

基于改进YOLOv8的轻量化藏药材植物检测算法

罗志凌, 李东, 刘晓静

(青海大学计算机技术与应用学院, 青海 西宁 810016)

摘要: 由于藏药生长环境恶劣, 人工识别藏药材植物非常困难, 提出一种基于改进YOLOv8的轻量化检测算法LTP-YOLO检测户外环境中的藏药材植物。首先, 利用MobileViT替换YOLOv8特征提取网络来减少算法参数量和计算量。其次, 引入内容感知特征重组上采样算子CARAFE帮助算法在上采样时感知上下文信息。再次, 提出多尺度融合注意力机制MFA建立局部跨通道交互提升算法检测精度。实验表明, 所提算法将参数量从3.02 MB减少到1.28 MB, 计算量从8.2 GFLOPs减少到5.8 GFLOPs, 在自建的藏药植物图像数据集上mAP@.5相较于YOLOv8更优。证明了该算法可满足移动设备的高精度、低计算部署需求, 并在各种密集植物检测任务中显示出广阔的应用前景。

关键词: 藏药材植物; 深度学习; 目标检测; 轻量化算法; YOLOv8

DOI: 10.11907/rjdk.241753

开放科学(资源服务)标识码(OSID):

中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(2025)001-0086-07



Lightweight Detection Algorithm for Tibetan Medicinal Plants Based on Improved YOLOv8

LUO Zhiling, LI Dong, LIU Xiaojing

(School of Computer Technology and Application, Qinghai University, Xining 810016, China)

Abstract: Due to the harsh growth environment of Tibetan medicine, it is very difficult to manually identify Tibetan medicinal plants. A lightweight detection algorithm LTP-YOLO based on improved YOLOv8 is proposed to detect Tibetan medicinal plants in outdoor environments. Firstly, replace the YOLOv8 feature extraction network with MobileViT to reduce the number of algorithm parameters and computational complexity. Secondly, the introduction of the content aware feature recombination upsampling operator CARAFE helps the algorithm perceive contextual information during upsampling. Once again, a multi-scale fusion attention mechanism MFA is proposed to establish local cross channel interaction to improve the detection accuracy of the algorithm. The experiment shows that the proposed algorithm reduces the parameter size from 3.02 MB to 1.28 MB and the computational complexity from 8.2 GFLOPs to 5.8 GFLOPs on a self built Tibetan medicine plant image dataset. Compared to YOLOv8, the mAP@.5 of the proposed algorithm is superior, demonstrating its ability to meet the high-precision and low computational deployment requirements of mobile devices, and demonstrating broad application prospects in various intensive plant detection tasks.

Key Words: Tibetan medicinal plants; deep learning; object detection; lightweight algorithm; YOLOv8

0 引言

藏药作为我国历史丰富、极具影响的民族药之一, 因

其独特的生态环境具有重要的文化和医学价值。然而, 藏药材植物通常生长在青藏高原人烟稀少的地区, 因此人工识别工作难度较大。为了解决上述问题, 本文提出一种量化的藏药材植物检测方法, 并尝试将其部署在移动设备

收稿日期: 2024-09-02

扫描二维码阅读全文:



基金项目: 国家自然科学基金项目(62366043); 青海省科技厅2023年十大国家级创新平台项目子课题(ZYYS DPT-2023-02)

作者简介: 罗志凌(1995-), 男, 青海大学计算机技术与应用学院硕士研究生, 研究方向为深度学习与目标检测; 李东(1987-), 男, 博士, 青海大学计算机技术与应用学院副教授、硕士生导师, 研究方向为神经网络和医学图像处理; 刘晓静(1980-), 女, 硕士, 青海大学计算机技术与应用学院教授、硕士生导师, 研究方向为计算机视觉与神经网络。本文通讯作者: 刘晓静。

上便捷、高效地识别藏药材植物,有效推动了藏药知识文化的传播。

近年来,随着深度学习理论的不断更新,卷积神经网络在图像特征提取方面表现优异,基于深度学习的目标检测技术已在植物检测场景中成为主流,主要分为以 Faster R-CNN 为代表的双阶段算法与以 YOLO 为代表的单阶段算法。其中,单阶段算法相较于双阶段算法省去了候选框的生成与预筛选步骤,在检测速度与效率上表现更好。

在植物检测算法的研究上,尽管已有文献提出的算法取得了较好的检测精度,但大多数算法仅在实验室中进行测试,在野外环境和多目标环境的实验较少。此外,目前主流的目标检测算法结构复杂、计算量较大,难以部署到算力有限的移动设备中,从而限制了该技术在野外植物检测场景中应用。

藏药因其独特的生态环境与文化差异,鲜有文献针对藏药材植物进行目标检测。为此,本文基于目前主流的单阶段目标检测算法 YOLOv8,为算力有限的移动检测设备研制出一种具有良好精准度的轻量化的藏药材植物检测算法 LTP-YOLO。首先,比较多个轻量化特征提取网络,从中选取检测效果最佳的 MobileViT 来替换 YOLOv8 中的特征提取网络以减少算法参数量和计算量,使其在无显著精度损失的情况下能部署在算力有限的移动设备上。其次,引入上采样内容感知特征重组网络结构(Content-Aware ReAssembly of Feature, CARAFE)使算法在更大的感受野上感知上下文信息来提升算法检测精度,并通过极少的计算开销生成实时自适应卷积核,从而实现特定内容的感知处理。最后,本文提出多尺度融合注意力机制(Multi-scale Fusion Attention, MFA),使算法在没有通道降维的情况下建立局部跨通道交互,以提升算法在复杂背景下检测目标的精度。

本文主要贡献有以下几点:①使用数据增强技术搭建了由 43 种常见藏药材植物共 10 940 张图像组成的藏药材植物图像数据集,为相关研究提供了关键的数据支撑;②将 YOLOv8 的特征提取网络替换为轻量化网络 MobileViT,极大减少了算法参数量和计算量,使其能更方便地部署在算力受限的移动检测设备中;③引入上采样内容感知特征重组网络 CARAFE,并通过多尺度融合注意力机制 MFA 提升算法检测精度,使本文算法在藏药材植物检测任务中的准确率不低于 YOLOv8 算法。

1 相关工作

目前,植物识别分类和目标检测算法主要分为传统图像处理 and 深度学习。传统图像处理算法通过相应的数学方式得到植物颜色、形状和纹理等特征描述符以获得分类结果,但由于其人为成本较高且精度难以达到预期水平,因此无法满足实际应用需求。随着机器学习技术发展和

相关硬件能力提升,基于深度学习的目标检测算法为植物的检测识别提供了快速有效的解决方案,例如以 Faster R-CNN 为代表的双阶段算法、以 YOLO 为代表的单阶段算法^[1-2]。其中,YOLO 等单阶段算法相较于双阶段算法省去了候选框生成与预选操作,在略微牺牲准确度的情况下大幅提升了检测速度,但因其较复杂的结构和较大参数量、计算量,导致该模型难以部署到算力有限的移动设备上。

为此,不少学者尝试在保证较高目标检测性能的前提下,对 YOLO 系列进行轻量化改进。例如,胡俊峰等^[3]使用部分卷积替换 YOLOv8 中的 3×3 卷积,以解决网络冗余和计算量复杂的问题。李小祥等^[4]使用 Bottleneck Transformer 替换 YOLOv8 中的 C2f 模块,在降低开销的同时提升模型特征提取能力。同时,在自然背景下进行植物检测也是一项具有挑战性的任务。朱永宁等^[5]使用 Faster R-CNN 算法识别枸杞开花和结果阶段的 mAP 为 74%。Zhou 等^[6]在 YOLOv7 模型中引入 RepGhost 网络与 ECA 注意力机制提出一种改进的火龙果检测算法。针对因植物分布密集、互相遮挡而导致的漏检问题,Bodla 等^[7]提出 Soft-NMS 方法,采用一种衰减的算法检测所有预测框,即在衰减重叠度高的其他预测框得分时保留预测框而非直接删除。针对密集植物场景尺度变化明显而导致的小目标漏检问题,Duan 等^[8]提出基于 Anchor-free 的 CenterNet 算法省去边框回归操作,直接输出目标中心点坐标及宽高等信息。

然而,当前植物识别主流研究方向主要在特定场景下检测分类常见植物,很少有在野外场景中检测藏药材植物。实验表明,尽管当前表现最佳的网络模型在简单背景下的 mAP 能达到 90% 以上,但在野外环境中由于植物分布密集且受到互相遮挡、光照等因素会导致降低模型泛化能力、检测精度和性能。为了解决以上问题,本文提出一种基于改进 YOLOv8 的轻量化藏医药植物检测方法 LTP-YOLO,整体网络结构如图 1 所示(彩图扫 OSID 码可见,下同)。

2 LTP-YOLO 方法

2.1 轻量化特征提取网络

MobileViT 是苹果公司 2021 年发表的轻量级骨干网络,由普通卷积、MV2、MobileViT block 组成,网络结构如图 2 所示。具体为,在 Transformer 模型中引入卷积神经网络模块解决传统 Vision Transformer(ViT)模型算力要求高、缺少空间归纳偏置和模型训练困难等问题^[9-11]。MobileViT 结合了 CNN 与 Transformer 的优点,克服了 Transformer 网络结构复杂、计算量大的问题,且性能相较于现有轻量级 CNN 网络更好。

MV2 模块指 MobileNetV2 中的倒残差结构。在倒残差结构中,特征图首先经过 1×1 卷积提升维度以引入更多特

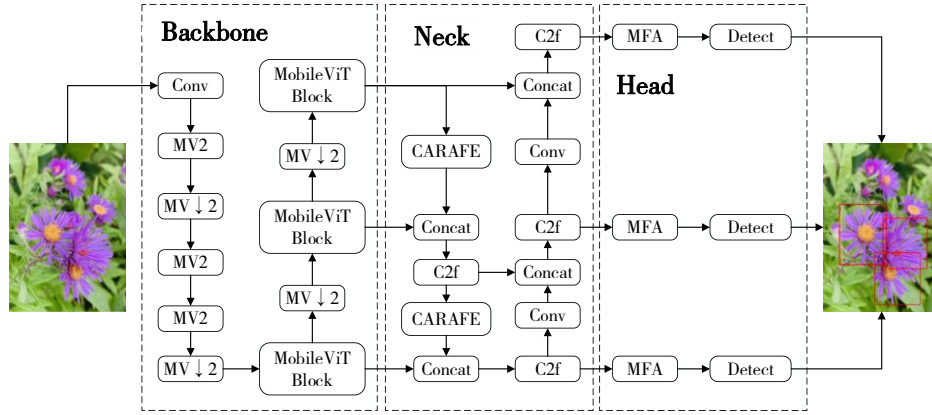


Fig. 1 Network structure of LTP-YOLO algorithm

图1 LTP-YOLO算法网络结构

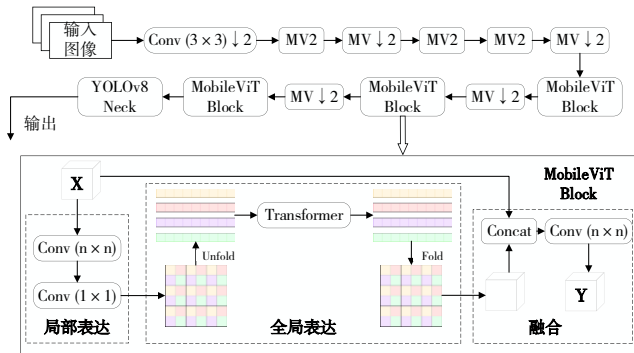


Fig. 2 Network structure of MobileViT

图2 MobileViT网络结构

征表达,因为将更高维度的特征图输入激活函数后只会损失一部分信息^[12];其次使用深度可分离卷积进行特征提取和降维,↓2表示在特征提取时使用步长为2的深度卷积进行下采样操作;最后在倒残差结构中通过通道扩张与引入深度可分离卷积在轻量级网络中实现高效的特征提取。

MobileViT Block 由特征局部表达、全局表达和融合 3 部分组成,旨在通过更少的参数建模输入特征图中的局部和全局特征。在特征局部表达部分,输入 MobileViT Block 的特征图经过一个 $n \times n$ 的卷积块,其中卷积块的尺寸 n 远小于特征图尺寸,之后经过一个 1×1 的卷积块进行通道升维。在全局表达部分,特征图被划分为 N 个 patch(见图 3),在进行自注意力计算时每个 patch 中的 token 只与其他 patch 中相同位置的 token 进行自注意力计算。这样操作的原因是图像数据自身存在大量的数据冗余(一张图像的每个像素点与其周围像素点非常近似,且图像分辨率越高时差异越小),因此 MobileViT 进行自注意力计算时可忽略一些像素点。在全局表达部分,Unfold、fold 负责将特征图重新排列为计算自注意力所需的向量格式。其中,Unfold 将特征图中每个 patch 中位置相同的 token 拼成二维向量序列送入 Transformer 进行建模;Fold 则将计算完成的向量重新调整为原始形状。在融合部分,首先将经过局部、全局表达的特征图与原始特征图进行拼接,其次使用 $n \times n$ 卷积

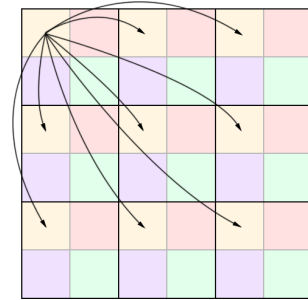


Fig. 3 Self attention mechanism in MobileViT

图3 MobileViT 中的自注意力机制

块进行降维以维持通道数不变,并将其送入下一模块。

2.2 CARAFE 上采样算子

在 YOLOv8 目标检测算法中特征聚合网络采用最邻近插值上采样方法,但该方法仅基于像素空间位置来确定上采样核,未充分利用特征信息。为了解决这一问题,WANG 等^[13]提出具有较大的感受野的内容感知特征重组 CARAFE 上采样算子,以更好的利用周围特征信息生成基于输入内容的特定上采样核,从而有效提升算法识别精度。特别是,CARAFE 上采样算子仅引入了少量参数和计算量,从而保证了模型的轻量化。

CARAFE 上采样算子分为分别为上采样核预测模块和特征重组模块,网络结构如图 4 所示。上采样核预测模块负责输入特征图 X 的语义信息,以生成基于特征内容的特定上采样核。对于形状为 $H \times W \times C$ 的输入特征图,CARAFE 首先利用 1×1 卷积将通道数压缩到 C_m 以减小后续步骤的计算量。例如,将上采样倍数为设置 σ 、上采样核尺寸设置为 $k_{up} \times k_{up}$ (更大的上采样核具有更大的感受野,但会引入更多参数),则对于压缩后尺寸为 $H \times W \times C_m$ 输入特征图,CARAFE 首先使用卷积层预测上采样核得到通道数为 $\sigma^2 k_{up}^2$ 的特征图;其次将输出图在通道维度展开得到形状为 $\sigma H \times \sigma W \times k_{up}^2$ 的上采样核 Y 。

内容感知重组模块负责利用得到的上采样核进行特征重组,以提取目标特征并完成上采样操作。对于每个重组内核 Y_i ,内容感知重组模块通过加权求与运算重组区域内特征,具体运算操作如式(1)所示。

$$X'_l = \sum_{n=-r}^r \sum_{m=-r}^r Y_{l'(n,m)} \cdot X_{(i+n, j+m)} \quad (1)$$

式中: $r = \frac{k_{up}}{2}$ 。

对于目标位置 l' , 以 $l=(i, j)$ 为中心的相应正方形区域 $N(X_l, k_{up})$ 中的每个像素, 基于特征内容而非位置的距离对上采样像素 l' 作出不同贡献。最后, 经过内容感知重组模块后可得到形状为 $\sigma H \times \sigma W \times C$ 的输出特征图 X' 。

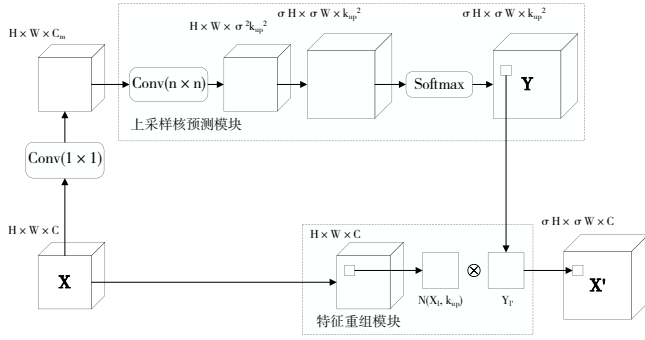


Fig. 4 Network structure of CARAFE upsampling operator

图 4 CARAFE 上采样算子网络结构

2.3 MFA 注意力机制

为了充分利用输入特征图中丰富的语义特征, 本文提出可有效融合藏药材植物图像特征的多尺度融合注意力机制 MFA (见图 5)。MFA 在 CA 注意力机制的基础上设计了一个用于建立短期和长期依赖关系的多尺度并行子网络, 并通过跨空间学习方法融合两个并行子网络的输出特征图。

首先, 将给定的输入特征图在通道维度上划分为 G 个子特征图以学习不同语义。其次, 将分组特征图分别送入并行的 1×1 和 3×3 分支来提取注意力权重特征图。在 1×1 分支中通过两个 1D 全局平均池化操作沿着水平、垂直空间方向对特征图进行编码, 具体公式如式 (2)、式 (3) 所示。再次, 与 CA 注意力类似将两个编码特征在通道维度上进行拼接, 并使用 1×1 卷积进行权重学习。最后, 将学习到的权重分解为两个向量后输入 Sigmoid 函数, 通过乘法将

两个注意力图聚合在一起。在 3×3 分支中, 仅堆叠单个 MV2 模块对子特征图进行升维, 然后通过 3×3 卷积以较大的局部感受野收集更多空间信息。

$$z_c^H(H) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(H, i) \quad (2)$$

$$z_c^W(W) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, W) \quad (3)$$

式中: C 表示输入通道的数量; H 、 W 表示输入特征的空间维度; x_c 为第 c 个通道处的输入特征。

跨空间学习模块将不同空间、维度方向的信息进行聚合, 以实现更丰富的特征融合。首先, 跨空间学习模块使用 2D 全局平均池化将来自 1×1 分支和 3×3 分支输出的全局空间信息进行编码。其中, 1×1 分支的输出图在池化前需要经过 GroupNorm 操作, 2D 全局池化操作如式 (4) 所示。

$$z_c = \frac{1}{H \times W} \sum_j \sum_i x_c(i, j) \quad (4)$$

首先, 通过 2D 全局平均池化操作编码全局信息, 并对长程依赖性进行建模; 其次, 使用 Softmax 函数拟合线性变换并分别与另一条分支的输出图进行矩阵点积运算, 以融合不同空间和维度方向的特征信息; 最后, 将两条分支得到的空间注意力图相加得到最终的注意力权重特征图。

3 实验结果与分析

3.1 实验平台与评价指标

本文操作系统为 Linux, CPU 为 Intel Xeon CPU E5-2603 v4 @ 1.70 GHz, 内存为 16 GB, 4 块 Nvidia GTX1080Ti 显卡以分布式数据并行模式进行多卡分布式训练, 深度学习框架为 Pytorch-GPU 1.13.0, CUDA 版本为 11.6, 实验批量大小设置为 32, 训练周期为 500 个 epoch, 输入图像尺寸为 640×640 。

本文采用平均精度均值 mAP 评估算法检测性能, mAP@.5 表示当 IoU 阈值设置为 0.5 时所有类别的平均检测精度, mAP@[.5:.95] 表示不同 IoU 阈值 (0.5~0.95, 步长为 0.05) 上的平均 mAP, n 表示检测到的类别总数。mAP 计

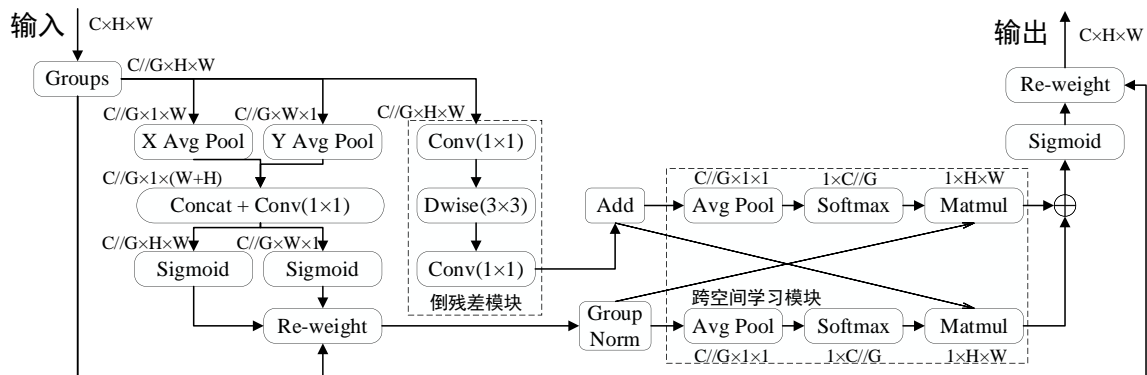


Fig. 5 Network structure of MFA mechanism

图 5 MFA 注意力机制网络结构

计算公式如式(5)、式(6)所示。

$$AP = \int_0^1 P(R) dr \tag{5}$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \tag{6}$$

3.2 数据集搭建

传统藏药植物主要分布在人口稀少的青藏高原地区,受阳光、时间和天气等因素影响较大,为手工收集植物图像带来了一定的挑战。因此,本文首先采用网络爬虫技术,从中国植物图像库网站收集藏药植物图像,共获得43种不同藏药植物23 525张;然后基于专业知识对图像进行手动筛选以消除重复和图像质量差的图像;最后使用Labelimg标注工具对藏药材植物进行标注。本文经过上述数据清理操作后的得到包括党参、紫菀等43种藏药材植物共8 084张图片,为构建藏药植物图像数据集提供了关键的数据支持。

由于部分藏药植物物种十分稀有,因此数据集中存在标签数量分布不均的情况,为此采用镜像、旋转和噪声添加等技术对标签数量较少的藏药材图像进行数据增强,以缓解数据集标签数量分布不均的问题。数据增强后数据集共包括43种藏药植物10 904张,按照7:1:2的比例随机划分训练集、验证集和测试集,分别为7 633张、1 090张和2 181张。图6为数据增强后每种植物物种标签数量的变化情况。

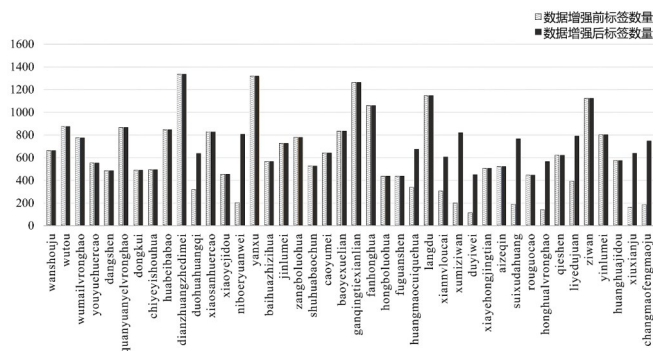


Fig. 6 Number of labels after data augmentation

图6 数据增强后的标签数量

3.3 实验结果

3.3.1 轻量化网络比较

为了验证不同轻量化特征提取网络在藏药材植物目标检测任务上的性能,选取FasterNet、ShuffleNet、MobileNet和MobileViT等为代表的主流轻量化网络进行比较,实验结果如表1所示。由此可知,大部分轻量化特征提取网络随着参数量、计算量减少,检测精度明显下降。以华为提出的VanillaNet为例,其参数量、计算量分别为1.17 MB和3.5 GFLOPs,相较于基准分别下降61.3%和57.3%,检测精度下降也十分严重,mAP@.5、mAP@[.5:.95]分别下降6.4%和9.2%,难以胜任藏药材植物的检测工作。MobileNetV3、EfficientNet和MobileViT-xxs等网络虽然在参数量、

计算量上有所减少,但引入神经网络后的计算开销对通用硬件不够友好,导致其FPS有所下降。例如,MobileNetV3引入了深度可分离卷积,大量1×1卷积会频繁访问内存而降低运行速度;MobileViT因Transformer模块引入了大量自注意力运算,导致其在FPS上的表现较差。

本文综合考虑上述性能指标后选择MobileViT-xxs作为藏药材植物检测模型的特征提取网络,原因为MobileViT-xxs的参数量、计算量分别为基准模型的41.3%和69.5%时,其mAP@.5、mAP@[.5:.95]只下降了0.5%和1.6%,对藏药材植物仍具有良好的检测精度。需要注意的是,虽然MobileViT-xxs的检测精度优于基准模型,能满足野外场景下对目标进行实时推理检测的要求,但引入了大量的计算开销会导致FPS下降明显,当FPS下降到63时部署到算力受限的移动设备中将难以满足野外实时推理的需求。

Table 1 Experimental results of different lightweight feature extraction networks

表1 不同轻量化特征提取网络实验结果

骨干网络	mAP@.5/%	mAP@[.5:.95]/%	Parameter/10 ⁶	GFLOPs	FPS
DarkNet53	85.7	70.1	3.02	8.2	204
FasterNet ^[14]	83.8	66.6	1.76	5.1	232
ShuffleNetV2 ^[15]	79.9	62.7	1.72	5.0	250
MobileNetV3 ^[16]	82.3	65.0	2.36	5.8	164
VanillaNet ^[17]	79.3	60.9	1.17	3.5	278
EfficientNet ^[18]	83.5	66.8	1.92	5.7	176
MobileViT-xxs ^[19]	85.2	68.5	1.25	5.7	88
MobileViT-xs	87.8	72.5	2.22	13.1	63

3.3.2 注意力机制比较

为了验证不同注意力机制对藏药材植物检测精度的影响,选取多种主流注意力机制进行实验,结果如表2所示。由此可知,注意力机制可通过自适应分配权重使算法关注目标区域以减少背景噪声干扰,从而提升算法检测准确性和效率。全局注意力机制(Global Attention Mechanism, GAM)、多尺度融合注意力机制(Multi Scale Fusion Attention Mechanism, MFA)在藏药材植物检测任务上具有良好表现,GAM虽然在mAP@[.5:.95]上达到了70.6%,意味着其可使算法具有更好的精度和鲁棒性,但参数量、计算量分别达到3.59 MB和9.3 GFLOPs,与本文研发轻量化检测算法的目的相背。为此,在特征融合网络中的3个检测分支后使用MFA注意力机制,以指导算法检测更重要的目标区域。

为了深入探索算法的可解释性,使用Grad-CAM方法生成梯度可视化热力图(见图7),以直观观察算法在目标检测时的注意力分布情况及模型对不同区域的关注程度^[30]。由此可见,本文算法在任务关键区域更集中,能准确识别、定位植物目标所在区域,而YOLOv8、CBAM等其他注意力机制在关注点上的表现较为分散,精确性较差。

Table 2 Experimental results of different attention mechanisms

表 2 不同注意力机制实验结果

注意力机制	mAP@.5/%	mAP@[.5:.95]/%	Parameter/ 10^6	GFLOPs
-	85.7	70.1	3.02	8.2
SE ^[20]	85.5	69.3	3.03	8.2
CBAM ^[21]	85.1	68.9	3.10	8.3
ECA ^[22]	85.7	69.9	3.02	8.2
SA ^[23]	85.8	69.9	3.02	8.2
CA ^[24]	85.6	69.5	3.03	8.3
SimAM ^[25]	84.6	67.7	3.02	8.2
Triplet ^[26]	85.2	68.5	3.02	8.4
GAM ^[27]	85.9	70.6	3.59	9.3
NAM ^[28]	85.6	69.5	3.02	8.2
SGE ^[29]	85.2	69.6	3.02	8.2
MFA(Ours)	86.2	70.2	3.03	8.3

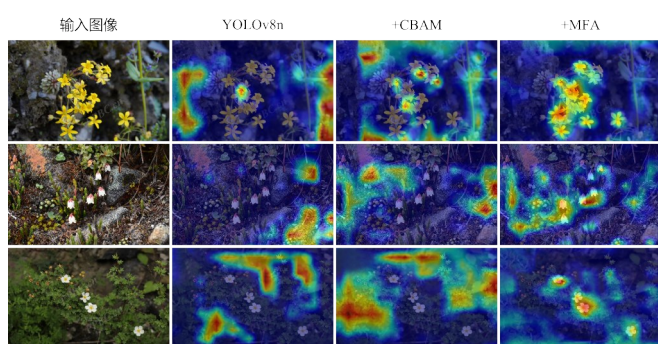
**Fig. 7** Grad-CAM heatmap with different attention mechanisms

图 7 不同注意力机制的 Grad-CAM 热力图

3.3.3 消融实验

为了验证本文算法各改进模块的有效性,以 YOLOv8n 算法为基准,通过多个改进模块的不同组合进行消融实验,结果如表 3 所示。

Table 3 Module ablation experiment

表 3 模块消融实验

MobileViT	CARAFE	MFA	mAP@.5/%	mAP@[.5:.95]/%	Parameter/ 10^6	GFLOPs	FPS
-	-	-	85.7	70.1	3.02	8.2	204
√	-	-	85.2	68.5	1.25	5.7	88
-	√	-	86.3	70.6	3.15	8.7	185
-	-	√	86.2	70.2	3.03	8.3	176
√	-	√	85.4	68.7	1.25	5.7	85
-	√	√	86.7	70.7	3.20	9.0	154
√	√	√	85.8	69.8	1.28	5.8	84

由表 3 可知,将 YOLOv8n 原始算法的骨干网络替换为 MobileViT-xxs 网络后算法参数量、计算量、精度均明显降低,尤其是 mAP@[.5:.95] 下降了 1.6%; 分别加入 CARAFE、MFA 模块后算法检测精度、参数量、计算量均小幅增加,特别是 CARAFE 引入的额外参数量、计算量较为明显,原因为 CARAFE 上采样针对不同特征层采用不同的上采样核,相较于传统插值上采样更注重特征全局信息,而非仅关注局部信息,因此相应的参数量、浮点计算数均有

所增加。

综上,本文改进后的轻量化检测算法相较于原始 YOLOv8n 基准算法,在参数量、计算量分别下降 57.6%、29.2% 的情况下,依旧保持着良好的检测精度,mAP@[.5:.95] 相较于基准算法仅降低 0.3%,而 mAP@.5 还提升了 0.1%。虽然在 FPS 指标上存在明显降低,但依然能满足野外环境下实时检测的需求。

此外,为了直观展示本文算法在检测性能上的优势,将其与 YOLOv8n、以 FasterNet 作为骨干网络的 YOLOv8n 及 LTP-YOLO 算法在藏药材植物中代表性的图像中进行比较,实验结果如图 8 所示。其中,图 8(a) 为匙叶翼手花;图 8(b) 为甘青铁线莲;图 8(c) 为小伞虎耳草;图 8(d) 为岩须。

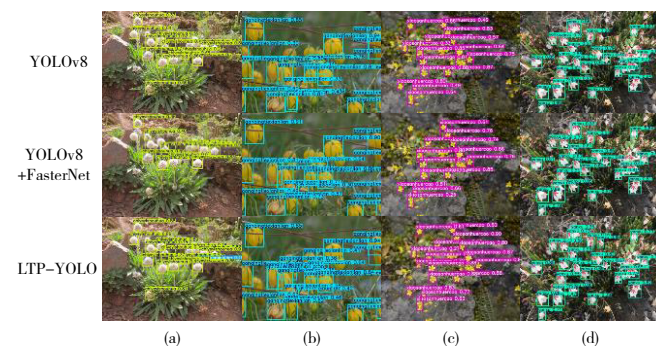
**Fig. 8** Detection effect of different algorithms on various Tibetan medicinal plants

图 8 不同算法对各种藏药材植物的检测效果

由图 8 可见,尽管使用 FasterNet 作为骨干网络的 YOLOv8n 在降低参数量和提升 FPS 方面表现良好,但在检测精度方面表现欠佳,尤其在检测匙叶翼手花、甘青铁线莲和小伞虎耳草时存在明显的漏检情况。LTP-YOLO 总体上保持了与 YOLOv8n 算法相同的检测精度,在某些情况下检测性能甚至更好,尤其在检测甘青铁线莲时漏检情况更少,但在一些特殊情况下仍会出现漏检现象。

4 结语

本文为移动检测设备提出了一种轻量化藏药材植物目标检测算法。首先将 YOLOv8n 的特征提取网络替换为 MobileViT-xxs,使算法参数量、计算量分别减少 1.74 MB 和 2.4 GFLOPs,分别为基准算法的 42.3% 和 70.7%。其次,在本文算法中引入 CARAFE 上采样算子和 MFA 注意力机制,使其达到了与原本 YOLOv8 相当检测精度。实验表明,所提算法在藏药材植物检测任务中表现良好,在相关任务场景中具有广阔的应用前景。

尽管本文算法在藏药材植物图像数据集上展现出良好的检测精度,但仍存在漏检现象,且算法的 FPS 相较于 YOLOv8n 算法下降明显。未来,将进一步提升算法检测精度和 FPS 性能,旨在让该算法成为户外环境中检测密集藏药材植物对象的高效解决方案。

参考文献:

- [1] REN S, 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*, 2016, 39(6): 1137–1149.
- [2] 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.
- [3] HU J F, LI B C, ZHU H, et al. Improved YOLOv8 lightweight UAV target detection algorithm [J]. *Computer Engineering and Applications*, 2024, 60(8): 182–191.
胡峻峰, 李柏聪, 朱昊, 等. 改进YOLOv8的轻量化无人机目标检测算法[J]. *计算机工程与应用*, 2024, 60(8): 182–191.
- [4] LI X X, ZHANG J, QIN K B, et al. A lightweight method for apple leaf disease detection in complex environments based on improved YOLOv8 [J/OL]. *Journal of Nanjing Agricultural University*, 2024–08–29. <http://kns.cnki.net/kcms/detail/32.1148.s.20240619.0952.002.html>.
李小祥, 张洁, 秦柯贝, 等. 基于改进YOLOv8的轻量化复杂环境苹果叶片病害检测方法[J/OL]. *南京农业大学学报*, 2024–08–29. <http://kns.cnki.net/kcms/detail/32.1148.s.20240619.0952.002.html>.
- [5] ZHU Y N, ZHOU W, YANG Y, et al. Automatic identification technology of lyceum barbarum flowering period and fruit ripening period based on Faster R-CNN [J]. *Chinese Journal of Agrometeorology*, 2020, 41(10): 668–677.
朱永宁, 周望, 杨洋, 等. 基于Faster R-CNN的枸杞开花期与果实成熟期识别技术[J]. *中国农业气象*, 2020, 41(10): 668–677.
- [6] ZHOU J, ZHANG Y, WANG J. Rde-YOLOv7: an improved model based on YOLOv7 for better performance in detecting dragon fruits [J]. *Agronomy*, 2023, 13(4): 1042.
- [7] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS—improving object detection with one line of code [C]// *Proceedings of the IEEE International Conference on Computer Vision*, 2017: 5561–5569.
- [8] DUAN K W, BAI S, XIE L X, et al. CenterNet: keypoint triplets for object detection [C]// *Proceedings of the IEEE International Conference on Computer Vision*, 2019: 6569–6578.
- [9] HAN K, WANG Y H, CHEN H T, et al. A survey on vision transformer [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(1): 87–110.
- [10] ALZUBAIDI L, ZHANG J L, HUMAIDI A J, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions [J]. *Journal of Big Data*, 2021, 8: 1–74.
- [11] KHAN S, NASEER M, HAYAT M, et al. Transformers in vision: a survey [DB/OL]. <https://arxiv.org/abs/2101.01169>.
- [12] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetv2: inverted residuals and linear bottlenecks [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 4510–4520.
- [13] WANG J Q, CHEN K, XU R, et al. Carafe: content-aware reassembly of features [C]// *Proceedings of the IEEE International Conference on Computer Vision*, 2019: 3007–3016.
- [14] CHEN J R, KAO S, HE H, et al. Run, don't walk: chasing higher flops for faster neural networks [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2023: 12021–12031.
- [15] MA N N, ZHANG X Y, ZHENG H T, et al. Shufflenetv2: practical guidelines for efficient cnn architecture design [C]// *Proceedings of the European Conference on Computer Vision*, 2018: 116–131.
- [16] HOWARD A, SANDLER M, CHU G, et al. Searching for mobileNetv3 [C]// *Proceedings of the IEEE International Conference on Computer Vision*, 2019: 1314–1324.
- [17] CHEN H T, WANG Y H, GUO J Y, et al. VanillaNet: the power of minimalism in deep learning [DB/OL]. <https://arxiv.org/abs/2305.12972>.
- [18] TAN M X. EfficientNet: rethinking model scaling for convolutional neural networks [DB/OL]. <https://arxiv.org/abs/1905.11946>.
- [19] MEHTA S, RASTEGARI M. Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer [DB/OL]. <https://arxiv.org/abs/2110.02178>.
- [20] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141.
- [21] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]// *Proceedings of the European Conference on Computer Vision*, 2018: 3–19.
- [22] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020: 11534–11542.
- [23] ZHANG Q L, YANG Y B. SA-net: shuffle attention for deep convolutional neural networks [C]// *ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2021: 2235–2239.
- [24] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021: 13713–13722.
- [25] YANG L X, ZHANG R Y, LI L D, et al. Simam: a simple, parameter-free attention module for convolutional neural networks [C]// *International Conference on Machine Learning*, 2021: 11863–11874.
- [26] MISRA D, NALAMADA T, ARASANIPALAI A U, et al. Rotate to attend: convolutional triplet attention module [C]// *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, 2021: 3139–3148.
- [27] LIU Y C, SHAO Z R, HOFFMANN N. Global attention mechanism: retain information to enhance channel-spatial interactions [DB/OL]. <https://arxiv.org/abs/2112.05561>.
- [28] LIU Y C, SHAO Z R, TENG Y Y, et al. NAM: normalization-based attention module [DB/OL]. <https://arxiv.org/abs/2111.12419>.
- [29] LI X, HU X L, YANG J. Spatial group-wise enhance: improving semantic feature learning in convolutional networks [DB/OL]. <https://arxiv.org/abs/1905.09646>.
- [30] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization [J]. *International Journal of Computer Vision*, 2020, 128: 336–359.

(责任编辑:刘嘉文)