doi:10.6041/j.issn.1000-1298.2025.11.054

基于改进 YOLO v8n 的双目视觉自动套杯奶牛乳头空间定位方法

王　娟[1,2]　李梦洁[1]　刘雅举[3]　付鑫培[1]　李斯睿[1]

(1.河北农业大学机电工程学院, 保定 071001; 2.河北省畜禽养殖智能装备与新能源利用重点实验室, 保定 071001;
3.河北农业大学渤海学院, 沧州 061100)

摘要: 奶牛挤奶时人工套杯劳动强度大、生产效率低,而自动套杯成本较高,对中小型养殖场来说存在一定经济负担。为满足中小型养殖场需求,针对挤奶器快速准确自动套杯工作中挤奶环境复杂、奶牛乳头形态各异等问题,本文提出了一种基于改进 YOLO v8n 的双目视觉定位新方法,能够实现对奶牛乳头的准确检测和快速定位。将 YOLO v8n 的 CSPDarknet 主干网络替换为轻量级的 FasterNet 主干网络、在颈部网络构建 P2 特征层、在 Detect 部分与 C2f 部分添加 EMA 注意力机制,提高了模型对奶牛乳头的检测精度与速度。利用 RANSAC 算法对 SURF 特征点进行优化,减少了乳头与乳房相似导致的不匹配问题。基于改进 YOLO v8n 模型,结合双目视觉获取奶牛乳头的三维空间信息。在自制数据集上进行消融试验和对比试验,结果表明,改进 YOLO v8n 模型的 mAP@0.5 为 98.62%,精确率为 97.23%,召回率为 96.69%,与原始 YOLO v8n 模型相比,各项指标分别提高 2.17、3.28、3.65 个百分点;参数量和帧率分别为原始 YOLO v8n 模型的 0.55、2.31 倍,显著提升了模型的检测性能。经统计分析,奶牛乳头定位的平均绝对偏差为 0.011 7 m,方差为 0.000 1 m^2,标准差为 0.011 9 m,能够满足下一阶段挤奶器自动套杯要求。

关键词: 奶牛; 自动套杯; 乳头识别; 乳头空间定位; 改进 YOLO v8n; 双目视觉

中图分类号: S24　　**文献标识码**: A　　**文章编号**: 1000-1298(2025)11-0560-12　　**OSID**:

Binocular Vision-based Spatial Positioning Method for Cow Automatic Teat-cup-attachment Utilizing Improved YOLO v8n

WANG Juan[1,2]　LI Mengjie[1]　LIU Yaju[3]　FU Xinpei[1]　LI Sirui[1]

(1. *College of Mechanical and Electrical Engineering*, *Hebei Agricultural University*, *Baoding* 071001, *China*
2. *Hebei Key Laboratory of Intelligent Equipment and New Energy Utilization of Livestock and Poultry Breeding*, *Baoding* 071001, *China*
3. *Bohai College*, *Hebei Agricultural University*, *Cangzhou* 061100, *China*)

Abstract: Manual teat-cup-attachment is labor-intensive and has low production efficiency, while automatic attachment involves higher costs, posing a certain economic burden for small and medium-sized dairy farms. To meet the needs of these farms and address challenges such as the complex milking environment and varying teat morphologies in rapid and accurate automatic teat-cup-attachment for milking machines, a novel binocular vision positioning method was proposed based on an improved YOLO v8n. This method enabled accurate detection and rapid localization of cow teats. The CSPDarknet backbone network of YOLO v8n was replaced with the lightweight FasterNet backbone network, a P2 feature layer was constructed in the neck network, and an EMA attention mechanism was added to the Detect and C2f sections, enhancing the model's detection accuracy and speed for cow teats. The RANSAC algorithm was employed to optimize SURF feature points, reducing mismatches caused by the similarity between teats and udders. Based on the improved YOLO v8n model, binocular vision was used to acquire three-dimensional spatial information of cow teats. Ablation and comparative experiments conducted on a self-built dataset showed that the improved YOLO v8n model achieved a mAP@0.5 of 98.62%, a precision of 97.23%, and a recall of 96.69%, representing improvements of 2.17, 3.28, and 3.65 percentage points, respectively, compared with that of the original YOLO v8n model. The

收稿日期: 2024-07-24　修回日期: 2024-10-30

基金项目: 河北省重点研发计划项目(19227213D)

作者简介: 王娟(1981—),女,副教授,博士,主要从事图像处理与模式识别、智能化检测技术研究,E-mail: wangjuan@hebau.edu.cn

parameter count was reduced to approximately half, and the frame rate was increased by a factor of 2.31, significantly enhancing the model's detection performance. Statistical analysis revealed that the average absolute deviation of teat positioning was 0.011 7 m, with a variance of 0.000 1 m^2 and a standard deviation of 0.011 9 m, meeting the requirements for the next stage of automatic teat cup attachment in milking machines.

Key words: cow; automatic teat-cup-attachment; teat detection; teat spatial location; improved YOLO v8n; binocular vision

0　引言

挤奶是奶牛养殖过程中劳动强度最大的环节，实现自动化挤奶对降低生产成本具有重要意义[1]。其中，自动套杯技术是提升挤奶效率的关键，但目前中小型养殖场受限于设备成本，仍广泛采用效率低的人工操作方式。因此，开发适用于中小规模牧场的低成本、高精度自动套杯系统具有迫切的实际需求。

实现自动套杯的核心在于对奶牛乳头的快速、准确识别与定位。基于深度学习[2]的目标检测已在农业视觉任务如果实识别[3]、家畜典型行为监测[4]等方面展现出显著优势。在奶牛乳头检测方面，PORTER 等[5]开发了预测奶牛末端分数的深度学习系统，RASTOGI 等[6]实验了 Haar 级联分类器和 YOLO 算法两种乳头检测方法，夏事成等[7]改进了 YOLO v5 模型，席横流等[8]提出了一种改进的 CG-YOLO v5 目标识别算法，李小明等[9]通过立体匹配与三维重建等实现仿真奶牛乳头识别，LU 等[10]采用 R^2Faster R-CNN 算法实现了旋转奶牛乳头的识别。然而现有研究多集中于末端执行器改进[11-13]，针对复杂环境下乳头检测与定位的研究较少，且存在检测精度低、速度慢或模型参数过大等问题。

奶牛乳头检测面临三大挑战：①缺乏高质量奶牛乳头图像的公共数据集，用于自动套杯装置的奶牛乳头识别与定位研究仍处于探索阶段。②奶牛挤奶厅环境复杂，存在光线变化、遮挡和背景干扰，给奶牛乳头检测带来困难。③奶牛乳头为小目标且与乳房颜色相近，易导致误匹配，产生较大的定位误差。YU 等[14]提出的 FS-YOLO v4 模型检测奶牛乳头的精度和检测时间均优于现有模型，但参数过大，难以部署嵌入式终端。

针对挤奶厅环境复杂、奶牛乳头形态各异等问题，提出一种基于 YOLO v8n 的方法，以轻量级的 FasterNet 网络替换 YOLO v8n 的 CSPDarknet 主干网络；在颈部网络构建 P2 特征层，并在 Detect 部分与 C2f 部分添加高效多尺度注意力机制（Efficient multi-scale attention，EMA），以提高模型对奶牛乳头的检测精度与速度；利用随机采样一致算法（Random sample consensus，RANSAC）对加速稳健特征（Speeded up robust features，SURF）进行优化，以减少乳头与乳房相似导致的不匹配问题。以期构建用于自动套杯装置的奶牛乳头识别与定位模型，实现对奶牛乳头准确且快速的自动定位。

1　材料与方法

1.1　数据集构建

目前对于奶牛乳头识别方面的研究主要应用于奶牛乳房炎诊断[15-18]，且奶牛挤奶器自动套杯技术大多针对挤奶器末端执行器进行研究。基于目标检测的奶牛挤奶器自动套杯技术研究相对较少，且缺乏可用的公共数据集。本研究自建奶牛乳头图像数据集，图像采集自河北省保定市徐水区某奶牛养殖场（39°07′18″N，115°32′21″E）。采集对象为泌乳期荷斯坦奶牛，采集设备为智能手机（NTH-AN00 型）和双目相机（HBV-1780-2 S2.0 型），采集时间为 2023 年 10 月 24 日 11:00—14:00。智能手机分辨率为 3 904 像素×2 928 像素，等效焦距为 35 mm，所采集图像经预处理后，调整为 YOLO v8n 模型所需的标准尺寸（640 像素×640 像素）。在工作人员对奶牛乳头进行清洁后进行仰视拍摄，位置距离奶牛乳头 0.3～0.5 m。先后对 400 多头奶牛进行拍摄，共采集 5 315 幅图像，去除模糊、重叠等低质量图像后共筛选出 1 272 幅。为防止训练模型过拟合，对筛选的图像进行添加噪声、改变亮度、裁剪、平移、旋转、镜像等数据增强处理，最终构建了 7 632 幅图像的数据集供后续试验。利用 LabelImg 软件对获取的数据集进行标注，每幅图像标注完成后均会产生一个. xml 格式的标签文件。为提高模型的泛化能力，采用随机分配的方式，对标签文件按照比例 8∶1∶1 划分为训练集（6 106 幅）、测试集（763 幅）和验证集（763 幅）。

1.2　YOLO v8n 模型改进

改进 YOLO v8n 模型结构如图 1 所示，主要分为主干网络、颈部网络和头部网络 3 部分。其中，黄色虚线框框选部分为将 CSPDarknet 主干网络替换为轻量级的 FasterNet 主干网络；蓝色虚线框框选部

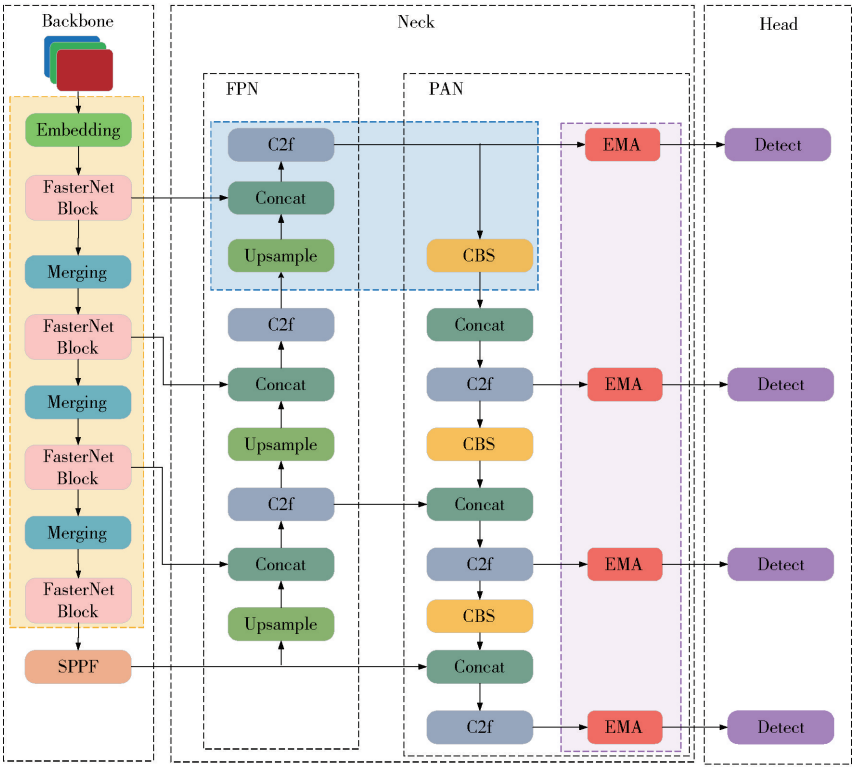


图 1 改进 YOLO v8n 网络结构图

Fig. 1 Improved YOLO v8n network structure diagram

分为在颈部网络构建的 P2 特征层;紫色虚线框框选部分为添加的 EMA 注意力机制模块。

1.2.1 替换 FasterNet 主干网络

YOLO v8n 原始模型采用 CSPDarknet 作为主干网络进行特征提取,该设计通过跨阶段的局部连接有效降低了计算负担,但复杂的网络结构和大量的卷积层限制了其在计算资源受限的边缘设备上的应用。目前轻量级的目标检测任务中,模型通常使用 MobileNet[19]、ShuffleNet[20]和 GhostNet[21]等架构,采用深度卷积(Depthwise convolution,DWConv)或组卷积(Group convolution,GConv)来提取空间特征。然而,这些卷积在减少浮点运算量的过程中,运算符经常遭受增加内存访问的副作用。部分卷积(Partial convolution,PConv)能够减少掩码区域对卷积结果的影响,使模型更专注于有效像素;逐点卷积(Pointwise convolution,PWConv)能够有效整合通道信息,同时减少浮点运算量和参数量。为了提高模型的检测速度并降低内存访问和计算冗余,本文使用由 PConv 和 PWConv 构建的轻量级 FasterNet 替换原始 YOLO v8n 的主干特征提取网络,如图 1 所示(黄色虚线框框选部分)。

FasterNet 为轻量级网络[22],其核心思想是保持高速度与轻量级的同时,提高特征表达能力,其网络结构如图 2 所示,其中,C 表示特征图通道数,H、W 分别表示特征图的高与宽。该模型包含 3 类模块,其中嵌入层(Embedding layer)和融合层(Merging layer)分别由 4×4 卷积和 2×2 卷积组成,起下采样和扩充通道数的作用。此外,每个 FasterNet 块都由

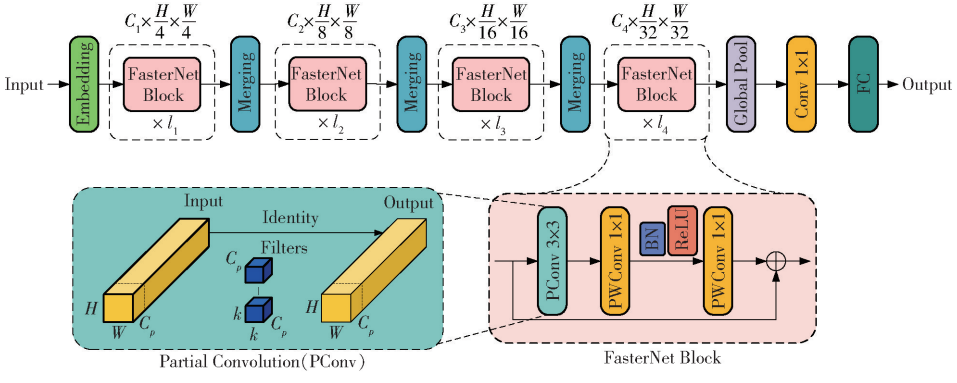


图 2 FasterNet 网络结构图

Fig. 2 FasterNet structure diagram

1 个 PConv 和后续的 2 个 PWConv 构成。网络归一化层采用批量归一化(Batch normalization,BN),以提高推理速度;激活层放置在每个中间 PWConv 之后,以保持特征多样性并减少延迟。最后 3 个层,即全局平均池化、卷积 1×1 和全连接层,用于特征转换和分类。

1.2.2　构建 P2 特征层

小目标在原始输入图像中的像素数量较少,深层特征图感受野较大,会使小目标在特征图上的表示变得稀疏,导致细节特征信息缺失。为了在复杂背景下捕获更多小目标特征信息,提高小目标检测的准确率,本文在 YOLO v8n 颈部网络结构中构建一个 160×160 的小目标检测层 P2,如图 1 所示(蓝色虚线框框选部分)。P2 特征层生成的 160×160 特征图包含大量奶牛乳头纹理和细节信息,分别通过卷积和下采样与主干网络和颈部网络中另外 3 个特征层进行多尺度特征融合,使得深层特征图能够得到更充足的小目标信息,进而提高复杂背景下对小目标的特征捕获能力[23]。另外,增加与 P2 特征层对应的检测头,提高在复杂背景下对小目标的检测能力。

1.2.3　添加 EMA 注意力机制

在奶牛乳头检测任务中,光线变化可能会导致图像亮度分布不均,影响奶牛乳头与奶牛乳房以及周围环境的对比度。奶牛本身具有高度的动态性,导致奶牛乳头的边缘和轮廓变得模糊,降低奶牛乳头检测的准确性。挤奶厅环境错综复杂,这些背景元素会在图像中引入大量冗余信息和干扰,从而导致奶牛乳头难以从复杂背景中分离出来。近年来,注意力机制在计算机视觉领域得到了广泛应用[24-25],其基本原理是通过调整不同图像区域的特征权重,提升模型对重要区域的关注程度。

EMA 注意力机制[26]在不降低卷积运算中通道维数的同时,生成像素级注意力高级特征图。不同于现有的 SENet(Squeeze-and-excitation networks)通道注意力机制将输入特征层通过 2D 全局池化转化为单个特征向量,EMA 注意力机制将位置信息嵌入到通道注意力中,全面关注特征层的空间信息和通道信息。另一种常用的注意力机制 CBAM 采用通道注意力与空间注意力独立运行的方式,只能有效地捕获短程信息,难以建立长距离通道依赖性。EMA 沿用 CA(Coordinate attention)注意力机制[27]的处理方式,将通道注意力分解为两个沿着不同方向聚合特征的 1D 特征编码,用于对具有空间位置信息的跨通道依赖性进行建模。但 CA 注意力机制忽略了整个空间位置之间相互作用的重要性,1×1 卷积的有限感受野阻碍了局部跨通道相互作用的建模和上下文信息的利用。综合考虑,本文选择添加 EMA 注意力机制,以加强模型对奶牛乳头特征的关注度。

EMA 注意力机制提出了一种新的跨空间学习方法,并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,如图 3 所示,其中,G 表示组数,“//”表示将特征图在通道方向分为 G 组,形成 G 个子空间,每个子空间都进行相同的操作。

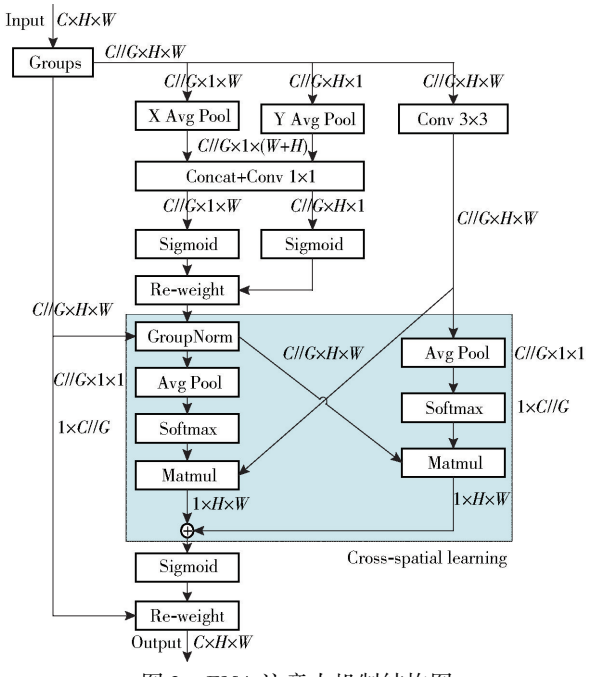


图 3　EMA 注意力机制结构图

Fig. 3　Structure diagram of EMA

首先,利用沿垂直维度方向的 1D 平均池化在空间上捕获长距离的依赖性;其次,在水平维度方向进行 1D 全局平均池化,也可以视为沿垂直维度方向的位置信息集合,因此能够保留水平方向精确的位置信息;同时,增加一个卷积核大小为 3×3 的并行分支,聚合多尺度的空间结构信息,可以互补地应用于输入的特征图,有效地建立长短期的依赖关系,增强对感兴趣物体特征信息的关注度。

头部网络负责生成最终的检测结果,同时颈部网络的 C2f 部分处理的是经过多层卷积提取的特征图,这些特征图包含了丰富的深层语义信息。在头部网络的 Detect 部分与颈部网络的 C2f 部分添加 EMA 注意力机制,如图 1 所示(紫色虚线框框选部分),可以更好地对这些特征融合的深层语义信息进行加权处理,提升对目标特征的捕捉能力,增强模型对全局信息的利用率,有助于提高检测准确率。

1.3　双目视觉定位

1.3.1　测距原理

使用分辨率 2 560 像素×720 像素、焦距 2.1 mm、100°无畸变、基线为 60 mm 的 USB 双目摄

像头从不同角度拍摄目标,然后根据三角形测距原理,通过计算视差得到物体三维空间位置,如图 4 所示。

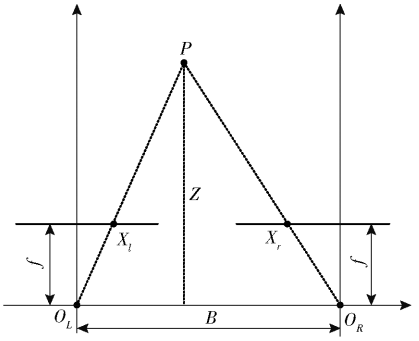


图 4 三角形测距原理图

Fig. 4 Principle diagram of triangle ranging

在光学中心分别为 O_L 和 O_R 的理想双目视觉模型中,任意一点 P 在摄像头左右图像上形成的像点在 X 轴上的坐标分别为 X_l 和 X_r,根据三角形测距原理得

$$\frac{B}{Z}=\frac{B-(X_l-X_r)}{Z-f} \tag{1}$$

整理可得

$$Z=\frac{fB}{d} \tag{2}$$

其中 $$d=x_l-x_r$$

式中 B——双目基线长度,mm

Z——深度值,mm

f——左右摄像头焦距,像素

d——双目视差,像素

1.3.2 立体匹配

想要通过二维图像坐标信息获取物体的三维空间信息,需要通过对相机标定来获取相机的参数信息。本研究使用 Matlab 中的 toolbox_calib 工具箱对双目相机进行参数标定,得到相机的内外参数及旋转平移矩阵,用于图像的立体校正。在完成双目相机的立体校正后,需要对左右相机采集到的图像进行立体匹配。

(1)特征点提取

由于奶牛乳头尺寸较小,特征信息稀疏,本研究选择最精确的点特征进行立体匹配。点特征提取算法主要有 Harris 算法、ORB (Oriented FAST and rotated BRIEF)算法[28]、SIFT(Scale-invariant feature transform)算法[29]和 SURF 算法[30]。Harris 算法是最早的特征点检测算法之一,主要通过计算图像梯度的自相关矩阵特征值来检测图像中的角点,其对图像的尺度变化并不敏感,无法有效检测不同尺度的特征点。ORB 算法通过组合 FAST(Features from accelerated segment test)和 BRIEF(Binary robust independent elementary features)达到了相对较快的计算速度,通过引入方向信息和多尺度金字塔实现了旋转不变性和一定的尺度不变性。与 SIFT 算法和 SURF 算法相比,ORB 算法在特征点描述的细节和准确性上匹配精度稍低。SIFT 算法通过在尺度空间中检测关键点并计算特征向量,可以检测和描述不同尺度下的特征点,在处理图像旋转、缩放、光照变化等方面鲁棒性较好。然而,SIFT 算法计算复杂度较高,处理速度相对较慢。SURF 算法通过修改滤波器尺寸和模糊度得到不同层级影像,在保持 SIFT 算子优良性能特点的基础上解决了计算复杂度高、耗时长等问题,对光照变化和噪声有很好的鲁棒性,提升了算法的执行效率,为算法在实时计算机视觉系统中应用提供了可能。

(2)特征点匹配

特征点匹配是进行精确立体匹配的基础。在完成特征点提取后,需要对左右相机拍摄的图像进行特征点匹配,建立 2 幅立体图像中对应特征点之间的关联,即确定其在不同视角下的相对位置。快速最近邻搜索包(Fast library for approximate nearest neighbors,FLANN)匹配算法使用近似最近邻搜索方法,大幅提高了特征匹配速度,同时还能够保持较高的匹配精度[31]。

2 结果与分析

2.1 试验环境

模型训练环境配置如表 1 所示,参数设置如表 2 所示。

表 1 训练环境配置

Tab. 1 Training environment configuration

训练环境配置	型号
处理器	E5 -2680 v4
显卡	RTX 3090
操作系统	Windows 10
深度学习框架	TensorFlow 2. 7. 0
网络架构	CUDA 11. 0
加速库	cuDNN 8. 0. 5
编译环境	Python 3. 8. 16

表 2 模型参数设置

Tab. 2 Model parameters setting

参数	数值/类型
输入图像尺寸/(像素×像素)	640×640
训练轮次	200
等待轮次	50
训练批次大小	16
权值保存间隔	5
工作线程数	4
优化器	SGD
学习率下降方式	cos

2.2　评价指标

选取精确率(Precision)、召回率(Recall)、平均精度均值(mAP)、参数量、帧率和浮点运算量对模型进行评估。其中平均精度均值采用 mAP@0.5，表示在交并比阈值为 0.5 时计算的平均精度均值。

2.3　结果与分析

2.3.1　试验过程

对 YOLO v8n 原始模型进行训练，训练轮次为 300，其中训练轮次-mAP 曲线和训练轮次-损失值曲线分别如图 5、6 所示。

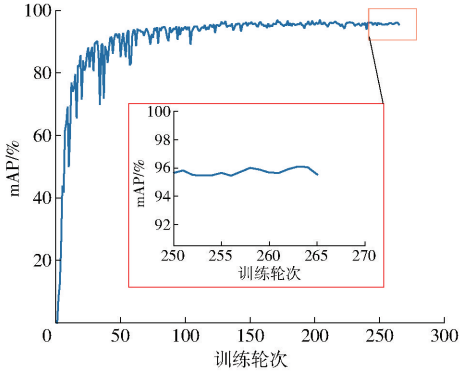


图 5　训练轮次-mAP 曲线

Fig. 5　Epoch - mAP curves

由图 5 和图 6 可知，模型在第 265 次训练时出现早停，等待轮次设置为 50，即若连续 50 次训练内模型性能未出现显著提升则自动终止训练，由此表明 YOLO v8n 原始模型在第 215 次训练之前已出现最优训练结果。由图 5 可以看出模型的 mAP 在 200 次训练之后未出现明显变化，由图 6 可以看出模型的验证损失值和训练损失值在第 200 次训练开始逐渐收敛。为确定模型最优世代，选择 200、210、215 次训练的权值分别对 YOLO v8n 原始模型的验证集进行验证，结果如表 3 所示。

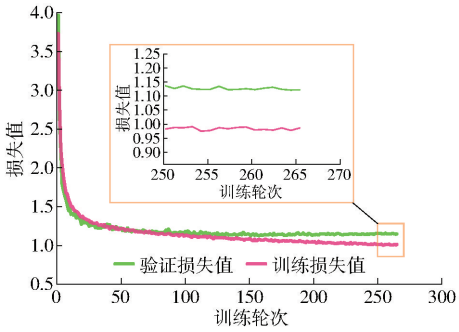


图 6　训练轮次-损失值曲线

Fig. 6　Epoch - loss curves

表 3　不同训练轮次训练结果对比

Tab. 3　Comparison of different epochs training results

%

训练轮次	mAP@0.5	精确率	召回率
200	96.45	93.95	93.04
210	96.22	93.61	91.97
215	95.37	91.52	91.73

由表 3 可知，训练轮次为 200 时的 mAP@0.5、精确率和召回率均高于 210、215，因此本文选择训练轮次为 200。

2.3.2　消融试验

2.3.2.1　数据分析

为验证各改进点的有效性，共进行 5 组消融试验，各项评价指标结果如表 4 所示，其中“√”表示所采用的改进点。

表 4　消融试验结果

Tab. 4　Ablation experiment results

组别	YOLO v8n	FasterNet	P2	EMA	参数量	浮点运算量	帧率/(f·s^{-1})	mAP@0.5/%	精确率/%	召回率/%
N1	√				1.148×10^{7}	8.2×10^{9}	62.07	96.45	93.95	93.04
N2		√			6.660×10^{6}	5.1×10^{9}	171.62	96.58	95.75	93.48
N3		√	√		6.340×10^{6}	9.2×10^{9}	126.57	97.73	96.74	94.73
N4		√		√	6.670×10^{6}	5.2×10^{9}	161.61	97.61	96.64	95.09
N5		√	√	√	6.350×10^{6}	9.3×10^{9}	143.22	98.62	97.23	96.69

图 6 为 N1 组 YOLO v8n 原始模型的试验结果，参数量为 1.148×10^{7}，浮点运算量为 8.2×10^{9}，帧率为 62.07 f/s，mAP@0.5 为 96.45%，精确率为 93.95%，召回率为 93.04%，作为评价指标基准。

N2 组将 YOLO v8n 主干网络替换为 FasterNet 网络，mAP@0.5、精确率和召回率均有小幅增加，参数量和浮点运算量大幅减少，帧率提升至 171.62 f/s，模型在保持轻量化的同时实现了对奶牛乳头的快速检测，为后续移植到移动设备打下基础。

N3 组是在 YOLO v8n 主干网络替换为 FasterNet 网络的基础上，在颈部网络构建 P2 特征层，浮点运算量增加至 9.2×10^{9}，相较于 N2 组模型，帧率出现小幅降低，但参数量有所减少。这是由于 P2 层需要进行额外的前向传播计算，包括卷积操作和特征融合，增加了模型在推理时的计算负担，降低了模型的检测速度。其次，P2 特征层是相对更早期的高分辨率特征图，包含更多的细节信息，其通道数比后续特征图少，意味着在 Detect 部分使用 P2 特

征图时,需要处理的通道数会减少,从而导致总参数量减少。同时,在 Detect 部分中,通常会将多个特征层进行融合操作,引入 P2 特征层后,在特征融合阶段进行了优化,减少了一些不必要的参数,从而使得整体参数量减少。相较于 YOLO v8n 原始模型,mAP@0.5 提高 1.28 个百分点,精确率提高 2.79 个百分点,召回率提高 1.69 个百分点。相较于 N2 组模型,mAP@0.5 提高 1.15 个百分点,精确率提高 0.99 个百分点,召回率提高 1.25 个百分点。由表 4 可知,构建 P2 特征层后,模型的 mAP@0.5、精确率和召回率都显著提升,说明改进模型能够更全面地覆盖目标区域,更好地捕捉到真正存在的目标,降低了奶牛乳头的漏检率,提高了模型对奶牛乳头小目标的检测性能。

N4 组在 YOLO v8n 主干网络替换为 FasterNet 网络基础上添加 EMA 注意力机制,相较于 YOLO v8n 原始模型,mAP@0.5 提高 1.16 个百分点,精确率提高 2.69 个百分点,召回率提高 2.05 个百分点。相较于 N2 组模型增加了小部分参数量和计算量,帧率也小幅下降,但 mAP@0.5 提高 1.03 个百分点,精确率提高 0.89 个百分点,召回率提高 1.61 个百分点。可见,在加入 EMA 注意力机制后,模型更加聚焦于目标区域,加强对乳头区域的特征提取,降低了复杂背景的干扰,提高了奶牛乳头检测的准确性。

N5 组是集合所有改进点的改进 YOLO v8n 模型。由表 4 可知,改进模型各指标均优于 YOLO v8n 原始模型;相较于 N2 组模型,虽然浮点运算量有所增加,帧率出现小幅降低,但是模型参数量大幅减少,同时 mAP@0.5 提高 2.04 个百分点,精确率提高 1.48 个百分点,召回率提高 3.21 个百分点。本文模型整体优于其他模型,更具轻量化且实现了多尺度特征融合,提高了奶牛乳头检测的准确性和泛化性,更具可行性和有效性。

2.3.2.2 可视化分析

构建 P2 特征层前后,模型对验证集中小目标奶牛乳头的漏误检测效果如图 7 所示。其中绿色框线表示正检,蓝色框线表示误检,红色框线表示漏检。由图 7 可以看出,在未构建 P2 特征层时,YOLO v8n 模型对奶牛乳头存在漏检和误检情况;构建 P2 特征层后,YOLO v8n-P2 能够对奶牛乳头进行精准识别。结果表明,在颈部网络构建 P2 特征层可以捕获到小目标奶牛乳头的特征信息,实现了多尺度特征融合,提高了模型对小目标的检测性能。

在相同训练环境和训练参数情况下,为验证 EMA 注意力机制对奶牛乳头检测精度以及关注度

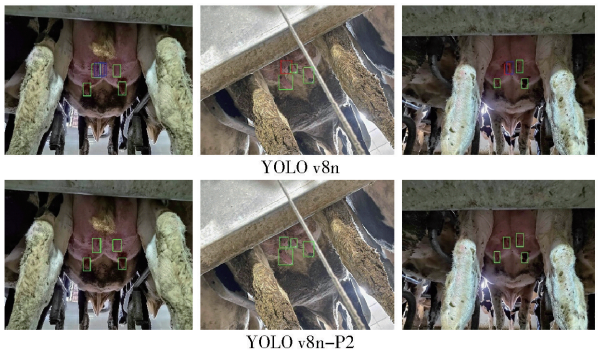


图 7 构建 P2 特征层前后目标漏误检测效果图

Fig. 7 Effect diagram of missed and false detection of targets before and after constructing P2 feature layer

的影响,通过热力图(Gradient-weighted class activation mapping,Grad-CAM)可视化的方式对比了添加注意力机制前后网络模型对奶牛乳头目标区域的敏感度情况。其中,蓝色表示低权重区域,红色表示高权重区域,热力图颜色从蓝色到红色过渡表示模型在特定区域对于预测目标的关注程度逐渐增加。从图 8 可以看出,添加 EMA 注意力机制后,奶牛乳头预测框置信度均有所提升,对奶牛乳头的关注度也均有提高,感知区域更加准确。

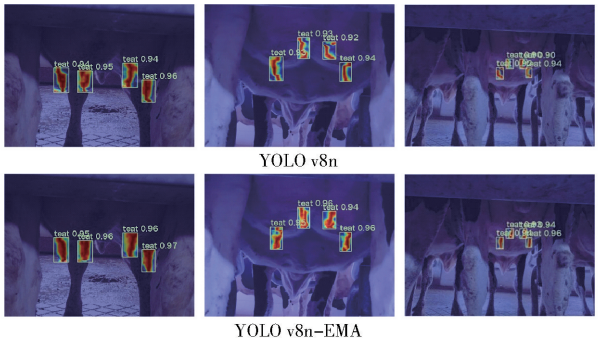
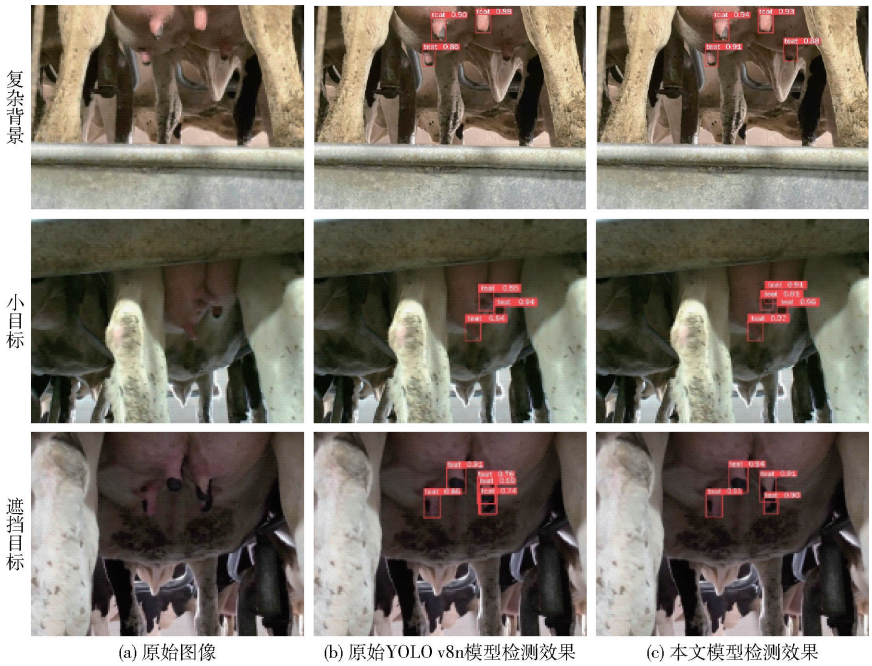


图 8 添加 EMA 注意力机制前后关注度变化热力图

Fig. 8 Visual Grad-CAM of attention change before and after adding EMA attention mechanism

奶牛乳头检测过程中存在背景复杂、目标较小以及目标遮挡等情况,图 9 为采用 YOLO v8n 原始模型和本文模型对以上 3 种情况下奶牛乳头进行检测的结果。

由图 9 可以看出,原始 YOLO v8n 模型对于复杂背景会出现漏检,原因在于原始模型对于目标的关注度较低,缺少足够的表征能力捕获奶牛乳头细节信息。原始模型对奶牛乳头小目标检测也出现误检情况,图 9b 将 2 个交叠的奶牛乳头识别为 1 个,对遮挡的奶牛乳头进行了重复识别,且检测置信度较低。由图 9c 可以看出,本文模型能够更加准确地识别复杂背景下目标,区分与目标有相似特征的背景,未出现误检;小目标检测准确度得到提升,且置信度明显提高;遮挡导致的漏检情况也得到改善,显



(a) 原始图像　(b) 原始YOLO v8n模型检测效果　(c) 本文模型检测效果

图 9　检测效果对比

Fig. 9　Comparison of detection effect

著提升了目标置信度,整体提高了模型对奶牛乳头的检测性能。

2.3.3　对比试验

为更深入地研究本文模型在奶牛乳头检测任务中的性能,在相同的数据集和训练参数下,与 SSD、Faster R－CNN、YOLOv 5s、YOLO v7、YOLO X、YOLO v8n、YOLO v9 以及 YOLO v10 进行了对比试验,结果如表 5 所示。

表 5　不同检测模型对比试验结果

Tab. 5　Comparative experiments results of different detection models

模型	参数量	浮点运算量	帧率/(f·s^{-1})	mAP@0.5/%	精确率/%	召回率/%
SSD	2.629×10^{7}	$6.275\,0\times10^{10}$	37.89	91.03	95.27	54.93
Faster R－CNN	1.371×10^{8}	$3.702\,1\times10^{11}$	9.48	92.82	61.38	97.27
YOLO v5s	4.706×10^{7}	$1.159\,2\times10^{11}$	46.59	95.72	95.16	93.99
YOLO v7	3.762×10^{7}	$1.064\,7\times10^{11}$	43.35	96.34	96.52	93.04
YOLO X	5.421×10^{7}	$1.560\,1\times10^{11}$	38.79	95.69	97.86	91.79
YOLO v8n	1.148×10^{7}	$8.200\,0\times10^{9}$	62.07	96.45	93.95	93.04
YOLO v9	1.190×10^{8}	$1.176\,0\times10^{11}$	47.84	95.64	94.31	90.83
YOLO v10	1.134×10^{7}	$8.400\,0\times10^{9}$	52.85	93.97	95.44	90.54
本文模型	6.350×10^{6}	$9.300\,0\times10^{9}$	143.22	98.62	97.23	96.69

由表 5 可知,相较于 SSD、Faster R－CNN、YOLO v5s、YOLO v7、YOLO X、YOLO v9 模型,本文模型的参数量、浮点运算量指标大幅下降,帧率大幅提升;虽然浮点运算量略高于 YOLO v8n 和 YOLO v10,但参数量和帧率分别为原始 YOLO v8n 模型的 0.55、2.31 倍;另外,mAP@0.5、精确率、召回率均优于 SSD、YOLO v5s、YOLO v7、YOLO v8n、YOLO v9 和 YOLO v10 模型,与 YOLO v8n 模型相比,分别提高 2.17、3.28、3.65 个百分点;精确率略低于 YOLO X,但 mAP@0.5 较 YOLO X 模型高 2.93 个百分点。综上所述,本文模型在小幅增加浮点运算量的情况下,参数量大幅减少,帧率和 mAP@0.5 显著提升,对复杂背景下的奶牛乳头具有更强的检测能力。

2.3.4　双目视觉定位试验

2.3.4.1　特征点提取试验

图 10 为应用 Harris 算法、ORB 算法、SIFT 算法和 SURF 算法分别对左摄像头采集的奶牛乳头图像进行特征提取效果。从图 10 可以看出,SURF 算法提取的特征点个数最多,效果更加细致,因此选择 SURF 算法对奶牛乳头进行特征提取。

2.3.4.2　特征点匹配试验

利用 SURF 算法对左右摄像头采集的图像分别进行特征点提取,如图 11 所示。然后采用 FLANN



图 10 特征点提取效果图

Fig. 10 Effect diagram of feature point extraction



图 11 特征提取后的左右摄像头图像

Fig. 11 Left and right camera images after feature extraction

算法进行特征点匹配,即在左、右摄像头拍摄的图像中找出同一像素点的对应关系,如图 12 所示。由于奶牛腹部特征过于相似,出现了许多匹配错误的情况。



图 12 特征点匹配效果图

Fig. 12 Effect diagram of feature point matching

为了提高匹配精度,引入匹配约束来消除匹配错误点对。RANSAC[32]是一种用于从包含大量“外点”的数据中准确估计数学模型参数的迭代方法。该算法通过随机选择数据子集估计模型参数,然后使用该模型对所有数据点进行测试,将符合模型的点视为“内点”,不符合的点视为“外点”。通过多次迭代,算法最终选择内点数量最多的模型作为最佳估计结果。“外点”一般指数据中的噪声,比如匹配中出现的错误匹配点。因此,本文利用 RANSAC 算法剔除上述匹配中出现的大量错误匹配点对,匹配效果如图 13 所示。

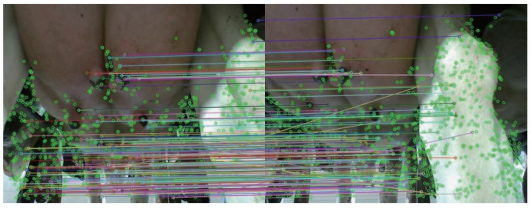


图 13 基于 RANSAC 算法的特征点匹配图

Fig. 13 Feature point matching map based on RANSAC algorithm

2.3.4.3 奶牛乳头定位试验

为了获得奶牛乳头的三维坐标,需要知道左右摄像头图像中奶牛乳头的视差。单应性矩阵 $\boldsymbol{H}$ 反映了双目相机中左右摄像头图像对应点之间的映射关系,其可通过优化后的立体匹配特征点得到。假设一对优化后的匹配特征点,在左摄像头图像中的像素坐标为 $P_l(x_l,y_l)$,在右摄像头图像中的对应点为 $P_r(x_r,y_r)$,单应性矩阵为 $\boldsymbol{H}$,则

$$\begin{bmatrix} x_r \\ y_r \\ 1 \end{bmatrix} = s\boldsymbol{H}\begin{bmatrix} x_l \\ y_l \\ 1 \end{bmatrix} = s\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}\begin{bmatrix} x_l \\ y_l \\ 1 \end{bmatrix} \tag{3}$$

将式(3)展开为

$$\begin{cases} x_r = s(h_{11}x_l + h_{12}y_l + h_{13}) \\ y_r = s(h_{21}x_l + h_{22}y_l + h_{23}) \\ 1 = s(h_{31}x_l + h_{32}y_l + h_{33}) \end{cases} \tag{4}$$

其中 s 为比例因子($s \neq 0$)。将式(4)中的第 3 个方程代入前 2 个方程,则矩阵 $\boldsymbol{H}$ 中有 8 个未知参数,由于每对匹配点可以提供 2 个方程,因此通过 4 对匹配点即可获得单应性矩阵 $\boldsymbol{H}$。

改进 YOLO v8n 模型可以检测到左摄像头目标中心点的像素坐标 $P_l(x_l,y_l)$,将其代入式(4),获取右摄像头对应点的像素坐标 $P_r(x_r,y_r)$,则图像的视差 $d = x_l - x_r$。由式(2)可得目标的深度值,根据相机参数将奶牛乳头的像素坐标转换为三维空间坐标,得到奶牛乳头的三维空间信息。在试验牛场对奶牛乳头进行定位试验,效果图如图 14 所示。

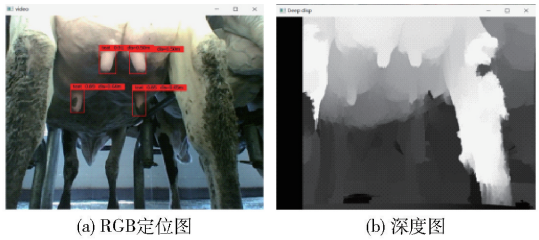


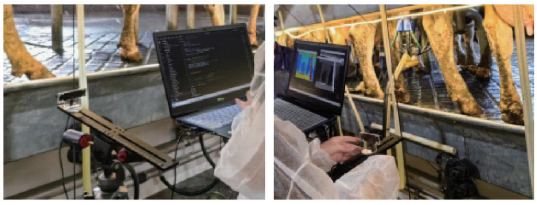
图 14 定位试验效果图

Fig. 14 Effect diagrams of positioning test

评价双目视觉系统可靠性最重要的指标是空间定位下 Z 轴坐标精度,选取左摄像头坐标系作为世界坐标系,验证空间定位下 Z 轴坐标精度。为提高算法的泛化性,随机选取 10 头奶牛,对奶牛乳头进

行空间定位。如图 15 所示，将试验平台的相机导轨和云台调整到 Z 轴移动方向，并保持 X、Y 方向位置不变。试验得到的 Z 轴方向上移动前后奶牛乳头的空间三维坐标及定位误差如表 6 所示。

由表 6 可知，奶牛乳头空间定位的最大误差和最小误差分别为 0.014 2、0.005 7 m。经统计分析，数据的平均绝对偏差为 0.011 7 m，方差为



(a) 移动前　　(b) 移动后

图 15　相机移动前后示意图

Fig. 15　Schematics before and after moving camera

表 6　奶牛乳头空间定位误差分析

Tab. 6　Spatial positioning error analysis of teat　　m

乳头编号	相机移动前乳头坐标	相机移动后乳头坐标	测量差值	实际差值	定位误差
1	(-0.008 5, -0.083 4, 0.459 8)	(-0.014 1, -0.076 9, 0.548 5)	0.088 7	0.100 0	-0.011 3
2	(0.074 5, -0.091 9, 0.468 1)	(0.072 7, -0.094 6, 0.581 2)	0.113 1	0.100 0	0.013 1
3	(0.017 9, -0.026 5, 0.615 3)	(0.018 5, -0.026 8, 0.702 6)	0.087 3	0.100 0	-0.012 7
4	(0.091 4, -0.020 5, 0.638 6)	(0.091 1, -0.023 0, 0.752 5)	0.113 9	0.100 0	0.013 9
5	(0.070 4, 0.157 1, 0.630 4)	(0.074 2, 0.160 5, 0.743 4)	0.113 0	0.100 0	0.013 0
6	(0.082 1, 0.121 2, 0.464 8)	(0.081 6, 0.120 8, 0.552 5)	0.087 7	0.100 0	-0.012 3
7	(0.150 3, 0.161 6, 0.622 6)	(0.153 1, 0.160 7, 0.736 4)	0.113 8	0.100 0	0.013 8
8	(0.168 7, 0.128 3, 0.459 8)	(0.167 7, 0.128 2, 0.572 5)	0.112 7	0.100 0	0.012 7
9	(0.459 2, 0.180 6, 0.854 6)	(0.451 8, 0.173 6, 0.968 7)	0.114 1	0.100 0	0.014 1
10	(0.526 9, 0.184 5, 0.876 9)	(0.518 6, 0.184 7, 0.964 0)	0.087 1	0.100 0	-0.012 9
11	(0.433 8, 0.167 3, 0.962 9)	(0.440 6, 0.165 2, 1.077 1)	0.114 2	0.100 0	0.014 2
12	(0.540 6, 0.165 6, 0.976 9)	(0.539 1, 0.164 9, 1.063 7)	0.086 8	0.100 0	-0.013 2
13	(0.296 1, 0.114 3, 0.555 2)	(0.296 8, 0.109 8, 0.665 7)	0.110 5	0.100 0	0.010 5
14	(0.395 1, 0.114 9, 0.600 6)	(0.395 6, 0.114 7, 0.689 6)	0.089 0	0.100 0	-0.011 0
15	(0.286 5, 0.109 6, 0.707 3)	(0.286 5, 0.109 6, 0.796 9)	0.089 6	0.100 0	-0.010 4
16	(0.409 8, 0.108 1, 0.707 0)	(0.410 2, 0.107 2, 0.797 2)	0.090 2	0.100 0	-0.009 8
17	(-0.235 5, -0.001 7, 0.506 8)	(-0.237 0, -0.005 6, 0.612 5)	0.105 7	0.100 0	0.005 7
18	(-0.194 0, -0.001 7, 0.509 8)	(-0.198 7, -0.004 1, 0.602 5)	0.092 7	0.100 0	-0.007 3
19	(-0.239 8, 0.015 6, 0.659 5)	(-0.239 6, 0.012 8, 0.765 4)	0.105 9	0.100 0	0.005 9
20	(-0.174 2, 0.017 0, 0.649 3)	(-0.173 4, 0.168 7, 0.756 2)	0.106 9	0.100 0	0.006 9
21	(-0.233 7, 0.049 7, 0.505 2)	(-0.232 4, 0.049 8, 0.593 6)	0.088 4	0.100 0	-0.011 6
22	(-0.175 4, 0.053 6, 0.499 8)	(0.072 7, -0.094 6, 0.611 9)	0.112 1	0.100 0	0.012 1
23	(-0.218 2, 0.072 5, 0.642 0)	(0.018 5, -0.026 8, 0.730 1)	0.088 1	0.100 0	-0.011 9
24	(-0.152 6, 0.081 6, 0.638 2)	(0.091 1, -0.023 0, 0.750 7)	0.112 5	0.100 0	0.012 5
25	(0.073 6, 0.043 4, 0.506 8)	(0.073 5, 0.040 3, 0.594 1)	0.087 3	0.100 0	-0.012 7
26	(0.127 7, 0.047 5, 0.512 0)	(0.126 1, 0.048 4, 0.567 1)	0.111 9	0.100 0	0.011 9
27	(0.098 8, 0.050 8, 0.684 4)	(0.108 1, 0.053 3, 0.797 0)	0.112 6	0.100 0	0.012 6
28	(0.128 7, 0.058 3, 0.707 0)	(0.124 6, 0.057 5, 0.793 7)	0.086 7	0.100 0	-0.013 3
29	(0.019 9, 0.021 0, 0.452 2)	(0.021 1, 0.021 9, 0.540 5)	0.088 3	0.100 0	-0.011 7
30	(0.114 6, 0.027 5, 0.456 3)	(0.116 5, 0.024 7, 0.542 7)	0.086 4	0.100 0	-0.013 6
31	(-0.014 1, 0.032 7, 0.567 6)	(-0.014 2, 0.032 1, 0.679 8)	0.112 2	0.100 0	0.012 2
32	(0.119 0, 0.030 3, 0.565 4)	(0.119 9, 0.031 2, 0.677 8)	0.112 4	0.100 0	0.012 4
33	(-0.042 2, 0.070 4, 0.527 9)	(-0.043 9, 0.071 9, 0.638 9)	0.111 0	0.100 0	0.011 0
34	(0.103 1, 0.071 3, 0.535 3)	(0.101 5, 0.076 2, 0.624 5)	0.089 2	0.100 0	-0.010 8
35	(-0.053 2, 0.081 6, 0.659 8)	(-0.054 1, 0.082 3, 0.748 4)	0.088 6	0.100 0	-0.011 4
36	(0.110 0, 0.082 9, 0.660 6)	(0.113 4, 0.084 9, 0.746 9)	0.086 3	0.100 0	-0.013 7
37	(-0.176 4, -0.012 6, 0.452 6)	(-0.168 8, -0.017 4, 0.540 8)	0.088 2	0.100 0	-0.011 8
38	(-0.106 4, -0.018 1, 0.453 7)	(-0.107 3, -0.011 7, 0.566 1)	0.112 4	0.100 0	0.012 4
39	(-0.181 8, -0.030 6, 0.654 0)	(-0.186 3, -0.032 5, 0.742 5)	0.088 5	0.100 0	-0.011 5
40	(-0.139 3, -0.033 6, 0.652 6)	(-0.037 8, -0.038 2, 0.764 6)	0.112 0	0.100 0	0.012 0

0.000 1 m^2,标准差为 0.011 9 m,符合奶牛自动挤奶设备对乳头定位精度的行业标准要求(通常允许定位误差小于等于 0.02 m),表明该方法可用于奶牛乳头定位。

3 结论

(1)针对复杂环境中奶牛乳头识别与定位模型复杂度高、计算量大、耗时长和检测精度低等问题,提出了一种基于改进 YOLO v8n 的双目视觉定位新方法,在自制数据集上验证了改进模型的检测效果。与原始 YOLO v8n 模型相比,本文模型的 mAP@0.5 提高 2.17 个百分点,精确率提高 3.28 个百分点,召回率增加 3.65 个百分点,模型参数量减少 5.13 × 10^6,图像刷新帧率提升至 143.22 f/s。本文模型能够实现对奶牛乳头的快速准确检测,满足对奶牛乳头的检测要求。

(2)利用具有较强鲁棒性和快速性的 SURF 算法,从复杂背景中提取奶牛乳头特征点。利用 FLANN 匹配算法结合 RANSAC 算法对 SURF 特征点进行优化,剔除了部分错误匹配点对,提高了特征点匹配的准确度。

(3)将改进 YOLO v8n 奶牛乳头检测模型与双目视觉结合,实现了对奶牛乳头的精准检测与快速定位,并验证了其空间定位精度。结果表明,定位平均绝对偏差为 0.011 7 m,方差为 0.000 1 m^2,标准差为 0.011 9 m,能够满足下一阶段挤奶器自动套杯要求。

参 考 文 献

[1] COLLISHAW A, JANZEN S, MULLALLY C, et al. A review of livestock development interventions' impacts on household welfare in low- and middle-income countries[J]. Global Food Security, 2023, 38: 100704.

[2] QIU S, CUI X, PING Z, et al. Deep learning techniques in intelligent fault diagnosis and prognosis for industrial systems: a review[J]. Sensors, 2023, 23(3): 1305.

[3] 宋怀波, 尚钰莹, 何东健. 果实目标深度学习识别技术研究进展[J]. 农业机械学报, 2023, 54(1): 1-19.
SONG Huaibo, SHANG Yuying, HE Dongjian. Review on deep learning technology for fruit target recognition[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(1): 1-19. (in Chinese)

[4] 张宏鸣, 孙扬, 赵春平, 等. 反刍家畜典型行为监测与生理状况识别方法研究综述[J]. 农业机械学报, 2023, 54(3): 1-21.
ZHANG Hongmin, SUN Yang, ZHAO Chunping, et al. Review on typical behavior monitoring and physiological condition identification methods for ruminant livestock[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(3): 1-21. (in Chinese)

[5] PORTER I R, WIELAND M, BASRAN P S. Feasibility of the use of deep learning classification of teat-end condition in Holstein cattle[J]. Journal of Dairy Science, 2021, 104(4): 4529-4536.

[6] RASTOGI A, RYUH B S. Teat detection algorithm: YOLO vs. Haar-cascade[J]. Journal of Mechanical Science and Technology, 2019, 33(4): 1869-1874.

[7] 夏事成, 王磊, 王成军, 等. 基于改进 YOLOv5 的奶牛乳头目标检测与数据集构建[J]. 安徽科技学院学报, 2023, 37(5):64-70.
XIA Shicheng, WANG Lei, WANG Chengjun, et al. Nipples object detection and dataset construction based on improved YOLOv5[J]. Journal of Anhui Science and Technology University, 2023, 37(5): 64-70. (in Chinese)

[8] 席横流, 王磊, 王成军, 等. 基于深度学习的奶牛乳头检测方法研究[J]. 内蒙古民族大学学报(自然科学版), 2024, 39(3): 58-66.
XI Hengliu, WANG Lei, WANG Chengjun, et al. Research on detection method of cow's nipple based on deep learning[J]. Journal of Inner Mongolia Minzu University (Natural Sciences Edition),2024, 39(3): 58-66. (in Chinese)

[9] 李小明, 杨开锁, 李军辉, 等. 基于双目立体视觉的挤奶机自动奶杯套杯技术研究[J]. 中国奶牛, 2020(12): 42-45.
LI Xiaoming, YANG Kaisuo, LI Junhui, et al. Research on automatic teatcup attaching technology of milking machine based on binocular stereo vision[J]. China Dairy Cattle, 2020(12): 42-45. (in Chinese)

[10] LU Z, ZHAO M, LUO J, et al. Automatic teat detection for rotary milking system based on deep learning algorithms[J]. Computers and Electronics in Agriculture, 2021,189: 106391.

[11] 常博斌, 毕亮亮. 基于 ANSYS 的挤奶机器人末端执行器可靠性研究[J]. 电子测试, 2021(21): 130-131, 96.
CHANG Bobin, BI Liangliang. End effector of milking robot based on ANSYS reliability research[J]. Electronic Test, 2021(21): 130-131, 96. (in Chinese)

[12] 王成军,李少强. 基于 TRIZ 理论的转盘式挤奶机器人结构设计[J]. 科学技术与工程, 2022, 22(7): 2770-2775.
WANG Chengjun, LI Shaoqiang. Design and analysis of a new rotary milking robot based on TRIZ theory[J]. Science Technology and Engineering, 2022, 22(7): 2770-2775. (in Chinese)

[13] 李硕, 王成军, 李少强. 七自由度挤奶机器人运动特性分析[J]. 兰州文理学院学报(自然科学版), 2023, 37(2): 67-72.
LI Shuo, WNAG Chengjun, LI Shaoqiang. Analysis of motion characteristics of 7-DOF milking robot[J]. Journal of Lanzhou University of Arts and Science (Natural Sciences), 2023, 37(2): 67-72. (in Chinese)

[14] YU Z, LIU Y, YU S, et al. Teat detection of dairy cows based on deep learning neural network FS－YOLOv4 model[J]. Computers and Electronics in Agriculture, 2022, 200: 107224.

[15] 郭艳娇, 杨圣慧, 迟宇, 等. 基于热红外图像的奶牛乳区温度分布与乳房炎识别方法[J]. 农业工程学报, 2022, 38(2): 250－259.
GUO Yanjiao, YANG Shenghui, CHI Yu, et al. Recognizing mastitis using temperature distribution from thermal infrared images in cow udder regions[J]. Transactions of the CSAE, 2022, 38(2): 250－259. (in Chinese)

[16] 王彦超, 康熙, 李孟飞, 等. 基于改进 YOLO v3－tiny 的奶牛乳房炎自动检测方法[J]. 农业机械学报, 2021, 52(增刊): 276－283.
WANG Yanchao, KANG Xi, LI Mengfei, et al. Automatic detection method for dairy cow mastitis based on improved YOLO v3－tiny[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(Supp.): 276－283. (in Chinese)

[17] 郭荣美. 基于热红外图像的奶牛乳房炎自动检测方法研究[D]. 保定: 河北农业大学, 2023.
GUO Rongmei. Study on automatic detection method of dairy cow mastitis based on thermal infrared image[D]. Baoding: Hebei Agricultural University, 2023. (in Chinese)

[18] 宋子琪. 基于热红外图像的奶牛乳房炎检测方法研究[D]. 杨凌: 西北农林科技大学, 2021.
SONG Ziqi. Detection of dairy cow mastitis based on thermal infrared image[D]. Yangling: Northwest A&F University, 2021. (in Chinese)

[19] MITICHE I, MORISON G, NESBITT A, et al. Classification of partial discharge signals by combining adaptive local iterative filtering and entropy features[J]. Sensors, 2018, 18(2): 406.

[20] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition: IEEE, 2018: 6848－6856.

[21] HAN K, WANG Y, TIAN Q, et al. GhostNet: more features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 1577－1586.

[22] CHEN J, KAO S, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): IEEE, 2023: 12021－12031.

[23] LIM J, ASTRID M, YOON H, et al. Small object detection using context and attention[C]//2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). IEEE, 2021: 181－186.

[24] 白强, 高荣华, 赵春江, 等. 基于改进 YOLOV5s 网络的奶牛多尺度行为识别方法[J]. 农业工程学报, 2022, 38(12): 163－172.
BAI Qiang, GAO Ronghua, ZHAO Chunjiang, et al. Multi-scale behavior recognition method for dairy cows based on improved YOLOV5s network[J]. Transactions of the CSAE, 2022, 38(12): 163－172. (in Chinese)

[25] 郭阳阳, 洪文浩, 丁屹, 等. 基于坐标注意力机制和 YOLO v5s 模型的山羊脸部检测方法[J]. 农业机械学报, 2023, 54(7): 313－321.
GUO Yangyang, HONG Wenhao, DING Yi, et al. Goat face detection method by combining coordinate attention mechanism and YOLO v5s model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(7): 313－321. (in Chinese)

[26] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning[C]//ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1－5.

[27] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2021: 13708－13717.

[28] BELMESSAOUD N M, BENTOUTOU Y, EL-MEZOUAR M C. FPGA implementation of feature detection and matching using ORB[J]. Microprocessors and Microsystems, 2022, 94: 104666.

[29] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91－110.

[30] HERBERT B, ANDREAS E, TINNE T, et al. Speeded-up robust features (SURF)[J]. Computer Vision & Image Understanding, 2008, 110(3): 346－359.

[31] 毛清华, 王孟寒, 胡鑫, 等. 改进 SURF－FLANN 的综采工作面视频拼接特征提取与匹配算法[J]. 煤炭学报, 2025, 50(6):3224－3234.
MAO Qinghua, WANG Menghan, HU Xin, et al. Improved SURF－FLANN feature extraction and matching algorithm for video stitching of fully-mechanized working face[J]. Journal of China Coal Society, 2025,50(6): 3224－3234. (in Chinese)

[32] 褚福春, 宫金良, 张彦斐. 基于密度自适应的 RANSAC 非结构化环境下果园机器人导航[J]. 华南农业大学学报, 2022, 43(5): 99－107.
CHU Fuchun, GONG Jinliang, ZHANG Yanfei. Orchard robot navigation in unstructured environment based on density adaptive RANSAC[J]. Journal of South China Agricultural University, 2022, 43(5): 99－107. (in Chinese)