

追踪提示结合视觉大模型微调的猪只实例分割方法

孙立博 张泽昀 秦文虎
(东南大学仪器科学与工程学院, 南京 210096)

摘要: 针对当前主流猪只感知算法存在追踪与分割任务缺乏协同、依赖高成本人工标注的难题,提出了一种端到端猪只追踪与分割框架。该方法将追踪算法输出的锚框作为视觉大模型 SAM (Segment anything model) 的动态提示,以生成时序连贯且身份一致的个体掩模序列。通过结合低秩自适应 (Low rank adaptation, LoRA) 与猪只多尺度特征适配器 (PF Adapter),在大规模数据集上对 SAM 进行高效微调,使该模型仅需少量标注样本,即可在性能上超越传统的有监督学习算法,有效克服了对高昂人工标注的依赖。实验结果表明,微调后的 SAM 模型在两种不同的数据集上的平均像素准确率、交并比和 Dice 相似指数分别达到了 91.24%、85.06% 和 86.18%,参数量相较于 SAM 基准模型只增加了 3%。改进模型输出的猪只体态和运动轨迹信息可用于猪只评估和母猪分娩预警等,为智慧养殖提供技术支撑。

关键词: 猪只; 实例分割; 追踪提示; 视觉大模型 (SAM); LoRA; PF Adapter

中图分类号: TP391 **文献标识码:** A **文章编号:** 1000-1298(2026)15-0036-10

OSID:



Pig Instance Segmentation Based on Tracking Prompts and Fine-tuned Large Vision Models

Sun Libo Zhang Zeyun Qin Wenhui
(School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China)

Abstract: Aiming to address the challenges of lack of coordination between tracking and segmentation tasks and reliance on costly manual annotation in current mainstream pig perception algorithms, a novel end-to-end framework was proposed. Specifically, the proposed method ingeniously leveraged the bounding boxes output by the multi-object tracking algorithm as dynamic spatial prompts for the segment anything model (SAM). This integration facilitated the generation of temporally coherent and identity-consistent individual mask sequences across video frames, bridging the gap between localization and pixel-level segmentation. To adapt the vision foundation model to the specific agricultural domain, the low-rank adaptation (LoRA) was combined with a custom-designed pig multi-scale feature adapter (PF Adapter). The model required only a minimal number of annotated samples to achieve robust feature extraction, ultimately surpassing the performance of traditional fully supervised learning algorithms and effectively overcoming the bottleneck of expensive data annotation. Comprehensive experimental results demonstrated the superiority of the proposed framework. Evaluated across two distinct datasets, the fine-tuned SAM achieved impressive performance metrics, with pixel accuracy (PA) of 91.24%, an intersection over union (IoU) of 85.06%, and a dice similarity coefficient of 86.18%. Remarkably, this significant performance enhancement was achieved with merely a 3% increase in number of parameters compared with the baseline SAM architecture. Furthermore, the pig body posture and movement trajectory information output by the improved model can be used for pig evaluation and sow farrowing warning, providing technical support for smart farming.

Key words: pig; instance segmentation; tracking prompts; large vision model (SAM); LoRA; PF Adapter

收稿日期: 2026-03-17 修回日期: 2026-04-12

基金项目: 江苏现代农业产业单项技术研发项目 (CX(23)3120)

作者简介: 孙立博 (1983—), 女, 副教授, 博士, 主要从事计算机视觉和人工智能研究, E-mail: sunlibo@seu.edu.cn

通信作者: 秦文虎 (1969—), 男, 教授, 博士, 主要从事农业物联网感知技术研究, E-mail: qinwenhu@seu.edu.cn

0 引言

在现代生猪集约化养殖模式下,高密度的猪群环境使得个体健康状况的实时监测与精细化管理成为行业亟待解决的难题^[1-2]。现代生猪养殖业亟需保障猪只健康和提升生产效率。随着传统养殖向智慧化模式的加速转型,行业通过集成先进传感、数据分析与人工智能等技术,实现对猪只个体及养殖环境的精细化监测与管理。

然而,实际养殖复杂场景使得常规视觉监测面临巨大挑战^[3]。为此,高精度实例分割已成为推动精准畜牧业发展的关键^[4-5]。通过实例分割提取像素级轮廓以获取关键形态学数据,从而为猪只体况精细评估、异常行为识别以及母猪临产预警等智能化应用奠定坚实基础。

近年来,视觉基础模型^[6-7]凭借其强大的泛化能力和小样本学习特性,已成为推动农业智能化发展的关键技术。例如,Cao 等^[8]以麦穗作为参考对象,评估了 SAM 在作物语义分割方面的性能。Saeidifar 等^[9]通过预处理技术微调视觉大模型,使其能够自动选择目标点作为 SAM 的提示点,这提高了鸡只的识别精度以及鸡蛋的分割精度。

在生猪养殖领域,猪只的体态蕴含着丰富的行为特征信息,获取多视角猪只逐帧掩模对于猪只状态监测和健康评估意义重大。目前,传统方法将猪只追踪与实例分割任务分开处理,在猪只追踪领域,Zhang 等^[10]提出了一种分层数据关联框架,该框架集成了 CNN 的检测器与 Bytetrack,实现了在低光照等挑战性条件下提升了猪只追踪的鲁棒性。Cowton 等^[11]将 Faster R-CNN 与两种实时多目标追踪算法相结合,构建了一个全自动的猪只定位与追踪系统。Tran 等^[12]采用了 YOLO v7 结合 DeepSORT 来进行实时的猪只检测与追踪,从而对健康猪只日常行为轨迹进行建模。Zhou 等^[13]提出了一种结合光流法和多目标追踪算法来追踪猪只的方法。通过结合光流与追踪算法,能够精确估计每只猪的运动情况。

在猪只实例分割方面,赵宇亮等^[14]采用 DeepLabCut 算法来解决低对比度图像输入条件下的实例分割问题,并以此为基础设计了一种猪只体尺快速测量方法。Lu 等^[15]提出了一种基于 Swin Transformer 的框架,该框架即使在低分辨率条件下也能精确分割个体猪只,精度达到 86.9%。Tu 等^[16]扩展了 Mask Scoring R-CNN 框架应用于群养猪只,并在复杂环境中展示了强大的鲁棒性,F1 分数高达 0.940 5。王海燕等^[17]提出一种名为 PigNet 的深度卷积神经网络模型,以解决群养猪只因身体

粘连造成的分割瓶颈,在验证集上的分割准确率达到到了 85.40%。由于分割过程无法结合猪只追踪提供的时序运动与身份信息,仅依赖单帧图像,导致无法在视频中为每个特定猪只生成时序连贯、身份明确的高精度掩模序列。此外,基于有监督学习的猪只追踪和实例分割算法依赖大量的人工标注,这一标注过程不仅耗时、成本高昂且效率低下,同时还易引入主观偏差,从而限制了这些算法在大规模猪只养殖场场景下的应用。

为解决上述问题,本文提出一种端到端猪只追踪与分割框架,其核心是将追踪结果作为分割的引导提示。首先利用 YOLO v5s 与 ByteTrack 算法实现对猪只的连续跟踪与 ID 分配,然后将追踪框作为提示输入视觉大模型(SAM),以驱动其生成时序连贯且身份一致的高精度猪只掩模。为了更好地将 SAM 从通用分割任务迁移至猪只实例分割任务,本研究提出一种结合低秩适应(Low rank adaptation, LoRA)与猪只多尺度特征适配器(Pig multi-scale feature adapter, PF Adapter)的高效微调策略,使模型仅需在少量标注样本下进行训练,即可在性能上超越传统监督式模型,有效克服人工标注成本高的瓶颈。

1 猪只追踪与实例分割

1.1 方法框架

本文提出了一种基于视觉大模型的猪只追踪与实例分割方法。利用 YOLO v5s 结合 ByteTrack 的方法对猪只进行跟踪,然后再将追踪框作为动态提示,引导微调后的 SAM 对每一帧中的猪只个体进行实例分割。最终得到的结果是视频流中个体猪只的高精度序列化掩模。系统分为两部分:猪只追踪模块和猪只实例分割模块,算法框架如图 1 所示。

1.2 猪只追踪算法

猪只跟踪采用了两阶段的多目标跟踪框架。猪只检测器部分,参考了 Tu 等^[18]的研究成果,采用 YOLO v5s 作为猪只检测器。在自建数据集上,对 YOLO v5s 模型进行了训练,以确保其在实际养殖场场景中的准确性和鲁棒性。前后帧检测框匹配算法基于 ByteTrack^[19]构建。

ByteTrack 算法是构建于基于交并比(IoU)的锚框匹配策略之上。在目标被部分或完全遮挡的场景下,ByteTrack 通过有效利用低置信度检测框与轨迹的二次匹配策略,在多目标跟踪任务中展现出更强的鲁棒性。该方法可有效缓解猪舍密集遮挡场景下的目标 ID 频繁切换与丢失问题,从而大幅提升了在高遮挡条件下的跟踪鲁棒性。

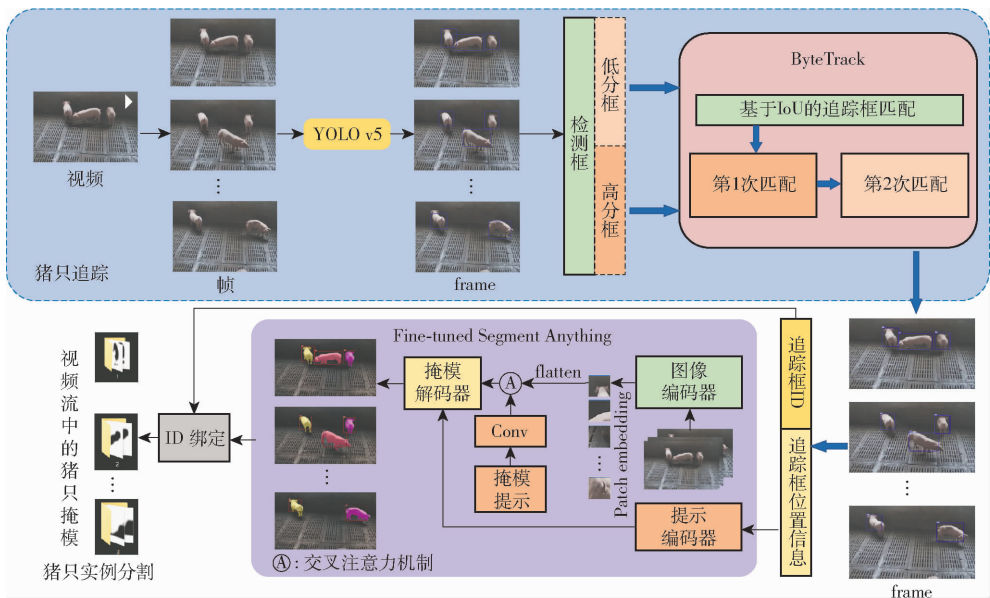


图 1 基于视觉大模型 SAM 的猪只追踪与实例分割算法框图

Fig. 1 Pig tracking and instance segmentation based on large vision model SAM

1.3 猪只实例分割算法

本研究采用视觉大模型 SAM^[6] 进行猪只实例分割。该方法将 1.2 节中追踪器所生成的追踪框坐标作为提示输入到 SAM 中, SAM 则会根据这些提示执行实例分割。为了提升 SAM 的分割精度, 采用 LoRA 对 SAM 主干的参数进行了微调。同时, 还在 SAM 的模型掩模解码器 (Mask Decoder) 的 MLP 预测层前集成了 PF Adapter。经过 LoRA 微调后的 SAM 的 ViT-B 图像编码器共有 12 层 Transformer Block。本研究将其按深度划分为 3 个特征提取阶段: LR_Block 1~4、LR_Block 5~8 与 LR_Block 9~12。分别对应底层空间特征提取阶段和高层全局语义特征提取阶段。这 3 个阶段输出的图像特征与解码器进行多头注意力交互, 并将增强后的特征注入 PF Adapter 进行多尺度特征融合; 显著增强了模型在猪只尺度多变场景下猪只实例分割鲁棒性。微调后 SAM 结构如图 2 所示。

1.3.1 低秩矩阵微调模块 LR_block

本研究以 SAM 的 ViT-B 架构为基础框架, 针对其视觉 Transformer 图像编码器模块提出了一种 LoRA^[20] 的参数微调模块 LR_Block。该方法在保持预训练模型参数冻结的前提下, 仅需微调约 5% 的可训练参数即可实现模型优化。LR_Block 及其内部详细架构如图 3 所示。

在 SAM 图像编码器的前馈网络的 MLP 层插入了低秩矩阵 A 和 B , 其中两个矩阵的秩 $r=8$, 在训练时仅更新低秩矩阵 A 、 B 的参数。对于 SAM 图像编码器的原始权重 W_0 , 将其分解成 W_0 与动态调节参数 ΔW 之和。在训练的过程中冻结 W_0 的基本参

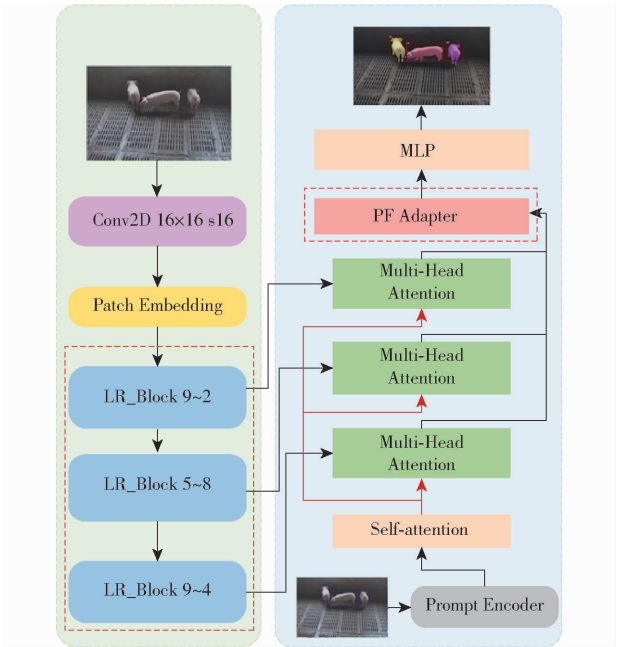


图 2 微调后 SAM 框图

Fig. 2 Structure of fine-tuned SAM

数, 使其不接受梯度更新, 只更新低秩矩阵 A 和 B 中的参数。最终对于原始的前向传播过程, 微调后的前向传播过程公式为

$$h = (W_0 + \Delta W)x = W_0x + \Delta Wx = W_0x + \frac{\alpha}{r}ABx \quad (1)$$

- 式中 h ——前向传播过程
 x ——输入特征
 α ——低秩矩阵缩放因子
 r ——低秩矩阵的秩

在模型训练的初始阶段, 采用随机高斯分布对矩阵 A 进行初始化, 同时将矩阵 B 初始化为零矩阵。因此, 在训练开始, ΔW 的值为零矩阵。随后,

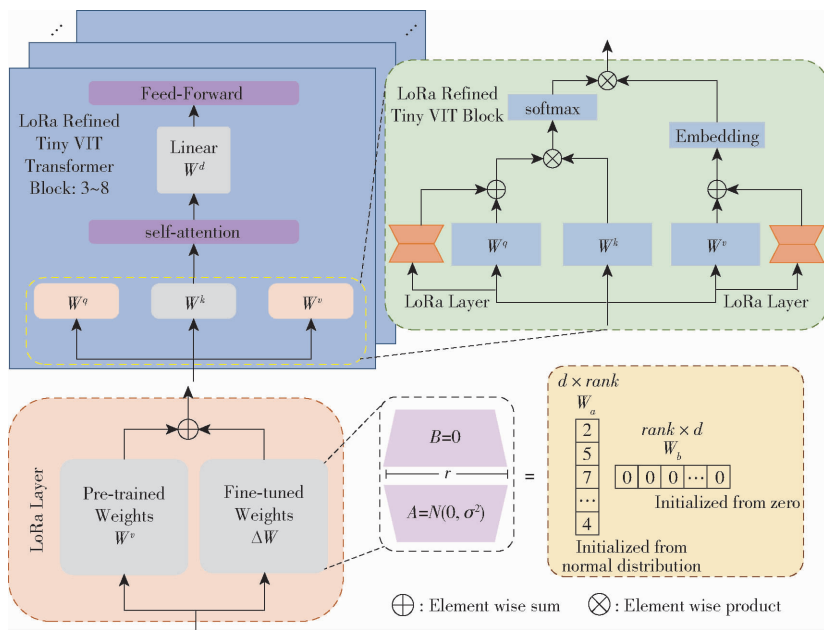


图 3 LR_block 架构

Fig. 3 Architecture of LoRA refined VIT block (LR_block)

依据公式

$$\Delta W = \frac{\alpha}{r} \mathbf{A} \mathbf{B} \quad (2)$$

对 ΔW 进行缩放操作, 其中 α 是与 r 相关联的一个常数。得益于 Adam 优化器的特性, 调整学习率基本等同于调整 α 。因此, 可以将 α 固定初始设定值, 无需后续修改。

1.3.2 猪只多尺度特征适配器 PF Adapter

为了让 SAM 骨干网络更好地适应猪实例分割任务, 本研究提出了 PF Adapter。该模块旨在高效融合由编码器输出的多尺度特征, 从而为后续的图像解码器和 MLP 提供输入。PF Adapter 参考了 FPN^[21] 的自顶向下多尺度融合思想, 使用上采样与侧向连接的方式将高层语义信息逐级传递到低层图像特征中; 同时又通过下采样与逐层聚合的方式, 将低层图像特征中的空间细节反馈到 FPN 结构的深层节点中。

具体来说, 首先对底层空间特征 B3 进行上采样, 得到 32×32 的特征图 C2, 然后通过 DCFF 模块对特征图 C2 与特征图 B2 进行特征融合, 并经过上采样得到 64×64 的特征图 C1。特征图 C1 与特征图 B1 通过 DCFF 模块进行特征融合, 下采样后得到特征图 D1。特征图 D1 与特征图 C1 通过逐元素相加和 C2F 模块进行深度特征融合, 下采样后得到 32×32 的特征图 D2。最后, 特征图 D2 与特征图 C2 通过逐元素相加和 C2F 模块进行深度特征融合, 下采样后得到 16×16 的特征图 D3。此特征被送入后续的线性层以实现高质量的实例分割, 其详细结构如图 4a 所示。

DCFF 模块使用残差结构来保留初始信息, 并采用可变形卷积自适应地对齐几何变换下的特征。DCFF 的框架如图 4c 所示。给定一个高层级特征图 F_h 和一个低层级特征图 F_l , 该模块首先应用 1×1 卷积, 随后进行通道拼接和一个 3×3 卷积来融合两种特征, 公式为

$$F_{fused} = \text{Conv}_{3 \times 3}([\text{Conv}_{1 \times 1}(F_h), \text{Conv}_{1 \times 1}(F_l)]) \quad (3)$$

式中 F_{fused} ——融合后的特征
 F_h ——高层特征 F_l ——低层特征
 $\text{Conv}_{3 \times 3}$ —— 3×3 卷积层
 $\text{Conv}_{1 \times 1}$ —— 1×1 卷积层

此外, 为增强空间适应性, DCFF 模块中应用了可变形卷积 T_{def} , 公式为

$$T_{def}(F_{fused}, p) = \sum_{k \in R} w_k F_{fused}(p + k + \Delta_{pk}) \quad (4)$$

式中 R ——规则采样网格 w_k ——卷积权重
 Δ_{pk} ——模型学习到的空间偏移量
 p ——当前像素位置
 k ——第 k_{th} 个采样权重

输出是通过对原始特征与增强后的特征进行残差融合计算得出的, F_{out} 计算公式为

$$F_{out} = F_{fused} + T_{def}(F_{fused}) \quad (5)$$

DCFF 模块增强了空间和语义特征的表达能力, 从而为下游的解码任务生成了高质量的多尺度表示。

PF Adapter 的下采样路径中应用了 YOLO v8 中的 C2F 模块, 如图 4b 所示。它将输入通道分割: 一部分通过一个轻量级的 BottleNeck 模块堆栈, 而另

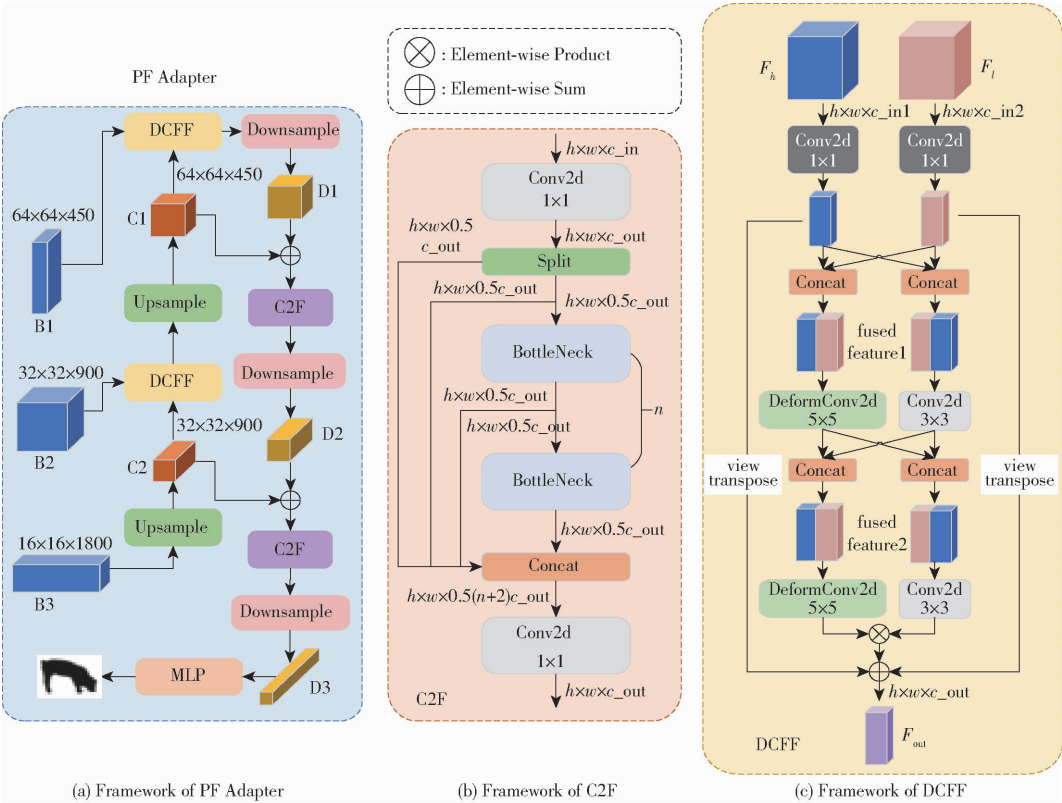


图 4 猪只特征适配器 (PF Adapter) 及其核心模块架构图

Fig.4 Architecture of proposed PF adapter and its core modules

一部分则直接绕过。两部分的输出随后被拼接,并通过一个 1×1 卷积进行融合。

1.3.3 LR_Block 和 PF Adapter 的联合训练策略和损失函数

在模型参数初始化阶段,对于 LR_Block 中引入的低秩矩阵 $\mathbf{A}$,采用标准高斯分布进行随机初始化,而低秩矩阵 $\mathbf{B}$ 则初始化为零矩阵。对于 PF Adapter 内的模块,则使用标准的随机初始化方法,并将可变形卷积的偏移量设置为 0.5。

训练策略为两阶段的训练策略。第 1 阶段是 PF Adapter 的训练。在此阶段,SAM 骨干网络及 LR_Block 被冻结。仅对 PF Adapter 进行训练,采用较高的初始学习率,并通过余弦退火策略逐步降低。这确保了 PF Adapter 能够快速收敛并高效地融合多尺度特征,同时避免了对骨干网络参数的干扰。第 2 阶段是 LR_Block 与 PF Adapter 的协同优化。在 PF Adapter 初步收敛后,LR_Block 的低秩矩阵 $\mathbf{A}$ 和 $\mathbf{B}$ 被解冻,并与 PF Adapter 进行联合优化,而 SAM 骨干网络的其余部分仍然保持冻结。PF Adapter 的参数维持一个较高的学习率,而 LR_Block 的学习率则设置得较低以进行微调,从而防止破坏 PF Adapter 预先学到的特征。

损失函数主要是用于 LR_Block 低秩矩阵中的参数更新。本研究选择了 $Loss(Dice)$ 作为猪只实例

分割的损失函数。这是因为在猪圈场景下,背景像素往往占据绝大部分,如果只使用交叉熵损失函数会导致模型倾向于将所有像素都预测为背景。而 $Loss(Dice)$ 能更好地处理这种不平衡情况,使得模型能够更好地区分前景和背景。 $Loss(Dice)$ 对于边界像素的预测有较好的效果。其计算公式为

$$Loss(Dice) = 1 - \frac{2 \sum P \cap Y}{\sum |P| + \sum |Y|}$$

(6)

式中 $\sum |P|$ ——预测结果中标签的值
 $\sum |Y|$ ——真实结果中标签的值

1.4 猪只掩模与运动轨迹应用分析

本研究提出的端到端猪只追踪与分割方法能够生成高质量的多视角妊娠母猪掩模与运动轨迹,为后续的健康监测与母猪妊娠状态分析提供了可靠的数据基础。

为确保基于掩模的猪只形态学参数测量的几何准确性,本研究参考文献[22]的研究,引入透视变换对原始掩模进行畸变校正。校正前后样例如图 5 所示。

首先,在原始图像中提取猪圈边界的 4 个角点作为基准参考点,通过建立其与标准矩形平面(尺寸为 $H \times W$)的映射关系,构建原始透视视角向标准侧视视角的投影矩阵 $\mathbf{M}$ 。映射公式为

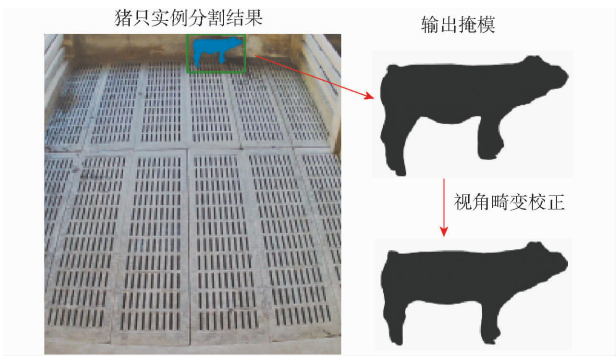


图 5 猪只掩模畸变校正样例

Fig. 5 Example diagram of pig mask distortion correction

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \mathbf{M} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{7}$$

式中 (x', y', z') ——投影变换后的齐次坐标

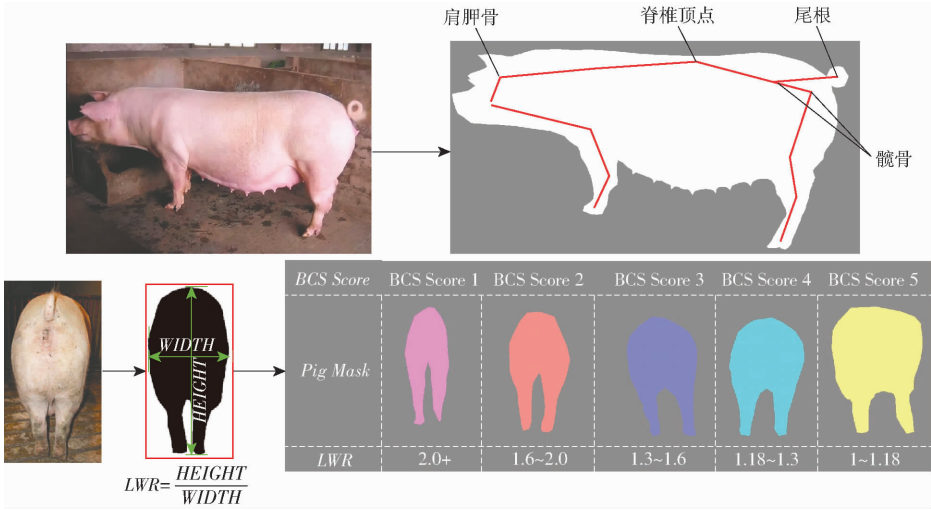


图 6 基于掩模分析的猪只状态监测应用实例

Fig. 6 Application example of pig state detection based on mask analysis

的母猪作为重点监测对象,实施全天候连续追踪。重点监测母猪一系列重复且空间集中高的异常运动特征,如原地高频绕圈、持续刨地等^[24],并将其作为判断母猪是否即将分娩的关键依据。

本研究提出的算法可以实时记录追踪框的中心点坐标,从而得到母猪的运动轨迹。通过分析同一时间窗口内母猪的运动轨迹特征,可识别分娩前兆的行为模式,进而实现母猪分娩的提前预警。如图 7 所示, ID1 和 ID3 分别代表处于静卧与进食状态的猪只,其轨迹特征表现为活动范围有限、移动频率较低。而 ID2 所对应的母猪则呈现出较高的空间重复性和高频绕圈行为,明显偏离正常的行为范式,具备明显的分娩信号。

2 实验

2.1 数据集构建

数据集主要由两部分组成:一部分采自江苏省

(x, y) ——原始图像中的像素坐标

在获得投影矩阵 $\mathbf{M}$ 后,将猪只掩模的所有像素点统一映射至标准坐标系中。并通过齐次坐标归一化,得到二维物理坐标 (x'', y'') ,其公式为

$$(x'', y'') = \left(\frac{x'}{z'}, \frac{y'}{z'} \right) \tag{8}$$

在对掩模进行视角变化之后,结合关键点检测技术提取肩胛骨、尾根等骨架特征。通过融合标准体长、体宽构建多参数体质量回归模型^[23],实现了猪只体重的评估。同时,利用尾部视角的掩模最小外接矩形长宽比(LWR)作为体况评分指标,也为自动化猪只体况评分提供了新的途径。基于掩模分析的猪只状态监测应用实例如图 6 所示。

对妊娠母猪的分娩预警,可以结合本研究所提算法进行针对性的监测。选取妊娠周期超过 15 周

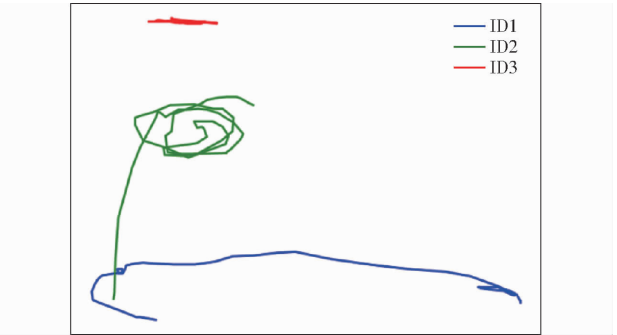


图 7 怀孕母猪历史运动轨迹示意图

Fig. 7 Diagram of historical movement trajectory of sows

镇江市丹徒区的私有数据,另一部分选自 Posta 等^[25]的公开数据集。

本研究所使用的私有数据集采自江苏省镇江市丹徒区的实验猪场。猪舍中单间猪圈尺寸为 6.4 m × 3.2 m,每圈猪只平均为 8 头。实验对象品种均为大白猪,生理阶段处于育肥期;其中部分实验母猪处于妊娠中期(怀孕 6 ~ 7 周)。

本研究在一天中的 4 个关键时段(早上、上午、下午、夜间)连续 24 h 录制视频,单个视频平均时长为 100 min。图像采集方式是通过 RGB 相机进行猪圈视频采集,相机采用导轨吊装的方式架设在猪圈上方 2.5 m 处,RGB 相机的型号为海康威视 MV-CL042-91GC,摄像头通过网线连接到交换机,交换机将采集到的视频传输到本地服务器中,并在这台服务器上进行处理和分析。视频数据以 24 f/s 的速率进行采集,并全天连续记录。夜晚场景下,启用该相机的夜景模式进行拍摄。为防止数据冗余和过拟合,采用动态抽帧策略(高频活动场景间隔 20 s,低频活动场景间隔 120 s)对采集到的视频切帧以获取图像。私有数据集共计 2 600 幅图像。此外,本研究选用 Psota 等^[25]提供的大白猪公开数据集作为补充,依据图像相似度

进行严格筛选,最终筛选出 1 200 幅图像。两部分数据集最终整合成规模为 3 800 幅图像的猪只数据集。

在数据标注阶段,首先将收集整理的全部 3 800 幅图像尺寸统一调整为 640 像素 × 640 像素,并按比例 3:1:1 划分为训练集、验证集和测试集,使用 CVAT 工具完成标注,构建了猪只追踪数据集。随后,为了满足更精细的分析需求,从 3 800 幅图像中进一步筛选了 800 幅具有代表性的高质量图像作为实例分割数据集,使用 X-Anylabeling 工具进行实例分割标注后,构建了猪只实例分割数据集。猪只实例分割数据集同样遵循相同的尺寸标准和划分比例。图 8 为猪只追踪和实例分割标注示意图,图中上、中、下行分别为原始输入图像、目标追踪结果、实例分割掩模结果。

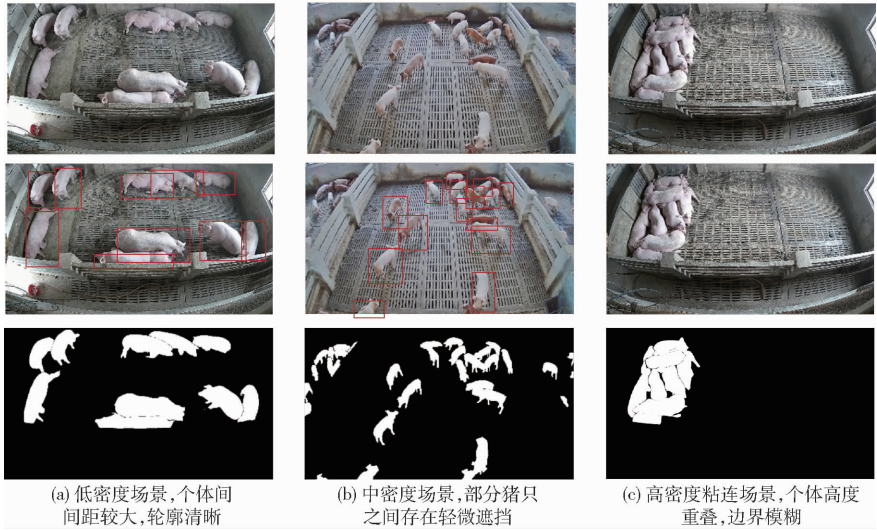


图 8 猪只追踪和实例分割标注示意图

Fig. 8 Sample of pig tracking and instance segmentation labeling

2.2 猪只追踪模型训练结果

将 ByteTrack 算法与主流的 YOLO v5s 检测器进行了有效整合。该组合模型使用了 2.1 节中描述的猪只追踪数据集进行训练。训练过程中,图像被统一标准化为 640 像素 × 640 像素,训练 200 个轮次,训练过程中采用余弦退火策略来调整学习率。

在测试集上,该模型在猪只多目标跟踪任务中取得了优越的性能,其中,多目标追踪准确率(MOTA)达到 91.9%,多目标追踪框精确率(MOTP)为 23.6%。实验结果表明,该模型在追踪稳定性和追踪框位置精度上均表现出色。这种高精度的追踪能力为下游的 SAM 提供了高质量的提示,从而为基于 SAM 的猪只实例分割任务奠定了坚实的基础。

2.3 猪只实例分割对比实验

在猪只实例分割实验部分,选取了 5 种代表性

的主流有监督学习模型:Mask R-CNN^[26]、YOLO v8-seg^[27]、SOLO v2^[28]、Sotr^[29]、YOLO 11-seg 与本研究提出的模型进行性能对比。这些模型均在 2.1 节描述的实例分割数据集上进行了训练。所有图像被统一标准化为 640 像素 × 640 像素,并采用余弦退火学习率策略训练了 200 个周期。

在对比实验中,经过微调的 SAM 模型展现了出色的分割精度,该模型在两种不同的测试场景中均优于所有基线模型。对比实验结果如表 1 所示。

在以猪只频繁遮挡为特征的场景 1 中,该模型实现了 90.85% 的像素准确率(PA)、84.63% 的交并比(IoU)和 85.97% 的 Dice 系数。在主要挑战为猪只间高度粘连的场景 2 中,3 项指标分别达到了 91.63%、85.48% 和 86.39%。相较于场景 1,该模型在场景 2 上相较于有监督算法的领先幅度有所下降。这是因为猪只静卧状态下的图像之间相似度较

高,这为基于 CNN 的监督式模型创造了理想的拟合条件。基于 Transformer 的 Sotr 模型由于在小数据集上收敛性较差且收敛速度最慢,在所有精度指标上一直排名最后。实时性能上,YOLO 11-seg 表现最佳,达到 28 f/s。本研究所提出的模型实现了 10 f/s 的中等推理速度,略优于 Mask R-CNN 的 8 f/s。

表 1 猪只实例分割对比实验结果

Tab.1 Results of comparative experiment on pig instance segmentation on two test sets

	模型	PA/	IoU/	Dice	帧率/
		%	%	系数/%	(f·s ⁻¹)
场景 1	Mask R-CNN	83.56	77.11	79.15	8
	YOLO v8-seg	77.21	70.32	71.87	23
	SOLO v2	84.56	78.27	79.22	17
	Sotr	70.12	65.34	60.46	12
	YOLO 11-seg	78.26	74.18	71.64	28
	Ours	90.85	84.63	85.97	10
场景 2	Mask R-CNN	88.84	82.19	83.31	8
	YOLO v8-seg	82.54	78.49	83.45	23
	SOLO v2	86.75	80.67	82.29	17
	Sotr	76.61	70.49	74.27	12
	YOLO 11-seg	83.62	80.04	83.91	28
	Ours	91.63	85.48	86.39	10

2.4 猪只实例分割消融实验

为了减少计算负担并提高推理速度,本研究选择了 SAM 模型中最小的主干网络版本 ViT-B 进行消融实验,以比较其在微调前后的性能。消融实验的结果见表 2,其可视化结果如图 9 所示。

在场景 1 中,两头猪只相互啃咬,这给实例分割带来了很大的挑战。在集成了 PF_Adapter 和 LR_Block 模块后,模型输出了更准确的分割结果,能精确区分两头猪之间的边界(S1)。在场景 2 中,受右侧墙壁背景的干扰,靠近该区域的猪只在分割过程中出现了严重的背景泄漏问题。在加入了 PF_Adapter 和 LR_Block 模块后,模型有效地减少了墙壁干扰,并生成了更精确的猪只掩模(S2)。在场景 3

中,针对远距离的小尺度目标,基准模型难以保留精细细节。在加入了 PF_Adapter 和 LR_Block 模块后,改进后的模型能够精准分割出这些目标,并完整保留了肢体轮廓等细节特征(S3)。在场景 4 中,两头猪只高度粘连,添加 PF_Adapter 和 LR_Block 模块后,模型不仅能有效分割粘连的猪只,还能保留腿部等重要结构细节(S4)。

表 2 猪只实例分割消融实验结果

Tab.2 Results of ablation experiment on pig instance segmentation on test sets

LR_	PF	PA/	IoU/	Dice	参数量
Block	Adapter	%	%	系数/%	
×	×	83.34	79.91	81.26	9.65 × 10 ⁶
√	×	89.16	82.51	82.81	9.65 × 10 ⁶
√	√	90.85	84.63	85.97	9.94 × 10 ⁶

注:√表示采用此模块,×表示不采用此模块。

2.5 输入提示对实例分割性能的影响

为了评估微调后的 SAM 模型对其输入提示的敏感性,设计了一项由两部分组成的受控实验。本研究在 2.1 节详述的实例分割测试集上进行了实验,实验结果如表 3 所示。可视化结果如图 10 所示。

首先,通过模拟常见的提示框误差来分析模型对提示框几何误差的容忍性。为此,设计了 3 种提示干扰策略:随机均匀缩放,模拟过大的提示框;随机平移,模拟位置偏差;随机收缩与平移,模拟目标不完整捕获。随后,将模型在这些受干扰提示下的分割性能,与使用精确真实边界框作为提示时的性能进行对比。其次,研究了模型对不同提示方式的响应。本研究比较了两种提示下的分割性能:仅使用锚框作为提示;在锚框的基础上,补充锚框中心点作为提示。

在输入为均匀缩放的追踪框时,仅使用边界框的提示非常容易受到背景泄漏的影响,从而影响分割准确性。相比之下,“锚框+点”的提示方式显示

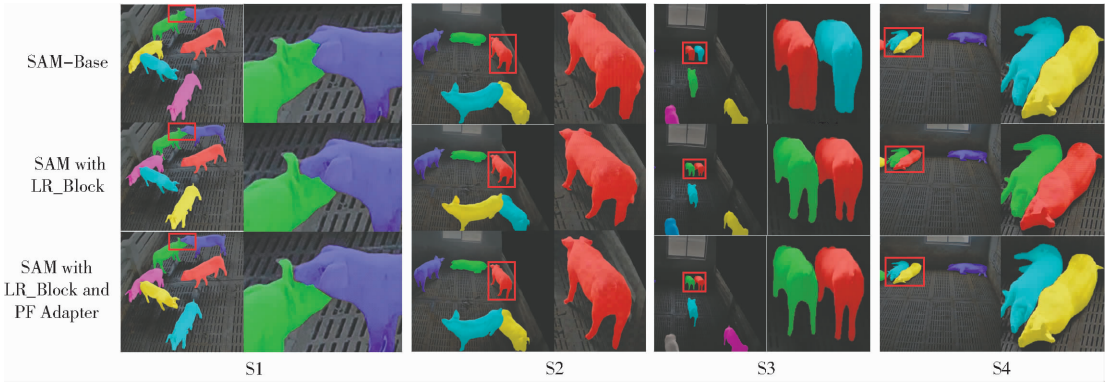


图 9 消融实验的可视化结果

Fig.9 Visualization of ablation study results under various challenging scenarios

表 3 不同输入提示时实例分割性能

Tab.3 Segmentation performance of tracking box precision and prompting modalities

提示模式		PA/ %	IoU/ %	Dice 系数/%	推理时 间/ms
精确边界框	锚框	90.63	84.67	85.51	95.27
	锚框 + 点	90.85	84.63	85.97	98.63
随机均匀缩放	锚框	75.33	70.15	71.27	95.27
	锚框 + 点	84.28	77.39	80.24	98.63
随机平移	锚框	73.83	65.25	68.87	95.27
	锚框 + 点	81.15	76.49	77.26	98.63
随机收缩与平移	锚框	72.74	64.38	67.64	95.27
	锚框 + 点	79.74	70.43	75.76	98.63

出更好的鲁棒性,由于锚框中心点在缩放过程偏移较小,这一特性有效补偿了因边界框过大所带来的定位模糊。

当追踪框受到随机平移干扰时,在杂乱场景中,仅使用边界框的提示容易出现精度下降问题,例如分割出完全错误的猪只个体。而“锚框 + 点”的提示策略通过使用锚框中心点来消除目标的歧义,显著减轻了这种问题。

在输入为随机缩放且平移的猪只追踪框下,仅使用边界框的提示往往会产生不完整的掩模。相反,采用“锚框 + 点”的提示策略能够利用锚框中心点信息引导模型输出更完整的猪只掩模。

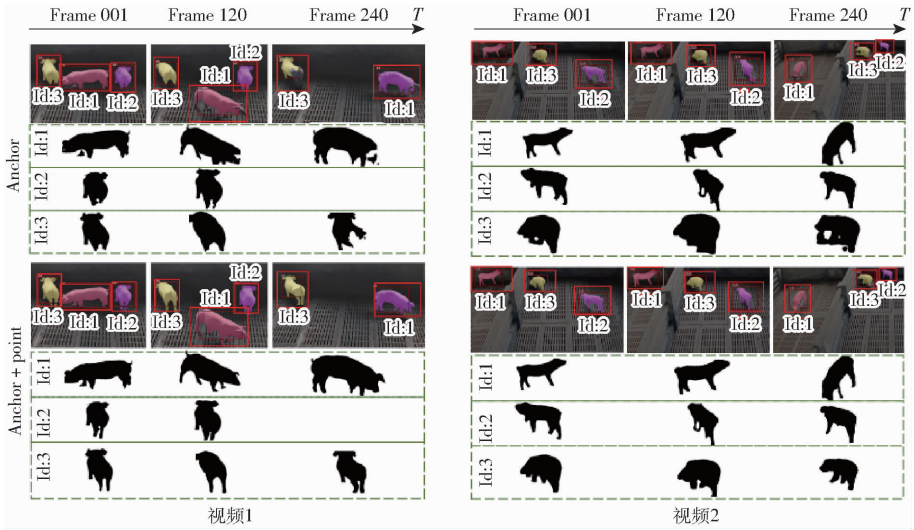


图 10 不同追踪框精度和提示方式下猪只实例分割的可视化结果

Fig. 10 Visualization of segmentation results on two prompt modalities at different tracking box accuracies

3 结论

(1)提出基于追踪框提示的 SAM 实例分割框架:将多目标追踪生成的边界框作为 SAM 的动态提示,在少量人工标注下实现了时序连贯且身份一致的猪只掩模生成,大幅降低了对大规模标注数据的依赖。

(2)构建基于 LoRA 的参数高效微调策略:仅需微调 SAM 图像编码器约 5% 的参数,即实现了从

通用分割向猪只分割任务的高效迁移,成功兼顾了训练效率与分割精度。

(3)设计猪只多尺度特征适配器(PF Adapter)。针对固定视角下猪只尺度不一的问题,通过在掩模解码器前引入双向多尺度特征融合,显著提升了分割鲁棒性。实验表明,微调后的模型像素准确率(PA)、交并比(IoU)和 Dice 系数分别达 91.24%、85.06% 和 86.18%,参数量相较于 SAM 基准模型只增加了 3%,整体性能优于传统有监督模型。

参 考 文 献

[1] Silapachote P, Srisuphab A, Banchongthanakit W. An embedded system device to monitor farrowing [C] // 2018 5th International Conference on Advanced Informatics; Concept Theory and Applications (ICAICTA). IEEE, 2018: 208 – 213.

[2] Vranken E, Berckmams D. Precision livestock farming for pigs[J]. Animal Frontiers, 2017, 7(1): 32 – 37.

[3] García R, Aguilar J, Toro M, et al. A systematic literature review on the use of machine learning in precision livestock farming [J]. Computers and Electronics in Agriculture, 2020, 179: 105826.

[4] Koknaroglu H, Akunal T. Animal welfare: an animal science approach[J]. Meat Science, 2013, 95(4): 821 – 827.

[5] Akkem Y, Biswas S K, Varanasi A. Smart farming using artificial intelligence: a review[J]. Engineering Applications of Artificial Intelligence, 2023, 120: 105899.

[6] Kirillov A, Mintun E, Ravi N, et al. Segment anything[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 4015 – 4026.

- [7] Li J, Li D, Savarese S, et al. Blip - 2: Bootstrapping language-image pre-training with frozen image encoders and large language models[C]//International Conference on Machine Learning. PMLR, 2023: 19730 - 19742.
- [8] Cao S, Xu B, Zhou W, et al. The blessing of depth anything: an almost unsupervised approach to crop segmentation with depth-informed pseudo labeling[J]. Plant Phenomics, 2025, 7(1): 100005.
- [9] Saeidifar M, Li G, Chai L, et al. Zero-shot image segmentation for monitoring thermal conditions of individual cage-free laying hens[J]. Computers and Electronics in Agriculture, 2024, 226: 109436.
- [10] Zhang L, Gray H, Ye X, et al. Automatic individual pig detection and tracking in pig farms[J]. Sensors, 2019, 19(5): 1188.
- [11] Cowton J, Kyriazakis I, Bacardit J. Automated individual pig localisation, tracking and behaviour metric extraction using deep learning[J]. IEEE Access, 2019, 7: 108049 - 108060.
- [12] Tran D, Thanh N. Pig health abnormality detection based on behavior patterns in activity periods using deep learning[J]. International Journal of Advanced Computer Science and Applications, 2023, 14(5): 603 - 610.
- [13] Zhou H, Chung S, Kakar J K, et al. Pig movement estimation by integrating optical flow with a multi-object tracking model [J]. Sensors, 2023, 23(23): 9499.
- [14] 赵宇亮, 曾繁国, 贾楠, 等. 基于 DeepLabCut 算法的猪只体尺快速测量方法研究[J]. 农业机械学报, 2023, 54(2): 249 - 255, 292.
- Zhao Yuliang, Zeng Fanguo, Jia Nan, et al. Rapid measurements of pig body size based on DeepLabCut algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(2): 249 - 255, 292. (in Chinese)
- [15] Lu J, Wang W, Zhao K, et al. Recognition and segmentation of individual pigs based on Swin Transformer[J]. Animal Genetics, 2022, 53(6): 794 - 802.
- [16] Tu S, Liu H, Li J, et al. Instance segmentation based on mask scoring R - CNN for group-housed pigs [C]//2020 International Conference on Computer Engineering and Application (ICCEA). IEEE, 2020: 458 - 462.
- [17] 王海燕, 江烨皓, 黎煊, 等. 基于弱监督数据集的猪只图像实例分割[J]. 农业机械学报, 2023, 54(10): 255 - 265.
- Wang Haiyan, Jiang Yehao, Li Xuan, et al. Pig image instance segmentation based on weakly supervised dataset[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(10): 255 - 265. (in Chinese)
- [18] Tu S, Cai Y, Liang Y, et al. Tracking and monitoring of individual pig behavior based on YOLO v5 - Byte[J]. Computers and Electronics in Agriculture, 2024, 221: 108997.
- [19] Zhang Y, Sun P, Jiang Y, et al. Bytetrack: multi-object tracking by associating every detection box [C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 1 - 21.
- [20] Hu E J, Shen Y, Wallis P, et al. LoRA: low-rank adaptation of large language models[J]. ICLR, 2022, 1(2): 3.
- [21] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117 - 2125.
- [22] 秦文虎, 杨雨锟, 孙立博. 一种生猪体尺视觉测量的特征点选取及测量方法: 202210538688.3[P]. 2025 - 04 - 29.
- [23] Ji X, Li Q, Guo K, et al. A machine learning-based method for pig weight estimation and the PIGRGB-weight dataset[J]. Agriculture, 2025, 15(8): 814.
- [24] Wischner D, Kemper N, Krieter J. Nest-building behaviour in sows and consequences for pig husbandry[J]. Livestock Science, 2009, 124(1 - 3): 1 - 8.
- [25] Psota E T, Schmidt T, Mote B, et al. Long-term tracking of group-housed livestock using keypoint detection and map estimation for individual animal identification[J]. Sensors, 2020, 20(13): 3670.
- [26] He K, Gkioxari G, Dollár P, et al. Mask R - CNN [C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961 - 2969.
- [27] Bolya D, Zhou C, Xiao F, et al. Yolact: real-time instance segmentation [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9157 - 9166.
- [28] Wang X, Zhang R, Kong T, et al. SOLOV2: dynamic and fast instance segmentation[J]. Advances in Neural Information Processing Systems, 2020, 33: 17721 - 17732.
- [29] Guo R, Niu D, Qu L, et al. Sotr: segmenting objects with transformers [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 7157 - 7166.