DOI: 10.12398/j.issn.2096-7217.2026.02.017

基于YOLO模型的动物体尺关键点检测研究

刘地华[1], 司秀丽[1,2], 张树敏[3], 马丽[4*]

(1. 吉林农业大学信息技术学院,吉林长春,130118; 2. 吉林省农业人工智能研究基地,吉林长春,130118; 3. 吉林农业大学动物科学技术学院,吉林长春,130118; 4. 无锡学院,江苏无锡,214105)

摘要:对动物体尺测量关键点进行精确定位,实现智能畜牧健康监控和自动化测量具有重要意义。现有关键点检测模型往往存在:参数量较大,不利于在边缘设备上的部署;部分轻量化模型的检测精度难以满足定位精度的要求。为此,研究设计一种基于改进YOLOv8-Pose架构的轻量化模型,在保证实时性的同时有效提升了关键点检测精度。该模型在主干网络的浅层以及3个检测头之前集成坐标注意力机制,将空间坐标信息有效编码到通道注意力中,增强模型对关键点空间分布感知能力,提升关键点定位的精确度和稳定性。浅层主干网络引入强制锐化模块,强化边缘和轮廓等细节信息,平衡全局语义信息与局部细节信息的表达能力。此外,模型在浅层主干网络引入强制锐化模块,强化边缘和轮廓等细节信息,平衡全局语义信息与局部细节信息的表达能力。另外,模型通过分支方式引入更高分辨率的特征图,向3个检测头提供细粒度的空间信息,弥补深层特征图在空间分辨率上的不足。试验结果表明,所提出的改进模型在复杂马匹关键点检测任务中精度达到82.6%,召回率为81.3%,*mAP*50-95提升至58.1%。相较于YOLOv8-Pose,模型精度、召回率和*mAP*50-95分别提高了21.0%、39.2%和45.3%,在复杂姿态和多变场景下表现出更强的关键点定位能力。模型参数量由3.361 M增加至4.611 M,增幅约为37.2%;浮点计算量由9.6×10^9增加至10.4×10^9,增幅仅为8.3%。在保持推理速度基本不受影响的前提下,模型性能实现了显著提升,表明该方法以较小的模型代价换取较大的性能收益,在精度与效率之间取得良好的平衡,对智能畜牧养殖和自动测量等应用具有重要推动作用。

关键词:YOLOv8-Pose;动物关键点检测;局部特征增强;坐标注意力;智能畜牧

中图分类号:S24 **文献标识码:**A **文章编号:**2096-7217(2026)02-0182-11

刘地华, 司秀丽, 张树敏, 马丽. 基于YOLO模型的动物体尺关键点检测研究[J]. 智能化农业装备学报(中英文), 2026, 7(2): 182－192

Liu Dihua, Si Xiuli, Zhang shumin, Ma Li. Animal body measurement keypoint detection based on YOLOv8-Pose [J]. Journal of Intelligent Agricultural Mechanization, 2026, 7(2): 182－192

0 引言

近年来,关键点检测作为一种介于目标检测与精细结构建模之间的中间表示方法,已被广泛应用于农业与畜牧视觉感知任务中[1-3]。现有研究表明,其应用大体可分为以粗略定位为目标的姿态与行为分析[4-6],以及以高精度定位为基础的几何测量与三维重建两类。前者主要用于目标结构建模与状态识别,对像素级定位精度要求相对有限。相比之下,在农业表型分析[7-8]与动物体尺测量等任务中,关键点不再仅用于结构描述,而是直接作为几何测量与空间计算的核心基准点。例如,在玉米植株参数提取及牲畜体长、体高等体尺测量任务中,无论采用单目、双目还是多深度相机系统[9-11],最终的距离与尺度计算均高度依赖关键测量点的精确定位。已有研究指出,关键点二维定位误差会在视差计算与深度恢复过程中被显著放大,从而成为制约测量精度与系统稳定性的主要误差来源[1,12-14]。因此,在此类应用场景下,关键点检测实质上承担了跨视角匹配与三维计算锚点的功能。

然而,对于复杂姿态与多关键点分布等任务场景,高精度模型通常依赖复杂网络结构和较高计算开销,难以满足农业与畜牧系统对实时性和稳定运行的要求;而轻量化模型虽具备较好的推理效率,却往往难以提供稳定可靠的关键点定位精度。随着农业生产和畜牧养殖向智能化、自动化方向发展,关键点检测系统逐步从实验室研究走向实际场景部署,这使得模型在保

收稿日期:2025-12-25 修回日期:2026-02-10

基金项目:国家重点研发计划子课题(20200136);吉林省科技发展计划项目(20240303066)

第一作者:刘地华,男,1998年生,江西吉安人,硕士研究生;研究方向为智慧农业。E-mail: liudihuaemail@163.com

*通信作者:马丽,女,1980年生,吉林长春人,博士,教授;研究方向为智慧农业。E-mail: mali@cwxu.edu.cn

证一定定位精度的同时，还必须具备推理效率高、结构简洁、易于部署等特性。已有大量研究表明，基于YOLO[15]的单阶段目标检测模型已在农业与畜牧视觉任务中实现了稳定的实时部署和良好的工程适配性[16-18]。在现有关键点检测框架中，以YOLO-Pose为代表的单阶段方法在推理速度、工程适配性和部署成熟度方面具有明显优势，因此成为当前实现关键点检测工程化落地的合理选择。

综上所述，在智能化农业测量与动物体尺估计等高精度应用场景中，高效且稳定的关键点精确定位能力已成为制约系统性能的核心瓶颈。现有YOLO-Pose方法虽然在工程部署和实时性方面具有优势，但其在动物精确定位方面仍有提升空间。一方面，YOLO-Pose对关键点没有显示空间建模。另外，因为共享主干，也容易发生多任务的特征优化冲突。基于此，本研究以YOLOv8-Pose为基础模型，对其进行针对性改进，以提升复杂场景和姿态下关键点检测的精度和稳定性，为视觉测量应用提供更加可靠的技术支撑。

1　材料与方法

1.1　试验材料

本研究采用的试验数据集为马匹关键点数据集，来源于公开数据集AnimalPose[19]。为保证数据质量，剔除了人眼难以辨识关键点以及马匹在图像中占比过小、无法进行有效关键点标注与测量的223个样本，最终获得了一个包含500张图像的马匹关键点数据集合。图1展示了不同姿态、不同光线和不同环境下马匹图像标注样本。



图1　不同姿态、不同光线和不同环境下的马匹关键点

Figure 1　Keypoints of horses under different poses, lighting conditions, and environmental settings

该数据集中包含不同场景的马匹图像，图像之间不存在时序关联，具有较高的场景多样性。马匹在图像中呈现出丰富的姿态变化，涵盖站立、行走、奔跑等多种动作形式，同时伴随复杂多变的光照条件。此外，不同马匹在体型和品种方面存在明显差异，使得整体数据分布更加复杂，对关键点检测模型的鲁棒性提出了较高要求。在关键点标注方面，数据集共标注了包括眼睛、耳朵及四肢关节在内的20个关键点。表1描述了数据集中关键点序号代表的部位。

表1　数据集20个点关键点对应关系

Table 1　Correspondence of keypoints for 20 points in the dataset

关键点序号	代表的部位
0	左眼
1	右眼
2	鼻子
3	左耳
4	右耳
5	左前肘关节
6	右前肘关节
7	左后肘关节
8	右后肘关节
9	左前膝关节
10	右前膝关节
11	左后膝关节
12	右后膝关节
13	左前蹄
14	右前蹄
15	左后蹄
16	右后蹄
17	喉咙部位
18	鬐甲
19	尾根

1.2　数据集预处理

作为一种典型的监督学习关键点检测模型，YOLOv8-Pose对训练数据标注质量具有较高敏感性。标注误差所引入的标签噪声可能影响模型的特征学习过程并降低预测精度。为减小标签噪声对模型训练的干扰，本文借助可视化检查脚本对500张图像的关键点标注进行了逐一核查与修正，从而提高训练数据的准确性与一致性。由于数据量相对较少，难以涵盖所有可能的姿态和环境变化，模型容易发生过拟合，泛化能力不足。

为此，本研究对数据集进行了离线数据增强，主要采用左右和上下翻转以扩大关键点分布范围，利用自适应直方图均衡化增强图像局部对比度，突出边缘细

节,以及使用上下对称裁剪、左右对称裁剪、中心裁剪增强以强化局部信息和局部特征学习。离线数据增强之后,再次去除极端结果,原有500张标注图像扩充至4 257张,数据集随机划分为训练集(3 326张)、验证集(698张)和测试集(233张)。同时,本研究不使用模型在线数据翻转增强,开启了在线数据增强中的马赛克策略,以进一步提升数据多样性和模型鲁棒性。图2展示了增强后的图像样本。

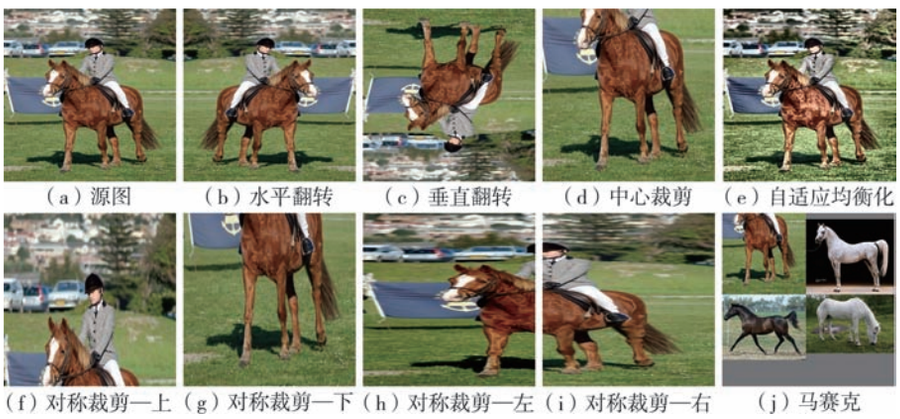


图2 所有用到的增强方式

Figure 2 All applied data augmentation methods

2 改进YOLOv8n-Pose检测模型

2.1 YOLO-Pose检测模型

YOLO-Pose是在YOLO目标检测框架基础上发展而来的关键点检测模型,最早由Maji等[20]基于YOLOv5提出,通过将目标检测与关键点回归统一到单阶段网络中,实现了无热图的端到端多人姿态估计。该方法以直接回归关键点坐标的方式替代传统两阶段或基于热图的检测范式,在模型结构简洁性和推理效率方面具有明显优势。

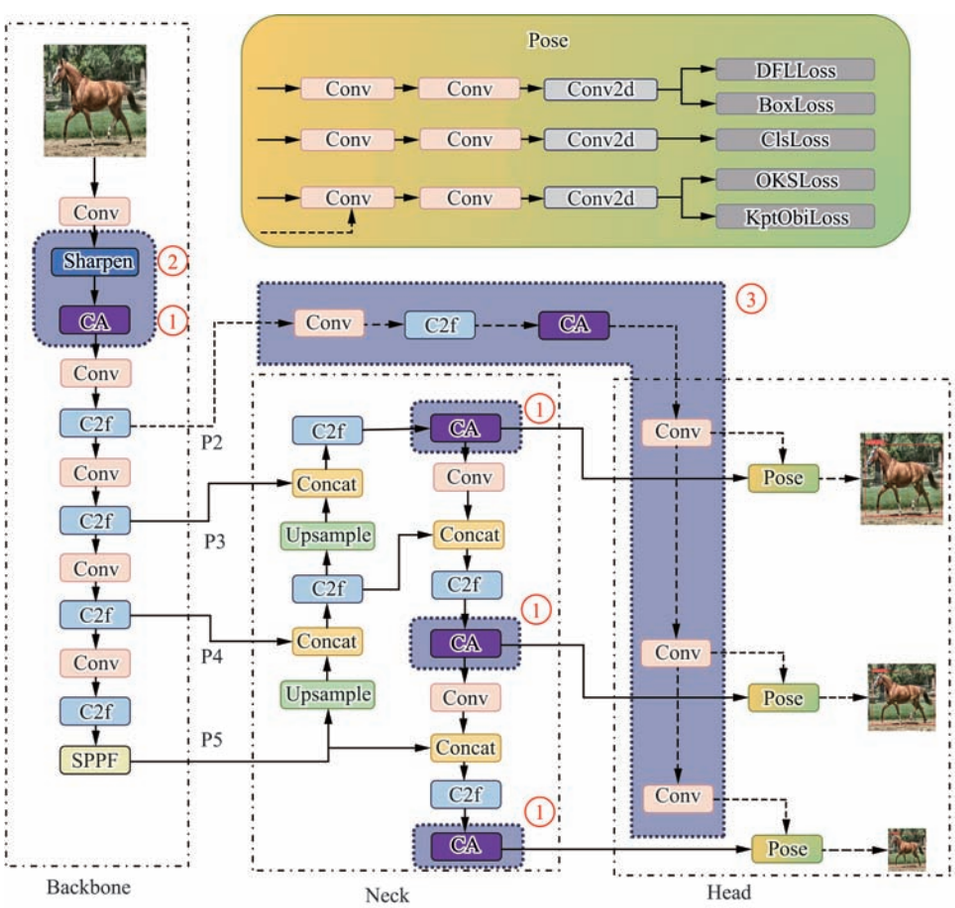


图3 改进的YOLOv8-Pose结构图

Figure 3 Architecture diagram of the improved YOLOv8-Pose model

基于上述分析，本文围绕YOLOv8-Pose结构展开改进，针对关键点检测任务的特点设计了一系列改进，旨在提升模型在复杂环境中的关键点检测性能和稳定性，为智能畜牧领域的视觉测量任务提供更加可靠的技术支撑。图3为改进后的结构示意图，虚线框内为本文所作修改。

2.2　添加坐标注意力机制

在实际畜牧体尺测量场景中，关键点往往沿目标的长轴方向呈现非均匀分布，例如体长相关关键点主要沿水平方向排列，而体高相关关键点则沿垂直方向分布。此外，目标在不同拍摄距离与姿态下，其尺度变化显著，导致关键点在特征图中的相对位置差异较大。现有YOLO-Pose中的特征主干主要关注全局语义响应，缺乏对关键点空间方向性的显式建模能力，在目标尺度变化或姿态倾斜时，容易出现关键点响应分布模糊、定位偏移等问题。因此，在本研究中，关键点检测不仅需要语义区分能力，还需要显式建模关键点在水平与垂直方向上的空间分布特征，以增强模型对结构性几何信息的感知能力。基于上述考虑，本研究引入了坐标注意力机制[21](coordinate attention, CA)。坐标注意力机制通过将二维空间信息沿水平方向和垂直方向进行解耦建模，分别对特征图在高度和宽度方向进行全局聚合，生成具有方向感知能力的注意力权重，并与原始特征进行逐元素融合。该过程在建模通道间依赖关系的同时，保留了关键的空间位置信息，从而实现对不同空间位置特征响应的自适应增强。具体的实现过程如图4所示。

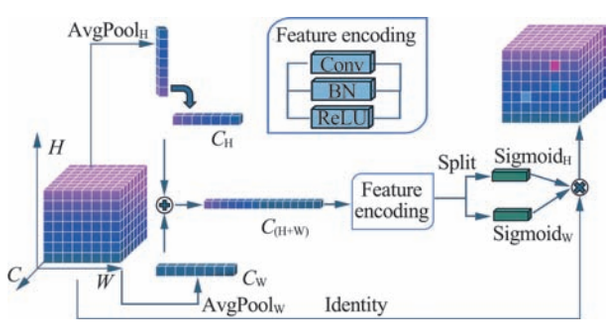


图4　坐标注意力机制

Figure 4　Coordinate attention

本文在YOLOv8-Pose主干网络的第二层以及3个检测头之前引入坐标注意力机制，以增强模型对关键点空间结构的建模能力(见图3中①)。该机制在特征提取阶段显式引入空间位置信息，引导网络重点关注与关键点定位密切相关的结构区域。

2.3　引入锐化模块

在动物体尺估计等应用场景中，关键点通常分布于目标结构的边界或局部几何突变位置，例如关节拐点、躯干端点及轮廓极值区域。这类关键点的准确定位高度依赖于局部边缘与轮廓等高频几何信息[22-24]。然而，在YOLO-Pose框架中，边框检测与关键点回归任务共享同一特征主干，特征提取过程更侧重目标区域的语义一致性。在连续卷积、下采样及多尺度特征融合过程中，与关键点定位密切相关的浅层边缘与局部几何信息容易被平滑化，从而削弱模型对精细结构的表达能力，导致关键点在边界区域出现定位抖动或系统性偏移。针对上述问题，本文在特征提取阶段引入了一种针对关键点任务的局部几何增强锐化模块(图3中②)。该模块旨在补偿多任务共享特征主干中高频几何信息的损失，通过显式增强特征图中与边界和局部结构变化相关的响应，使关键点预测更加集中于真实几何位置，从而提升后续测量任务中的定位精度与稳定性。

具体而言，锐化模块采用基于拉普拉斯算子[25]的固定权重卷积，对特征图各通道进行高频强化处理，在不引入额外下采样或空间压缩的前提下，突出边缘、轮廓及局部强度变化显著的区域，抑制过度平滑的冗余特征。与依赖学习权重的常规卷积不同，该模块通过显式建模局部几何变化，为关键点分支提供更加清晰且判别性更强的结构特征支持。其具体实现流程如图5所示。

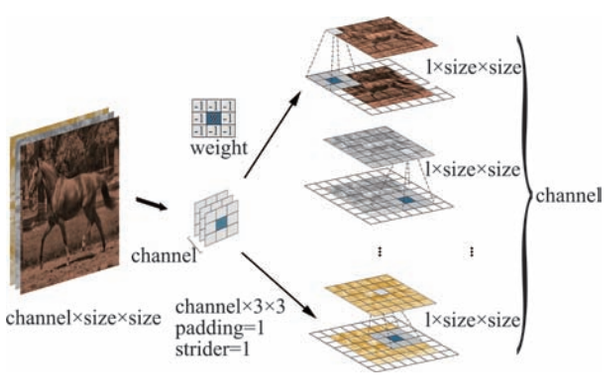


图5　锐化模块

Figure 5　Sharpen module

2.4　Pose分支单独引入高分辨率特征

在实际采集条件下，为覆盖完整目标并保证多目标并行处理，农业与畜牧图像通常具有较大的视场范围，单个目标在图像中所占比例有限，导致部分关键点在高层特征图中仅对应极少的像素响应。现有YOLO-Pose关键点预测主要依赖P3及以上特征层，其空间分辨率较低，对于体型较小或远距离目标，关键点在特征空间中容易发生量化误差，限制了像素级定位精度。然而，在体尺测量任务中，即使是亚像素级的二维误差，也可能在深度恢复过程中被显著放大。因此，有必要在模型中引入更高分辨率的特征表示，以保

留细粒度空间信息。

基于上述需求,本文在特征金字塔结构中引入高分辨率P2分支,用于增强关键点分支对细粒度空间结构的感知能力(见图3中③)。相较于仅依赖P3及以上特征层的关键点预测方式,P2分支能够在较少语义抽象的前提下保留更加精细的位置信息,从而有效提升小尺度目标及边界关键点的定位精度。此外,作为面向关键点检测的独立特征分支,P2分支为关键点任务提供了一条更偏向局部几何信息的特征路径,在一定程度上减少了其对高层语义特征的依赖,从而缓解了边框检测与关键点检测在共享主干上的特征竞争问题。

2.5 模型训练与算法性能指标

2.5.1 模型训练

本研究试验均在统一的软硬件环境下完成。试验平台基于Ubuntu 24.04 LTS操作系统,硬件配置包括14核vCPU、120 GB系统内存,以及一块24 GB显存的NVIDIA GPU。深度学习加速环境采用CUDA 11.8,模型训练与推理基于PyTorch 2.1.0框架,使用Python 3.10作为开发语言。在模型设置方面,本文以Ultralytics 8.0.183作为改进版本,并在其基础上进行结构改进与对比试验。所有输入图像均统一调整为640×640的分辨率,以保证训练与测试阶段的一致性。在训练策略上,模型优化采用随机梯度下降优化器,初始学习率设置为0.01,动量参数设为0.937。训练过程中批大小设为16,每个模型均训练200轮次,以确保模型充分收敛并获得稳定的性能表现。

2.5.2 评价指标

在模型的训练损失函数曲线可用于反映训练过程的收敛特性与优化稳定性,其下降趋势及最终收敛水平在一定程度上决定了模型的性能上限。在性能评估阶段,本文主要采用精确率(*Precision*)、召回率(*Recall*)、平均精确率(*average precision*,*AP*)和均值平均精确率(*mean average precision*,*mAP*)等指标,对模型在关键点检测任务中的整体性能进行定量评估,见式(1)—式(4)。当统计所有任务指标时,边框检测任务使用Box表示,关键点检测关键点任务使用Point表示。在关键点检测任务中,模型需要首先正确定位目标实例,并在此基础上预测其对应的关键点位置。对于第i类目标,其平均精度$AP(i)$定义为在指定IoU阈值范围内精确率与召回率(*Precision-Recall*)曲线下的面积。均值平均精确率mAP则通过对所有类别AP值取平均值获得,用于综合评估模型在多类别关键点检测任务中的整体性能。本文采用mAP50-95作为主要评价指标,该指标在10个不同IoU阈值(0.50, 0.55, 0.60, …, 0.95)下分别计算AP并取平均值,能够更全面地反映模型在不同定位精度要求下的关键点检测能力。此外,为评估模型在实际部署场景中的运行效率,本文引入模型参数量(*Params*)、计算量以及推理速度作为辅助评价指标。通过综合精度指标与效率指标,可对模型在关键点检测任务中的性能进行全面评估。

$$Precision=\frac{TP}{TP+FP} \tag{1}$$

$$Recall=\frac{TP}{TP+FN} \tag{2}$$

$$AP=\int_0^1 Precision(Recall)\,\mathrm{d}Recall \tag{3}$$

$$mAP=\frac{1}{N}\sum_{i=1}^{N}AP(i) \tag{4}$$

式中:TP——在给定边框或者关键点的IoU阈值下,正确预测的边框或者关键点的数量;

FP——在给定边框或者关键点的IoU阈值下,错误预测的边框或者关键点的数量;

FN——在给定边框或者关键点的IoU阈值下,未能预测的边框或者关键点的数量。

3 试验结果与分析

3.1 消融试验

首先,在基线模型上引入了坐标注意力机制(A)后,模型在目标检测与关键点检测任务上均取得明显提升。其中,目标检测精度与召回率均有显著提高,关键点检测的整体精度提升幅度为7%,同时模型参数量和计算量基本保持不变。随后,引入锐化模块(B)后,关键点检测性能获得一定提升,关键点召回率提高16.5%,但对目标检测性能的提升有限,整体增益相对较小。在模型结构中加入P2高分辨率特征图分支(C)后,关键点检测精度得到进一步提升,关键点mAP50-95提升30%,但同时带来了模型参数量和计算量的增加。在组合试验中,BC组合相较于基线模型,使关键点检测精度提升11%,关键点召回率提升29.7%,验证了高分辨率特征与局部细节增强模块之间的互补性。AB组合在保持模型规模不变的前提下,使目标检测精度提升7%,同时关键点检测精度提升约13.7%。进一步地,AC组合在关键点检测任务上取得更大提升,关键点精度与召回率分别提升13.5%和35.9%,表明坐标注意力机制与高分辨率特征分支在多关键点建模方面具有显著协同作用。

综合3种改进模块后,ABC模型在整体性能上达到最优。与基线模型相比,目标检测精度提升7.1%,关键点检测精度提升21%,关键点检测召回率提升

39.2%，关键点 *mAP*50-95 提升 45.3%。在可接受的计算开销增加下，模型在多关键点复杂姿态场景中的检测性能得到显著增强。消融试验结果如表 2 所示，部分关键点检测结果如图 6 所示。

表 2　消融试验结果

Table 2　Results of ablation experiments

Model	Box			Point			参数量/M	浮点计算量/($\times10^9$)	推理速度/(帧·s^{-1})
	Precision/%	*Recall*/%	*mAP*50-95/%	*Precision*/%	*Recall*/%	*mAP* 50-95/%			
YOLOv8-Pose	91.6	0.85	71.6	61.6	42.1	12.8	3.361	9.6	120
A	96.3	90.4	81.2	79.6	69.6	35.2	3.374	9.6	118
B	90.5	84.3	70.6	68.5	58.6	40.6	3.361	9.6	119
C	91	82	71.2	69.6	57.8	42.8	4.599	10.3	110
BC	90.8	83	70.7	72.6	71.8	50.5	4.599	10.3	108
AB	98.6	92.8	80.6	73.5	76.9	52.6	3.374	9.6	117
AC	97.4	95	81	75.1	78	56.4	4.611	10.4	108
ABC	98.7	95.6	84.3	82.6	81.3	58.1	4.611	10.4	105



图 6　消融试验可视化

Figure 6　Visualization of ablation study results

图 7 为消融试验中不同模型在关键点任务上的训练集与验证集损失曲线。引入改进模块后的模型在训练和验证阶段均表现出更快的收敛速度，并在训练后期达到更低的损失水平，表明所提出的改进有助于优化关键点回归的学习过程并提升模型的收敛下限。

3.2　与其他算法对比分析试验

为了验证所改进方法的有效性，本研究选取了 YOLO 系列中具有代表性的多个版本（YOLOv3、YOLOv5、YOLOv7[26]、YOLOv8、YOLOv11、YOLOv12）进行对比。YOLO 系列以其轻量化设计和快速的推理速度广泛应用于目标检测任务，且各版本在结构和性能上存在差异，能够反映轻量目标检测模型的性能水平。与此同时，本研究还对比了当前主流的关键点检测和姿态估计模型，包括 Keypoint R-CNN[27]、CenterNet[28]、HRNet[29]和 DEKR[30]，这些模型在关键点定位精度和检测准确率方面均处于领先水平。

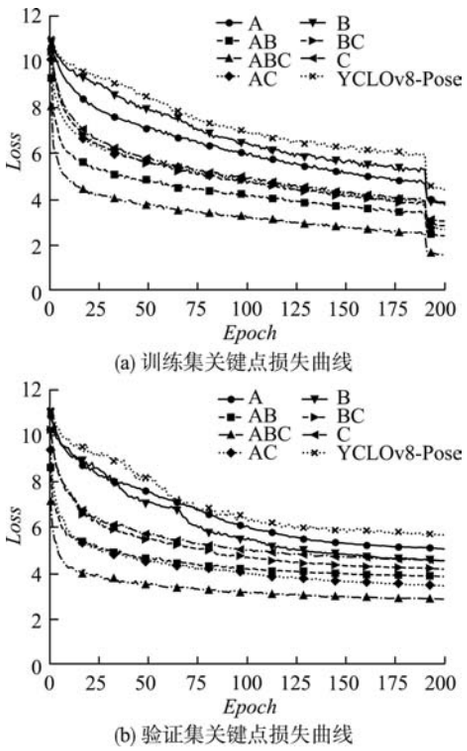


图 7 消融试验中在验证集和训练集的关键点损失曲线

Figure 7 Keypoint loss curves on the training and validation sets in the ablation experiments

表 3 对比试验结果表

Table 3 Comparison experiment results

Model	精确率/%	召回率/%	*mAP*50-95/%	参数量/M	浮点计算量/($\times10^9$)	推理速度/(帧·s^{-1})
YOLOv3	58	49.2	13.5	4.413	14.8	95
YOLOv5	56	50.3	12.1	2.693	7.9	136
YOLOv7	62	57	26	36.6	104	33
YOLOv8	61.6	42.1	12.8	3.361	9.6	120
YOLOv11	53	48.3	11.8	2.94	7.8	140
YOLOv12	51.4	47.6	11.1	2.919	7.9	133
Keypoint R-CNN	68	57.6	46	34.5	45.4	50
CenterNet	80	79.9	48.7	151.5	131	23
HRNet	89	87.8	85	63.5	131	21
DEKR	81.3	80.2	82	60	32	70
Ours	82.6	81.3	58.1	4.611	10.4	108

对比结果如表3所示,YOLO同系列模型在参数量和计算复杂度方面较低,适合实时性要求较高的场景,但关键点检测性能有限,*mAP*50-95指标普遍较低,最高仅为13.5%;主流关键点检测模型则表现出明显优势,*Precision*和*Recall*均高于80%,其中HRNet以85%的*mAP*领先;而本文所改进的模型在参数量、计算量上和计算速度与大部分YOLO系列相当,显著低于大部分模型,但关键点检测性能大幅提升,达到58.1%的*mAP*50-95,精确率和召回率均超过80%,实现了准确度与效率的良好平衡,体现了所提方法在保证轻量化的同时显著提升检测性能的能力,具备较强的应用价值。

4 讨论

为进一步分析各改进模块在关键点检测任务中的作用差异,本文统计了不同消融配置下模型在测试集上的各关键点平均像素误差,相关结果如表4所示。

表 4 不同消融配置下测试集各关键点的平均像素误差

Table 4 Mean pixel error of each keypoint on the test set under different ablation settings

序号	YOLOv8-pose	A	B	C	BC	AB	AC	ABC
0	90.21	45.1	54.12	40.7	28.49	27.55	20.35	19.31
1	90.27	45.13	54.16	42.36	29.62	26.56	21.18	20.06
2	88.49	44.24	44.25	40.49	28.34	24.12	24.24	21.86
3	87.79	43.89	52.74	35.79	25.05	21.94	17.89	16.1
4	88.84	44.42	53.04	36.83	25.78	26.21	18.41	16.57
5	67.99	47.59	40.74	30.5	21.35	23.79	18.3	16.47
6	63.12	44.18	37.87	31.12	21.78	22.09	18.67	16.8
7	76.77	53.73	46.02	34.35	24.04	26.86	20.61	18.5
8	72.09	50.46	43.54	30.19	21.13	25.23	18.11	17.3
9	80.86	56.6	40.43	42.86	25.71	24.16	25.71	24.14
10	78.52	54.9	39.26	42.5	25.5	25.35	25.5	22.95
11	90.63	63.4	45.15	37.65	22.59	26.43	22.59	20.33
12	89.79	62.85	44.85	38.79	23.24	29.18	23.27	20.94
13	99.49	49.74	49.75	42.9	25.74	23.82	21.45	19.3
14	99.75	49.87	49.85	45.7	27.42	25.96	22.85	20.56
15	98.86	49.43	49.43	48.8	29.28	27.72	24.4	21.96
16	99.39	49.69	49.95	48.9	29.34	25.85	24.45	22
17	97.99	63.69	48.95	43.1	25.86	28.57	25.86	23.27
18	89.41	58.11	44.05	47.21	28.32	25.88	22.32	20.04
19	99.54	49.77	49.77	47.3	33.11	24.85	23.65	21.28

需要说明的是,本文所研究的关键点定位任务与人体姿态估计等通用姿态任务在问题设定上存在一定差异。在实际测量场景中,关键点检测仅针对可见关键点展开,不可见关键点(包括因遮挡导致的不可见点)不参与位置回归与像素误差计算,模型训练过程中也仅对关键点可见性进行约束。因此,本文中关于性

能提升的分析主要集中在可见关键点的定位精度上。基于该设定，本文进一步从关键点空间分布位置及其在不同姿态下的几何变化特性出发，对各改进模块的作用进行了讨论。

4.1　坐标注意力机制的影响

在YOLO-Pose框架中，关键点坐标回归并非基于所有候选锚点，而是基于正样本分配阶段筛选得到的有效锚点。具体而言，模型首先在目标边框内部对锚点进行筛选，并依据*IoU*或匹配策略选取与真实目标重叠度较高的锚点作为正样本，用于后续的边框与关键点回归。由于该筛选过程以目标边框重叠关系为主要依据，最终保留下来的正锚点在空间分布上通常集中于目标主体区域，整体上接近于物体的几何中心位置。基于这一特性，本文将关键点到目标几何中心（亦可视为正锚点聚集区域）的空间距离，作为衡量关键点“远离目标中心”程度的参考指标。对于马匹关键点而言，耳朵、眼睛、马蹄及尾根等部位在静止状态下本身即位于目标轮廓或肢体末端区域，相较于躯干中心具有更大的空间偏移；同时，在实际运动或姿态变化过程中，马匹四肢摆动幅度较大，头部在低头或抬头等动作下亦会产生明显位移，使得上述关键点在图像中的空间位置变化更为剧烈。因此，这类关键点可视为相对于正锚点分布区域距离更远且对姿态变化更为敏感的一类关键点。由表4可见，在仅引入CA注意力机制的情况下，模型在马蹄、耳朵等关键点上的误差下降幅度，相较于肘关节、膝关节等更接近目标中心的关键点更为显著。由于P2分支中同样引入了CA模块，其整体提升趋势与单独引入CA的结果保持一致，进一步表明CA注意力机制对远离目标中心且空间变化较大的关键点具有更显著的增强作用。

4.2　锐化模块的影响

锐化模块的核心作用在于强化特征中的边界轮廓与高频信息，从而提升模型对局部几何结构的辨识能力。然而，该过程在增强目标轮廓的同时，也不可避免地存在对噪声或背景边缘产生误增强的潜在风险，尤其是在纹理变化较为剧烈的区域。因而，锐化模块的输入特征质量对其实际效果具有重要影响，其前序特征提取与筛选过程在一定程度上起到了抑制无关高频噪声的过滤作用。

从表4的统计结果可以观察到，单独引入锐化模块时，模型在马匹四肢、鬐甲及鼻部等轮廓特征较为明显的关键点上获得了相对更显著的误差降低。进一步结合CA注意力机制及高分辨率分支后，锐化模块的增益得到进一步放大，整体误差水平持续下降。这一结果表明，通过注意力机制的约束以及预训练特征的引导，可以在强化关键点相关高频结构信息的同时，有效抑制锐化模块可能带来的噪声放大问题，从而提升其在关键点定位任务中的整体有效性。

4.3　改进模型在视觉测量领域的影响

在畜牧生产场景中，动物体尺测量、姿态监测及生长状态评估对自动化与实时性提出了更高要求。本文所提出的关键点检测改进方法，在保持较低模型复杂度与良好推理效率的前提下，显著提升了复杂姿态条件下关键点的定位稳定性，尤其在肢体末端及轮廓关键点上表现出更强的鲁棒性。基于上述特性，该方法能够有效支持基于单目RGB图像的非接触式体尺测量与姿态分析，在实际养殖环境中具备良好的应用可行性，可为规模化养殖条件下的个体生长监测与精细化管理提供技术支撑。在此基础上，若进一步结合双目或多视角成像系统，所提出的关键点检测模型有望作为稳定的前端感知模块，为高精度体尺测量与三维结构重建提供可靠的关键点约束。相较于单目测量方式，多视角条件下可在尺度一致性、深度估计精度及复杂姿态下的几何稳定性方面获得进一步提升，从而更好地适应自由运动、视角变化较大等实际养殖场景。

5　结论

1）本研究提出了一种改进的YOLOv8-Pose关键点检测模型，在复杂姿态和多尺度目标场景下表现出较好的检测性能。通过在网络中引入坐标注意力机制、特征锐化模块以及高分辨率P2检测分支，模型在关键点定位精度和整体检测稳定性方面均取得了明显提升。试验结果表明，改进模型在保持较低参数量和计算复杂度的前提下，在*Precision*、*Recall*以及*mAP*50-95等指标上均优于原始YOLOv8-Pose，并在性能与效率之间实现了较好的平衡，验证了所提方法在轻量化关键点检测任务中的有效性。

2）本研究的主要创新体现在针对关键点检测任务特点的结构设计与优化上。坐标注意力机制通过在空间维度上显式建模关键点的位置信息与通道关系，增强了模型对关键点空间分布和结构约束的感知能力；特征锐化模块强化了边缘与局部细节特征，有助于缓解关键点回归过程中细粒度信息被弱化的问题；引入的P2高分辨率分支进一步补充了浅层细节特征，使模型在复杂姿态场景下具备更强的关键点定位能力。上述改进在不显著增加模型复杂度的情况下，有效提升了YOLOv8-Pose的整体性能。

3）尽管改进模型在试验数据集上取得了较为理想的效果，但仍存在一定的局限性。一方面，关键点检测精度仍有进一步提升空间，尤其在姿态变化剧烈、关键

点分布密集的情况下,部分关键点仍存在定位偏差,说明模型在细粒度空间约束建模和高难度关键点回归方面仍有不足;另一方面,当前试验主要基于特定数据集进行评估,尚未在更大规模、更多场景以及更复杂姿态分布的数据上进行充分验证,模型的泛化能力仍需进一步检验。此外,本文工作主要聚焦于单帧图像条件下的关键点检测性能,尚未系统分析模型在视频流或实际应用场景中的时序稳定性与推理效率,在一定程度上限制了结论的普适性。

4)未来研究可从多个方向展开。首先,有必要采集更多贴近真实养殖环境的图像数据,在光照变化、背景复杂度、遮挡情况和姿态多样性等方面增强数据分布的真实性,以提升模型在实际应用中的鲁棒性和泛化能力。其次,可引入更多场景、多姿态和多尺度的数据进行训练与测试,进一步完善模型的关键点表达能力。此外,结合时序信息或视频级建模方法,有望提升关键点检测在动态场景下的稳定性。同时,还可探索更高效的注意力机制或结构重参数化策略,在进一步压缩模型规模的同时持续提升关键点检测精度。最后,将改进模型部署于实际应用系统中,如集成至采摘机器人上,对其在真实环境下的实时性、稳定性和工程可行性进行验证,将有助于推动轻量化关键点检测方法的实际落地。

参 考 文 献

[1] Norton T, Berckmans D, Vranken E, et al. Precision livestock farming: Building 'digital representations' to bring the animals closer to the farmer [J]. Animal, 2019, 13(12): 3009-3017.

[2] 李熙莹, 刘恒, 苏淑妍, 等. 无接触式猪体体尺测量技术研究进展[J]. 华中农业大学学报, 2024, 43(5): 167-177.
Li Xiying, Liu Heng, Su Shuyan, et al. Progress on non-contact measurement of pig body size [J]. Journal of Huazhong Agricultural University, 2024, 43(5): 167-177.

[3] 王克林, 刘龙申, 陈金鑫,等. 人工智能在畜禽养殖机器人中的应用与展望[J]. 智能化农业装备学报(中英文), 2025, 6(2): 44-57.
Wang Kelin, Liu Longshen, Chen Jinxin, et al. Application and prospect of artificial intelligence in livestock and poultry farming robots [J]. Journal of Intelligent Agricultural Mechanization, 2025, 6(2): 44-57.

[4] Russello H, Pezzuolo A, Guarino M, et al. Video-based automatic lameness detection of dairy cows using pose estimation and multiple locomotion traits [J]. Computers and Electronics in Agriculture, 2024, 223: 109040.

[5] Zhao S, Li J, Wang Y, et al. Deep learning-based approach for behavior recognition and pose estimation in cage-reared ducks: leveraging cage-net elimination to propel precision improvement [J]. Computers and Electronics in Agriculture, 2025, 236: 110469.

[6] 董力中, 孟祥宝, 潘明, 等. 基于姿态与时序特征的猪只行为识别方法[J]. 农业工程学报, 2022, 38(5): 148-157.
Dong Lizhong, Meng Xiangbao, Pan Ming, et al. Recognizing pig behavior on posture and temporal featuresusing computer vision [J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(5): 148-157.

[7] 刘莫尘, 褚镇源, 崔明诗, 等. 基于改进YOLO v8-Pose的红熟期草莓识别和果柄关键点检测[J/OL]. 农业机械学报, 1-12 [2025-12-15]. http://kns.cnki.net/kcms/detail/11.1964.S.20230911.2034.046.html.

[8] Pan Y, Chang J, Dong Z, et al. PFLO: a high-throughput pose estimation model for field maize based on YOLO architecture [J]. Plant methods, 2025, 21(1): 51.

[9] Xu Z, Liu Y, Wang H, et al. A geodesic distance regression-based semantic keypoints detection method for pig point clouds and body size measurement [J]. Computers and Electronics in Agriculture, 2025, 234: 110285.

[10] 李明煌, 苏力德, 张永, 等. 基于改进YOLOv8n-pose和三维点云分析的蒙古马体尺自动测量方法[J]. 智慧农业(中英文), 2024, 6(4): 91-102.
Li Minghuang, Su Lide, Zhang Yong, et al. Automatic measurement of mongolian horse body based on improved YOLOv8n-pose and 3D point cloud analysis [J]. Smart Agriculture, 2024, 6(4): 91-102.

[11] 李琦, 刘伟, 赵建敏. 基于双目视觉及Mask RCNN的牛体尺无接触测量[J]. 黑龙江畜牧兽医, 2020(12): 46-50, 159-160.
Li Qi, Liu Wei, Zhao Jianmin. Non-contact measurement of bovine body size based on binocular vision and Mask RCNN [J]. Heilongjiang Animal Science and Veterinary Medicine, 2020(12): 46-50, 159-160.

[12] Ao D, Hao G, Jie L, et al. Automatic livestock body measurement based on keypoint detection with multiple depth cameras [J]. Computers and Electronics in Agriculture, 2022, 198: 107059.

[13] Li R, Zhang Y, Wang X, et al. Automated measurement of beef cattle body size via key point detection and monocular depth estimation [J]. Expert Systems with Applications, 2024, 244: 123042.

[14] Luo X, Zhang H, Li Y, et al. Automated measurement of livestock body based on pose normalisation using statistical shape model [J]. Biosystems Engineering, 2023, 227: 36-51.

[15] Sapkota R, Flores-Calero M, Qureshi R, et al. YOLO advances to its genesis: a decadal and comprehensive review of the You Only Look Once (YOLO) series [J]. Artificial Intelligence Review, 2025, 58(9): 274.

[16] 吴擎，韦润轩，周乐，等．基于改进YOLOv5s的轻量化鲜茶叶识别方法[J]．智能化农业装备学报(中英文)，2025，6(1)：1－14.
Wu Qing，Wei Runxuan，Zhou Le，et al. Lightweight fresh tea leaf recognition method based on improved YOLOv5s [J]. Journal of Intelligent Agricultural Mechanization，2025，6(1)：1－14.

[17] 蔡家一，刘世伟，单龙祥，等．基于机器视觉与YOLO v5的裂纹蛋分拣机器人设计与试验[J]．智能化农业装备学报(中英文)，2025，6(1)：41－50.
Cai Jiayi，Liu Shiwei，Shan Longxiang，et al. Design and experiment of cracked egg sorting robot based on machine vision and YOLO v5 [J]. Journal of Intelligent Agricultural Mechanization，2025，6(1)：41－50.

[18] 陈家政，付根平，黄伟锋，等．基于改进YOLOv5s的云机器人视觉导航方法[J]．智能化农业装备学报(中英文)，2025，6(3)：98－110.
Chen Jiazheng，Fu Genping，Huang Weifeng，et al. Visual navigation method based on improved YOLOv5s and cloud robot [J]. Journal of Intelligent Agricultural Mechanization，2025，6(3)：98－110.

[19] Cao J，Tang H，Fang H S，et al. Cross-domain adaptation for animal pose estimation [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision，2019：9498－9507.

[20] Maji D，Nagori S，Mathew M，et al. YOLO-pose：enhancing YOLO for multi person pose estimation using object keypoint similarity loss [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2022：2637－2646.

[21] Hou Q，Zhou D，Feng J. Coordinate attention for efficient mobile network design [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2021：13713－13722.

[22] Zeiler M D，Fergus R. Visualizing and understanding convolutional networks [C]// European Conference on Computer Vision，Cham：Springer International Publishing，2014：818－833.

[23] Selvaraju R R，Cogswell M，Das A，et al. Grad-cam：Visual explanations from deep networks via gradient-based localization [C]// Proceedings of the IEEE International Conference on Computer Vision，2017：618－626.

[24] Wang H，Wang Z，Du M，et al. Score-CAM：Score-weighted visual explanations for convolutional neural networks [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops，2020：24－25.

[25] Cho Y，Kim D，Saeed S，et al. Keypoint detection using higher order Laplacian of Gaussian [J]. IEEE Access，2020，8：10416－10425.

[26] Wang C Y，Bochkovskiy A，Liao H Y M. YOLOv7：trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2023：7464－7475.

[27] He K，Gkioxari G，Dollár P，et al. Mask R-CNN [C]// Proceedings of the IEEE International Conference on Computer Vision，2017：2961－2969.

[28] Duan K，Bai S，Xie L，et al. Centernet：Keypoint triplets for object detection [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision，2019：6569－6578.

[29] Sun K，Xiao B，Liu D，et al. Deep high-resolution representation learning for human pose estimation [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2019：5693－5703.

[30] Geng Z，Sun K，Xiao B，et al. Bottom-up human pose estimation via disentangled keypoint regression [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2021：14676－14686.

Animal body measurement keypoint detection based on YOLOv8-Pose

Liu Dihua[1]，Si Xiuli[1, 2]，Zhang shumin[3]，Ma Li[4*]

(1. *School of Information Technology*，*Jilin Agricultural University*，*Changchun* 130118，*China*；2. *Jilin Agricultural Artificial Intelligence Research Base*，*Changchun* 130118，*China*；3. *School of Animal Science and Technology*，*Jilin Agricultural University*，*Changchun* 130118，*China*；4. *Wuxi University*，*Wuxi* 214105，*China*)

Abstract: Accurate localization of keypoints for animal body measurement is of great importance for intelligent livestock

health monitoring and automated measurement systems. However, existing keypoint detection models often suffer from two major limitations: a large number of parameters which are unfavorable for deployment on edge devices, and insufficient localization accuracy in some lightweight models that fails to meet the precision requirements of body measurement tasks. To address these challenges, this paper proposes a lightweight keypoint detection model based on an improved YOLOv8-Pose architecture, which effectively enhances detection accuracy while maintaining real-time performance. In the proposed model, coordinate attention mechanisms are integrated into the shallow layers of the backbone network and before the three heads, enabling spatial coordinate information to be effectively embedded into channel attention. This design enhances the model ability to perceive the spatial distribution of keypoints, thereby improving the accuracy and stability of keypoint localization. In addition, a forced sharpening module is introduced in the shallow backbone to strengthen edge and contour features, facilitating a better balance between global semantic representation and local detail preservation. Furthermore, a high-resolution feature branch is incorporated to provide fine-grained spatial information to the three detection heads, compensating for the loss of spatial resolution in deeper feature maps. Experimental results demonstrate that the proposed model achieves a precision of 82.6%, a recall of 81.3%, and an *mAP*50-95 of 58.1% on complex horse keypoint detection tasks. Compared with YOLOv8-Pose, the proposed method achieves improvements of 21.0%, 39.2%, and 45.3% in *precision*, *recall*, and *mAP*50-95, respectively, exhibiting superior robustness and localization capability under complex poses and diverse environmental conditions. In terms of model complexity, the number of parameters increases from 3.361 M to 4.611 M, corresponding to an increase of approximately 37.2%, while the floating-point operations rises from 9.6×10^9 to 10.4×10^9, an increase of only 8.3%. Despite this limited computational overhead, the proposed model delivers substantial performance gains without significantly affecting inference speed. These results indicate that the proposed approach achieves a favorable trade-off between accuracy and efficiency by incurring a relatively small model cost for considerable performance improvement, demonstrating its strong potential for intelligent livestock farming and automated body measurement applications.

Keywords: YOLOv8-Pose; animal keypoint detection; local feature enhancement; coordinate attention; intelligent livestock farming

Foundation items: Subproject of the National Key R&D Program of China (20200136); Jilin Provincial Science and Technology Development Program Project (20240303066)

First Author: Liu Dihua, E-mail: liudihuaemail@163.com

* **Corresponding Author:** Ma Li, E-mail: mali@cwxu.edu.cn