doi:10.6041/j.issn.1000-1298.2026.18.029

基于改进 SAMF－YOLO v10 的火龙果目标检测方法

曾　达　曾德真　覃胜林　韦钙兴
（南宁理工学院大数据与人工智能学院，南宁 530100）

摘要：为解决火龙果果实在自然生长环境中因尺寸差异大、数量密集导致的重叠遮挡问题，本研究提出了一种基于 YOLO v10 改进的协同注意多特征融合目标检测模型（SAMF－YOLO v10）。在特征提取中，使用多尺度特征融合（Multi-scale feature fusion，MFC）模块代替传统的卷积模块，以增强模型对不同尺度特征的表达能力并提升细节信息的捕获效果。同时，加入协同通道空间注意力（Synergistic effects between spatial and channel attention，SCSA）模块，通过对不同空间位置及不同方向特征之间的关系进行建模，实现特征之间的协同增强，从而提高对遮挡目标和小目标的特征表达能力。结果表明，SAMF－YOLO v10 模型在火龙果数据集上的检测精确率为 90.18%，召回率为 86.28%，平均精度均值（mAP@0.5）为 93.03%。与 YOLO v10 模型相比，SAMF－YOLO v10 在精确率、召回率和 mAP@0.5 上分别提升 3.17、4.93、3.87 个百分点；所提模型能够有效整合多层次和关键信息，从而提升火龙果图像目标检测性能。

关键词：火龙果；目标检测；YOLO v10；多尺度特征融合；协同注意力
中图分类号：TP391；S126　　**文献标识码**：A　　**文章编号**：1000-1298(2026)18-0346-09　　**OSID**：

Pitaya Fruit Target Detection Method Based on Improved SAMF－YOLO v10

Zeng Da　Zeng Dezhen　Qin Shenglin　Wei Gaixing
(*Big Data and Artificial Intelligence College*, *Nanning College of Technology*, *Nanning* 530100, *China*)

Abstract: Aiming to tackle the issues of mutual overlap and occlusion of pitaya fruits under complex natural growth environments, which are primarily induced by drastic scale variation and dense distribution of fruit targets, an improved collaborative attention-based multi-feature fusion object detection model named SAMF－YOLO v10 was developed on the basis of the original YOLO v10 architecture. In the feature extraction process, the multi-scale feature fusion (MFC) module was used instead of the traditional convolution module, which can enhance the model's ability to express features of different scales and improve the capture of detailed information. Meanwhile, the synergistic effects between spatial and channel attention (SCSA) module was incorporated, it can capture the relationships between features at different spatial positions and in different directions to achieve collaborative enhancement of features, thereby improving the feature expression ability for occluded targets and small targets. The results showed that the SAMF－YOLO v10 model achieved a precision of 90.18%, a recall of 86.28%, and an mAP@0.5 of 93.03% on the pitaya fruit dataset. Compared with the original YOLO v10 model, SAMF－YOLO v10 improved the precision, recall, and mAP@0.5 indicators by 3.17, 4.93, and 3.87 percentage points, respectively. The experimental results demonstrated that the newly proposed model was able to effectively integrate multi-level feature information and critical semantic content, which further enhanced and optimized the overall target detection performance for pitaya fruit images.

Key words: pitaya fruit; target detection; YOLO v10; multi-scale feature fusion; synergistic attention

收稿日期：2026－04－03　修回日期：2026－04－27
基金项目：广西高校中青年教师科研基础能力提升项目（2024KY1761）
作者简介：曾达（1990—），男，讲师，主要从事深度学习和图像处理研究，E-mail：779102186@qq.com
通信作者：曾德真（1988—），男，高级工程师，主要从事智慧农业研究，E-mail：dezhenzeng@163.com

0　引言

火龙果是我国重要经济作物之一，具有较高的营养价值和经济效益[1-2]。广西地区凭借独特的地理与气候条件，已成为我国火龙果的主要生产区域。随着种植规模扩大，传统人工采摘已难满足效率和成本要求，火龙果采摘智能化发展成为必然趋势[3]。高效准确的火龙果识别检测是实现火龙果智能采收的关键。

早期农作物果实检测方法主要依赖人工设计的图像特征，如纹理、颜色和形状等，并结合传统机器学习分类器完成目标识别任务[4-5]。Wang 等[6]提出了一种结合像素阈值分割与几何中心匹配的荔枝识别检测方法。Sari 等[7]提出了一种基于朴素贝叶斯分类器和局部二值模式特征的木瓜检测方法。然而，此类传统方法高度依赖人工特征设计，且多针对特定场景进行优化，在光照变化剧烈、遮挡严重等复杂自然环境下，其检测性能和泛化能力往往受到明显限制。

相比传统方法，深度学习在农作物果实目标检测中具有更强的特征表征能力，并能更好地适应尺度变化、遮挡及复杂背景。其中，YOLO 系列算法因其端到端结构和较高的检测效率，得到了迅速发展并被广泛应用于水果目标检测领域[8-9]。Chen 等[10]基于 YOLO v5s 模型提出 YOLO－COF，通过引入 K－means＋＋和坐标注意力机制提升茶油果检测性能。Li 等[11]在 YOLO v7 模型基础上引入轻量级 GHOST 卷积和坐标注意力模块，实现复杂自然环境下的草莓检测。Shi 等[12]提出了 YOLO v7－SAP 苹果检测网络。在火龙果目标检测方面，Qiu 等[13]提出 GSE－YOLO 检测火龙果的成熟度，在 YOLO v8n 模型基础上，使用 GhostConv 替换卷积层，并引入注意力机制，在花芽期、未成熟期、半成熟期和成熟期中 mAP@0.5 达 90.9%。Qiu 等[14]提出 YOLO v8n－CBN 模型，用于开花期、未成熟和成熟期的火龙果检测。然而火龙果在不同成熟期的形态差异大，现有模型在多尺度特征融合方面不够，易丢失细节特征。Li 等[15]结合 MobileNet 提出 MobileNet－YOLO 实现火龙果检测。刘进一等[16]提出了基于 YOLO v8 模型的多尺度加权特征融合网络，实现了火龙果的检测和定位。Nan 等[17]提出基于 MobileNetV3 的轻量级 Faster R－CNN 模型，在不同图像类型的火龙果数据集上 mAP 最高达到 92.9%。在实际的火龙果目标检测场景中，果实尺度变化明显、目标分布密集且常受到枝叶遮挡，极大增加了检测难度。此外，现有检测模型在精度、复杂度和检测速度之间难以取得理想的平衡，这限制了其在复杂真实火龙果种植环境中的应用。

针对上述问题，本文基于 YOLO v10 模型，提出火龙果目标检测模型 SAMF－YOLO v10。通过 MFC 模块获取多尺度信息，以充分利用特征。为增强不同语义层级间的信息交互并强化关键特征的传递，引入 SCSA 模块以提升复杂场景下火龙果检测性能。

1　材料与方法

1.1　数据获取与数据集构建

数据采自广西壮族自治区南宁市隆安县火龙果种植园，使用佳能 EOS 850D 型相机获取。在不同时间段、不同光照条件以及多个视角进行拍摄，最终共采集火龙果图像 1 106 幅。为了便于后续训练和测试，使用 LabelImg 标注软件对图像中的火龙果进行标记，每幅图像标注完成后，生成相应的 XML 格式标注文件。将数据集按比例 7∶3 随机分为训练集和测试集。图 1 为不同光照条件和背景环境下采集的火龙果图像。



图 1　数据集部分图像示例

Fig. 1　Sample images from dataset

1.2　模型构建

1.2.1　SAMF－YOLO v10 模型

相比于其他 YOLO 系列模型，YOLO v10 模型[18]通过融合多尺度特征和采用高效的轻量化网络架构，显著增强了对上下文信息的捕捉能力。在保持高精度的同时显著提升了推理速度，并降低了计算资源消耗。因此，在 YOLO v10 模型基础上，融合 MFC 模块和 SCSA 模块，提出了 SAMF－YOLO v10 模型以实现复杂场景下的火龙果检测。模型结构如图 2 所示。

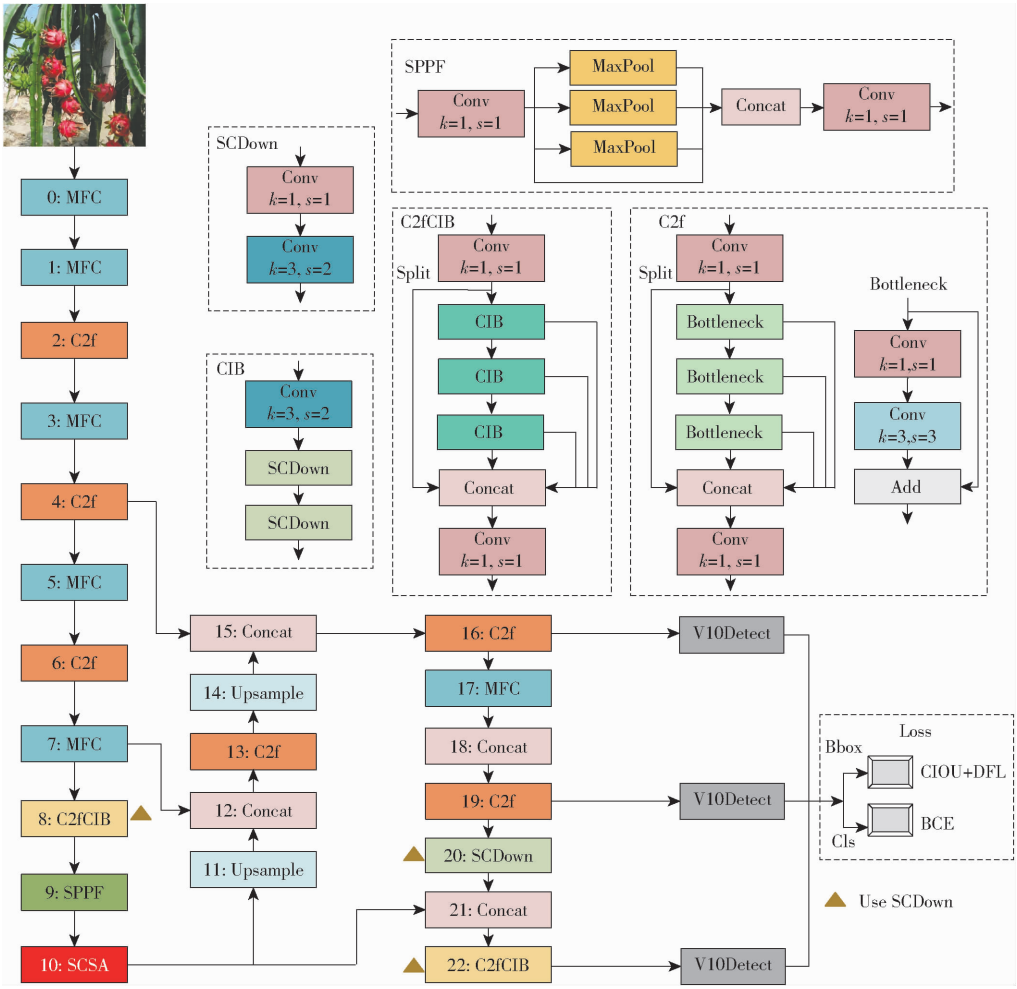


图 2 SAMF－YOLO v10 结构

Fig. 2 SAMF－YOLO v10 structure

火龙果图像输入到 SAMF－YOLO v10 模型后，首先经过 MFC 模块，该模块替代了基线 YOLO v10 中的传统卷积操作，通过在不同尺度上提取局部特征并进行融合，实现更高效的特征表达能力。随后，特征被输入到 C2f（Cross-stage partial-connection with 2 convolutions）模块。该模块通过对输入特征进行分支划分，使用 1×1 卷积和 3×3 卷积进行特征堆叠和融合，能够保留更加丰富的梯度信息并减少计算量。在此基础上，网络进一步引入 C2fCIB 模块，该模块是在 C2f 结构上结合 CIB（Channel interaction block）结构形成改进的模块，通过 3×3 卷积、1×1 卷积、3×3 卷积加强通道间的信息交互。按“3 个 C2f 模块＋1 个 C2fCIB 模块”模式进行交替堆叠。其中，浅层阶段主要采用 C2f 模块以增强基础特征表达能力，而在中后层阶段引入 C2fCIB 模块以增强上下文建模与信息交互能力。通过 SPPF（Spatial pyramid pooling fast）模块对特征进行处理。该模块利用多次连续的最大池化操作，在不同感受野尺度上提取特征信息并进行融合，从而有效扩大网络的感受野，提高模型对不同尺度目标的感知能力。随后，本研究提出的 SCSA 注意力模块被用于替代原模型中的 PSA（Partial self-attention）注意力机制。SCSA 模块能够同时对特征图的通道维度和空间维度进行自适应加权，从而强化关键区域的特征响应，提高模型在目标遮挡或复杂背景环境下的检测能力。此外，引入由 1×1 卷积和 3×3 卷积组成的 SCDown（Spatial-channel decoupled downsampling）模块替代传统的步长为 2 的 3×3 卷积操作，减少特征损失并提升检测性能。值得注意的是，SCDown 模块并未在所有下采样位置都替换，而是在第 8、20、22 阶段中进行替换，在每个 MFC 模块后的位置依然使用传统的步长卷积下采样方法。最后将提取的特征进行聚合，输入到 YOLO v10 检测头，对火龙果目标的位置和类别进行预测，实现对不同尺度火龙果目标的精准检测。

1.2.2 MFC 模块

针对火龙果检测过程存在的光照强度变化对果实颜色影响显著以及小目标特征模糊、难以识别等问题，本研究提出了 MFC 模块。该模块通过融合来自不同区域的全局上下文信息，有效增强模型对小目标的特征表达与感知能力，从而提升检测性能，具

体结构如图 3 所示。

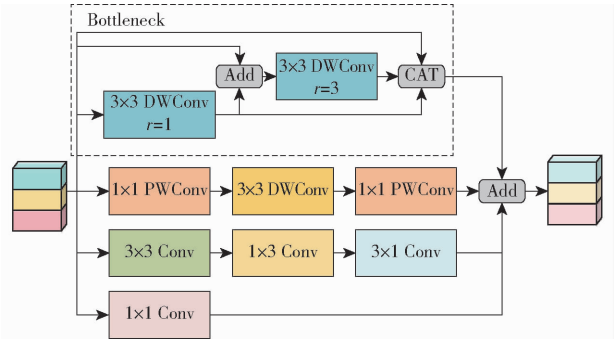


图 3　MFC 模块结构

Fig. 3　MFC module structure

首先，输入通过 4 个分支进行处理。第 1 个分支将输入张量分成 3 个子集，每个特征图子集（除第 1 个子集外）被传递给 3 × 3 DWConv（Depthwise convolution）实现高效的局部特征提取，并结合逐元素相加以提升特征表达能力。在这过程中，使用膨胀率为 3 的 3 × 3 DWConv 进一步增加感受野，允许模型更有效地捕获局部区域内的特征，然后再与原始特征进行融合操作，有助于保留原始表征进一步强化局部空间特征的表征能力。第 2 个分支主要由 1 × 1 PWConv（Pointwise convolution）、3 × 3 DWConv、1 × 1 PWConv 构成，不仅可以降低参数量，还能保留有效的特征表示。第 3 个分支主要由 3 × 3 Conv、1 × 3 Conv、3 × 1 Conv 构成，利用 1 × 3、3 × 1 非对称卷积以不同方式处理输入数据的不同方向，有助于模型更好地适应火龙果形状的不对称特征。第 4 个分支由 1 × 1 Conv 构成，为全局信息。最后将这 4 个分支的信息相加生成最终输出。通过将这些分支结构灵活地集成到网络中，既保持了模型效率，又可进一步丰富模型对输入数据的抽象表达。具体操作为

$$A = f_{DWconv}^{3\times3} FB = A \oplus FC = f_{DWconv}^{3\times3,r=3}(B) \tag{1}$$

$$X_1 = \mathrm{CAT}(C, A, F) \tag{2}$$

$$X_2 = f_{PWconv}^{1\times1}(f_{DWconv}^{3\times3}(f_{PWconv}^{1\times1}F)) \tag{3}$$

$$X_3 = f_{Conv}^{3\times1}(f_{Conv}^{1\times3}(f_{Conv}^{3\times3}F)) \tag{4}$$

$$X_4 = f_{Conv}^{1\times1}F \tag{5}$$

$$Y = X_1 \oplus X_2 \oplus X_3 \oplus X_4 \tag{6}$$

式中　$f_{DWconv}^{3\times3}$——3 × 3 深度卷积

$f_{DWconv}^{3\times3,r=3}$——膨胀率为 3 的 3 × 3 深度卷积

$f_{PWconv}^{1\times1}$——1 × 1 逐点卷积

$f_{Conv}^{3\times1}$——内核大小为 3 × 1 的卷积运算

$f_{Conv}^{1\times3}$——内核大小为 1 × 3 的卷积运算

$f_{Conv}^{3\times3}$——内核大小为 3 × 3 的卷积运算

$f_{Conv}^{1\times1}$——内核大小为 1 × 1 的卷积运算

$\oplus$——元素相加操作

CAT——特征融合操作

F——输入特征　　Y——输出特征

A——$r = 1$ 时，深度卷积结果

B——A 与输入特征 F 相加结果

C——$r = 3$ 时，对 B 进行深度卷积结果

X_1、X_2、X_3、X_4——分支 1、2、3、4 的输出结果

MFC 模块通过引入不同卷积操作提取多尺度特征，使模型能够在同一层级同时捕获局部细节与全局结构信息，从而增强特征表达能力，特别是更有效地识别在遮挡或重叠情况下难以分辨的火龙果目标。

1.2.3　SCSA 模块

注意力机制能够引导模型聚焦于输入中最相关的区域，减少冗余信息干扰。由于自然环境中火龙果枝叶常遮挡果实，为提升对重叠与遮挡目标的识别能力，引入 SCSA 模块[19]，对关键目标赋予更高权重，同时抑制背景信息，其结构如图 4 所示。

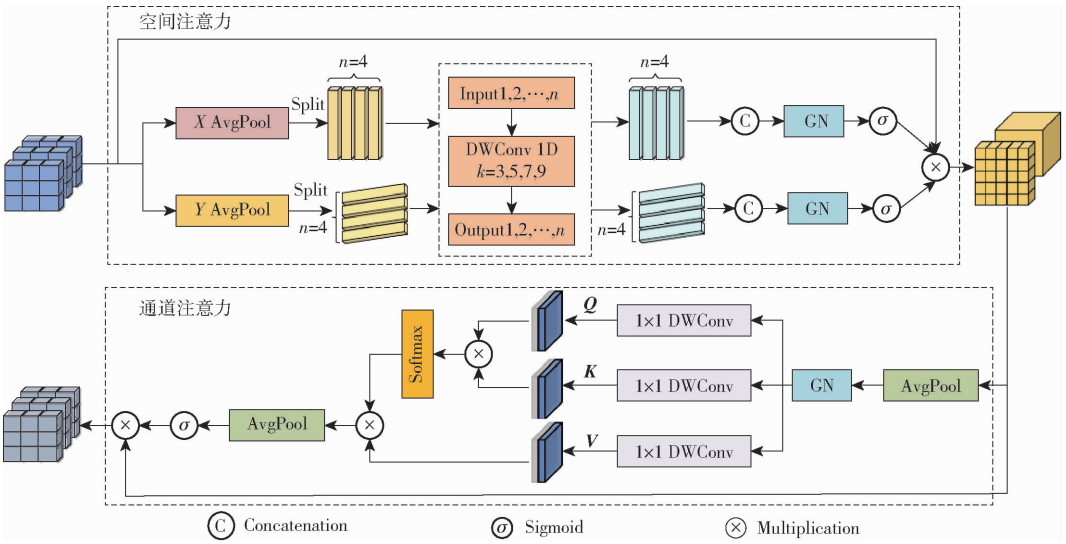


图 4　SCSA 模块结构

Fig. 4　SCSA module structure

SCSA 模块主要分为 2 个部分,即通道注意力部分和空间注意力部分。在通道注意力部分,对输入特征F_{in}在高度 H 和宽度 W 维度上分别进行平均池化得到F_H和F_W,从而生成 2 个一维序列结构。为了捕捉不同的空间分布和上下文信息,将特征集划分为 4 个大小相同、互相独立的子特征F_H^i和F_W^i。该过程可表示为

$$F_H^i = F_H\left[:,(i-1)\frac{C}{K}:i\frac{C}{K},:\right] \tag{7}$$

$$F_W^i = F_W\left[:,(i-1)\frac{C}{K}:i\frac{C}{K},:\right] \tag{8}$$

其中 $i \in [1,4]$

式中 F^i——第 i 个子集特征

C——通道数

随后对每个子特征分别应用深度方向的一维卷积(卷积核尺寸为 3、5、7 和 9),以提取不同尺度的语义信息。提取完成后,对子特征进行组归一化(Group normalization,GN),再通过 Sigmoid 激活函数生成通道注意力图。最终,将得到的注意力图与原始输入特征逐元素相乘,得到通道注意力的输出F_C。

在空间注意力部分,对F_C应用平均池化和 GN 操作得到F_P,使用多个 1 × 1 DWConv 分别提取特征,接着基于单头自注意力计算通道间的相似性,将特征F_{proj}分别输入到 3 个相互独立的线性映射层中,进而生成 Query($\boldsymbol{Q}$)、Key($\boldsymbol{K}$)和 Value($\boldsymbol{V}$)3 个向量。将向量 $\boldsymbol{Q}$ 与向量 $\boldsymbol{K}$ 进行矩阵相乘,所得结果再通过 Softmax 函数进行归一化处理,由此计算出注意力权重矩阵,该矩阵反映不同空间位置之间的相关性。把计算得到的注意力权重矩阵与向量 $\boldsymbol{V}$ 相乘,得到 1 个融合了全局依赖信息的上下文特征表示F_{att}。该过程可表示为

$$F_{att} = \text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{C}}\right)\boldsymbol{V} \tag{9}$$

其中

$$\begin{cases}\boldsymbol{Q} = F_{proj}^{Q}(F_p)\\ \boldsymbol{K} = F_{proj}^{K}(F_p)\\ \boldsymbol{V} = F_{proj}^{V}(F_p)\end{cases} \tag{10}$$

式中 F_{proj}^{Q}——查询特征 1 × 1 DWConv 卷积

F_{proj}^{K}——键特征 1 × 1 DWConv 卷积

F_{proj}^{V}——值特征 1 × 1 DWConv 卷积

然后通过平均池化的渐进压缩策略对特征进行压缩,以引导通道间的协同效应。经过 Sigmoid 激活函数生成注意力图,最终与通道注意力输出逐元素相乘,生成 SCSA 模块的最终输出特征F_{out}。

网络的 SPPF 模块通过多次连续的最大池化操作,有效扩大了网络的感受野,在 SPPF 模块后引入 SCSA 模块,可进一步提升模型对关键区域的关注程度,增强模型在复杂场景中的特征表达能力,尤其在面对小目标或存在部分遮挡的火龙果目标时,能够更加准确地实现目标的检测。

2 结果与分析

2.1 试验条件

本试验在 Windows 11 操作系统环境下开展,硬件平台采用 NVIDIA GeForce GTX 4090 显卡。软件环境基于 Conda 进行配置,主要包括 Python 3.12.9、PyTorch 2.5.0 以及 CUDA 12.3。优化器为 SGD,初始学习率为 0.01,权重衰减系数为 0.005,训练周期为 200,批量大小为 16。在训练过程中,应用随机旋转、缩放、改变亮度等预处理操作进行数据扩增。在输入模型之前将所有图像统一缩放为 640 像素 ×640 像素,并进行归一化处理。

2.2 评估指标

采用精确率(Precision)、召回率(Recall)以及平均精度均值(mean Average Precision,mAP)对模型检测性能进行定量评估。其中平均精度均值使用 mAP@0.5,表示在阈值为 0.5 时的平均精度。使用参数量、浮点运算量及推理时间对模型的计算复杂度与运行效率进行评估。

2.3 不同检测模型性能测试

使用几种经典的目标检测模型,如 YOLO v5、YOLO v8、YOLO v9[20]、YOLO v10[18]、YOLO 11[21]和 YOLO 12[22],对火龙果图像进行目标检测,结果见表 1。由表 1 可看出,相比于其他基础检测模型,YOLO v10 效果最佳,精确率为 87.01%,召回率为 81.35%,mAP@0.5 为 89.16%。所提 SAMF-YOLO v10 在关键指标上均优于 YOLO v10,证实了 MFC 模块和 SCSA 模块相结合的有效性。其中,SAMF-YOLO v10 精确率为 90.18%,召回率为 86.28%,mAP@0.5 为 93.03%,分别比 YOLO v10 提高 3.17、4.93、3.87 个百分点。SAMF-YOLO v10 模型参数量为 7.28×10^6,浮点运算量为 2.741×10^{10},推理时间为 9.12 ms。尽管改进的模型在计算复杂度方面并非最优,但其他指标优于其他模型。因此,SAMF-YOLO v10 模型的整体检测性能更优,能够满足火龙果检测的实际需求。

不同模型火龙果检测结果如图 5 所示,由图 5 可看出,SAMF-YOLO v10 模型能够更清晰地聚焦于火龙果目标区域,显著抑制由枝叶背景干扰引起的噪声。同时能够有效地增强对小目标,尤其是远处或部分遮挡的火龙果感知能力,验证了模型在火龙果目标检测任务中的有效性。

表 1　不同模型性能测试结果

Tab. 1　Performance test results of different models

模型	精确率/%	召回率/%	mAP@0.5/%	参数量	浮点运算量	推理时间/ms
YOLO v5	83.26	75.81	84.65	9.15×10^{6}	2.421×10^{10}	9.74
YOLO v8	84.99	79.95	87.13	1.12×10^{7}	2.885×10^{10}	11.64
YOLO v9	85.51	80.48	88.22	7.31×10^{6}	2.763×10^{10}	10.03
YOLO v10	87.01	81.35	89.16	8.12×10^{6}	2.514×10^{10}	8.42
YOLO 11	85.79	78.32	86.64	9.45×10^{6}	2.176×10^{10}	7.50
YOLO 12	85.84	79.46	87.66	9.28×10^{6}	2.173×10^{10}	7.14
SAMF－YOLO v10	90.18	86.28	93.03	7.28×10^{6}	2.741×10^{10}	9.12



图 5　不同模型检测结果

Fig. 5　Detection result of different models

2.4　消融试验

2.4.1　不同模块性能测试

表 2 为 YOLO v10 模型分别与 MFC 模块和 SCSA 模块进行不同组合得到的检测结果。由表 2 可知，当 MFC 模块独立用作特征提取时，即位于第 0、1、3、5、7、17 阶段中，精确率为 87.56%，召回率为 82.15%，mAP@0.5 为 89.78%，相较于 YOLO v10 模型有所提升。当仅在第 10 阶段中使用 SCSA 模块时，各项指标性能同样得到了改善，说明 SCSA 模块显著增强了模型对上下文信息的理解，使火龙果检测和定位更加精确。当 2 个模块联合使用时，效果最好，mAP@0.5 增至 93.03%，表明该模型能够提高火龙果检测效果。

表 2　不同模块性能测试结果

Tab. 2　Performance test results of different modules

%

模型	精确率	召回率	mAP@0.5
MFC	87.56	82.15	89.78
SCSA	88.13	83.74	90.79
SAMF－YOLO v10	90.18	86.28	93.03

2.4.2　MFC 模块中不同分支性能测试

为了验证 MFC 模块中不同分支的有效性，进行了一系列试验，并采用以下定义：①分支 1 主要包含 3×3 DWConv 和膨胀率为 3 的 3×3 DWConv。②分支 2 包含 1×1 PWConv、3×3 DWConv、1×1 PWConv。③分支 3 包含 3×3 Conv、1×3 Conv、3×1 Conv。④分支 4 仅包含 1×1 Conv。⑤所有分支的输出进行相加操作。试验中的数据集、优化器、训练周期等设置均相同。由表 3 可看出，单独使用每个分支的表现均未能超过所有分支输出相加的效果。因此，通过将不同分支的特征进行融合，模型能够更充分地捕捉多层次的特征信息，验证了 MFC 模块中

多分支结构的有效性和互补性。

表 3 不同分支性能测试结果

Tab. 3 Performance test results of different branches %

类型	精确率	召回率	mAP@0.5
分支 1	88.51	85.48	91.23
分支 2	88.07	84.52	91.21
分支 3	87.92	83.58	90.63
分支 4	87.01	82.67	90.05
所有分支输出相加	90.18	86.28	93.03

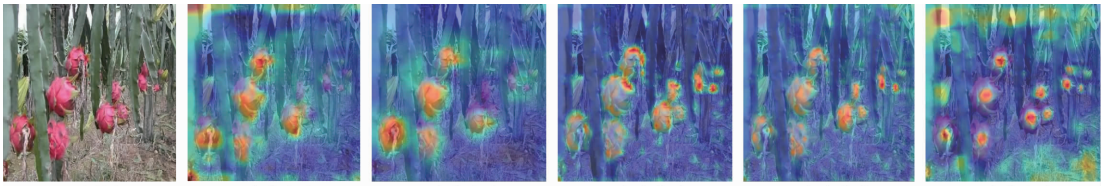
2.4.3 MFC 模块不同膨胀率性能测试

为了验证在 MFC 模块中所使用膨胀率为 3 的 3×3 DWConv 的有效性,分别设置膨胀率为 1、2、3、4 和 5 进行了试验,结果如表 4 所示。由表 4 可看出,当膨胀率为 3 时性能最佳;结果表明,合理的膨胀率设置能够有效增强网络的感受野建模能力。当膨胀率进一步增大时,尽管感受野得到扩展,但模型更易引入与目标无关的背景信息,从而导致噪声干扰增加。通过可视化热图对第 17 层输出进行了展示,如图 6 所示。由图 6 可观察到,随着膨胀率增大,网络注意力分布逐渐扩散到更多区域,尤其是一些与目标无关的背景区域。

表 4 不同膨胀率测试结果

Tab. 4 Test results of different dilation rates %

膨胀率	精确率	召回率	mAP@0.5
1	88.58	84.95	90.04
2	89.85	85.20	92.67
3	90.18	86.28	93.03
4	90.02	85.32	92.70
5	88.62	83.74	91.58



(a) 原图 (b) 膨胀率为1 (c) 膨胀率为2 (d) 膨胀率为3 (e) 膨胀率为4 (f) 膨胀率为5

图 6 不同膨胀率可视化热图

Fig. 6 Visual heat maps of different dilation rates

2.4.4 不同注意力机制性能测试

与其他常用的注意力机制进行对比,包括 SE(Squeeze-and-excitation module)[23]、CBAM(Convolutional block attention module)[24]、ECA(Efficient channel attention module)[25] 和 PSA(Polarized self-attention module)[26],添加位置为主干网络第 10 层,其余部分(第 0、1、3、5、7、17 阶段)中 MFC 模块不变,结果如表 5 所示。由表 5 可看出,使用 SCSA 注意力机制效果最好,这得益于 SCSA 模块利用空间注意力模块和通道注意力模块之间的协同效应,不仅能够增强在不同尺度上表达空间信息的能力,而且能更清晰地聚焦于标记区域,减少背景噪声。同时 SCSA 在模型参数量,浮点运算量以及推理速度上较为合理,在性能和效率之间取得较好平衡的模型。

表 5 不同注意力机制结果

Tab. 5 Results of different attention mechanisms

注意力机制	精确率/%	召回率/%	mAP@0.5/%	参数量	浮点运算量	推理时间/ms
SE	85.01	77.47	85.46	8.45×10^6	2.906×10^{10}	10.08
CBAM	86.70	80.11	88.02	8.92×10^6	2.913×10^{10}	10.55
ECA	86.24	79.43	87.61	7.14×10^6	2.702×10^{10}	8.67
PSA	81.87	76.76	83.96	9.07×10^6	3.076×10^{10}	11.07
SCSA	90.18	86.28	93.03	7.28×10^6	2.741×10^{10}	9.12

进一步对不同注意力机制在相同计算复杂度下的性能进行了分析。通过调整各个注意力模块参数配置,使得不同模型参数量保持接近,从而避免了模型规模差异导致的性能偏差。从表 6 可看出,SCSA 注意力机制依然表现优越,显著提升了模型在复杂任务中的性能。

表 6 同等计算复杂度结果

Tab. 6 Results with the same computational complexity

参数	SE	CBAM	ECA	PSA	SCSA
参数量	7.21×10^6	7.06×10^6	7.35×10^6	7.29×10^6	7.28×10^6
mAP@0.5/%	87.08	87.84	87.97	82.81	93.03

2.5　5 折交叉验证

为了验证模型稳健性，对 YOLO v10 和 SAMF－YOLO v10 模型进行 5 折交叉验证，旨在评估模型在特定数据集上的拟合能力，并确保其在不同数据子集上表现一致。每个数据集都被分为 5 个分层子集，且采用相同超参数，以确保公平性。由表 7 可知，在不同折中，YOLO v10 和 SAMF－YOLO v10 模型性能保持一致，且在各指标均达到预期的稳定水平。这一系列结果充分证明了模型稳定性和可靠性，避免了过拟合或数据集偏差导致的性能波动。

表 7　不同模块 5 折交叉验证结果

Tab.7　Five-fold cross-validation results of different modules　%

模型	类型	精确率	召回率	mAP@0.5
YOLO v10	第 1 折	84.39	80.13	87.83
	第 2 折	87.28	83.46	89.86
	第 3 折	86.75	82.28	88.48
	第 4 折	87.87	84.08	90.01
	第 5 折	85.01	81.16	88.83
	平均	86.26	82.22	89.23
SAMF－YOLO v10	第 1 折	89.26	84.83	91.85
	第 2 折	89.70	84.99	92.01
	第 3 折	90.07	85.24	92.96
	第 4 折	91.51	87.48	94.23
	第 5 折	90.33	85.32	93.08
	平均	90.17	85.57	92.83

2.6　准确率和漏检率

为了进一步验证 SAMF－YOLO v10 模型在火龙果目标检测任务中应对各种干扰因素的能力，以单幅图像中火龙果的正确检测数量和漏检数量计算准确率和漏检率，综合评估模型性能，统计结果如表 8 所示。由表 8 可看出，YOLO v10 模型未能很好地检测出因多目标重叠被错误合并的火龙果，存在准确率偏低、漏检率较高的问题，而 SAMF－YOLO v10 在复杂果园环境下表现出更高的检测准确率和更低的漏检率，有效提升了火龙果目标检测整体性能。

表 8　火龙果检测准确率和漏检率

Tab.8　Accuracy and missed detection rate of pitaya detection　%

模型	准确率	漏检率
YOLO v10	89.52	10.48
SAMF－YOLO v10	94.43	5.57

3　结论

（1）在 YOLO v10 模型基础上，提出了一种协同注意多特征融合目标检测模型 SAMF－YOLO v10。该模型通过引入 MFC 模块与 SCSA 模块，显著提升了对全局上下文信息的感知能力，同时增强了对关键特征的捕获与利用效率。

（2）试验结果表明，SAMF－YOLO v10 模型在检测精确率、召回率、平均精度均值 mAP@0.5 与计算资源消耗之间实现了良好平衡，其整体性能明显优于 YOLO v10 模型。

（3）所提 SAMF－YOLO v10 模型为复杂果园环境下火龙果高精度、实时检测提供了一种高效且可靠的解决方案。

参　考　文　献

[1] 王蔚婕，唐道邦，曹清明，等. 火龙果的营养价值及加工现状[J]. 中国食物与营养，2019，25(4)：27－30.
Wang Weijie, Tang Daobang, Cao Qingming, et al. Nutritional value and processing status of pitaya fruit[J]. Chinese Food and Nutrition, 2019, 25(4): 27－30. (in Chinese)

[2] 薛庆鑫. 火龙果营养价值高常吃对身体有益[N]. 中国食品安全报，2023－07－08(B04).

[3] 林伟，陈敏忠，林海荣，等. 火龙果种植技术现状及发展建议探讨[J]. 农业开发与装备，2023(11)：160－162.
Lin Wei, Chen Minzhong, Lin Hairong, et al. Discussion on the current status and development recommendations for pitaya fruit cultivation techniques[J]. Agricultural Development and Equipment, 2023(11): 160－162. (in Chinese)

[4] 霍迎秋，赵士超，赵国淇，等. 基于无人机多光谱的猕猴桃园冠层叶绿素含量检测方法[J]. 农业机械学报，2024，55(9)：297－307.
Huo Yingqiu, Zhao Shichao, Zhao Guoqi, et al. Detection method of chlorophyll content in canopy of kiwifruit orchard based on UAV[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(9): 297－307. (in Chinese)

[5] Zhang C L, Zhang K F, Ge L Z, et al. A method for organs classification and fruit counting on pomegranate trees based on multi-features fusion and support vector machine by 3D point cloud[J]. Scientia Horticulturae, 2020, 278: 109791.

[6] Wang C L, Tang Y C, Zou X J, et al. Recognition and matching of clustered mature litchi fruits using binocular charge-coupled device (CCD) color cameras[J]. Sensors, 2017, 17(11): 2564.

[7] Sari C A, Sari I P, Rachmawanto E H, et al. Papaya fruit type classification using LBP features extraction and naive bayes classifier[C]// 2020 International Seminar on Application for Technology of Information and Communication (iSemantic). IEEE, 2020: 28－33.

[8] 徐婷婷，宋亮，卢学鹤，等. 基于 YOLO v7－RA 的火龙果品质与成熟度双指标检测方法[J]. 农业机械学报，2024，55(7)：

405 – 414.
Xu Tingting, Song Liang, Lu Xuehe, et al. Dual-index detection method of pitaya quality and maturity based on YOLO v7 – RA [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(7): 405 – 414. (in Chinese)
[9] 张奕,朱雅丽,高健,等. 基于改进 YOLO v12 的无人机视角下冰雹灾后香梨落果检测方法[J]. 农业机械学报,2025, 56(11): 453 – 460.
Zhang Yan, Zhu Yali, Gao Jian, et al. Detection of hail-damaged fallen fragrant pears from UAV viewpoint based on improved YOLO v12[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(11): 453 – 460. (in Chinese)
[10] Chen S, Zou X, Zhou X, et al. Study on fusion clustering and improved YOLO v5 algorithm based on multiple occlusion of *Camellia oleifera* fruit[J]. Computers and Electronics in Agriculture, 2023, 206: 107706.
[11] Li C, Wu H, Zhang T, et al. Lightweight network of multi-stage strawberry detection based on improved YOLO v7 – Tiny[J]. Agriculture, 2024, 14(7): 1132.
[12] Shi B, Hou C, Xia X, et al. Improved young fruiting apples target recognition method based on YOLO v7 model[J]. Neurocomputing, 2025, 623: 129186.
[13] Qiu Z, Huang Z, Mo D, et al. GSE – YOLO: a lightweight and high-precision model for identifying the ripeness of pitaya (Dragon Fruit) based on the YOLO v8n improvement[J]. Horticulturae, 2024, 10(8): 852.
[14] Qiu Z, Zhuo S, Li M, et al. An efficient detection of the pitaya growth status based on the YOLO v8n – CBN model[J]. Horticulturae, 2024, 10(9): 899.
[15] Li X, Qin Y, Wang F, et al. Pitaya detection in orchards using the MobileNet – YOLO model[C]//2020 39th Chinese Control Conference (CCC). IEEE, 2020: 6274 – 6278.
[16] 刘进一,晏伏山,董赫,等. 基于改进 MBS – YOLO v8 的火龙果目标检测与定位方法[J]. 农业机械学报,2025,56(5): 425 – 432.
Liu Jinyi, Yan Fushan, Dong He, et al. Pitaya fruit target detection and localization method based on improved MBS – YOLO v8[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(5): 425 – 432. (in Chinese)
[17] Nan Y, Zhang H, Zheng J, et al. Pitaya detection using an improved lightweight Faster R – CNN based on MobileNetV3 in densely planted pitaya orchards[J]. Journal of Agricultural Engineering, 2025, 56(4): 1886.
[18] Wang A, Chen H, Liu L, et al. Yolov10: real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 107984 – 108011.
[19] Si Y, Xu H, Zhu X, et al. SCSA: exploring the synergistic effects between spatial and channel attention[J]. Neurocomputing, 2025, 634: 129866.
[20] Wang C Y, Yeh I H, Mark L H Y. Yolov9: learning what you want to learn using programmable gradient information[C]// European Conference on Computer Vision, 2024: 1 – 21.
[21] Khanam R, Hussain M. Yolov11: an overview of the key architectural enhancements[PP/OL]. V1. arXiv (2024 – 10 – 23) [2026 – 04 – 03]. https://doi.org/10.48550/arXiv.2410.17725.
[22] Tian Y, Ye Q, Doermann D. Yolov12: attention-centric real-time object detectors[PP/OL]. V1. arXiv(2025 – 02 – 18) [2026 – 04 – 03]. https://doi.org/10.48550/arXiv.2502.12524.
[23] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132 – 7141.
[24] Woo S, Park J, Lee J Y, et al. Cbam: convolutional block attention module[C]// Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3 – 19.
[25] Wang Q, Wu B, Zhu P, et al. ECA – Net: efficient channel attention for deep convolutional neural networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534 – 11542.
[26] Liu H, Liu F, Fan X, et al. Polarized self-attention: towards high-quality pixel-wise mapping[J]. Neurocomputing, 2022, 506: 158 – 167.