doi:10.6041/j.issn.1000-1298.2025.05.042

基于改进 YOLO 11 的多尺度板栗果实识别方法

李　茂[1,2]　肖洋轶[1,2]　宗望远[1,2]

(1.华中农业大学工学院,武汉 430070;2.农业农村部长江中下游农业装备重点实验室,武汉 430070)

摘要:为解决现阶段自然条件下板栗目标尺度不一带来的检测局限性,本文基于改进 YOLO 11 模型提出一种多尺度板栗果实识别方法 YOLO 11-MCS。提出了多尺度关键特征聚合模块(MKFA),并将其引入 C3k2 模块,构建 C3k2-MKFA 特征提取模块,有效捕捉不同尺度特征信息;提出了 CGAFPN 网络,通过内容引导注意力模块引入小目标检测层,弥补了原生算法在多尺度、小目标检测中存在的不足;提出了共享卷积分离批量归一化检测头(SCSB),采用共享卷积和分离批量归一化结构,实现跨尺度特征高效提取,增强了不同尺度特征一致性。试验结果表明,改进模型板栗识别准确率为 88.2%,召回率为 79.2%,平均精度为 87.2%,相较于原始 YOLO 11s 模型,准确率、召回率、平均精度分别提升 0.8、5.9、5.5 个百分点。采用通道式特征蒸馏后模型平均精度为 84.7%,模型内存占用量为 6.0 MB,经 Infer 推理库在 Jetson Nano 上部署后,检测时间为 23 ms/幅,满足板栗识别要求。

关键词:YOLO 11-MCS;板栗;多尺度;目标识别;知识蒸馏

中图分类号:TP391.4;S24　　**文献标识码**:A　　**文章编号**:1000-1298(2025)05-0443-12　　**OSID**:

Multi-scale Chestnut Detection Method Based on Improved YOLO 11 Model

LI Mao[1,2]　XIAO Yangyi[1,2]　ZONG Wangyuan[1,2]

(1. *College of Engineering*, *Huazhong Agricultural University*, *Wuhan* 430070, *China*

2. *Key Laboratory of Agriculture Equipment in Mid-lower Yangtze River*, *Ministry of Agriculture and Rural Affairs*, *Wuhan* 430070, *China*)

Abstract: Aiming to address the current limitations in detecting chestnut of varying scales under natural conditions, an innovative multi-scale chestnut detection method was introduced, YOLO 11-MCS, based on an improved YOLO 11 model. Firstly, a novel multi-scale key feature aggregation (MKFA) module was proposed, which was integrated into the C3k2 module to form the C3k2-MKFA feature extraction module, effectively capturing features at different scales, enhancing multi-scale feature extraction capabilities. Subsequently, the CGAFPN network was introduced, which incorporated a small object detection layer through a content-guided attention module and increased the contribution proportion of chestnut small object to multi-scale object, overcoming the deficiencies of the original algorithm in multi-scale and small object detection. Finally, a shared convolution separated batch normalization detection head (SCSB) was presented, utilizing shared convolution and separated batch normalization structures to efficiently extract cross-scale features and enhance feature consistency across different scales, effectively improved the performance of multi-scale object detection. Experimental results demonstrated that the improved model achieved a chestnut detection precision of 88.2%, a recall rate of 79.2%, and an average precision of 87.2%, which had improvements of 0.8, 5.9, and 5.5 percentage points, respectively, compared with the original YOLO 11 network. The model with channel-wise feature distillation achieved an average precision of 84.7%, with a model size of 6.0 MB. When deployed on the Jetson Nano using the Infer inference library, the detection speed was 23 ms per image, meeting the requirements for chestnut detection.

Key words: YOLO 11-MCS; chestnut; multi-scale; object detection; knowledge distillation

收稿日期:2024-12-11　修回日期:2025-01-22

基金项目:湖北省重点研发计划项目(2020BED027)和湖北省自然科学基金项目(2023AFB871)

作者简介:李茂(1995—),男,博士生,主要从事现代农业电气化与自动化研究,E-mail:leemone@webmail.hzau.edu.cn

通信作者:宗望远(1969—),男,教授,博士生导师,主要从事现代农业装备设计与测控研究,E-mail:zwy@mail.hzau.edu.cn

0 引言

板栗作为一种重要经济作物,在世界各地均有种植[1],具有丰富的营养价值和经济价值[2]。我国作为板栗重要种植区域,板栗种植面积和产量均位于世界第一[3]。

传统板栗检测识别主要在板栗分选、分级等固定环境,姜洪喆等[4]采用可见光/近红外光谱检测分析,通过采集光谱并对比不同处理方法分别建立产地与霉变的偏最小二乘判别模型,交叉验证后产地模型预测集准确率为 97.12%、霉变模型预测集判别准确率为 94.74%。MARTINEZ 等[5]利用太赫兹时域高光谱成像(THz-TDHIS)与无监督学习相结合的方法识别霉变板栗,采用主成分分析、K 均值聚类和凝聚聚类,区分板栗的霉变区域,结果表明,不仅可以标记霉变区域,还能量化板栗感染程度。展慧等[6]利用 BP 神经网络方法建立了板栗分级模型。试验结果表明,识别正确率达到 91.67%。周竹等[7]采用近红外漫反射法采集合格板栗和霉变板栗光谱,比较板栗近红外光谱在 6 种预处理方法下所建模型识别正确率。结果表明,经矢量归一化预处理所建模型识别效果最好。

随着深度学习发展,板栗目标识别方法逐渐向更为复杂的卷积神经网络模型转变[8-9]。深度学习技术能够克服传统方法局限性,在特征提取和模式识别方面具有显著优势。

随着深度学习方法不断突破和计算能力不断提升,在农业领域使用场景更加多样化[10-11],涌现出越来越多的检测模型,其中基于 Transformer 的 RT-DETR[12]和 YOLO 系列模型具有较高的检测性能,YOLO 系列模型因部署体积小,检测精度高广泛用于农业目标识别领域,通过将目标检测问题转换为回归问题,能够在单个前向传播过程中同时预测多个目标类别和位置,显著提升了检测速度。ARAKAWA 等[13]使用 YOLO v4 对板栗果实进行计数,用 UAV 收集的 RGB 图像训练模型,与人工计数相比,R^2 为 0.98,均方根误差为 6.3,结果表明,YOLO 算法结合无人机 RGB 影像是一种精确、高效的树上栗果检测方法。YOLO 模型兼容改进特性展示了在农业应用领域的潜力,也为进一步算法优化提供了基础。

在农业智能化背景下,目标识别逐渐从单一目标、单一场景的需求转向环境复杂、目标尺度不一的应用场景[14-15]。YOLO 系列模型在多尺度检测存在检测不足的劣势[16],许多研究者开始关注 YOLO 模型多尺度检测技术改进,旨在提高对不同大小目标的识别能力[17-19]。通过多尺度聚合、优化特征融合结构、加强不同尺度特征共享等方法可有效提升 YOLO 模型在多尺度特征的检测能力。

本文基于 YOLO 11[20]提出一种多尺度复杂背景下板栗目标识别方法 YOLO 11-MCS。利用多尺度特征聚合模块(Multi-scale key feature aggregation, MKFA)进行特征提取,并构建 C3k2-MKFA 模块,有效捕捉不同尺度特征信息;引导注意力融合网络(Content guides attention FPN, CGAFPN)实现更优的特征融合,通过引导网络关注更重要的特征,提高模型准确性;提出共享卷积分离批量归一化检测头(Shared convolution separation batch normalize, SCSB),采用共享卷积和耦合头的检测头结构,以提升检测速度和精度,减少计算复杂度;最后,通过通道级知识蒸馏(Channel-wise distllation, CWD)以有效减小模型权重,使算法在边缘设备上也能实现高效运行。

1 材料与方法

1.1 数据集建立

自然条件下多尺度板栗图像于 2022 年 9 月 2—8 日(08:00—18:00)在湖北省板栗主产区黄冈市罗田县和红安县、孝感市安陆市、荆门市京山市等地采集。2022 年 9 月 23 日、2023 年 9 月 15 日和 2024 年 9 月 24 日,在华中农业大学板栗种植园采集部分数据。采集设备为 Canon 200D 型和 Fuji X-T20 型相机。由于自然环境背景复杂性和板栗果实目标尺度多样性,采集过程中选择不同光线条件、拍摄时间、拍摄角度、拍摄距离等,进一步增加目标数据多样性。同时,由于 YOLO 系列模型在小目标检测中存在不足,增加数据集中板栗小目标比例。采集的不同尺度目标图像如图 1 所示。



(a) 小目标 (b) 中目标 (c) 大目标

图 1 不同尺度板栗果实采集图像

Fig. 1 Acquisition images of chestnut fruits at different scales

板栗图像采用 Makesense[21]标注目标标签,生成 txt 标签文件,构建 YOLO 数据集。多尺度板栗数据集按照比例 8∶1∶1 划分训练集、验证集、测试集。数据集中小中大目标标签量与分布如表 1 所示。

板栗目标小中大分类按标签框面积与原图面积比值界定,小目标比值小于 0.002 5,中目标比值为 0.002 5~0.022 5 之间,大目标比值大于 0.225。板栗目标标签与原图宽高比值分布如图 2 所示。

表 1　多尺度板栗目标数据集构成

Tab. 1　Composition of multi-scale chestnut target dataset

类别	图像数量	标签数量	小目标	中目标	大目标
训练集	3 615	59 586	43 326	11 493	4 767
验证集	452	7 303	5 429	1 411	463
测试集	452	6 322	4 529	1 326	467

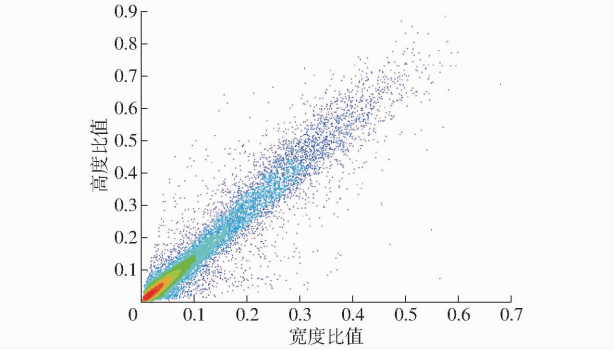

图 2　板栗目标尺度分布图

Fig. 2　Target scale distribution of chestnut

1.2　多尺度板栗目标识别方法

1.2.1　网络模型改进

YOLO 11 为沿袭 YOLO v5[22]、YOLO v8[23]的架构基础上的最新一代实时目标检测器。相较于 YOLO v8，YOLO 11 主要改进有：①用 C3k2 模块替换 C2f 模块，通过引入更多的卷积层和注意力机制，显著增强了模型对特征的提取能力，增强了模型对复杂场景的适应性。②在 SPPF 后面增加类注意力机制 C2PSA 层，能够自适应的聚焦于重要特征，提升模型特征表达能力。③利用深度可分离卷积（DWConv）[24]替换检测头内部类别检测卷积，减小模型计算量的同时保持跨通道信息流动。④调整模型深度和宽度参数，进一步优化模型结构，平衡模型检测精度和检测速度。

在 YOLO 11s 模型基础上，提出多尺度关键特征聚合模块 MKFA，改进 C3k2 模块，构建 C3k2－MKFA 模块，提高板栗目标多尺度特征聚合能力；提出 CGAFPN 结构，通过内容引导注意力机制引小尺度目标检测层，进一步提高多尺度目标识别能力；提出 SCSB 检测头，采用共享卷积和分离批量归一化结构，多尺度结果级联组合，实现跨层级特征融合，增强不同尺度特征提取能力。改进 YOLO 11－MCS 结构如图 3 所示。图中 Conv 为标准卷积，DWConv 为可分离卷积层，SPPF 为空间金字塔池化模块，Upsample 为上采样模块；Contact 为特征连接模块，CGAFusion 为内容引导注意力融合模块，SCSB Detect 为共享卷积分离批量归一化检测头，MKFA 为多尺度关键信息融合机制。

1.2.1.1　多尺度关键特征聚合模块

自然条件下板栗小目标信息较多，小目标识别依赖于细粒度特征细节信息和丰富的语义信息。YOLO 11 网络中 C3k2 模块中采用多个 C3k 模块提取特征信息，每个 C3k 模块采用 2 个并行卷积和多层卷积，随着卷积层数逐渐增多，特征图细节逐渐加深。浅层特征图尺寸较大，分辨率较高，含有更多的目标位置，细节信息，经过多层卷积，特征提取逐渐加深，深层特征图具有更强的语义信息，但特征图尺寸较小，分辨率较低，含有特征细节信息减少。小目标在特征图中所占比例较小，自身特征不足以分辨位置信息，对上下文信息依赖性更高，C3k2 结构在跨层多尺度特征融合上有局限性。

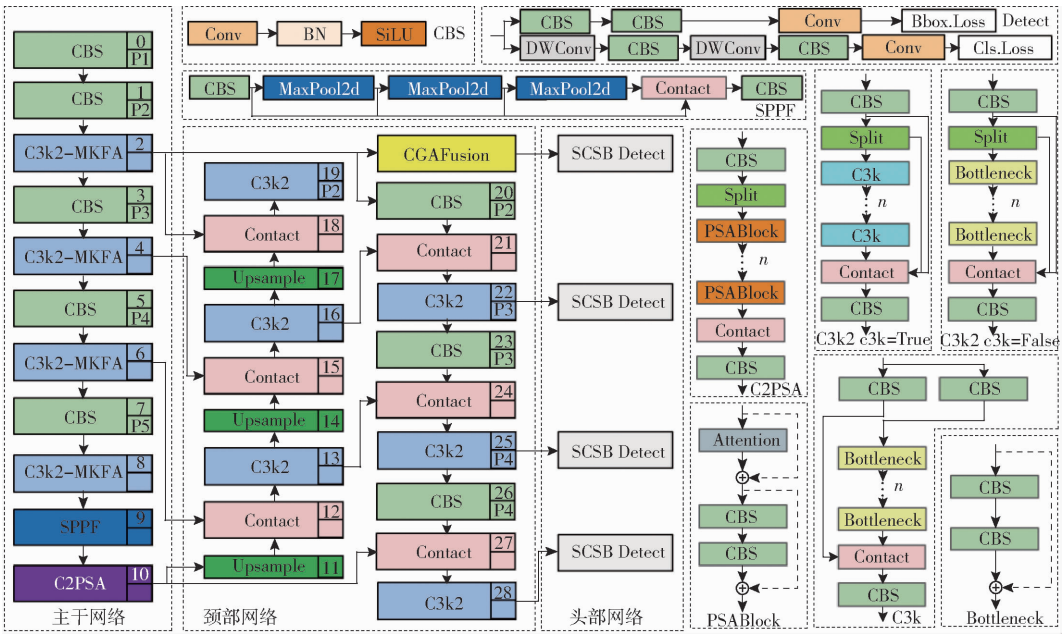

图 3　YOLO 11－MCS 网络结构图

Fig. 3　YOLO 11－MCS network structure diagram

同时特征图中低阶通道特征和高阶通道信息具有较高的相似性,在卷积过程中全部参与会造成资源浪费和计算冗余,图 4 为同一特征图沿通道提取的特征可视化。

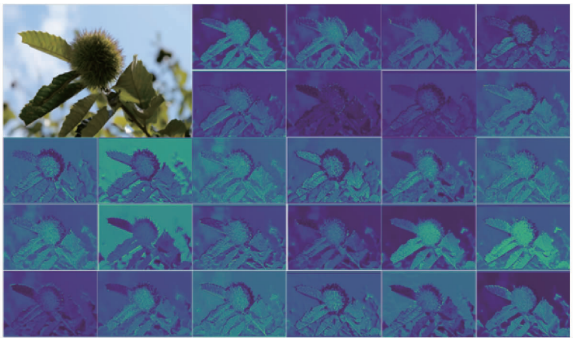

图 4 同一特征图沿通道提取的可视化

Fig. 4 Visualization of features extracted along channel from the same feature map

根据部分卷积[25]和多尺度特征聚合[26]理念,提出 MKFA 多尺度关键特征聚合模块,结构如图 5 所示。

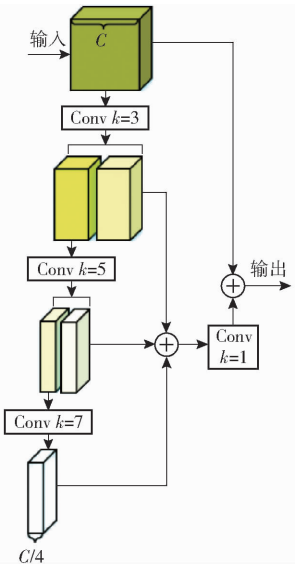

图 5 MKFA 结构图

Fig. 5 MKFA structure

模块定义 4 个卷积层,特征图 X 先经过 3×3 卷积输出特征图 X_1,按照通道分割为浅层特征图 X_1' 和深层特征图 X_1'' 2 个部分,X_1' 经过 5×5 卷积输出特征图 X_2,同样分割为 X_2' 和深层特征图 X_2'' 两个部分,X_2' 经过 7×7 卷积层输出特征图 X_3,将经过多次卷积的浅层特征信息 X_3 和含有深层语义信息 X_1''、X_2'' 进行拼接,通过 1×1 卷积后输出特征图 X'。

MKFA 表达式为

$$X_1', X_1'' = Split(f_{Conv}^{3\times3}(X)) \tag{1}$$

$$X_2', X_2'' = Split(f_{Conv}^{5\times5}(X_1')) \tag{2}$$

$$X_3 = f_{Conv}^{7\times7}(X_2') \tag{3}$$

$$X' = Cat(X, f_{Conv}^{1\times1}(Cat(X_1'', X_2'', X_3'))) \tag{4}$$

在 YOLO 11 原有的 C3k2 模块中嵌入 MKFA 多尺度特征提取模块,构建 C3k2 - MKFA 模块,结构如图 6 所示。

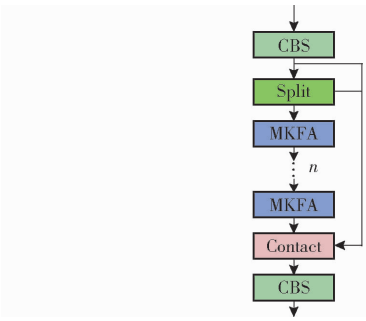

图 6 C3k2 - MKAF 结构

Fig. 6 C3k2 - MKAF structure

改进 C3k2 - MKFA 模块可以缩减卷积过程计算开支,提取板栗目标关键特征,同时通过引入多尺度特征,有效提取到不同尺度特征,增强模型对上下文信息捕捉能力,提升模型对多尺度板栗目标检测性能。

1.2.1.2 CGAFPN 特征融合网络

原始 YOLO 11 颈部网络主要关注较高层次的特征,低层次特征引入不足,对板栗小目标特征提取能力不足,缺乏对小尺度特征图综合处理能力,无法有效捕捉板栗目标上下文信息,同时在不同尺度特征融合上不充分,没有更深入的跨尺度信息交互,自然条件下多尺度板栗目标检测效果不佳。

提出的 CGAFPN 网络,引入小尺度特征图检测层,使用 CGA 模块优化特征信息组合,更好地整合全局和局部信息。内容引导注意力机制(Content-guided attention,CGA)[27]是一种由粗粒度信息到细粒度信息的处理过程,生成粗略的空间注意力图,根据输入通道进行细化,产生最终的空间注意力图,使用输入特征内容来引导注意力图生成,从而更加关注每个通道的唯一特征,可以更好地重新校准特征,学习通道特定的注意力图,关注通道间板栗目标分布差异,充分混合通道注意权重和空间注意权重,保证信息交互。CGA 模块结构如图 7 所示。

由图 7 可得,接收输入特征 $X \in R^{C\times H\times W_1}$,生成与输入特征尺寸相同的空间重要性图(Spatial important map, SIM),通道注意 SIMs 和空间注意 SIMs 计算式为

$$W_c = f_{Conv}^{1\times1}(\max(0, f_{Conv}^{1\times1}(X_{GAP}^{C}))) \tag{5}$$

$$W_s = f_{Conv}^{7\times7}([X_{GAP}^{S}, X_{GMP}^{S}]) \tag{6}$$

式中 $f_{Conv}^{k\times k}$——卷积核为 $k\times k$ 卷积层

$\max(0,x)$——ReLU 激活函数

$[\,]$——通道级连接操作

X_{GAP}^{C}——跨空间维度全局平均池化处理特征

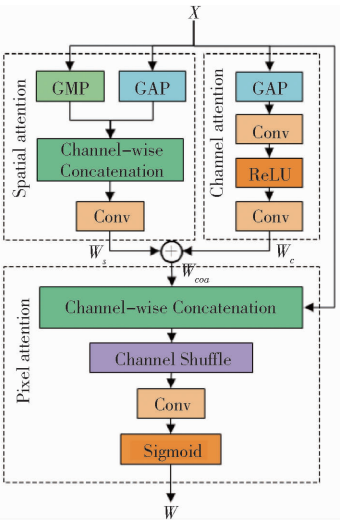

图 7　CGA 模块结构

Fig. 7　CGA module structure

X_{GAP}^{S}——跨通道维度全局平均池化处理特征

X_{GMP}^{S}——全局最大池化处理特征

为了减小计算参数量，第 1 个 1×1 卷积用比例系数 r 将通道 C 维度减小到 C/r，第 2 个 1×1 的卷积将通道维度恢复到 C，取 r 为 16。遵循广播规则将 W_c 和 W_s 结合在一起，获得粗略的 SIMs $W_{coa} \in R^{C\times H\times W}$。$W_{coa}$ 计算式为

$$W_{coa} = W_c + W_s \tag{7}$$

为获取最终优化的 SIMs W，利用输入特征内容作为引导，通过深度通道清洗操作将 W_{coa} 和 X 以交替形式重新排列，再经过卷积核为 7×7 的组卷积，最后 Sigmoid 激活得到最终输出 SIMs W。SIMs W 计算式为

$$W = \sigma(f_{Conv}^{7\times 7}(CS([X, W_{cos}]))) \tag{8}$$

式中　σ——Sigmid 激活函数

$f_{Conv}^{7\times 7}()$——卷积核为 $k\times k$ 组卷积层

$CS()$——深度通道清洗操作

浅层特征对多尺度目标识别具有不可忽视的作用，但在经过许多中间层后逐渐失去影响，特征融合可增强信息从浅层向深层流动，有利于特征保留和梯度方向传播。但简单的融合方法是元素相加，存在接收野失配问题，因此基于 CGA 的特征混合（CGAFusion）通过学习到的空间权重，自适应融合浅层特征和相应的深层特征。CGAFusion 结构如图 8 所示。

CGAFusion 将浅层特征和相应的高层特征输入到 CGA 中计算特征调制的空间权重，采用加权求和的方式组合，同时通过跳层连接增加输入特征，以缓解梯度消失问题，对融合后特征图采用 1×1 的卷积，得到最终特征 X_{fuse} 为

$$X_{fuse} = f_{Conv}^{1\times 1}(X_S W + X_D(1-W) + X_S + X_D) \tag{9}$$

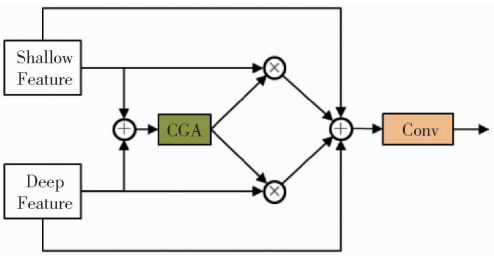

图 8　CGAFusion 结构图

Fig. 8　Structure diagram of CGAFusion

式中　X_D——高层特征　　X_S——浅层特征

原网络存在小尺度目标识别性能不足的问题，为提高多尺度板栗目标识别能力，改进原始网络颈部，提出 CGAFPN 网络，通过 CGA 将小尺度目标层融入，增强多尺度目标识别能力，改进前后网络结构如图 9 所示。

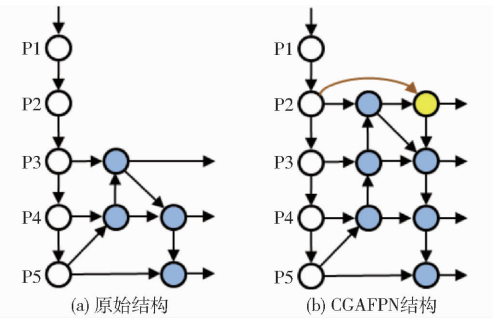

图 9　特征融合网络结构

Fig. 9　Feature fusion network structure

相较于原始网络，CGAFPN 引入浅层特征图，同时引入更多的特征融合操作，增强模型多尺度特征识别性能；加入 CGA 模块，优化了特征信息组合，更好地整合全局和局部信息。

1.2.1.3　SCSB 检测头

原始 YOLO 11 检测头采用解耦二分支结构，其中分类检测头由 2 个深度可分离卷积构成，位置检测由 2 个常规卷积构成，2 个分支均由独立卷积架构构成，难以充分共享和整合特征信息。同时由于特征图在不同层级上的独立处理方式，在处理复杂背景和小目标时表征不够全面，特征细节捕获和融合的能力不足。此外，Detect 结构中的多次卷积计算增加了推理时间和计算开销，不利于部署在低算力设备或实时检测任务中。

提出了 SCSB 检测头，采用共享卷积和耦合头的结构，共享卷积能够在不同尺度特征层之间传递信息，提升了特征的多任务共享性，增强了模型对小目标特征的敏感性，有效减小计算冗余，同时耦合头的协同处理使得分类和回归任务在同一特征空间中互为补充，进一步加强了分类和回归协同性，使得模型在单类板栗目标检测中表现更佳。SCSB 检测头结构如图 10 所示，图中 BMS 为批量归一化激活操

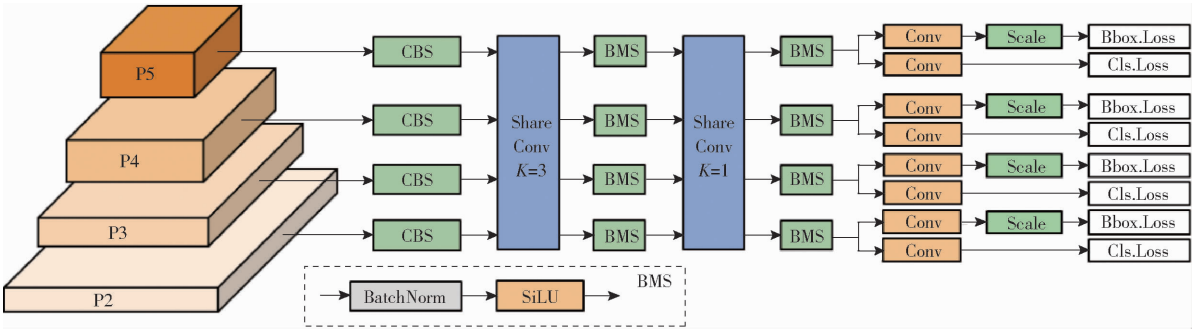

图 10 SCSB 检测头结构图

Fig. 10 Structure diagram of SCSB detection head

作,Share Conv 为共享卷积模块,Scale 为尺度平衡模块。

具体实现过程为:

经过特征融合的 4 个不同尺度特征图,通过常规卷积模块整合特征图中的信息,传递至共享卷积模块进行初步特征提取,共享卷积由一个 3×3 的卷积和一个 1×1 的卷积组成,通过跨层共享特征,增强各尺度特征的相关性。输入特征图为 4 种不同尺度特征层$\{X_i \mid i \in \{1,2,3,4\}\}$,每层特征图 X_i 经过共享卷积生成中间特征 $\widetilde{X}_i$,即

$$\widetilde{X}_i = \sigma(W * X_i + b) \tag{10}$$

式中 $*$——卷积操作 b——偏置

W——权重 σ——激活函数

在每一层共享卷积后引入分离批量归一化模块(Separate BatchNorm)使得不同尺度特征图在共享卷积后表现得到独立调整,防止过度融合带来的信息丢失,对于每个尺度,经过 BN 层结果为

$$BN(\widetilde{X}_i) = \frac{\widetilde{X}_i - \mu_i}{\sqrt{\sigma_i^2 + \varepsilon}}\gamma_i + \beta_i \tag{11}$$

式中 μ_i——$\widetilde{X}_i$ 均值 σ_i——$\widetilde{X}_i$ 标准差

ε——极小数

γ_i、β_i——可学习参数

随后模型采用 2 个卷积分支分别用于类别分类和位置回归,分别生成类别置信度和目标的回归坐标。随后在位置回归卷积后应用尺度平衡(Scale)模块对每层特征图分配自适应权重,从而确保对不同大小的输入特征进行适应性调节,增强小目标在多尺度特征融合中得到增强。分类卷积输出 X_i' 和回归卷积输出 X_i'' 表达式为

$$X_i' = f_{Conv}^{1\times1}(BN(\widetilde{X}_i)) \tag{12}$$

$$X_i'' = f_{Conv}^{1\times1}(BN(\widetilde{X}_i))S_i \tag{13}$$

最终在推理阶段将各尺度特征图预测结果进行级联组合,为单类板栗小目标在不同尺度特征图中得到一致性检测和回归,为多尺度板栗目标检测提供了高效、稳定的结构。

1.2.2 通道式特征知识蒸馏

知识蒸馏通过从大型网络传递知识,可提高小型网络性能,满足需要在受限的设备上部署高精度板栗目标识别模型的应用场景,逻辑知识蒸馏使用教师模型输出的逻辑训练学生模型,而特征知识蒸馏将教师模型特征层输出作为特征表示让学生模型学习,可以更好地学习更丰富的特征表示。

使用通道式特征知识蒸馏(Channel-wise knowledge distillation, CWD)[28]作为蒸馏方法,通过通道级特征传递,可提升学生模型对复杂环境中的关键特征捕捉能力,利用散度 KL 评估教师模型和学生模型之间的通道分布差异,在多尺度板栗目标识别的场景中,从教师模型多个特征层中捕捉到不同尺度特征信息,增强蒸馏目标多样性,同时通过最小化教师模型和学生模型之间特征图距离,促进知识有效迁移,保持板栗目标在不同尺度下保持一致,增强模型多尺度输入稳定性。CWD 结构如图 11 所示。

CWD 对教师模型和学生模型对应的通道激活值转化成概率分布,之后计算相应通道概率分布之间的散度 Asymmetric KL 作为损失,CWD 损失函数为

$$L_{CWD} = KL(\phi(y^{(T)}) \parallel \phi(y^{(S)})) \tag{14}$$

式中 $y^{(T)}$——教师网络激活图

$y^{(S)}$——学生网络激活图

$\phi(\,)$——将激活值转换为概率分布函数

$\phi(\,)$使用 Softmax 归一化,消除教师网路和学生网络之间的数值差异,$\phi(\,)$转换函数为

$$\phi(y_C) = \frac{e^{\left(\frac{y_{C,i}}{t}\right)}}{\sum_{i=1}^{WH} e^{\left(\frac{y_{C,i}}{t}\right)}} \tag{15}$$

式中 C——通道 t——温度超参

i——通道中的索引位置

通过概率分布度量方式衡量差异,散度 KL 计算式为

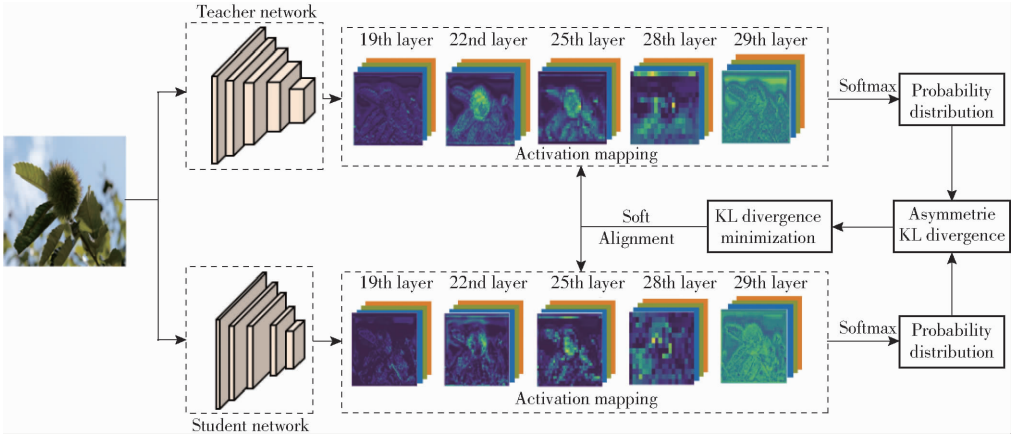

图 11　通道级知识蒸馏过程图

Fig. 11　Diagram of channel-level knowledge distillation process

$$KL(y^{(T)} \parallel y^{(S)}) = \frac{t^2}{C}\sum_{c=1}^{C}\sum_{i=1}^{WH}\phi(y_{c,i}^{(T)})\lg\frac{\phi(y_{c,i}^{(T)})}{\phi(y_{c,i}^{(S)})} \tag{16}$$

散度 KL 是一个非对称度量，由式(16)可知，若 $\phi(y_{c,i}^{(T)})$ 很大，则 $\phi(y_{c,i}^{(S)})$ 要与 $\phi(y_{c,i}^{(T)})$ 近似才能使 KL 散度最小，而当 $\phi(y_{c,i}^{(T)})$ 很小时，散度 KL 却不会使 $\phi(y_{c,i}^{(S)})$ 一直变小，对最小化 $\phi(y_{c,i}^{(S)})$ 的关注较少，因此，学生网络更趋向于在前景突出特征分布与教师网络趋于近似，而教师网络背景区域对应的特征对学习的影响较小。

通道级蒸馏能够使学生模型更好地捕捉复杂环境中的细微变化，有效提高模型特征表示能力和鲁棒性，更好地适应多尺度和复杂自然环境。

1.3　训练模型平台

训练模型平台硬件配置为 i9 9900K CPU，Nvidia A4000 GPU，8 GB 显卡，软件环境为 Win10 系统，64 位，Python 3.9 版本，Pytorch 2.0.0 框架，Cuda 11.7 版本。训练参数如表 2 所示。

表 2　训练参数

Tab. 2　Training parameters

参数	数值/类型
初始学习率	0.01
优化器	SGD
优化器动量	0.937
优化器权重衰减系数	0.000 5
训练批次数	4
迭代周期数	300

2　试验结果与分析

2.1　消融试验

消融试验结果如表 3 所示。从表 3 可知，改进 MKFA 特征提取模块后，提升了多尺度特征提取能力；改进 CGAFPN 网络后，平衡小目标在多尺度目标的贡献比重；改进 SCSB 检测头后，实现特征跨尺度提取，增强多尺度特征一致性。

表 3　消融试验结果

Tab. 3　Results of ablation test

基线网络	MKFA	CGAFPN	SCSB	精确率/%	召回率/%	平均精度/%	内存占用量/MB
YOLO 11s	×	×	×	87.4	73.1	81.7	19.2
YOLO 11s	√	×	×	87.0	74.4	82.6	19.5
YOLO 11s	×	√	×	87.7	78.7	86.6	21.3
YOLO 11s	×	×	√	87.4	74.0	82.2	19.9
YOLO 11s	√	√	×	88.7	78.7	87.0	21.7
YOLO 11s	×	√	√	88.8	78.2	86.8	22.2
YOLO 11s	√	×	√	87.5	74.8	83.5	20.0
YOLO 11s	√	√	√	88.2	79.2	87.2	22.5

注：√表示使用此模块，×表示未使用此模块。

由表 3 可知，经过 300 次训练迭代后，改进网络精确率、召回率和平均精度分别为 88.2%、79.2%、87.2%。较原 YOLO 11s 网络精确率、召回率和平均精度分别提升 0.8、6.1、5.5 个百分点，有效提升板栗目标检出率和正确率，改进 MKFA 特征提取模块后，召回率和平均精度分别提升 1.3、0.9 个百分点，有效提升多尺度目标提取能力；改进 CGAFPN 特征融合网络后，召回率和平均精度分别提升 5.6、4.9 个百分点，有效增加小尺度在多尺度目标中的贡献比重，显著增加不同尺度板栗目标检测能力；改进 SCSB 检测头后，召回率和平均精度分别提升 0.9、0.5 个百分点，借助跨尺度特征提高不同尺度特征一致性；共同改进 MKFA 模块和 CGAFPN 网络后，召回率和平均精度分别提升 5.6、5.3 个百分点，有效提升板栗目标特征提取和特征融合利用；共同改进 CGAFPN 网络和 SCSB 检测头后，召回率和平均精度分别提升 5.1、5.1 个百分点，进一步提升板栗多尺度目

标识别能力。

2.2 可视化试验

为观察改进 C3k2 - MKFA 模块对于板栗目标关键特征的提取能力,增强改进方法可解释性,用高分辨率类激活图[29](High-Resolution class activation mapping, HiResCAM)对输入图像生成类激活热力图。通过特定类别的分类权重与激活特征图逐像素相乘,直接计算出特征对目标类别的贡献,生成初始热力图,将初始热力图插值至输入图像的原始分辨率以提高细节表达能力,从而保留目标区域的高分辨率信息。使用 HiResCAM 生成的热力图可以表征上述改进后对预测输出的贡献分布,揭示每个位置对提取特征的重要程度,颜色越深表示原始图像对应区域对网络的响应越高,贡献越大,颜色分布位置、深度同样表示算法改进对于特征提取能力的影响。

图 12 为改进 MKFA 模块前后板栗目标特征热力图,通过热力图观察 MKFA 模块提取的板栗目标特征对于整体预测的贡献情况,评估改进的有效性。

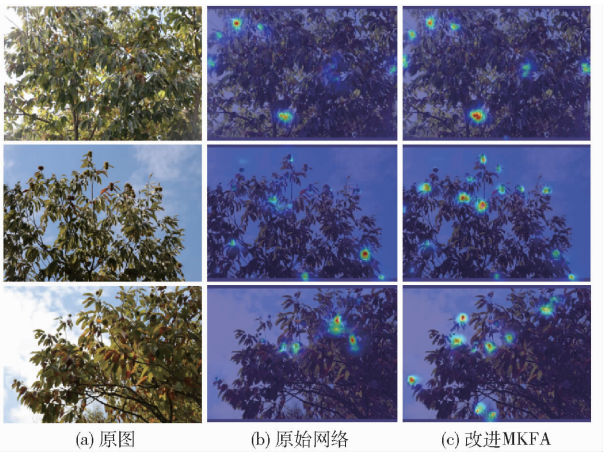

图 12 特征提取效果热力图

Fig. 12 Heatmaps of feature extraction effect

由图 12 可知,原始 YOLO 11s 网络存在特征提取能力不足、提取不够全面和特征错误贡献的情况,改进 MKFA 模块后,正确特征对网络的贡献更强,同时对预测结果产生影响位置更多,更全面,检测效果更好。

2.3 特征融合网络不同结构对比

为探索最佳 CGAFPN 网络结构,分别在不同尺度的特征层输出层和主干对应特征层增加跨层连接,共同进行 CGAFusion 融合操作,图 13 为融合不同尺度特征层结构图。

为验证融合不同尺度特征层的 CGAFPN 网络检测效果,以 YOLO 11 为基线网络,开展对比试验。试验结果如表 4 所示。

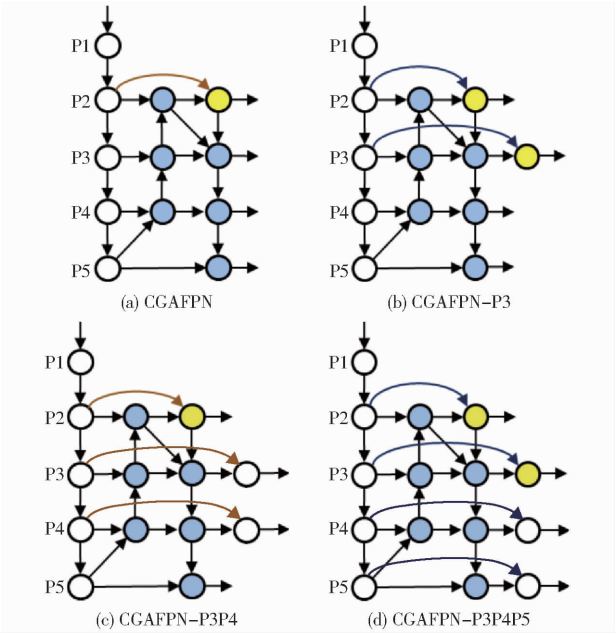

图 13 融合不同特征层结构

Fig. 13 Fusion of different feature layer structures

表 4 不同特征融合层对比试验结果

Tab. 4 Comparative tests of different features fusion layers

模型	*P*/%	*R*/%	mAP/%	内存占用量/MB
YOLO 11s	87.4	73.1	81.7	19.2
CGAFPN	88.2	79.2	87.2	22.5
CGAFPN - P3	88.1	78.7	87.2	22.7
CGAFPN - P3P4	87.7	79.1	87.0	22.9
CGAFPN - P3P4P5	87.9	79.5	87.2	23.7

由表 4 可知,改进 CGAFPN 对板栗目标的识别性能优于融合更多特征层,同时体积更小,原因是各特征层的组合和融合根据任务需求和网络结构细致调优,不同特征图对板栗目标的关注点不一样,改进 CGAFPN 能够有效利用细节特征的注意力信息,而在其他尺度同时增加 CGA 特征融合,会使得模型在高层次语义信息上更为分散,引入更多的冗余信息,不同层次特征间信息干扰增加,影响模型在细节和全局信息的平衡,削弱了原始特征区分能力,从而影响模型整体检测性能。

2.4 不同知识蒸馏方法对比试验

知识蒸馏可分为特征蒸馏和逻辑蒸馏。特征蒸馏通过教师模型和学生模型在中间特征表示之间建立约束,使得学生模型能够学习到教师模型的深层特征表示。逻辑蒸馏通过学生模型学习教师模型的输出分布,从而提高学生模型对类别间关系的理解。

为获取最好的模型压缩结果,对比 BCKD[30]、L1、L2 等逻辑蒸馏方法和 CWD、MIMIC[31]、MGD[32] 等特征蒸馏方法对改进模型的知识整理效果,结果如表 5 所示。

表 5　不同蒸馏方法压缩效果

Tab. 5　Compression effects of different distillation methods

蒸馏类型	P/%	R/%	mAP/%	内存占用量/MB
BCKD	87.1	74.0	83.8	6.0
L1	86.3	74.5	83.5	6.0
L2	87.5	73.6	83.7	6.0
CWD	87.6	75.6	84.7	6.0
MIMIC	87.6	74.1	84.1	6.0
MGD	86.8	74.4	83.5	6.0

学生模型与教师模型主要区别在于深度和宽度系数不同，教师模型在各特征层表示的细节和丰富度上更优。特征蒸馏通过直接传递教师在不同尺度上提取的特征，可以保留更多空间细节和结构信息，学生模型可以学习到多尺度下的特征表达方式，有利于保持小尺度目标和大尺度目标特征一致性，学生模型尺度感知能力更强，逻辑蒸馏对语义信息学习较为模糊，特征蒸馏下的学生模型可以更好地学习细粒度局部信息和语义细节信息，尤其是自然条件下背景复杂，含有大量小目标的多尺度板栗果实数据集，特征辨别力和定位精度更高。

2.5　主流网络对比试验

对比当下 SSD[33]、RT－DETR、EfficientDet[34]、YOLO v5m、YOLO v8m、YOLO v9m[35]、YOLO v10m[36]、YOLO 11m 以及本文改进网络，对比结果如表 6 所示。

由表 6 可知，改进网络 YOLO 11－MCS 模型内存占用量最小，平均精度相较于 SSD、RT－DETR、EfficientDet、YOLO v5m、YOLO v8m、YOLO v9m、YOLO v10m、YOLO 11m 分别高 22.4、0.3、30.5、3.5、3.7、3.7、4.0、4.8 个百分点。CWD 蒸馏后模型

表 6　主流网络对比试验

Tab. 6　Comparative tests of mainstream networks

网络	精确率/%	召回率/%	平均精度/%	内存占用量/MB
SSD	76.8	52.8	64.8	92.8
RT－DETR	88.1	79.0	86.9	35.5
EfficientDet	66.7	58.5	56.7	51.6
YOLO 11m	88.2	75.5	83.7	40.5
YOLO v10m	87.4	75.3	83.5	50.8
YOLO v9m	88.2	74.7	83.5	69.7
YOLO v8m	87.8	75.3	83.2	52.0
YOLO v5m	88.1	70.6	82.4	42.1
本文改进	88.2	79.2	87.2	22.5
本文改进＋CWD	87.6	75.6	84.7	6.0

平均精度分别高于 SSD、EfficientDet、YOLO v5m、YOLO v8m、YOLO v9m、YOLO v10m、YOLO 11m 模型 19.9、28、1.0、1.2、1.2、1.5、2.3 个百分点，模型内存占用量只有 6 MB。改进网络具有良好的板栗多尺度目标检测性能，蒸馏压缩后模型依旧具有较好的检测能力。

针对检测效果较好的 YOLO 11m、YOLO v10m、YOLO v9m、YOLO v8m、YOLO v5m、RT－DETR、改进 YOLO 11－MCS 开展小尺度目标检测效果对比试验，图 14 为各模型检测效果。

由图 14 可知，待检目标共 9 个，其中改进 YOLO 11－MCS 模型检出 8 个目标，经过 CWD 蒸馏后的改进 YOLO 11－MCS 模型检出 8 个目标，RT－DETR 检出 6 个目标，YOLO 11m 检出 6 个目标，YOLO v10m 检出 7 个目标，YOLO v9m 检出 4 个目标，YOLO v8m 检出 4 个目标，YOLO v5m 检出 7 个目标。所有模型均存在 1 处板栗小目标漏检，RT－



图 14　不同主流模型检测结果

Fig. 14　Detection results of different mainstream models

DETR、YOLO 11m、YOLO v9m、YOLO v8m、YOLO v5m 对于栗果互相重叠目标均存在 1 或 2 处漏检。

改进 YOLO 11 – MCS 对于极难检出小目标依旧存在不足,蒸馏后模型检测能力并未有较大下降,满足板栗多尺度目标识别要求。

2.6 多尺度板栗目标检测效果

为验证不同环境下多尺度板栗目标检测效果,选择阴天、逆光等较差检测条件下大、中、小尺度板栗目标进行改进前后对比试验,结果如图 15 所示。

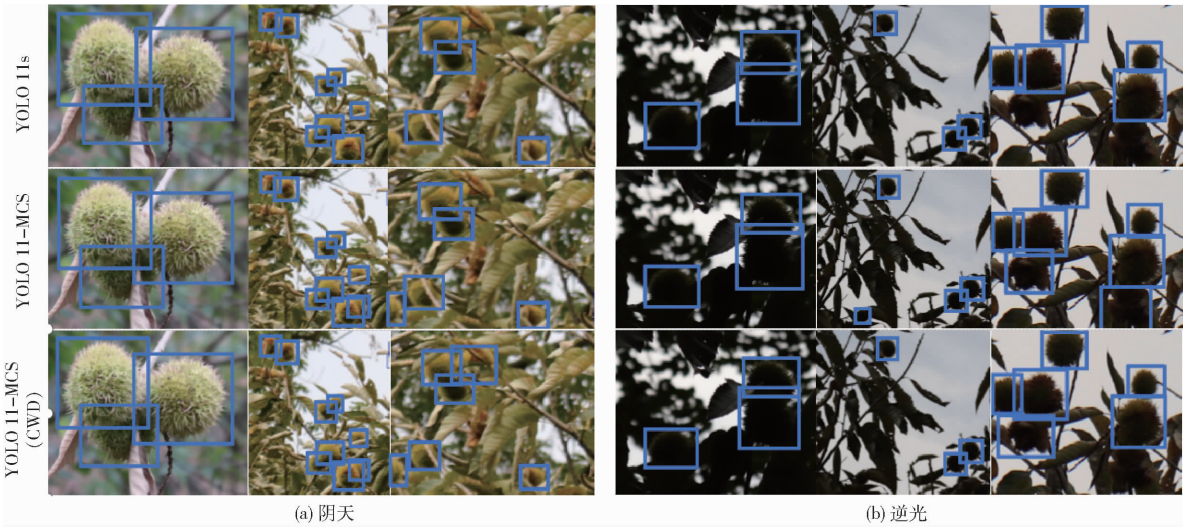

图 15 多尺度板栗目标检测效果对比

Fig. 15 Comparison of multi-scale chestnut target detection results

由图 15 可知,在阴天光线不足和逆光较难检测条件下,改进前后及整理后模型在大尺度板栗目标下均能检出;小尺度板栗目标检测中,阴天和逆光条件下,改进 YOLO 11 – MCS 均能检出,蒸馏后模型均存在 1 处漏检,YOLO 11s 均存在 2 处漏检,中尺度板栗目标检测中,阴天条件下,改进 YOLO – MCS 和蒸馏后模型均能全部检出,YOLO 11s 在遮挡目标处存在 1 处漏检,逆光条件下,改进 YOLO – MCS 能全部检出,蒸馏后模型和 YOLO 11s 存在 1 处漏检。

试验结果表明,改进后模型在复杂自然条件中表现出更强的适应性和鲁棒性,在阴天光线不足、逆光的较难检测条件下,改进前 YOLO 11s 模型在处理小尺度目标检测时表现出局限性,尤其在低光照情况下容易产生漏检和误检。而改进后的模型通过强化多尺度提取、特征融合和检测,有效提升了对大、中、小尺度板栗目标检测能力。

2.7 边缘设备部署效果

使用 Infer 推理库,基于 TensorRT 加速器,将蒸馏压缩后的模型部署在 Jetson Nano 上。在 Jetson Nano 上使用 Infer 推理库部署 YOLO 11 模型可以充分利用设备的 NVIDIA GPU 和 TensorRT 加速能力。

首先,将用 CWD 蒸馏后改进 YOLO 11 – MCS 模型权重从 PyTorch 格式(.pt)导出为 ONNX 格式(.ONNX),通过 ONNX Simplifier 优化模型结构,将模型转换为 TensorRT 推理格式(.engine),使用 TensorRT 引擎实现高性能推理。经 Jetson Nano 部署后模型检测速度有了较大提升,为了对比部署前后检测速度差异,开展对比试验,结果如表 7 所示。

表 7 模型部署检测结果

Tab. 7 Model deployment detection results

模型	计算量	检测时间/(ms·幅$^{-1}$)
YOLO 11 – MCS	6.09×10^{10}	1 727.0
YOLO 11 – MCS(CWD)	1.58×10^{10}	824.5
YOLO 11 – MCS(CWD) + Infer	1.58×10^{10}	23.0

由表 7 可知,在 Jetson Nano,改进 YOLO 11 – MCS 计算量较大,检测速度较低,经过蒸馏后改进模型计算量和检测速度分别提升 3.85 倍和 2.09 倍,知识蒸馏有效减小模型体积,缩减计算开销,提升计算速度。

使用 Infer 推理库部署后,检测时间为 23 ms/幅,较改进 YOLO 11 – MCS 速度提升 75.09 倍,较蒸馏压缩后模型速度提升 35.85 倍,满足计算力不足的边缘设备部署要求。

3 结论

(1)提出了一种基于 YOLO 11 的复杂条件下多尺度板栗果实识别方法,较原始模型平均精度提升 5.5 个百分点,召回率提升 6.1 个百分点,对于多尺度板栗目标检测能力有了较大提升,相较于 SSD、RT – DETR、EfficientDet、YOLO v5m、YOLO v8m、

YOLO v9m、YOLO v10m、YOLO 11m 分别高 22.4、0.3、30.5、3.5、3.7、3.7、4.0、4.8 个百分点。

(2)提出了 MKFA 多尺度关键特征聚合模块，并构建了 C3k2－MKFA 特征提取模块，提高了不同尺度板栗目标特征捕捉能力；提出了 CGAFPN 特征融合网络，提高了板栗小目标在多尺度目标的贡献比重；提出了 SCSB 检测头，进一步增强了不同尺度板栗目标特征提取能力，共同改进后平均精度提升 5.5 个百分点。

(3)通过 CWD 特征知识蒸馏，模型体积缩小 3.75 倍，通过 Infer 推理库在 Jetson Nano 部署后，每幅图像推理时间缩短至 23 ms，较未蒸馏压缩前模型推理速度提升 75.09 倍，较未使用 Infer 推理库部署前的模型推理速度提升 35.85 倍。

参 考 文 献

[1] MASSANTINI R, MOSCETTI R, FRANGIPANE M T. Evaluating progress of chestnut quality: a review of recent developments [J]. Trends in Food Science & Technology, 2021, 113: 245－254.

[2] FRANGIPANE M T, MASSANTINI R, CORONA P. Better selection of chestnut cultivars: experimenting the sensory characterization of the Marrone chestnuts compared to the "Chataigne" chestnuts[J]. European Food Research and Technology, 2024, 250(3): 961－969.

[3] LI R, SHARMA A K, ZHU J, et al. Nutritional biology of chestnuts: a perspective review[J]. Food Chemistry, 2022, 395: 133575.

[4] 姜洪喆,谭烽,李兴鹏,等.转盘式板栗可见/近红外光谱检测分选系统研究[J].农业机械学报,2024,55(12):462－469.
JIANG Hongsheng, TAN Feng, LI Xingpeng, et al. Research and development of rotary sorting system and device for chestnuts based on visible/near-infrared spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(12): 462－469. (in Chinese)

[5] MARTINEZ A, DI SARNO V, MADDALONI P, et al. Chestnut quality classification by THz time-domain hyperspectral imaging combined with unsupervised learning analysis[J]. Food Control, 2025, 168: 110878.

[6] 展慧,李小昱,王为,等.基于机器视觉的板栗分级检测方法[J].农业工程学报,2010,26(4):327－331.
ZHAN Hui, LI Xiaoyu, WANG Wei, et al. Determination of chestnuts grading based on machine vision[J]. Transactions of the CSAE, 2010, 26(4): 327－331. (in Chinese)

[7] 周竹,刘洁,李小昱,等.霉变板栗的近红外光谱和神经网络方法判别[J].农业机械学报,2009,40(增刊):109－112.
ZHOU Zhu, LIU Jie, LI Xiaoyu, et al. Discrimination of moldy chinese chestnut based on artificial neural network and near infrared spectra[J]. Transactions of the Chinese Society for Agricultural Machinery, 2009, 40(Supp.): 109－112. (in Chinese)

[8] COMBA L, BIGLIA A, SOPEGNO A, et al. Convolutional neural network based detection of chestnut burrs in UAV aerial imagery[C]// Conference of the Italian Society of Agricultural Engineering. Cham: Springer International Publishing, 2022: 501－508.

[9] SUN Y, HAO Z, GUO Z, et al. Detection and mapping of chestnut using deep learning from high-resolution UAV-based RGB imagery[J]. Remote Sensing, 2023, 15(20):4923.

[10] WANG J, SU Y, YAO J, et al. Apple rapid recognition and processing method based on an improved version of YOLOv5[J]. Ecological Informatic, 2023,77: 102191.

[11] QI J, LIU X, LIU K, et al. An improved YOLOv5 model based on visual attention mechanism: application to recognition of tomato virus disease[J]. Computers and Electronics in Agriculture, 2022,194: 106780.

[12] ZHAO Y, LV W, XU S, et al. Detrs beat yolos on real-time object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965－16974.

[13] ARAKAWA T, TANAKA T S T, KAMIO S. Detection of on-tree chestnut fruits using deep learning and RGB unmanned aerial vehicle imagery for estimation of yield and fruit load[J]. Agronomy Journal, 2023, 116: 973－981.

[14] KAMILARIS A, PRENAFETA-BOLDÚ F X. Deep learning in agriculture: a survey[J]. Computers and Electronics in Agriculture, 2018, 147: 70－90.

[15] ZHAO Z Q, ZHENG P, XU S, et al. Object detection with deep learning: a review[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(11): 3212－3232.

[16] TERVEN J, CÓRDOVA-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.

[17] ZHANG Y, LIU X, ZHAO X. YOLO－SE: a universal real-time multi-scale object detection framework[C]// 2024 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE, 2024: 1－6.

[18] ZHOU S, LONG S. YOLO－LF: a lightweight multi-scale feature fusion algorithm for wheat spike detection[J]. Journal of Real-Time Image Processing, 2024, 21(4): 148.

[19] RUAN Z, WANG H, CAO J, et al. Cross-scale feature fusion connection for a YOLO detector[J]. IET Computer Vision,

2022, 16(2): 99 - 110.
[20] SAPKOTA R, MENG Z, KARKEE M. Synthetic meets authentic: leveraging LLM generated datasets for YOLO 11 and YOLOv10-based apple detection through machine vision sensors[J]. Smart Agricultural Technology, 2024, 9: 100614.
[21] 张昌松,李伟. 基于 YOLO v5 的直播油菜激光间苗系统设计与试验[J]. 农业机械学报,2024,55(4):40 - 52.
ZHANG Changsong, LI Wei. Research and experiment of direct seeding rape laser seedling system based on YOLO v5[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(4):40 - 52. (in Chinese)
[22] 李尚平,郑创锐,文春明,等. 基于无人机 RGB 图像与改进 YOLO v5s 的宿根蔗缺苗定位方法[J]. 农业机械学报,2024,55(12):57 - 70.
LI Shangping, ZHENG Chuangrui, WEN Chunming, et al. Method for locating missing ratoon sugarcane seedlings based on RGB images from unmanned aerial vehicles and improve YOLO v5s[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(12):57 - 70. (in Chinese)
[23] 刘双喜,张玮平,胡宪亮,等. 基于改进 YOLO v8s 的水稻种植机械作业质量检测[J]. 农业机械学报,2024,55(增刊 1):61 - 70.
LIU Shuangxi, ZHANG Weiping, HU Xianliang, et al. Rice planting machinery operation quality detection based on improved YOLO v8s[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(Supp. 1):61 - 70. (in Chinese)
[24] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251 - 1258.
[25] 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 Vision and Pattern Recognition, 2023: 12021 - 12031.
[26] 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.
[27] CHEN Z, HE Z, LU Z. DEA - Net: single image dehazing based on detail-enhanced convolution and content-guided attention [J]. IEEE Transactions on Image Processing, 2024,33:1002 - 1015.
[28] SHU C, LIU Y, GAO J, et al. Channel-wise knowledge distillation for dense prediction[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 5311 - 5320.
[29] DRAELOS R L, CARIN L. Use HiResCAM instead of Grad - CAM for faithful explanations of convolutional neural networks [J]. arXiv preprint arXiv:2011.08891, 2020.
[30] YANG L, ZHOU X, LI X, et al. Bridging cross-task protocol inconsistency for distillation in dense object detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 17175 - 17184.
[31] WEI Y, PAN X, QIN H, et al. Quantization mimic: towards very tiny CNN for object detection[C]// Proceedings of the European Conference on Computer Vision (ECCV), 2018: 267 - 283.
[32] YANG Z, LI Z, SHAO M, et al. Masked generative distillation[C]// European Conference on Computer Vision, 2022: 53 - 69.
[33] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: single shot multibox detector[C]// Proceedings of the 14th European Conference on Computer Vision, 2016: 21 - 37.
[34] TAN M, PANG R, LE Q V. Efficientdet: scalable and efficient object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781 - 10790.
[35] WANG C, YEH I H, MARK LIAO H. Yolov9: learning what you want to learn using programmable gradient information[C]// European Conference on Computer Vision, 2025: 1 - 21.
[36] WANG A, CHEN H, LIU L, et al. Yolov10: real-time end-to-end object detection[J]. arXiv Preprint, arXiv:2405.14458, 2024.