doi:10.6041/j.issn.1000-1298.2025.09.054

基于 SDE－YOLO 的矮砧密植化果园苹果检测方法

朱立成[1,2]　王文贝[1,2]　赵　博[1,2]　韩振浩[2,3]　高建波[1,2]　陈凯康[1,2]　冯旭光[1,2]

（1. 中国农业机械化科学研究院集团有限公司，北京 100083；2. 农业装备技术全国重点实验室，北京 100083；
3. 中国农业机械化科学研究院集团有限公司山东分公司，青岛 266200）

摘要：矮砧密植化苹果园是未来机器人采摘的典型应用场景，面临复杂光照、果实重叠和枝叶遮挡等挑战，精准检测果实是苹果采摘机器人的关键核心技术之一。为进一步提高矮砧密植化种植的果园中苹果的检测准确性和鲁棒性，提出一种基于 SDE－YOLO 的矮砧密植果园苹果检测模型。构建包含不同光照环境、遮挡状态的果实数据集，并对果实遮挡类型进行了统计学分类。然后，通过在骨干网络中设计复合特征提取结构，将后两层 C2f 模块替换为 Swin Transformer，增强模型建立长程依赖的能力，有效提升密集场景下的检测性能；同时主干融入 EMA 注意力机制，通过不降维的通道重构方式实现像素级自适应注意力分配，有效抑制枝叶等背景干扰，降低计算复杂度；在特征融合网络中引入 DCN v2 模块，通过动态可变形卷积提升对不同形态和姿态苹果的检测能力。最后利用 Grad－CAM 方法产生目标检测热力图，形成有效特征可视化语言，提高模型关注区域的理解能力。结果表明，SDE－YOLO 精确率、召回率和平均精度均值分别达到 88.9%、86.6% 和 94.2%，相比基线模型分别提高 2.0、1.7、3.3 个百分点，模型参数量减少 9.38%。通过与其他主流目标检测模型的对比，SDE－YOLO 在处理光照变化、果实重叠遮挡和枝叶遮挡等复杂场景时表现出更好的性能。采用本文方法可在矮砧密植化果园对苹果果实进行较准确的果实检测，为苹果采摘机器人提供有效的目标定位信息。

关键词：苹果；复杂光照；重叠遮挡；枝叶遮挡；卷积神经网络；目标检测

中图分类号：TP391.4　　**文献标识码**：A　　**文章编号**：1000-1298(2025)09-0638-10　　**OSID**：

Detection Method of Apples in Dwarf Rootstock Densification Orchards Based on SDE－YOLO

ZHU Licheng[1,2]　WANG Wenbei[1,2]　ZHAO Bo[1,2]　HAN Zhenhao[2,3]　GAO Jianbo[1,2]
CHEN Kaikang[1,2]　FENG Xuguang[1,2]

(1. *Chinese Academy of Agricultural Mechanization Sciences Group Co., Ltd., Beijing* 100083, *China*
2. *State Key Laboratory of Agricultural Equipment Technology, Beijing* 100083, *China*
3. *Shandong Branch, Chinese Academy of Agricultural Mechanization Sciences Group Co., Ltd., Qingdao* 266200, *China*)

Abstract: The apple orchard with dwarf rootstock densification is a typical application scenario for future robotic picking, facing challenges such as complex light, fruit overlapping and branch and leaf shading, and accurate fruit detection is one of the key core technologies for apple picking robots. An enhanced apple detection model termed SDE－YOLO was proposed to improve detection accuracy and robustness in such environments. Initially, a comprehensive dataset encompassing various lighting conditions and occlusion states was constructed, with systematic categorization of occlusion patterns. The model architecture incorporated several key improvements: replacement of the last two C2f modules with Swin Transformer in the backbone network to enhance long-range dependency modeling and dense scene detection capabilities; integration of EMA attention mechanism in the backbone network, achieving pixel-level adaptive attention allocation through non-dimensional channel reconstruction to effectively suppress background interference; introduction of DCN v2 module in the feature fusion network to improve detection of apples with varying morphologies and poses. Experimental results demonstrated that SDE－

收稿日期：2025－02－13　修回日期：2025－04－17

基金项目：山东省农机研发制造推广应用一体化试点项目(NJYTHSD－202323)和中国机械工业集团有限公司科研专项(ZDZX2023－2)

作者简介：朱立成(1984—)，男，正高级工程师，主要从事农业机器人技术研究，E-mail：zhulicheng@caams.org.cn

通信作者：赵博(1981—)，男，研究员，博士，主要从事农业机器人和农业工程技术研究，E-mail：zhaoboshi@126.com

YOLO achieved precision, recall, and the average precision of 88.9%, 86.6%, and 94.2% respectively, representing improvements of 2.0, 1.7, 3.3 percentage points over the baseline model, while reducing model parameters by 9.38%. Comparative analysis with mainstream detection models confirmed SDE－YOLO's superior performance in handling illumination variations and occlusions, providing accurate fruit detection for robotic harvesting in dwarf rootstock orchards.

Key words: apple; complex illumination; overlapping shading; branch and leaf shading; convolutional neural network; target detection

0　引言

苹果作为全球产量最大的水果之一，在农业经济中扮演着重要角色。近年来，中国的苹果产业迅速发展，种植面积广、产量居全球前列。随着市场需求增加和劳动力成本上升，传统果园管理特别是采摘环节面临效率低、成本高和易损伤果实等挑战[1]，制约了鲜果产业效益的进一步提升，亟需新技术介入以提高管理与作业质量。

为应对这些挑战，矮砧密植化种植模式因其高产、优质、易管理等优势成为现代果园的主要发展方向。该模式通过控制树势、密植栽培，提高了单位面积的产量，但也带来了果实分布密集、枝叶交错遮挡复杂情况，增加了智能检测的难度。深度学习的快速发展，为解决这一问题提供了新思路。智能设备在果园管理中的应用能够显著提升作业效率，减轻劳动负担，同时确保不影响果实质量的精准作业。相关应用包括智能产量预测[2]、智能疏果[3]、自动化采摘[4]和精准喷药[5]等，这些技术的应用不仅提高了果园管理的效率和精度，也为实现可持续农业生产提供了技术支持。

在目标检测算法方面，从早期的两阶段算法（如 R－CNN[6]、Faster R－CNN[7] 和 Mask R－CNN[8]）到一阶段检测算法（如 SSD[9] 和 YOLO[10]（You only look once）系列），检测效率和精度不断提升。在传统果园环境下，已有学者进行广泛探索。例如，王丹丹等[11]将 K-means 与 Normalized Cut 算法相结合应对果实重叠问题，对无遮挡和被遮挡苹果的平均分割误差分别为 3.15% 和 5.24%，但该方法仅适用于简单背景条件。TIAN 等[12]提出改进 YOLO v3 网络用于果园苹果检测，通过引入 DenseNet 结构处理低分辨率特征层，提高了特征利用率和检测性能，但在密集种植环境下仍存在检测盲区。徐利锋等[13]提出的多尺度特征提取框架虽改善了小目标检测问题，但高内存占用限制了实际应用。CHEN 等[14]提出 Des－YOLO v4 苹果检测算法，通过借鉴 DenseNet 网络结构优化 YOLO v4 模型，设计基于平均精度损失的分类损失函数，并采用 Soft－NMS 算法解决目标漏检问题，但在低光环境下性能显著下降，难以应用于复杂环境。

针对矮砧密植化果园的特殊环境，冯青春等[15]提出了基于改进 YOLO v4 的多任务深度卷积网络和深度点云辅助果实质心定位方法，在一定程度上提升了果实检测与定位的鲁棒性，但在极端遮挡与动态光照条件下仍存在漏检和误检问题。整体来看，尽管已有研究取得一定成果，但多数集中于传统果园环境，对矮砧密植果园的目标检测研究仍较为稀缺。考虑到其作业空间规整、具备机械化采摘潜力，针对该类果园开展高精度目标检测研究具有重要意义，不仅有助于提升检测准确性，也为实现智能采摘与无人果园提供技术支撑。

针对上述问题，本文提出 SDE－YOLO 模型。首先对矮砧密植化果园苹果各类生长情况进行图像采集，增加数据集种类，构建真实采摘场景；针对果实重叠和遮挡问题，通过在骨干网络中引入 Swin Transformer 替换原有的 C2f，构建复合特征提取架构，增强模型建立长程特征依赖的能力。针对枝叶背景干扰和光照变化问题，在主干网络中融入高效多尺度注意力机制，采用不降维的通道重构方式，实现像素级注意力分配，抑制无关区域干扰，提升目标特征的区分能力。同时，针对果实形态和姿态多样性问题，在特征融合网络中引入 DCN v2，提升对不同形态、姿态苹果的适应性。最后通过热力图可视化关注区域，验证模型有效性。以期提高模型在密集果园环境中的检测性能，并为智慧农业提供技术支持。

1　材料与方法

1.1　图像采集与数据集构建

数据于 2023 年 10 月在山东省淄博市巴以智慧果园采集。该果园采用高纺锤树形的矮砧密植栽培模式，树体结构紧凑，行间通道宽敞，该栽培模式还为机械化作业提供了便利，从而提升整体管理效率。树木行距 3.0～4.0 m，株距 1.0～1.5 m，行间可通行宽度 1.8～2.7 m，平均冠径约 1.2 m，无枝树干高度 0.5～0.6 m。如图 1 所示。

图像拍摄于晴天和阴天，拍摄阶段包括上午、中午、下午、黄昏。图像采集设备为 Intel ®



图 1 矮砧密植标准化果园

Fig. 1 Standardized orchard of dwarf rootstock

RealSense™ D455 型相机,图像分辨率为 1 280 像素 × 720 像素,在不同拍摄距离下(0.5 ~ 1.0 m)选择多种角度(俯拍、仰拍、平拍)进行图像采集,将获得的图像进行人工筛选,剔除过度曝光、过于模糊、严重遮挡等影响果实形状特征提取的图像,以提升数据质量,经筛选处理后共获得数据集 2 700 幅,包含顺光、逆光、树枝遮挡、树叶遮挡、果实重叠、单目标等多种情景。数据集组成如表 1 所示,图 2 为复杂环境中不同条件下的数据集样本。

表 1 矮砧密植化果园苹果数据集组成

Tab. 1 Composition of apple dataset for dwarf rootstock densification orchards

样本	树枝遮挡	树叶遮挡	果实重叠	总计
顺光	377	757	378	1 512
逆光	360	446	382	1 188
总计	737	1 203	760	2 700



图 2 矮砧密植化果园苹果样本

Fig. 2 Dwarf rootstock densification orchard apple sample

使用 LabelImg 标注软件对数据集进行人工标注,标注格式为 YOLO,将标注后数据集按照比例 8∶1∶1随机划分为训练集、验证集、测试集。

1.2 算法模型构建

1.2.1 YOLO v8

YOLO 作为一种单阶段目标检测模型,采用单个卷积神经网络(Convolution neural network,CNN)对图像进行处理,能够同时输出物体的类别和位置坐标。该模型通过端到端的方式实现目标定位与分类,显著提升了检测速度,尤其在实时性要求较高的场景中表现出色。YOLO v8[16] 是 YOLO 系列中经典的算法,其采用多种优化策略来实现速度和精度之间的平衡,共包括 5 个架构:YOLO v8n、YOLO v8s、YOLO v8m、YOLO v8l、YOLO v8x。

本研究选择 YOLO v8n 作为基础网络,主要基于以下考虑:尽管 YOLO 11 模型已相继推出,但其在标准数据集上的平均精度均值相较 YOLO v8n 仅提升约 2.2 个百分点,且在 GPU 计算性能方面表现欠佳。在实际应用场景中,该微小的性能增益难以弥补额外的计算开销。因此,本研究通过针对性的网络改进策略,以期在 YOLO v8n 原有基础上实现性能的进一步提升。其次,考虑到矮砧密植化果园中复杂的光照环境和果实遮挡问题,模型的稳定性和优化潜力显得尤为重要。YOLO v8 作为经过广泛验证的框架,不仅提供了良好的性能基础,也为后续优化提供了充分改进空间,针对特定应用场景,在保持较低计算复杂度的同时,有效提升模型的检测性能。

1.2.2 SDE - YOLO

YOLO v8n 网络主要由主干网络(Backbone)、特征融合网络(Neck)和预测头(Head)3 部分组成。针对矮砧密植化果园苹果数据集的特点,在网络的 3 个关键部分进行重构,以提高目标检测的准确性。SDE - YOLO 模型的结构如图 3 所示。

(1)针对矮砧密植环境下果实重叠和遮挡问题,在主干网络中设计复合特征提取结构,将部分 C2f 模块替换为 Swin Transformer。既保留了 C2f 模块对局部特征的精细提取能力,又通过 Swin Transformer 增强了模型建立长程依赖关系的能力,能够有效捕获像素间的远程关联,提升了模型在矮砧密植种植环境下对果实重叠和遮挡场景的检测能力,减少因重叠或遮挡导致的漏检情况。

(2)针对复杂背景干扰和光照变化问题,在主干网络的关键特征提取层中融入高效多尺度注意力(Efficient multi-scale attention,EMA),其结构如图 3 所示,其中,H、W 表示特征图高度、宽度,C 表示通道数,G 表示分组数,C // G 表示分组后每个子组的通道数。该机制通过不降维的通道重构方式,实现了像素级的自适应注意力分配。这使得模型能够更有效地抑制密集树叶、枝干等背景干扰,并且增强了不同光照条件下的特征表达能力,从而提高了模型在光照变化环境中的鲁棒性。

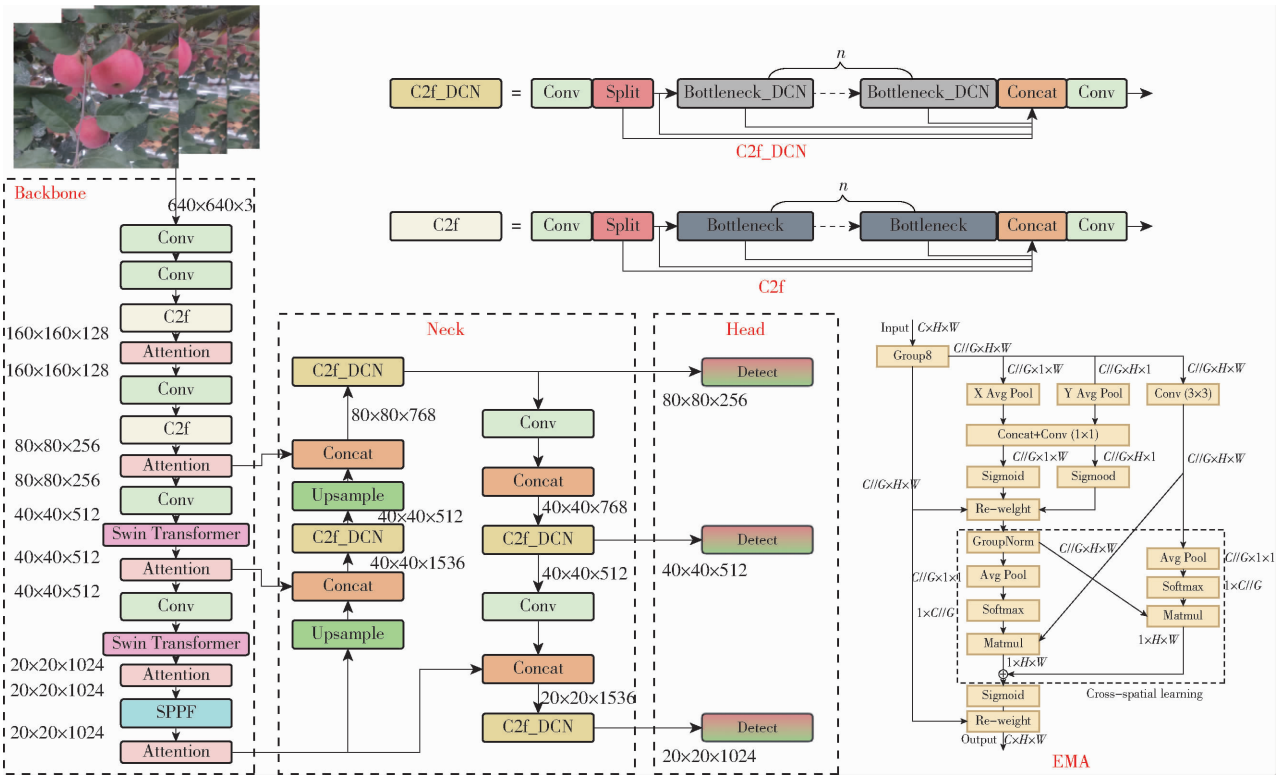


图 3　SDE－YOLO 模型结构图

Fig. 3　Network structure diagram of SDE－YOLO

（3）针对苹果形态和姿态多样性问题，在 Neck 网络中策略性地引入 DCN v2（Deformable convolutional networks v2）模块。通过其动态可变形卷积特性，实现了卷积核采样位置的自适应调整，从而增强了模型对不同生长状态和观察角度下果实的检测能力，提升了对多样目标的特征提取及识别效果，有效应对因生长环境和视角差异导致的果实形态变化问题。

1.2.3　Swin Transformer 模块

深入分析 YOLO v8 主干网络结构发现，后两层 C2f 模块作为深层特征提取单元，在网络架构中承担着高层语义特征提取和长距离特征关联的关键任务。这一位置对于矮砧密植化果园中的目标检测尤为重要，主要是由于果实的密集种植模式和复杂的生长环境对模型的特征提取能力提出了更高要求。然而，传统 C2f 模块仅依靠卷积层堆叠形成的局部感受野，在处理矮砧密植环境下果实间复杂的空间关系时存在明显局限性，难以有效建立远程特征依赖，导致重叠果实检测精度下降。基于此，本研究构建了一种新型复合特征提取架构，提出将这两层 C2f 模块修改为具有强大长程依赖建模能力的 Swin Transformer[17] 结构。

在所提改进结构中，Swin Transformer 通过其独特的层次化设计和注意力机制，有效克服了上述局限性。具体而言，窗口多头自注意力（Window multi-head self-attention，W－MSA）模块在局部窗口内构建像素级特征关联，而偏移窗口多头自注意力（Shifted window multi-head self-attention，SW－MSA）则通过巧妙的窗口移动机制实现了跨区域的信息交互。这种设计不仅显著降低了计算复杂度，更重要的是增强了模型对不同尺度下果实特征的感知能力。特别是在处理被枝叶遮挡或果实重叠等复杂场景时，移动窗口机制能够有效建立远程特征关联，大幅提升检测性能。

本文复合特征提取结构如图 4 所示，优势主要体现在以下方面：①保留的 C2f 继续承担局部特征的提取任务，保证了模型对目标细节特征的捕获能力。②新增的 Swin Transformer 针对密集果实重叠问题，通过其层次化的窗口注意力机制增强了对全局信息的感知，两种特征提取机制形成互补，使网络能够同时获取果实的局部细节和全局上下文信息，同时通过自注意力机制增强了对不同光照条件下的特征适应能力，在光照变化剧烈的场景中保持了稳定的检测性能。③Swin Transformer 中引入的相对位置偏置（Relative position bias）机制进一步增强了网络对果实空间分布特征的捕获能力，这对提升密集场景下的检测准确性具有重要意义。

1.2.4　高效多尺度注意力机制

深入分析矮砧密植化果园环境发现，复杂的背景干扰和光照变化严重影响目标检测性能。特别是

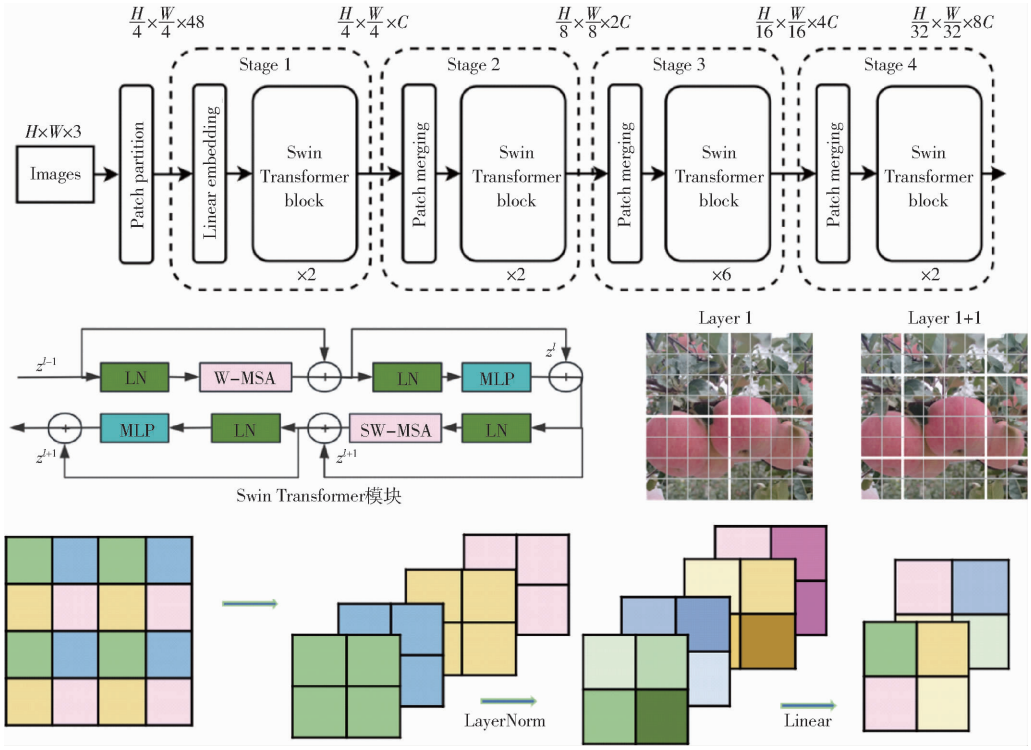


图 4 Swin Transforme 模块结构图

Fig. 4 Structure diagram of Swin Transforme module

在果实密集生长区域,枝叶遮挡和果实重叠导致背景特征与目标特征难以有效区分,影响检测准确率。传统注意力机制虽可增强目标区域关注度,但通道降维操作造成关键特征信息损失[18],难以满足复杂果园场景的检测需求。

针对上述问题,通过在 C2f、Swin Transformer 及 SPPF 结构后嵌入高效多尺度注意力模块[19],构建了层次化的注意力增强网络。EMA 模块采用不降维的特征重塑方式,具有以下技术创新:①设计并行的 1×1 和 3×3 卷积子网络,实现多尺度特征的自适应融合。②引入二维全局平均池化操作,增强感兴趣区域的特征表达。③通过 cross-spatial learning 机制实现像素级注意力分配。

这种创新设计在技术实现上具有显著优势:首先,多尺度特征融合提升了模型对不同生长阶段苹果的识别能力;其次,像素级注意力分配机制有效抑制了复杂背景干扰,提高了逆光和弱光环境下的检测灵敏度;最后,特征维度的完整保留确保了细节信息不被损失。具体网络流程如下:

(1)特征分组:EMA 模块将输入特征图 $X \in \mathbf{R}^{C\times H\times W}$(其中 $\mathbf{R}$ 表示实数空间,即特征图中元素的取值为实数)沿通道维度分为 G 组,得到 G 组子特征图 $X=(X_0,X_1,\cdots,X_i,\cdots,X_{G-1})$,其中每个子特征图 $X_i \in \mathbf{R}^{\frac{C}{G}\times H\times W}$。这种分组策略有助于均匀分布空间语义特征,为后续的多尺度处理奠定基础。

(2)进行并行子网络处理:EMA 模块包含 1×1 分支和 3×3 分支,前者用于捕获局部跨通道交互,保持通道信息的完整性,后者用于提取多尺度特征,增强空间信息的表达。

1×1 分支:首先对子特征图进行水平和垂直方向的一维全局平均池化,池化后的特征沿垂直维度连接,通过 1×1 卷积建立通道和空间位置信息之间的联系。经 Sigmoid 激活后生成水平和垂直维度的注意力权重图。上述过程计算式为

$$z_c^H(H)=\frac{1}{W}\sum_{0\leqslant i\leqslant W}x_c(H,i) \tag{1}$$

$$z_c^W(W)=\frac{1}{H}\sum_{0\leqslant j\leqslant H}x_c(j,W) \tag{2}$$

使用 3×3 卷积核处理子特征图,提取更大感受野的特征信息,经过 Conv 3×3 操作后,进行组归一化(GroupNorm)处理。

(3)跨空间学习模块:该模块融合 2 个分支的输出,生成最终的空间注意力权重图。对 1×1 分支输出进行全局平均池化和 Softmax 处理,与 3×3 分支输出相乘,得到第 1 个空间注意力图。对 3×3 分支输出进行类似处理,与 1×1 分支输出(经组归一化)相乘,得到第 2 个空间注意力图。2 个注意力图相加并经 Sigmoid 激活归一化。公式为

$$z_c=\frac{1}{HW}\sum_{j}^{H}\sum_{i}^{W}x_c(i,j) \tag{3}$$

式中 z_c——在第 c 个通道的全局特征表示

x_c——第 c 个通道的特征图

(4)特征加权与输出：将原始特征图与得到的注意力权重图逐元素相乘，得到突出重要区域的特征图。

1.2.5　可变形卷积网络

在果园环境中，自然生长和观察视角的差异导致果实呈现多种形态特征，加之枝叶遮挡和复杂背景的干扰，传统卷积网络采用的固定采样模式难以适应这种复杂的几何变化。针对这些问题，本研究采用了可变形卷积网络(Deformable convolutional networks，DCN)[20]，通过引入可学习的偏移量，实现对输入特征的自适应采样，有效克服了传统卷积在处理不规则物体和复杂几何变换时的局限性。

DCN v2[21]在 DCN v1 的基础上进行了关键改进。在每个采样点不仅加入偏移量，还引入了权重系数，用于区分感兴趣区域的重要性。DCN v1 仅学习调整卷积核位置的 offset，而 DCN v2 通过加入采样点权重，显著增强了对目标几何形变的建模能力。DCN v1 和 DCN v2 分别表示为

$$y_1(p) = \sum_{k=1}^{K} \omega_k x(p + p_k + \Delta p_k) \tag{4}$$

$$y_2(p) = \sum_{k=1}^{K} \omega_k x(p + p_k + \Delta p_k)\Delta m_k \tag{5}$$

式中　$y_1(p)$、$y_2(p)$——输出特征图在空间位置 p 处的特征响应

x——输入特征图

p——输出特征图的空间位置(同时作为 DCN 卷积操作的中心参考位置，用于确定输入特征图上的动态采样区域)

p_k——预先指定的偏移量

K——卷积核中采样位置的数量

ω_k——位置 k 上卷积核权重

Δp_k——偏移量，用于调整采样位置，适应局部特征的变形

Δm_k——权重系数，范围为[0,1]

基于上述 DCN v2 的优势，本研究设计了 C2f_DCN 模块。如图 3 所示，该模块采用 Split－Bottleneck_DCN－Concat 的新型架构，在 Neck 部分的 4 个 C2f 模块中将标准 3×3 卷积替换为 DCN v2。具体而言，通过 Split 操作将输入特征分为 2 条并行路径：一路经过 Bottleneck_DCN 进行深度特征提取，利用可学习的偏移量和权重系数增强对目标的感知能力；另一路保留 CSPNet 的跨层连接，维持计算效率。这种复合特征提取机制在提升特征表达能力的同时，有效平衡了计算开销。此外，多个 C2f_DCN 模块通过 Upsample 和 Concat 操作构建特征融合网络，其中 DCN v2 的可变形采样机制显著增强了模型对果实形变、旋转和遮挡的适应能力。

1.3　模型训练与评价

1.3.1　模型训练环境及策略

训练环境采用 Ubuntu 20.04 操作系统、Intel ® Xeon(R) Silver 4214R CPU @ 2.40 GHz (12 vCPU) 中央处理器和 NVIDIA GeForce RTX 3080 Ti (12 GB) GPU 的硬件平台，深度学习框架采用 PyTorch 1.11.0，编程平台为 PyCharm，编程语言为 Python 3.8，所有对比算法均在相同环境下运行。

在训练过程中，输入图像尺寸为 640 像素×640 像素，并配合使用 Mosaic 数据增强。在模型训练过程中，初始学习率为 0.001，配合优化器 Adam 调整学习率。设定权重衰减率为 0.000 5，动量设为 0.935。其余超参数相关配置采用 YOLO v8n 的 yaml 配置文件中的默认配置。

1.3.2　评价指标

采用精确率(Precision，P)、召回率(Recall，R)、F1 分数、参数量和平均精度均值(mean Average Precision，mAP)等作为评价指标。mAP@0.5 是所有类别 IoU 阈值为 0.5 时平均精度的平均值。

2　结果与分析

2.1　消融实验

为验证 SDE－YOLO 模型中各改进模块的有效性，基于基线模型(YOLO v8n)进行消融实验，结果如表 2 所示。

表 2　消融实验结果

Tab.2　Results of ablation experiment

Swin Transformer	EMA	DCN v2	精确率/%	召回率/%	参数量	mAP@0.5/%
×	×	×	86.9	84.9	3.2×10^6	90.9
√	×	×	87.3	86.7	2.8×10^6	93.1
√	√	×	86.7	88.2	2.9×10^6	94.1
√	√	√	88.9	86.6	2.9×10^6	94.2

注：√表示采用此模块，×表示不采用此模块。

引入 Swin Transformer 模块后，模型的 mAP@0.5 从基线的 90.9% 提升至 93.1%(提高 2.2 个百分点)，精确率达到 87.3%；在此基础上集成 EMA 注意力机制，使 mAP@0.5 进一步提高至 94.1%(提高 3.2 个百分点)，召回率显著提升至 88.2%；最后引入 DCN v2 模块，模型的 mAP@0.5 达到 94.2%(提高 3.3 个百分点)，精确率提升至 88.9%。实验结果表明，改进模型显著提升了在矮砧密植化果园复杂场景下的检测性能。在性能提升的同时，改进模型参数量减少了 9.38%。首先，在主干网络中引

入的复合结构虽然增加了 Swin Transformer 分支,但通过该设计减少了原有卷积路径的参数量;其次,EMA 注意力模块采用不降维的特征重塑方式,通过 1×1 和 3×3 并行卷积的轻量级设计,在保持特征提取能力的同时降低了参数开销;最后,C2f_DCN 模块通过保留 CSPNet 的部分通道连接策略,在引入可变形卷积增强特征提取能力的同时,实现了参数量的优化。这种模块设计不仅提升了模型性能,还实现了计算效率的提高。

综上所述,SDE-YOLO 检测模型的 P、R、mAP@0.5 分别达到 88.9%、86.6%、94.2%,较基线模型高 2.0、1.7、3.3 个百分点,且模型参数量减少 9.38%。

2.2 检测热力图定量分析

热力图可视化是通过颜色强度来表示模型在图像中不同区域的关注程度。在目标检测的应用中,用 Grad-CAM(Gradient-weighted class activation mapping)[22]技术生成目标检测热力图,有助于直观地了解模型如何在复杂背景下识别目标。利用此技术产生目标检测热力图并进行定量分析。为系统性评估模型性能,采用以下关键定量指标:①目标区域高响应占比(Target high-response region ratio, THRR):反映模型对目标区域特征提取的精确程度。②背景区域高响应占比(Background high-response region ratio, BHRR):用于衡量模型对背景噪声的抑制能力。③目标-背景响应比(Target-background response ratio, TBR):该指标综合评估模型区分目标和背景的能力,越高表示模型的辨别能力越强。

为验证模型改进的有效性,本文通过热力图对比(图 5)以及定量分析(表 3),对复杂场景下的目标检测性能进行系统评估。

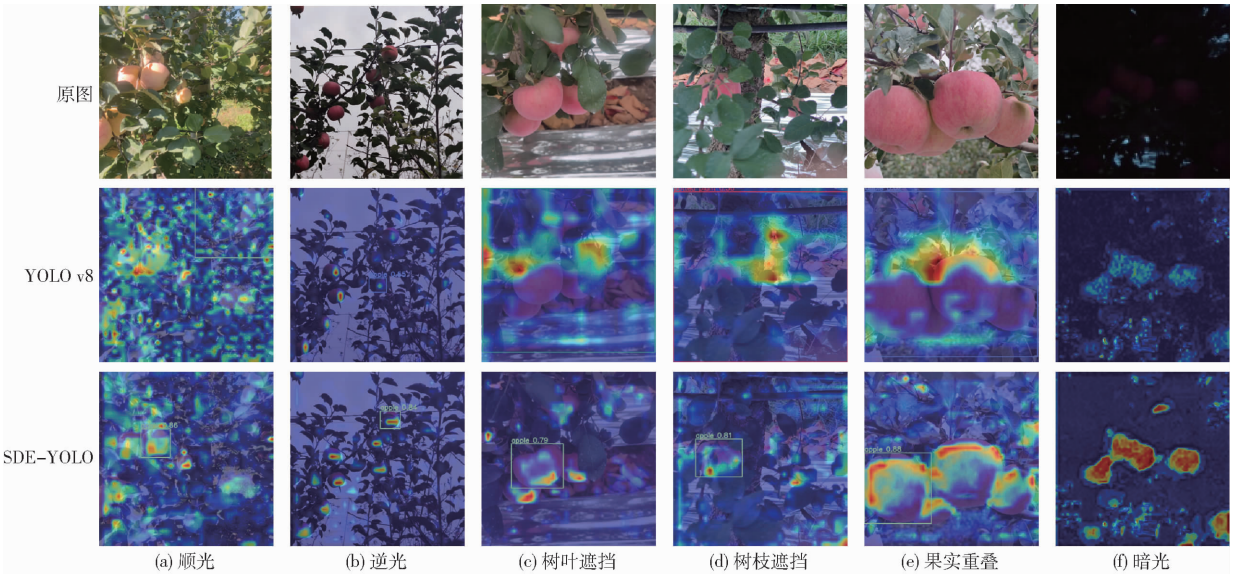


图 5 矮砧密植化果园复杂环境下模型检测热力图

Fig. 5 Heatmaps of model detection in complex environments of dwarf rootstock orchards

表 3 热力图高响应区域占比统计

Tab. 3 Proportion of high-response regions in heatmaps

参数	模型	顺光	逆光	树叶遮挡	树枝遮挡	果实重叠	暗光
目标区域高响应占比(THRR)/%	YOLO v8	70.11	10.33	31.97	56.42	31.81	77.10
	SDE-YOLO	76.17	23.42	62.16	70.44	68.80	80.25
背景区域高响应占比(BHRR)/%	YOLO v8	40.16	41.62	40.91	38.34	44.29	27.69
	SDE-YOLO	35.30	40.52	34.22	37.89	21.02	26.21
目标-背景高响应比(TBR)	YOLO v8	1.75	0.25	0.78	1.47	0.72	2.78
	SDE-YOLO	2.16	0.58	1.82	1.86	3.27	3.06

在遮挡场景中,模型对目标特征的表征能力显著增强,树叶遮挡条件下,THRR 从基准模型 YOLO v8 的 31.97% 提升至 62.16%,同时 BHRR 由 40.91% 降至 34.22%,TBR 从 0.78 提升至 1.82;果实重叠场景中 THRR 从 31.81% 提升至 68.80%,TBR 由 0.72 增至 3.27。结合热力图可视化分析表明,Swin Transformer 通过建立跨遮挡区域的长程像素依赖关系,有效恢复被遮挡目标的结构信息;DCN v2 模块则通过动态调整卷积核采样位置,在树枝遮挡场景中使 THRR 提升 14.02 个百分点,表明其对

遮挡苹果局部形变特征的捕捉能力。

针对光照复杂性，本文提出的模型在逆光环境下 THRR 从 10.33% 提升至 23.42%，TBR 由 0.25 增至 0.58；暗光场景中 TBR 从 2.78 提升至 3.06。实验分析表明，EMA 模块通过多尺度通道重构与空间注意力协同优化，在保留空间信息的同时增强弱纹理特征响应，逆光与暗光条件下的 BHRR 分别降低 1.10、1.48 个百分点，验证了模型对背景噪声的抑制能力。

进一步分析表 3 可知，SDE－YOLO 模型的 THRR 值和 TBR 值在全部场景中均呈系统性提升，各模块的协同作用形成层次化优化。多维度指标和热力图可视化分析不仅定量验证了模型的改进效果，更直观展示了 SDE－YOLO 模型在复杂果园环境中的技术优势，为精准农业目标检测提供了有价值的技术解决方案。

2.3　不同网络模型对比

为验证所提模型的有效性，本文使用 SSD、RT－DETR[23]、YOLO v5n、YOLO v6n[24]、YOLO v8n、YOLO 11n[25] 以及 YOLO v8 改进模型（YOLO v8－BiFormer），分别在苹果数据集上进行训练和测试，结果如表 4 所示。

表 4　不同模型对比结果

Tab.4　Comparison results of different models　%

模型	精确率	召回率	F1 分数	mAP@0.5
SSD	86.9	82.9	84.9	85.1
RT－DETR	87.7	76.3	81.6	87.5
YOLO v5n	88.1	78.6	83.1	85.6
YOLO v6n	84.5	77.0	80.6	86.4
YOLO v8n	86.9	84.9	85.9	90.9
YOLO v8－BiFormer	87.0	77.9	82.2	87.8
YOLO 11n	85.8	85.2	85.5	92.6
SDE－YOLO	88.9	86.6	87.7	94.2

与传统 SSD 模型相比，SDE－YOLO 精确率、召回率和 F1 分数分别提高 2.0、3.7、2.8 个百分点，mAP@0.5 提升 9.1 个百分点，表明其在复杂背景干扰场景下的检测能力得到显著提升。在光照变化和果实重叠环境下，SSD 表现欠佳，而 SDE－YOLO 通过引入 Swin Transformer 增强了长程依赖关系的建立，并且利用 DCN v2 模块实现了对不同生长状态和角度下果实的自适应特征提取。相较于 RT－DETR 及同系列的 YOLO v5n 和 YOLO v6n，SDE－YOLO 展现出全面性能优势，特别是在果实重叠场景中，其精确率提高 0.8～4.4 个百分点，召回率提高 8.0～10.3 个百分点，F1 分数提升 4.6～7.1 个百分点，mAP@0.5 提高 6.4～8.6 个百分点。性能显著提升主要得益于 DCN v2 模块的动态可变形卷积特性，其有效地应对了果实形态变化带来的检测挑战。

即使在光照剧烈变化条件下，SDE－YOLO 仍保持高检测精度，相比 YOLO 11n，该模型各项指标分别提升 3.1、1.4、2.2、1.6 个百分点，这一性能提升主要得益于 EMA 注意力机制的引入，该机制通过像素级自适应注意力分配，增强了模型在不同光照环境中的鲁棒性。尽管 YOLO v8－BiFormer 在精确率上略有提升，但其召回率和 mAP@0.5 低于基线模型，这表明其结构优化在此应用场景中未充分发挥效果。相比之下，SDE－YOLO 在多项评估指标上均优于 YOLO v8－BiFormer 和 YOLO 11n，进一步验证了所提模型更适用于矮砧密植化果园苹果检测，能够实现矮砧密植果园环境下苹果目标的快速准确识别。

2.4　复杂环境不同模型检测效果对比

在现代果园管理中，准确的果实检测技术对提高生产效率和果品质量至关重要。然而，果园环境的复杂性为目标检测系统带来了诸多挑战。不同的光照条件、遮挡情况及果实生长状态都会显著影响检测精度。

为全面评估 SDE－YOLO 模型在矮砧密植化果园环境中的性能，本研究开展了系统的对比实验，结果见表 5。传统的 SSD 模型在各类复杂场景中表现欠佳，尤其在低光照条件下，其 mAP@0.5 仅为 77.5%。RT－DETR 虽通过 Transformer 结构在果实重叠场景中获得一定提升（mAP@0.5 达 88.7%），但在复杂背景下仍存在较高误检率。

表 5　矮砧密植化果园复杂环境下不同模型 mAP@0.5 对比结果

Tab.5　Comparison of detection performance of different models in complex environments of dwarf rootstock orchards　%

模型	顺光	逆光	树叶遮挡	树枝遮挡	果实重叠	暗光
SSD	80.3	78.1	81.8	80.6	83.0	77.5
RT－DETR	85.5	82.2	85.9	83.3	88.7	80.2
YOLO v5n	84.5	80.6	85.4	82.7	84.8	78.1
YOLO v6n	85.1	84.8	86.3	83.8	85.1	84.3
YOLO v8n	88.3	85.0	90.3	87.0	89.4	84.9
YOLO 11n	88.7	86.4	91.0	89.0	90.2	85.7
SDE－YOLO	91.1	87.1	91.6	90.8	92.7	86.4

YOLO 系列模型通过版本迭代在各场景中的性能不断提升。检测结果如图 6 所示，图中黄色圈出区域为漏检，绿色圈出区域为误检。最新版 YOLO 11n 在树叶遮挡环境下 mAP@0.5 达到 91.0%，在

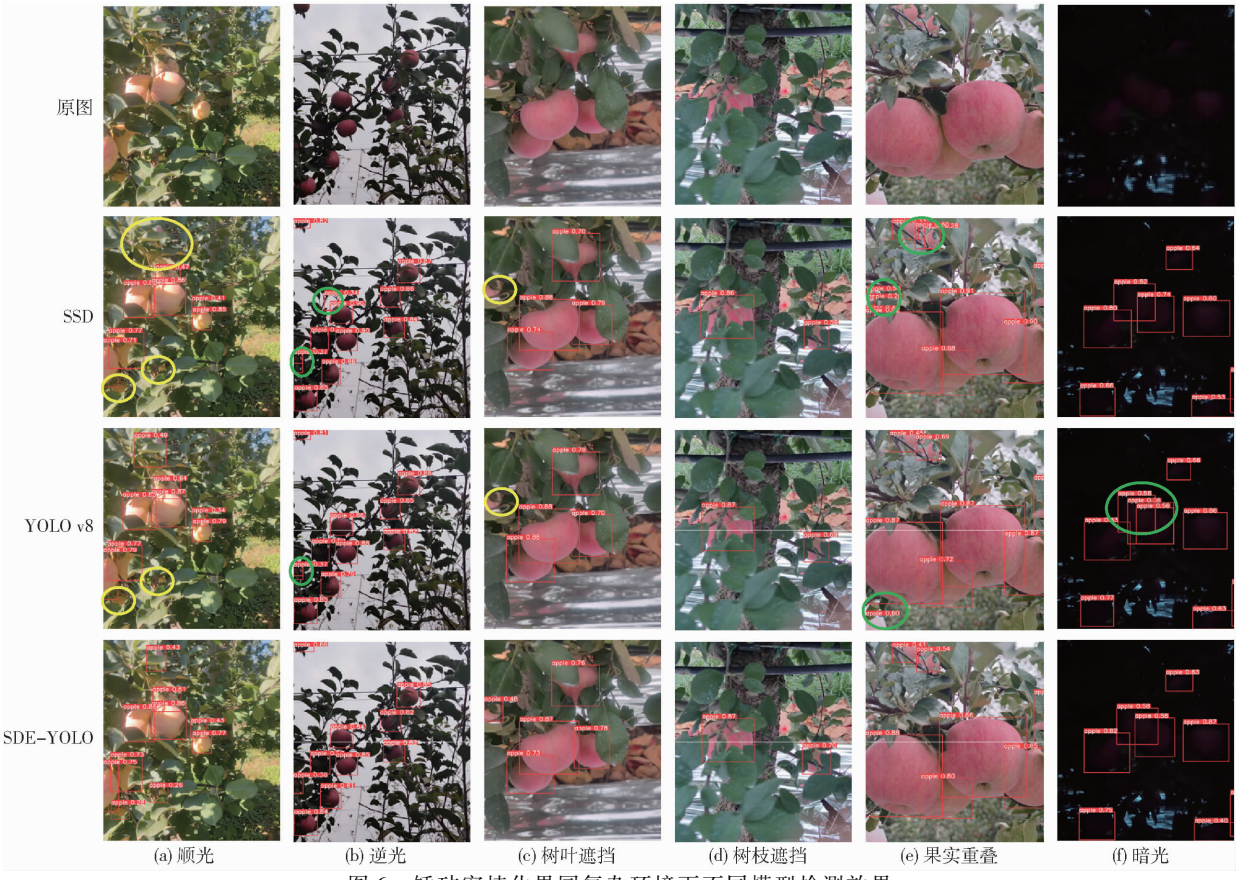


图 6 矮砧密植化果园复杂环境下不同模型检测效果

Fig. 6 Detection results of different models in complex orchard environments with dwarf rootstock orchards

果实重叠场景中达到 90.2%,展现了较强的特征提取能力。然而,其在光照变化剧烈的情况下表现仍存在一定局限性。所提 SDE-YOLO 通过综合优化结构,针对性地解决了矮砧密植化果园环境中的特定挑战:

针对果实重叠和遮挡问题,通过 Swin Transformer 增强长程依赖关系建立能力,既保留了对局部特征的精细提取,又提升了对重叠和遮挡场景的检测能力,使 mAP@0.5 在果实重叠场景下达到 92.7%,较 YOLO 11n 提高 2.5 个百分点,有效减少了漏检情况。

针对复杂背景干扰和光照变化问题,本研究引入了 EMA 注意力机制,实现了像素级自适应注意力分配,这一机制有效抑制了背景干扰并增强了不同光照条件下的特征表达能力,结果表明,在逆光条件下和树叶遮挡场景下,mAP@0.5 分别达到 87.1% 和 91.6%,较 YOLO 11n 分别提高 0.7、0.6 个百分点。

在 Neck 网络中策略性引入 DCN v2 模块,通过动态可变形卷积特性适应卷积核采样位置,从而增强了对不同生长状态和观察角度下果实的检测能力,这一改进使得 mAP@0.5 在树枝遮挡场景下达到 90.8%,有效应对了苹果形态和姿态多样性挑战。

综合对比实验结果,SDE-YOLO 在复杂环境下的检测精度和鲁棒性均优于现有模型,尤其在果实重叠与光照变化剧烈的场景中,展现出更强的特征提取与目标识别能力,为矮砧密植果园果实检测提供了更可靠的解决方案。

3 结论

(1)针对矮砧密植化果园中苹果采摘任务面临的精度低、鲁棒性差等技术难题,提出了 SDE-YOLO 目标检测模型。通过构建涵盖不同光照强度、多种拍摄角度以及各种遮挡程度的多样化数据集,使模型训练数据更贴近实际采摘场景,从而有效提升了模型对果实的精准识别能力。

(2)通过在骨干网络中引入 Swin Transformer,增强了模型在密集场景下建立长程特征依赖的能力,显著改善了果实重叠与遮挡场景下的检测性能;主干网络嵌入 EMA 注意力模块实现了像素级自适应特征加权,有效抑制了枝叶等非目标区域干扰并提升了模型在光照变化条件下的适应能力;在特征融合网络中整合 DCN v2 提升了对多样形态目标的自适应特征提取能力。这些针对性优化策略共同提高了算法在复杂果园环境中的检测性能,有效降低

了误检率和漏检率。

(3)在自建的矮砧密植化果园苹果数据集上进行测试,结果表明,SDE－YOLO 模型精确率、召回率和平均精度均值分别达到 88.9%、86.6% 和 94.2%,相比基线模型 YOLO v8n 分别提高 2.0、1.7、3.3 个百分点,模型参数量减少 9.38%。本研究在提升目标检测精度的同时,有效降低了模型的计算复杂度,实现了性能与效率的均衡提升。

参 考 文 献

[1] 王杰,闫肖肖. 水果采摘装置的发展[J]. 科技创新与应用, 2018(30):78－79.

[2] OIKONOMIDIS A,CATAL C,KASSAHUN A. Deep learning for crop yield prediction: a systematic literature review[J]. New Zealand Journal of Crop and Horticultural Science, 2023, 51(1): 1－26.

[3] WANG D,HE D. Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning[J]. Biosystems Engineering, 2021, 210: 271－281.

[4] BAI Y,MAO S,ZHOU J, et al. Clustered tomato detection and picking point location using machine learning-aided image analysis for automatic robotic harvesting[J]. Precision Agriculture, 2023, 24(2): 727－743.

[5] DANBAKI C A, ONYEMACHI N C, GADO D S M, et al. Precision agriculture technology: a literature review[J]. Asian J. Adv. Res. Rep., 2020, 14: 30－34.

[6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition,2014: 580－587.

[7] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137－1149.

[8] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]// 2017 IEEE International Conference on Computer Vision (ICCV), 2017: 2980－2988.

[9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]// European Conference on Computer Vision. Cham: Springer International Publishing, 2016: 21－37.

[10] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779－788.

[11] 王丹丹, 徐越, 宋怀波,等. 融合 K-means 与 Ncut 算法的无遮挡双重叠苹果目标分割与重建[J]. 农业工程学报, 2015, 31(10): 227－234.
WANG Dandan, XU Yue, SONG Huaibo, et al. Fusion of K-means and Ncut algorithm to realize segmentation and reconstruction of two overlapped apples without blocking by branches and leaves[J]. Transactions of the CSAE, 2015, 31(10): 227－234. (in Chinese)

[12] TIAN Y, YANG G, WANG Z, et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model[J]. Computers and Electronics in Agriculture, 2019, 157: 417－426.

[13] 徐利锋,黄海帆,丁维龙,等.基于改进 DenseNet 的水果小目标检测[J].浙江大学学报(工学版),2021,55(2):377－385.
XU Lifeng, HUANG Haifan, DING Weilong, et al. Detection of small fruit target based on improved DenseNet[J]. Journal of Zhejiang University(Engineering Science),2021,55(2):377－385. (in Chinese)

[14] CHEN W, ZHANG J, GUO B, et al. An apple detection method based on Des-YOLO v4 algorithm for harvesting robots in complex environment[J]. Mathematical Problems in Engineering, 2021, 2021(1): 7351470.

[15] 冯青春,赵春江,李涛,等. 苹果四臂采摘机器人系统设计与试验[J]. 农业工程学报,2023,39(13):25－33.
FENG Qingchun, ZHAO Chunjiang, LI Tao,et al. Design and test of a four-arm apple harvesting robot[J]. Transactions of the CSAE, 2023, 39(13):25－33. (in Chinese)

[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] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012－10022.

[18] GUO M H, XU T X, LIU J J, et al. Attention mechanisms in computer vision: a survey[J]. Computational Visual Media, 2022, 8(3): 331－368.

[19] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning[C]// ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1－5.

[20] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 764－773.

[21] ZHU X, HU H, LIN S, et al. Deformable convnets v2: more deformable, better results[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9308－9316.

[22] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: visual explanations from deep networks via gradient-based localization[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 618－626.

[23] 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.

[24] LI C, LI L, JIANG H, et al. YOLOv6: a single-stage object detection framework for industrial applications[J]. arXiv Preprint, arXiv:2209.02976, 2022.

[25] KHANAM R, HUSSAIN M. Yolov11: an overview of the key architectural enhancements[J]. arXiv Preprint, arXiv:2410.17725, 2024.