量增大，使得模型检测速度都有不同程度的降低，但注意力模块 SimAM 不仅保持模型轻量化，在达到更高的识别准确率的同时还对模型检测速度的影响较少。

表 2 改进算法与原始算法评价指标对比

Table 2 Comparison of evaluation indicators between improved algorithms and original algorithms

算法模型 Algorithm model	mAP/%	Precision/%	Recall/%	Fps/s
YOLOv7	96.4	93.7	92.1	102.3
YOLOv7+GhostConv	97.2	94.3	91.2	141.2
YOLOv7+GhostConv+SimAM	98.3	95.1	91.7	133.5
YOLOv7+GhostConv+SimAM+SoftNMS	98.8	95.4	92.3	131.8

表 3 相同位置添加不同注意力机制的评价指标结果对比

Table 3 Comparison of evaluation indicators for adding different attention mechanisms at the same position

算法模型 Algorithm model	mAP/%	Precision/%	Recall/%	Fps/s
YOLOv7+SoftNMS+GhostConv	98.1	94.3	93.1	132.4
YOLOv7+SoftNMS+GhostConv+SimAM	98.8	95.4	92.3	131.8
YOLOv7+SoftNMS+GhostConv+SE	97.2	93.9	92.1	115.2
YOLOv7+SoftNMS+GhostConv+CBAM	98.3	95.6	90.1	113.1

2.3 改进模型与其他目标检测模型对比结果与分析

对比表 4 中实验结果可得，本文算法相比于其他主流检测模型，有第二快的检测速度，同时保持了最高的检测精度。相比于原始 YOLOv7 模型检测精度提高 2.4 个百分点，比 YOLOv5s 高 5.5 个百分点，比 YOLOv8m 高 3.7 个百分点。与

传统的 two-stage 算法 Faster RCNN 相比具有较大的检测速度优势，检测精度也有极大提升，与原始模型检测速度相比提升 28.4 个百分点。综上所述，本文提出的改进模型在保持轻量化的同时，有着最高的检测精度，保持较好实时性，整体表现较为突出，证明本文算法的可行性与优越性。

表4 不同算法的评价指标结果对比
Table 4 Comparison of evaluation index results for different algorithms

算法模型 Algorithm model	mAP/%	Precision/%	Recall/%	Fps/s
Faster RCNN	86.1	84.7	82.1	25.0
YOLOv5s	93.3	91.9	89.2	144.9
YOLOv7	96.4	93.7	92.1	103.4
YOLOv8m	95.1	93.1	90.4	74.2
改进模型	98.8	95.4	92.3	131.8



图6 不同模型在部分测试图片上的检测结果

Fig. 6 Detection results of different models on part of the test image

本文对改进后算法的效果在不同光照条件的场景下进行对比展示，各模型在部分测试图片上的检测结果见图6。

作为双阶段目标算法，Faster-RCNN模型对于远距离小目标识别精度较差，并且对于颜色和

果实相近的树枝和树叶有误检情况；YOLOv5s模型鲁棒性较差，少量远距离小目标和大面积遮挡目标无法识别；未经改善的YOLOv7模型对于高密度果实存在漏检情况；YOLOv8m模型的检测准确度和最优模型相比差距较小，但检测速度较慢，

无法满足实际要求; 而改进模型在能够满足实际检测速度的同时能够正确识别到不同光照场景下小目标以及密集目标, 有较好的鲁棒性; 目标置信度最高。符合实验预期效果。

3 结论与讨论

本文针对自然环境下的成熟猕猴桃的识别检测, 依据它实时性和准确性的要求, 提出一种基于改进 YOLOv7 的猕猴桃轻量级实时算法改进, 通过将主干网络的卷积替换为轻量化的 Ghostconv 模块, 引入注意力机制 SimAM 并将原始 NMS 改为 Soft-NMS, 有效解决原始模型检测速度不高以及高重叠目标漏检问题, 这对实时监测采摘, 提高采摘效率具有重要意义。

通过实验认证改进后的模型在实时性上具有一定的优势, 其平均精度 (mAP) 达到 98.8%, 相比原始 YOLOv7 模型提高 2.4 个百分点。在维持高精度的同时参数量大大降低, 优于现有的主流算法, 具有一定的推广价值。但模型的参数量和计算量仍有一定的进步空间, 后期研究将逐步优化改进模型, 提升精确率; 同时扩充数据集, 增加更为复杂的训练图片以提升模型的鲁棒性。

参考文献

[1] 秦志伟. 科技助力猕猴桃产业“熟了”[N]. 中国科学报, 2018-09-05(8).

[2] 慕军营, 陈军, 孙高杰. 基于机器视觉的猕猴桃特征参数提取[J]. 农机化研究, 2014, 36(6): 138-142.

[3] 朱智惟, 单建华, 余贤海. 基于 YOLOv5s 的番茄采摘机器人目标检测技术[J]. 传感器与微系统, 2023, 42(6): 129-132.

[4] 薛勇, 王立扬, 张瑜. 基于 GoogLeNet 深度迁移学习的苹果缺陷检测方法[J]. 农业机械学报, 2020, 51(7): 30-35.

[5] 刘媛. 基于深度学习的葡萄叶片病害识别方法研究

[D]. 兰州: 甘肃农业大学, 2018.

[6] GAO J, DAI S, HUANG J. Kiwifruit Detection Method in Orchard via an Improved Light-Weight YOLOv4[J]. Agronomy, 2022, 12(9): 2081.

[7] 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, Vancouver: IEEE, 2023: 7464-7475.

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

[9] YANG L, ZHANG R Y, LI L. Simam: A simple, parameter-free attention module for convolutional neural networks[C]//International conference on machine learning, Vienna, Austria: ICML, 2021: 11863-11874.

[10] 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.

[11] WOO S, PARK J, LEE J Y. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV), Cham, Switzerland: Springer, 2018: 3-19.

[12] BODLA N, SINGH B, CHELLAPPA R. Soft-NMS-improving object detection with one line of code[C]//Proceedings of the IEEE international conference on computer vision Venice, Italy: (ICCV, 2017: 5561-5569.

[13] HE Y, ZHANG X, SAVVIDES M, et al. Softer-nms: Rethinking bounding box regression for accurate object detection. arXiv 2018 [J]. arXiv preprint arXiv: 1809.08545.

[14] JIANG B, LUO R, MAO J. Acquisition of localization confidence for accurate object detection[C]//Proceedings of the European conference on computer vision (ECCV), Cham, Switzerland: Springer, 2018: 784-799.

[15] ZHENG Z, WANG P, LIU W. Distance-IoU loss: Faster and better learning for bounding box regression[C]//Proceedings of the AAAI conference on artificial intelligence, Palo Alto, California USA: AAAI Press, 2020: 12993-13000.