光电智能识别与感知

DOI：10.16818/j.issn1001－5868.20260426001

复杂茶园场景的轻量化茶芽采摘点检测模型

张澎涛，于　帅，黄建平，张宪奇，张佳薇*

（东北林业大学 计算机与控制工程学院，哈尔滨 150040）

摘　要： 针对光电成像中微小、遮挡、高速目标的实时定位难题，传统姿态估计模型在保持微小关键点特征与轻量化推理之间难以平衡。文章以复杂茶园中的茶芽采摘点定位为典型验证场景，提出一种基于改进YOLOv8n-Pose的轻量化姿态估计模型。首先，设计并行化补丁感知注意力模块，增强网络对细微连接处特征的捕获能力；其次，构建多源特征融合模块与递进式双向循环颈部架构，实现全尺度定位特征的高效互联与互补增强；最后，引入基于交并比加权的样本动态加权损失，提升模型对遮挡、模糊等困难样本的学习关注度。在自建数据集上的实验结果表明，关键点检测平均精度均值达79.1%，参数量仅为2.59 M，计算量为8.3 GFLOPs。经TensorRT加速后，该模型在边缘设备上实现106.84 FPS的推理速度。

关键词： 光电成像；目标定位；姿态估计；轻量化；关键点检测

中图分类号： TP391.41　　**文章编号：** 1001－5868(2026)05－0827－10

A Lightweight Detection Model for Tender Tea Shoot Picking Points in Complex Tea Garden Scenes

ZHANG Pengtao, YU Shuai, HUANG Jianping, ZHANG Xianqi, ZHANG Jiawei

(**College of Computer and Control Engineering, Northeast Forestry University, Harbin 150040, CHN**)

Abstract: To address the common challenge of real-time, high-precision localization of tiny, occluded, and fast-moving targets in photoelectric imaging systems, conventional pose estimation models often fail to balance the detection of subtle features with lightweight deployment. This study takes the picking point localization of tender tea shoots in complex tea gardens as a representative scenario, and a lightweight pose estimation model based on an improved YOLOv8n-Pose is proposed. Firstly, a parallelized patch-aware attention module is designed to enhance the capture of delicate features at junctions. Secondly, a multi-source feature fusion module and a progressive bidirectional recurring neck architecture are constructed for efficient full-scale feature interconnection. Finally, a slide-weighted loss function based on IoU is introduced to focus learning on hard samples. Experiments on a custom dataset demonstrated a keypoint mAP of 79.1%, with only 2.59 M parameters and 8.3 GFLOPs. Deployed on an edge device with TensorRT, the model achieved 106.84 FPS.

Key words: photoelectric imaging; target localization; pose estimation; lightweight; keypoint detection

收稿日期：2026－04－26.

* **通信作者：**张佳薇

0 引言

在工业视觉检测、自动驾驶与精准农业等光电成像应用领域，系统需在非结构化环境中对微小、运动、易遮挡目标实现毫米级定位[1]。然而，传统目标检测仅提供边界框，其粗粒度信息无法满足精密抓取或交互的需求。姿态估计虽能给出关键点坐标，但主流高精度模型多依赖庞大网络架构，面临微小目标特征丢失、空间定位偏移及边缘部署算力受限等问题[2]。姿态估计的核心难题在于：高精度特征提取依赖深层语义，而微小目标定位依赖浅层细节，这些细节在网络下采样过程中极易丢失；同时，对于复杂背景与光照变化所产生的困难样本，常规损失函数难以充分挖掘其潜在信息。对于非结构化茶园测试环境，茶芽采摘点的直径仅数毫米，与老叶、枝干严重交叠，且户外光照变化剧烈。在此场景下突破微小目标定位与模型轻量化瓶颈，对推动光电成像系统中通用鲁棒关键点检测技术的发展具有重要意义。

针对关键点检测，技术路线主要分为基于热力图与基于回归两类。热力图类虽空间泛化能力强，但其受限于固有量化误差和高计算开销。回归类端到端地预测坐标，具备亚像素精度与轻量化优势，更符合边缘部署需求。其中，YOLO-Pose等回归范式在农业采摘领域已初步验证其有效性，Shuai等[3]以YOLOv5为骨干网络构建改进型YOLO-Tea模型，通过引入上采样算子并融合CBAM卷积注意力机制，有效增强了特征提取能力。在自建数据集上的实验结果表明，该模型成功实现了茶芽的6个关键点的精准定位，其平均精度较基准模型提升了5.26%。在茶树嫩芽三维空间定位方面，Wang等[4]提出ECS-Tea算法，基于YOLOv11-Pose结合生物启发式优化策略，实现了茶芽的高精度检测与定位，在复杂光照与遮挡条件下保持良好的鲁棒性。Wang等[5]则从系统集成角度出发，融合YOLO网络与Mamba图像分割技术，采用单目相机俯视视角完成茶芽的自动采摘作业，采摘成功率显著提升。这些研究为姿态估计驱动的采摘点定位提供了多样化技术路径。Chen等[6]针对采摘点定位问题，提出一种基于深度卷积神经网络的定位方法，通过构建包含上、下采样网络的FCN识别采摘模块对Faster R-CNN模型进行改进，使改进后网络的定位准确率达到84.91%，较改进前显著提升了5.91%。此外，Chen等[7]探索了多算法融合策略，提出一种结合YOLO-v3算法、语义分割、骨架提取及最小外接矩形的采摘点定位方法，并据此设计了基于PC端与微控制器协同控制的智能末端执行器，实验显示其采摘点提取精度达83%，实际采摘成功率为80%。为进一步提升三维空间内的定位能力，Chen等[8]提出一种融合深度相机与改进YOLOv5模型的茶芽3D姿态估计与检测算法，该算法基于YOLOv5l网络，通过引入GhostConv和CAM坐标注意力机制及BiFPN结构进行优化，将检测模型的mAP提升至85.2%，各项指标均优于原始模型。针对名优茶机械采摘环境复杂、定位困难的问题，Zhang等[9]提出一种融合DeepLabv3+高精度分割特性与YOLOv7快速检测特性的MDY7-3PTB模型，该模型确立“先分割、后检测、再定位”的级联处理流程，实现对茶芽采摘点的准确识别。

鉴于此，本文以回归类YOLOv8n-Pose为基线，针对光电成像系统中的定位难题，提出了一种轻量化高精度模型。主要研究工作如下：(1)提出并行化补丁感知注意力模块，作为一种通用的微小目标特征增强机制；(2)设计多源双向特征融合金字塔网络，通过MFM模块与循环架构，为全尺度信息交互提供新范式；(3)引入基于IoU加权的Slide Loss，自适应提升困难样本贡献。在茶芽数据集上，模型以极小参数量实现高精度定位，并在边缘设备上完成实时验证。

1 材料与方法

1.1 数据集构建

为验证模型在复杂场景下的检测性能，实验基于复杂茶园嫩芽采摘测试场景，构建了名优茶嫩芽关键点数据集。数据采集于山东省日照市东港区某名优茶核心产区，采集周期覆盖阴、雨、晴等多种天气条件及不同光照时段；拍摄视角涵盖俯视、侧视与近景，以模拟机械臂作业时的动态观测变化；采集设备采用固定参数(分辨率1 920×1 080像素，微距对焦模式)，获取连续视频流后经10 FPS抽帧及人工筛选，获得1 210张高质量图像。数据集涵盖了修剪与未修剪茶树、不同遮挡程度以及复杂背景等多种场景，部分图像如图1所示。



图 1　茶芽数据集的部分图像

标注工作基于 Labelme 进行，同步标注目标边界框与采摘关键点（茎叶连接处）。标注数据转换为 YOLOv8 标准 txt 格式，并按 8∶2 比例随机划分为训练集（968 张）与验证集（242 张）。为缓解样本有限及类别分布不均带来的过拟合风险，训练阶段引入随机几何变换（旋转、缩放、平移）及光照扰动（亮度、对比度变化）等数据增强策略。

1.2　YOLO-AMT 网络架构

本文所提出的 YOLO-AMT 模型结构，如图 2 所示。该模型整体架构由主干、颈部和解耦检测头三部分组成。在主干网络中，采用 C2f_PPA 模块替换原 C2f 模块，强化细节特征；在颈部网络中，构建 MFM 模块与“递进式双向循环”Neck 结构，优化多尺度特征融合路径；检测头沿用解耦设计，实现边界框回归与关键点坐标预测的同步输出。

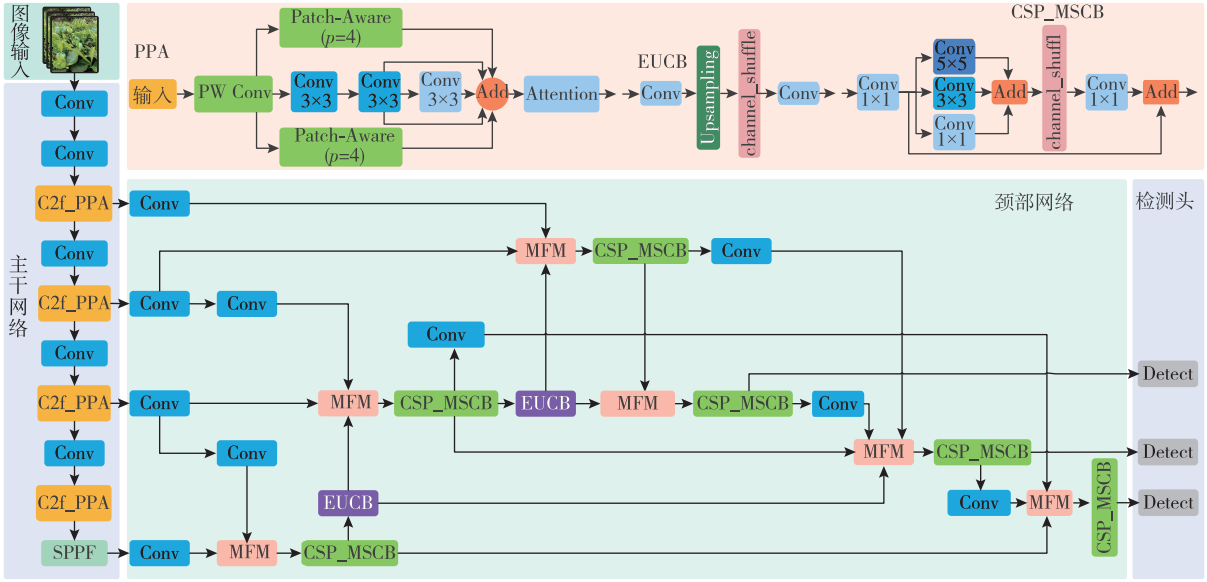
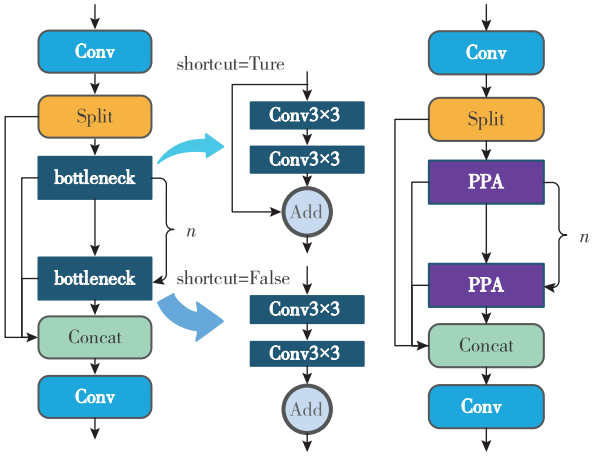


图 2　YOLO-AMT 模型结构

1.2.1　C2f_PPA 模块设计

针对常规卷积难以兼顾局部细节与全局语义的问题，本文设计的 C2f_PPA 模块重构了原始 C2f 结构，其网络架构如图 3 所示。输入特征经 1×1 卷积升维后分为两条路径：一条为跨层直连分支，用于保持浅层信息的完整性；另一条通过级联的并行补丁感知注意力单元进行深度特征挖掘，最后将两条路径的输出沿通道维度拼接并通过 1×1 降维融合。该设计思路受 Wang 等[10]在轻量化病害检测中注意力机制应用研究的启发，借鉴了特征分离-重组思想，将局部细节感知与全局语义建模进行并行整合。同时参考 Zhong 等[11]提出的双骨干网络解耦自注意力机制，在保持轻量化的前提下增强多尺度特征表达能力。



（a）C2f模块结构　　（b）C2f_PPA 模块结构

图 3　C2f_PPA 模块网络结构

如图4所示，核心计算单元PPA模块采用三分支并行架构：跳跃连接分支直接透传原始特征；多尺度局部-全局注意力（LGA）分支整合不同尺寸的补丁特征，以解析细粒度纹理与粗粒度语义；串行卷积分支逐步提取空间几何属性。三条分支的输出在末端逐元素相加融合。其中，LGA机制对特征图进行无重叠切片，经MLP映射至高维空间，并通过可学习提示向量与变换矩阵弥合补丁间空间割裂。在融合阶段，引入高效通道注意力与空间注意力，对特征进行联合重标定。最后，依次应用Dropout、批归一化及SiLU激活函数来提升泛化能力并加速收敛。

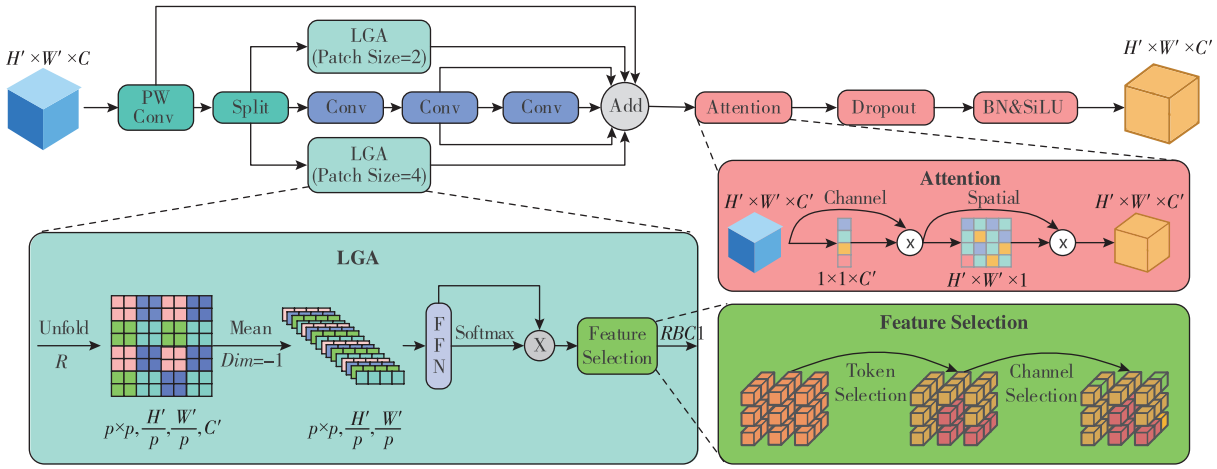


图4 PPA模块网络结构

1.2.2 多源双向对称多尺度特征融合金字塔颈部网络设计

针对传统FPN单向传递损耗与信息同质化问题，提出多源双向对称多尺度特征融合金字塔网络（MBSMFFPN）。如图5所示，该颈部网络由MFM模块与“递进式双向循环”架构构成。多尺度跨层特征融合是提升小目标检测性能的关键技术路径。Liu等[12]提出的LACF-YOLO通过轻量级注意力机制实现跨尺度特征高效融合，证明了双向信息流动在缓解小目标信息丢失中的有效性。Li等[13]在人体姿态估计任务中将混合注意力与多尺度特征融合结合，基于改进的YOLOv8-Pose显著提升了遮挡场景下的关键点定位精度。上述研究为本文MBSMFFPN的设计提供了重要参考。

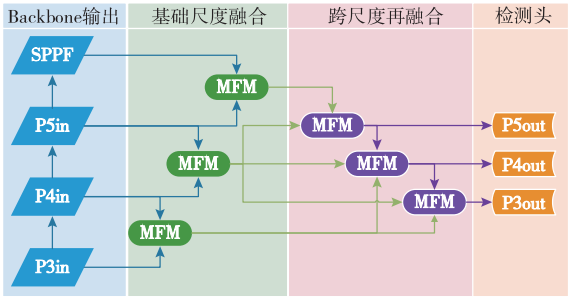


图5 “递进式双向循环”Neck架构示意图

图5所示的颈部网络整体架构遵循“高层重语义、低层重细节”原则建立双向交互通路。

（1）第一轮，基础尺度融合。以骨干网络输出的多尺度特征为输入，沿用自上而下的语义传递路径与自下而上的定位增强路径，构建初始特征金字塔。与传统PANet不同，本文在各融合节点均采用多流调制模块（MFM）替代简单的拼接或逐元素相加操作，通过全局注意力对各输入支路的贡献进行加权调制，生成具有明确功能定位的基础特征。此轮输出作为第二轮循环增强的初始状态。

（2）第二轮，跨尺度再融合（循环增强）。在第一轮基础特征之上，启动第二轮信息传递：将已生成的金字塔各层特征重新注入融合路径，并引入跨层级跳跃连接，使非相邻尺度之间能够直接交互。该设计打破了传统FPN/PAN仅允许相邻层信息交换的约束，使深层语义与浅层细节在整个金字塔内实现直接对话。第二轮传递完成后，特征再次经由MFM进行调制整合，输出的最终特征金字塔在语义鲁棒性和空间精确性上均得到递进式提升。

两轮融合首尾相连，构成从基础融合到互补增强，最后再到基础融合的闭环往复：第一轮提供稳定的初始表征，第二轮在此基础上进行跨尺度细节并注入和语义补偿，输出结果可视为对初始特征的修正。这种“递进式双向循环”设计有效缓

解了传统单向特征流中的语义鸿沟，使各个尺度的特征均能充分共享全局上下文信息。为清晰呈现本文结构与经典特征金字塔的差异，将本文递进式双向循环 Neck 与传统 PANet、BiFPN 在融合机制、连接方式与信息流特征等方面进行了系统性对比，结果如表 1 所示。

表 1　本文递进式双向循环 Neck 与传统特征金字塔的性能对比

网络	特征流次数	第一轮功能	第二轮功能	跨尺度连接方式	融合机制	信息流闭环	语义鸿沟缓解
PANet	单次 FPN＋单次 PAN	FPN 语义传递	PAN 定位传递（仅相邻层）	仅相邻层（P3→P4→P5）	逐元素相加或通道拼接	开环（单向或单次往返）	FPN 自上而下单向缓解
BiFPN	可重复多次，但每次路径相同	加权双向融合（无功能分化）	重复同一双向路径	允许跳跃连接，但权重固定	快速归一化加权融合	可重复，但无明确的“循环”反馈	双向加权缓解
本文	两轮功能分化的完整双向流	通过 MFM 生成功能定位基础特征	跨层级再融合，包括 P3↔P5 等非相邻连接	相邻层＋跨层级跳跃连接，动态注意力加权	MFM 全局注意力动态调制	两轮咬合闭环：基础融合→互补增强→再调制	全尺度信息闭环，误差逐轮衰减

从表 1 可以看出，本文的“递进式”体现在两轮融合的功能递进：首轮建立基础，次轮注入跨层互补信息；“循环”体现为特征信息在多尺度节点间的往复流动与再调制，形成持续优化的闭环机制，而非简单的重复堆叠。该设计在保持轻量化的前提下，显著提升多尺度特征的表征质量，为后续的目标检测与姿态估计任务奠定了高质量的特征基础。

如图 6 所示，在循环通路关键节点嵌入 MFM 模块，对多源输入特征做自适应加权聚合。通过对齐通道与空间尺寸，以可学习注意力权重动态重标定各特征贡献，逐元素相加融合，有效弥合语义鸿沟。

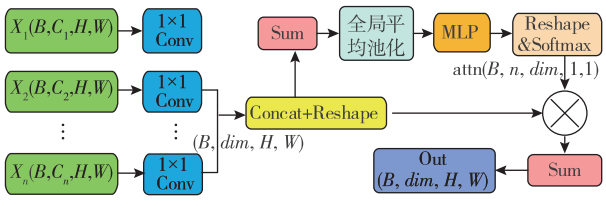


图 6　MFM 模块示意图

1.2.3　Slide Loss 损失函数

为使模型聚焦遮挡、模糊等困难样本，Zhang 等[14]提出基于 IoU 加权的 Slide Loss。该算法通过动态强调训练中代表性不足和被遮挡的困难样本，在不牺牲召回率的前提下提升检测精度。

该函数根据预测框与真实框的 IoU 动态调整权重：定位越差的样本，获得的权重越大。首先定义当前批次所有候选框 IoU 均值 μ 为自适应阈值，即：

$$\mu=\frac{1}{N}\sum_{i=1}^{N}\mathrm{IoU}_i \tag{1}$$

将样本分为“高质量简单样本”（IoU＞μ）与“低质量困难样本”（IoU＜μ）。为了加强对阈值附近样本的关注，采用高斯形式的权重函数，即：

$$\omega(\mathrm{IoU}_i)=1+\alpha\cdot\exp\left[-\frac{(\mathrm{IoU}_i-\mu)^2}{2\sigma^2}\right] \tag{2}$$

式中，α 为峰值加权系数，σ 控制影响区间宽度。结合标准二元交叉熵（BCE）损失，最终 Slide Loss 定义为：

$$L_{\mathrm{Slide}}=\frac{1}{N}\sum_{i=1}^{N}\omega(\mathrm{IoU}_i)\cdot L_{\mathrm{BCE}}(p_i,y_i) \tag{3}$$

该机制迫使模型在训练中优先优化决策边界附近的样本，解决样本不平衡问题。

2　实验结果与分析

2.1　实验设置

实验平台配置：硬件采用 Intel（R）Xeon（R）Gold 6 330 CPU@2.00 GHz，512 GB RAM，NVIDIA RTX4 090（24 GB）；软件采用 Ubuntu 20.04.6 LTS，Python 3.9，PyTorch 2.3.1＋cu 181。

训练参数：模型输入图像分辨率为 640×640，批次大小设置为 64，初始学习率为 0.01，最终学习率为 0.000 1，权重衰减系数为 0.000 5，动量为 0.937，训练轮次共设置 300 轮，设置早停机制的耐心值为 50，使用自适应动量的随机梯度下降算法（SGD）作为模型优化器，开启马赛克数据增强功能。

2.2　横向对比实验

选取 YOLO 系列中的多版本模型进行横向对比，实验结果详见表 2 与表 3。在目标与关键点检测上，将本模型与 YOLOv8/v10/v11/v13 系列 n/s 版本轻量及超轻量模型对比。总体而言，本文模型以 2.59 M 的参数量和 8.3 GFLOPs 的计算量，在各项指标上取得最优或接近最优表现。与基线 YOLOv8n-pose 相比，本文模型在零额外计算开销的前提下，在所有关键评估维度上实现一致且显著的性能跃升，尤其是关键点检测 $mAP@0.5{:}0.95$ 从 74.2% 提升至 79.1%。即便面对参数量高出一个量级的 YOLOv8s-pose，本文模型也以更少的资源消耗实现关键点精度的反超。在与超轻量的

YOLOv11n和YOLOv13n对比时，本文模型在姿态估计指标上优势明显，验证了其参数的高效性。

表2 不同模型的横纵对比实验目标检测结果

模型	参数量/M	计算量/GFLOPs	*P*/%	*R*/%	*mAP*@0.5/%	*mAP*@0.5:0.95/%
YOLOv8n-pose	3 077 771	8.3	80.0	68.0	76.6	56.8
YOLOv8s-pose	11 412 651	29.4	84.5	68.7	78.7	59.6
YOLOv10n-pose	2 766 734	8.5	79.2	67.8	76.9	57.0
YOLOv11n-pose	2 662 059	6.7	79.3	71.4	78.1	58.1
YOLOv13n-pose	2 520 018	6.4	77.6	67.4	76.3	63.6
本文	2 586 337	8.3	82.6	73.8	80.9	61.2

表3 不同模型的横纵对比实验目标关键点检测结果

模型	参数量/M	计算量/GFLOPs	*P*/%	*R*/%	*mAP*@0.5/%	*mAP*@0.5:0.95/%
YOLOv8n-pose	3 077 771	8.3	80.7	68.5	76.6	74.2
YOLOv8s-pose	11 412 651	29.4	85.2	69.4	79.3	76.5
YOLOv10n-pose	2 766 734	8.5	80.5	68.8	77.9	75.4
YOLOv11n-pose	2 662 059	6.7	80.6	72.3	79.7	76.0
YOLOv13n-pose	2 520 018	6.4	69.0	78.1	75.5	71.7
本文	2 586 337	8.3	83.5	74.8	82.1	79.1

2.3 消融实验

消融实验系统评估了C2f_PPA，MBSMFFPN和Slide Loss的贡献，实验结果如表4与表5所示，其中勾选标记（√）表示采用该项改进策略。

表4 不同模型消融实验目标检测结果

C2f_PPA	MBSMFFPN	Slide Loss	参数量/M	计算量/GFLOPs	*P*/%	*R*/%	*mAP*@0.5/%	*mAP*@0.5:0.95/%
			3 077 771	8.3	80.0	68.0	76.6	56.8
√			3 489 313	9.3	85.6	67.6	79.3	60.3
	√		2 174 795	7.3	84.0	69.5	80.5	61.8
√	√		2 586 337	8.3	84.2	73.4	81.0	61.4
√	√	√	2 586 337	8.3	82.6	73.8	80.9	61.2

表5 不同模型消融实验目标关键点检测结果

C2f_PPA	MBSMFFPN	Slide Loss	参数量/M	计算量/GFLOPs	*P*/%	*R*/%	*mAP*@0.5/%	*mAP*@0.5:0.95/%
			3 077 771	8.3	80.7	68.5	76.6	74.2
√			3 489 313	9.3	86.1	68.0	80.4	77.2
	√		2 174 795	7.3	83.3	70.5	81.3	78.7
√	√		2 586 337	8.3	84.7	73.8	81.7	78.9
√	√	√	2 586 337	8.3	83.5	74.8	82.1	79.1

由表4与表5可知，单独引入C2f_PPA模块后，因并行补丁感知注意力单元中新增的MLP映射层、可学习提示向量及通道/空间注意力等结构，参数量较基线增加约0.41 M；单独引入MBSMFFPN时，由于该颈部网络采用轻量化的多源特征融合模块，并移除了原FPN结构中部分冗余卷积与跨层连接，参数量较基线减少约0.91 M。两者分别作用于主干网络与颈部网络，结构上无重叠、参数上无共享。因此，当同时引入C2f_PPA与MBSMFFPN时，组合模型的参数量近似为二者变化的线性叠加，即基线参数量加C2f_PPA的增量、减MBSMFFPN的减量，理论值约为2.58 M，实际为2.59 M，与预期高度吻合，表明不存在参数互斥或结构冲突。类似地，计算量也遵循独立叠加规律：C2f_PPA使计算量从8.3 GFLOPs增至9.3 GFLOPs，MBSMFFPN使其降至7.3 GFLOPs，二者组合后计算量回归至8.3 GFLOPs，说明两项改进在计算开销上同样实现了有效平衡。

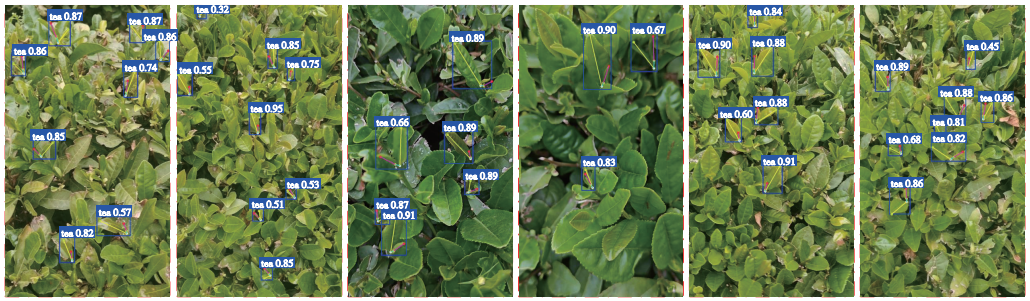
2.4 实际检测结果分析

如图7所示的实际检测可视化结果表明，本文模型检测框覆盖完整，紧密贴合目标轮廓，尤其在高密集重叠区域仍能精确区分各实例，而对比模型均有不同程度的漏检现象。在关键点定位方面，本文模型标注准确定位在茎叶连接解剖位置，对比模

型则出现位置偏移或关键点映射错误。



（a）原图



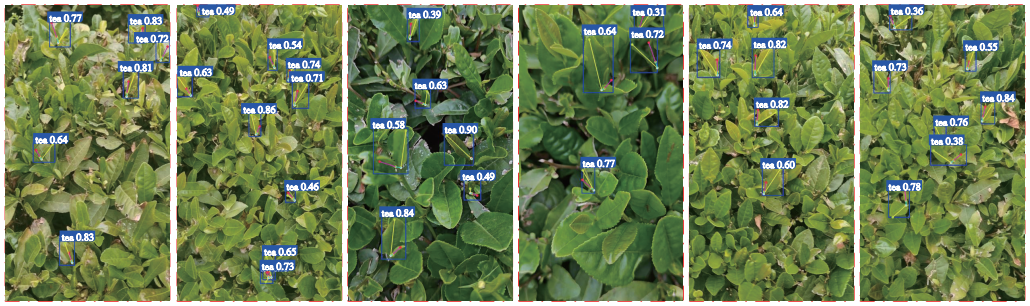
（b）YOLO-AMT



（c）YOLOv8n-pose



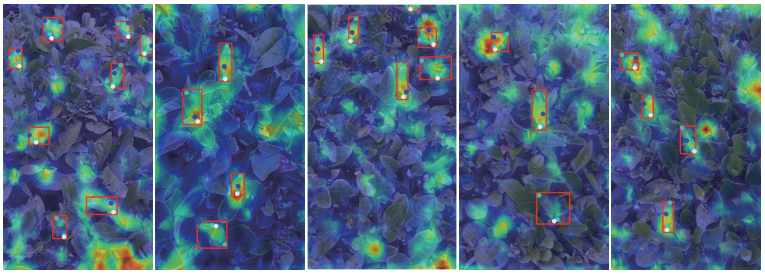
（d）YOLOv11n-pose



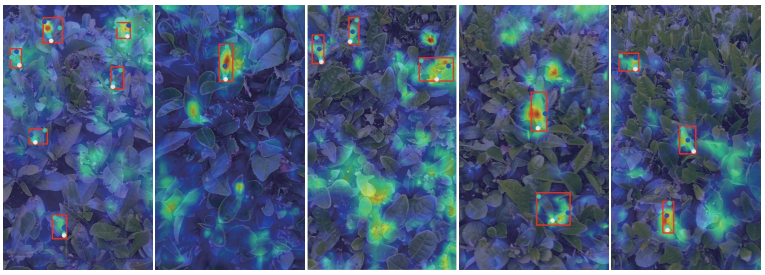
（e）YOLOv13n-pose

图 7　不同模型的实际检测结果对比图

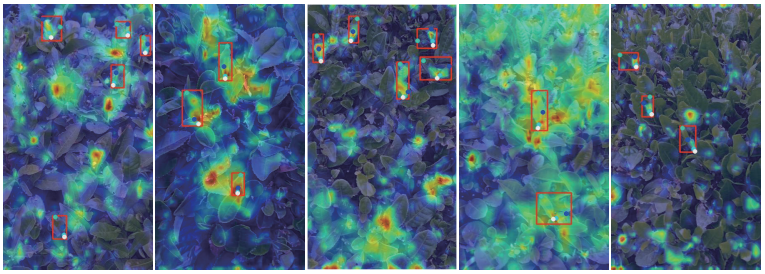
图8所示为不同模型的Grad-CAM输出结果。Grad-CAM热力图分析从特征响应层面解释了其中原因：本文模型实现了对嫩芽区域的精确聚焦，有效抑制了周围老叶及复杂背景的干扰；而对比模型的注意力则存在发散、偏移或被背景信息干扰的问题。该结果证实了本文改进模块使特征提取具有更强的抗干扰能力和感知稳定性。Selvaraju等[15]提出的Grad-CAM方法通过梯度反向传播实现深度网络决策区域的可视化解释，已成为评估模型注意力合理性的重要工具。本文热力图对比进一步证实，各项改进能有效引导网络聚焦于嫩芽茎叶连接处等任务关键区域。



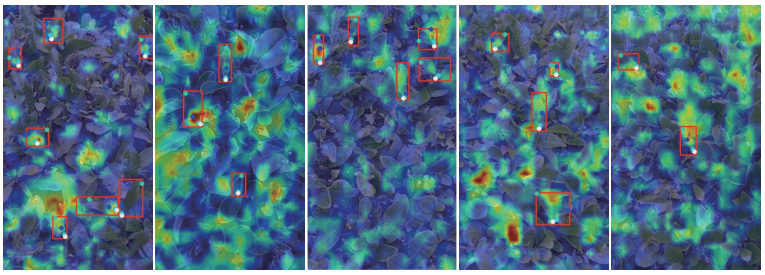
（a）YOLO-AMT



（b）YOLOv8n-pose



（c）YOLOv11n-pose



（d）YOLOv13n-pose

图8　不同模型的Grad-CAM输出结果

3　基于NVIDIA Jetson Orin Nano的部署

为验证边缘部署性能，在Jetson Orin Nano Super平台（67 TOPS算力8 GB LPDDR5内存）上完成部署测试。部署策略采用PC端构建转换、嵌入式端实测的流程：先将PyTorch模型导出为ONNX，再利用TensorRT进行FP16精度量化优化生成推理引擎，最终在Jetson平台上加载运行。整体流程如图9所示。

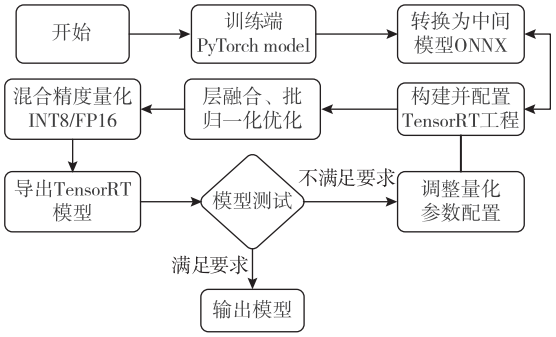


图 9　模型转换过程

针对 Jetson 平台，实验过程中通过执行 sudo nvpmodel-m0 与 sudo jetson_clocks 指令，强制将 CPU 与 GPU 锁定在最高频率运行。该操作旨在关闭处理器的动态调频机制，消除因负载波动引起的频率抖动，最大程度减少由系统策略差异引入的测试误差，确保性能评估结果的精准度与一致性。具体的软硬件环境配置参数如表 6 所示。

表 6　实验软硬件环境配置

配置项	Jetson Orin Nano Super平台
操作系统	Ubuntu 20.04(JetPack 6.0)
推理引擎	TensorRT(v8.6.1.6)
Python版本	3.9.25
电源模式	性能模式(15 W)
散热条件	主动散热风扇

3.1　部署实测结果

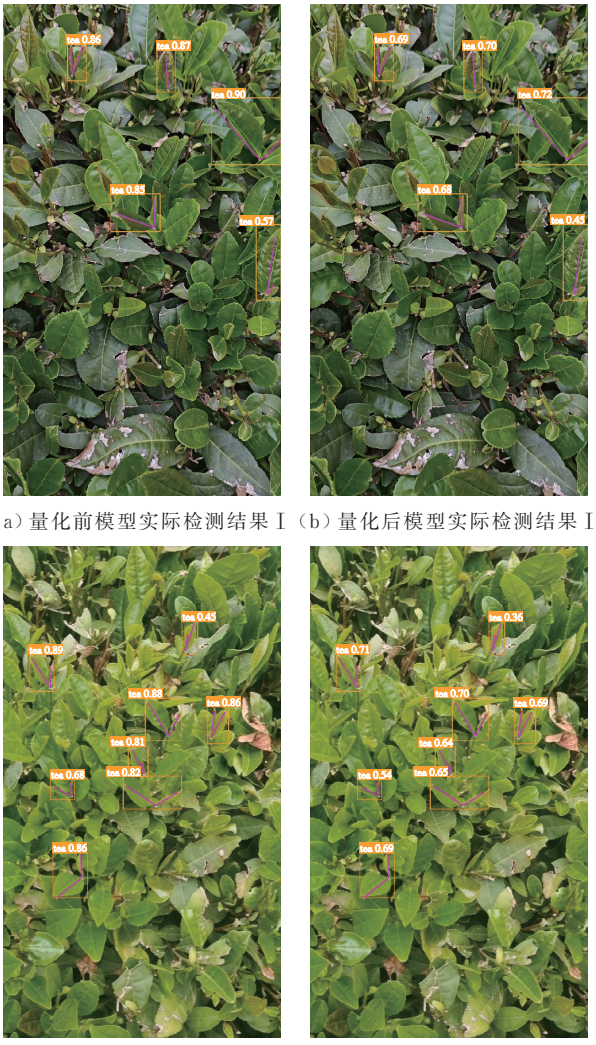
所有实验在最大性能模式下进行，并对比了模型量化前后的性能差异，如表 7 所示。

表 7　模型使用 TensorRT 量化前后的性能对比

参数	YOLOv8n-pose		YOLO-AMT	
	无量化	有量化	无量化	有量化
mAP@0.5	78.69%	69.71%	80.90%	76.70%
mAP@0.5:0.95	61.56%	53.18%	61.20%	56.73%
Pose_mAP@0.5	79.28%	72.20%	82.08%	78.00%
Pose_mAP@0.5:0.95	76.94%	67.73%	79.12%	74.82%
总耗时	12.97 ms	6.82 ms	35.97 ms	9.36 ms
推理帧率	77.13 FPS	146.59 FPS	27.80 FPS	106.84 FPS

YOLO-AMT 模型经 FP16 量化后，总推理耗时从 35.97 ms 降至 9.36 ms，推理帧率从 27.80 FPS 提升至 106.84 FPS，实现近 4 倍的加速比，满足实时作业需求。在精度方面，量化后的 Pose_mAP@0.5 为 78.00%，较量化前下降约 4%，但仍维持在较高水平。相比之下，基准模型 YOLOv8n-pose 量化后 Pose_mAP@0.5 下降达 7.08%（由 79.28% 降至 72.20%）。分析表明，尽管 YOLO-AMT 的 FLOPs 与基线持平，因 PPA 模块中包含张量切片等操作，在 Jetson Orin Nano 上未经深度优化的算子执行效率较低，导致其在未量化时推理耗时高于基线。然而，经 FP16 量化后，模型效率显著提升，最终实现 106.84 FPS 的实时推理性能，表明该轻量化架构在量化部署中具有较大应用潜力。

为验证 TensorRT 量化后的 YOLO-AMT 模型检测效果，本文在测试集中选取代表性样本进行实检。如图 10 所示，在模型进行 TensorRT 量化后，图中检测框置信度相比于量化前模型最高降低了近 20%。因此，在较高置信度阈值的情况下，量化操作后容易出现漏检情况。这表明在选择量化数据格式时，需要根据边缘设备的实际算力情况，权衡模型精确率与检测速度，以确保最佳的检测效果。



（a）量化前模型实际检测结果Ⅰ（b）量化后模型实际检测结果Ⅰ

（c）量化前模型实际检测结果Ⅱ（d）量化后模型实际检测结果Ⅱ

图 10　量化前后模型检测结果

4 结论

针对光电成像系统中运动目标实时高精度检测与轻量化部署之间的矛盾，本文以复杂茶园嫩芽采摘点定位为验证场景，提出一种基于改进YOLOv8n-Pose的轻量化姿态估计模型YOLO-AMT。该模型引入C2f_PPA模块、MBSMFFPN多尺度融合颈部网络及IoU加权Slide Loss三项核心改进，实验结果表明，所提出模型在自建数据集上实现关键点检测*mAP*@0.5∶0.95达79.1%，参数量仅为2.59 M，计算量为8.3 GFLOPs，在各评估维度上均优于同级别基线模型。在NVIDIA Jetson Orin Nano边缘平台上，经TensorRT FP16量化后推理速度达106.84 FPS，实现高精度与实时性的有效平衡。

参考文献：

[1] 陈中举，李和平，许浩然．基于LID-YOLO的小目标昆虫轻量化检测算法[J]．半导体光电，2024，45(6)：945-952.
Chen Zhongju，Li Heping，Xu Haoran. Light weight detection algorithm for small target insects based on lightweight insect detection-YOLO [J]. Semiconductor Optoelectronics，2024，45(6)：945-952.

[2] 谷学静，郭志斌．层内调节特征金字塔的密集人群姿态估计算法[J]．半导体光电，2024，45(6)：931-938.
Gu Xuejing，Guo Zhibin. Dense crowd pose estimation algorithm for in-layer adjustment feature pyramid [J]. Semiconductor Optoelectronics，2024，45(6)：931-938.

[3] Shuai L，Mu J，Jiang X，et al. An improved YOLOv5-based method for multi-species tea shoot detection and picking point location in complex backgrounds[J]. Biosystems Engineering，2023，231：117-132.

[4] Wang J，Li W，Xu J，et al. ECS-tea：a bio-inspired high-precision detection and localization algorithm for young shoots of Pu-erh tea [J]. Frontiers in Plant Science，2025，16：1697209.

[5] Wang Z，Niu H，Zhang J，et al. Automated tea shoot picking using the YOLO network and Mamba images segmentation for top-view detection with a monocular camera [J]. Journal of Agricultural Engineering，2026，57(1)：1637.

[6] Chen Y T，Chen S F. Localizing plucking points of tea leaves using deep convolutional neural networks [J]. Computers and Electronics in Agriculture，2020，171：105298.

[7] Chen C，Lu J，Zhou M，et al. A YOLOv3-based computer vision system for identification of tea buds and the picking point [J]. Computers and Electronics in Agriculture，2022，198：107116.

[8] Chen Z，Chen J，Li Y，et al. Tea bud detection and 3D pose estimation in the field with a depth camera based on improved YOLOv5 and the optimal pose-vertices search method [J]. Agriculture，2023，13(7)：1405.

[9] Zhang F，Sun H，Xie S，et al. A tea bud segmentation，detection and picking point localization based on the MDY7-3PTB model [J]. Frontiers in Plant Science，2023，14：1199473.

[10] Wang J，Ma S，Wang Z，et al. Improved lightweight YOLOv8 model for rice disease detection in multi-scale scenarios[J]. Agronomy，2025，15(2)：445.

[11] Zhong M，Wei L，Mo H. YOLOv8n-DSDL：a lightweight dual-backbone network with decoupled self-attention for cotton maturity detection[J]. Journal of King Saud University Computer and Information Sciences，2025，37(5)：88.

[12] Liu S，Shao F，Chu W，et al. An improved YOLOv8-based lightweight attention mechanism for cross-scale feature fusion[J]. Remote Sensing，2025，17(6)：1044.

[13] 谷学静，栗燕茹，杨蓝潇．结合混合注意力与多尺度特征的人体姿态估计[J]．计算机工程与科学，2026，48(3)：531-539.
Gu Xuejing，Li Yanru，Yang Lanxiao. Human pose estimation combining mixed attention and multi-scale feature [J]. Computer Engineering & Science，2026，48(3)：531-539.

[14] Zhang J，White T S，Zhang H，et al. MS-YOLO：infrared object detection for edge deployment *via* MobileNetV4 and SlideLoss[C]//2025 International Joint Conference on Neural Networks，2025：1-8.

[15] 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(2)：336-359.

作者简介：

张澎涛(1980—)，副教授，博士研究生，硕士生导师，主要从事计算机视觉、智能检测与故障诊断领域的研究；

张佳薇(1975—)，教授，博士研究生，博士生导师，主要从事多传感器数据融合、林业信息智能检测、林业物联网等领域的研究。