doi:10.6041/j.issn.1000-1298.2026.06.027

基于 SAB－YOLO 模型的生菜生长期识别方法

林开颜[1,2]　王先浪[1,2]　牛程远[1,2]　吴军辉[1,2]　陈　杰[1,2]　杨学军[1,2]
（1. 同济大学现代农业科学与工程研究院，上海 201804；2. 同济大学电子与信息工程学院，上海 201804）

摘要：针对作物生长期难以自动化识别、传统机器学习识别方法的精度有限、不同阶段尺度差异大导致识别准确度较低的问题，本文提出一种生长期识别模型 SAB－YOLO（Self attention based－YOLO）。将 YOLO v5 特征提取网络替换为基于自注意力机制的 Swin Transformer 网络，增强模型对全局特征的捕捉能力；并将 Neck 网络改进为跨层连接更为密集的 AFPN 结构，改善多尺度特征融合效果；提出了将卷积和自注意力机制结合的 CTF（Convolutional transformer fusion）模块，并应用在检测头位置以增强全局特征；最后将损失函数改为 Inner－SIoU。试验结果表明，改进模型在测试集上精确率达到 88.5%、mAP_0.5 为 92.1%，提升了生菜图像生长期识别精度。研究结果为作物生长期识别提供了新的技术方案，对精准农业发展具有实践价值。

关键词：生菜；生长期；SAB－YOLO 模型；识别方法
中图分类号：S126；S43　　**文献标识码**：A　　**文章编号**：1000-1298(2026)06-0290-10　　**OSID**：

Lettuce Growth Period Recognition Method Based on SAB－YOLO Model

LIN Kaiyan[1,2]　WANG Xianlang[1,2]　NIU Chengyuan[1,2]　WU Junhui[1,2]　CHEN Jie[1,2]　YANG Xuejun[1,2]
(1. *Modern Agricultural Science and Engineering Institute*, *Tongji University*, *Shanghai* 201804, *China*
2. *College of Electronic and Information Engineering*, *Tongji University*, *Shanghai* 201804, *China*)

Abstract: In container-based vertical agricultural production systems, supplementary lighting is a key technical approach for regulating crop growth, optimizing resource utilization, and improving production efficiency. However, most existing lighting control strategies rely on fixed time cycles or empirical parameters, lacking effective perception of crop growth stages, which limits their adaptability and precision under dynamic growth conditions. To address these challenges, a lettuce growth stage recognition model named self attention based－YOLO (SAB－YOLO) was proposed to realize accurate and automated identification of crop growth periods in complex visual environments. The proposed model was developed by introducing multiple structural improvements to the YOLO v5 framework. Firstly, the conventional convolutional backbone was replaced with a Swin Transformer network based on self-attention mechanisms, which enhanced the ability of the model to capture long-range dependencies and global semantic information. Secondly, an asymptotic feature pyramid network (AFPN) with denser cross-layer connections was adopted in the Neck to strengthen multi-scale feature fusion and improve robustness to large scale variations among different growth stages. Furthermore, a convolution transformer fusion (CTF) module that integrated convolutional operations with self-attention was designed and embedded into the detection head to further enhance global contextual representation. In addition, the Inner－SIoU loss function was employed to improve bounding box regression accuracy and accelerate model convergence. Experimental results on a mixed dataset collected from open-source platforms and a container-based plant factory showed that the proposed model achieved a precision of 88.5% and an mAP_0.5 of 92.1%, outperforming the baseline YOLO v5 model. Furthermore, an intelligent supplementary lighting system based on growth stage recognition was designed and validated, demonstrating the practical applicability of the proposed method in precision agriculture.

Key words: lettuce; growth period; SAB－YOLO model; recognition method

收稿日期：2025－08－18　修回日期：2025－10－11
基金项目：上海市科委科技创新行动计划项目(23N21900400)
作者简介：林开颜(1975—)，男，副教授，主要从事农业计算机视觉技术研究，E-mail：linkaiyan@tongji.edu.cn

0　引言

人工光植物工厂是一种高密度、高效利用的立体化、集约化现代农业生产方式，由计算机自动精准控制作物生长过程中的光照强度、温度、相对湿度、CO_2浓度等关键环境参数，以实现农作物周年连续生产，是现代农业发展的重要方向。植物工厂种植中，通过对作物生长发育阶段的自动识别，并按照作物生长需求精准调节补光系统光强、光质和光周期，不仅在节省能耗的同时保障作物品质，还能避免过度补光导致叶片边缘灼伤而影响品质[1－3]。

目前对作物生长情况识别方法主要有人工观察法和图像处理法 2 种。人工观察法存在主观性强、效率低、实时性差等局限，不利于垂直农业中对作物环境参数的实时调控和自动化生产。图像处理法通常基于作物生长过程中拍摄的二维图像数据，经处理后计算作物投影面积、株高、叶片尺寸、叶面积、群体覆盖指数等指标，通过回归分析等方法建立特征参数与作物生长信息之间的数学关系。DU 等[4]提出基于图像的多品种莴苣高通量检测与表型评价方法，为作物表型精准图像分析提供了参考，并验证了图像处理方法相比人工观察在效率和客观性上的优势。李修华等[5]将生菜图像从背景中分割出来，计算颜色、形状、纹理 3 大类共 39 个表型指标，进而构建生菜叶片面积指数与鲜质量预测模型，结果表明自动化计算方法可替代人工估测。SAKAMOTO 等[1]设计了一种作物物候监测系统，用于获取玉米和大豆发育阶段昼夜连续时间序列的可见光和近红外图像，进而计算得到作物叶面积指数、生物量等参数，并与 MODIS 卫星系统反演的作物生长情况进行了比较，两者相关性较好，进一步证明了自动监测可行性。刘林等[6]在提取生菜图像表型特征参数后，提出了一种基于表型的鲜质量估算方法，用于生菜生长期评估，避免了人工检测低效与破坏性。然而，这类基于人工设计特征的机器学习方法在实际应用中仍存在一定局限：①人工提取的颜色、形状、纹理等特征在跨品种或不同环境条件下的泛化能力较差，模型往往需要重新设计和调优[1－2]。②特征构建与参数选择过程通常较为复杂，限制了方法的可扩展性[4]。③部分方法需要对植物进行取样或标定，可能对植株本身造成一定程度的破坏，难以满足植物工厂中对作物实时、无损检测的需求[5]。随着深度学习技术发展，国内外许多学者开始将基于深度学习的图像检测算法应用在农作物生长期识别领域。陈杰杰[2]使用深度学习模型 VGGNet 实现了对生菜图像育苗期、生长期、发棵期和成熟期 4 个生长阶段的识别。STEPHI 等[7]针对小麦不同生长阶段尺度差异较大的问题，通过在卷积神经网络中引入多头注意力机制，实现了对小麦图像冠根期、分蘖期、营养生长期、拔节期、孕穗期、开花期和灌浆期 7 个生长阶段的有效识别。XU 等[8]将蘑菇划分为恢复期、原基形成期、现蕾期、伸长期和成熟期 5 个生长阶段，并使用基于深度卷积的 EfficientNet 网络准确识别了蘑菇图像生长阶段。扶兰兰等[3]利用基于自注意力机制的 Swin Transformer 模型，实现了对玉米图像苗期、拔节期、小喇叭口期和大喇叭口期 4 个生长阶段的有效识别，相比于 AlexNet、VGG16、GoogleNet 模型精度分别提高 6.9、2.7、2.0 个百分点。

近年来，已有学者尝试基于无人机与神经结构搜索实现多阶段作物的实时识别，证明了深度学习在提高检测速度与精度方面的潜力[9]；也有研究提出在垂直农场环境中利用 ROI 区域预测与分割方法对作物生长状态进行监测，以提升阶段识别效率[10]。然而，在高密度种植和复杂光照背景下，传统卷积神经网络方法容易受到背景干扰、遮挡及小目标特征缺失的影响，识别精度明显下降[11－12]。为解决这些问题，部分研究利用轻量化 CNN 或注意力机制对小麦、玉米等作物的生长期进行了识别，精度可达 97% 以上，但仍存在对环境适应性不足的局限[13]。在目标检测任务中，基于 CNN 的 YOLO[14]模型被广泛使用：王兴旺等[15]提出一种基于 YOLO v8－STSF 的水稻害虫智能识别方法，通过引入 Swin Transformer 模块增强骨干网络多尺度特征提取能力，并结合分布移位卷积和改进损失函数提升密集小目标检测精度；王泰华等[16]通过对 YOLO v5s 进行结构改进并引入注意力机制与优化后处理策略，有效提升复杂稻田环境下水稻害虫检测精度，体现了深度目标检测模型在农业视觉识别任务中的应用潜力。

YOLO 是一种 one－stage 目标检测算法，能够较快识别出图像中物体类别和边界框。YOLO v5 在 YOLO 算法基础上改进优化，其性能和精度都得到了极大提升[17]。然而，已有研究表明，YOLO 系列模型仍存在一些局限性：其卷积特征提取器依赖局部感受野，难以充分建模植株整体与局部之间的全局关系，在叶片密集和互相遮挡场景下识别性能下降[12,18－19]；其多尺度特征融合结构在小目标与尺度差异较大目标的检测中存在不足[11]；常用的 CIoU 边界框回归损失在复杂目标形态下收敛速度慢、定位精度受限[20]。在垂直农业场景的生长期识别任务中尤为突出，因为作物密度高、光照复杂且不同生

长期株型差异显著。尽管 YOLO 已迭代至 YOLO v12[21],其模型复杂度和计算开销较大,不利于资源受限的植物工厂环境下部署。综合考虑检测精度、速度及工程应用的成熟度,选择 YOLO v5 作为基础网络,并在此基础上提出改进的 SAB-YOLO 模型,以增强全局特征建模能力、优化多尺度特征融合效果、提升小目标检测与边界框回归精度,以期适应垂直农业场景下作物生长期识别任务。

1 材料和方法

1.1 数据来源

数据集来源于 2 个部分:①开源平台 Roboflow Universe 提供的生菜图像数据。②实验室自建的植物工厂种植环境下采集的数据。为了满足生长期识别需求,结合农业生产实际,将生菜生长过程划分为出芽期、幼苗期、快速生长期、品质形成期和采收调控期 5 个阶段[22]。开源数据集图像主要通过固定角度相机拍摄获得,通常从俯视角度(45°~90°)采集植株整体形态信息;实验室数据则采用安装在集装箱植物工厂顶部与侧面的萤石高清相机(及华为 Mate 50 Pro 型手机)定期采集,拍摄位置包括顶部俯视图和侧视图,以保证对叶片展开度、株高等特征的全面覆盖,由于采用红、蓝 LED 补光,故背景偏红色。不同阶段生菜图像如图 1 所示。

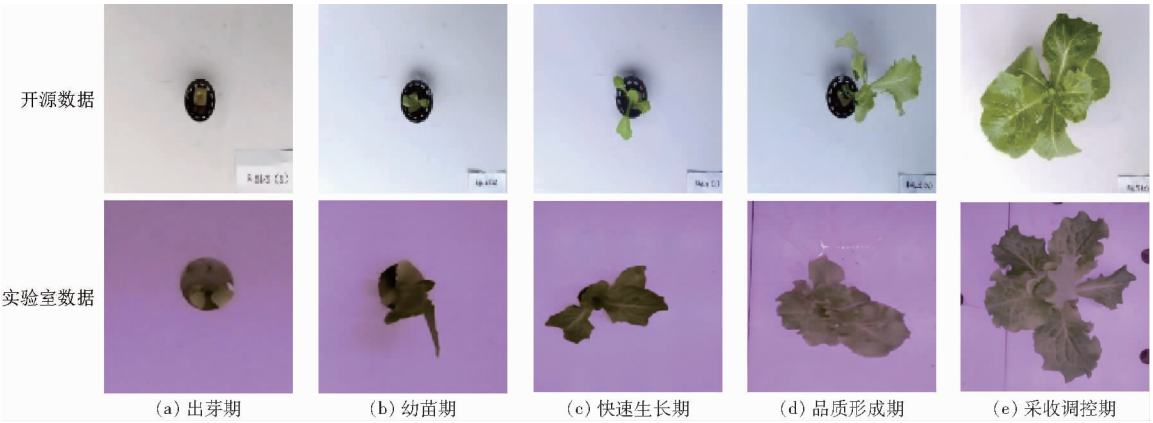

图 1 不同阶段生菜图像

Fig. 1 Images of lettuce at different growth stages

1.2 数据预处理

在数据规模上,共获取原始图像 1 250 幅,其中 Roboflow Universe 数据约占 40%,实验室采集数据约占 60%。为保证训练与验证的合理性,数据集按比例 7∶2∶1 划分为训练集、验证集和测试集。随后通过随机旋转、平移、镜像、滤波模糊和添加噪声等数据增强手段,将数据集扩充至 6 240 幅图像,每幅图像分辨率为 640 像素×640 像素。各生长阶段数据集如表 1 所示。

表 1 数据集分布

Tab. 1 Dataset distribution

数据集	出芽期	幼苗期	快速生长期	品质形成期	采收调控期	总数
训练集	714	1 065	1 065	1 030	491	4 365
验证集	330	328	259	254	94	1 265
测试集	147	90	147	147	79	610
总数	1 191	1 483	1 471	1 431	664	6 240

1.3 模型构建

提出一种基于 YOLO v5 的改进模型 SAB-YOLO,其网络结构如图 2 所示。将 YOLO v5 的特征提取网络替换为信息提取能力更强的 Swin Transformer 网络[23];改进 Neck 层网络结构,将原本的路径聚合特征金字塔网络(Path aggregation feature pyramid network, PAFPN)结构替换为更加密集的渐进特征金字塔网络(Asymptotic feature pyramid network, AFPN)[24],以增强 Neck 网络对于多尺度特征的融合能力;为进一步提高卷积网络对于小尺寸、多尺度以及高重叠的目标特征学习能力,提出了一种融合卷积与自注意力的卷积 Transformer 模块(Convolutional transformer fusion, CTF),并将其引入 YOLO v5 网络;最后引入 Inner-SIoU[25-26] 损失函数,以提高模型精确度和收敛速度。

1.3.1 Swin Transformer 特征提取网络

YOLO v5 所采用的 Darknet-53 Backbone 基于小卷积核构建,注重局部特征提取。然而,单纯依赖局部感受野导致模型较难捕获图像中不同区域间的全局语义关系[18]。在叶片密集或遮挡严重的垂直农业场景中,这种局部感受机制往往导致模型难以准确区分叶片与背景或辨识叶片间遮挡关系[19]。因此,为了提高模型在复杂环境下的识别性能,采用 Swin Transformer 替代原骨干网络,该网络基于移动窗口的多头自注意力 W-MSA(Window multi-head self-attention)和滑动窗口注意力 SW-MSA(Shifted window multi-head self-attention)机制实现跨窗口信

息交互，在降低计算复杂度的同时保持全局建模能力，其网络结构如图 3 所示。

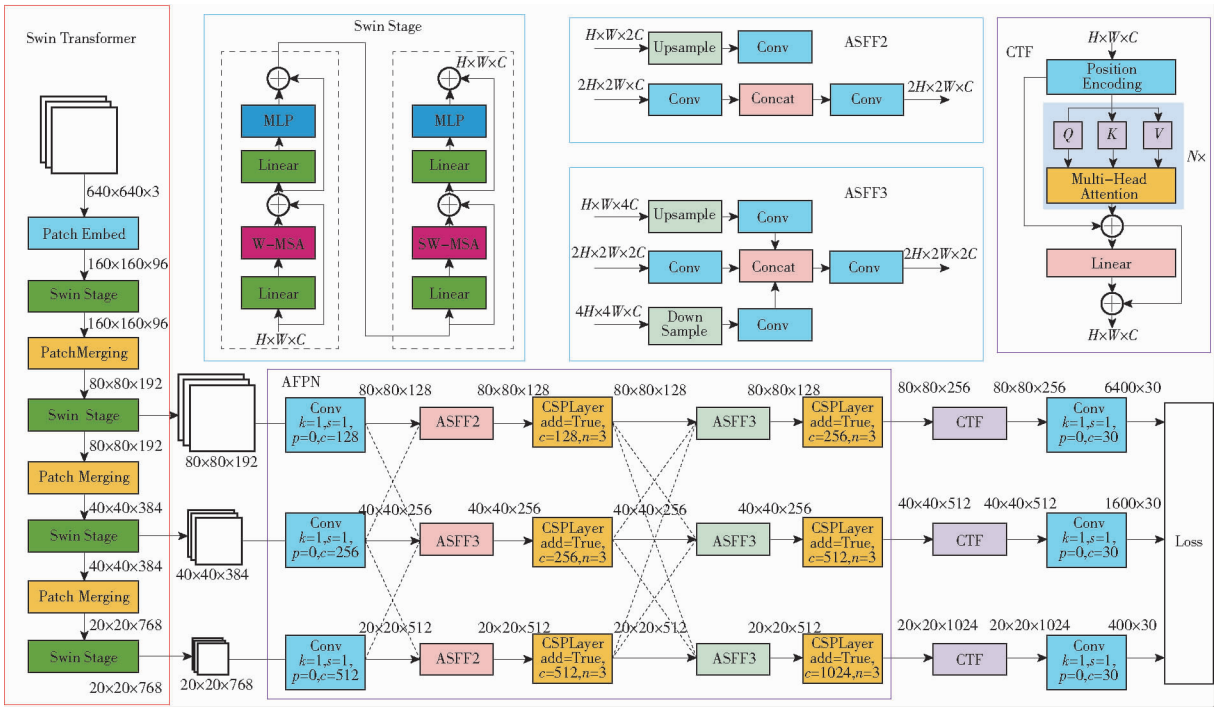

图 2　SAB－YOLO 网络结构图

Fig. 2　Diagram of SAB－YOLO network architecture

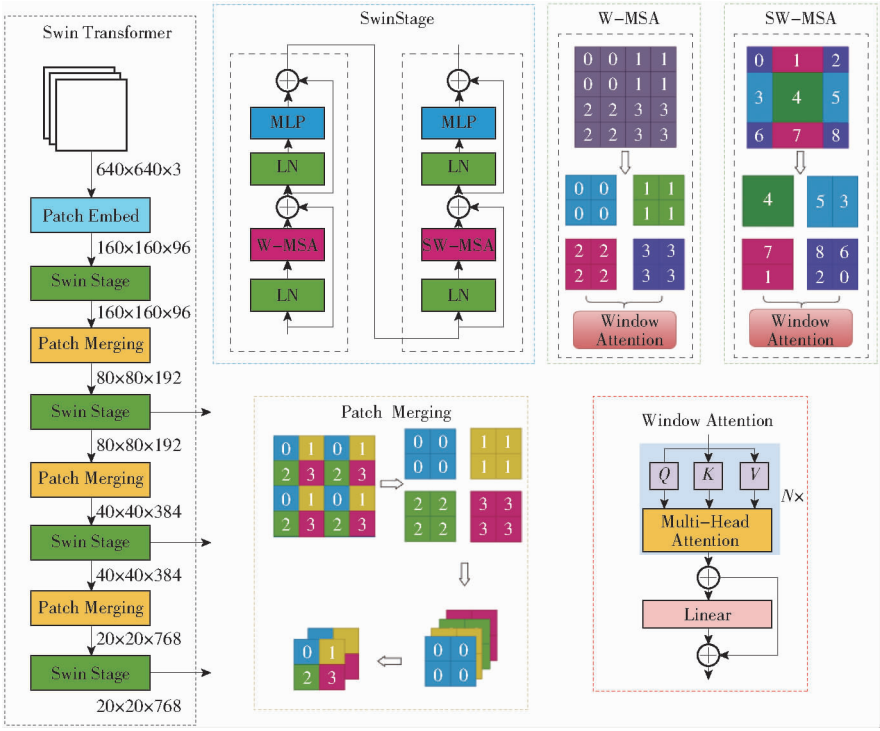

图 3　Swin Transformer 网络结构图

Fig. 3　Diagram of Swin Transformer network architecture

由图 3 可见，其由 Patch Embed 层、多级 Swin Stage 层和 Patch Merging 层构成渐进式特征金字塔。将 Swin Transformer 作为 YOLO v5 的特征提取骨干网络，采用四阶段层次化结构的 Swin Stage 作为特征提取器。输入图像经 4×4 非重叠切片嵌入后，依次通过 4 个 Swin Stage 模块，各阶段通过 Patch Merging 模块实现 2 倍下采样，最终输出特征图尺寸分别为输入图像的 1/4、1/8、1/16 和 1/32。输入长×宽×通道数（长、宽单位为像素）为 640×640×3 的图像经过 Swin Transformer Backbone 后输出 80×80×192、40×40×384 和 20×20×768 的特征图，分别对应小、中、大目标的检测。

1.3.2 渐进特征金字塔网络

在 YOLO v5 中,其多尺度融合网络 PAFPN 在进行跨尺度融合时,容易稀释浅层细节,导致跨层之间的特征交互减弱,不利于小目标检测[24,27]。AFPN 是一种针对多尺度目标检测任务设计的特征融合模块,其核心思想是通过动态学习空间权重,实现不同层级特征图的有效融合,以增强模型对多尺度目标的感知能力[24]。如图 2 所示,与传统的特征金字塔网络不同,AFPN 引入更加关注相邻尺度特征的自适应尺度特征融合模块(Adaptive scale feature fusion,ASFF)。该模块通过密集跨层连接与动态权重分配,使网络能够根据不同位置的语义重要性自适应调整融合策略。具体而言,AFPN 在相邻特征级别之间增加了双向(ASFF2)和三向连接(ASFF3),其结构如图 2 所示。其中,ASFF2 层通过动态权重叠加相邻层之间的有效信息;ASFF3 层则采用三叉戟式连接结构进行 3 个尺度上特征图动态融合,从而更有效地捕获多尺度特征。AFPN 在融合不同特征时采用加权特征融合机制,将不同级别的特征按照权重进行组合,在训练过程中不断拟合出最优权重,从而实现更好的多尺度特征融合效果。

构建 AFPN 作为 YOLO v5 的多尺度融合网络,在特征提取网络输出大、中、小 3 个尺度的特征图后,大、小尺度特征图通过 ASFF2 模块进行相邻层级之间的特征融合,中尺度特征则利用 ASFF3 融合小尺度和大尺度上的信息,从而提高模型多尺度泛化能力。首次融合后,各个尺度上的特征图输入 CSPLayer 残差模块以避免梯度消失;再通过 ASFF3 模块实现 3 个尺度上的二次融合;最终输出的 3 个尺度特征图都充分引入各个尺度之间的有效信息,增强了模型对于多尺度目标的识别能力。

1.3.3 注意力机制改进

传统的注意力机制如挤压激励网络(Squeeze-and-excitation networks,SE)、卷积块注意力模块(Convolutional block attention module,CBAM)、高效通道注意力模块(Efficient channel attention,ECA)[28-30]等,主要基于局部感受野构建通道或空间注意力。这类机制难以有效捕获图像全局区域关联性,对跨区域之间的特征建模能力不足[31-33]。引入基于 Transformer 的自注意力机制,通过计算图像不同区域之间的权重,隐式建模区域间依赖关系,使网络能够自主学习图像中关键区域与背景区域的差异化表征,显著增强对目标主体特征的聚焦能力。

基于 Transformer 的自注意力机制模块整合进卷积网络中,作为 CTF 结构,并添加于 Head 层之前,CTF 结构如图 2 所示。这种整合方式实现了卷积网络对于局部特征的感知优势与自注意力机制全局上下文建模能力的有机融合。CTF 结构不仅保留了卷积操作对局部细节特征的提取能力,同时赋予了检测头进行跨区域关联的能力,从而提升模型整体表征能力。

1.3.4 损失函数改进

YOLO v5 采用的边界框回归损失函数为 CIoU。CIoU 在 IoU 基础上引入中心点距离和长宽比惩罚,在一定程度上提升了定位精度。然而,已有研究指出 CIoU 仍存在局限性:其计算过程未显式建模预测框与真实框的角度对齐关系,导致在目标形状差异较大时收敛速度偏慢;同时对长宽比的强制约束易造成过拟合,使得模型在复杂场景下的泛化能力不足[18,34]。因此,近年来研究者提出了 SIoU、EIoU 等改进型损失函数,以提升边界框回归几何建模能力和收敛效率。

内部斯库拉交并比损失函数(Inner - Scylla IoU,Inner_SIoU)[18]通过引入辅助缩放框、角度对齐惩罚、方向感知距离惩罚和动态形状惩罚,改善了收敛速度和检测精度。其计算式为

$$\Gamma_S = 1 - I_p + \frac{\Delta_a + \Delta_d + \Delta_s}{2} \tag{1}$$

式中 Δ_a——角度成本

Δ_d——距离成本

Δ_s——形状成本

I_p——内部框交集、并集面积比

其中I_p,即 Inner_IoU,是一种通过不同尺度的辅助边界框来计算损失的新型 IoU 度量方法。传统 IoU 对边界框整体重叠敏感,但对中心对齐的细微差异缺乏区分度。Inner_IoU 通过缩放原始框生成内部区域,迫使模型关注中心对齐,显著提升了小目标检测和边界框回归精度。

(1)交集面积 I_t 计算式为

$$I_t = \max(0,\min(x_{1,\max},x_{2,\max})) - \max(x_{1,\min},x_{2,\min}) \cdot \max(0,\min(y_{1,\max},y_{2,\max})) - \max(y_{1,\min},y_{2,\min}) \tag{2}$$

式中 x_1、y_1——预测框中心坐标

x_2、y_2——真实框中心坐标

$x_{1,\min}$、$x_{1,\max}$、$y_{1,\min}$、$y_{1,\max}$——预测框缩小后内部框中心坐标

$x_{2,\min}$、$x_{2,\max}$、$y_{2,\min}$、$y_{2,\max}$——真实框缩小后内部框中心坐标

(2)内边框并集面积 I_u 计算式为

$$I_u = w_1 h_1 r^2 + w_2 h_2 r^2 - I_t \tag{3}$$

式中 w_1、h_1——预测框宽度和高度

w_2、h_2——真实框宽度和高度

r——内部框缩放比例

(3)交集、并集面积比计算式为

$$I_p = \frac{I_t}{I_u} \tag{4}$$

角度成本Δ_a为中心连线与坐标轴偏差，计算式为

$$\Delta_a = \cos\left(2\arcsin(\sin\alpha) - \frac{\pi}{2}\right) \tag{5}$$

其中

$$\sin\alpha = \frac{|x_o|}{\sqrt{x_o^2 + y_o^2}} \tag{6}$$

$$\begin{cases} x_o = \dfrac{x_2 - x_1}{2} \\ y_o = \dfrac{y_2 - y_1}{2} \end{cases} \tag{7}$$

式中　α——中心连线与水平轴夹角

x_o、y_o——中心点偏移量，即预测框与真实框中心坐标差异

距离成本Δ_d为动态调整中心点偏移距离的成本权重，计算式为

$$\Delta_d = 2 - e^{-\gamma\rho_x} - e^{-\gamma\rho_y} \tag{8}$$

其中

$$\rho_x = \left(\frac{x_o}{c_w}\right)^2 \qquad \rho_y = \left(\frac{y_o}{c_h}\right)^2$$

式中　ρ_x、ρ_y——归一化距离偏移，用来衡量中心点偏移与包围框尺寸比例

c_w、c_h——最小包围框宽、高

形状成本Δ_s衡量宽高比差异，计算式为

$$\Delta_s = (1 - e^{-w_w})^4 + (1 - e^{-w_h})^4 \tag{9}$$

其中

$$\begin{cases} w_w = \dfrac{|w_1 - w_2|}{\max(w_1, w_2)} \\ w_h = \dfrac{|h_1 - h_2|}{\max(h_1, h_2)} \end{cases} \tag{10}$$

式中　w_w、w_h——宽高比差异，反映预测框与真实框之间尺寸差异

改进的 Inner_SIoU 损失函数根据目标尺寸自动调整损失计算策略，对于小目标可放大内部区域损失贡献，缓解小目标因像素少导致的漏检问题；对于大目标可抑制边缘区域权重，避免大目标边缘噪声影响定位精度，可有效提升边界框回归精度和训练效率。

2　试验

2.1　试验环境及参数设置

试验操作系统为 Ubuntu 16.04，GPU 为 NVDIA GeForce A800，显存大小为 80 GB；CPU 为 Inter(R) Xeon(R) E5－2618L；使用的深度学习框架是 Pytorch 1.9.0。初始学习率设为 0.01，动量为 0.9，批处理尺寸为 8，迭代次数为 100。

2.2　结果与分析

2.2.1　不同主干特征提取网络性能对比

通过在相同环境下复现典型主干网络(YOLO v5－Darknet53、Swin Transformer、EfficientNet 等)性能进行对比。同时，结合已有作物生长期识别相关研究[3,5,13,24]，验证了所提方法在垂直农业环境下的有效性。选取 YOLO v5－Darknet53、Swin Transformer、EfficientNet、RepVGGNet、C2FNet(Context-aware cross-level fusion network)和 YOLO v12[35-38]6 种典型骨干网络进行对比，在测试集上试验结果如表 2 所示。由表 2 可知，Swin Transformer 精确率和召回率，较 RepVGGNet 提升 11.3、0.5 百分点。从鲁棒性上看，Swin Transformer 的 mAP_0.5 和 mAP_0.5:0.95 达到 89.0% 和 72.7%，mAP_0.5:0.95 较 RepVGGNet 提升 1.6 个百分点，表明其在不同 IoU 阈值下具有更好的定位稳定性。相关研究表明，Swin Transformer 等分层视觉 Transformer 在处理复杂背景和小目标任务时，能够有效缓解 CNN 结构中易出现的特征混淆和局部依赖不足问题[23,33]。

表 2　不同主干网络试验结果

Tab.2　Experimental results of different backbone networks

主干网络模型	参数量	精确率/%	召回率/%	mAP_0.5/%	mAP_0.5:0.95/%
YOLO v5－Darknet53	4.61×10^7	73.7	76.8	81.5	66.0
EfficientNet	2.29×10^7	77.9	75.1	86.2	51.4
RepVGGNet	4.68×10^7	68.9	87.6	88.3	71.1
C2FNet	6.10×10^7	62.6	77.2	83.4	60.7
YOLO v12	2.64×10^7	69.2	82.9	80.1	65.4
Swin－Transformer	4.68×10^7	80.2	88.1	89.0	72.7

尽管 Swin Transformer 在精确率、召回率及 mAP 等指标上均表现出明显优势，但其参数量也较其他模型有所增加，意味着更强的特征表达能力与更复杂的网络结构，从而在性能上获得提升。Swin Transformer 在一定程度上反映模型复杂度提升带来的性能增益。此外，较大的模型规模也会带来更高

的计算与存储开销,在实际部署到资源受限的植物工厂场景时可能受到限制。

2.2.2 不同多尺度融合网络性能对比

选取 YOLO v5 - PAFPN、双向特征金字塔网络(Bidirectional feature pyramid network,BiFPN)、全局特征金字塔网络(Global feature pyramid network,GFPN)、空间注意力特征金字塔网络(Spatial attention feature pyramid network,SAFPN)[39-41]和 AFPN 5 种多尺度融合网络作为 Neck 层进行对比试验,在测试集上试验结果如表 3 所示。在 5 种融合网络中 AFPN 召回率为 84.9%,表明 AFPN 能有效减少漏检,同时 mAP_0.5 达 89.3%,说明在 IoU 阈值较低时,AFPN 对生菜多个阶段图像的识别效果最优;精确率 AFPN 虽不及 BiFPN,但较 YOLO v5 - PAFPN 仍提升 6.1 个百分点。综合来看,AFPN 在精确度、召回率上有着明显优势。

表 3 不同融合网络性能试验结果

Tab.3 Experimental results of different fusion networks

颈部模型	参数量	精确率/%	召回率/%	mAP_0.5/%	mAP_0.5:0.95/%
YOLO v5 - PAFPN	4.61×10^7	73.7	76.8	81.5	66.0
BiFPN	4.61×10^7	82.1	81.6	84.0	65.3
GFPN	6.68×10^7	78.9	84.6	85.9	60.4
SAFPN	4.84×10^7	65.2	81.2	80.2	53.9
AFPN	4.33×10^7	79.8	84.9	89.3	65.8

2.2.3 不同注意力机制性能对比

选取 SE、ECA 和 CBAM 3 种常用的注意力机制进行对比试验,在测试集上试验结果如表 4 所示。由表 4 可知,CTF 模型召回率达 84.1%,明显高于其他注意力机制。表明 CTF 在识别生菜不同生长阶段图像上更加全面,能够有效降低漏检率。此外 CTF 的 mAP_0.5:0.95 也达到 66.0%,表明其在各种 IoU 阈值下表现更好,说明其稳定性和鲁棒性,在实际应用中能够更好地处理复杂场景,特别是目标密集场景。整体来看,CTF 模型在多项性能指标中展现出了较强的综合能力,能够更好地完成生菜生长期识别。

表 4 不同注意力机制性能试验结果

Tab.4 Experimental results of different attention mechanisms

注意力模型	参数量	精确率/%	召回率/%	mAP_0.5/%	mAP_0.5:0.95/%
YOLO v5	4.61×10^7	78.4	67.1	77.6	54.7
SE	4.61×10^7	75.4	75.9	87.0	65.5
ECA	4.61×10^7	83.0	75.3	88.0	64.9
CBAM	4.63×10^7	73.8	78.8	83.1	64.2
CTF	5.23×10^7	76.2	84.1	87.1	66.0

2.2.4 不同损失函数性能对比

选取 CIoU、Inner - CIoU、内部加权交并比(Inner - Wise IoU,Inner - WIoU)和内聚高效交并比(Inner - Focal and Efficient IoU,Inner - EIoU)[42-43]4 种损失函数进行对比试验,在测试集上试验结果如表 5 所示。Inner - SIoU mAP_0.5 为 90.4%,显示了其在 IoU 为 0.5 时的良好检测能力。同时,Inner - SIoU 的 mAP_0.5:0.95 达 74.3%,表明其在多种重叠阈值下均能保持较高的检测性能,展示了在较复杂场景中的适应性和稳定性。

表 5 不同损失函数性能试验结果

Tab.5 Experimental results of different loss functions

%

注意力模型	精确率	召回率	mAP_0.5	mAP_0.5:0.95
YOLO v5 - CIoU	78.4	67.1	77.6	54.7
Inner - CIoU	71.7	83.7	85.1	70.4
Inner - WIoU	77.2	88.3	87.3	72.3
Inner - EIoU	68.9	84.5	86.9	70.8
Inner - SIoU	82.5	85.9	90.4	74.3

2.2.5 消融试验

通过逐步引入 Swin Transformer Backbone、AFPN、CTF 和 Inner - SIoU 损失函数 4 个核心模块,验证各个改进对目标检测性能的贡献。测试集上试验结果如表 6 所示。由表 6 可知,模型 mAP_0.5 达 92.1%及 mAP_0.5:0.95 达 75.9%,精确率达 88.5%及召回率达 87.0%,验证了各模块协同作用。

表 6 消融试验结果

Tab.6 Experimental results of ablation experiment

%

主干网络	颈部	注意力机制	损失函数	精确率	召回率	mAP_0.5	mAP_0.5:0.95
×	×	×	×	78.4	67.1	77.7	54.7
Swin Transformer	×	×	×	80.2	88.1	89.0	72.7
×	AFPN	×	×	79.8	84.9	89.3	65.8
×	×	CTF	×	76.2	84.1	87.1	66.0
×	×	×	Inner - SIoU	82.5	85.9	90.4	74.3
Swin Transformer	AFPN	×	×	78.1	80.2	83.2	67.4
Swin Transformer	AFPN	CTF	×	75.5	89.6	87.1	71.7
Swin Transformer	AFPN	CTF	Inner - SIoU	88.5	87.0	92.1	75.9

注:×表示未使用模块。

3　基于生长期识别的智能补光系统设计

SAB－YOLO 模型可用于植物工厂、垂直农业等场景的作物生长期自动识别。已有研究表明，基于作物生长阶段的光照调控能够在保障产量和品质的同时实现节能[2]。例如在幼苗期需较低光强、较高相对湿度与适中温度以促进根系发育；在快速生长期需较高光强、适当降低相对湿度并提高营养液浓度以加快光合产物积累；在品质形成期则需控制光周期和温度以避免早薹抽薹、保持风味品质。为拓展该模型的应用，在前述生长期识别模型基础上，以集装箱垂直农业为场景，设计一套智能补光系统，可根据生菜所处生长阶段自动调节光强、光质和光周期，以优化光环境并提升资源利用效率。基于生长期识别智能补光系统由图像采集模块、Web 环境控制模块、生长检测模块、数据存储模块和智能调控模块 5 部分组成，如图 4 所示。

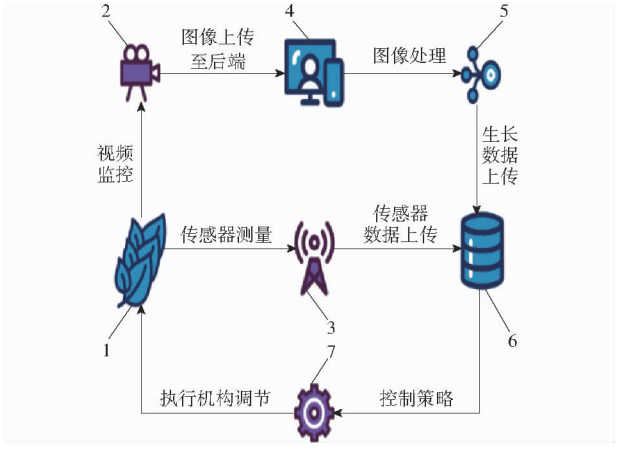

图 4　智能补光系统架构

Fig. 7　Intelligent supplementary lighting system architecture

1. 生菜　2. 相机　3. 传感器　4. Web 环境控制模块　5. 生长检测模块　6. 数据存储模块　7. 智能调控模块

由图 4 可知，系统通过相机以及传感器定期收集生菜生长图像和环境数据，作为系统基础信息；生长检测模块即为本文生长期识别模型 SAB－YOLO，服务器将接收到的生菜生长实时图像进行预处理后，通过生长期识别模型评估当前生菜所处的生长阶段；数据存储模块将当前生菜生长信息、分阶段补光配方以及控制策略进行存储；最后，智能调控模块则基于收到的控制参数，动态调节 LED 灯具的电压、光周期、空调的设定温度、配肥机的 EC/pH 设定阈值等参数，从而实时调控生菜最佳生长环境。

生菜分阶段补光模型所用补光方案如表 7 所示，通过 SAB－YOLO 算法识别作物生长期优化环境调控决策，从而减少无效补光能源损耗，实现节能

表 7　分阶段补光方案

Tab. 7　Staged supplementary lighting scheme

生长阶段	出芽期	幼苗期	快速生长期	品质形成期	采收调控期
预估周期/d	0～5	6～15	16～30	31～35	36
光质(红:蓝:绿)	1:0:0	4:1:0	3:1:0	2:1:1	1:1:0
光照强度/(μmol·m^{-2}·s^{-1})	50～100	100～200	300～400	250～300	150～200
光周期/h	12	14	16	14	12

效果，同时针对作物当前需求，动态调节环境参数，达到优化生产品质的目的。该模型应用于上海市崇明岛光明母港花博会基地集装箱式植物工厂生产作业中，植物工厂软硬件控制系统如图 5 所示。系统于 2025 年 5 月 18 日开展了约 40 d 的种植试验，在集装箱两侧放置 5 层栽培架，每层种植意大利生菜约 120 棵，栽培架上每层各安装 1 个相机，相机视场范围内有 2～4 棵生菜，选择拍摄效果较好的 2 棵作为监测对象，从播种后第 3 天开始与人工同步监测，共记录 35 d。控制环境温度 18～25℃，相对湿度 65%～75%，通过监测 EC、pH 值控制配肥机自动配肥。在配备 NVIDIA RTX 4060 GPU 的服务器上，模型平均推理速度为 55 f/s，单帧推理耗时约18.2 ms；从集装箱图像数据采集、经网络传输到云端分析、再到树莓派推理结果发送，全流程平均响应时间约为 1.8 s，总体识别精度约 82.1%，满足实时性和鲁棒性要求，适用于集装箱种植环境检测任务。

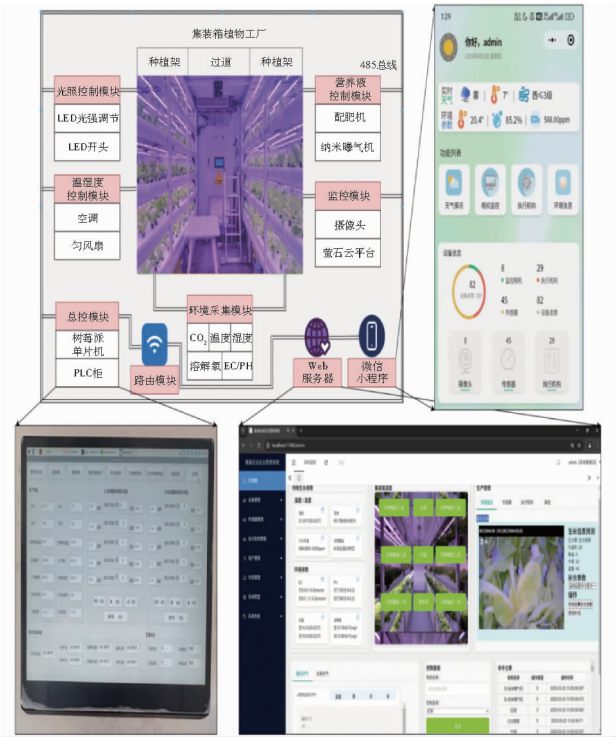

图 5　集装箱植物工厂软硬件控制系统

Fig. 5　Software and hardware control system for container-based plant factories

4 结论

(1)提出了一种基于 YOLO v5 的改进模型 SAB - YOLO,该模型在 5 个生长阶段平均识别精度达到 88.5%,比 YOLO v5 - Darknet53 提高 10.1 个百分点,表明引入 Swin Transformer 的全局建模能力和 AFPN 的多尺度特征融合确实有效增强了对复杂背景下目标的识别性能。

(2)消融试验结果表明,改进后模型精确率达到 88.5%、mAP_0.5 为 92.1%,进一步验证了各改进模块的贡献。

(3)以上海市崇明岛光明母港花博会基地的集装箱垂直农业环境为应用场景,设计了一种基于该模型的补光控制系统,并进行了初步验证,结果表明,本文 SAB - YOLO 模型可用于垂直农业场景的叶菜生长期自动识别。

参 考 文 献

[1] SAKAMOTO T, GITELSON A A, NGUY-ROBERTSON A L, et al. An alternative method using digital cameras for continuous monitoring of crop status[J]. Agricultural and Forest Meteorology, 2011, 154 - 155:113 - 126.

[2] 陈杰杰. 智能植物工厂数据处理与预测方法研究[D]. 天津:天津职业技术师范大学, 2020.
CHEN Jiejie. Research on data processing and prediction methods for intelligent plant factories [D]. Tianjin: Tianjin University of Vocational and Technical Teacher Education, 2020. (in Chinese)

[3] 扶兰兰, 黄昊, 王恒, 等. 基于 Swin Transformer 模型的玉米生长期分类[J]. 农业工程学报, 2022, 38(14): 191 - 200.
FU Lanlan, HUANG Hao, WANG Heng, et al. Corn growth stage classification based on Swin Transformer model[J]. Transactions of the CSAE, 2022, 38(14): 191 - 200. (in Chinese)

[4] DU Jianjun, LU Xianju, FAN Jiangchuan, et al. Image-based high-throughput detection and phenotype evaluation method for multiple lettuce varieties[J]. Frontiers in Plant Science, 2020, 11: 563386.

[5] 李修华, 项志伟, 郭新宇, 等. 基于图像的生菜表型高通量获取方法[J]. 江苏农业科学, 2022, 50(20): 1 - 9.
LI Xiuhua, XIANG Zhiwei, GUO Xinyu, et al. High-throughput phenotypic acquisition method for lettuce based on image analysis[J]. Jiangsu Agricultural Sciences, 2022, 50(20): 1 - 9. (in Chinese)

[6] 刘林, 苑进, 张岩, 等. 日光温室基质培生菜鲜质量无损估算方法[J]. 农业机械学报, 2021, 52(9): 230 - 240.
LIU Lin, YUAN Jin, ZHANG Yan, et al. Non-destructive estimation method for fresh quality of lettuce grown in sunlight greenhouses using substrate cultivation[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(9): 230 - 240. (in Chinese)

[7] STEPHI S, NAIR J J. Wheat disease detection and growth stage monitoring using deep learning architectures[C]// 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2023: 1 - 5.

[8] XU Z, YANNG S, HUANG J, et al. Study on classification of seafood mushroom growth stages based on deep learning[C]// 2023 3rd International Conference on Electronic Information Engineering and Computer Science (EIECS), 2023: 1088 - 1091.

[9] LIMA F, OLIVEIRA R, DIAS J, et al. Deep learning structure for real-time crop monitoring based on neural architecture search and UAV[J]. Brazilian Archives of Biology and Technology, 2023, 66: e23230046.

[10] KANG H, PARK J, KIM H, et al. Crop growth monitoring system in vertical farms based on region-of-interest prediction[J]. Agriculture, 2022, 12(5): 656.

[11] LIU H, WU J, LI X, et al. Computer vision-based recognition of small targets in complex agricultural environments[J]. Applied Sciences, 2024, 15(15): 8438.

[12] WU D, MA X, ZHANG J, et al. Applications of computer vision in agriculture: current challenges and future trends[J]. Frontiers in Plant Science, 2022, 13: 865271.

[13] YU Y, LI C, LIU F, et al. Wheat growth stage identification method based on multimodal data[J]. European Journal of Agronomy, 2024, 152: 126823.

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

[15] 王兴旺,查海涅,卢浩男,等. 基于 YOLO v8 STSF 的多类别害虫识别算法与监测系统研究[J]. 农业机械学报,2025,56(6):228 - 236.
WANG Xingwang, ZHA Hainie, LU Haonan, et al. Multi-category pest identification algorithm and monitoring system based on YOLO v8 STSF[J]. Transactions of the Chinese Society for Agricultural Machinery,2025,56(6):228 - 236. (in Chinese)

[16] 王泰华,郭亚州,张家乐,等. 基于改进 YOLO v5s 的水稻害虫识别研究[J]. 农业机械学报,2024,55(11):39 - 48.
WANG Taihua, GUO Yazhou, ZHANG Jiale, et al. Rice pest identification based on improved YOLO v5s[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(11):39 - 48. (in Chinese)

[17] SHAO Y H, ZHANG D, CHE H Y, et al. A review of YOLO object detection based on deep learning[J]. Journal of Electronics & Information Technology, 2022, 44(10): 3697 - 3708.

[18] ZHORA G. SIoU loss: more powerful learning for bounding box regression[J]. arXiv preprint arXiv: 2205.12740,2022.

[19] ZHOU L, WANG H, ZHANG X, et al. A review on crop pest classification: challenges and advances with deep learning[J].

arXiv preprint arXiv:2507.01494, 2025.
[20] GEVORGYAN Z. SIoU loss: more powerful learning for bounding box regression[J]. arXiv preprintarXiv:2205.12740, 2022.
[21] KHANAM R,HUSSAIN M. A review of YOLO v12: attention-based enhancements *vs*. previous versions[J]. arXiv preprint arXiv:2504.11995,2025
[22] LI H, ZHAO Y, ZHANG J, et al. Image-based monitoring of lettuce growth in plant factories using machine vision techniques [J]. Computers and Electronics in Agriculture, 2021, 188:106350.
[23] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted wind-ows[C]//18th IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 9992－10002.
[24] YANG G, LEI J, ZHU Z, et al. AFPN: asymptotic feature pyramid network for object detection[C]//2023 IEEE International Conference on Systems, Man, and Cybernetics, 2023: 2184－2189.
[25] DANG Yuanjie, CHEN Shuailin, MIAO Haochun, et al. Target detection from drone perspectives: enhancing YOLO v5_3s with SIoU loss and SPD modules[C]//2023 Cross Strait Radio Science and Wireless Technology Conference, 2023: 1－3.
[26] CAO L, WANG Q, LUO Y, et al. YOLO－TSL: a lightweight target detection algorithm for UAV infrared images based on triplet attention and slim-neck[J]. Infrared Physics and Technology, 2024,141:105487.
[27] ZHANG Y, WANG J, LI X, et al. MHAF－YOLO: multi-branch heterogeneous auxiliary fusion network for object detection [J]. arXiv preprint arXiv:2502.04656, 2025.
[28] HU J, SHENN L, SUN G. Squeeze-and-excitation networks[C]//31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 7132－7141.
[29] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//15th European Conference on Computer Vision (ECCV), 2018: 3－19.
[30] WANNG Q, WU B, ZHU P, et al. ECA－Net: efficient channel attention for deep convolutional neural networks[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 11531－11539.
[31] CHENG J, DONG L, LAPATA M. An empirical study of spatial attention mechanisms in deep networks[J]. arXiv preprint arXiv:1904.05873, 2019.
[32] FU J, LIU J, TIAN H, et al. DANet: dual attention network for scene segmentation[C]//CVPR, 2019:3146－3154.
[33] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
[34] ZHANG H, XU C, ZHANG S. Inner－IoU: more effective intersection over union loss with auxiliary bounding box[J]. arXiv preprint arXiv:2311.02877, 2023.
[35] TANN M, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks[C]//36th International Conference on Machine Learning, 2019:10691－10700.
[36] DINNG X, ZHANNG X, MA N, et al. RepVGG: making VGG-style convnets great again[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2021: 13728－13737.
[37] CHEN Genng, LIU Sijie, SUN Yujia, et al. Camouflaged object detection via context-aware cross-level fusion[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(10): 6981－6993.
[38] TIAN Y, YE Q, DOERMANN D, et al. YOLOv12: attention-centric real-time object detectors[J]. arXiv preprint arXiv: 2502.12524, 2025.
[39] TAN Mingxing, PANG Ruoming, LE Quocv. EfficientNet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2020:10778－10787.
[40] LIU L, WANG R, XIE C, et al. A global activated feature pyramid network for tiny pest detection in the wild[J]. Machine Vision and Applications, 2022, 33(5):76－90.
[41] ZHOU Zhiying, XIAO Jianqiang, SATOSHI Y. SAFPN: self adapted feature pyramid networks for object detection[C]//2021 IEEE 10th Global Conference on Consumer Electronics, 2021: 658－660.
[42] LIU Zhigang, SUN Baoshan, BI Kaiyu. Optimization of YOLO v7 based on PCONV, SE attention and Wise－IoU[J]. International Journal of Computational Intelligence and Applications, 2024, 23(1): 147－173.
[43] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient IoU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506(28): 146－157.