

基于轻量化 Mamba-YOLO 模型的梨表面缺陷检测方法

修贤超¹, 费士祺^{1,2,3}, 黄文倩^{2,3}, 李楠^{1*}, 苗中华¹

(1. 上海大学机电工程与自动化学院, 上海 200444, 中国; 2. 北京市农林科学院智能装备技术研究中心, 北京 100097, 中国; 3. 北京市农林科学院信息技术研究中心, 北京 100097, 中国)

摘要: [目的/意义] 针对当前砀山梨表面缺陷因尺度小而导致检测精度差的问题, 本研究提出了一种基于改进 Mamba-YOLO 的轻量化高精度模型, 旨在实现检测精度与效率的平衡。[方法] 首先, 采用动态上采样模块, 相较于现有 Mamba-YOLO 的上采样模块具有更少的参数量和浮点运算次数, 可在保障模型计算效率的同时, 提升对缺陷细节信息的保留能力。其次, 提出频率自适应空洞卷积, 通过动态调整卷积核尺寸, 使网络依据输入局部特征自适应选择匹配的卷积核, 从而增强对缺陷的特征提取能力。最后, 融合压缩和激励模块和通道混合器卷积门控线性单元, 同时引入多尺寸卷积核提取多尺度特征, 进一步提升模型对局部细节的捕捉能力与鲁棒性。[结果和讨论] 改进后的算法在砀山梨测试集上经过评估, 平均精度均值达到了 95.1%, 帧率达到了 72 帧/s。与 YOLOv8n、Gold-YOLO-N 和 YOLOv12n 相比, 平均精度均值分别高出了 4.7、5.3 和 6.3 个百分点; 与基准 Mamba-YOLO-T 相比, 平均精度均值提升了 3.4 个百分点, 帧率提高了 10.8 个百分点。[结论] 改进模型在提升综合检测性能的同时降低了计算复杂度与参数量, 可为轻量化梨表面缺陷检测研究提供可靠的算法支撑。

关键词: Mamba-YOLO; 缺陷检测; 图像识别; 频率自适应空洞卷积; 卷积核; 动态上采样模块

中图分类号: TP391

文献标识码: A

文章编号: SA202508022

引用格式: 修贤超, 费士祺, 黄文倩, 李楠, 苗中华. 基于轻量化 Mamba-YOLO 模型的梨表面缺陷检测方法[J]. 智慧农业(中英文), 2026, 8(2): 147-157. DOI: 10.12133/j.smartag.SA202508022

XIU Xianchao, FEI Shiqi, HUANG Wenqian, LI Nan, MIAO Zhonghua. A Lightweight Method for Pear Surface Defect Detection Based on Improved Mamba-YOLO Architecture[J]. Smart Agriculture, 2026, 8(2): 147-157. DOI: 10.12133/j.smartag.SA202508022 (in Chinese with English abstract)

0 引言

梨是一种常见的水果, 富含维生素和矿物质。它能够维持人体细胞的健康状态, 具有丰富的营养价值^[1]。中国作为梨生产与出口大国, 其产量与出口量长期位居世界首位^[2]。传统的梨分级检测主要依靠人工经验, 通过视觉和触觉等感官手段进行判断, 不仅效率较低, 而且会受限于主观因素影响, 导致检测结果的准确性有限。此外, 人工操作处理梨的过程中可能会造成不同程度的物理损伤, 从而影响梨的外观质量和市场价值。因此, 开发自动、高效且可靠的梨分级技术已成为行业迫切的需求。基于计算机视觉与机器学习技术的自动分类方法^[3], 虽然能够提供比人工判断更为客观的检测结

果, 但其效率依旧难以满足工业化流水线日益增长的实时性需求。在此背景下, 开展可靠的梨分级检测研究, 保障产品质量并增强市场竞争力, 成了重要的研究方向。

随着神经网络和深度学习技术的快速发展, 当前相关的分级研究已经从传统算法转向基于深度学习的目标检测算法^[4-6], 如 YOLO 系列模型^[7-9]。KARKI 等^[10]利用迁移学习在 ImageNet 数据集上对 VGG19 (Visual Geometry Group 19)、Inception V3、ResNet50 (Residual Network with 50 layers) 和 DenseNet121 (Densely Connected Convolutional Network with 121 layers) 等架构进行预训练, 并采用微调 and 特征提取技术来检测草莓病害, 在

收稿日期: 2025-08-21

基金项目: 国家重点研发计划项目 (2024YFB4707400); 上海市重点科技攻关项目 (24N32800100)

作者简介: 修贤超, 博士, 副教授, 研究方向为人工智能与具身智能。E-mail: xcxiu@shu.edu.cn

*通信作者: 李楠, 博士, 讲师, 研究方向为智能装备与机器人技术。E-mail: linan2019@shu.edu.cn

copyright©2026 by the authors

(登录 www.smartag.net.cn 免费获取电子版全文)

ResNet50 架构上取得了高达 94.4% 的准确率。然而, ResNet50 作为特征提取网络所构成的双阶段检测模型不仅需要更多计算资源, 而且训练时间过长, 不适合实际部署。

YOLO 系列模型由于网络结构简单、推理速度快, 非常适合实时监控和轻量级部署, 在农业领域得到了广泛的应用。例如, 周宏平等^[11]提出了一种基于迁移学习和 YOLOv8n 算法的油茶果分类识别方法, 有效降低了视觉引导式油茶果采摘机器人采摘被遮挡油茶果时造成的果树和抓取装置损伤。陈俊霖等^[12]提出了一种基于轻量化 YOLOv8s 的草莓穴盘苗分级识别和定位方法, 有效克服了穴盘苗越界生长带来的识别和定位干扰。黎祖胜等^[13]针对自然环境中荔枝虫害识别效率低的问题, 提出一种基于改进 YOLOv10n 的轻量化目标检测模型 YOLO-LP (YOLO-Litchi Pests), 实现对小目标的高效聚焦, 增强目标与背景的区分能力, 同时减少参数和计算量, 为荔枝虫害检测的实际应用提供了有效的参考。针对采后芦笋在销售前人工分级成本高、效率低的问题, 杨启良等^[14]提出了一种基于改进 YOLOv11 模型的采后芦笋分级方法, 并引入了 ECAM (Efficient Channel Attention Module)^[15]以提高网络的特征提取能力。HU 等^[16]通过集成注意力机制优化了 YOLOv5^[17]模型来提高柑橘表面缺陷的检测准确性, 改进算法的平均精度均值相较于基准模型提高了 5.8 个百分点。谭厚森等^[18]针对非结构化环境下香梨识别准确率低、检测速度慢的问题, 在 YOLOv8 模型的基础上, 通过使用 PConv (Partial Convolution)^[19]模块替换部分冗余的 C2f (CSPDarknet53 to 2-Stage FPN)^[20]模块以减少模型的参数量, 同时引入 simSPPF (Simplified Spatial Pooling Fast) 和 Inner IOU (Intersection Over Union)^[21]来进一步提高模型的推理速度与性能。

综上所述, 深度学习在农作物缺陷检测领域已经取得了显著的成果^[22]。但是, 目前关于梨缺陷检测的研究仍有较大的改进空间。例如双阶段模型参数量庞大和推理速度缓慢, YOLO 模型对梨的小目标缺陷检测精度不足, 这些严重影响了实际生产线的性能和效率^[22]。因此, 本研究基于 Mamba^[23]与 YOLO 相结合的 Mamba-YOLO^[24]模型, 引入了卷积门控线性单元 (Convolutional Gated Linear Unit, CGLU)^[25]模块、动态上采样 (Dynamic Upsampling, Dysample)^[26]模块和频率自适应空洞卷积 (Frequency-Adaptive Dilated Convolution,

FADC)^[27]模块, 提出了一种新的梨表面缺陷检测方法, 提升综合检测性能的同时降低了计算复杂度与参数量。

1 数据获取与预处理

1.1 图像采集设备

试验平台为北京市农林科学院智能装备技术研究中心自主研发的果托式多表面品质检测分级生产线。该生产线由输送果托、可编程逻辑控制器 (Programmable Logic Controller, PLC) 控制单元、分辨率为 $1\,920 \times 1\,080$ 的彩色工业相机 (acA1920-40gc, Basler, 德国)、带状 LED 光源、漫反射光箱, 以及包含分选软件的工业计算机等部分。梨由输送果托输送至光箱检测区域, 通过机器视觉系统采集梨图像, 经过工业计算机中的分选软件检测梨是否含有缺陷, 并将检测结果发送至 PLC 控制单元, 最后在卸料区完成梨的分选。采集系统示意图如图 1 所示。

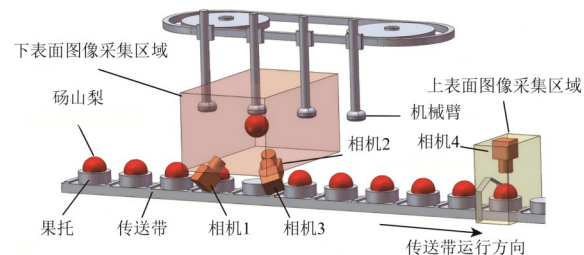


图 1 梨图像采集系统的示意图

Fig. 1 Schematic diagram of pear image acquisition system

1.2 数据采集

本研究采用的品种为安徽砀山梨, 每张图像为同一个梨的四个不同角度。利用 Labelimg 标记图像中的缺陷区域框来构建训练数据集, 花萼的标签为“calyx”, 紫盖缺陷的标签为“cap”, 锈斑缺陷的标签为“rust spot”, 霉斑的标签为“mold”。其中紫盖缺陷为果梗及其周围因日灼导致的褐色放射状纹路; 锈斑缺陷多为果柄附近的不规则斑块; 霉斑为区域性黑色斑点, 表面带有些许绒毛。花萼虽不是表面缺陷, 但标注并让模型学习后可以避免干扰其他缺陷的识别, 从而降低模型的误检率。图 2 为砀山梨的缺陷图像。

为了提升模型训练的精度与泛化能力, 人工剔除质量差的数据集图像。本数据集共有 1 000 张图像, 按 8 : 1 : 1 划分为训练集、验证集与测试集。此外, 为了增强模型的训练效果, 对数据集进行了

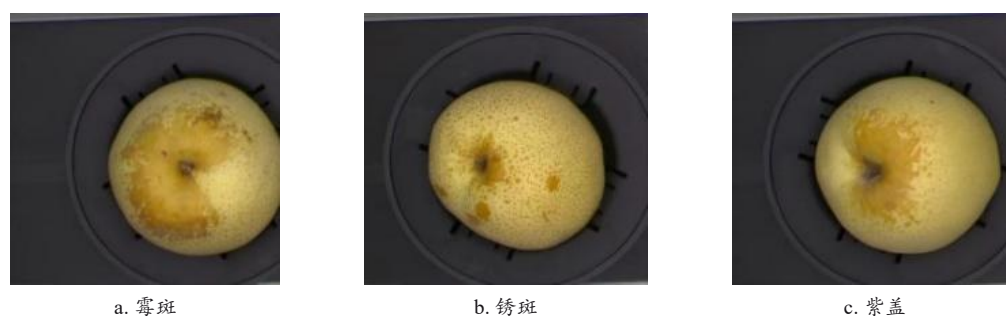


图2 砀山梨缺陷的类别图

Fig. 2 Classification diagram of dangshan pear defects

数据增强，包括旋转、裁剪、镜像、改变亮度等操作。图3为部分梨数据集图像。

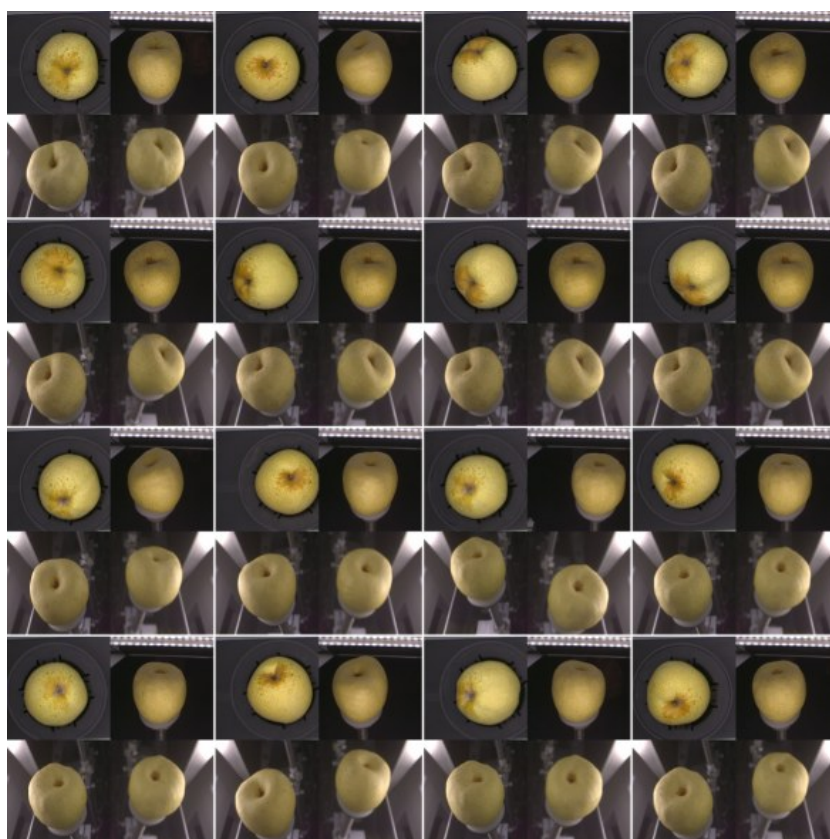


图3 部分梨数据集的图像

Fig. 3 Partial images of the pear dataset

2 研究方法

2.1 Mamba-YOLO 模型结构

在计算机视觉领域，Vision Mamba^[28]首次将状态空间模型^[29]（State Space Model, SSM）引入视觉任务，提出了一种基于SSM的视觉主干模型。该模型通过交叉扫描模块对图像进行选择扫描，从而增强了视觉特征的处理能力。

Mamba-YOLO模型是一种以YOLOv8为基础模型并采用SSM的创新架构。得益于Mamba的计

算复杂度随输入大小线性增长的优势，Mamba-YOLO模型可有效解决YOLO模型因引入自注意力机制而导致计算开销过大的问题，因此更加适用于目标检测任务。

与YOLO模型结构相似，Mamba-YOLO模型由骨干网络、颈部层、头部检测层三部分组成。首先输入层对需要检测的图像进行预处理，然后输入骨干网络层进行特征提取，将其分为三个不同尺寸的特征图。随后把特征图输入颈部层处理，将得到的三个不同大小的特征图进行预测，最后输出目标

检测的结果。

2.2 Mamba-YOLO 模型改进

为了提高 Mamba-YOLO 模型的检测精度, 本研究提出了一种有效的改进算法, 记为 Mamba-YOLO-FC。将 Mamba-YOLO 模型中的 ODSS 模块替换为 FCSS (Feature Convolutional Structured) 模块, 改进了 ODSS 模块中的 LS (Local Spatial) 模

块与 RG (Res Gated) 模块, 旨在能够更好地捕捉局部特征, 进而提升模型小目标检测的性能, 同时提高模型的鲁棒性, 减少噪声和扰动的影响。将颈部层的 Upsample 模块替换为 Dysample 模块, 不仅减少了模型的参数量和浮点运算次数, 同时提升了模型的检测精度。本研究提出的梨表面缺陷检测模型 Mamba-YOLO-FC 的结构如图 4 所示。

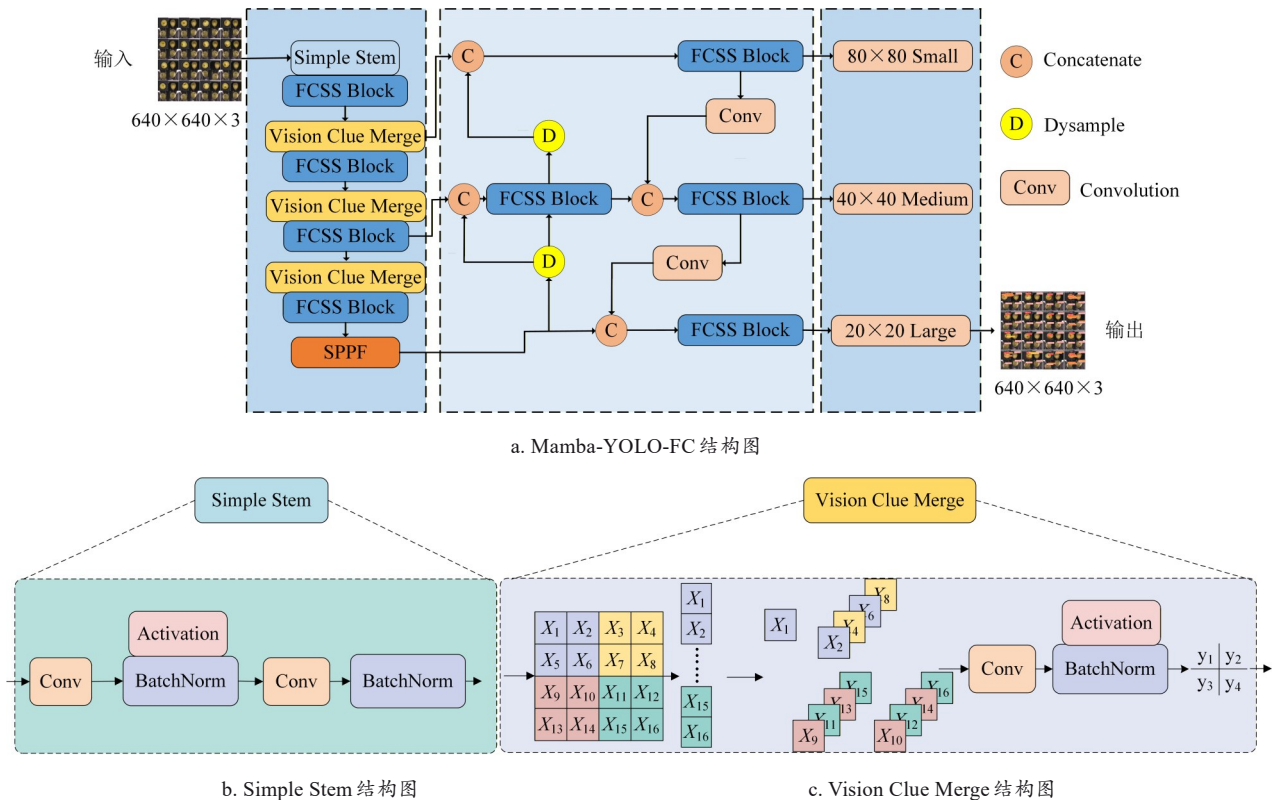


图 4 Mamba-YOLO-FC 整体架构图

Fig. 4 Illustration of the proposed Mamba-YOLO-FC

2.2.1 Dysample 上采样器

上采样器是目标检测领域中的关键技术之一。它能够通过恢复特征图的分辨率, 帮助模型在复杂场景中实现精确定位与多尺度目标检测。具体而言, 上采样器通过多尺度特征融合机制, 如 FPN (Feature Pyramid Network) 中的“自上而下”结构, 使高层语义信息与底层细节特征互补, 以此提升模型的检测精度与鲁棒性。但是在砀山梨的表面缺陷检测实验中发现, 许多小目标缺陷信息并没有被捕捉, 特征没有得到充分的提取, 这对砀山梨缺陷检测产生了一定的影响。

于是, 本研究引入了一种轻量级且高效的动态上采样器, 即 Dysample。尽管内容感知特征重组 (Content-Aware ReAssembly of FEatures, CA-

RAFE)^[30]、特征对齐与蒸馏增强 (Feature Alignment and Distillation Enhancement, FADE)^[31] 和尺度感知金字塔注意力 (Scale-Aware Pyramid Attention, SAPA)^[32] 等基于内核的动态上采样器显著提升了检测精度, 但因为耗时的动态内核中额外子网络导致检测效率大幅降低。针对梨表面缺陷实时检测所需精度与效率兼顾问题, 在本研究提出的 Mamba-YOLO-FC 模型中, Dysample 模块通过动态因子自适应调整偏移量的幅度, 进而增强网络的灵活性, 如公式 (1) 所示。

$$O = 0.5 \times \text{sigmoid}(\text{linear}(X)) \times \text{linear}(X) \quad (1)$$

式中: X 是输入的特征图; linear 表示线性投影; sigmoid 为函数。将输入特征图以通道维度分为多组, 每组独立生成偏移量, 单独计算后合并再

用于重采样。将偏移量 O 加到原始网格 G 上，生成动态采样点 S ，如公式 (2) 所示。

$$S = G + O \quad (2)$$

随后使用 PyTorch 的 `grid_sample` 函数进行双线性插值采样， X' 为输出的特征图，如公式 (3) 所示。

$$X' = \text{grid_sample}(X, S) \quad (3)$$

Dysample 通过动态生成采样点的偏移量，结合

双线性插值的重采样机制，实现了高效且轻量的上采样。其设计摒弃了复杂的动态卷积核，仅需少量线性层和分组操作，即可兼顾高精度和低计算开销。

2.2.2 频率自适应空洞卷积

Mamba-YOLO 模型为了增强特征的提取能力，提出了使用 ODSS 模块来代替 C2f 模块。在此基础上，本研究改进 ODSS 模块为 FCSS 模块，如图 5a 所示。

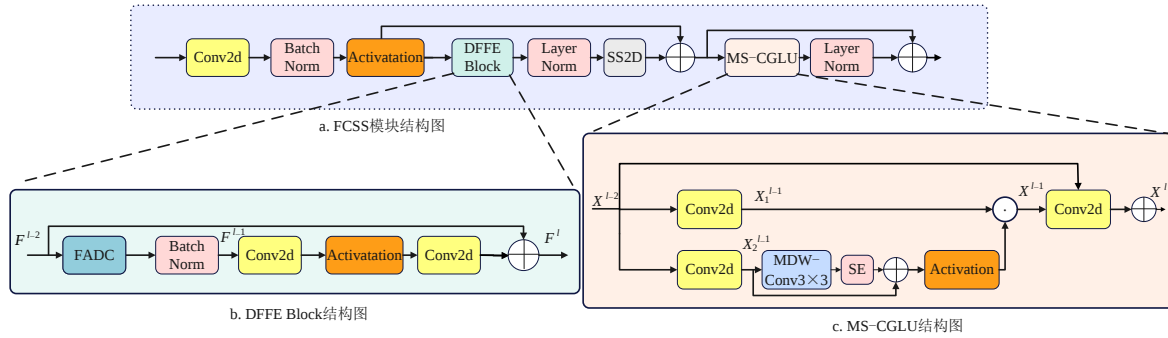


图 5 FCSS 模块整体架构图

Fig. 5 Illustration of the FCSS Block architecture

在梨表面缺陷检测中，缺陷往往表现为高频局部特征，而传统卷积神经网络存在特征捕捉不足与频率响应失衡的问题。随着扩张率增大，卷积核频率响应降低，带宽变窄，进而限制了对高频信息的处理能力。为此，改进 FADC 模块，通过动态调整卷积核的大小，使其在高频区域采用小的卷积核，在低频区域采用大的卷积核，实现多频段特征的协同优化，如图 5b 所示。

具体来说，首先将给定的输入特征 F^{l-2} 进行频率自适应空洞卷积，有效提取输入特征图的局部空间信息。然后进行批处理的归一化，减少过拟合的同时还有一定程度的正则化效果，得到中间状态 F^{l-1} ，如公式 (4) 所示。

$$F^{l-1} = \text{BN}(\text{FADC}(F^{l-2})) \quad (4)$$

式中： BN 为批归一化。中间状态 F^{l-1} 进行 1×1 卷积，混合了通道信息，可以从输入的特征图中获得更多的信息。然后通过激活函数 GeLU (Gaussian Error Linear Unit) 增强特征的信息，如公式 (5) 所示。

$$F^l = \text{Conv}_{1 \times 1}(\Phi(\text{Conv}_{1 \times 1}(F^{l-1}))) \oplus F^{l-2} \quad (5)$$

式中： Φ 为非线性函数 GeLU； $\oplus$ 表示残差拼接的元素加法； $\text{Conv}_{1 \times 1}$ 为 1×1 卷积层。最后将处理后的信息与残差网络的信息融合，以此使模型

具有更好的鲁棒性。

2.2.3 卷积门控线性单元

最初的 MLP (Multi-Layer Perceptron) 结构仍是目前最广泛采用的全连接层，它在各种视觉任务中表现突出，但面临信息混合不足的问题。针对这一问题，引入了结合卷积和门控线性单元 (Gated Linear Unit, GLU) 的通道混合器 CGLU 模块。

在 CGLU 的基础上，考虑仅通过单尺度的 3×3 深度卷积捕获局部特征，可能对输入特征的信息感知不足。为此，引入了多尺度机制，这样不同尺度的深度卷积可以实现更强的特征表达能力。同时，传统门控机制存在全局上下文信息建模不足的问题。针对这一局限性，提出了一种改进方案：采用融合通道注意力机制的 CGLU 模块，并将其扩展为多尺度版本的 MS-CGLU (Multi-Scale Convolutional Gated Linear Unit)，替代原始 ODSS 模块中的 RG 模块。通过添加不同尺寸的卷积核获取多尺度特征，再经过加权融合实现更强的特征表达。这一改进策略提升了网络捕获全局上下文信息和局部特征细节的能力，从而提升了模型的表征能力和鲁棒性。改进后的模块结构如图 5c 所示。该模块从输入创建两个分支，记为 X_1^{l-1} 与 X_2^{l-1} ，并分别对其进行 1×1 卷积，如公式 (6) 和公式 (7) 所示。

$$X_1^{l-1} = \text{Conv}_{1 \times 1}(X^{l-2}) \quad (6)$$

$$X_2^{l-1} = \text{Conv}_{1 \times 1}(X^{l-2}) \quad (7)$$

将 X_2^{l-1} 进行多重深度卷积增强特征表达能力处理后, 将残差拼接得到的 X_2^{l-1} 与 X_1^{l-1} 进行门控机制的元素乘法, 如公式 (8) 所示。

$$X^{l-1} = X_1^{l-1} \odot \Phi(\text{MDWConv}_{3 \times 3}(X_2^{l-1}) \oplus X_2^{l-1}) \quad (8)$$

式中: $\odot$ 为门控机制的元素乘法; MDWConv 为多尺度深度卷积。通过 1×1 的卷积与全局特征进行细化, 混合通道信息。最后通过残差拼接得到的原始输入 X^{l-2} 的进行求和, 如公式 (9) 所示。

$$X^l = \text{Conv}_{1 \times 1}(X^{l-1}) \oplus X^{l-2} \quad (9)$$

该模块可以获得更丰富的全局特征信息, 且不会带来过多的计算量。

3 结果与分析

3.1 训练方法与环境

实验环境的硬件和软件配置如下: GPU 为 NVIDIA GeForce RTX 4090, CPU 为 AMD EPYC 9354, CUDA 11.8.89, Python 3.10, PyTorch 2.0.1。输入图像像素尺寸为 640×640 , 初始学习率为 0.01, 优化器选择为 Adam, Batchsize 为 16, Epoch 为 300。所有模型均在相同的硬件环境和初始化条件下进行, 旨在消除因配置差异引入的偏差, 从而确保实验结论的有效性。

3.2 评价指标

模型性能的评估体系通常包含检测效率与检测精度。在检测效率方面, 模型参数量反映了可训练参数的总数, 直接影响内存占用; GFLOPs (十亿次浮点运算) 量化了前向推理的计算复杂度, 数值

越高, 表示模型对硬件要求越高。在检测精度方面, F_1 值 (F_1 -Score) 是一个综合评估模型分类性能的重要指标, 它结合了精确率 (Precision, P) 和召回率 (Recall, R), 如公式 (10) ~ 公式 (12) 所示。

$$P = \text{TP} / (\text{TP} + \text{FP}) \quad (10)$$

$$R = \text{TP} / (\text{TP} + \text{FN}) \quad (11)$$

$$F_1 = 2PR / (P + R) \quad (12)$$

式中: TP 为检测正确的正样本数量; FP 为检测错误的正样本数量; FN 为检测错误的负样本数量。平均精度均值 (Mean Average Precision, mAP) 是 AP (Average Precision) 的平均值。通常来说, mAP 值越高, 模型检测效果越好。计算 AP 与 mAP 如公式 (13) 和公式 (14) 所示。

$$\text{AP} = \int_0^1 P(R) dR \quad (13)$$

$$\text{mAP} = \frac{\sum_{i=1}^C \text{AP}_i}{C} \quad (14)$$

式中: C 为检测类别数, 实验中为花萼、霉斑、紫盖和锈斑, 因此共有 $C = 4$ 类。

3.3 消融实验分析

根据表 1 的消融实验结果可以看出, B+Dysample (Baseline+Dysample) 模型在引入 Dysample 模块后帧率 (Frames Per Second, FPS) 提高 53.8%, F_1 显著上升, $\text{mAP}_{0.5}$ 提高 1.2 个百分点。虽然精确率略微下降, 但是召回率上升, 说明模型的泛化能力在增强。由于 F_1 是精确率和召回率的调和平均值, 其上升说明召回率的提升幅度显著高于精确率的下降幅度。这表明, Dysample 模块在精度不低于其他高精度动态上采样器的情况下, 推理速度提升, 实现了模型检测效率的优化。

表 1 梨表面缺陷检测研究消融实验的结果

Table 1 Results of ablation experiments for pear surface defect detection research

模型	FADC	CGLU	Dysample	精确率/%	召回率/%	$\text{mAP}_{0.5}/\%$	$\text{mAP}_{0.5,0.95}/\%$	F_1 值/%	FPS/(帧/s)
B	×	×	×	95.7	83.9	91.7	53.2	89.4	65
B+Dysample	×	×	√	92.6	92.9	92.9	53.5	92.7	100
B+FADC	√	×	×	93.1	90.4	93.1	54.7	91.7	56
B+CGLU	×	√	×	97.1	88.4	92.2	53.7	92.5	38
B+CGLU+Dysample	×	√	√	95.5	87.4	93.8	53.9	91.3	51
B+FADC+Dysample	√	×	√	88.4	92.0	92.3	54.0	90.2	57
B+FADC+CGLU	√	√	×	93.7	89.2	93.0	55.2	91.4	41
B+FADC+CGLU+Dysample	√	√	√	95.1	91.1	95.1	56.6	93.1	72

注: Mamba-YOLO-T 模型为 Baseline 模型 (简称为 B)。√ 表示引入该模块, × 表示未引入该模块。

B+FADC模型相较于B模型， $mAP_{0.5}$ 提高1.4个百分点， $mAP_{0.5:0.95}$ 提高1.2个百分点，FPS下降13.8%， F_1 提高2.3个百分点。B+CGLU模型单独引入CGLU模块， mAP 值有少量增长，FPS大幅下降41.5%，但精确率和召回率都有显著提升， F_1 值提高3.1个百分点。B+CGLU+Dysample和B+FADC+Dysample模型情况相似， $mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别提高2.1、0.6和0.7、0.8个百分点，FPS下降27%和12.3%。

B+FADC+CGLU模型同时引入CGLU模块与FADC模块， $mAP_{0.5}$ 提高1.3个百分点， $mAP_{0.5:0.95}$ 提高2个百分点， F_1 提高2个百分点，FPS下降37%。可以看出，CGLU与FADC对于精度和模型的泛化性与鲁棒性有明显的提升，但由于引入更加复杂的卷积层，模型的计算量大幅增加，FPS下降。面对需要实时性的梨表面缺陷检测任务，推理速度也是不能忽略的一环。

B+FADC+CGLU+Dysample模型通过引入CGLU、FADC和Dysample三个模块，实现了检测性能的全面提升。在检测精度方面， $mAP_{0.5}$ 达到95.1%， $mAP_{0.5:0.95}$ 达到56.6%，较B模型提高3.4个百分点；在推理效率方面，达到72帧/s，比B模型提高10.8%。

从上述实验中可以看出，在梨表面缺陷检测任务中，单独引入某一模块并不能提高模型的性能，甚至会导致性能的降低。这意味着在模型改进时需要进行多方面的考量，不同的模块是否兼容，这样才能根据需求提升模型的效果。本研究提出的Mamba-YOLO-FC模型在工业化流水线的梨缺陷检测与分级分拣任务中，各项指标均满足实际需求。

3.4 与其他模型对比

本研究在梨的缺陷检测数据集上对Mamba-YOLO-FC模型与YOLO系列模型和不同规模的Mamba-YOLO模型分别进行了训练，表2和表3展示了对比结果。

实验结果表明，本研究提出的Mamba-YOLO-FC在模型性能方面表现良好，以少量参数量与计算量的增加，获得了 mAP 的大幅提升。与YOLOv5n、YOLOv6n、YOLOv7-tiny和YOLOv8n模型相比， $mAP_{0.5}$ 分别有8.3、4.8、4.4和4.7个百分点的提高，与当前目标检测领域表现优异的Gold-YOLO-N模型相比，展现出更优的性能，其 $mAP_{0.5}$

表2 梨表面缺陷检测研究YOLO系列模型对比的实验结果
Table 2 Experimental results of YOLO-series models for pear surface defect detection research

模型	$mAP_{0.5}/\%$	$mAP_{0.5:0.95}/\%$	参数量/M	计算量/GFLOPs
YOLOv5n	86.8	49.1	1.9	4.5
YOLOv6n	90.3	48.4	4.7	4.7
YOLOv7-tiny	90.7	49.3	6.2	13.7
YOLOv8n	90.4	50.5	3.2	34.1
Gold-YOLO-N	89.8	51.0	5.6	12.1
YOLOv12n	88.8	51.9	2.6	6.5
Mamba-YOLO-FC	95.1	56.6	9.1	28.4

表3 不同规模的Mamba-YOLO模型对比的实验结果
Table 3 Experimental results of Mamba-YOLO models at various scales

模型	$mAP_{0.5}/\%$	$mAP_{0.5:0.95}/\%$	参数量/M	计算量/GFLOPs
Mamba-YOLO-T	91.7	53.2	6.1	14.3
Mamba-YOLO-B	91.9	53.5	21.8	49.7
Mamba-YOLO-L	94.2	54.0	57.6	156.2
Mamba-YOLO-FC	95.1	56.6	9.1	28.4

和 $mAP_{0.5:0.95}$ 分别提高5.3个和5.6个百分点。即使与YOLO系列目前最新版本的YOLOv12n相比， $mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 仍分别提高6.3个和4.7个百分点。

进一步，与基准模型Mamba-YOLO-T相比，本方法仅增加3 M参数量，却使 $mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 均提高3.4个百分点，实现了精度与效率的良好平衡。此外，在与Mamba-YOLO同系列更大规模的模型对比实验中，Mamba-YOLO-FC仍然表现出显著的优势：在参数量仅为Mamba-YOLO-B的41.7%和Mamba-YOLO-L的15.7%的情况下，计算量分别仅为二者的57.1%和18.1%，而 $mAP_{0.5}$ 却分别提高3.2和1.4个百分点， $mAP_{0.5:0.95}$ 分别提高3.1和2.6个百分点。

综合以上实验结果，Mamba-YOLO-FC模型在保持轻量级架构的同时实现了检测精度的显著提升。该模型在参数量、计算复杂度和检测精度三者之间得到了平衡，同步优化了效率和性能，满足梨表面缺陷检测任务对模型性能和实时性的双重需求，具有重要的实际应用价值。

3.5 不同的特征提取模块对比

为了验证本研究提出的FCSS模块在梨表面缺

陷检测任务中的性能，将FCSS模块和传统YOLO模型主要使用的C2f模块、C3模块和基准模型使用的ODSS模块分别作为模型的特征提取模块，在同

一数据集上进行训练，再将各项指标进行对比，结果如表4所示。

表4 梨表面缺陷检测研究改进特征提取模块的实验结果

Table 4 Performance of the enhanced feature extraction module for pear surface defect detection research

特征提取模块	精确率/%	召回率/%	mAP _{0.5} /%	mAP _{0.5:0.95} /%	F ₁ 值/%	计算量/GFLOPs
C2f	90.9	93.6	93.4	54.5	92.2	33.1
C3	94.0	85.8	92.8	54.3	89.7	23.6
ODSS	95.7	83.9	91.7	53.2	89.4	14.3
FCSS	95.1	91.1	95.1	55.6	93.1	28.4

实验结果表明，提出的FCSS模块在检测性能上展现出显著的优势。FCSS模块相较于传统YOLO模型常用的C2f和C3模块，mAP_{0.5}分别提高1.7和2.3个百分点，mAP_{0.5:0.95}分别提高1.1和1.3个百分点。FCSS模块与基准模型中的ODSS模块相比，mAP_{0.5}和mAP_{0.5:0.95}分别提高3.4和2.4个百分点。

在模型分类性能方面，FCSS模块同样表现出色。实验数据显示，F₁相较于C2f、C3和ODSS模块分别提高0.9、3.4和3.7个百分点，这一结果充分验证了FCSS模块在保持高检测精度的同时，能够有效降低漏检率和误检率，以少量计算量的增加，显著提升了模型的整体性能。

3.6 模型训练损失图

损失曲线如图6所示，其中Bounding box loss表示边界框损失，Classification loss表示分类损失。可以看出Mamba-YOLO-FC模型相较于基准模型Mamba-YOLO-T的损失曲线呈现出更加优秀的收敛趋势。

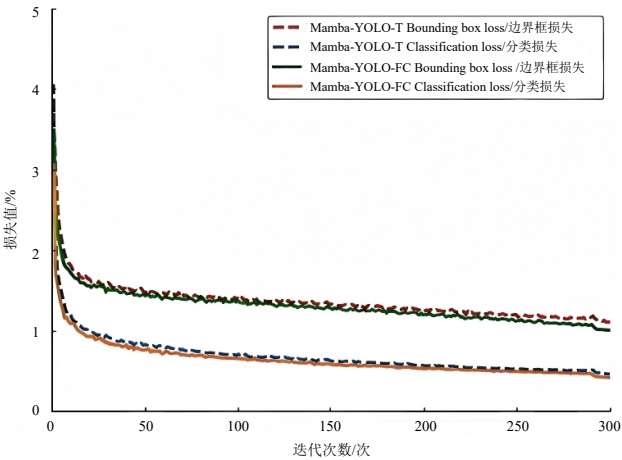


图6 梨表面缺陷检测研究损失的曲线图

Fig. 6 Diagram of loss curves for pear surface defect detection research

具体而言，在前50个训练周期内，损失值呈现快速下降态势，表明模型在此阶段迅速学习到有效的特征表示；随后损失值逐渐稳定，模型得到了收敛。这种收敛特性表明，本研究提出的Mamba-YOLO-FC模型具有良好的学习效率和训练稳定性。损失曲线的平滑过渡和稳定在较低水平，充分验证了模型超参数设置的合理性，以及模型架构优秀的鲁棒性。实验结果表明，Mamba-YOLO-FC模型成功完成了训练过程，为后续对模型性能的评估奠定了基础。

3.7 模型检测效果可视化分析

为了验证模型Mamba-YOLO-FC的在线检测效果，在实际砀山梨数据集上进行了实验，缺陷检测效果图如图7所示。

如图7所示，在不同视角的情况下，模型对锈斑这种易于混淆的缺陷能够有效地检测出来，这反映了模型良好的检测性能。同时在图中也能发现，由于图像采集角度问题，会导致漏检的发生。

3.8 模型检测错误情况分析

在果实缺陷检测中，正常梨样本的误判现象主要源于其花萼结构与缺陷区域在形态特征上的相似性。针对该问题，可通过扩充训练数据集的多样性促进模型对不同区域特征的辨识能力，从而增强分类鲁棒性。而对于缺陷梨样本的识别误差问题，研究发现与小尺寸样本的特征表征不足密切相关。由于梨缺陷尺寸较小，当果实整体尺寸较小时，其表面缺陷区域过小，使模型特征检测不足，进而产生漏检的风险。此外，尺寸较小的梨在摩擦传输带翻滚过程中速度较快，易造成图像采集系统未能有效捕捉到缺陷区域的图像，导致缺陷未能被检测。需要说明的是，在实际生产场景中，尺寸较小的梨会依据尺寸标准作为外果，该类梨果不参与正常等级



注:红色边框标记紫盖区域;粉色边框标记花萼区域;橙色边框标记锈斑区域;黄色边框标记霉斑区域。

图7 Mamba-YOLO-FC模型检测的结果

Fig. 7 Detection results of the Mamba-YOLO-FC model

的分选。同时还应注意,果实的摆放可能会导致图像采集角度问题,如图7中第三行第二列的图所示,这同样会导致漏检的发生。

4 结 论

本研究针对梨表面缺陷目标小、易与花萼区域混淆的问题进行了探讨,提出了一种兼顾检测精度与推理速度的Mamba-YOLO-FC模型,主要结论如下。

1) 首先,设计了新的FCSS特征提取模块来替代基础模型的ODSS模块,将FADC模块和CGLU模块加以改进并引入到FCSS模块中,为模型提供了更好的特征提取能力和更强的鲁棒性。其次,引入了新的上采样模块,使用Dysample模块取代原颈部网络中的Upsample模块,不仅使参数量与浮点运算次数得到了一定程度的下降,而且模型检测效率得到了提升。

2) 所提出的Mamba-YOLO-FC模型与YOLOv5n、YOLOv6n、YOLOv7-tiny、YOLOv8n、Gold-YOLO-N和YOLOv12n模型相比,mAP_{0.5}分别有8.3、4.8、4.4、4.7、5.3和6.3个百分点的提高。与Mamba-YOLO同系列更大规模的模型相比,Mamba-YOLO-FC模型在梨表面缺陷检测任务中依旧能够展现出更高的精度与效率,具备良好的综合性能。

3) 提出的Mamba-YOLO-FC算法精确率达到了95.1%,召回率达到了91.1%,mAP_{0.5}达到了95.1%,mAP_{0.5:0.95}达到了56.6%,帧率达到了72帧/s。因此能够快速精准地检测出缺陷梨,具有良好的稳定性与泛化能力,对梨表面缺陷检测具有积极的意义。

需要指出的是,在砀山梨表面缺陷检测的任务中仍存在以下局限性:1) 数据集规模相对有限,可能影响模型在训练过程中的特征学习充分性,未来需进一步扩充数据以提升模型性能。2) 构建小

目标缺陷数据集,以深入验证模型在小目标检测能力方面的提升效果。

利益冲突声明: 本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献:

- [1] ZHANG H, LAI L S, GU J H, et al. Applications of near-infrared spectroscopy in pear quality assessment: A comprehensive review [J]. *Journal of Food Process Engineering*, 2025, 48(5): e70086.
- [2] 王文辉, 王国平, 田路明, 等. 新中国果树科学研究 70 年: 梨[J]. *果树学报*, 2019, 36(10): 1273-1282.
WANG W H, WANG G P, TIAN L M, et al. Fruit scientific research in New China in the past 70 years: Pear[J]. *Journal of Fruit Science*, 2019, 36(10): 1273-1282.
- [3] RAWAT W, WANG Z H. Deep convolutional neural networks for image classification: a comprehensive review[J]. *Neural Computation*, 2017, 29(9): 2352-2449.
- [4] 蒋雪松, 计恺豪, 姜洪喆, 等. 深度学习在林果品质无损检测中的研究进展[J]. *农业工程学报*, 2024, 40(17): 1-16.
JIANG X S, JI K H, JIANG H Z, et al. Research progress of non-destructive detection of forest fruit quality using deep learning[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2024, 40(17): 1-16.
- [5] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. arXiv: 1409.1556, 2014.
- [6] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2016: 770-778.
- [7] BOCHKOVSKIY A, WANG C Y, LIAO H M. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. arXiv: 2004.10934, 2020.
- [8] LI C Y, LI L L, JIANG H L, et al. YOLOv6: A single-stage object detection framework for industrial applications[EB/OL]. arXiv: 2209.02976, 2022.
- [9] SAPKOTA R, FLORES-CALERO M, QURESHI R, et al. YOLO advances to its genesis: A decadal and comprehensive review of the You Only Look Once (YOLO) series[J]. *Artificial Intelligence Review*, 2025, 58(9): 274.
- [10] KARKI S, BASAK J K, TAMRAKAR N, et al. Strawberry disease detection using transfer learning of deep convolutional neural networks[J]. *Scientia Horticulturae*, 2024, 332: 113241.
- [11] 周宏平, 金寿祥, 周磊, 等. 基于迁移学习与 YOLOv8n 的田间油茶果分类识别[J]. *农业工程学报*, 2023, 39(20): 159-166.
ZHOU H P, JIN S X, ZHOU L, et al. Classification and recognition of *Camellia oleifera* fruit in the field based on transfer learning and YOLOv8n[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2023, 39(20): 159-166.
- [12] 陈俊霖, 赵鹏, 曹先林, 等. 基于通道剪枝的轻量化 YOLOv8s 草莓穴盘苗分级检测与定位方法[J]. *智慧农业(中英文)*, 2024, 6(6): 132-143.
CHEN J L, ZHAO P, CAO X L, et al. Lightweight YOLOv8s-based strawberry plug seedling grading detection and localization via channel pruning[J]. *Smart Agriculture*, 2024, 6(6): 132-143.
- [13] 黎祖胜, 唐吉深, 匡迎春. 基于改进 YOLOv10n 的轻量化荔枝虫害小目标检测模型[J]. *智慧农业(中英文)*, 2025, 7(2): 146-159.
LI Z S, TANG J S, KUANG Y C. A lightweight model for detecting small targets of *Litchi* pests based on improved YOLOv10n [J]. *Smart Agriculture*, 2025, 7(2): 146-159.
- [14] 杨启良, 禹璐, 梁嘉平. 基于改进 YOLOv11 的采后芦笋分级检测方法[J]. *智慧农业(中英文)*, 2025, 7(4): 84-94.
YANG Q L, YU L, LIANG J P. Grading *Asparagus officinalis* L. using improved YOLOv11[J]. *Smart Agriculture*, 2025, 7(4): 84-94.
- [15] LI L T, ZHAO Y D. Tea disease identification based on ECA attention mechanism ResNet50 network[J]. *Frontiers in Plant Science*, 2025, 16: 1489655.
- [16] HU W X, XIONG J T, LIANG J H, et al. A method of *Citrus* epidermis defects detection based on an improved YOLOv5[J]. *Bio-systems Engineering*, 2023, 227: 19-35.
- [17] INBAR O, SHAHAR M, GIDRON J, et al. Analyzing the secondary wastewater-treatment process using Faster R-CNN and YOLOv5 object detection algorithms[J]. *Journal of Cleaner Production*, 2023, 416: 137913.
- [18] 谭厚森, 马文宏, 田原, 等. 基于改进 YOLOv8n 的香梨目标检测方法[J]. *农业工程学报*, 2024, 40(11): 178-185.
TAN H S, MA W H, TIAN Y, et al. Improved YOLOv8n object detection of fragrant pears[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2024, 40(11): 178-185.
- [19] CHEN J R, KAO S H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]// 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2023: 12021-12031.
- [20] TERVEN J, CORDOVA-ESPARZA D M, ROMERO-GONZÁLEZ J A. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS[J]. *Machine Learning and Knowledge Extraction*, 2023, 5(4): 1680-1716.
- [21] ZHANG H, XU C, ZHANG S J. Inner-IoU: More effective intersection over union loss with auxiliary bounding box[EB/OL]. arXiv: 2311.02877, 2023.
- [22] MENGHANI G. Efficient deep learning: A survey on making deep learning models smaller, faster, and better[J]. *ACM Computing Surveys*, 2023, 55(12): 1-37.
- [23] ZHU L H, LIAO B C, ZHANG Q, et al. Vision mamba: Efficient visual representation learning with bidirectional state space model [EB/OL]. arXiv: 2401.09417, 2024.
- [24] WANG Z Y, LI C, XU H Y, et al. Mamba YOLO: A simple baseline for object detection with state space model[EB/OL]. arXiv: 2406.05835, 2024.
- [25] SHI D. TransNeXt: Robust foveal visual perception for vision transformers[C]// 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2024: 17773-17783.
- [26] LIU W Z, LU H, FU H T, et al. Learning to upsample by learning to sample[C]// 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2023: 6004-6014.
- [27] CHEN L W, GU L, ZHENG D Z, et al. Frequency-adaptive dilated convolution for semantic segmentation[C]// 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2024: 3414-3425.
- [28] GU A, GOEL K, RE C. Efficiently modeling long sequences with structured state spaces[EB/OL]. arXiv: 2111.00396, 2021.
- [29] LIU Y, TIAN Y J, ZHAO Y Z, et al. VMamba: Visual state space model[EB/OL]. arXiv: 2401.10166, 2024.
- [30] WANG J Q, CHEN K, XU R, et al. CARAFE: Content-aware Re-Assembly of Features[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2019: 3007-3016.
- [31] LU H, LIU W Z, FU H T, et al. FADE: Fusing the Assets of Decoder and Encoder for Task-Agnostic Upsampling[C]// Computer Vision-ECCV 2022. Cham, Germany: Springer, 2022: 231-247.
- [32] LU H, LIU W Z, YE Z X, et al. SAPA: Similarity-aware point affiliation for feature upsampling[EB/OL]. arXiv: 2209.12866, 2022.

A Lightweight Method for Pear Surface Defect Detection Based on Improved Mamba-YOLO Architecture

XIU Xianchao¹, FEI Shiqi^{1,2,3}, HUANG Wenqian^{2,3}, LI Nan^{1*}, MIAO Zhonghua¹

(1. School of Mechanic Engineering and Automation, Shanghai University, Shanghai 200444, China; 2. Intelligent Equipment Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; 3. Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China)

Abstract: [Objective] Pears are a common fruit rich in vitamins and minerals. Traditional pear grading primarily relies on manual inspection, which is not only laborious but also susceptible to subjective factors, leading to unstable and inaccurate results. Furthermore, manual operations may cause varying degrees of physical damage to pears, affecting their appearance and market value. Therefore, developing an automated, efficient, and reliable pear grading technology has become an urgent demand in the industry. To address the current problem of poor detection accuracy caused by the small scale of surface defects in Dangshan pears, a lightweight high-precision model was proposed based on an improved Mamba-YOLO architecture, aiming to balance detection accuracy and efficiency. [Methods] The dataset comprised 1 000 images, which were partitioned into training, validation, and test sets in an 8:1:1 ratio. The following improvements were made to the network architecture. Firstly, a dynamic upsampling (Dysample) module was adopted. Compared to the existing upsampling module in Mamba-YOLO, the Dysample module featured fewer parameters and floating-point operations (FLOPs). Its design eliminated complex dynamic convolution kernels, requiring only a small number of linear layers and grouping operations, thereby preserving computational efficiency while enhancing the retention of defect details. Secondly, regarding pear surface defect detection, defects often exhibited high-frequency local features, whereas traditional convolutional neural networks (CNNs) suffer from insufficient feature capture and imbalanced frequency response. As the dilation rate increased, the frequency response of the convolution kernel decreased and its bandwidth narrowed, consequently limiting its ability to process high-frequency information. Therefore, a frequency-adaptive dilated convolution (FADC) module was proposed, which dynamically adjusted the convolution kernel size, enabling the network to adaptively select matching kernels based on local input features. Smaller kernels were used in high-frequency regions, and larger kernels in low-frequency regions, thereby achieving collaborative optimization of multi-band features and enhancing the ability to extract defect features. Finally, considering that using only single-scale depthwise convolutions to capture local features might lead to insufficient perception of input feature information, and that traditional gating mechanisms may lack adequate global context information modeling, the squeeze-and-excitation module was fused with a channel mixer based on the convolutional gated linear unit (CGLU). This combination was extended into a multi-scale version termed MS-CGLU. By incorporating convolutional kernels of different sizes to extract multi-scale features, followed by weighted fusion, stronger feature representation was achieved. [Results and Discussions] The proposed method was rigorously evaluated on the dangshan pear test set. Ablation experiments demonstrated that introducing the CGLU, FADC, and Dysample enhanced detection performance, confirming the effectiveness of these modules. Compared to YOLOv8n, Gold-YOLO-N, and YOLOv12n, the mean average precision (mAP) was higher by 4.7, 5.3, and 6.3 percentage points, respectively. Compared to the baseline Mamba-YOLO-T, the mAP increased by 3.4 percentage points and the frames per second improved by 10.8 percentage points. Furthermore, in comparative experiments with larger-scale models from the same Mamba-YOLO series, the proposed algorithm still demonstrated significant advantages, i.e., its parameter count was only 41.7% of Mamba-YOLO-B and 15.7% of Mamba-YOLO-L, and its FLOPs was merely 57.1% and 18.1% of the respective models, yet it achieved increases in mAP@0.5 of 3.2% and 1.4%, and increases in mAP@0.5:0.95 of 3.1% and 2.6%, respectively. [Conclusions] This research developed a high-precision and lightweight algorithm for detecting surface defects on Dangshan pears. It achieved a superior balance between detection accuracy and inference speed, significantly outperforming relevant lightweight benchmarks and even larger models within its own family in terms of efficiency. This work can provide reliable algorithmic support for lightweight detection research of pear surface defects.

Keywords: Mamba-YOLO; defect detection; image recognition; frequency-adaptive dilated convolution (FADC); convolutional kernels; dynamic upsampling (Dysample)

Foundation items: National Key Research and Development Program of China (2024YFB4707400); Shanghai Key Science and Technology Project (24N32800100)

First author: XIU Xianchao, E-mail:xcxiu@shu.edu.cn

***Corresponding author:** LI Nan, E-mail: linan2019@shu.edu.cn