

基于改进 YOLOv7 的密集褐菇检测算法研究

谭善恒, 许宗华

(青岛科技大学 信息科学技术学院, 山东 青岛 266000)

摘要: 在褐菇种植工业化的复杂环境下, 针对采摘机器人在密集生长的褐菇群的实时检测精度与速度低、误检率高等问题, 提出一种基于改进 YOLOv7 的密集褐菇检测算法; 为了防止网络退化, 提高网络的检测精度与速度, 降低网络的计算成本, 引入 ELAN_PS 模块替换原 ELAN 模块; 使用 AFPN 网络代替原网络的 Neck 部分进行多尺度融合, 为特征图分配不同的空间权重, 提高模型对密集目标的划分能力; 引入 MDIoU 损失函数作为算法的边界框损失函数, 优化网络训练的收敛速度, 提高模型对密集褐菇个体的检测精度; 将改进后的算法在自建的工业化种植褐菇数据集上进行训练与测试, 与原 YOLOv7 相比, 模型检测速度提高了 2.1%, 检测精确度提高了 4.9%, 平均精度 $mAP@0.5$ 提高了 9.1%。

关键词: 褐菇; 密集目标检测; YOLOv7; AFPN; ELAN_PS; MPDIoU

Research on Dense Detection Algorithm for Brown Mushroom Based on Improved YOLOv7

TAN Shanheng, XU Zonghua

(Qingdao University of Science and Technology, Qingdao 266000, China)

Abstract: In the complex environment of industrialized brown mushroom cultivation, to address issues such as low real-time detection accuracy and speed, and high false detection rates of harvesting robots in densely growing brown mushroom clusters, this paper proposes a dense brown mushroom detection algorithm based on an improved YOLOv7. To prevent network degradation, enhance the network's detection accuracy and speed, and reduce computational costs, the ELAN_PS module is introduced to replace the original ELAN module; The AFPN network is used to replace the Neck part of the original network for multi-scale fusion, allocating different spatial weights to feature maps, thereby enhancing the model's ability to distinguish dense targets; The MDIoU loss function is introduced as the bounding box loss function for the algorithm, optimizing the network training convergence speed and improving the model's detection accuracy for individual dense brown mushrooms; The improved algorithm is trained and tested on a self-built industrialized cultivated brown mushroom dataset. Compared with the original YOLOv7, the model increases the detection speed by 2.1%, the detection accuracy by 4.9%, and the average precision $mAP@0.5$ by 9.1%.

Keywords: brown mushroom; dense detection; YOLOv7; AFPN; ELAN_PS; MPDIoU

0 引言

褐菇所含有的营养物质多样且含量极高, 具有极佳的食用价值。2002 年以来, 我国食用菌产业发展迅速, 食用菌总产量由 2002 年的 1 458 万吨上升至 2020 年的 6 135 万吨^[1], 可以看出我国的食用蘑菇产业具有极大的潜力, 食用蘑菇的种植规模扩增的速度也会越来越快。在褐菇工业化生产过程中, 褐菇采摘是其中最重要的环节之一。当前我国大部分的企业仍以人工采摘作为

主要的采摘方式, 采摘过程中工人需要在褐菇种植仓库中昼夜观察、攀爬高架采摘褐菇, 劳动强度大, 采摘效率较低。褐菇的生长环境与其他蘑菇的生长环境存在差异, 同时期种植的褐菇生长密集与重叠, 部分褐菇倾斜甚至倒下, 菌丝大片缠绕。所以, 要提高在这种复杂环境下进行褐菇采摘的效率, 研究出一种能在复杂环境下进行褐菇精准与快速识别的算法正当其时。

当前, 基于深度学习的目标检测算法的检测性能出类拔萃, 主要分为基于候选区域的双阶段检测算法与基

收稿日期: 2024-03-05; 修回日期: 2024-04-11。

作者简介: 谭善恒(1999-), 男, 硕士研究生。

引用格式: 谭善恒, 许宗华. 基于改进 YOLOv7 的密集褐菇检测算法研究[J]. 计算机测量与控制, 2025, 33(5): 143-151, 161.

于回归的一阶段检测算法。其中双阶段检测算法的主要代表是以文献 [2] 中在 2014 年提出的 R-CNN 检测算法为基础思想的 Fast R-CNN^[3]、Mask RCNN^[4]、Faster R-CNN^[5]等算法。双阶段检测算法虽然在检测精度上有着较好的表现,但其检测速度较慢,模型体积较大,不适用于菌菇采摘这类实时检测任务。

一阶段目标检测算法主要包括 SSD^[6]算法、RetinaNe^[7]算法与 YOLO 系列算法,这类算法的主要特点是神经网络处理输入图片时跳过寻找独立地、显示地提取候选区域过程,直接输入图像检测到其中存在的目标物体的类别和位置信息。因此,一阶段检测算法的检测速度远超过双阶段检测算法,更适用于有实时需求的任务场景。

文献 [8] 在分割检测到双孢蘑菇的基础上,分别提取双孢蘑菇表面的语义特征,选择 OVR-SVMs 为分类算法对褐斑病阶段和人工或机械采摘损伤的双孢蘑菇做分类,其分类成功率达 94%。文献 [9] 在 YOLOv3 的基础上提出一种轻量级的香菇检测模型,计了一种称为 shuffle 自适应空间特征金字塔网络 (ASA-FPN) 的颈部网络,构建轻量级的 GhostNet16 代替 DarkNet53 作为主干网络,有效提高了网络的检测精度。文献 [10] 中提出一种改进的 YOLOv5s 算法,在骨干网络中融入 RFBSE 模块,使网络关注重点区域,通过施加通道注意力机制,增强对黑皮鸡枞菌特征表达能力,然后设计多分支采样 DCSP 池化模块,加强局部信息与全局信息的融合,最后在颈部网络采用 RFP 结构,增强鸡枞菌样本多遮挡场景下的检测能力,在一定程度上提高了模型的检测性能,其中对比原 YOLOv5s 模型算法, *mAP*、精确率、召回率分别提高 2.7、3.8、3.9 个百分点。文献 [11] 中出了一种基于 YOLOv5 模型的 OMM-YOLO (ostreatus measure modle-YOLO) 平菇目标检测与分类模型,通过在模型的 Backbone 层添加 ECA 注意力机制模块,提高了模型对平菇的检测准确度,然后在算法 Neck 层中引入 BiFPN 网络结构,加强模型对图像特征的提取能力,最终实验结果显示,改进模型的检测平均精度提升了 4.5%,检测精度提升了 1.1%。文献 [12] 中提出一种基于 YOLOv5 (YOLOv5-TL) 迁移学习结合褐菇三维边缘信息直径动态估测法的褐菇原位识别定位技术,通过 YOLOv5-TL 实现褐菇快速识别,再针对锚框区域褐菇图像进行图像增强、去噪、自适应等处理进行褐菇边缘定位与坐标提取,然后通过直径动态估测法实现褐菇尺寸的精确测量和中心点定位,最终模型的单帧图像处理时间为 50 ms,菇盖的尺寸测量平均精度为 97.28%,在光照情况下褐菇识别成功率高达 100%。虽然这种测量法对褐菇的识别率很低,但是对于密集、倾斜生长的褐菇群中的个体识别没

有涉及,其每秒检测速度 FPS 值仅为 20,模型处理图片速度较慢,还有很大的提升空间。文献 [13] 针对遮挡率高、重叠率高的夜间樱桃识别问题,在 YOLOv7 的基础上提出 YOLOv7-Cherry 算法,其引入 Soft NMS 算法使用衰减函数计算置信度,而不是直接进行暴力归零。YOLOv7-Cherry 对夜间樱桃图像进行检测后再对果实检测结果使用 Soft NMS 算法实现余去除的操作,有效减少遮挡对樱桃目标造成的损失,其检测精度相对于原模型来说提升了 7.7%,召回率提升了 10%。

目前,基于深度学习的目标检测算法在农业检测领域取得了较大的进展。但是食用菌在工业化种植环境下,二维机器视觉在食用菌生产中应用存在食用菌姿态复杂、成像环境复杂和对相机像素要求较高等问题,造成二维采集的图像容易受环境影响,并且图像特征不明显,导致实际检测效果不理想,在检测和控制食用菌生长、自动采摘和品质分级等方面发挥作用有限^[14]。同属于食用菌的褐菇培养基大多凹凸不平,布满菌丝,且褐菇生长以密集的褐菇群为主,褐菇重叠率高以及边缘粘连不清晰。对此,本文提出一种基于改进 YOLOv7 褐菇检测算法,主要改进如下:

- 1) 采用高效轻量化的 ELAN_PS 模块,增强了骨干网络的特征提取的能力。
- 2) 采用渐进特征金字塔网络 AFPN 替换原模型的 SPSPCSPC 模块,提高模型对密集目标的划分能力。
- 3) 改进原模型的边界框损失函数,降低漏检率与优化模型收敛速度。

1 改进的 YOLOv7 算法

本文在 YOLOv7^[15] 中的 Backbone 网络部分引入 ELAN_PS 结构,降低模型计算复杂度以及冗余特征,优化信息路径;然后采用 AFPN 模块替换原模型的 SP-PCSP 模块,减少非相邻特征图之间的较大语义差距,极大地避免特征信息在传输过程中丢失或者退化,提高模型对密集目标的划分能力;最后,引入 MPDIoU 作为模型的边界框损失函数,有效提高模型收敛速度以及检测精度。改进后的网络结构如图 1 所示。

1.1 高效轻量化的 ELAN_PS 结构

Efficient Layer Aggregation Network (ELAN) 是高效聚合层网络,主要由多个 CBS 模块分层堆叠组成,每个 CBS 由常规卷积 Convolution (Conv), Batch Normalization (BN), 激活函数 Sigmoid Weighted Liner Unit^[16] (SiLu) 组成,通过残差结构的思想增加网络深度,利用计算块中的堆栈结构优化了整个网络的梯度长度,并含有丰富的梯度流信息。ELAN 网络在处理输入数据时展现出强大的特征提取能力,使得网络可以在不增加额外复杂架构的情况下实现更高的检测精度。

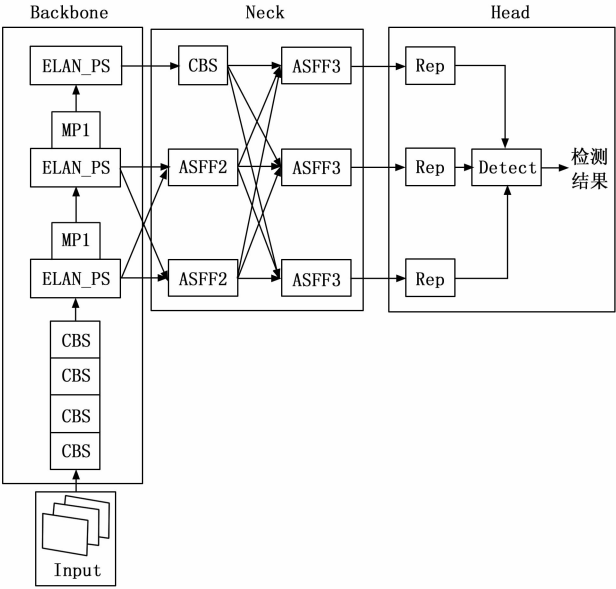


图 1 改进 YOLOv7 结构图

由于 ELAN 模块含有多层常规卷积, 不可避免地 为模型增加了特征冗余与计算成本, 从而影响模型的 运行速度与性能。考虑到复杂环境下密集菌群对模型的 检测精度与实时性的需求, 本设计从减轻模型计算负荷, 提高模型对特征图的提取能力的角度出发, 引入低成本 逐点卷积算子 Partial Convolution (PConv) 与空间和通 道重建卷积 Spatial and Channel reconstruction Convolution (SCConv) 提出高效轻量化的 ELAN_PS 模块。具 体的 ELAN_PS 结构如图 2 所示。

1) PConv 卷积算子:

学者们为了设计快速神经网络, 通常会将许多工作 都集中在提升每秒低浮点运算 (FLOPs) 的效率, 然 而模型计算中低的 FLOPs 主要是由于运算符的频繁内 存访问, 尤其是在多层次特征提取的深度卷积模型中。 文献 [17] 在 2023 年提出的 Fasternet 中, 首次提出了 逐点卷积算子 PConv, 通过利用特征图在不同通道之间 具有高度相似性的特征冗余进一步优化成本, 有效减少 冗余计算和内存访问, 提升模型的检测速度与检测精 确度。

PConv 卷积算子相较于常规卷积在特征提取的计算 方式上有所不同。如图 3、图 4 所示。对于卷积层共有 m 个 Filter, 卷积核大小为 $k \times k \times c$, 则常规卷积的 FLOPs 为 $h \times w \times k^2 \times l$, 式中, h 和 w 是特征图的宽和 高, l 是常规卷积作用的通道数。在实际实现时 PConv 算子的通道数 c 与 l 的比为 $r = \frac{c}{l} = \frac{1}{4}$, 所以 PConv 算 子的 FLOPs 仅仅为常规卷积的 1/16。

相应的, 常规卷积的内存访问数量由式 (1) 所示:

$$h \times w \times 2l + k^2 \times l^2 \approx h \times w \times 2l \quad (1)$$

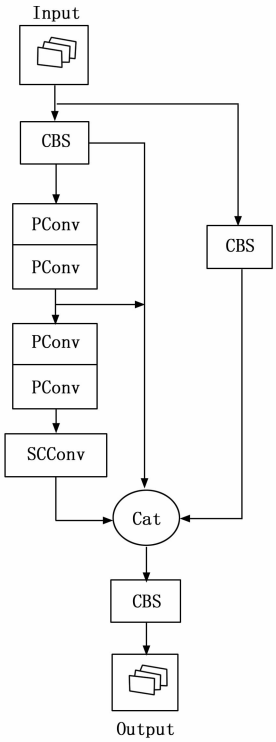


图 2 ELAN_PS 模块结构图

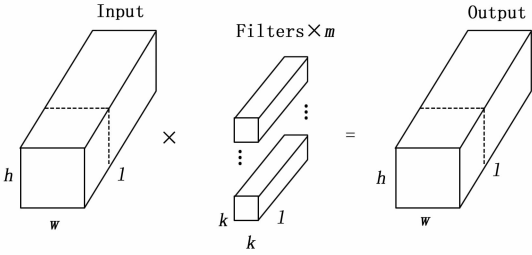


图 3 常规卷积运算图

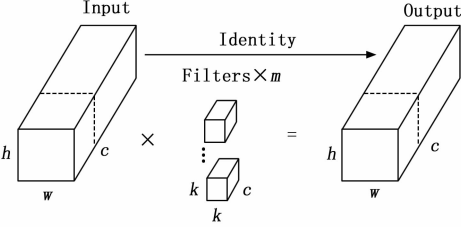


图 4 PConv 卷积算子运算图

由于 $l=4c$, 所以 PConv 算子的内存访问数量仅仅 为常规卷积的 1/4。

因此, 引入 PConv 替换 CBS 模块, 有效降低模型的 计算量与参数量, 提高检测速度。此外, PConv 能够 更好地利用计算设备的能力, 并在同时提取空间特征方 面表现出良好的效果^[18]。

2) SCConv 卷积模块:

文献 [19] 中于 2023 年提出一个即插即用的 SC-

Conv 卷积模块, 其中包含两个单元, 一个名为空间重构单元 SRU (Spatial Reconstruction Unit), 另一个名为通道重构单元 CRU (Channel Reconstruction Unit)。虽然 SCConv 卷积模块的设计目标在于减少视觉任务中由于冗余特征提取而产生的计算成本, 但是其通过在空间与通道方向的特征重构增强了模型对特征图细化识别能力。SCConv 结构如图 5 所示。

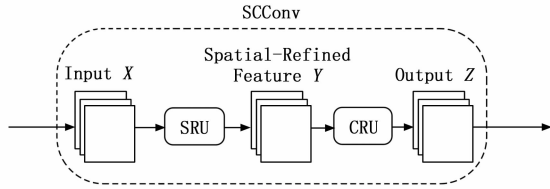


图 5 SCConv 卷积算子结构

在这个卷积算子中, 上一层卷积输入的特征 X 先经过空间重构单元 SRU, 得到空间细化的特征 Y 。再经过通道重构单元 CRU, 得到通道提炼的特征 Z 作为输出。SRU 单元主要分为特征分离与特征重建两个步骤。首先将输入的信息丰富的特征图与信息较少的特征图分离开来, 然后将信息丰富的特征与信息较少的特征相加重建, 生成信息更丰富的特征。

SRU 的结构如图 6 所示。对于上一层输入第 N 批次的 feature map X , 其通道数为 M , 大小为 $H \times W$ 。

首先, SRU 单元对 X 进行组归一化 (GN), 通过减去平均值 μ 并除以标准差 δ 来标准化输入特征 X , 具体计算过程如式 (2) 所示:

$$X_{out} = F_{GN} = \gamma \frac{X - \mu}{\sqrt{\delta^2 + \epsilon}} + \beta \quad (2)$$

其中: ϵ 是用于保证稳定性的极小常数, 可训练的参数 γ 代表特征图中更丰富的空间信息, 从而得出多个

特征图的重要性权重 W_γ :

$$W_\gamma = \frac{\gamma_i}{\sum_{j=1}^C \gamma_j} \quad (3)$$

设置对应 W_γ 归一化的阈值 x , 将映射到 $(0, 1)$ 范围的值与 x 比较, 高于 x 则得到有效信息权重 $W_1 = 1$, 否则得到冗余信息权重 $W_2 = 0$ 。然后将 W_1 与 W_2 特征图 X 分别进行逐元素点乘得到 X^{w1} 与 X^{w2} 。最后将二者通过交叉重建, 将加权后的两个不同信息特征沿通道拼接起来得到空间精细化特征图 X^w :

$$\begin{cases} X_1^w = W_1 \times X \\ X_2^w = W_2 \times X \\ X_{11}^w + X_{22}^w = X^{w1} \\ X_{12}^w + X_{21}^w = X^{w2} \\ X^{w1} \cup X^{w2} = X^w \end{cases} \quad (4)$$

其中: “ $\times$ ” 是元素乘法, “+” 是元素加法, “ $\cup$ ” 是求并集。经过 SRU 单元获得精细化特征 Y 之后, 输入到 CRU 单元进行特征通道提炼。CRU 的结构如图 7 所示, 其使用分割—转换—融合方法来降低通道冗余。将 Y 分割为不同通道数的两部分, 一部分通道数为 αC ($0 < \alpha < 1$), 其余的通道数为 $(1 - \alpha)C$, 随后分别输入两个 1×1 的卷积层得到 X_{up} 与 X_{low} 。然后将输入的 X_{up} 分别进行 GWC 和 PWC 操作之后再通过逐像素相加得到丰富特征 Y_1 , 将输入的 X_{low} 进行 PWC 计算之后再与本身取并集得到补充特征 Y_2 。

最后使用简化的 SKNet^[20] 方法来自适应的合并 Y_1 与 Y_2 。通过对 Y_1 与 Y_2 使用全局平均池化技术得到 S_1 与 S_2 , 这一步的目的是将特征图的全局空间信息和通道统计信息结合起来, 随后对 S_1 与 S_2 经过 Softmax 得到特征权重向量 Ψ_1 与 Ψ_2 , 基于两个特征权重向量合并

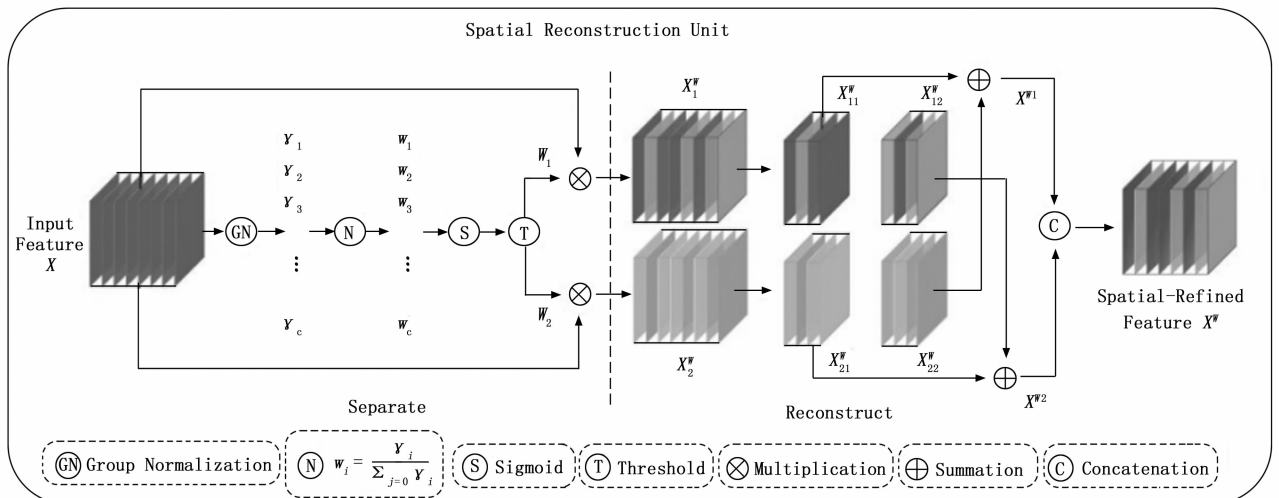


图 6 SRU 结构图

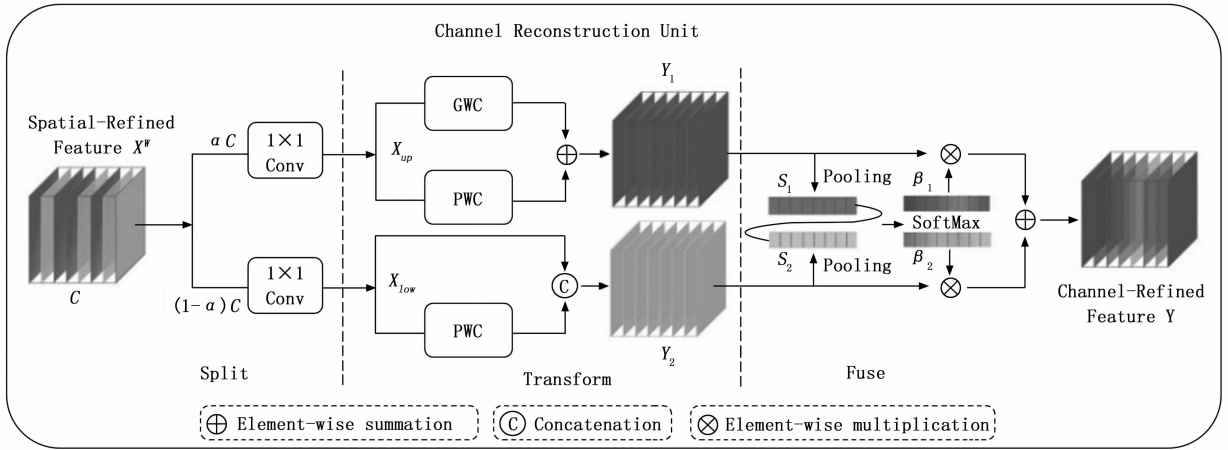


图 7 CRU 结构图

特征 Y_1 与 Y_2 得到空间通道精细化特征 Y :

$$Y = \Psi_1 Y_1 + \Psi_2 Y_2 \quad (5)$$

通过 SCConv 卷积能够有效地限制特征冗余, 并增强了特征表示的能力。经过以上阐述, 对原 ELAN 模块进行改进主要针对其计算资源需求与特征提取能力, 现总结结论如下:

原 ELAN 模块结构如图 8 所示, 它通过控制最短和最长的梯度路径, 使网络能够学习到更多的特征。其中最短路径包括一个 1×1 的常规卷积, 最长路径包括一个 1×1 的常规卷积与 4 个 3×3 的常规卷积, 经过计算可以得知, 常规卷积的 FLOPs 为 PConv 的 16 倍之多, 这表明在同等计算资源下, 使用 PConv 卷积算子可以节约更多的计算成本, 提高算法的运行速度。

虽然 PConv 在一定程度上降低了模型的计算量, 但是在深度的学习网络中不可避免的因为特征的多层次嵌套提取而在空间与通道方向是丢失一定的特征, 可能减弱了模型对特征的表征能力。

卷积层的输出特征图以及输入通道都存在高、低频分量。其中低频分量支撑的是目标的整体轮廓, 高频分量则更关注细节部分。SCConv 卷积算子在进一步减轻模型负荷的基础上精细化的提取特征图的空间与通道方向上的特征, 更关注特征图的高频分量, 使得模型获取具有判别力的特征表示, 增强了模型对目标边缘的识别能力。

1.2 特征融合模块

经过改进主干网络自下而上的特征提取之后的特征图含有丰富的空间与通道信息, 但在此过程中, 图像的分辨率会逐渐下降, 这导致图像中相对较小或者重叠的目标细节信息容易丢失或者退化。文献 [21] 在 2023 年提出的渐近特征金字塔网络 (AFPN, asymptotic feature pyramid network) 来支持非相邻层的特征直接交互, 使得浅层特征 (Low-Level) 与深层特征 (High-

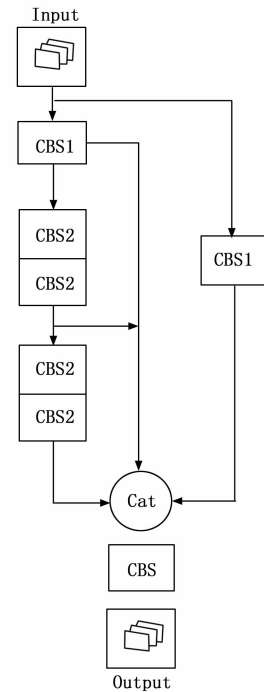


图 8 ELAN 模块结构图

Level) 可以互相补充丢失的细节特征。AFPN 中引入了 HRNet^[22] 反复融合浅层特征和深层特征以生成更丰富的顶层特征 (Top-Level) 的特点, 通过多阶段不同级别特征的交叉式渐进自适应融合方式, 将所有输入金字塔网络的 Low-Level 特征至顶层特征进行邻级融合, 避免了多级传输中的细节信息丢失或者退化。

AFPN 从模型的骨干网络中提取到不同级别的特征图, 首先将 Low-Level 的两个特征图输入到特征金字塔网络中进行融合, 随后逐位添加 High-Level 特征图、Top-Level 特征图。然后经过特征自适应空间融合过程, 将产生一组对应的多尺度饱满特征, 最后将这组多尺度特征输入 Head 检测端进行预测。如图 9 所示, 由

骨干网络输入的特征图经过两轮特征权重自适应与特征提取,不同 Level 特征的语义信息在渐进融合过程中更加接近,最终获得的一组饱满的特征 $D = \{D_1, D_2, D_3\}$ 。

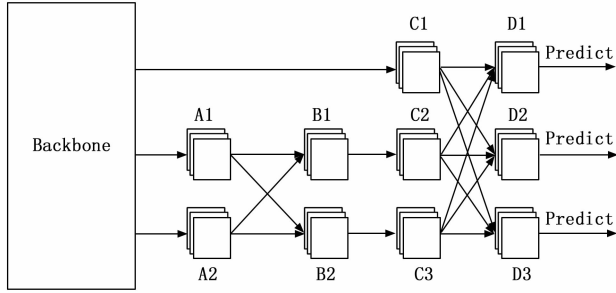


图 9 AFPN 网络示意图

为了突出关键 Level 的重要特征占比, AFPN 借鉴了 Adaptively Spatial Feature Fusion (ASFF) 网络每个尺度特征图的融合空间权重的自适应地学习方式,为不同层次的特征分配不同的空间权重,以此来增强关键层次特征的重要性,并减轻了来自不同目标的矛盾信息的影响。具体结构如图 10 所示。

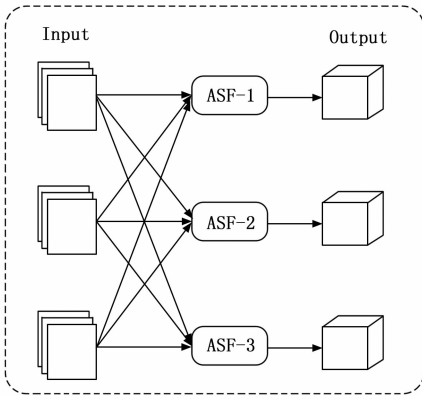


图 10 自适应分配权重示意图

多级特征图融合后的结果特征图的空间位置 (i, j) 处的向量是融合前 3 个特征图 (i, j) 处的向量的加权融合,特征图的空间重要性权重是由网络自适应学习到的,它们在所有通道空间都是共享的。使 $x_{ij}^{n \rightarrow l}$ 表示为特征图 n 到特征图 l 的位置 (i, j) 处的融合特征向量,通过多级特征的自适应空间融合获得结果特征向量 y_{ij}^l ,由特征向量 $x_{ij}^{1 \rightarrow l}$ 、 $x_{ij}^{2 \rightarrow l}$ 和 $x_{ij}^{3 \rightarrow l}$ 的线性组合定义如下:

$$y_{ij}^l = \alpha_{ij}^l x_{ij}^{1 \rightarrow l} + \beta_{ij}^l x_{ij}^{2 \rightarrow l} + \gamma_{ij}^l x_{ij}^{3 \rightarrow l} \quad (6)$$

$$\alpha_{ij}^l + \beta_{ij}^l + \gamma_{ij}^l = 1 \quad (7)$$

其中, α_{ij}^l 、 β_{ij}^l 和 γ_{ij}^l 表示各个特征 n 在结果特征图 l 的空间权重 $\alpha_{ij}^l, \beta_{ij}^l, \gamma_{ij}^l \in [0, 1]$ 。

密集、重叠的褐菇群作为工业化种植褐菇的主要生长方式,本身就给模型的检测能力带来了较大的挑战,再加上褐菇颜色相近,重叠的边缘不易被检测,使用原

始的 YOLOv7 算法无法对以菇群为主的褐菇进行有效检测。本文设计将原 YOLOv7 的 SPPCSPC 模块替换为 AFPN 网络,对非相邻层次特征图进行自适应融合并极大地降低了它们之间的较大语义差距,有效防止褐菇边缘、褐菇根部等重要特征信息在传输过程中丢失或者退化,增强了算法对密集褐菇的特征学习能力,提高算法的检测精度与泛化能力。

1.3 MPDIoU 损失函数

目前大多数性能优异的目标检测算法依赖于边界框回归 (BBR, bounding box regression) 损失函数用于确定检测目标的具体位置。然而,现有的大多数 BBR 损失函数在预测框与实际标注框具有相同的宽高比但宽度和高度值完全不同的情况下具有相同的值,这降低了边界框回归的收敛速度和准确性。MPDIoU^[23] 是一种基于最小点距离的新型边界框回归损失函数,其直接最小化预测边界框与实际标注边界框之间的左上角和右下角点距离来确定目标位置。

MPDIoU 的实现过程为:对于大小为 $w \times h$ 的输入图像, (x_1^A, y_1^A) 和 (x_2^A, y_2^A) 分别为预测框 A 的左上角与右下角的点坐标, (x_1^B, y_1^B) 和 (x_2^B, y_2^B) 分别为实际标注框 B 的左上角与右下角的点坐标。首先,计算这两个矩形框的左上角与右下角的点距离如下:

$$d_1^2 = (x_1^B - x_1^A)^2 + (y_1^B - y_1^A)^2 \quad (8)$$

$$d_2^2 = (x_2^B - x_2^A)^2 + (y_2^B - y_2^A)^2 \quad (9)$$

然后,计算出他们之间的最小点距离交并比为:

$$MPD = IoU - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2} \quad (10)$$

其中: IoU 是两个矩形框的交并比大小,通常作为预测框与实际标注框之间得到重叠程度。在模型训练阶段,通过最小化 MPDIoU 损失函数的值,使得模型预测的每个边界框 $\mathbf{B}_{prd} = [x^{prd}, y^{prd}, w^{prd}, h^{prd}]$ 趋近于目标的实际边界框 $\mathbf{B}_{gt} = [x^{gt}, y^{gt}, w^{gt}, h^{gt}]$ 。损失函数表达式如下:

$$\mathbb{L}_{MPDIoU} = 1 - MPDIoU \quad (11)$$

与大多数 IoU 相比, $MPDIoU$ 简化了两个边界框之间的相似性比较,在多边界框重叠或非重叠的任务中表现更为突出。所以本文采用 $MPDIoU$ 替换原来的 $CIOU$ ^[24],既可以提高模型在多重重叠褐菇群检测中保持良好的精度,又提高了模型的鲁棒性与训练速度。

2 实验结果与分析

2.1 研究对象

在褐菇种植工业化环境下,为了辅助褐菇采摘机器人对褐菇个体进行精确采摘,就需要视觉模块主对褐菇进行精准识别。但是,工厂种植环境下的褐菇生长以密集菇群为主,同时期种植的褐菇生长密集与重叠,褐菇

边缘不清晰，由于培养土在褐菇生长过程中被破坏，导致部分褐菇倾斜甚至倒下，褐菇的生长方式如图 11 所示，其中情况 A 展示了褐菇重叠与生长倾斜；情况 B 展示了褐菇完全倒下，无法观察到菇盖；情况 C 展示了在强光照射下褐菇边缘不清晰的情况。

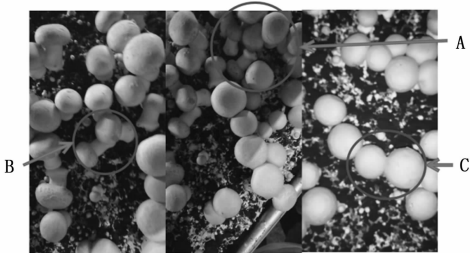


图 11 褐菇生长方式示意图

褐菇簇的密集生长通常可以分为高密度程度与一般密集程度。如图 12 所示，高密度程度的褐菇簇存在多种褐菇的生长姿态，比如说生长倾斜、多个菇盖重叠、褐菇个体遮挡甚至倒下等情况；而一般密集程度的褐菇簇存在轻微重叠现象，相对而言更容易被检测，如图 13 所示。

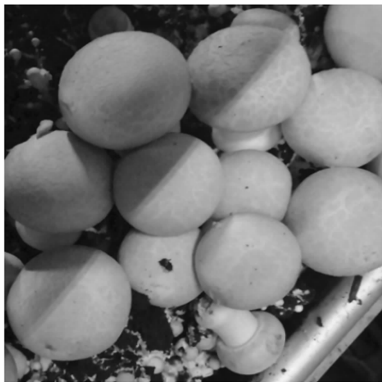


图 12 高密度程度的褐菇簇

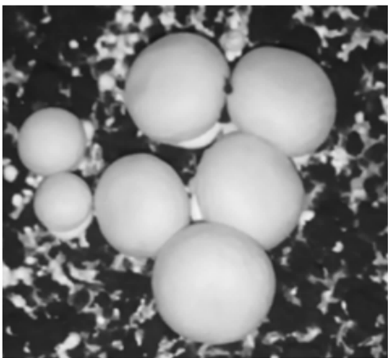


图 13 一般密集程度的褐菇簇

为了给褐菇采摘机器人提供有效而精准的褐菇识别模块，针对褐菇的多种生长姿态、形式识别困难等问题，提出本文改进的 YOLOv7 算法以提高褐菇簇中的

褐菇个体识别率并降低误检率。

2.2 数据集的采集与制作

本文以种植厂房内的褐菇作为研究对象，通过录制种植基地内多个生长时期的褐菇群视频作为数据源，然后设置一定的帧数间隔截取图片作为数据集。鉴于检测视角需要包括检测斜向生长以及贴地生长褐菇，所以采用与种植层垂直拍摄以及拍摄视角与菇床呈 30 度的夹角两种角度拍摄。

本文主要针对密集生长的褐菇不易检测问题，所以采集的视频大多数是高密度褐菇簇，极少部分是一般密集褐菇簇。其中高密度褐菇簇包含正常生长、斜向生长、倒地生长与粘连生长的完整单个褐菇或双褐菇；一般密集褐菇簇仅包括轻微重叠、倾斜等情况，将其作为平衡图像特征部分输入模型中进行训练。通过这种方式一共获取褐菇图片 735 张，其中垂直拍摄的图片为 486 张，斜向拍摄的图片为 249 张。所有图片都采用 LabelImg 工具进行标注，标注的所有标签均为“mushroom”，标注页面如图 14 所示。为了提高检测模型的鲁棒性，对 735 张图片分别进行图片锐度调整、图像翻转、图像对比度调整、图像模糊等 5 种方式，最后得到 3 675 张图片。最后将数据集分别划分为训练集、测试集与验证集，划分的比例为 6：2：2。

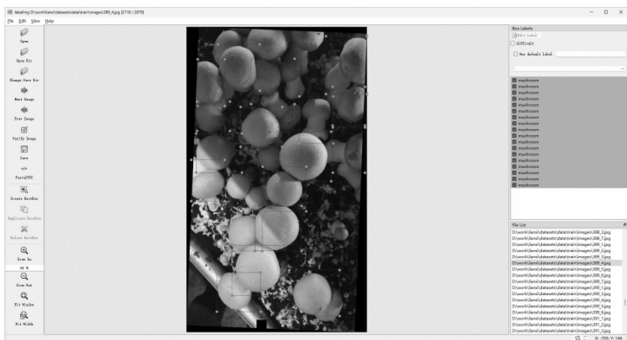


图 14 数据标注页面

2.3 训练平台与参数设置

本文实验的模型训练与测试均在实验室搭建的训练平台完成。实验训练平台为 Ubuntu 18.04 操作系统，搭载型号为 GeForce RTX 4080 的 GPU；CPU 型号为 Intel (R) Core (TM) i7-13700；同时，采用 pytorch 深度学习框架，使用 12.2 版本的 CUDA，开发软件为 python 3.9，开发编译器为 Visual Studio Code 2019。

模型训练的轮次大小为 200 (epochs)、初始学习率 (learning rate) 为 0.01、输入的图片大小 (imgsz) 为 640×640、每轮训练的批次大小 (batch-size) 为 16、模型训练采用无初始权重方式。

2.4 评价指标

为了对改进模型进行准确的评价，本文采用目标检

测算法领域内公认度最高的召回率 Recall (R)，精准率 Precision (P)，每秒内图片的检测数量 FPS 以及 *IOU* 值为 0.5 时的平均精度 Mean Average Precision (*mAP*) 作为本文实验的评价指标。其中，每个评价指标的定义方式如下所示：

$$R = \frac{TP}{TP + FN} \tag{12}$$

$$P = \frac{TP}{TP + FP} \tag{13}$$

$$mAP = \frac{1}{n} \sum_1^N \int_0^1 P(R) d(R) \tag{14}$$

其中：真正例 TP (True Positives) 表示褐菇目标被模型正确识别的总数量；假反例 FN (False Negatives) 表示褐菇目标没有被模型识别的总数量，即模型漏检褐菇；假正例 FP (False Positives) 表示非褐菇目标被模型误报为褐菇的数量，即漏报量；*N* 表示模型识别目标的类别数量，在本文中只对褐菇个体进行识别，所以 *N* 等于 1。

2.5 ELAN 模块的改进分析

在 ELAN 模块中引入新的卷积算子有多个不同的位置可以嵌入与替换，为了证明本改进的有效性，将 PConv 与 SCConv 分别嵌入 ELAN 模块的不同位置，以组成不同的方案进行对比。具体方案如表 1 所示，一共有 4 个对比方案。PConv 与 SCConv 在 ELAN 模块中的具体嵌入或替换位置 A、B、C、D 如图 15 所示。

表 1 实验方案

编号	说明
①	PC 替换位置 A SCC 放入位置 C
②	PC 替换位置 B SCC 放入位置 C
③	PC 替换位置 A,B
④	PC 替换位置 A,B SCC 放入位置 D

以上方案的模型训练均使用 2.1 中的自制数据集，采用召回率 *R*，精准率 *P* 以及 *mAP@0.5* 作为评价指标，模型训练均不使用初始权重。实验结果如表 2 所示。

表 2 实验结果

方案	Recall	Precision	<i>mAP@0.5</i>
①	0.878	0.935	0.902
②	0.861	0.940	0.893
③	0.909	0.924	0.922
④	0.898	0.965	0.907

方案①、②的整体数值都比方案③、④要差一些，可能在位置 A 替换；方案③的召回率与 *mAP@0.5* 虽然比方案④要高，但是模型的精确率却比方案④低 4.1%。在模型性能相差不是很大的情况下，选择精准率较高的方案在褐菇检测作业中更有性价比，所以方案

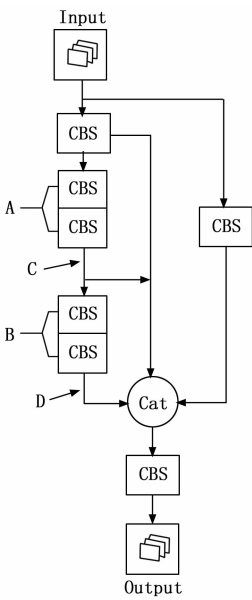


图 15 ELAN 位置图

④的结构更具有优越性。

2.6 消融实验

为了验证本文改进模型的合理性，对比基准模型进行消融实验，通过在原 YOLOv7 模型中逐步增添或者替换 ELAN_PS 结构、APFN 模块、MPDIoU 损失函数的方式，将得出的各个模型的性能评价指标与本改进进行对比，完成消融实验。本次实验均使用相同的数据集、实验环境以及模型参数。消融实验结果如表 3 所示。

表 3 消融实验结果

Model	Recall	Precision	<i>mAP@0.5</i>	FPS
v7	0.932	0.931	0.898	92
v7+MPDIoU	0.926	0.948	0.886	95
v7+AFPN	0.891	0.944	0.913	97
v7+ELAN_PS	0.898	0.965	0.907	89
v7+AFPN+MPD	0.939	0.950	0.874	106
本改进	0.981	0.989	0.983	97

从表 3 中可以看出，将原 YOLOv7 模型中的边界框回归损失函数替换为 MPDIoU 之后，检测精准度上升 1.7%，但是召回率略有下降；同样的，使用 APFN 模块替换原来的 SPPCSPC 模块后，模型的检测精度提高了 1.3%，召回率与平均精度都有所损失；在主干网络插入 ELAN_PS 模块替换原 ELAN 模块后，模型对目标的检测精度提高了 3.4%，但是模型的检测速度略有损失；将 APFN 模块与 MPDIoU 损失函数同时加入到原 YOLOv7 模型中，召回率、预测精度等都略有提升，其中检测精准度提高了 1.9%，检测速度提高了 15.2%；本改进在原 YOLOv7 的基础上融合了 3 种改进方式，改进后模型的精准率提高了 7.8%，检测速度

提高了 2.1%，召回率与平均精度都有明显的提升。通过消融实验，证明了本文提出的改进方式有效提升了 YOLOv7 模型在密集褐菇检测作业中的性能。

2.7 不同目标检测算法的对比实验

为了进一步分析与展现本改进模型在密集褐菇检测方面的有效性与优越性，本研究在 3.1 中自制数据集的基础上，将本改进算法与原 YOLOv7 算法、SSD、YOLOv3^[25]、YOLOv4^[26]、YOLOv5、YOLOv5 改进版、Faster R-CNN 算法进行对比实验。本节实验的评价指标分别是模型的平均精度 $mAP@0.5$ 与每秒检测帧数 FPS。实验结果如表 4 所示。

表 4 不同目标检测算法的性能对比

Model	$mAP@0.5$	FPS
YOLOv7	0.898	92
SSD	0.819	47
YOLOv3	0.814	41
YOLOv4	0.865	76
YOLOv5	0.886	83
Faster R-CNN	0.945	17
YOLOv5+MPD	0.804	76
YOLOv5+AFPN	0.851	88
YOLOv5+MPD+AFPN	0.834	86
本改进	0.983	97

从表 4 可以看出，本改进模型的 $mAP@0.5$ 值与 FPS 大小都优越与其他目标检测算法。其中，本改进模型在 IOU 阈值为 0.5 的平均精度分别比原 YOLOv7 算法、SSD、YOLOv3、YOLOv4、YOLOv5、Faster R-CNN 提升了 9.1%、16.4%、16.9%、11.8%、9.7% 与 3.8%，说明本改进模型在模型的性能上确实比其他常用模型具有优越性。另一方面，模型的检测速度直接影响到模型的实时性和可用性，相较于表 2 中 FPS 值最高的主流算法模型 YOLOv5 与原 YOLOv7，本文所提出的改进模型的 FPS 值比 YOLOv5 高 16.8%，比原 YOLOv7 高 2.1%。最后使用改进 YOLOv5 与本改进进行对比，在表中可以看出，使用 MPDIoU 与 AFPN 的改进 YOLOv5 算法并没有太突出的表现，甚至还低于原 YOLOv5 的性能数据，这可能是因为模结构的不匹配或者是计算冗余过多导致的。

综上所述，本文提出的改进模型在满足实时检测的要求之外，还能拥有高达 98.9% 的检测精度，同时模型的性能也优越于其他目标检测算法，整体表现突出，有效证明了本改进是具有优越性的。

3 结束语

工厂种植环境下的褐菇生长复杂，同时期种植的褐菇生长密集与重叠，由于培养土在褐菇生长过程中被破

坏，导致部分褐菇倾斜甚至倒下。视觉检测模块是从上往下进行检测的，褐菇边缘不清晰，重叠率高，导致当前主流的目标检测模型的识别精度不高，同时误检率太高。针对这些问题，本文以 YOLOv7 为基准模型，在原模型的骨干网络中引入 ELAN_PS 模块、改进模型的框回归损失函数以及使用 AFPN 模块引入原模型颈部网络。实验结果证明，改进后的算法在褐菇培养基的褐菇识别精准率提高了 4.9%，同时模型的 FPS 值保持在 97，在保持良好实时性的同时，改进的模型拥有高精度度与较好的性能，相对于其他目标检测算法更适用于实时褐菇个体检测。但是，在一些亮度过低、视觉可见度低的环境中，这些外部因素会对本改进算法的识别效果产生一定的影响，降低检测精度，因此本文算法还有较大的改进空间。在未来的工作中，可以对恶劣环境下的褐菇或者菌类识别做进一步研究，将恶劣环境因素做量化处理后加入检测模型中进行训练，以增强模型对环境因素的抵抗能力。

参考文献:

[1] 程嘉宁, 张俊飏, 颜廷武, 等. 中国食用菌产业地理集聚特征及空间演变趋势 [J]. 食药菌, 2023, 31 (1): 18 - 23.

[2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Richfeature hierarchies for accurate object detection and semanticsegmentation [C] //IEEE Conference on Computer Vision and Pattern Recognition (CVPR) New York: EEE Press, 2014: 580 - 587.

[3] GIRSHICK R. FAST R-CNN [C] //Proceedings of the CVPR2017: 1440 - 1448.

[4] HE K, GKIOXARI G, PIOTR DOLLÁr, et al. Mask R-CNN [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017. DOI: 10. 1109/TPAMI. 2018. 2844175.

[5] REN S Q, HE K M, 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, 2017, 39 (6): 1137 - 1149.

[6] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C] //Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21 - 37.

[7] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C] //Proceedings of the IEEE international conference on computer vision. 2017: 2980 - 2988.

(下转第 161 页)