

基于YOLOv8的中草药目标检测模型改进： 融合多注意力机制与大核卷积

沈珈毅¹, 韩刘婧¹, 王杰文², 何淳榆³

¹盐城工学院信息工程学院, 江苏 盐城

²宾夕法尼亚州立大学计算机科学与工程系, 美国, 斯泰特科利奇

³南京理工大学经济管理学院, 江苏 南京

收稿日期: 2025年8月8日; 录用日期: 2025年9月8日; 发布日期: 2025年9月16日

摘要

中草药作为中华文明的重要组成部分, 拥有数千年的悠久历史, 在传统医学体系上有着举足轻重的地位。随着现代科技的发展, 中草药的质量检测与类型识别已成为推动产业升级的关键环节。针对传统中草药人工采集、挑拣和分类的过程中效率低下、准确率不高等问题, 本文提出了一种基于YOLOv8s的改进目标检测模型YOLOv8s-LCD。该模型在YOLOv8s的主干网络基础上, 融合了多种注意力机制和大核卷积结构, 采用了大核注意力模块(LKA)、动态注意力模块(DyHead)、坐标注意力机制(CoordAttention)、CBAM模块与RepLK大核卷积。本文选取了白茯苓、白芍、人参等45类常见的中草药作为检测目标, 构建了包含一万张图像的中草药数据集, 并与原始YOLOv8s、YOLOv8s-ECA、YOLOv8s-GCNet和YOLOv8s-CondConv四个模型进行了对比。结果表明, YOLOv8s-LCD在mAP@0.5和mAP@0.5:0.95指标上均显著优于其他模型。相较于原始的YOLOv8s模型, YOLOv8s-LCD在mAP@0.5、mAP@0.5:0.95上分别提升3.7%和4.2%。上述结果验证了多重注意力机制与大核卷积融合的有效性。该研究为中草药的智能识别提供了一种高效、精准的技术方案, 具有良好的实际应用价值与推广前景。

关键词

YOLOv8, LKA, DyHead, RepLK, CoordAttention, 目标检测, 中草药

Improved Model for Chinese Herbal Medicine Object Detection Based on YOLOv8: Integrating Multi-Attention Mechanisms and Large-Kernel Convolutions

Jiayi Shen¹, Liuqing Han¹, Jiewen Wang², Chunyu He³

¹School of Information Engineering, Yancheng Institute of Technology, Yancheng Jiangsu

文章引用: 沈珈毅, 韩刘婧, 王杰文, 何淳榆. 基于YOLOv8的中草药目标检测模型改进: 融合多注意力机制与大核卷积[J]. 计算机科学与应用, 2025, 15(9): 161-173. DOI: 10.12677/csa.2025.159234

Abstract

Chinese herbal medicine, as a vital component of Chinese civilization, has a rich history spanning thousands of years and plays a significant role in traditional medical systems. With the rapid advancement of modern technologies, the quality assessment and classification of Chinese herbs have become critical tasks in driving industrial modernization. However, conventional methods for collecting, sorting, and classifying herbs are typically labor-intensive, inefficient, and prone to human error. To address these challenges, this study proposes an enhanced object detection model, termed YOLOv8s-LCD, based on the YOLOv8s architecture. The proposed model integrates multiple attention mechanisms and large kernel convolution modules into the YOLOv8s backbone, incorporating Large Kernel Attention (LKA), Dynamic Head (DyHead), Coordinate Attention (CoordAttention), Convolutional Block Attention Module (CBAM), and RepLK convolutional blocks. A custom dataset comprising 10,000 images across 45 commonly used Chinese herbal categories—including *Poria cocos*, *Paeonia lactiflora*, and *Panax ginseng*—was developed for evaluation. Comparative experiments were conducted against the baseline YOLOv8s, as well as three existing variants: YOLOv8s-ECA, YOLOv8s-GCNet, and YOLOv8s-CondConv. Experimental results demonstrate that YOLOv8s-LCD achieves superior performance, attaining improvements of 3.7% in mAP@0.5 and 4.2% in mAP@0.5:0.95 over the original YOLOv8s model. These findings validate the effectiveness of combining multi-attention mechanisms with large kernel convolutions for enhanced feature extraction and detection accuracy. The proposed method offers a robust and efficient solution for the intelligent recognition of Chinese herbal medicine, with strong potential for real-world deployment and industrial application.

Keywords

YOLOv8, LKA, DyHead, RepLK, CoordAttention, Object Detection, Chinese Herbal Medicine

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

中草药作为中华民族几千年传统医学的重要载体，历经数代传承与实践积淀，在防病治病、调养生息等方面发挥着举足轻重的作用。随着中医药现代化的发展，中药材产业规模不断扩大，其品质检测与种类识别问题成为制约产业链智能化、标准化水平提升的关键因素。传统中草药的采集、分拣及分类工作仍然以人工为主，存在着效率低、准确率不稳定等问题，难以适应现代中药加工和流通的高质量发展需求。近年来，图像识别和深度学习技术的快速发展为中草药的智能识别提供了新路径。目前，已有研究尝试将图像处理与机器学习方法应用于中药材分类任务，如基于 SVM、K-means 聚类等传统算法，在特定场景下可实现一定程度的识别效果。Liu 等[1]提出了细粒度特征增强的 CHMRN 网络，通过引入自顶向下的特征融合模块整合多尺度语义信息捕捉全面的上下文特征，并设计了自底向上的通道信息补偿模块，以增强细粒度特征的表达能力，确保能够准确捕捉中药类别之间的细微差异的方法；Mao 等[2]提

出了一种基于深度颜色特征融合的中草药菊花图像质量评价方法；Pan 等[3]提出了一种以 YOLOv5 为基础架构，引入聚合像素聚焦注意力机制和卷积门控线性单元(Gated Linear Unit, GLU)，模拟生物视觉系统，强化模型信息混合和自然视觉感知能力的方法；Zhang 等[4]提出了一种采用残差神经网络结合卷积块注意力模块，并应用融合局部二值模式特征以及迁移学习策略来提升模型的识别准确率的方法。然而，这类方法通常依赖手工特征提取，泛化能力弱，面对背景复杂、类别较多、形态相似度高的中草药图像识别任务时，识别性能较低。随着深度神经网络的发展，卷积神经网络(CNN)在图像分类与检测任务中取得了显著成果，尤其是目标检测算法在农业和医疗图像识别领域广泛应用，表现出良好的端到端学习与特征表达能力。YOLO (You Only Look Once)系列作为目标检测领域的主流算法，其兼顾检测速度与精度，且表现突出。YOLOv8 相较于之前的版本，优化了结构设计与性能，具备较强的泛化能力和实用性。然而，在中草药目标检测这一具体任务中，仍面临多类别、相似外观、小目标等挑战，导致检测精度仍有待提升。此外，当前针对中草药图像的目标检测研究相对较少，仍缺乏系统性的结构优化探索与模块组合的相关实验。为提升 YOLOv8 在中草药检测任务中的表现，本文提出一种基于 YOLOv8s 的改进模型，即 YOLOv8s-LCD。该模型在 YOLOv8s 主干网络的架构上，融合了多种注意力机制与大核卷积结构，集成大核注意力模块(LKA)、动态注意力模块(DyHead)、坐标注意力机制(CoordAttention)、CBAM 模块以及 RepLK 大核卷积模块，有效增强了模型对中草药小目标与相似类别的区分能力。本实验构建了包含 45 类常见中草药、总计 1 万张图像的中草药目标检测数据集，并分别将该模型与 YOLOv8s 原始模型以及 YOLOv8s-ECA、YOLOv8s-GCNet、YOLOv8s-CondConv 等改进模型进行了对比实验。实验结果表明，YOLOv8s-LCD 在 mAP@0.5 和 mAP@0.5:0.95 指标上均显著优于其他模型，验证了多重注意力机制与大核卷积结构在提升检测精度方面的有效性。本文提出的方法为中草药图像智能识别提供了一种高效、准确的解决方案，具有良好的实际应用价值与研究参考意义。

2. YOLOv8 神经网络

YOLOv8 是由 Ultralytics 于 2023 年 1 月 10 日发布的 YOLO 系列的新一代模型，代表了当时目标检测技术的前沿水平。相较于 YOLOv5，YOLOv8 在网络结构、算法设计与任务适应性等方面均进行了系统性的优化，显著提升了模型的检测精度、推理速度及部署灵活性。况且 YOLOv8 并非仅限于目标检测任务，还支持包括目标检测、实例分割、图像分类及关键点与姿态估计在内的多种视觉任务，体现出高度统一的多任务视觉建模能力。YOLOv8 提供了五种不同规模的预训练模型，分别为 Nano、Small、Medium、Large 与 Extra Large，用户可根据实际应用场景对性能与资源需求之间进行选择。其中，大模型(如 L 和 X)在保持较高计算效率的前提下，进一步提升了检测精度，适用于高性能平台部署。此外，YOLOv8 摒弃了传统 Anchor-based (基于锚框)检测机制，转而采用 Anchor-free (无锚框)设计，使模型结构更加简洁，检测过程更准确、快速，提高了鲁棒性，且对不同目标尺度与形状具备更强的适应能力。在网络结构方面，YOLOv8 引入了新型的 Backbone 与 Neck 模块，替代 YOLOv5 所采用的 CSPDarknet 架构，并融合 C2f 等轻量高效的模块设计，有效提升了特征表达能力与计算效率。与此同时，YOLOv8 继续保留了优化后的空间金字塔池化模块(SPPF)，以增强对多尺度目标的感知能力。模型同时支持基于 P5 (输入分辨率为 640)与 P6 (输入分辨率为 1280)的检测结构，适应不同复杂度的检测任务。此外，YOLOv8 还集成了基于 YOLACT 框架改进的实例分割能力，进一步拓展了其在多任务视觉场景下的适用范围。综上所述，YOLOv8 凭借其模块化、高性能与多任务适配的特性，已成为当前主流视觉任务中一种极具竞争力的端到端解决方案。

Backbone 是 YOLOv8 的特征提取模块，负责从输入图像中逐层提取丰富的视觉特征。与 YOLOv5 相比，YOLOv8 的主干结构不再使用传统的 CSPDarknet，而是引入了更轻量化、更高效的模块组合，用于

提升网络对空间、语义和长距离依赖关系的建模能力。除此此外，YOLOv8 同样保留了 SPPF (Spatial Pyramid Pooling Fast)模块，凭借多尺度感受野增强目标感知能力，改善模型对不同尺寸目标的识别效果。整体上，YOLOv8 的 Backbone 更注重特征层次的表达与感受野拓展，并融合注意力机制，有效提升了对复杂场景中目标的建模能力。

Neck 模块主要用于整合来自 Backbone 不同层级的特征图，实现多尺度特征融合，从而增强模型对不同尺寸目标的检测能力。YOLOv8 中 Neck 结构灵活性更强，不再局限于传统的 FPN + PAN 架构 (Feature Pyramid Network + Path Aggregation Network)，而是结合了多种注意力机制，进一步加强了高层语义与底层细节之间的交互。此外，YOLOv8 的 Neck 设计强化了语义和边缘特征之间的信息流通，从而更好地检测小目标和遮挡目标。

Head 是最终完成目标检测预测的模块，负责输出每个候选目标的边界框位置、置信度和类别信息。与 YOLOv5 不同，YOLOv8 抛弃了 Anchor-based (锚框)机制，采用了 Anchor-Free (无锚)结构，换成了目前主流的解耦头结构 (Decoupled-Head)，使得网络更轻量，推理更高效，且能适应更复杂的目标形状和尺度。检测头在三个不同尺度上进行输出，分别对应小、中、大目标的检测。每个尺度的输出都会经过非极大值抑制 (NMS) 处理，去除重复检测框，最终输出高质量目标候选，其结构如图 1 所示。

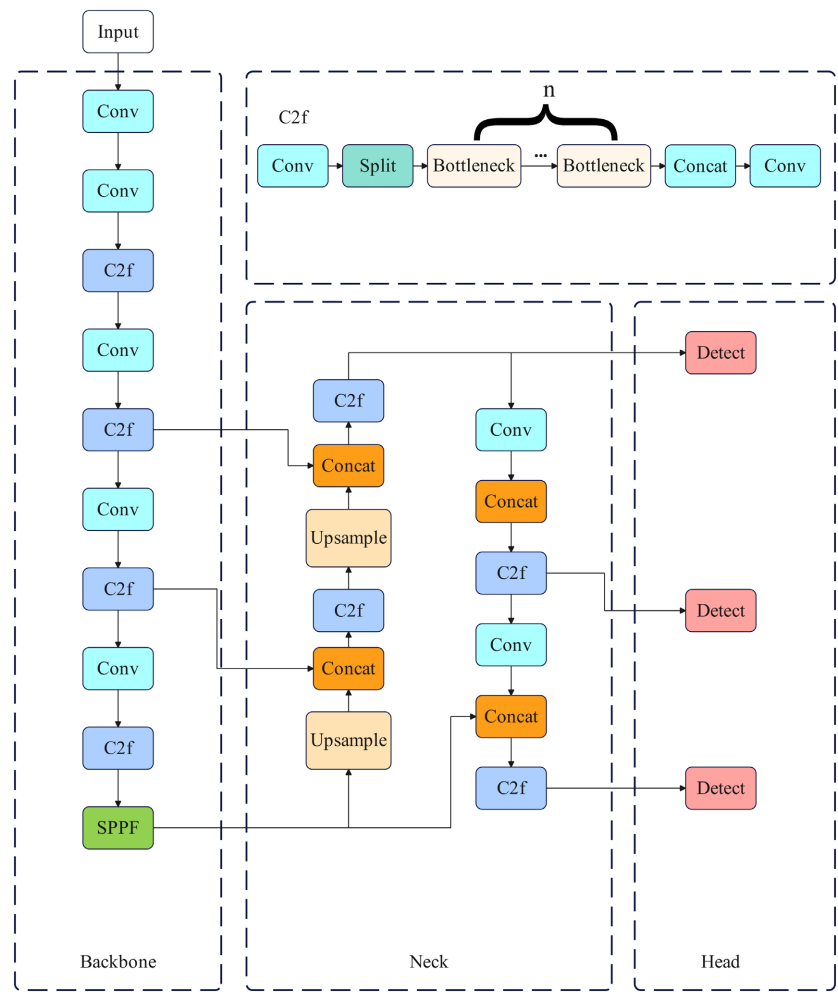


Figure 1. Overall structure of the YOLOv8 network
图 1. YOLOv8 网络整体结构

3. YOLOv8s 模型改进

3.1. LKA 大核注意力机制

LKA (Large Kernel Attention)大核注意力模块是一种提升卷积神经网络感受野和建模长距离依赖能力的轻量级注意力机制，最早由 RepLKNet 提出，旨在不显著增加计算量的前提下，增强模型对大尺度目标和全局上下文的理解能力。传统卷积感受野较小，对大范围依赖建模能力弱，而直接使用大卷积核(如 31×31 、 51×51)虽然能提升感受野，但参数量和计算量极大。LKA 采用分解策略，将大核卷积拆解为组合形式，用较小的卷积核模拟大卷积的效果，在提升感受野的同时，保持高效计算。LKA 模块引入了卷积核分解(Kernel Decomposition)策略，将原本计算量庞大的大核卷积分解为多个计算复杂度较低的卷积操作组合。具体而言，LKA 通常由三部分组成：Depthwise Dilated Convolution (深度可分离空洞卷积)：通过引入空洞率(dilation rate)在不增加参数量的前提下有效扩展感受野；Depthwise Convolution (深度可分离卷积)：对空间特征进行进一步精细建模；Pointwise Convolution (1×1 卷积)：实现通道间的信息交互与融合。这种组合方式不仅能够模拟大核卷积的感受野效果，还保持了较低的运算复杂度，从而在边缘设备或对推理速度要求较高的场景中具备极大的应用潜力，其结构如图 2 所示。

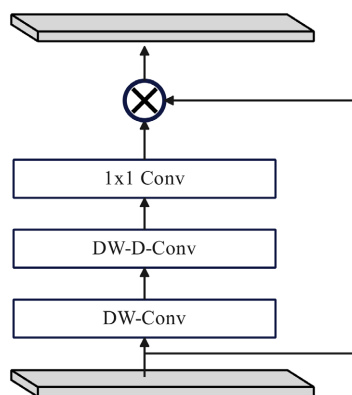


Figure 2. Structure of the Large Kernel Attention (LKA) module
图 2. LKA 大核注意力模块结构

3.2. RepLK 大核卷积

RepLKNet (RepLargeKernel Network)是由商汤科技(SenseTime)于 2022 年提出的一种新型卷积神经网络架构，其核心思想是通过显著扩大卷积核尺寸(例如 31×31 、 51×51 等)以拓展网络的感受野，从而提升模型对全局上下文信息的建模能力。与传统卷积神经网络普遍使用的小卷积核(如 3×3)相比，RepLKNet 的大核设计使得其在处理具有复杂背景或需要长距离依赖建模的视觉任务时具备更优的性能表现。为了在保持这种全局感知能力的同时，避免引入大量参数和计算开销，RepLKNet 引入了结构重参数化 (Structural Re-parameterization)技术，即在训练阶段使用包含分支结构的复杂网络以提升表达能力，而在推理阶段将其重参数化为等效的简单结构(如单一大卷积核)，以实现更快的推理速度和更低的计算资源占用。这种“训练 - 推理分离”的设计不仅提升了模型的可部署性，也保持了精度与效率之间的平衡。此外，RepLKNet 在多个视觉任务(如图像分类、目标检测、实例分割等)中都展现出强大的性能表现，特别是在不依赖 Transformer 或注意力机制的前提下，仅凭卷积结构便达到了与主流视觉 Transformer 相媲美甚至更优的效果。因此，RepLKNet 的提出不仅推动了 CNN 架构的发展，也为构建高性能、高效率的视觉模型提供了新的方向，其具体结构如图 3 所示。

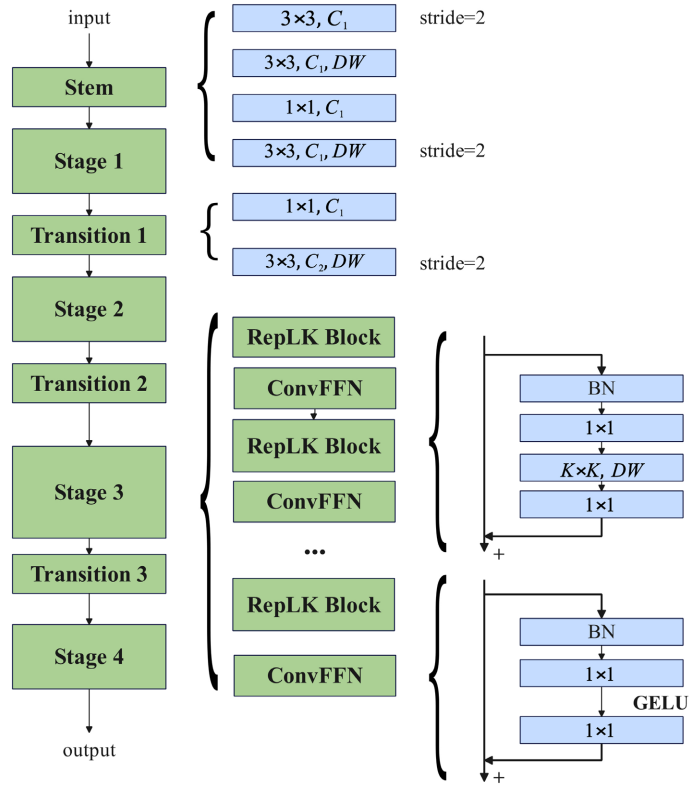


Figure 3. Structure of the ReplK large kernel convolution
图 3. ReplK 大核卷积的结构

3.3. DyHead 动态注意力模块

DyHead (Dynamic Head)是由加州大学伯克利分校(UC Berkeley)与亚马逊(Amazon)联合研究团队于2021年提出的一种高效且通用的动态注意力机制模块,主要应用于目标检测任务中,以提升特征金字塔中不同尺度和语义层之间的特征交互能力。该模块的设计理念是通过在空间、通道和任务维度上引入注意力机制,实现对多维度特征的动态建模与信息聚合,从而增强检测头对目标多样性(如尺度差异、语义复杂度、位置变化等)的适应性。与传统静态结构相比,DyHead 能够根据输入图像的内容灵活地调整注意力分布,使模型更加关注于关键区域或有辨别性的语义特征。它在保持较低计算代价的同时,有效提升了模型对小目标、遮挡目标和多类复杂场景的检测性能。DyHead 通常作为替代或增强 FPN/PAN 模块的补充单元,被广泛集成于如 YOLOv8、FCOS、RetinaNet、Faster R-CNN 等现代检测框架中,用于优化检测头的特征融合过程,进一步提升检测精度与鲁棒性。该机制在多个主流数据集上取得了显著性能提升,证明了其强大的通用性与实用价值,其结构如图 4、图 5 所示[5]。相关公式如下:

$$\pi_L(F) \cdot F = \sigma \left(f \left(\frac{1}{SC} \sum_{s,c} F \right) \right) \cdot F$$

$$\pi_S(F) \cdot F = \frac{1}{L} \sum_{l=1}^L \sum_{k=1}^K w_{l,k} \cdot F(l; p_k + \Delta p_k; c) \cdot \Delta m_k$$

$$\pi_C(F) \cdot F = \max(\alpha^1(F) \cdot F_C + \beta^1(F), \alpha^2(F) \cdot F_C + \beta^2(F))$$

$$W(F) = \pi_C(\pi_S(\pi_L(F) \cdot F) \cdot F) \cdot F$$

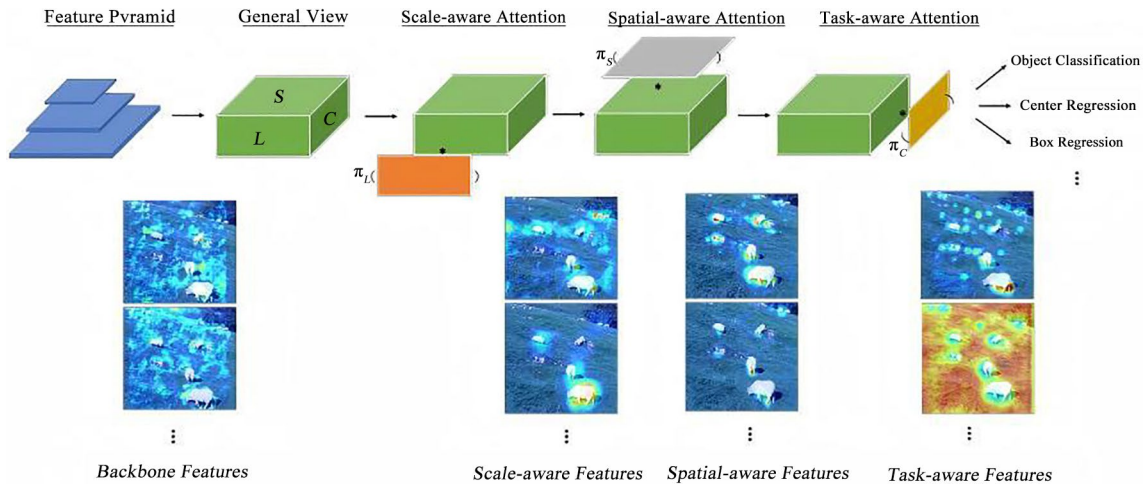


Figure 4. Dynamic Head approach
图 4. DyHead 方法

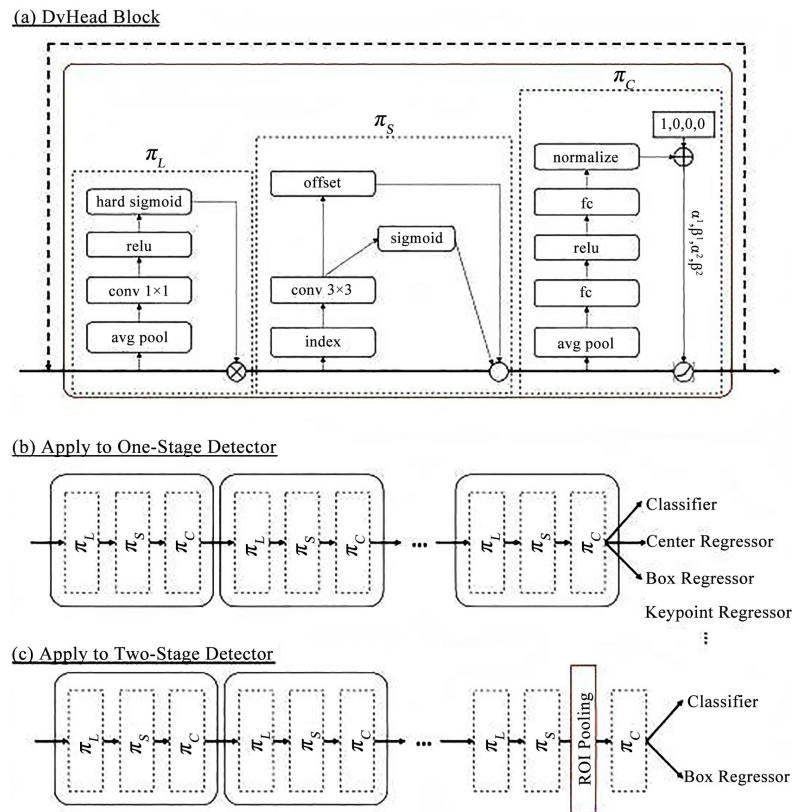


Figure 5. Implementation and Application of DyHead
图 5. DyHead 实现与应用

3.4. CoordAttention 坐标注意力机制

Coordinate Attention (坐标注意力)是一种高效轻量的注意力机制，由华为诺亚方舟实验室于 2021 年提出，旨在解决传统通道注意力机制(如 SE、CBAM)在空间信息建模能力方面的局限。传统通道注意力通常忽略了特征图中的位置信息，仅关注全局的通道关系，导致其在处理具有显著空间结构特征的目标(如细长、

细粒度目标)时表现不佳。为此, CoordAttention 创新性地引入空间位置编码, 将输入特征图沿两个空间维度(高度和宽度)分别进行方向编码, 使得注意力机制能够同时捕捉空间位置信息与通道间依赖关系。通过这种分离式的注意力编码方式, CoordAttention 能够在保持计算效率的同时, 有效提升模型对目标空间结构的建模能力, 实现更精细的目标定位与识别。该机制不仅提升了模型对细粒度目标、长条形物体等结构性目标的检测效果, 还因其高效性和模块兼容性, 广泛应用于轻量级神经网络和移动端视觉模型中, 并在如 YOLOv8、MobileNetV3、EfficientNet 等结构中展现出良好的性能增益。其相关公式如下所示:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w)$$

$$f = \delta \left(F_1 \left(\begin{bmatrix} z^h, z^w \end{bmatrix} \right) \right)$$

$$g^h = \sigma \left(F_h \left(f^h \right) \right)$$

$$g^w = \sigma \left(F_w \left(f^w \right) \right)$$

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j)$$

3.5. CBAM 注意力机制

CBAM 注意力机制(Convolutional Block Attention Module) [6]是一种结合了通道注意力与空间注意力的轻量级注意力机制。其核心思想是通过顺序地建模通道和空间两个维度上的注意力信息, 使网络能够自适应地关注“重要的特征通道”和“关键的空间位置”, 从而增强有用信息、抑制冗余干扰, 进一步提升模型的特征表示能力与检测性能。CBAM 模块主要包含两个子模块: 通道注意力模块(Channel Attention Module)和空间注意力模块(Spatial Attention Module), 按顺序串联而成。在通道注意力模块中, 分别使用全局最大池化和全局平均池化对输入特征图进行压缩, 获得两个通道描述向量。随后, 这两个向量通过一个共享的多层感知机(MLP)进行处理并相加, 再经 Sigmoid 函数生成通道注意力权重, 对输入特征图在通道维度上进行加权增强。在空间注意力模块中, 首先在通道维度上对特征图进行最大池化和平均池化操作, 将其压缩为两个空间注意力图。然后将二者拼接后送入一个 7×7 卷积层, 并通过 Sigmoid 激活函数生成空间注意力图, 用于对通道增强后的特征图在空间位置上进一步加权调整, 其结构如图 6 所示。

$$M_c(F) = \sigma \left(MLP \left(AvgPool(F) \right) + MLP \left(MaxPool(F) \right) \right)$$

$$M_s(F) = \sigma \left(f^{7 \times 7} \left(\begin{bmatrix} AvgPool(F); MaxPool(F) \end{bmatrix} \right) \right)$$

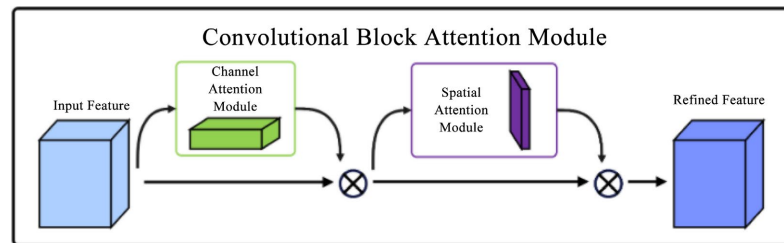


Figure 6. Structure of the CBAM attention mechanism
图 6. CBAM 注意力机制结构

3.6. YOLOv8-LCD 模型

为进一步提升 YOLOv8 在多尺度目标,尤其是小目标及复杂背景条件下的中草药检测能力,本文提出了一种基于多注意力机制和结构增强模块的改进型目标检测模型——YOLOv8-LCD。该模型融合了五类主流结构增强技术:大核注意力机制、坐标注意力机制、重参数化大卷积结构、CBAM 注意力机制以及动态注意力机制,以提升模型的全局建模能力、空间感知能力及特征自适应整合能力。在具体实现上,本文设计并集成了五类结构增强模块: C3_LKA、C3_Coord、C3_RepLK、DyHead 和 CBAM。具体而言, C3_LKA、C3_Coord 与 C3_RepLK 模块是在原始 YOLOv8 C3 模块的基础上引入相应注意力机制或大卷积结构,通过模块化设计实现特征增强的深度集成。该策略不仅增强了深层网络的空间依赖建模能力,还提高了模型对中草药中细粒度和易遮挡目标的检测敏感性。与部分研究仅在局部引入注意力机制不同,本文采用了系统性的结构替换策略,将 YOLOv8 中多个关键语义层级中的 C3 模块,全面替换为功能增强后的 C3_LKA、C3_Coord 与 C3_RepLK 模块,确保多层次特征图中显著通道与关键空间区域的信息表达能力均得到均衡提升。在 Neck 模块中引入 DyHead 结构和 CBAM 模块,通过引导模型在空间、通道与任务维度进行动态加权,实现更具自适应性的多尺度特征融合。与此同时,还能联合建模通道与空间注意力,进一步增强模型对复杂背景中细小目标的辨别能力,尤其在目标密集或形态多变场景下具有显著优势。整体上, YOLOv8-LCD 在保留原 YOLOv8 网络拓扑结构与轻量级特性的前提下,通过精细化的结构改造与注意力机制集成,显著提升了从浅层到底层特征的表达能力,具备良好的检测性能与部署适应性。在实验部分,本文基于 YOLOv8s 架构构建改进模型,并分别与原始 YOLOv8 及不同模块组合版本进行对比实验,系统分析各结构增强机制对检测精度、模型复杂度及推理效率的综合影响,其结构如图 7 所示。

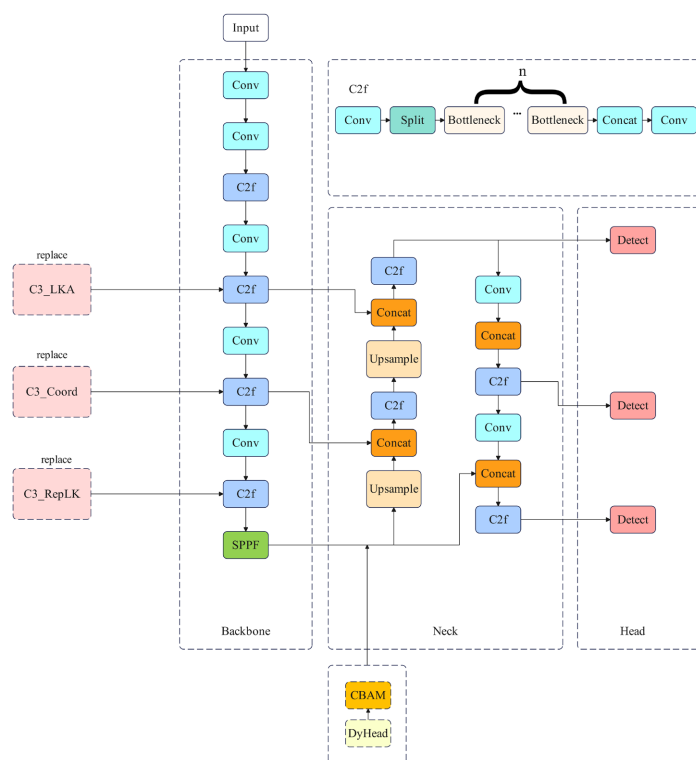


Figure 7. Schematic diagram of module replacement in the improved model

图 7. 改进模型的模块替换示意图

4. 实验结果与分析

4.1. 硬件及其参数设置

本实验使用 PyTorch 深度学习框架以及 vscode 开源集成开发环境进行模型构建与训练。硬件采用 Intel Core i5-12400F 处理器、32GB RAM 内存，搭载 NVIDIA GeForce RTX 3090 GPU，支持 CUDA 11.8 加速。软件方面采用 PyTorch 2.1.0 与 Python 3.10。完成了 YOLOv8s-LCD 模型、YOLOv8s 原始模型及其改进模型(YOLOv8s-ECA, YOLOv8s-GCNet, YOLOv8s-CondConv)的训练与测试工作，确保了实验的高效性与模型在高性能计算设备下的训练稳定性。如表 1 所示：

Table 1. Configuration of experimental software and hardware equipment

表 1. 实验软硬件设备配置

CPU	RAM	GPU	Cuda	Pytorch	python
I5-12400f	32.0G	NVIDIAGeForceRTX3090	11.8	2.1.0	3.10

4.2. 评价指标

本实验采用 mAP@0.5、mAP@0.5:0.95、准确率(Precision, P)、召回率(Recall, R)、Param(M)、Flops(G) 以及 fps 七项指标对模型的检测性能进行综合评估。其中，mAP@0.5 反映了在 IoU 阈值为 0.5 时模型整体的平均精度，mAP@0.5:0.95 更全面地考察了不同 IoU 下的检测能力，能够体现模型对目标定位精度的要求。准确率用于衡量模型检测出的正样本中有多少是实际为正的，召回率则表示模型成功检测出的正样本占有所有实际正样本的比例。通过这些指标的综合分析，能够有效评估 YOLOv8s-LCD 模型及其相关版本(YOLOv8s-ECA, YOLOv8s-GCNet, YOLOv8s-CondConv, 原始 YOLOv8s)在中草药目标检测任务中的检测精度、鲁棒性与稳定性。

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

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

$$mAP@0.5 = \frac{1}{N} \sum_{i=1}^N AP_i^{IoU=0.5}$$

$$mAP@0.5:0.95 = \frac{1}{10N} \sum_{i=1}^N \sum_{j=0}^9 AP_i^{IoU=0.5+0.05j}$$

4.3. 对比实验结果

本实验对比分析了 YOLOv8s-LCD 与其他相关改进模型在中草药检测任务中的性能表现，评价指标包括：准确率(Precision)、召回率(Recall)、mAP@0.5、mAP@0.5:0.95、Param(M)、Flops(G)以及 fps。实验结果如下表 2 所示：

Table 2. Comparison of detection performance of different models

表 2. 不同模型的检测性能对比

Model	Precision	Recall	map@0.5	map@0.5:0.95	Param(M)	Flops(G)	fps
YOLOv8s	0.815	0.791	0.842	0.724	11.2	28.7	110
YOLOv8s-LCD	0.839	0.831	0.879	0.766	8.2	24.0	98

续表

YOLOv8s-ECA	0.791	0.799	0.840	0.722	11.2	28.7	112
YOLOv8s-GCNet	0.798	0.806	0.847	0.726	11.3	28.9	106
YOLOv8s-CondConv	0.837	0.824	0.871	0.760	35.9	429.4	50

图 8 展示了 YOLOv8s-LCD 模型、原始 YOLOv8s 及其三种改进版本(YOLOv8s-CondConv、YOLOv8s-GCNet、YOLOv8s-ECA)在相同中草药图像上的检测结果。根据表 2 可见，各模型在中草药检测任务中的表现存在显著差异。本实验对 YOLOv8s-LCD 与其他相关改进模型在中草药检测任务中的表现进行了系统对比，结果如表 2 所示。从整体结果来看，原始 YOLOv8s 在精度与速度上均表现较为均衡，但在复杂背景与小尺度目标检测中存在漏检和置信度不足的问题。改进模型中，YOLOv8s-LCD 在四项精度指标上均取得最佳成绩(Precision 0.839、Recall 0.831、mAP@0.5 0.879、mAP@0.5:0.95 0.766)，同时其参数量(8.2M)和计算量(24.0G)也显著低于基线模型，尽管推理速度略降至 98 fps，但仍保持接近实时，体现了精度与效率的优良平衡。YOLOv8s-CondConv 在精度表现上接近 YOLOv8s-LCD (mAP@0.5 0.871、mAP@0.5:0.95 0.760)，但其参数量和计算量大幅增加(35.9M、429.4G)，导致推理速度骤降至 50 fps，限制了实际应用。YOLOv8s-GCNet 在 Recall 和 mAP 上优于基线模型(0.806、0.847)，并保持较快的速度(106 fps)，但对复杂背景下的细粒度特征刻画仍不足。YOLOv8s-ECA 尽管在推理速度上最快(112 fps)，计算开销几乎不变，但其精度指标整体低于基线 YOLOv8s，说明仅通道注意力不足以显著提升检测性能。综上，YOLOv8s-LCD 在多模块协同作用下实现了检测精度的显著提升，并兼顾较低的模型复杂度和实时性优势，相较其他改进模型在中草药小目标与复杂背景检测任务中更具实用性与推广价值。

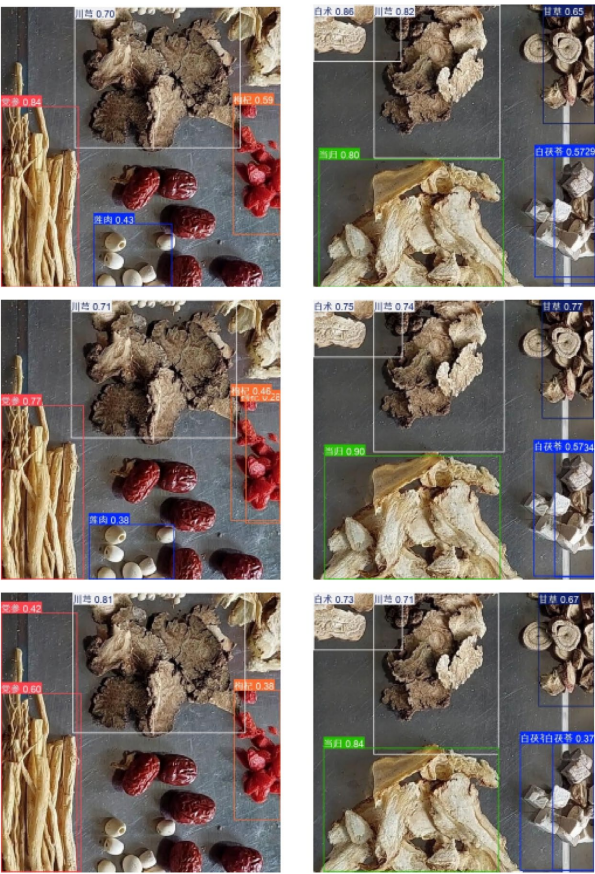




Figure 8. Visual comparison of detection results of different models
图 8. 不同模型的检测结果可视化对比

5. 讨论

本文针对中草药目标检测任务，围绕 YOLOv8s 在复杂背景及多尺度目标识别中的精度瓶颈和特征建模不足，提出了一种融合多注意力机制与大核卷积的改进方法。具体而言，本实验将大核卷积、坐标注意力与动态检测头等相结合，构建了 YOLOv8s-LCD 模型，并与 YOLOv8s-ECA、YOLOv8s-GCNet、YOLOv8s-CondConv 及原始 YOLOv8s 模型进行了对比实验。实验结果表明，YOLOv8s-LCD 在 Precision、Recall、mAP@0.5 及 mAP@0.5:0.95 四项指标上均取得最高成绩(0.839、0.831、0.879、0.766)，表明多模块融合能够显著提升模型的多尺度特征建模性能。模型的性能提升并非单一模块堆叠的直接结果，而是源于多机制在特征建模路径中的分工与耦合，针对中草药目标检测中“复杂背景、多尺度与细粒度相似”带来的识别瓶颈，本文提出的 YOLOv8s-LCD 并非简单通过堆叠注意力或卷积而获得提升，其优势应从模块间的分工与耦合机制来理解：一方面，大核卷积通过在训练期引入更大感受野和多分支表征、在推理期保持轻量化结构，增强了模型对长程上下文与形态先验的捕获能力，这一机制有助于在背景纹理与目标相似时提供额外的语义约束，从而抑制由局部噪声引发的错检；另一方面，坐标注意力将显式的位置信息注入通道选择过程，相当于为特征门控增加“软位置先验”，使模型在区分细长、片状或边界模糊的药材形态时既不丢失局部细节又能保持空间一致性；再次，DyHead 在多尺度特征融合阶段按内容自适应地重分配金字塔各层贡献，并在空间、通道与任务维度实施动态加权，这使得小尺度或被遮挡目标的信号在融合后被放大，从而改善召回与置信度校准。三者协同形成的“全局建模 - 位置感知 - 多尺度动态聚合”闭环既能抑制背景干扰又能保持或放大细粒度线索，解释了目标定位与小目标检出的同步改善。与之相比，仅做通道注意的 ECA 容易发生空间错配、侧重全局上下文的 GCNet 可能在汇聚时冲淡细节，而 CondConv 虽具更强的样本自适应能力但代价是显著的计算开销；因此本方法在表达力与部署效率之间取得了较优平衡。需要指出的是，该机制仍受限于极端密集遮挡与纹理级别的细微差异——在这些情形下，增加边界敏感的损失设计、引入超分辨率或分辨率自适应分支、以及采用跨域自训练/对抗适配来增强域泛化。YOLOv8s-LCD 模型在多注意力机制与结构增强模块的协同作用下，不仅提升了多尺度特征表达能力与定位精度，还兼顾了复杂背景下的稳定性，尤其适用于对高精度小目标检测要求较高的

中草药检测场景。未来可进一步结合更大规模、多样化的中草药数据集及高效特征增强方法，以提升模型的检测性能与泛化能力。

6. 结论

本实验结果表明，融合多注意力机制与大核卷积的 YOLOv8s-LCD 模型在中草药小目标检测中表现最优，其 Precision、Recall、mAP@0.5 和 mAP@0.5:0.95 分别达到 0.839、0.831、0.879 与 0.766，较原始 YOLOv8s 分别提升 2.4%、4.0%、3.7%和 4.2%，显著改善了检测精度与稳定性。YOLOv8s-CondConv 在 Precision、Recall、mAP@0.5 和 mAP@0.5:0.95 (0.837、0.824、0.871、0.760)方面相较于原始 YOLOv8s 也有提升，验证了动态卷积核在通道特征选择与小目标识别中的有效性，但较高的计算量限制了实时部署的可行性。YOLOv8s-GCNet 在 mAP@0.5 (0.847)接近原始模型，但 mAP@0.5:0.95 (0.726)相较于 YOLOv8s-LCD 和 YOLOv8s-CondConv 较低。YOLOv8s-ECA 整体性能低于 YOLOv8s-LCD、YOLOv8s-CondConv 与 YOLOv8s-GCNet，说明 ECA 模块的建模能力有限，缺乏空间注意力增强，存在过轻量化带来的“欠拟合”。综上所述，本文提出的 YOLOv8s-LCD 在多模块协同作用下兼顾了精度、鲁棒性与部署适应性，能够为中草药等小目标检测精度要求较高的应用场景提供有效技术支持。

参考文献

- [1] Liu, X.Y., Liang, J., Long, J.H., *et al.* (2025) Fine-Grained Chinese Herbal Medicine Image Classification Based on Feature Fusion and Channel Information Compensation. *Journal of Computer Applications*. <https://doi.org/10.11772/j.issn.1001-9081.2025050632>
- [2] Mao, Z., Yang, X., Wang, T., *et al.* (2025) Research on Intelligent Image Quality Assessment Method for Chrysanthemum Morifolium Ramat via Deep Color Feature Fusion. *Chinese Archives of Traditional Chinese Medicine*. <https://link.cnki.net/urlid/21.1546.R.20250514.1440.012>
- [3] Pan, Z.L., Wang, Y.S., You, Z. and Wang, Y.F. (2024) Chinese Herbal Medicine Recognition Based on Improved YOLOv5 Algorithm. *Technology of IoT & AI*, **56**, 64-68.
- [4] Zhang, C.H., Jiang, Y., Zhang, L. and Dai, J. (2024) Design and Implementation of Chinese Herbal Medicine Image Recognition Classification Model. *Journal of Fujian Computer*, **40**, 14-20. <https://doi.org/10.16707/j.cnki.fjpc.2024.05.003>
- [5] Dai, X., Chen, Y., Xiao, B., Chen, D., Liu, M., Yuan, L., *et al.* (2021). Dynamic Head: Unifying Object Detection Heads with Attentions. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 7369-7378. <https://doi.org/10.1109/cvpr46437.2021.00729>
- [6] Shen, J.Y., Ran, W.H., Huang, J.C., *et al.* (2025) Comparative Study on Rice Pest Detection Performance Using YOLOv5 Enhanced with Three C3 Attention Modules. *Computer Science and Application*, **15**, 41-49. <https://doi.org/10.12677/csa.2025.158196>