doi:10.6041/j.issn.1000-1298.2026.03.034

基于动态可变形卷积与注意力融合的密集桑蚕分割算法

文春明[1]　刘　康[1]　贾　讯[1]　徐　咏[2]　安　慧[2]　梁　湘[3]　廖义奎[4,5]

(1.广西智能视觉协作机器人工程研究中心,南宁 530006;2.广西高校人工智能与信息处理重点实验室,河池 546300;
3.亚热带农业生物资源保护与利用国家重点实验室,南宁 530004;
4.广西混杂计算与集成电路设计分析重点实验室,南宁 530006;5.广西智语人形机器人重点实验室,南宁 530006)

摘要:实现桑蚕自动化计数是蚕桑养殖智能化管理的关键环节。然而,由于桑蚕个体尺寸小、形态多变且在密集养殖环境下存在严重遮挡,人工肉眼计数的方法难以实现精确计数且费时费力,导致无法根据桑蚕数量精准投放桑叶和石灰粉以确保其精准进食量和消毒环境,桑叶投放过少影响桑蚕生长,过多则造成浪费与环境污染,石灰粉用量不足则消毒效果有限,过量又可能损害蚕体健康。为解决密集桑蚕计数困难并提高桑蚕检测与分割的准确性,本文提出了一种基于 YOLO 11n 改进的实例分割算法 DLC-YOLO,首先,在主干网络设计了 DysnakeConv 与 C3k2 融合结构,通过动态可变形卷积核沿桑蚕轮廓进行自适应形变,增强对桑蚕细长弯曲形态的特征提取能力。其次,在颈部网络引入了基于 CGAfusion 的内容引导注意力融合机制改变颈部原本的特征融合方式,在检测头输入端融合主干网络与特征金字塔的多尺度跨层特征,有效提升了多尺度特征的融合效率。最后,设计了 LSCSBD-Seg 的分割检测头,通过局部语义约束和边界细化模块,有效改善了密集桑蚕个体的边缘分割精度。为验证性能,构建了密集桑蚕实例分割数据集,涵盖不同养殖密度和光照条件。实验结果表明,DLC-YOLO 算法与 YOLO 11n 算法相比,检测精度 mAP50、mAP50:95 分别提高了 2.3、5.1 个百分点,分割精度 mAP50、mAP50:95 分别提高 3.0、2.9 个百分点,均优于原始 YOLO 11n 模型。

关键词:密集桑蚕计数;农业智能化管理;实例分割;多尺度特征融合;动态可变形卷积

中图分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1000-1298(2026)03-0353-12　　**OSID**:

Dense Silkworm Segmentation Algorithm Based on Dynamic Deformable Convolution and Attention Fusion

WEN Chunming[1]　LIU Kang[1]　JIA Xun[1]　XU Yong[2]　AN Hui[2]　LIANG Xiang[3]　LIAO Yikui[4,5]

(1. *Guangxi Engineering Research Center of Intelligent Vision and Collaborative Robotics*, *Nanning* 530006, *China*
2. *Guangxi Key Labratory of Artificial Intelligence and Information Processing*, *Hechi* 546300, *China*
3. *State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources*, *Nanning* 530004, *China*
4. *Guangxi Key Laboratory of Hybrid Computation and IC Design Analysis*, *Nanning* 530006, *China*
5. *Guangxi Key Labooratory of Zhiyu Humanoid Robot*, *Nanning* 530006, *China*)

Abstract: Automated counting of silkworms is crucial for intelligent sericulture management. However, due to small individual size, morphological variations, and severe occlusion in dense farming environments. This impeded precise distribution of mulberry leaves and lime powder for optimal feeding and disinfection. To address dense silkworm counting challenges and enhance detection and segmentation accuracy, DLC-YOLO, an improved instance segmentation algorithm was proposed based on YOLO 11n. Firstly, in the backbone network, a DysnakeConv and C3k2 fused structure was designed, utilizing dynamic deformable convolution kernels to adaptively trace silkworm contours, strengthening feature extraction for slender, curved morphologies. Secondly, in the neck network, a CGAfusion content-guided attention mechanism replaced the original feature fusion, integrating multi-scale cross-layer features from the backbone and feature pyramid at the detection head input. This significantly improved multi-scale

收稿日期:2025-08-20　修回日期:2025-09-20

基金项目:国家自然科学基金地区科学基金项目(62063006)、广西科技计划项目(桂科 AA24010001)、广西自然科学基金项目(2025GXNSFAA069536)和广西高校人工智能与信息处理重点实验室开放课题(2022GXZDSY009、2024GXZDSY017)

作者简介:文春明(1969—),男,教授,博士,主要从事人工智能、智能微系统等研究,E-mail:wenchunming@gxmzu.edu.cn

通信作者:徐咏(1989—),男,副教授,博士,主要从事嵌入式与智能控制研究,E-mail:xuyong@hcnu.edu.cn

feature fusion efficiency, particularly for dense small targets. Finally, an LSCSBD - Seg segmentation head incorporated local semantic constraints and boundary refinement modules to enhance edge segmentation precision in dense clusters. A dedicated dense silkworm instance segmentation dataset covering varied rearing densities and lighting conditions was constructed for validation. Experimental results demonstrated that DLC - YOLO outperformed YOLO 11n, with detection precision (mAP50, mAP50:95) increasing by 2.3 and 5.1 percentage points, and segmentation precision (mAP50, mAP50:95) improving by 3.0 and 2.9 percentage points, respectively.

Key words: dense silkworm counting; intelligent agricultural management; instance segmentation; multi-scale feature fusion; dynamic deformable convolution

0 引言

桑蚕养殖作为重要的传统农业产业,为丝绸产业提供核心原料,具有显著的经济价值[1-3]。桑蚕计数对精准投喂桑叶和石灰粉以保障桑蚕健康生长与环境消毒至关重要。然而,当前桑蚕生产仍以分散化小农户为主,技术渗透率低。依赖人工巡检和肉眼计数的方法不仅难以实现精确统计,且费时费力、易出错,导致无法根据实际桑蚕数量科学投喂与消毒。因此,发展高效、准确的自动化桑蚕计数技术,实现桑蚕养殖智能化管理,对提升养殖标准化水平、推动产业集约化发展具有重要意义[4]。

近年来深度学习驱动的高效视觉检测框架广泛应用于农业领域[5-7],尤其是对小目标的研究取得一定进展。何锐敏等[8]改进 Mask R-CNN 实现蚕体 96.83% 的识别准确率。KM-YOLO v4 在密集桑蚕检测中 mAP 达 69%[9]。马小林等[10]基于 YOLO v8n 的改进模型使 mAP50 达 93.2%。靳冲等[11]通过优化 YOLO v5 提升了农田害虫检测的鲁棒性。2024 年,本课题组采用多尺度残差网络融合 SENet 方法,使桑叶病害识别准确率达 98.72%[12]。尽管上述方法显著提升了检测效率,但在密集桑蚕养殖场景下面临形态多变、严重遮挡和个体粘连的挑战时,分割精度仍显不足。基于此,本文提出一种基于 YOLO 单阶段检测框架[13]开发的 DLC-YOLO 模型,相较于传统图像处理和现有研究中的局限,通过动态视觉感知实现个体定位和数量统计,在密集蚕群计数和复杂场景适应性方面具有一定优势。此模型重点针对密集桑蚕场景下的检测与分割难题进行系统性优化。通过构建包含动态形态适配[14]、多尺度特征增强[15]和边界精修的 3 个改进,形成完整的改进路径。

1 数据集

1.1 数据集获取

当前主流目标检测算法普遍采用监督学习范式,其性能高度依赖标注数据的质量与规模。为解决高精度标注数据的需求,数据采集自广西壮族自治区河池市宜州区庆远镇龙江路 254 号的桑蚕养殖数字化创新示范基地,桑蚕品种为两广二号,在温度为 22~26℃和相对湿度为 75%~80% 的桑蚕标准饲养环境中,通过 4 096 像素×3 072 像素的 REDMI K80 型相机采集桑蚕生态图像。为获取多样化数据,在环境光照强度 100~3 100 lx 范围内,选择上午、中午、下午 3 个时间段,系统性地改变拍摄距离与角度,最终获得 618 幅图像,覆盖了四龄和五龄桑蚕,其中密集群体占比 82%,完整记录四龄和五龄桑蚕的群体行为特征,特写样本占比 18%,精准捕捉个体形态学细节,如图 1 所示。



(a) 密集群体样本 (b) 特写样本

图 1 数据集样本示意图

Fig. 1 Illustration of dataset samples

1.2 数据预处理

采用规范化数据处理流程[16]对蚕体图像数据集进行系统构建,通过图像翻转、随机裁剪、调节亮度等数据增强方法,如图 2 所示,将数据集拓展到 2 845 幅图像,随后剔除模糊度过高及无目标个体的无效样本,保留 2 148 幅有效图像。数据集按比例 8∶1∶1 分层随机划分,其中训练集 1 718 幅图像、验证集 215 幅图像和测试集 215 幅图像,该划分方式有效保持了密集和特写样本在不同子集中均衡分布,为模型训练提供可靠的数据支撑。使用 LabelMe 以矢量多边形标注蚕体轮廓,并将 JSON 标注转换为 YOLO 格式的 TXT 文件,包含实例掩膜和边界框信息。

2 模型基础与改进

2.1 YOLO 11 模型整体架构

YOLO 算法的演进在 YOLO 11 的推出中达到



图 2　数据增强后的桑蚕图像

Fig. 2　Images of augmented silkworm

重要阶段，标志着实时目标检测技术新的进展。YOLO 11 作为 YOLO 系列的新一代模型，在保持高效单阶段推理的基础上，强化了多尺度融合与轻量化设计，在主干网络采用 C3k2 模块替代 C2F，通过动态卷积核重组增强特征捕获，颈部添加了 C2PSA 空间感知模块，融合通道与空间注意力，提升了多尺度目标表征，YOLO 11 的网络结构如图 3 所示。

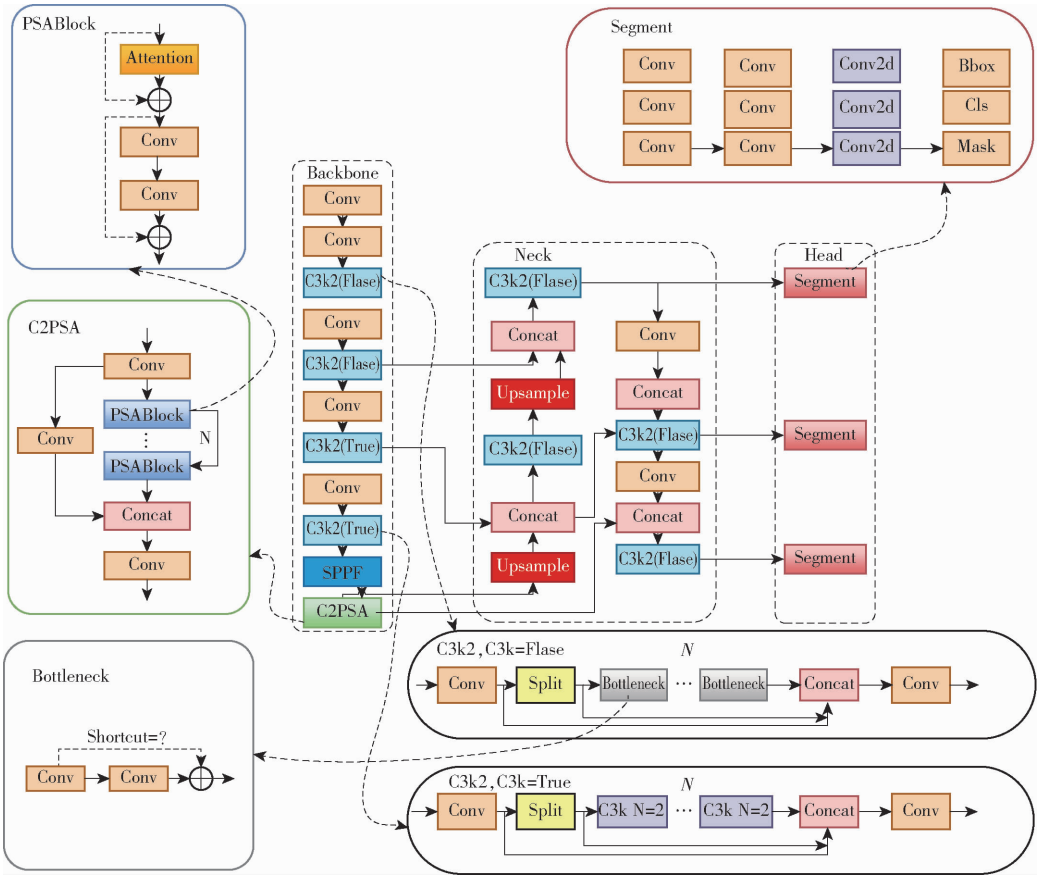


图 3　YOLO 11 结构图

Fig. 3　Structural diagram of YOLO 11

2.2　适于密集桑蚕实例分割的 DLC－YOLO 模型建构

为了提高性能并适应密集桑蚕分割，本文提出了 DLC－YOLO 模型，基于 YOLO 11n 进行了关键改进。首先，在主干网络中融合 DysnakeConv 与 C3k2 模块，取代原有结构，增强对复杂空间特征的捕捉能力，提升分割精度与模型适应性。其次，在颈部引入 DEA－Net 的内容引导注意力融合机制，优化特征拼接方式，实现更精准的多层次信息传递，加强密集实例的分割性能，这些改进有效提高了分割精度和模型适应性，满足密集场景下的分割需求。图 4 是改进后的 DLC－YOLO 结构图，其中用红框标出改进部分，方便更清晰地理解模型的改进之处。

2.3　耦合形态自适应的轴向可变形卷积模块优化密集桑蚕分割

针对密集桑蚕实例分割任务，YOLO 11n 中基础卷积结构对细长、弯曲且密集排列的生物形态缺乏针对性设计。传统 C3k2 模块采用标准 3×3 卷积构建残差块，其各向同性的方形感受野难以有效捕捉桑蚕长径比显著的形态特征，导致蚕体边缘分割模糊，尤其在头部与尾节转折处易出现断裂。此外，密集养殖场景下蚕体间存在大面积遮挡与粘连，原 Bottleneck 结构单一的空间特征提取模式对重叠区

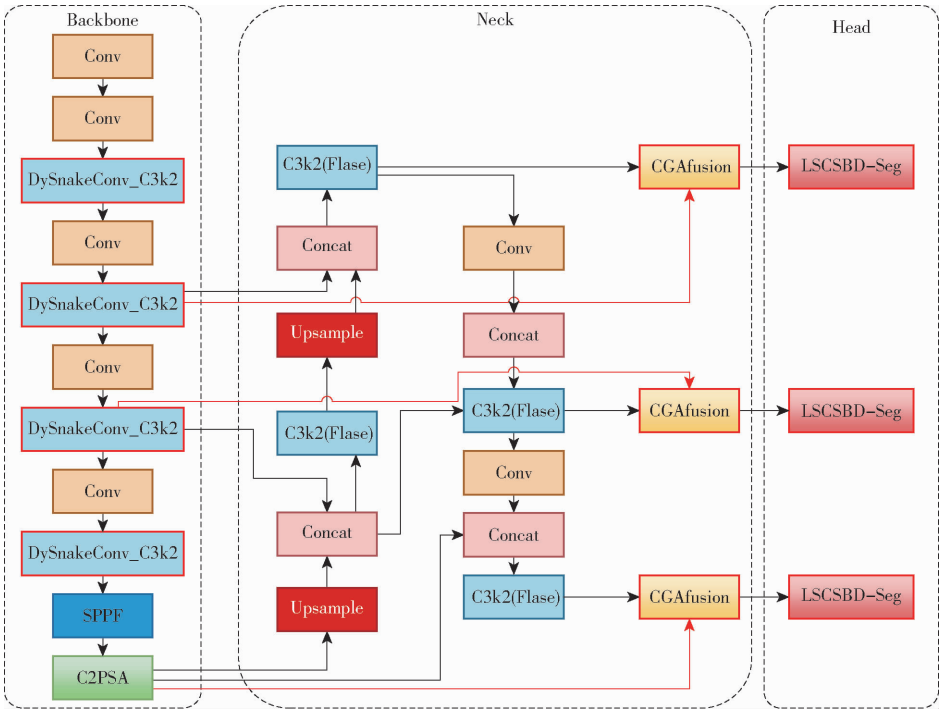


图 4 DLC－YOLO 结构图

Fig. 4 Structural diagram of DLC－YOLO

域缺乏动态适应能力,造成实例掩膜交叉污染。主干网络的特征融合机制也存在优化空间,传统残差连接难以聚合多方向上下文信息,在蚕体密集区域易丢失细小实例的语义特征。针对这一问题 DLC－YOLO 通过构建形态自适应的轴向可变形模块,从卷积算子层面实现形态驱动建模。DysnakeConv_C3k2 模块采用轴向动态卷积替代原 Bottleneck 结构中的标准卷积(图 5 中②部分),并行部署标准卷积、X 轴向 DSConv 与 Y 轴向 DSConv,其中 DSConv 通过可变形卷积核沿指定方向动态调整采样位置,在 X 轴方向,卷积核采样点从中心位置沿水平方向逐步偏移,最大延伸至 ±4 像素,同时通过垂直偏移量自适应弯曲轮廓。在 Y 轴方向,卷积核垂直延伸受限,但通过水平偏移累积主动贴合弧度。这种非对称形变将感受野从标准 3×3 卷积的正方形(长宽比 1∶1)转变为近似椭圆形,其水平跨度约 9 像素,垂直跨度约 1.3 像素,长宽比约为 7∶1,如图 6 所示。当处理弯曲蚕体时,X 轴向卷积核通过水平延伸覆盖蚕体全长以解决细长特性,Y 轴向则通过垂直形变贴合弯曲弧度以适配形态变化,二者协同捕获长条形目标的局部曲率特征。

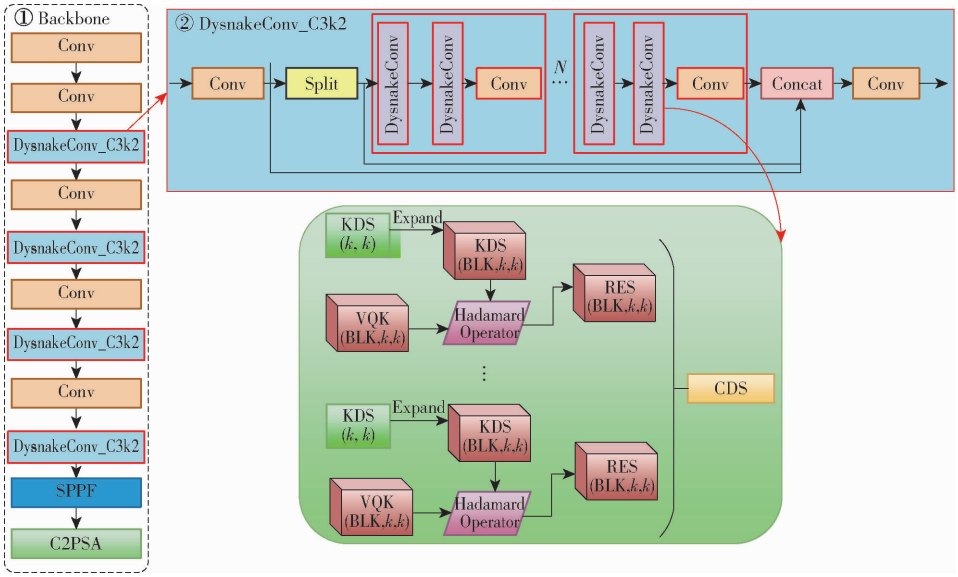


图 5 DysnakeConv_C3k2 结构图

Fig. 5 Structural diagram of DysnakeConv_C3k2

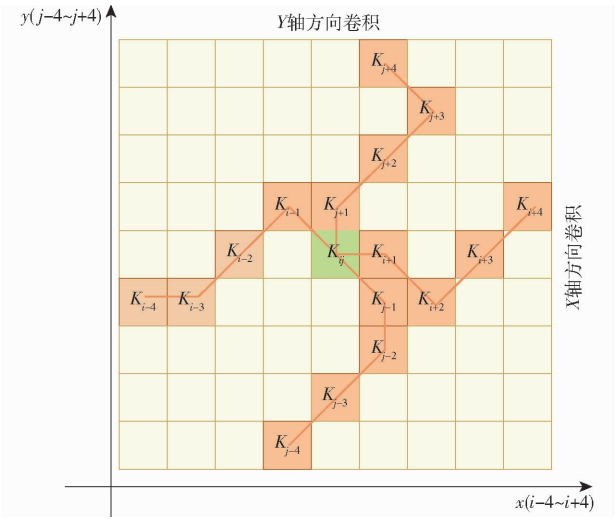


图 6　DysnakeConv_C3k2 可选择的感受野示意图

Fig. 6　Schematic of adaptive receptive field deformation in DysnakeConv_C3k2

在 X 轴方向的变化公式为

$$K_{i\pm c_i}=\begin{cases}(x_{i+c_i},y_{i+c_i})=\left(x_i+c_i,y_i+\sum_{i}^{i+c_i}\Delta y\right)\\(x_{i-c_i},y_{i-c_i})=\left(x_i-c_i,y_i+\sum_{i-c_i}^{i}\Delta y\right)\end{cases}\tag{1}$$

式中　c_i——与中心网格K_i的水平距离

$K_{i\pm c_i}$——卷积核中距离中心网格K_i水平偏移c_i个单位的网格位置坐标

i——中心网格K_i的索引

(x_i,y_i)——中心网格K_i的坐标

$\sum_{i}^{i+c_i}\Delta y$——从中心网格索引 i 到索引 $i+c_i$ 的网格位置在 Y 轴方向上累积的偏移量之和

$\sum_{i-c_i}^{i}\Delta y$——从索引 $i-c_i$ 到中心网格索引 i 的网格位置在 Y 轴方向上累积的偏移量之和

Δy——网格位置在 Y 轴方向的动态偏移量

在 Y 轴方向变化公式为

$$K_{j\pm c_j}=\begin{cases}(x_{j+c_j},y_{j+c_j})=\left(x_j+\sum_{j}^{j+c_j}\Delta x,y_j+c_j\right)\\(x_{j-c_j},y_{j-c_j})=\left(x_j+\sum_{j-c_j}^{j}\Delta x,y_j-c_j\right)\end{cases}\tag{2}$$

式中　$K_{j\pm c_j}$——卷积核中距离中心网格K_j垂直偏移c_j个单位的网格位置坐标

c_j——与中心网格K_j的垂直距离

j——中心网格K_j的索引

(x_j,y_j)——中心网格K_j的坐标

Δx——网格位置在 X 轴方向的动态偏移量

$\sum_{j}^{j+c_j}\Delta x$——从中心网格索引 j 到索引 $j+c_j$ 的网格位置在 X 轴方向上累积的偏移量之和

$\sum_{j-c_j}^{j}\Delta x$——从索引 $j-c_j$ 到中心网格索引 j 的网格位置在 X 轴方向上累积的偏移量之和

从中心位置K_i开始，远离中心网格的位置取决于前一个网格的位置，即K_{i+1}相对于K_i增加了偏移量 $\Delta=\{\delta\mid\delta\in[-1,1]\}$。

由于偏移量 Δ 通常是小数，然而坐标通常是整数形式，因此采用双线性插值，表示为

$$K=\sum_{K'}B(K',K)K'\tag{3}$$

式中　K——需要插值的位置（由式(1)或式(2)计算得到的小数坐标）

K'——K 周围 4 个最近的整数坐标像素位置

B——插值核函数

$B(K',K)$——插值核函数 B 在位置 K' 对于目标位置 K 的权重

$\sum_{K'}$ 表示对 K 周围 4 个整数坐标位置 K' 求和。

改进后的 DysnakeConv_C3k2 结构通过 X 轴向和 Y 轴向卷积核协同作用，使单个卷积核的有效感受野[17]长径比提升至 7∶1，较原 3×3 卷积提升 4 倍，配合双线性插值处理小数偏移（式(3)）有效增强对蚕体头尾关键点的定位精度。三向特征图经拼接后通过 1×1 卷积进行通道压缩，在保留多方向信息的同时控制参数增长，最终模块参数量仅增加 7.3×10^4 个。在主干网络中替换 4 个关键位置的 C3k2 模块（图 5 中①部分），以此形成层次化特征增强机制。浅层网络中的 DysnakeConv 侧重捕获蚕体轮廓的方向性纹理，其动态卷积核在蚕体区域自适应拉长为椭圆；深层模块专注于解决密集遮挡问题，通过轴向卷积的导向性特征提取，在重叠区域生成差异化的激活响应，从而提高了密集桑蚕场景下多尺度实例的形变适应性与遮挡穿透力。图 5 红框标记部分为替换部分以及替换后的 DysnakeConv_C3k2 的具体结构。

2.4　构建内容引导的注意力特征融合增强框架

在 YOLO 11 的密集桑蚕实例分割任务中，其核心瓶颈在于特征融合机制对密集小目标的空间敏感性和语义区分能力不足。YOLO 11 的颈部主要依赖传统的特征金字塔网络与路径聚合网络结构，通过简单的上采样和拼接实现多尺度特征融合。然而，桑蚕作为高密度、形态相似且尺寸微小的目标，

其实例边界在重叠或接触区域容易混淆,传统融合策略难以有效捕捉细粒度特征差异。存在问题如下:在特征融合过程中缺乏对通道和空间维度的动态注意力引导,导致背景噪声与前景目标在密集区域竞争性激活;固定权重的特征相加难以适应不同层级特征的语义分布差异,尤其在处理桑蚕这类具有相似纹理和颜色的目标时,易丢失局部关键信息;原模型对像素级位置敏感性的建模不足,无法在实例分割任务中精确区分相邻个体边缘。

针对上述问题,融入 DEA - Net 的 Content-guided attention fusion(CGAfusion)模块对 YOLO 11 颈部进行改进,通过多层次注意力协同机制实现更精细的特征融合。改进后的模型在密集桑蚕分割任务中展现出 3 方面核心优势:①通过内容引导的动态权重分配强化了目标区域的响应,抑制背景干扰。②多分支注意力机制协同挖掘通道、空间和像素维度的互补信息,提升特征表征的判别力。③自适应融合策略增强了对微小目标的边缘保持能力。改进步骤为:在 CGAfusion 模块中,首先通过初始特征相加($x+y$)保留原始特征分布。随后并行执行通道注意力(CA)和空间注意力(SA)计算。通道注意力采用压缩-激励结构,通过全局平均池化捕获通道间统计依赖关系,利用两层全连接层生成通道权重,使网络聚焦于桑蚕特有的颜色和纹理特征通道。空间注意力则通过空间金字塔结构提取多尺度上下文信息,强化密集区域的空间定位能力。两者输出的注意力图通过相加操作实现初步融合,此时形成的混合注意力已具备通道-空间联合感知能力。接着引入像素注意力(PA)模块,其独特之处在于将混合注意力作为先验信息输入,通过逐像素的仿射变换动态调整权重分布,实现特征选择的自适应性。最终通过残差连接和 1×1 卷积完成特征重整,该过程在保留原始特征信息的同时,使融合过程更关注桑蚕个体间的细微差异。

在特征金字塔的关键节点引入 CGAfusion 模块,取代原有的简单拼接操作。首先,在 P3 层融合时,将来自主干的浅层特征与特征金字塔[18]输出的深层特征输入 CGAfusion,利用浅层的高分辨率细节和深层的语义信息进行互补增强。其次,DysnakeConv_C3k2 模块中的残差结构为注意力机制提供梯度传播通路,避免深层网络训练中梯度衰减问题。最后,进入 LSCSBD - Seg 分割检测头的多尺度特征输入经过 CGAfusion 优化后,能够更好地平衡不同尺度的特征贡献,在 P5 层融合时,高层语义信息通过像素注意力精准定位桑蚕群体分布,而低层特征则保留个体边缘细节,图 7 为 CGAfusion 的结构图。

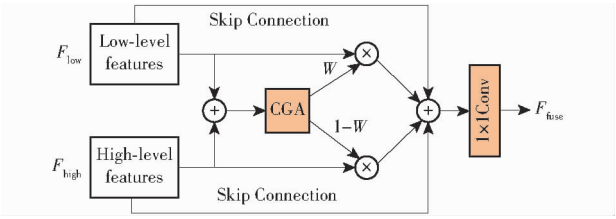


图 7 CGAfusion 结构图

Fig. 7 Structural diagram of CGAfusion

2.5 设计动态原型与掩膜优化检测头

YOLO 11n 采用的标准卷积运算[19]虽然通过参数共享机制降低了计算复杂度,但其固有的空间同质性特征提取模式难以应对密集场景下目标的形态多样性,特别是在处理桑蚕这类高密度重叠的实例时,传统卷积核无法有效区分相邻目标的边缘特征,导致分割精度显著下降。针对 YOLO 11n 网络在密集实例分割任务中存在的不足,本文作出有效改进,提出了 LSCSBD - Seg。

LSCSBD - Seg 在特征生成机制上引入动态原型生成模块[20],改进了传统卷积的静态参数共享特性。该模块采用双路径特征解耦架构,主路径通过 3×3 深度可分离卷积提取目标的几何结构特征,辅路径利用通道注意力机制捕捉实例的语义差异性,最终通过门控融合生成具有实例特异性的原型矩阵。相较于 YOLO 11n 的 128 维原型特征,将特征维度扩展至 256,使原型空间的信息容量有效提升。

针对掩膜生成的质量问题,LSCSBD - Seg 构建了层次化掩膜系数学习框架。传统方法依赖单层 3×3 卷积提取低层次特征,导致边缘细节丢失。LSCSBD - Seg 设计了三级级联的系数生成网络,首层 1×1 卷积进行通道压缩,将输入维度降至 1/4;中层 3×3 扩张卷积通过空洞率为 2 的核结构扩大感受野;末端 1×1 卷积实现跨通道特征聚合。该架构引入动态核选择机制,根据特征图的梯度分布自适应选择 3×3 或 5×5 卷积核,使密集目标的边缘精度有效提升。

LSCSBD - Seg 的另一个改进是多尺度特征融合。YOLO 11 虽然在处理不同尺度的特征图上有着较好的效果,但在分割任务中,多尺度特征融合的能力较弱。为了更好地应对密集目标的挑战,LSCSBD - Seg 在多个尺度上进行特征提取,并通过卷积操作进一步增强特征的空间信息。每个尺度的卷积核大小不同,这使得模型能够更加全面地理解不同尺寸目标的细节,从而提升了分割效果。通过这种多尺度的特征融合,LSCSBD - Seg 能够在处理不同尺度的目标时,保持高精度的分割性能。

计算效率优化层面,LSCSBD - Seg 引入了 Conv_GN 轻量化模块。该模块将组归一化与动态核剪枝

相结合，在检测分支采用 Ghost 卷积替代标准卷积，分割分支通过分析梯度幅值动态冻结 30% 的卷积核。配合隐藏层通道控制参数的引入，降低了检测头模型大小。

在任务协同优化机制上，提出了基于双模态特征同步学习框架的联合训练范式。通过构建并行的特征提取通道，使原型语义表征与目标定位特征实现空间-通道维度的解耦计算，解决了传统级联架构的时序依赖性瓶颈。在模型前向传播过程中，优化的检测子网络与分割特征生成器以互补式信息交互机制实现协同推理，其中检测分支输出空间置信度分布，分割分支通过三级级联系数生成网络与跨模态注意力门控生成对桑蚕这类细长形态敏感的特征映射，二者通过跨模态注意力门控实现权重共享。训练阶段设计监督矩阵，将目标定位置信度、实例轮廓系数和原型语义嵌入构成联合优化空间，利用动态梯度加权算法平衡不同任务的损失贡献度。推理阶段创新引入特征空间对齐策略，通过跨通道特征重组技术，实现检测框置信度与掩膜形态参数的几何一致性匹配。

通过原型生成增强、掩膜系数优化、跨尺度交互强化、轻量化计算重构和多任务协同训练进行改进，使 LSCSBD－Seg 在密集实例分割任务中的性能有效提升，此改进的详细结构如图 8 所示。

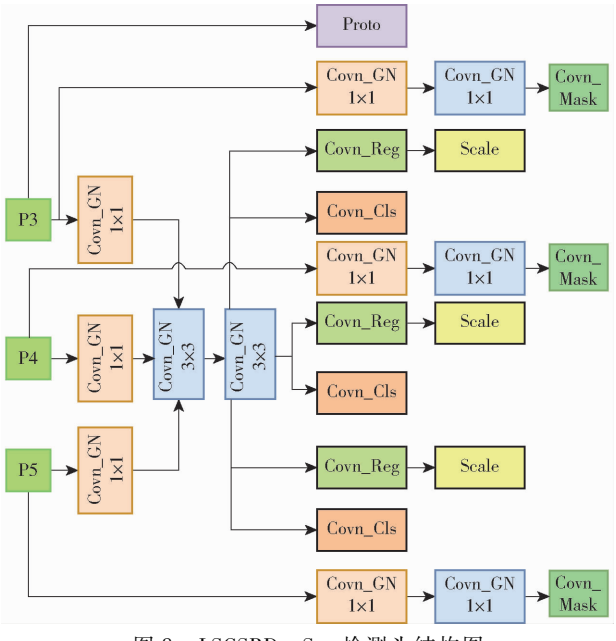


图 8　LSCSBD－Seg 检测头结构图

Fig. 8　Structural diagram of LSCSBD－Seg detection head

3　实验

3.1　实验环境与评价指标

3.1.1　实验环境

本研究在云计算服务器上进行模型训练，硬件配置包括搭载 24 GB 显存的 NVIDIA RTX 4090 GPU、基于 Intel Xeon Platinum 8352V 处理器的虚拟化计算单元以及 120 GB 大容量内存，为实验提供了稳定的计算环境保障，实验基于 Python 3.8 编程环境与 CUDA 11.8 计算架构，实验模型的各项超参数配置如下：输入图像采用 640 像素 ×640 像素的标准分辨率，基础学习速率初始值设为 0.01，采用随机梯度下降优化算法。综合模型规模与训练效率的平衡，训练过程设置每批次处理 16 个样本，并行数据加载线程数为 16，单帧最大检测目标数（max_det）限制为 1 000，置信度判定阈值为 0.25，整体训练轮次共计 150 次。

3.1.2　评价指标

采用目标检测与实例分割任务的标准评估指标对模型性能进行评估。针对 2 个任务分别量化了精确率（P）、召回率（R）、IoU 阈值为 0.5 时的平均精度均值（mAP50）、IoU 阈值为 0.75 时的平均精度均值（mAP75），以及 IoU 阈值在 0.5～0.95 范围内的平均精度均值（mAP50:95），同时通过参数量分析表征模型的计算效率，该综合性评估方案遵循计算机视觉研究的通用基准，确保结果的可比性与方法学的规范性。

3.2　消融实验

为验证本文提出的模块对桑蚕检测任务的有效性，在与 3.1 节相同训练参数下，通过控制变量法[21]进行消融实验，量化分析各改进模块的贡献。实验指标采用 mAP50、mAP50:95、精确率、召回率、模型内存占用量以及参数量，如表 1 所示，括号内的 B 表示目标检测结果，M 表示实例分割结果。

本消融实验通过逐步验证 DysnakeConv 与 C3k2 融合模块、CGAfusion 模块以及检测头 LSCSBD 对 YOLO 11 模型的改进效果，展示了各模块对目标检测与实例分割性能的差异化增益特性。

实验 1 作为基准测试组，采用原始 YOLO 11 模型获取实验结果，模型内存占用量为 5.7 MB，为后续改进提供参照。

实验 2 仅引入 DysnakeConv_C3k2 模块，模型内存占用量增至 6.4 MB，该模块通过可变形卷积增强对弯曲轮廓的建模能力，使目标检测 mAP50:95 提升 2.2 个百分点，实例分割的精确率与召回率均提高 1.4 个百分点，有效缓解了传统卷积核在细长目标上的适应性局限。

实验 3 仅采用 CGAfusion 模块，内存占用量为 6.6 MB，通过内容引导注意力机制优化特征融合，在密集场景下有效缓解了特征混淆和小目标漏检问题，实现了目标检测召回率提升 2 个百分点、

表 1 不同模块消融实验结果

Tab. 1 Ablative experimental results of different modules

实验序号	Dysnakeconv_C3k2	CGAfusion	LSCSBD－Seg	P/%	R/%	mAP50/%	mAP50:95/%	模型内存占用量/MB	参数量/个
1				89.8(B) 86.3(M)	81.9(B) 76.2(M)	90.4(B) 83.5(M)	65.4(B) 47.2(M)	5.7	2 842 803
2	√			90.6(B) 87.7(M)	83.0(B) 77.6(M)	91.2(B) 84.8(M)	67.6(B) 48.5(M)	6.4	3 135 171
3		√		90.9(B) 88.5(M)	83.9(B) 78.0(M)	91.6(B) 85.2(M)	67.8(B) 48.6(M)	6.6	3 287 100
4			√	90.1(B) 88.1(M)	82.7(B) 76.7(M)	91.1(B) 84.5(M)	67.1(B) 48.1(M)	5.2	2 601 780
5	√	√		91.2(B) 88.9(M)	84.1(B) 78.5(M)	91.8(B) 85.7(M)	68.3(B) 48.8(M)	7.0	3 478 148
6		√	√	91.3(B) 88.5(M)	84.5(B) 78.1(M)	91.7(B) 85.7(M)	68.7(B) 49.5(M)	5.9	2 952 189
7	√		√	90.8(B) 87.9(M)	83.5(B) 77.3(M)	91.4(B) 84.7(M)	67.7(B) 48.2(M)	5.7	2 792 844
8(本文模型)	√	√	√	91.4(B) 89.0(M)	85.4(B) 79.3(M)	92.7(B) 86.5(M)	70.5(B) 50.1(M)	6.6	3 236 861

注:√表示使用该模块。

mAP50:95 增长 2.4 个百分点,以及实例分割精确率增长 2.2 个百分点。

实验 4 仅引入 LSCSBD－Seg 检测头,内存占用量降至 5.2 MB,该检测头通过局部语义约束与边界细化缓解掩膜重叠问题,使目标检测 mAP50:95 提升 1.7 个百分点,实例分割精确率与 mAP50 分别提升 1.8、1.0 个百分点。

进一步的双模块组合实验(实验 5:DysnakeConv_C3k2 + CGAfusion;实验 6:CGAfusion + LSCSBD－Seg;实验 7:DysnakeConv_C3k2 + LSCSBD－Seg)表明模块间存在增益互补性。其中,实验 5 目标检测 mAP50:95 提升 2.9 个百分点,实例分割精确率提升 2.6 个百分点;实验 6 目标检测召回率和 mAP50:95 分别提升 2.6、3.3 个百分点;实验 7 目标检测 mAP50:95 和实例分割精确率分别提升 2.3、1.6 个百分点。

实验 8 集成 3 个模块,内存占用量为 6.6 MB,较基准仅增 15.8%,获得最优性能,目标检测精确率提升 1.6 个百分点,召回率提升 3.5 个百分点,mAP50 与 mAP50:95 分别增长 2.3、5.1 个百分点;实例分割精确率、召回率、mAP50 及 mAP50:95 分别提升 2.7、3.1、3.0、2.9 个百分点。各模块功能互补,DysnakeConv_C3k2 增强底层特征表达,CGAfusion 强化中高层语义关联,而 LSCSBD－Seg 提升端到端优化效率,此外,mAP50:95 显著提高表明多模块联合有效提升了模型鲁棒性。

图 9 为消融实验结果可视化图,此 3 个模块协同后达到了最好效果,其中 A 表示 Dysnakeconv_C3k2,B 表示 CGAfusion,C 表示 LSCSBD－Seg。

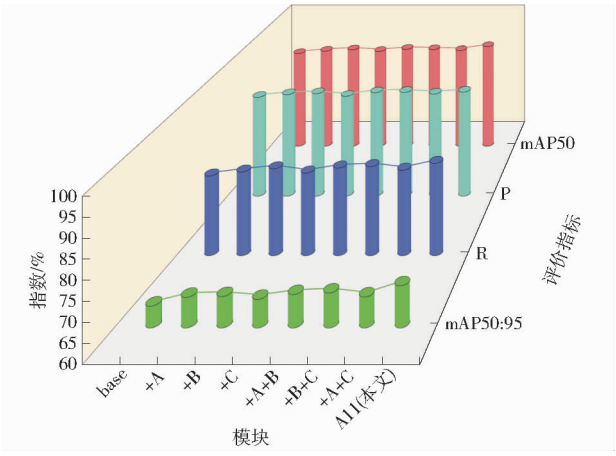


图 9 消融实验结果可视化图

Fig. 9 Visualization of ablative experiment results

3.3 对比实验

为评估 DLC－YOLO 在密集桑蚕检测与分割任务中的性能,选择当前主流的实例分割模型作为基准对比[22]方法,包括 YOLO 系列中的 YOLO v8－Seg、YOLO 11－Seg 和 YOLO 12－Seg,经典两阶段模型 Mask R－CNN,轻量化模型 Hyper－YOLO－Seg,实时实例分割模型 YOLACT 以及 Transformer 模型 RTDETR－r18 和 RTDETR－r50。所有对比模型均在同一训练集上复现并采用相同超参数配置。实验以 mAP50、mAP75、mAP50:95、模型内存占用量和检测速度等指标作为核心评估标准,同时引入精确率与召回率进行细粒度性能分析,各指标结果如表 2 所示。

表 2　不同模型在密集桑蚕场景中的检测性能

Tab. 2　Detection results of various models in dense silkworm scenarios

模型	mAP50/%	mAP75/%	mAP50:95/%	P/%	R/%	模型内存占用量/MB	检测速度/(f·s^{-1})
Mask R－CNN	89.3	76.4	70.6			491.5	9.1
Hyper－YOLO－Seg	91.4	75.9	67.6	87.6	77.8	7.8	33.8
YOLO v8－Seg	91.5	75.1	67.2	87.9	78.1	6.5	37.3
YOLO v11－Seg	90.4	73.0	65.4	86.3	76.2	5.7	35.8
YOLO v12－Seg	90.6	73.5	65.5	87.2	76.1	5.8	32.6
YOLACT	80.9	59.3	51.4				
RTDETR－r18	84.5	69.6	60.8	77.2	74.8	61.3	28.3
RTDETR－r50	87.8	71.2	62.3	81.5	77.6	96.3	31.6
本文模型	92.7	79.3	70.5	89.1	79.3	6.6	36.2

在密集桑蚕检测任务中，DLC－YOLO 以 mAP50 为 92.7%、mAP75 为 79.3% 和 mAP50:95 为 70.5% 的综合性能领先其他模型。相比模型内存占用量为 5.7 MB、检测速度为 35.8 f/s 的基线模型 YOLO 11－Seg，DLC－YOLO 在保持内存占用量 6.6 MB 和实时检测速度 36.2 f/s 的同时，实现 mAP50 提升 2.3 个百分点、mAP75 提升 6.3 个百分点、mAP50:95 提升 5.1 个百分点，其精确率 89.1% 和召回率 79.3% 分别优化 2.8、3.1 个百分点，验证了 DLC－YOLO 的有效性。经典两阶段模型 Mask R－CNN[23] 虽 mAP50:95 达到 70.6%，但其模型内存占用量 491.5 MB，远超 DLC－YOLO 的 6.6 MB，且检测速度为 9.1 f/s 远低于 DLC－YOLO 的 36.2 f/s，难以适配农业场景的轻量化部署需求；实时模型 YOLACT[24] 因单阶段设计局限，综合指标 mAP50:95 仅为 51.4%，与 DLC－YOLO 差距达 19.1 个百分点。与 YOLO 系列的最新版本相比，DLC－YOLO 在模型内存占用量仅 6.6 MB 和检测速度 36.2 f/s 的配置下，实现了多项性能提升。相较于 YOLO 12－Seg，DLC－YOLO 的 mAP75 和 mAP50:95 两个指标分别高出 5.8、5.0 个百分点；相比 YOLO v8－Seg，DLC－YOLO 的 mAP75 提升 4.2 个百分点，mAP50:95 提升 3.3 个百分点；而与 Hyper－YOLO－Seg[25] 相比，DLC－YOLO 在两项精度指标上也分别有 3.4 个百分点和 2.9 个百分点的优势。实时 Transformer 模型 RTDETR－r18 在精度、内存占用量和速度方面均落后于 DLC－YOLO；RTDETR－r50 虽然检测速度达到 31.6 f/s，但其模型内存占用量高达 96.3 MB，是 DLC－YOLO 的 14.6 倍，mAP50:95 也比 DLC－YOLO 低 8.2 个百分点。实验结果证明，DLC－YOLO 在保持模型内存占用量 6.6 MB 且检测速度满足实时检测的基础上，实现检测精度、分割质量和计算效率的协同增长，为农业场景实例分割任务的嵌入式实施提供了解决方案。

为直观展示模型改进效果，本文选取以蚕框和纸盒为背景的 2 幅密集桑蚕样本图像进行可视化对比，如图 10 所示，蓝色蚕体为正确检测，白色蚕体为漏检，所有检测图按同比例排版，清晰展现各模型在密集桑蚕中识别计数的效果。可视化图顶部呈现原始图像，原始图像 1 包含密集分布的 449 只桑蚕，原始图像 2 包含 468 只桑蚕，中部与底部依次展示 DLC－YOLO、YOLO v8－Seg、Hyper－YOLO－Seg、YOLO 11－Seg、Mask R－CNN 和 RTDETR－r50 这 6 种模型对 2 个样本的检测效果。实验数据如表 3 所示，在纸盒背景样本中，DLC－YOLO 以 424 只正确识别数处于领先，较 YOLO v8－Seg 的漏检率降低 2.89 个百分点，较 Mask R－CNN 的漏检率降低 6.45 个百分点；在蚕

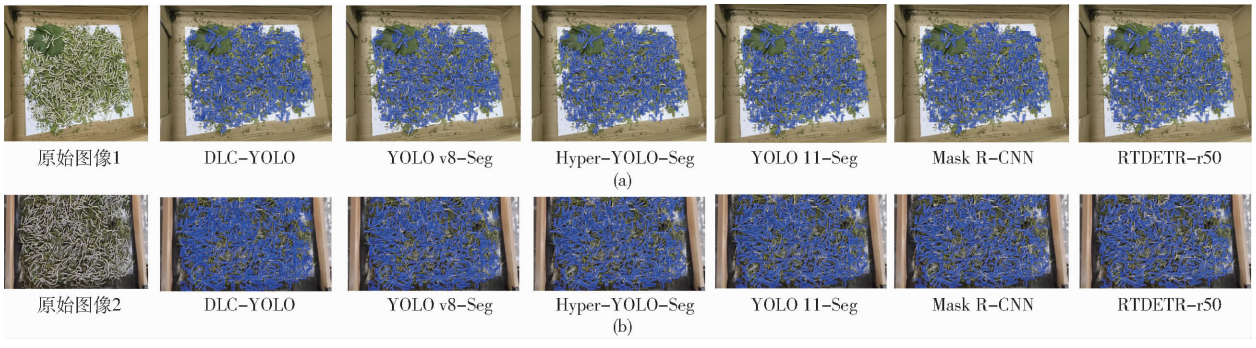


图 10　各模型识别效果图

Fig. 10　Recognition effects of improved and baseline models

表 3 各模型识别结果详细数据

Tab. 3 Detailed data of recognition results for each model

模型	纸盒样本(N_1 =449)			蚕框样本(N_2 =468)		
	检测数量	漏检数	漏检率/%	检测数量	漏检数	漏检率/%
YOLO v8 - Seg	411	38	8.46	428	40	8.55
Hyper - YOLO - Seg	403	46	10.24	401	67	14.32
YOLO 11 - Seg	406	43	9.58	431	37	7.91
Mask R - CNN	395	54	12.02	376	92	19.66
RTDETR - r50	374	75	16.70	355	113	24.15
DLC - YOLO	424	25	5.57	447	21	4.49

注:N_1、N_2分别为纸盒和蚕框样本中蚕的实际数量。

框背景样本中,DLC - YOLO 识别的桑蚕数量为 447,漏检率为 4.49%,比 YOLO 11 - Seg 的漏检率低 3.42 个百分点,且大幅领先 RTDETR - r50 的漏检率(19.66%)。结果表明,DLC - YOLO 对密集目标的适应性更强。

为验证 DLC - YOLO 算法在桑蚕计数任务中的实时性与实用性,本文基于 Python 3.9 开发环境,采用 PyQt5 框架设计桑蚕智能计数系统(图 11),并将其部署于桑蚕自动化设备的关键环节进行测试。该系统利用安装于蚕座进入桑叶投喂部分前的工业摄像头进行数据输入,型号为 RER - USBGS5MP01 - XCQ(图 12),通过 USB 3.0 协议连接便携式计算机,实时采集分辨率 2 592 像素 ×1 944 像素的视频流。在实际养殖密度下,蚕座以速度 0.2 m/s 通过视野,蚕座在传送带上的运行图如图 13 所示,之后系统对平均每帧含(270 ± 10)只、最大重叠率达 25% 的桑蚕进行实时目标检测与计数。其中重叠率 θ 定义为单帧图像中所有桑蚕最小外接矩形的相互重叠面积占蚕体总投影面积的百分比,计算公式为

$$\theta = \frac{\sum_{i \neq j} \mathrm{IOU}(B_i, B_j)\, A_{box}(B_i)}{\sum_{k} A_{box}(B_k)} \times 100\% \qquad (4)$$

式中 B_i、B_j——任意两蚕体的检测框($i \neq j$)

IOU( )——交并比函数

A_{box}——检测框面积

B_k——第 k 个蚕体的检测框

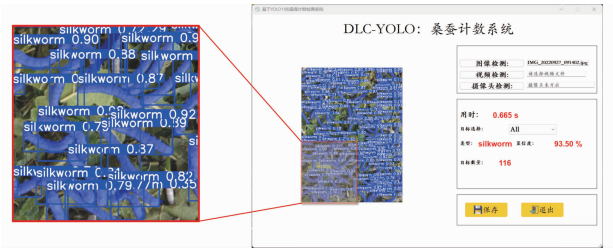


图 11 桑蚕计数系统

Fig. 11 Silkworm counting system



图 12 摄像头部署图

Fig. 12 Camera deployment diagram

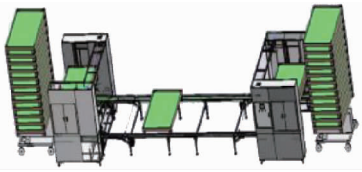


图 13 蚕坐输送图

Fig. 13 Operation of silkworm tray conveying

算法在输入分辨率 2 592 像素 ×1 944 像素下平均处理帧率达 36.2 f/s,满足实时处理要求,并且计数准确率为 93.5%,满足自动化设备需求且减少了人力,增加了计数效率。为进一步验证算法在实际环境下的适用性,在光照鲁棒性方面,本研究在 100 ~ 3 100 lx 范围内设置了 7 个均匀分布的光照等级(100、600、1 100、1 600、2 100、2 600、3 100 lx)进行系统测试,定量分析表明,模型在该区间内保持了较高的稳定性,在光照强度高于 500 lx 时模型表现最优,模型综合识别计数准确率达到 93.5%,mAP50 维持在 92.7% 左右,而在低于 200 lx 的弱光条件下,因图像噪声增加及对比度下降,模型综合识别计数准确率约为 90.8%,mAP50 降至约 88.3%,部分蚕体边缘细节提取能力略有减弱,但整体仍保持较好的可用性。软件实时解析图像,界面显示并保存带有右下角桑蚕数量标识和左下角检测时间戳的结果图像,充分验证了算法在实际场景下的鲁棒性,为桑蚕养殖场的数字化精细管理提供了有效的技术支撑。

4 讨论

本文提出了一种基于 YOLO 11n 改进的密集桑蚕实例分割模型 DLC - YOLO,针对桑蚕细长弯曲体态、多尺度特征融合效率低下及边缘粘连分割精度不足,从主干网络、颈部结构和分割检测头 3 方面进行了系统性优化。实验表明,DLC - YOLO 在检测精度、分割质量和模型内存占用量间实现了良好平衡,尤其在复杂养殖环境下的形态适应性、多尺度特征融合和边缘粘连处理方面表现较好,优于传统方法。

相较于原始 YOLO 11n，首先在主干网络中引入 DysnakeConv_C3k2 融合结构，通过动态可变形卷积核沿桑蚕细长体态进行自适应形变特征提取。这种做法避免了传统方形卷积核的形态限制，使得主干网络对弯曲桑蚕的轮廓跟踪能力得到提升，有效增强了对细长目标的特征表达。其次，颈部网络采用内容引导注意力融合机制（CGAfusion），重构多尺度特征金字塔的交互方式。通过通道注意力与空间注意力的协同优化，显著提升了密集区域的特征判别力，使小目标漏检率降到 4.49%。最后，设计的 LSCSBD－Seg 分割检测头通过局部语义约束和边界细化模块，减少了实例掩膜重叠误差，分割精度 mAP50 提升 3.0 个百分点，解决了传统方法因边缘粘连导致的计数误差累积问题。

尽管 DLC－YOLO 在密集场景中表现较好，但仍存在改进空间。例如，DysnakeConv 模块虽增强了对弯曲形态的捕捉能力，但在极端光照条件下的轮廓提取稳定性有待提升；CGAfusion 机制虽优化了特征融合效率，但对超密集场景的遮挡穿透能力仍受限于二维图像的空间信息损失。

未来研究可以引入光照不变性特征增强模块，结合多光谱数据融合提升复杂光照下的鲁棒性，探索三维点云或时序信息辅助分割，通过空间-时间联合建模解决超高密度遮挡问题；开发动态模型压缩策略，结合神经架构搜索与量化技术，在维持精度的前提下将模型规模压缩至 3 MB 以内，适配更多低成本嵌入式硬件。这些改进将进一步推动密集桑蚕计数技术从实验室验证向规模化农业应用的转化，为智慧蚕业的全面数字化提供有效技术支撑。

5　结论

（1）本文致力于提升密集桑蚕养殖场景下的自动化计数精度，突破传统方法在个体形态多变、密集遮挡和小目标检测中的局限性。为此，提出基于改进 YOLO 11n 的 DLC－YOLO 实例分割模型，通过 3 个改进实现检测精度与分割质量的协同优化。在主干网络设计动态蛇形卷积与跨阶段融合结构，采用可变形卷积核沿蚕体轮廓进行自适应形变，增强对细长弯曲形态的特征提取能力；颈部网络创新性地引入内容引导注意力融合机制，重构多尺度特征金字塔的跨层交互方式，提升密集小目标的特征表达能力；引入 LSCSBD－Seg 分割检测头通过原型生成与边界细化模块，有效解决个体边缘粘连问题。在自建的 2 148 幅图像密集桑蚕数据集上，DLC－YOLO 展现出较大优势，检测任务 mAP50 达 92.7%、mAP50:95 较原始模型提升 5.1 个百分点，分割任务 mAP50 达到 86.5%，较原始 YOLO 11n 提升 3.0 个百分点。结果显示，该模型在内存占用量仅 6.6 MB 时，其 mAP50 指标超越 Mask R－CNN 等主流模型，验证该模型的有效性。

（2）尽管 DLC－YOLO 在密集桑蚕检测与分割任务中表现出色，且在光照强度 100～3 100 lx 范围内能保持稳定的分割计数性能，但在极端光照条件或超高密度场景下的检测稳定性仍有优化空间。此外，模型对病态蚕体如体表异常皱缩或运动姿态畸变的适应性，仍需进一步验证。未来计划将模型部署至嵌入式蚕房监测终端，并探索多模态数据融合技术，进一步提升复杂养殖环境下的鲁棒性。

参　考　文　献

[1] 吴春标. 砷在桑-蚕生物链的富集迁移及对家蚕的影响[D]. 海口：海南大学，2024.
WU Chunbiao. The enrichment and migration of arsenic inmulberry-silkworm biological chain and its effect on silkworm[D]. Haikou: Hainan University, 2024. (in Chinese)

[2] 田文晗，何观平，刘玉增，等. 适配骨组织工程的蚕丝增强聚己内酯复合材料研究[J]. 中国修复重建外科杂志，2025，38(9)：1－7.
TIAN Wenhan, HE Guanping, LIU Yuzeng, et al. Study on silk reinforced polycaprolactone composites for bone tissue engineering[J]. Chinese Journal of Reparative and Reconstructive Surgery, 2025,38(9):1－7. (in Chinese)

[3] 虞崇江，何骥，黄梅梅，等. 粤桂协作背景下广西蚕桑茧丝绸产业高质量发展战略研究[J]. 广西蚕业，2024，61(2)：59－64.
YU Chongjiang, HE Ji, HUANG Meimei, et al. Research on high-quality development strategy of Guangxi sericulture and silk industry under Guangdong－Guangxi collaboration[J]. Guangxi Sericulture,2024,61(2):59－64. (in Chinese)

[4] 廖国森. 乡村振兴战略下桑蚕养殖业发展优势分析[J]. 南方农业，2021，15(12)：193－194.
LIAO Guosen. Development advantage analysis of mulberry sericulture under the rural revitalization strategy[J]. South China Agriculture, 2021, 15(12): 193－194. (in Chinese)

[5] 王金鹏，何萌，甄乾广，等. 基于改进 COF－YOLO v8n 的油茶果静态与动态检测计数方法[J]. 农业机械学报，2024，55(4)：193－203.
WANG Jinpeng, HE Meng, ZHEN Qianguang, et al. Camellia oleifera fruirt static and dynamic detection counting based on improved COF－YOLO v8n[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(4): 193－203. (in Chinese)

[6] 刘凯越,吴建军,李智慧,等. YOLO－Vega:一种面向复杂环境的轻量级番茄检测模型[J/OL]. 计算机工程与应用,2025,1－15. https://link.cnki.net/urlid/11.2127.TP.20250723.1424.019.
LIU Kaiyue, WU Jianjun, LI Zhihui, et al. YOLO－Vega: a lightweight tomato-detection model for complex environments[J/OL]. Computer Engineering and Applications, 2025: 1－15. https://link.cnki.net/urlid/11.2127.TP.20250723.1424.019. (in Chinese)

[7] 刘庆华,杨欣仪,接浩,等. 基于融合 GhostNetV2 的 YOLO v7 水稻籽粒检测[J]. 农业机械学报,2023,54(12):253－260,299.
LIU Qinghua, YANG Xinyi, JIE Hao, et al. Rice grain detection based on YOLO v7 fusion of GhostNetV2[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(12): 253－260,299. (in Chinese)

[8] 何锐敏, 郑可锋, 尉钦洋,等. 基于改进 Mask R－CNN 模型的工厂化养蚕蚕体识别与计数[J]. 智慧农业,2022,4(2):163－173.
HE Ruimin, ZHENG Kefeng, WEI Qinyang, et al. Identification and counting of silkworms in factory farm using improved Mask R－CNN model[J]. Smart Agriculture, 2022, 4(2): 163－173. (in Chinese)

[9] WEN C, WEN J, LI J, et al. Lightweight silkworm recognition based on multi-scale feature fusion[J]. Computers and Electronics in Agriculture, 2022, 200: 107234.

[10] 马小林,王梦麟,旷海兰,等. 基于 YOLOv8n 改进的蚕虫检测与计数方法[J]. 农业工程学报,2024,40(15):143－151.
MA Xiaolin, WANG Menglin, KUANG Hailan, et al. Detecting and counting silkworms using improved YOLOv8n[J]. Transactions of the CSAE, 2024, 40(15): 143－151. (in Chinese)

[11] 靳冲,陈首红. 基于改进 YOLO v5 的农田害虫检测计数方法研究[J]. 乡村科技,2025,16(14):140－143.
JIN Chong, CHEN Shouhong. Study on detection and counting method of farmland pests based on improved YOLO v5[J]. Rural Science and Technology, 2025, 16(14): 140－143. (in Chinese)

[12] WEN C, HE W, WU W, et al. Recognition of mulberry leaf diseases based on multi-scale residual network fusion SENet[J]. Plos One, 2024, 19(2): e0298700.

[13] 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.

[14] QI Y, HE Y, QI X, et al. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 6070－6079.

[15] CHEN Z, HE Z, LU Z M. DEA－Net: single image dehazing based on detail-enhanced convolution and content-guided attention[J]. IEEE Transactions on Image Processing, 2024, 33: 1002－1015.

[16] 毕泽洋,杨立伟,吕树盛,等. 基于 EDH－YOLO 的轻量型温室番茄检测方法[J]. 农业机械学报,2024,55(增刊2):246－254.
BI Zeyang, YANG Liwei, LÜ Shusheng, et al. Lightweight greenhouse tomato detection method based on EDH－YOLO[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(Supp.2): 246－254. (in Chinese)

[17] LUO W, LI Y, URTASUN R, et al. Understanding the effective receptive field in deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2016, 29:4905－4913.

[18] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117－2125.

[19] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278－2324.

[20] LIANG K, YIN Z, MIN M, et al. Learning dynamic prototypes for visual pattern debiasing[J]. International Journal of Computer Vision, 2024, 132(5): 1777－1799.

[21] 李兆东,章艳芳,汪蕴红,等. 基于 YC－YOLO v7 模型的油菜幼苗株数识别方法[J]. 农业机械学报,2024,55(12):322－332.
LI Zhaodong, ZHANG Yanfang, WANG Yunhong, et al. Identification of rapeseed seedling number based on YC－YOLO v7 model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(12): 322－332. (in Chinese)

[22] 李尚平,郑创锐,文春明,等. 基于改进 YOLO v5s 的甘蔗切种茎节特征识别定位技术[J]. 农业机械学报,2023,54(10):234－245,293.
LI Shangping, ZHENG Chuangrui, WEN Chunming, et al. Stem node feature recognition and positioning technology for transverse cutting of sugarcane based on improved YOLO v5s[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(10):234－245,293. (in Chinese)

[23] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R－CNN[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961－2969.

[24] BOLYA D, ZHOU C, XIAO F, et al. Yolact: real-time instance segmentation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9157－9166.

[25] FENG Y, HUANG J, DU S, et al. Hyper-yolo: when visual object detection meets hypergraph computation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024,47: 2388－2401.