

LI Wenqin, SUI Tingting, ZHANG Zhang, WU Huiying, CHANG Liang. An Accurate Detection Method for Fruits and Vegetables in the Space Station Cargo Bay (in Chinese). *Chinese Journal of Space Science*, 2026, 46(3): 749–761. CSTR:32142.14.cjss.2025–0080. DOI:10.11728/cjss2026.03.2025–0080

面向空间站货舱的果蔬精准检测方法*

李文琴¹ 随婷婷¹ 张璋^{2,3}
吴会英^{2,3} 常亮^{2,3}

1(上海电机学院电子信息学院 上海 201306)

2(中国科学院微小卫星创新研究院 上海 201203)

3(上海微小卫星工程中心 上海 201203)

摘要 机器视觉技术在空间站货舱管理的实际应用方面处于初步阶段,为了解决空间站货舱的狭小空间、遮挡和光照等问题导致的检测精度不足,提出一种基于 YOLO11 的空间站货舱果蔬检测改进算法——LEBR-YOLO. 该方法把卷积改进为结合空间信息和边缘信息的高效输入特征提取干层,同时添加双层注意力机制,提高了提取特征的能力. 引入改进的轻量级共享可变检测模块,提高了遮挡情况下的检测能力. 使用迁移学习作为优化模型的方法,弥补数据集的不足,提高泛化能力. 实验表明,该方法在自制果蔬类数据集上达到 95.3% 的准确率、88.6% 召回率和 93.9% mAP@0.5, 同时依然保持较低的模型复杂度,满足轻舟货运飞船在轨运行的需要. 该方法可有效应用于空间站水果、蔬菜类物品检测,提高了检测精度,有效减少了误检和漏检.

关键词 货舱果蔬检测, YOLO11, 特征提取, 共享卷积, 迁移学习

中图分类号 V19

An Accurate Detection Method for Fruits and Vegetables in the Space Station Cargo Bay

LI Wenqin¹ SUI Tingting¹ ZHANG Zhang^{2,3}
WU Huiying^{2,3} CHANG Liang^{2,3}

1(School of Electronic Information Engineering, Shanghai Dianji University, Shanghai 201306)

2(Innovation Academy for Microsatellites, Chinese Academy of Sciences, Shanghai 201203)

3(Shanghai Engineering Center for Microsatellites, Shanghai 201203)

Abstract Machine vision technology remains in its nascent stage for practical application in space station cargo bay management, with limited research currently addressing target detection in space station environments. To overcome detection accuracy limitations caused by confined spaces, obstructions,

* 国家自然科学基金项目资助 (62103256)

2025-05-16 收到原稿, 2025-11-27 收到修定稿

©The Author(s) 2026. This is an open access article under the CC-BY 4.0 License
(<https://creativecommons.org/licenses/by/4.0/>)

and lighting conditions in cargo bays, this study proposes an enhanced YOLOv11-based algorithm for fruit and vegetable detection: LEBR-YOLO. Drawing on successful implementations of existing visual detection techniques, this approach refines the original convolutional neural network architecture by integrating spatial and edge information *via* a dual-layer attention mechanism, thereby enhancing processing efficiency for high-resolution feature maps. Specifically, it improves feature extraction capabilities by modifying the original convolutional module into an efficient input feature extraction layer that fuses spatial and edge information. Concurrently, a dual-layer attention mechanism is incorporated to significantly boost the model's processing efficiency for high-resolution feature maps. An enhanced lightweight shared deformable detection module is introduced, which adopts a shared convolutional architecture combined with deformable convolutions; a dynamic adjustment mechanism integrating category loss and bounding box loss is also employed to improve detection performance under occlusion. Transfer learning is used as an optimization technique to compensate for dataset limitations, reducing computational costs while enhancing model generalization. Experiments demonstrate that this model significantly improves object detection under occlusion: on a custom fruit and vegetable dataset, it achieves 95.3% accuracy, 88.6% recall, and 93.9% mAP@0.5, while maintaining low model complexity. This meets the detection requirements for the Tianzhou cargo spacecraft during in-orbit operations. This approach proves highly effective for detecting fruits and vegetables in space stations, enhancing detection accuracy, substantially reducing false positives and false negatives, and elevating the automation level of on-board resource management.

Key words Cargo hold fruits and vegetables detection, YOLO11, Feature extraction, Shared convolution, Transfer learning

0 引言

随着中国空间站建设的推进和长期载人航天任务的开展,如何高效、安全地进行货物运输和管理成为了一个亟待解决的关键问题^[1]. 机器视觉和目标检测技术快速发展,基于机器视觉的检测方法在航天领域虽已进入探索阶段^[2],但目前针对空间站环境下的目标检测研究较少,航天领域的货物管理仍使用传统的人工方法^[3,4]. 由于空间站内的货物种类繁多,且涉及航天员的日常使用,要求对果蔬货物进行精确的存储和运输管理. 因此需要一种能适应空间站特殊环境的精确果蔬检测方法,来确定货舱内部果蔬种类及航天员的使用情况.

空间站货舱的环境对视觉检测系统提出了较高的检测精度要求,同时算法需要具备实时处理和响应能力. 空间站货舱内的复杂环境会对检测效果造成影响,主要包括:光照变化导致目标边缘模糊;狭小空间及物体堆叠产生的严重遮挡;背景复杂性造成特征干扰;微重力条件下物体漂浮或旋转,使检测目标姿态不稳定. 以上因素均对检测的鲁棒性提出了挑战. 因此从现有的视觉检测技术的成功实践出发,研究一种

更适合在空间站环境下使用的目标检测技术,对航天领域的智能化发展具有重要意义. YOLO 系列是目标检测领域中非常重要的一系列算法,其具有高效的实时处理能力和较高的准确率^[5],广泛应用于农业,工业等领域. Gai 等^[6]提出了一种基于 YOLOv8 的蓝莓果实检测方法,克服了果实特征相似和叶片遮挡等问题,有效提高了蓝莓检测的精度,提高了网络泛化能力. Jiang 等^[7]为解决在复杂环境下不同成熟度草莓目标检测精度较低的问题,实现复杂环境下的草莓成熟度快速检测,提出了一种草莓成熟度检测模型 PCIA-YOLO. 其模型不仅能够快速准确地检测复杂环境下不同成熟度的草莓,且更加适合部署在嵌入式设备上. 在遥感领域, Liu 等^[8]提出了 YOLO-class 的模型,通过优化小目标、密集目标和遮挡目标的检测,以及使用代表性批量归一化、Mish 激活函数和 Vari-Focal 损失函数来提高模型的分类准确性,最终在复杂背景、不同光照条件和不同拍摄角度下提高了检测精度和速度. Liu 等^[9]提出 YOLOv8-FDD 是一种改进的 YOLOv8 模型,通过优化检测头、增强特征交互和多尺度特征提取,实现了在保持实时性的同时提高车辆检测的准确性和泛化能力. 工业上, Sui 等^[10]设

计了一种基于 YOLOv8 改进的深度学习缺陷检测网络, 通过双路径注意力机制、多特征融合和并行空间金字塔池化, 提升铝型材表面小缺陷和高背景相似度缺陷的检测精度. Zheng 等^[11] 针对图像中背景复杂、目标小、分布密集等问题, 提出了一种改进的 DES-YOLO 方法. 通过引入可变形注意力模块 (DAM), 网络可动态关注关键区域, 提高物体识别和定位精度; 采用高效交并比 (EIoU) 损失函数, 减少低质量样本影响, 增强泛化能力和检测精度. Shao 等^[12] 依据不同的改进思路或应用场景对 YOLO 算法进行了系统的分类归纳, 为本文研究提供了理论参考.

与传统机器学习算法相比, 神经网络较大的劣势在于使用成本较高, 尤其是在处理图像识别、语音识别等实际问题时, 模型训练需要海量的训练数据支撑. 迁移学习在数据稀缺的情况下能够利用源域的丰富数据来训练模型, 减少对目标域大量标记数据的需求, 加速模型训练并提高其泛化能力. 其有助于模型快速适应新任务, 避免过拟合, 并在类别不平衡或数据分布不均的情况下提高模型性能^[13]. Huang 等^[14] 提出一种将迁移学习与传统 Q-learning 相结合的多智能体合作策略学习方法. Yan 等^[15] 提出一个基于深度迁移学习的多源异构数据融合框架, 利用迁移学习技术将预训练模型从源域迁移到目标域, 以实现工具磨损的跨场景监测. 上述研究表明, 迁移学习对于跨场景的目标检测也能有效提高模型的检测性能.

现有的目标检测算法在空间站此类场景中无法取得理想效果. 针对上述问题, 本文提出了一种基于 YOLO11 改进的检测方法, 探索能够提供适用于真实场景的稳健解决方案的算法. 针对货舱内部果蔬检测过程中面临的诸多问题, 做出了以下改进措施. 一是设计高效输入特征提取层 (Efficient Input Embedding Stem, EIEstem), 在传统卷积的基础上, 结合边缘检测 (Sobel 卷积) 与池化操作, 以增强特征提取能力; 二是引入双层路由注意力机制 (BiLevelRoutingAttention, BRA_nchw), 结合局部注意力与区域路由机制, 有效提升模型对高分辨率特征图的表达能力与处理效率; 三是设计轻量级共享可变形检测模块 (Lightweight Shared Deformable Detection, LSDDetect), 使用了共享卷积结构, 并且引进可变形卷积, 同时采用结合类别损失和边界框损失的动态调整机制, 以优化目标检测性能; 四是通过迁移学习将

收集的果蔬数据集运用到空间站果蔬检测中, 提升效率、节省成本, 并增强模型的准确性和泛化能力.

1 YOLO11

YOLO11 结合了准确性、速度和效率, 在边缘设备上更高效、更快^[16,17]. YOLO11 引入 C3K2 模块 (减少计算量同时保留残差结构特征)、C2 PSA 机制 (在 C2 模块中引入多头注意力增强特征交互), 在保持轻量化的同时提升特征表达能力. YOLOv8 强调轻量化与速度, YOLOv10 引入端到端解码机制, 而 YOLO11 更注重在边缘设备上的精度与效率平衡. 模型结构如图 1 所示. Backbone 通过卷积层和反卷积层提取特征, 利用残差连接和瓶颈结构减小网络规模, 并提高性能. 该部分采用 C3K2 模块为基本单元, 并引入 C2 PSA 机制, 在 C2 模块中嵌入了多头注意力机制, 使特征提取能力更强. Neck 融合来自 Backbone 的多尺度特征图, 并通过结合 FPN 和 PANet 结构进一步精炼这些特征. 在 Head 部分, YOLO11 通过引入两个深度可分卷积 (DWConv), 有效减少了参数量和计算量, 同时保持了跨通道的信息流动, 进一步提升了模型的效率和精度. 整体来看, 该模型在现有主流目标检测方法中, 其性能更适配空间站货舱在轨运行的实际需求. 因此尝试对 YOLO11 进行改进, 继续提高模型的精度, 以满足实际空间站货舱的检测需求.

2 LEBR-YOLO

针对空间站货舱环境的特殊性 & YOLO11 算法存在的缺陷, 设计一种面向空间站货舱内部果蔬类货物的精准检测方法. 图 2 为 LEBR-YOLO 模型结构.

2.1 高效输入特征提取层

在果蔬检测过程中, 由于手部遮挡以及光线问题, 原来的模型会出现误检的情况, 针对这个问题, 设计了高效输入特征提取层, 该模块使用 Sobel 卷积来提升模型对图像特征的提取和表示能力. 同时该模块使用轻量级设计, 在保持特征提取效果的同时, 还具备低延迟优势, 适合需要实时响应货舱的检测任务.

Sobel 算子是一种经典的边缘检测方法, 主要用于对图像的水平 and 垂直边缘提取^[18]. Sobel 的原始模

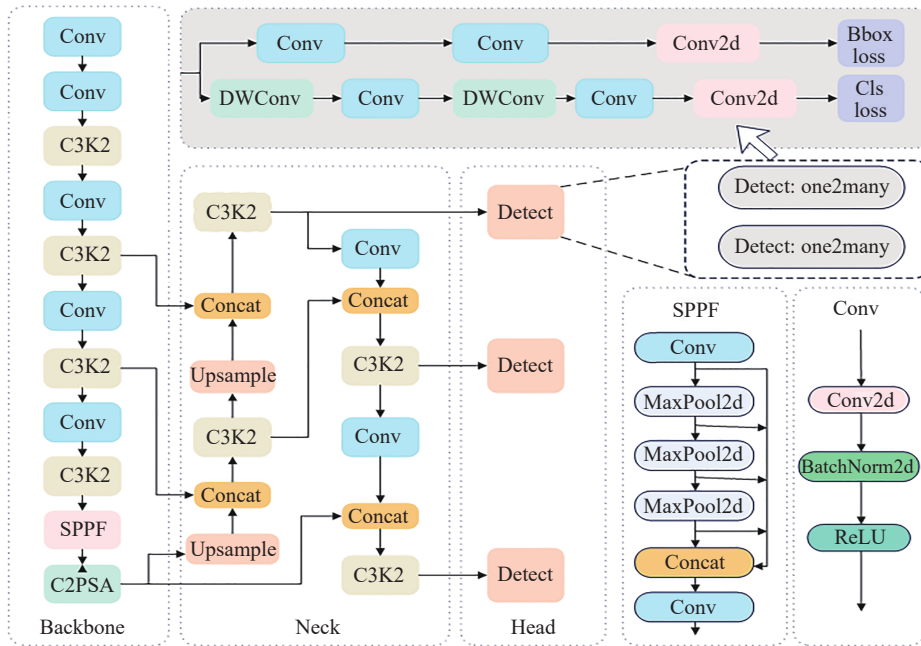


图 1 YOLO11 模型结构

Fig. 1 Structure of YOLO11 model

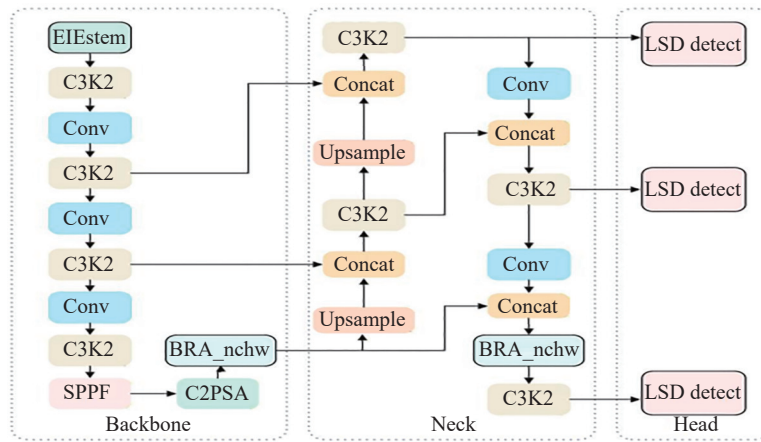


图 2 LEBR-YOLO 模型结构

Fig. 2 Structure of LEBR-YOLO model

型是两个 3×3 的卷积核, 这两个卷积核的设计主要是基于中心差分原理, 通过卷积核在图像上的滑动, 来计算像素邻域的梯度变化. 在 EIEstem 模块中, Sobel 算子作为特征提取器, 通过提取图像的边缘特征来区分物体. 水平梯度 (G_x) 和垂直梯度 (G_y) 的计算公式分别为

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad (1)$$

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}. \quad (2)$$

边缘强度 (Gradient magnitude, σ) 表示图像边缘的显著性, 可以通过水平和垂直梯度平方和的平方根来计算, 其表达式为

$$\sigma = \sqrt{G_x^2 + G_y^2}. \quad (3)$$

边缘方向 (Gradient direction, θ) 可以通过梯度的反正切来计算, 有

$$\theta = \arctan \left(\frac{G_y}{G_x} \right). \quad (4)$$

Sobel 卷积仅增加两个 3×3 卷积核, 与普通卷积相比较, 额外增加的 FLOPs (N_{FLOPs}) 计算公式为

$$N_{\text{FLOPs}} = 2HWC. \quad (5)$$

其中, H 和 W 为特征图尺寸, C 为通道数. 在 640×640 输入、3 通道的情况下, Sobel 卷积增加的 FLOPs 约占总 FLOPs 的不到 1%, 影响可以忽略不计.

EIEstem 作为模型中 Backbone 的初始特征提取模块, 输入图像首先经过步幅为 2 的卷积层进行初步特征提取, 减少计算量的同时扩大感受野, 接着分为两个分支: Sobel conv 分支提取边缘特征, Max pool 分支进行特征下采样, 提取局部极大值特征. Pooling 分支与 Sobel 分支互补, 使模型在保留边缘信息的同时也能捕捉整体的空间特征. 随后通过将两个分支的输出拼接, 形成融合的多尺度特征表示, 再经过卷积层进一步提炼特征. 其网络结构如图 3 所示, 这种多分支结构结合了卷积、边缘检测和池化操作, 实现了空间特征与边缘特征的互补融合, 既保持了图像的整体空间分布, 又增强了边缘细节的表达能力, 能够提取到更加丰富的特征表示, 同时保持高效的计算性能.

2.2 双层路由注意力机制

为提高在货舱站内模型检测的精度, 引入基于通道维度的双层路由注意力^[19], 优点是可以在通道维度上引导信息传播. 上层路由器通过全局自注意力机制对所有图像块进行交互, 并生成全局图像表示. 假设输入图像 X 被划分为 N 个图像块 $X_i (i \in \{1, 2, \dots, N\})$, 通过 Transformer 编码器提取查询 Q_i , 键 K_i 和值 V_i , 层路由器通过计算所有图像块之间的相似度, 生成全

局图像表示. 具体的注意力计算公式为

$$(Q, K, V)_{\text{attention, global}} = \left(\frac{QK^T}{\sqrt{d_k}} \right)_{\text{softmax}} V. \quad (6)$$

其中, attention 为注意力机制, Q 和 K 分别为查询矩阵和键矩阵; d_k 为键的维度; softmax 为归一化函数, 用于求取每个图像块间的权重. 这些权重用来对值矩阵 V 进行加权平均, 最终生成全局图像表示, 其表达式为

$$R_{\text{global}} = \sum_{i \in X} (Q_i, K_i, V_i)_{\text{attention, global}}. \quad (7)$$

其中, Q_i, K_i, V_i 分别为图像 i 的查询、键和值. 下层路由器则使用局部自注意力机制, 对每个图像块与其邻近的图像块进行交互, 生成相应的局部图像表示^[20]. 假设每个图块 i 通过的邻近区域为 N_i , 每个图像块 i 通过计算与邻近区域 N_i 中所有图像块的相似度, 来获取局部特征, 其表达式为

$$(Q_i, K_j, V_j)_{\text{attention, local}} = \left(\frac{Q_i K_j^T}{\sqrt{d_k}} \right)_{\text{softmax}} V_j, \quad \forall j \in N_i. \quad (8)$$

其中, Q_i, K_j, V_j 分别为图像 i 和邻近图像块 j 的查询、键和值. 每个图像块与其邻近的图像块进行交互, 生成局部图像表示, 其表达式为

$$R_{\text{local}} = \sum_{i \in X} \sum_{j \in N_i} (Q_i, K_j, V_j)_{\text{attention, local}}. \quad (9)$$

最后, 通过将全局图像 (R_{global}) 和局部图像 (R_{local}) 加权融合, 形成统一的图像表示 R , 其表达式为

$$R = \alpha R_{\text{global}} + (1 - \alpha) R_{\text{local}}. \quad (10)$$

其中, α 表示全局信息的权重, 通过调节 α , 可以灵活地平衡全局信息和局部信息的作用. 通过这种双层路由注意力机制, 同时捕捉全局和局部的特征信息, 提高了模型在目标检测任务中的性能.

2.3 轻量级共享可变形卷积检测模块

实际的空间站货舱环境存在人物遮挡、识别目标体积小等问题, 这些因素进一步增加了检测难度. 为了解决这一问题, 提出了轻量级共享可变形卷积检测模块. 通过使用共享卷积结构, 使该架构保持高效和轻量化, 同时通过对 Deformable Convolution V4 (DCNv4) 和 Group Normalization (GN) 的结合, 在遮挡场景下可动态调整采样点, 从而捕捉被部分遮挡的物体特征, 提高了模型对目标形变、遮挡和不同尺

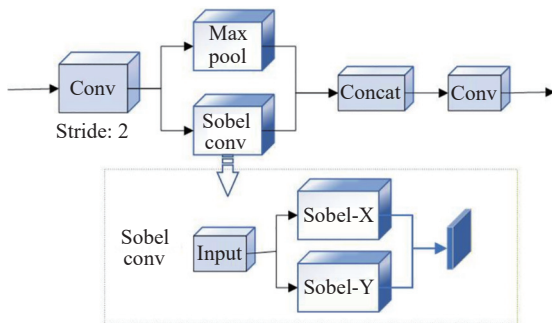


图 3 EIEstem 网络结构

Fig. 3 EIEstem network structure

度的适应能力. LSD 检测模块 (LSD Detect) 结构如图 4 所示.

改进的轻量级共享可变形检测模块使用了共享卷积结构^[21]. 每个输入特征图 (P3, P4, P5) 首先通过 1×1 的卷积层进行通道数调整, 以标准化不同尺度的特征图深度. 之后, 特征图接着通过共享的 3×3 可变形卷积层 (DCNv4) 进行空间特征提取, 由于 3×3 可变卷积采样点会根据输入特征图的内容进行自适应调整, 这使得模型能够在面对形变、遮挡或复杂背景时, 能更好地适应目标. 随后通过另一个 3×3 卷积层进行进一步的融合, 从而为分类和回归分析做准备. 分类分支 (Con_Cls) 用于预测每个位置的类别概率, 回归分支 (Conv_Reg) 则预测调整边界框的参数, 这些参数将通过分布式焦点损失 (DFL) 优化边界框的精确度, 通过缩放层调整预测框至适当的尺寸. 然后, 所有特征均通过解码算法转换为最终的边界框和类别预测. 其中缩放层的输出特征 X' 为

$$X' = X \cdot S. \quad (11)$$

其中, X 表示输入特征, S 为可学习的比例因子.

特征共享检测头的关键是通过利用权重共享卷积和在不同检测层之间共享特征提取器, 来显著减少检测头参数的数量. 对于模型轻量化设计造成的准确率损失, 通过广泛研究发现, GroupNorm (GN) 方法已被证明可以提高检测头的定位和分类性能^[22]. 因此, 在卷积中引入 GN 来调整通道数, 并在特征提取器中的卷积中引入 GN 来补偿检测精度. GN 的流程如下: 假设输入特征 X 的大小是 $N \cdot C \cdot H \cdot W$, 首先, GN 将通道数 C 划分为 G , 每个 G 包含 $C_G = C/G$ 通道, 每组的均值 μ_g 和方差 σ_g^2 分别为

$$\mu_g = \frac{1}{N \cdot C_G \cdot H \cdot W} \sum_{n=1}^N \sum_{c \in G_g} \sum_{h=1}^H \sum_{w=1}^W X(n, c, h, w), \quad (12)$$

$$\sigma_g^2 = \frac{1}{N \cdot C_G \cdot H \cdot W} \sum_{n=1}^N \sum_{c \in G_g} \sum_{h=1}^H \sum_{w=1}^W [X(n, c, h, w) - \mu_g]^2. \quad (13)$$

最后, 对每个元素进行归一化, 即

$$\hat{X}(n, c, h, w) = \frac{X(n, c, h, w) - \mu_g}{\sqrt{\sigma_g^2 + \varepsilon}}. \quad (14)$$

在批量归一化中, 为了让网络自主学习最优的尺度与偏移, 会引入如下仿射变换公式:

$$Y(n, c, h, w) = \gamma \hat{X}(n, c, h, w) + \beta. \quad (15)$$

其中, $\hat{X}(n, c, h, w)$ 是经过归一化后的特征图, $Y(n, c, h, w)$ 是仿射变换后的输出, 作为下一层的输入. γ 和 β 是大小为 $(1, C, 1, 1)$ 的可学习缩放参数和平移参数. 使用共享的 3×3 卷积层不仅可以减少参数数量, 降低计算成本, 还保持了对所有尺度特征的处理一致性.

可变形卷积是对传统卷积操作的一种扩展, 其通过引入动态的偏移量来调整卷积核在输入特征图上的采样位置. 在 Deformable Convolution V4 (DCNv4)^[23] 中, 卷积核的采样位置不固定, 通过学习得到的偏移量 (offsets) 和尺度因子 (scales) 来进行调整, 使卷积核可以根据输入数据的特征动态调整, 能更好地处理形变和复杂的图像变化. $y(i, j)$ 是输出特征图的某个像素值, 表示卷积操作在位置 (i, j) 处的结果, 其表达式为

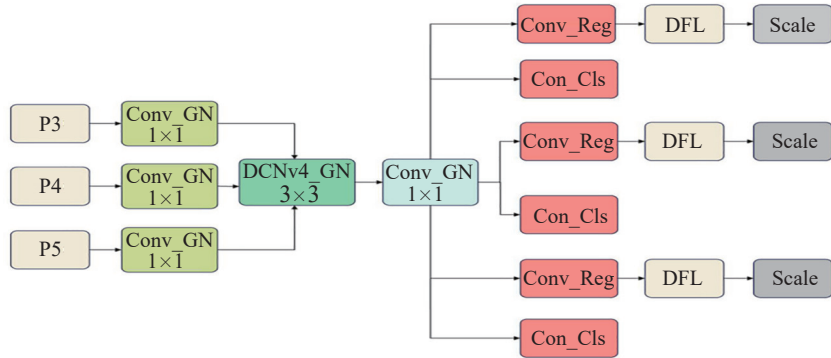


图 4 轻量级共享可变形检测模块

Fig. 4 Lightweight shared deformable detection module

$$y(i, j) = \sum_{m, n} K(m, n) \cdot X[i + m + \Delta m(i, j), j + n + \Delta n(i, j)]. \quad (16)$$

其中, $K(m, n)$ 表示 $k \times k$ 卷积核中第 (m, n) 个权重; $X(i + m, j + n)$ 为输入特征图 X 在位置 $(i + m, j + n)$ 处的像素值; $\Delta m(i, j)$ 和 $\Delta n(i, j)$ 分别为水平方向和垂直方向的偏移量, 表示卷积核采样点的动态调整. 在 Deformable Convolution V4 (DCNv4) 中, 进一步加入了尺度调整, $y(i, j)$ 可表示为

$$y(i, j) = \sum_{m, n} K(m, n) \cdot X\{i + S(i, j) \cdot [m + \Delta m(i, j)], j + S(i, j) \cdot [n + \Delta n(i, j)]\}. \quad (17)$$

其中, $S(i, j)$ 是一个动态学习因子, 表示每个卷积核位置的缩放比例. 当 $S(i, j) > 1$ 时, 卷积核的感受野增大; 当 $S(i, j) < 1$ 时, 卷积核的感受野缩小. 这种结构使得 DCNv4 在处理复杂环境、变形、遮挡问题时能够更加准确.

2.4 迁移学习

迁移学习利用将源领域获得的知识应用到另一个目标领域中, 来解决目标领域的数据不足或训练复杂度高的问题, 从而提高模型在目标任务上的表现^[24,25]. 在空间站这样的特殊环境下, 获取大量标注好的果蔬数据集成本比较高. 若直接从零开始训练一个深度学习模型可能导致过拟合, 并且训练时间长. 迁移学习通过利用现有的水果蔬菜数据上训练好的预训练模型, 将模型学到的通用特征迁移到空间站的检测任务上, 可以有效地减轻数据不足的影响, 且降低成本.

使用在网络上收集的蔬菜水果数据集作为预训练数据进行迁移学习. 先在整理的数据集上训练改进后的算法, 得到一个初始权重文件, 并将其作为预训练权重, 然后在自制的货舱数据集上再次训练. 训练果蔬数据集时, 将训练分为两个阶段, 冷冻阶段和解冻阶段. 在使用改进的模型训练之前, 需要对网络模型的参数进行设置, 实验设置训练轮数为 200, 批处理 (batch size) 为 16, 在前 40 轮训练中冻结网络主. 冷冻阶段: 冻结 Backbone 40 epochs, batch size 为 16, 初始学习率为 0.01, 优化器 Adam ($\beta_1=0.9$, $\beta_2=0.999$, weight decay= 10^{-4}). 解冻阶段: 解冻 Backbone 后继续训练 160 epochs, 学习率衰减至

0.001, 其他参数保持不变. 迁移学习流程如图 5 所示.

3 实验结果

3.1 实验数据

使用的数据集由两部分组成, 一部分是利用模拟空间站货舱环境下自制的水果蔬菜数据. 该数据集由海康威视摄像头录制的轻舟一号货舱视频中捕捉到的水果蔬菜图像组成, 这些视频被分割成 2325 张图像, 模拟了航天员在货舱中的拿取过程, 这些图像包括空间站可能出现的不同水果蔬菜类型共 18 种, 包括苹果、西红柿、卷心菜、胡萝卜、土豆、玉米等果蔬, 显示了不同位置、不同遮挡程度的果蔬图像, 如图 6 所示. 图像以 png 格式保存, 分辨率为 2560×1440 pixel, 使用 labelImg 进行数据注释. 完成注释后, 将数据集以 8:1:1 的比例划分为训练集、验证集、测试集.

另一部分数据集是通过从网络上收集的各种果蔬图片, 主要用于迁移学习的训练. 数据集共 15210 张, 包括苹果、香蕉、土豆、黄瓜等 20 种果蔬类型. 图像以 png 格式保存, 分辨率为 2560×1440 pixel, 如图 7 所示. 该数据集的目的是为模型提供一个基础的、通用的果蔬识别知识库, 从而加速在空间站货舱内果蔬识别的训练过程.

3.2 实验配置

实验环境配置如表 1 所示. 模型训练参数设置如

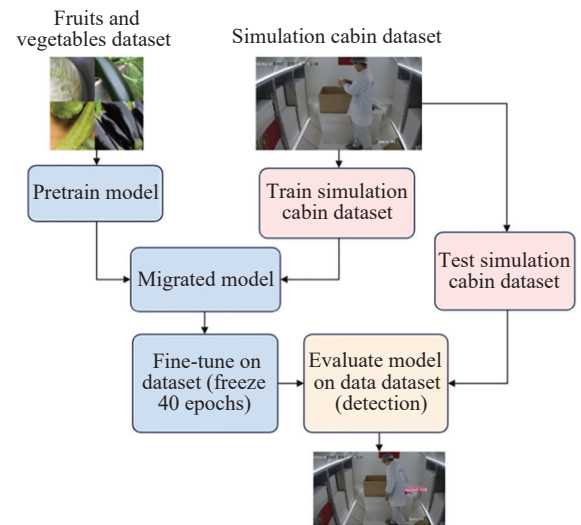


图 5 迁移学习流程

Fig. 5 Transfer learning process



图 6 空间站货舱数据集
Fig. 6 Space station cargo bay dataset

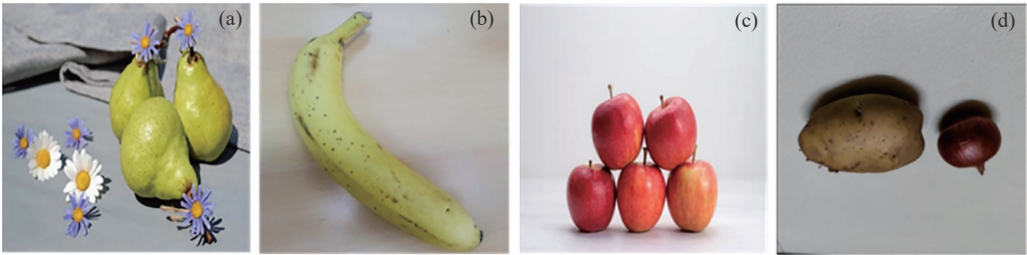


图 7 果蔬数据集
Fig. 7 Fruit and vegetable dataset

表 1 实验环境配置
Table 1 Configuration of experimental environment

Experimental environment	Configuration
Operating system	ubuntu22.04
Graphics Processing Unit (GPU)	RTX 3090 (24 GB)
Central Processing Unit (CPU)	Intel (R) Xeon (R) Gold 6330
Deep learning framework	PyTorch2.1
Computing platform	Cuda12.1
Random Access Memory (RAM)	60 GB
Interpreter	Python3.10

下. 训练周期为 200 个 epochs, batch size 设置为 16, 使用 4 个进程进行训练, 输入图像大小为 640×640 pixel. 模型使用 Adam 优化器进行参数优化, 初始学习率设置为 0.01.

3.3 评估指标

以精确率 P 、召回率 R 、平均精度均值 (m_{AP}) 作为关键指标来评估目标检测算法的性能. 精确率是模

型预测为正例的样本中实际为正例的比例, 精确率越高, 模型被误判的可能性越小, 其表达式为

$$P = \frac{T_{TP}}{T_{TP} + F_{TP}}. \tag{18}$$

其中, T_{TP} 和 F_{TP} 分别表示真正例和假正例的数量. 召回率是模型正确预测为正例的样本中实际为正例的比例, 其表达式为

$$R = \frac{T_{TP}}{T_{TP} + F_{FN}}. \quad (19)$$

其中, F_{FN} 表示假反例的数量.

平均精度 (P_{AP}) 是目标检测和分类任务中的关键指标之一, 其定义为 P - R 曲线下的面积, 通常使用数值积分来计算, 其表达式为

$$P_{AP} = \int_0^1 P(R) dR. \quad (20)$$

最后对所有类别的 AP 进行平均, 得到最终的 m_{AP} . 对于 N 个类别的检测目标, 各类平均精度均值 P_{mAP} 可表示为

$$P_{mAP} = \frac{\sum_{n=1}^N P_{AP_n}}{N}. \quad (21)$$

$m_{AP}@0.5$ 表示在 IoU 阈值取 0.5 时得到的平均精度均值, $m_{AP}@(0.5 \sim 0.95)$ 表示 IoU 阈值在 0.5~0.95 的范围内取得精度的平均值. 较高平均精度表示该模型可以在不同的置信度阈值下保持较高的检测准确率和召回率^[26].

3.4 性能分析

图 8 为 Precision-Recall 曲线. 从图 8 可以看出模型的检测性能比较出色, 平均精度均值达到了 93.9%. 各类别的精度-召回曲线普遍表现平滑, 表明模型的误检率较低, 同时能保证较高的召回率. 这说明在实际应用中, 模型能够较好地识别目标, 避免漏检的同时尽可能减少假阳性. 从整体来看, 模型在目

标检测任务中具有较强的竞争力和鲁棒性, 能够满足货舱智能化检测对精度和效率的要求, 是一个可靠的解决方案.

为了直观评估在实际场景应用中的模型性能, 对比改进前后的模型在测试集上的检测效果, 结果如图 9 所示. 在保证每次迭代的实验条件一致的情况下, 对比图 9 的检测结果可以发现, 改进后的模型在货舱环境中对物体的特征提取能力更强, 在小目标、有遮挡的情况下, 识别更准确, 改进了原模型误检的情况, 说明 LEBR-YOLO 模型在货舱环境下具备更精准的检测性能.

3.5 遮挡分组实验

为了进一步验证所提 LEBR-YOLO 模型在复杂空间站货舱环境下的鲁棒性, 基于测试集构造了遮挡分组实验. 以目标框之间的交并比 (IoU, Intersection over Union) 作为遮挡度量指标, 并根据目标被其他物体遮挡的比例, 将测试集中的样本划分为 4 个等级: 无遮挡 (0~10%)、轻度遮挡 (10%~30%)、中度遮挡 (30%~60%)、严重遮挡 (>60%). 对于某一目标 T_i , 其被其他目标遮挡的比例 R 为

$$R(A, B) = \frac{(A \cap B)_{\text{area}}}{(A \cup B)_{\text{area}}}. \quad (22)$$

其中, A 和 B 分别为两个目标的边界框; $(A \cap B)_{\text{area}}$ 表示两框交集面积; $(A \cup B)_{\text{area}}$ 表示两框并集面积; N 为图像中目标总数. 为了量化每个目标 T_i 被其他

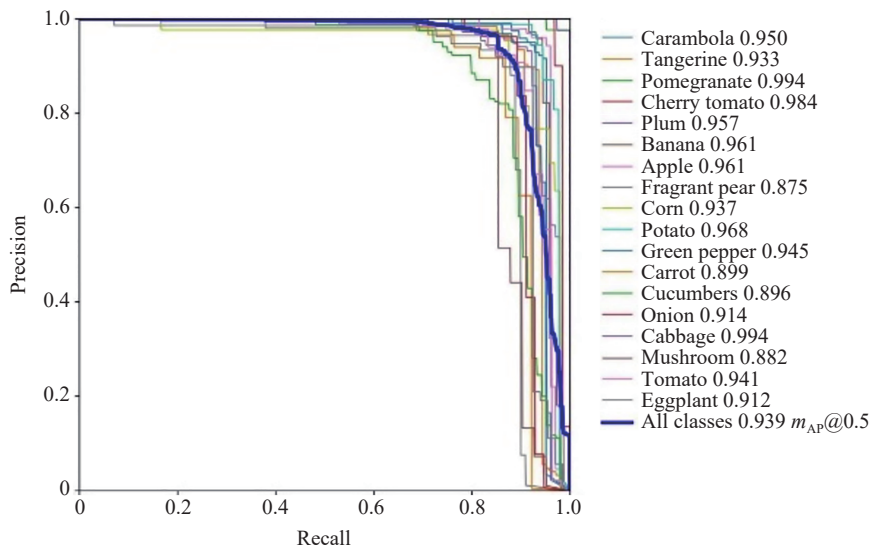


图 8 Precision-recall 曲线

Fig. 8 Precision-recall graph

目标遮挡或重叠的程度, 定义其重叠率 O_i 为

$$O_i = \frac{\sum_{j=1, j \neq i}^N (T_i \cap T_j)_{\text{area}}}{(T_i)_{\text{area}}}.$$

(23)

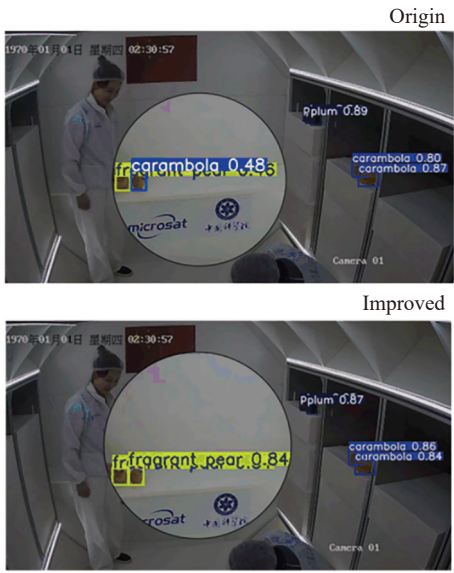


图 9 检测结果
Fig. 9 Test results

$O_i \in [0, 1]$, 数值越大表示遮挡越严重.

在划分数据后, 分别评估了基线 YOLO11 与本文改进的 LEBR-YOLO 模型在不同遮挡条件下的性能, 结果如表 2 所示. 随着遮挡程度加深, 所有模型的检测精度均有所下降, 这是空间站货舱复杂环境下不可避免的现象. 本文提出的 LEBR-YOLO 在各遮挡级别下均优于基线 YOLO11, 尤其在中度和严重遮挡条件下优势最为明显, $m_{AP}@0.5}$ 分别提升 2.0% 和 2.3%. 遮挡分组实验进一步验证了本文改进方法在空间站货舱这一特殊应用场景下的鲁棒性与实用价值.

3.6 消融实验

改进后的算法包括原模型 YOLO11 的模块、EIEstem 模块、BRA_nchw 模块和 LSDDetect 模块. 通过消融实验分析各模块的贡献, 结果如表 3 所示. 加入 EIEstem 模块使 $m_{AP}@0.5}$ 提升了 2%, 说明该模块提高了特征提取或特征融合的能力, 使模型对目标的识别更加精确; 加入 LSDDetect 检测模块, 准确率提高了 1.9%, $m_{AP}@0.5}$ 提高 1.2%, 说明该检测头在优化目标检测过程中的定位精度、边界框回归等方面做出了贡献, 帮助模型更好地识别和定位目标; 加入 BRA_nchw 注意力机制, 使 $m_{AP}@0.5}$ 提高 2%, 且召

表 2 遮挡对比试验
Table 2 Occultation comparison experiment

遮挡等级	YOLO11		LEBR-YOLO		$m_{AP}@0.5}$ 提升 幅度/(%)
	$m_{AP}@0.5/(%)$	$m_{AP}@0.5:0.95/(%)$	$m_{AP}@0.5/(%)$	$m_{AP}@ (0.5 \sim 0.95)/(%)$	
无遮挡	95.2	66.7	96.5	68.4	1.3
轻度遮挡	92.8	62.9	94.1	64.8	1.3
中度遮挡	89.8	58.9	91.8	61.0	2.0
严重遮挡	85.9	54.0	88.2	56.3	2.3

表 3 消融实验
Table 3 Ablation experiment

YOLO11 n	EIEstem	LSDDetect	BRA_nchw	$P/(%)$	$R/(%)$	$m_{AP}@0.5/(%)$	$m_{AP}@ (0.5 \sim 0.95)/(%)$
✓	—	—	—	91.2	85.6	91.2	61.1
✓	✓	—	—	92.8	88.1	93.2	61.5
✓	—	✓	—	93.1	87.2	92.4	60.8
✓	—	—	✓	93.2	88.4	93.2	62.2
✓	✓	✓	—	93.7	86.5	92.0	60.2
✓	✓	—	✓	93.3	88.0	93.1	61.2
✓	—	✓	✓	93.5	87.9	92.5	62.1
✓	✓	✓	✓	94.3	88.5	93.3	62.3

回率提高了 2.8%, 说明注意力机制可以加强特定区域或特征的关注度, 帮助模型更好地捕捉重要信息, 减少错过目标的情况. 加入两个模块的情况下, 模型的各项指标均有提升. 最后, 将三个模块均加入到原模型中后, 实验结果表明, 准确率提升了 3.1%, 召回率提升了 2.9%, $m_{AP}@0.5$ 提升了 2.1%, $m_{AP}@(0.5\sim 0.95)$ 提升了 1.2%. 这充分验证了各模块对模型整体性能提升具有必要性, 改进后的算法可以显著提高货舱内部果蔬检测的准确率, 且有效降低了误检与漏检.

3.7 注意力机制比较

为了显示双层路由注意力机制性能的优越性, 将不同的注意力机制进行对比试验. 从表 4 可知, 双层路由注意力机制 (BRA_nchw) 在各项指标上均表现较好, 尤其是在精度 (93.2%) 和召回率 (88.4%) 上具有明显优势, 表明该模型能够在保持高精度的同时有效捕获更多的真实目标. 较高的 $m_{AP}@(0.5\sim 0.95)$ (62.2%) 也意味着其能更好地处理更高难度的任务, 适合对检测精度要求较高的航天领域, 展现了其强大的适应性与鲁棒性.

3.8 迁移学习对模型性能的影响

为了能够充分展现迁移学习的优势, 分别做了两组实验数据的对比分析. 其中 A 表示 YOLO11; B 表示 YOLO11 结合迁移学习; C 表示改进的 YOLO11; D 表示改进的 YOLO11 结合迁移学习. 由表 5 可以

看出, 使用迁移学习使模型的准确率和召回率均获得较大幅度的提升, 并且改进后的模型在使用迁移学习的情况下, $m_{AP}@(0.5\sim 0.95)$ 又提升了 1.1%. 由此可见, 迁移学习对算法性能的提升有明显的积极作用.

3.9 模型对比实验

为了表明算法的先进性和有效性, 将 LEBR-YOLO 算法与现在主流的各个模型进行比较. 从表 6 的结果可以看出, LEBR-YOLO 算法在 $m_{AP}@0.5$ 指标上相比 YOLOv5, YOLOv8, YOLOv10 n, YOLO11 n 分别提高了 3.1%, 2.9%, 3.7%, 2.7%, 表现出更高的精度. 与 FasterRCNN 模型相比, LEBR-YOLO 在 mAP、召回率和参数大小方面均有显著提升. 尽管 LEBR-YOLO 模型在 Parameters, GFLOPs, FPS 这三项指标上并非最优, 但其综合性能出色, 在货舱环境下, 这种高效的检测能力能够保证高精度的同时, 减少计算资源的消耗, 满足实时性和资源限制的需求.

综上所述, LEBR-YOLO 算法是当前主流模型中在精度、速度和资源消耗之间实现最佳平衡的检测方案, 满足在轨运行的各项指标要求, 是一种专为航天环境优化的高效智能目标检测技术.

4 结论

针对空间站货舱内部复杂环境下的果蔬检测

表 4 注意力机制比较
Table 4 Comparison of attention mechanisms

Models	$P/(\%)$	$R/(\%)$	$m_{AP}@0.5/(\%)$	$m_{AP}@(0.5\sim 0.95)/(\%)$
CPCA	90.9	86.1	92.3	61.2
SimAm	92.3	85.9	91.9	61.2
Dattention	92.8	86.8	91.5	61.6
Local Window Attention	92.3	85.8	91.4	61.3
BRA_nchw	93.2	88.4	93.2	62.2

表 5 迁移学习对模型性能的影响
Table 5 Impact of transfer learning on model performance

Models	$P/(\%)$	$R/(\%)$	$m_{AP}@0.5/(\%)$	$m_{AP}@(0.5\sim 0.95)/(\%)$
A	91.2	85.6	91.2	61.1
B	93.9	87.4	92.7	62.0
C	94.3	88.5	93.3	62.3
D	95.3	88.6	93.9	63.4

表 6 模型对比
Table 6 Comparison of the models

Model	P/(%)	R/(%)	$m_{AP}@0.5/(%)$	$m_{AP}@ (0.5\sim0.95)/(%)$	Parameters ($\times 10^6$)	GFLOPs	FPS
YOLOv5	91.0	86.6	90.8	60.7	2.2	5.8	242.31
YOLOv8	91.2	85.5	91.0	60.0	3.0	8.1	153.81
YOLOv10 n	90.9	83.7	90.2	60.9	2.3	6.5	379.53
FasterRCNN	87.8	78.0	83.1	55.6	41.0	198	13.65
YOLO11 n	91.2	85.6	91.2	61.1	2.6	6.3	288.50
LEBR-YOLO	95.3	88.6	93.9	63.4	3.1	7.1	270.24

需求,提出一种面向空间站货舱的果蔬精准检测方法——LEBR-YOLO.通过结合传统卷积、边缘检测和池化操作改进初始特征提取模块,提高了边缘特征提取能力.引入双层路由注意力机制,该机制通过动态调整注意力区域,有效聚焦关键区域,提升了细粒度特征的学习能力.改进了检测模块,采用共享卷积结构并引入可变形卷积,提高复杂特征提取能力的同时保持了模型的轻量.同时又采用迁移学习弥补数据的稀缺性,从而优化了目标检测性能.这些改进使模型在货舱内部复杂的果蔬检测任务中,能够显著提升精度和鲁棒性,满足了轻舟一号货运飞船在轨运行需求,同时提高了空间站资源管理的自动化程度.未来可以结合计算机视觉和空间站机器人,实现自主决策,减少对地面控制的依赖,为航天生态闭环系统提供智能化支持.

利益冲突 所有作者均声明不存在利益冲突.

参考文献

[1] WANG Wei, GUO Pei, YIN Zhao. Development and outlook for space technology experiments on space stations[J]. *Journal of Space Science and Experiment*, 2024, **1**(1): 1-12 (王巍, 郭佩, 尹钊. 空间站航天技术试验发展与展望 [J]. *空间科学与试验学报*, 2024, **1**(1): 1-12)

[2] SHAN Xianying, ZHANG Lin, LI Zehui. Review of research progress in object detection driven by deep learning[J]. *Computer Engineering and Applications*, 2025, **61**(1): 24-41 (山显英, 张琳, 李泽慧. 深度学习驱动下的目标检测研究进展综述 [J]. *计算机工程与应用*, 2025, **61**(1): 24-41)

[3] PANT P, RAJAWAT A S, GOYAL S B, *et al*. AI based technologies for international space station and space data[C]//2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART).

Moradabad, India: IEEE, 2022: 19-25

[4] MING L, HONG L X, WEI W. Design and application of aerospace intelligent housing system of manned space station base on space-integrated-ground network[C]//2024 7th World Conference on Computing and Communication Technologies (WCCCT). Chengdu, China: IEEE, 2024: 255-259

[5] REDMON J, DIVVALA S, GIRSHICK R, *et al*. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016: 779-788

[6] GAI R L, LIU Y, XU G H. TL-YOLOv8: a blueberry fruit detection algorithm based on improved YOLOv8 and transfer learning[J]. *IEEE Access*, 2024, **12**: 86378-86390

[7] JIANG Yefan, JIANG Zhaohui, ZHU Rui, *et al*. Strawberry ripeness in a complex environment based on PCIA-YOLO Lightweight detection method[J/OL]. *Journal of Nanjing Agricultural University*, 2024: 1-15. (2024-10-08) [2025-01-08].<https://link.cnki.net/urlid/32.1148.S.20240930.1738.008> (姜业帆, 江朝晖, 朱瑞, 等. 基于 PCIA-YOLO 的复杂环境下草莓成熟度轻量化检测方法 [J/OL]. *南京农业大学学报*, 2024: 1-15. (2024-10-08)[2025-01-08]. <https://link.cnki.net/urlid/32.1148.S.20240930.1738.008>)

[8] LIU Z G, GAO Y, DU Q Q. YOLO-Class: detection and classification of aircraft targets in satellite remote sensing images based on YOLO-extract[J]. *IEEE Access*, 2023, **11**: 109179-109188

[9] LIU X J, WANG Y P, YU D X, *et al*. YOLOv8-FDD: a real-time vehicle detection method based on improved YOLOv8[J]. *IEEE Access*, 2024, **12**: 136280-136296

[10] SUI T T, WANG J W. DMPDD-Net: an effective defect detection method for aluminum profiles surface defect[J]. *IEEE Transactions on Instrumentation and Measurement*, 2025, **74**: 3500313

[11] ZHENG Huawei, WANG Fei, GAO Jianbang. ES-YOLO: a more accurate object detection method[J]. *Opto-Electronic Engineering*, 2024, **51**(11): 240212 (郑华伟, 王飞, 高建邦. DES-YOLO: 一种更精确的目标检测方法 [J]. *光电工程*, 2024, **51**(11): 240212)

[12] SHAO Yanhua, ZHANG Duo, CHU Hongyu, *et al*. A re-

- view of YOLO object detection based on deep learning[J]. *Journal of Electronics & Information Technology*, 2022, **44**(10): 3697-3708 (邵延华, 张铎, 楚红雨, 等. 基于深度学习的 YOLO 目标检测综述 [J]. *电子与信息学报*, 2022, **44**(10): 3697-3708)
- [13] HAO J B. Deep learning-based medical image analysis with explainable transfer learning[C]//2023 International Conference on Computer Engineering and Distance Learning (CEDL). Shanghai, China: IEEE, 2023: 106-109
- [14] HUANG B N, CHEN X, SUN Y P, *et al.* Multi-agent cooperative strategy learning method based on transfer learning[C]//2022 13th Asian Control Conference (ASCC). Jeju, Republic of Korea: IEEE, 2022: 1095-1100
- [15] YAN J H, WANG X F, ALI A. Deep transfer learning based multi-source heterogeneous data fusion with application to cross-scenario tool wear monitoring[C]//2021 7th International Conference on Mechanical Engineering and Automation Science (ICMEAS). Seoul, Republic of Korea: IEEE, 2021: 96-101
- [16] WANG A, CHEN H, LIU L H, *et al.* YOLOv10: real-time end-to-end object detection[EB/OL]. (2024-05-23)[2024-10-30]. <https://arxiv.org/abs/2405.14458>
- [17] SAPKOTA R, MENG Z C, CHURUVIJA M, *et al.* Comprehensive performance evaluation of YOLOv12, YOLO11, YOLOv10, YOLOv9 and YOLOv8 on detecting and counting fruitlet in complex orchard environments[EB/OL]. (2024-07-01)[2024-10-17]. <https://arxiv.org/abs/2407.12040>
- [18] WU T, WANG L W, ZHU J C. Image edge detection based on Sobel with morphology[C]//2021 IEEE 5th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). Xi'an, China: IEEE, 2021: 1216-1220
- [19] ZHU L, WANG X J, KE Z H, *et al.* BiFormer: vision transformer with bi-level routing attention[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE, 2023: 10323-10333
- [20] XIAO Lei, HU Peng, MA Junjie. Self-supervised monocular depth estimation model based on global information correlation under influence of local attention[J]. *Laser & Optoelectronics Progress*, 2025, **62**(8): 0815010 (肖磊, 胡鹏, 马俊杰. 局部注意力作用下基于全局信息关联的自监督单目深度估计模型 [J]. *激光与光电子学进展*, 2025, **62**(8): 0815010)
- [21] KHAN G, PIMBBLET K, WERTHEIM K, *et al.* A single shot multi-head gender, age, and landmarks detection using shared convolution features[C]//2024 29th International Conference on Automation and Computing (ICAC). Sunderland, United Kingdom: IEEE, 2024: 1-6
- [22] WU Y X, HE K M. Group normalization[C]//Proceedings of 15th European Conference on Computer Vision – EC-CV 2018. Munich: Springer, 2018: 3-19
- [23] XIONG Y W, LI Z Q, CHEN Y T, *et al.* Efficient deformable ConvNets: rethinking dynamic and sparse operator for vision applications[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2024: 5652-5661
- [24] LI Xinyao, LI Jingjing, ZHU Lei, *et al.* Efficient transfer learning of large models with limited resources: a survey[J]. *Chinese Journal of Computers*, 2024, **47**(11): 2491-2521 (李鑫尧, 李晶晶, 朱磊, 等. 资源受限的大模型高效迁移学习算法研究综述 [J]. *计算机学报*, 2024, **47**(11): 2491-2521)
- [25] ZHU Z D, LIN K X, JAIN A K, *et al.* Transfer learning in deep reinforcement learning: a survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, **45**(11): 13344-13362
- [26] HU Haining, HUANG Leiyang, YANG Honggang, *et al.* Improved YOLOv8n lightweight honeycomb ceramic defect-detection algorithm[J]. *Laser & Optoelectronics Progress*, 2024, **61**(22): 2212006 (胡海宁, 黄雷阳, 杨洪刚, 等. 改进 YOLOv8n 的轻量化蜂窝陶瓷缺陷检测算法 [J]. *激光与光电子学进展*, 2024, **61**(22): 2212006)

作者简介



李文琴 女, 1998 年出生, 上海电机学院硕士研究生, 主要从事深度学习、低成本货运飞船智能检测系统的研究。
E-mail: 15858329317@163.com



常亮 (通信作者) 男, 1980 年出生, 中国科学院微小卫星创新研究院研究员、博士生导师, 主要研究方向为人工智能、微纳卫星系统等领域。现任中国科学院微小卫星创新研究院轻舟货运飞船总设计师, 曾参与北斗导航卫星系统建设, 并担任“悟空”号暗物质探测卫星总体主任设计师。
E-mail: 13774362558@163.com