

基于背部特征的挤奶场景下奶牛识别方法

李泽昊^{1,2}, 王月明^{2,3*}, 蒲朝焱^{2,3}

(1. 内蒙古科技大学数智产业学院(网络安全学院), 包头 014010; 2. 草原畜牧业人工智能内蒙古自治区工程研究中心, 包头 014010; 3. 内蒙古科技大学自动化与电气工程学院, 包头 014010)

摘要: 针对奶牛养殖场挤奶场景下奶牛实时无接触身份识别需求, 该研究提出一种基于背部特征的奶牛实时身份识别方法。首先在挤奶场景下采集奶牛背部视频数据, 建立包含 793 张图片的目标检测数据集和包含 3 145 头奶牛、52 834 张图片的身份识别数据集, 构建 YOLOv8-DW 目标检测模型, 检测奶牛背部。然后, 训练对比 HRNet、EfficientNet、ConvNeXt、Swin Transformer 和 Swin Transformer V2 这 5 种骨干网络, 选择最优网络提取奶牛背部特征, 通过计算其与数据库中身份的相似度并与入库验证阈值对比, 得到单帧识别结果。最后, 设计多帧识别方法, 通过对单帧识别结果进行融合得到奶牛身份。YOLOv8-DW 目标检测模型平均精度均值达 99.3%, 处理速度达 169.49 帧/s; 最优网络 HRNet 在测试集中的排序结果平均精度均值为 99.76% 和 120.12 帧/s 的处理速度; 在对不同挤奶日期奶牛的识别中, 多帧识别下已入库奶牛的识别率达 97.7%, 未入库奶牛识别率达 94.0%。该研究结果可为挤奶场景下奶牛的识别提供参考。

关键词: 奶牛; YOLOv8; 身份识别; 背部特征; 挤奶场景

doi: 10.11975/j.issn.1002-6819.202408198

中图分类号: TP391.4

文献标志码: A

文章编号: 1002-6819(2025)-11-0201-08

李泽昊, 王月明, 蒲朝焱. 基于背部特征的挤奶场景下奶牛识别方法[J]. 农业工程学报, 2025, 41(11): 201-208. doi: 10.11975/j.issn.1002-6819.202408198 <http://www.tcsae.org>

LI Zehao, WANG Yueying, PU Chaoyi. Cow recognition method in milking scene based on back features[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2025, 41(11): 201-208. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202408198 <http://www.tcsae.org>

0 引言

奶牛身份的准确识别是实现奶牛健康预警、体况评估、精准饲喂、行为监测的先决条件, 是智能化养殖的必然要求^[1]。转盘式挤奶机具有操作人员少、设备使用效率高、占地面积小、观察牛体况方便等优点, 广泛应用于大型奶牛养殖场^[2]。现代奶牛养殖场为了掌握奶牛产奶水平, 提高牛奶质量, 需要在挤奶场景下对奶牛进行准确快速的身份识别。奶牛实时无接触身份识别技术不仅可以显著提升挤奶效率、优化奶牛管理, 还减少了奶牛应激反应, 提高了奶牛舒适度, 对提升养殖场运营效率和盈利能力有重要作用^[3]。

传统牛只身份识别方法有烙印法^[4]、冷烙法^[5]、耳痕法^[6]等, 这些方法可能对动物的身体造成损伤, 引起牛只行为和内分泌异常^[7], 随着动物识别技术的发展, 这些方法逐渐被淘汰。目前养殖场普遍采用射频识别技术(radio frequency identification, RFID)对牛只进行身份标识^[8], 常见的设备有电子耳标^[9]、智能脚环和智能项圈等^[10]。电子耳标需要打入奶牛耳部, 这会给奶牛带来一

定程度的疼痛, 并且存在引发过敏反应的风险, 损害动物福利; 佩戴智能脚环和智能项圈虽然相对舒适, 但成本较高, 而且绑带易断裂, 导致设备丢失, 给后续数据采集带来困难。相比之下, 基于计算机视觉的身份识别方法具有成本低、无接触、效率高等优点^[11], 在挤奶场景下, 能够实现每头奶牛的准确识别。

基于生物特征的动物身份识别技术具有高准确性和不可复制性, 是动物身份识别领域研究的热点^[12]。LARREGUI 等^[13]提出一种基于 HSV (hue saturation value) 颜色空间特性从复杂场景下的牛眼图像中准确分割出虹膜区域的算法, 平均分割精度达到了 89.57%, 但存在图像模糊、睫毛遮挡和患眼疾动物难以采集图像的缺陷, 难以应用于实际场景。SAYGILI 等^[14]建立了包含 2 430 张牛眼视网膜图像的公开数据集, 并开发了视网膜识别系统 CattNIS, 采用加速稳健特征(speeded up robust features, SURF)提取特征, 实现了 92.25% 的准确率。视网膜血管图像获取难度较大, 且实现应用还面临一定的挑战。KAUR 等^[15]采用三种特征提取算法提取牛唇纹特征, 结合三种分类器(决策树、K 近邻、随机森林)进行牛只的身份识别, 实现 97.23% 的识别准确率。基于传统图像处理和机器学习的身份识别方法依赖于单一生物特征, 往往需要专家手工标注特征, 在不同环境、不同拍摄条件下泛化能力有限, 图像噪声、光照变化以及遮挡等因素也可能影响特征的提取。

随着深度学习的迅速发展, YOLO (you only look once)^[16]、SSD (single shot multibox detector)^[17]、ResNet (residual network)^[18]、EfficientNet^[19]等模型在目标检

收稿日期: 2024-08-27 修订日期: 2025-01-29

基金项目: 内蒙古自治区自然科学基金项目(2024MS06026; 2024MS06022); 内蒙古自治区高校基本科研业务项目(2023QNJS193); 内蒙古自治区科技计划项目(2023YFSW0014)

作者简介: 李泽昊, 研究方向为智能图像处理与动物身份识别。

Email: 769166767@qq.com

*通信作者: 王月明, 博士, 教授, 研究方向为智能信息处理技术。Email: wym20017@imust.edu.cn

测领域展现出优良的性能^[20]。LI 等^[21]对 59 个深度学习模型进行了比较评估,最佳识别准确率为 98.7%,最快处理速度为单张图片 28.3 ms。刘世峰等^[22]采用 Inception ResNet V2 网络模型对牛脸进行识别,在 108 头牛 8 344 张图片的数据集上实现了 98.6% 的识别准确率。基于牛脸,牛鼻纹的识别方法准确率较高,但需要采集牛头部正、侧面图像,而奶牛头部时常晃动,数据采集难度大。为了解决这个问题,XU 等^[23]提出了一种提升牛脸识别准确率的方法,使用 MobileFaceNet 提取特征,通过构建嵌入增强模块(embedding enhancement module, EEM)和嵌入优化模块(embedding optimization module, EOM)改善特征分布并增强对脸部姿势变化的鲁棒性,实现了 98.69% 的识别准确率。通过躯干特征识别奶牛身份是一种适用于复杂场景的识别方法。WANG 等^[24]提出了一种 Open Pose Mask R-CNN 网络,通过提取牛的骨骼关键点和躯干区域特征实现不同视角下识别奶牛身份,达到了 90.2% 的准确率和 89.9% 的平均精度均值。

奶牛体纹图像便于采集,无需限制牛只姿势,得到国内外学者的广泛关注。ANDREW 等^[25]采用基于 softmax 的逆三元组损失函数对奶牛俯视视角的花纹特征进行识别,识别率达到 98.2%。FU 等^[26]改进了 ResNet50 模型,通过采集训练奶牛侧面花纹特征的方式对奶牛进行识别,识别率达 98.58%。WANG 等^[27]提出了一种超轻量级的适用于开源数据集的牛背图像识别方法,通过设计 LightCowsNet 骨干网络,实现了 94.26% 的识别准确率,模型大小仅 4.06 MB,适用于资源有限的边缘设备。XIAO 等^[28]提出了一种基于 YOLOx 和 CowbodyNet 的躺卧奶牛个体识别方法,通过检测牛舍中躺卧奶牛侧面图像,实现了 94.43% 的识别准确率。基于深度学习的奶牛识别方法因其优越的泛化能力、高自动化程度以及强大的特征提取能力,克服了传统方法的不足,成为国内外学者研究的热点^[29]。

本文以挤奶场景下的奶牛为对象,针对智慧养殖场挤奶场景下奶牛实时无接触身份识别需求,提出一种基于背部特征的奶牛实时身份识别方法。为了提取挤奶场景下奶牛背部图片,构建轻量化目标检测模型 YOLOv8-DW,训练并对比 HRNet (high-resolution network)^[30]、EfficientNet、Swin Transformer^[31]、Swin Transformer V2^[32] 和 ConvNeXt^[33] 网络,选择性能最优的网络提取奶牛特征向量,通过计算该特征向量与养殖场数据库奶牛相似度以及对比入库验证阈值得到单帧识别结果。为了提高识别的准确性,设计了一种多帧身份识别方法,对单帧识别结果进行融合得到奶牛身份,实现对挤奶场景下奶牛的实时身份识别。

1 材料与方法

1.1 数据处理

1.1.1 数据采集

本文视频数据于 2023 年 10 月采集于山东某奶牛养殖场。拍摄了多日的挤奶视频,整个牛群共包含 3 145 头奶牛,单次完整挤奶过程约为 5 h,视频总时长 30 h。视频采用俯视视角拍摄,分辨率为 2 560×1 440 像素,帧率为 25 帧/s。

本文采用 YOLOv8 作为数据预处理模型,对奶牛的背部图像进行检测和裁剪。检测结果如图 1 所示。

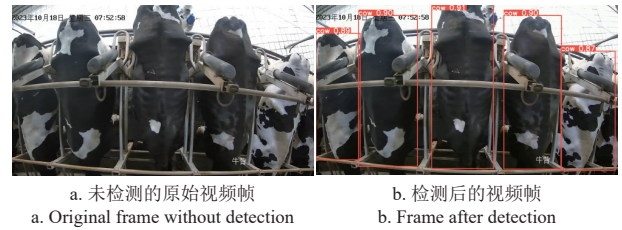


图 1 YOLOv8 检测效果

Fig.1 YOLOv8 detection result

1.1.2 目标检测数据集

为了准确提取每头奶牛的背部图片,从采集到的奶牛视频中选择稳定、清晰的视频片段,涵盖纯黑色、纯白色、纯红色和黑白相间等多种类型的奶牛背部,并且包含空挤奶位情况。

从这些视频片段中提取帧图像,从中筛选出奶牛背部轮廓清晰的 793 张图片,使用标注软件对图片中的奶牛进行标注。将标注完成后的图片按 7:3 划分为训练集和验证集,构建了一个包含 556 张训练图片和 237 张验证图片的奶牛目标检测数据集。

1.1.3 身份识别数据集

为了达到最佳特征提取效果,使计算机更好地识别奶牛身份,首先采用预先训练好的目标检测模型对视频数据进行处理,检测并裁剪奶牛经过测量区域过程中的多张背部图像,接下来将裁剪所得图像按照其所属的奶牛身份分别归档于独立的文件夹。将所有数据处理后,本文建立了包含 3 145 头牛、共计 52 834 张图片的奶牛身份识别数据集,其中黑白花色背部奶牛 2 907 头,纯黑色背部奶牛 218 头,纯白色背部奶牛 15 头,纯红色背部奶牛 5 头,分别占奶牛总数的 92.43%、6.93%、0.48% 和 0.16%,各种类型牛背部如图 2 所示。

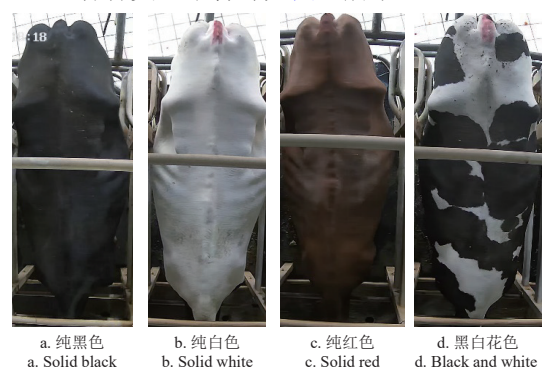


图 2 各类型牛背部示例

Fig.2 Examples of different types of cow backs

奶牛身份识别数据集划分为训练集、验证集和测试集。其中,70% 的奶牛图像数据被分配到训练集和验证集,共包含 2 202 头牛,37 731 张图片,剩余 30% 的奶牛图像数据则作为测试集,共有 943 头牛,15 103 张图片。

每头奶牛随机选取一张图片存放到验证集中,共 2 202 张图片,验证集用于在训练中调整模型超参数;其余 35 529 张图片放入训练集中,用于训练模型。为了评估特征提取网络的性能,将测试集分为 query 集和 gallery 集,测

试集中每头奶牛随机选取一张图片放入 query 集，共 943 张图片，作为待识别奶牛；测试集中剩余图片放入 gallery 集，共 14 160 张图片，gallery 集作为奶牛身份的参考数据库，供 query 集内的图像进行身份匹配。

1.2 识别方案

在转盘式挤奶机作业场景下，奶牛在挤奶过程中的移动轨迹和身体姿态相对稳定，有利于采集俯视视角的视频数据，基于这一特点，本文首先在挤奶位正上方安

装摄像头并设定测量区域，通过目标检测模型对进入摄像头测量区域的奶牛进行检测，提取奶牛背部的清晰图片，将图片输入特征提取网络，提取其特征向量；然后计算该向量与养殖场数据库中特征向量之间的余弦相似度得分，匹配相似度得分最高的身份，对比入库验证阈值后得到单帧识别结果；最后对多个单帧识别结果进行融合得到奶牛身份，实现奶牛实时无接触身份识别。奶牛身份识别整体方案如图 3 所示。

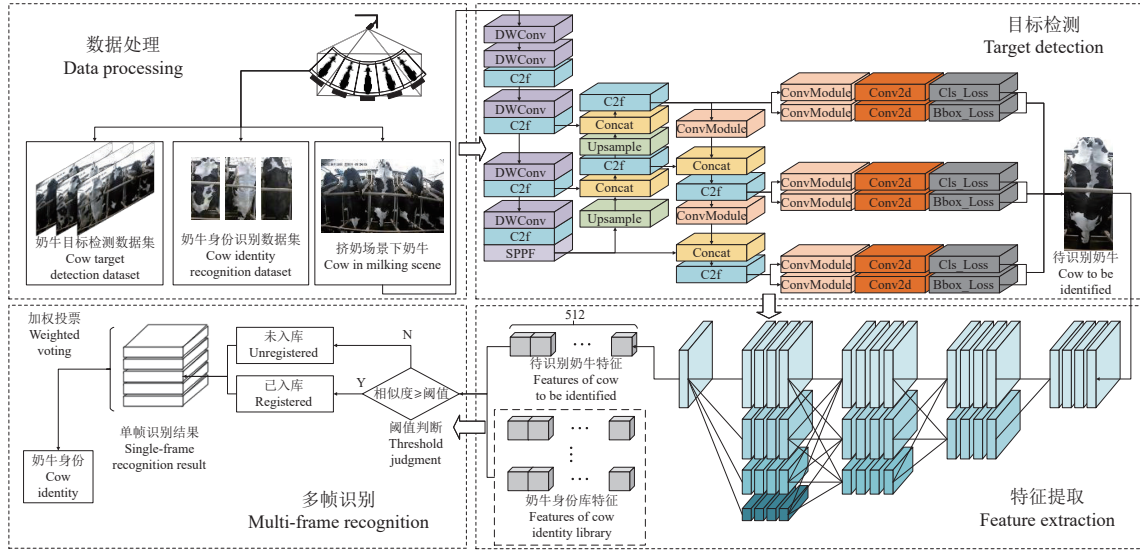


图 3 奶牛身份识别整体方案

Fig.3 Dairy cow identity recognition overall solution

1.2.1 测量区域

挤奶机在运行过程中沿顺时针方向旋转，将摄像头安转在挤奶位正上方，以垂直于地面的视角拍摄奶牛背部，所有奶牛会依次经过摄像头视野中央区域，此时可以观察到清晰完整的奶牛背部，位于此区域的牛背图像角度变化较小，无侧边栏杆或其他牛只躯干遮挡，而且几乎没有发生畸变，有利于牛背图像的获取，将此区域设定为目标检测的测量区域，测量区域的范围为

$$\frac{w}{3} + \frac{d}{2} \leq x \leq \frac{2w}{3} - \frac{d}{2} \quad (1)$$

式中 x 为奶牛检测框中心点水平位置（从画面左边缘起算，向右为正方向），像素； d 为奶牛的水平宽度，由目标检测模型计算，像素； w 为摄像头画面的总像素宽度，像素。

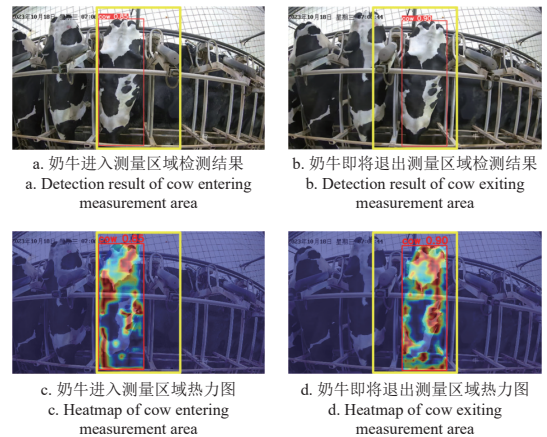
奶牛完全进入测量区域后，目标检测模型对检测到的奶牛提取背部图像，同一时刻仅有一头奶牛位于测量区域内。为了避免画面边缘信息不全和图像畸变带来的不利影响，仅选择位于测量区域内的奶牛背部图像进行检测和提取。测量区域内检测结果如图 4 所示。

1.2.2 目标检测模型

对视频流中的奶牛进行目标检测是奶牛身份识别的重要环节。YOLOv8 是一种实时高效的单阶段目标检测算法，在奶牛目标检测领域拥有优秀的性能^[34]。

针对不同的应用场景，YOLOv8 模型共有 5 种不同大小、宽度的模型，分别是 nano (n)、small (s)、medium (m)、large (l) 和 extra large (x)。由于对挤奶场景视

频流进行目标识别需要准确快速，本文以实时性最好的 YOLOv8n 为基础，引入深度可分离卷积（depthwise separable convolution，DWConv），构建更加轻量化的 YOLOv8-DW 模型。



注：黄色框代表测量区域，红色框代表检测结果。
Note: The yellow box represents the measurement area and the red box represents the detection results.

图 4 测量区域示例

Fig.4 Examples of measurement area

DWConv 由两个部分组成，分别是深度卷积（depthwise convolution）和逐点卷积（pointwise convolution），相较于传统卷积同时处理多通道输入的形式，DWConv 首先在深度卷积部分对每个输入通道独立进行卷积操作，以相同的通道数输出，以减少计算量，接着在逐点卷积

部分使用 1×1 的卷积核进行卷积, 将深度卷积的输出融合, 得到与普通卷积一样通道数和大小的特征图。DWConv 通过对输入通道的独立卷积, 以减小模型大小, 使模型能够更快速地对奶牛背部图像进行检测。

YOLOv8-DW 在 YOLOv8 的基础上分别将第 1、2、4、8 层的标准卷积替换为 DWConv, 以减少算力需求和模型大小, 同时保持良好的目标检测能力。

1.2.3 特征提取

为了选择最适合奶牛特征提取部分的骨干网络, 使用身份识别数据集训练了 5 种在目标检测领域拥有优异性能的网络, 包括 HRNet、EfficientNet、Swin Transformer、Swin Transformer V2 和 ConvNeXt, 并对它们在奶牛身份识别数据集上的性能进行对比。在训练过程中采用不同训练策略以达到最优效果, 基于训练集和测试集性能表现, 选择其中表现最优的网络作为奶牛身份识别特征提取模型的骨干网络。

使用测试集对模型进行评估时, 首先将 gallery 集中的 14 160 张图像依次输入到特征提取模型中, 得到一组已知身份牛图像的特征向量 v_1 , 接着将 query 集中的 943 张图像依次输入到模型中, 得到一组待识别身份牛图像的特征向量 v_2 , 依次计算 v_2 中每个向量与 v_1 中每个向量之间余弦相似度, 共产生 943 组结果, 每组结果按照相似度进行降序排序。

1.2.4 多帧识别方法

在实际应用场景中, 不同日期下奶牛的站姿可能发生变化; 人为标记和污迹也可能导致背部图像改变。为提高奶牛身份识别的准确性和鲁棒性, 本文提出一种多帧识别方法, 通过融合奶牛单帧识别结果来增加结果的可靠性。同时, 为了确保能够对未入库的奶牛进行准确识别, 在识别流程中引入了阈值判断机制。

奶牛正确身份和错误身份的相似度得分之间通常存在明显差异, 因此可以通过设置阈值来对奶牛是否建档入库进行判断。通过统计正确身份相似度最大值和错误身份相似度最大值, 利用核密度估计 (kernel density estimation, KDE) 分析二者的得分分布, 计算密度曲线的交点确定入库验证阈值。如果最高相似度得分小于阈值, 判断待识别奶牛身份不存在于数据库中, 该次单帧识别结果为未入库牛; 如果最高相似度得分大于阈值, 判断待识别奶牛身份存在于数据库中, 该次单帧识别结果为最高相似度得分对应身份。

首先对进入摄像头测量区域的待识别奶牛抽取多帧图像, 每帧图像进行特征匹配和阈值判断得到单帧识别结果, 通过判断奶牛位置是否按检测顺序递增认定识别结果是否属于同一头奶牛。

以单帧识别结果的相似度得分为权重, 对所有单帧识别结果进行加权投票, 选择票数最高的身份作为多帧识别结果; 当单帧识别结果为未入库牛时, 权重设置为入库验证阈值的值。

最后, 对于奶牛背部图像发生变化的情况, 实际应用中

使用最新奶牛背部图片和特征对数据库进行更新,

确保库中特征的时效性。

多帧识别时, 检测帧数越多识别结果越稳定, 但会带来更大的计算开销。实际应用中, 需要根据现场挤奶机转速、设备算力资源合理设置多帧识别的帧数。

1.3 损失函数和评价指标

本文采用多种损失函数来优化模型性能。采用完整交并比损失 (complete intersection over union, CIoU) 作为边界框回归的损失函数 (box loss), 使预测框更接近真实框; 采用分布式焦点损失 (distribution focal loss, DFL) 提高预测的准确性和鲁棒性; 使用二元交叉熵损失 (binary cross entropy loss, BCE Loss) 作为分类损失 (classification loss)。

采用平均精度均值 (mean average precision, mAP)、精确率 (P)、召回率 (R)、帧率 (frame rate)、参数量 (Param)、浮点运算 (FLOPs)、第 1 命中率 (Rank@1) 和前 5 命中率 (Rank@5) 作为评价指标。其中, 目标检测任务中 mAP 表示交并比阈值为 0.5 时所有类别的平均精度均值, 身份识别任务中 mAP 基于识别结果的排序精度计算。 P 衡量模型预测为正例的样本中真正正例的比例。 R 衡量所有真正正例中被正确预测为正例的比例。Rank@1 表示正确身份出现在检索结果排序中第 1 个位置的准确率, Rank@5 表示正确身份出现在检索结果排序中前 5 个位置的准确率。帧率评估图像数据处理速度, 代表每秒处理的图片帧数。

2 结果与分析

本文试验在 Ubuntu 18.04.6 操作系统下进行, CPU 型号为 Intel(R) Core(TM) i9-10900X, GPU 为 NVIDIA GeForce RTX 3 090 $\times 2$, 显存容量为 48 GB, 内存大小为 64 GB。python 版本为 3.8.10, torch 版本为 1.12.0, cuda 版本为 11.3。

首先对 YOLOv8-DW 以及 YOLOv8 各版本目标检测模型进行训练对比, 评估模型的准确性和实时性, 选择综合性能最好的版本; 其次, 比较 5 种骨干网络在奶牛身份识别数据集上的特征提取能力, 选出兼顾精度与效率的特征提取骨干网络; 最后, 通过试验确定入库验证阈值并验证多帧识别方法在奶牛身份识别任务的性能。

2.1 YOLOv8-DW 与 YOLOv8 各版本对比试验与分析

为了验证改进后 YOLOv8-DW 在奶牛目标检测任务中的准确性和实时性, 本文对 YOLOv8-DW 以及 YOLOv8 各版本目标检测模型进行对比试验。在训练过程中, 为了加快训练过程, 防止过拟合, 对批量大小、学习率、权重衰减等超参数进行了调整。同时为了提高模型在颜色变化、几何变化等方面的鲁棒性, 并增强其泛化能力, 应用了色彩增强、几何变换等数据增强策略。试验结果如表 1 所示。

结果显示, YOLOv8n 和 YOLOv8l 在精确度方面表现最佳, 达到了 99.4%; YOLOv8m、YOLOv8x 的召回率达 98.3%。YOLOv8 各版本及 YOLOv8-DW 的 mAP 均不低于 99%, 说明它们在奶牛目标检测任务中都有出

色的表现。YOLOv8-DW 在精确度和召回率表现良好，但并无明显优势。在处理速度方面，改进后的 YOLOv8-DW 达到了 169.49 帧/s。同时，得益于深度可分离卷积对输入通道的独立卷积处理和通过 1×1 卷积核融合特征

图，YOLOv8-DW 的参数量和算力需求进一步降低，参数量为 2.60 M，浮点运算仅为 7.2 G，均为所有模型中最低。综合准确性与实时性，本文选择 YOLOv8-DW 作为奶牛身份识别模型。

表 1 YOLOv8-DW 与 YOLOv8 各版本目标检测结果

Table 1 The target detection results of YOLOv8-DW and different versions of YOLOv8						
模型 Model	精确度 Precision P/%	召回率 Recall R/%	平均精度均值 Mean average precision mAP/%	帧率 Frame rate/(帧·s ⁻¹)	参数量 Parameters/M	浮点运算 Floating point operations FLOPs/G
YOLOv8n	99.4	97.8	99.3	156.25	3.01	8.1
YOLOv8s	98.8	94.5	99.3	74.63	11.13	28.4
YOLOv8m	98.8	98.3	99.4	45.05	25.84	79.3
YOLOv8l	99.4	96.0	99.0	34.84	34.84	165.7
YOLOv8x	98.7	98.3	99.3	20.53	68.12	258.5
YOLOv8-DW	98.9	96.4	99.3	169.49	2.60	7.2

2.2 骨干网络性能对比试验与分析

为了比较 5 种骨干网络的性能表现，使用奶牛身份识别数据集分别训练这 5 种骨干网络，训练迭代 180 次后各网络的损失变化如图 5 所示。

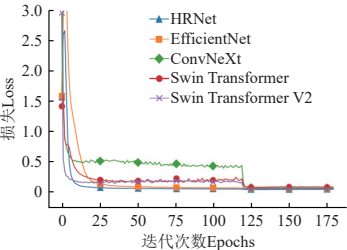


图 5 各模型损失曲线

Fig.5 Loss curves of different models

在训练过程中，为了加快训练收敛并防止过拟合，设置合适的学习率和权重衰减。同时，为了提升模型鲁棒性，采用颜色抖动、对比度增强和随机擦除等数据增强策略。

随着迭代次数的增加，各模型 loss 持续下降，训练结果逐渐接近最优解。为了更好地稳定训练并在后期进一步优化训练效果，在训练进行到总体迭代次数的 2/3 时，将学习率降低到原来的 1/10，以便模型以更小的步伐进行参数更新，从而更快收敛，因此在第 120 个 epoch 时，ConvNeXt、Swin Transformer 和 Swin Transformer V2 的 loss 曲线出现了明显下降。

在经过 180 次迭代后，所有模型的 loss 均已收敛。各模型在测试集上的识别结果如表 2 所示。

表 2 不同骨干网络识别结果

Table 2 Different backbone network identification results				
骨干网络 Backbone network	帧率 Frame rate/ (帧·s ⁻¹)	mAP/%	第 1 命中率 Rank@1 accuracy	前 5 命中率 Rank@5 accuracy
HRNet	120.12	99.76	1	1
EfficientNet(EN)	56.06	99.53	1	1
ConvNeXt(CN)	49.22	98.93	1	1
Swin Transformer(ST)	34.94	99.61	1	1
Swin Transformer V2(STV2)	28.53	99.10	0.998 9	1

在所有网络中，HRNet 的 mAP 达到 99.76%，处理速度为 120.12 帧/s，这主要得益于 HRNet 同时进行高分辨率特征和低分辨率特征提取，能够有效捕捉奶牛背部图像的细节。EfficientNet 通过同时优化网络的宽度、深度和分辨率，表现出不错的精度和处理速度。ConvNeXt

结合了卷积和 Transformer 的优势，在处理速度方面优于 Swin Transformer 和其 V2 版本。Swin Transformer 使用滑动窗口自注意力机制，实现了较高的精度，但窗口机制的设计增加了计算复杂度，限制了其在处理大尺寸图像时的速度，Swin Transformer V2 在此基础上进行了改进，增加更多创新的同时也增加了更多计算开销，导致其在提取牛背特征时处理速度略有降低。

在试验中，出现 HRNet、EfficientNet、ConvNeXt 和 Swin Transformer 的 Rank@1 值达到 1 的现象，这是由于身份识别数据集采用同一天的奶牛视频数据，奶牛的站姿一致，背部图像未发生明显变化，这种情况下，输入图像的差异主要体现在个体特征，而非姿势或背部图像的变化。在表 2 试验结果中平均精度均值低于 100% 的原因是在对奶牛匹配身份进行相似度排序时，部分相似度较高的错误身份排在了相似度较低的正确身份之前，导致整体精度被拉低，无法达到 100%。

由于纯黑色背部奶牛花纹特征不明显，人眼难以分辨。为了评估各网络模型在识别纯黑色背部奶牛的表现，本文从测试集中筛选出 79 头纯黑色背部奶牛作为样本进行不同骨干网络的对比测试，结果如图 6 所示。

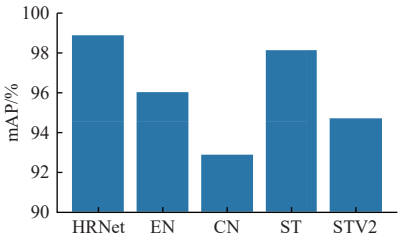


图 6 各骨干网络对纯黑色牛背奶牛识别结果的对比

Fig.6 Comparison of recognition results for solid black back dairy cows across different backbone networks

对于纯黑色牛背的奶牛识别，所有网络的精度都出现了不同程度的下降，但 HRNet 依然保持了良好的识别能力，对不同尺度并行处理的设计使得网络在处理纯黑色牛背图片时，依然能够提取背部的微小特征，mAP 达到了 98.88%。T 在处理纯黑色牛背特征时 mAP 达到了 98.12%，说明其能够有效识别纯黑色牛背特征。EN、CN 和 ST V2 则分别降低了 3.49、6.05 和 4.38 个百分点，下降幅度较大，说明这 3 种网络捕捉纯黑色牛背特征的能力略显不足。试验表明，HRNet 的实时性和准确性最好，适

合作为奶牛身份识别的特征提取骨干网络。

2.3 多帧识别方法试验与分析

为了测试奶牛身份识别方法在实际应用场景下对不同站姿，背部图像变化的奶牛的识别表现，本文使用在对比试验中表现最好的 HRNet 网络，对不同日期下的奶牛进行阈值计算试验和多帧识别试验。

首先，选取 400 头奶牛作为样本，分别统计所有奶牛正确身份和错误身份的余弦相似度最大值。利用核密度估计二者的分布并将结果可视化，如图 7 所示，正确身份的余弦相似度最大值主要集中在 0.8~0.9 之间，而错误身份余弦相似度最大值的峰值出现在 0.4~0.5 之间，两条曲线的交点为 0.620 4。为了最小化错误判断的风险，本文分别以交点及其正负 0.05 为阈值，进行单帧和多帧识别试验，通过试验确定最适合作为入库验证阈值的点。

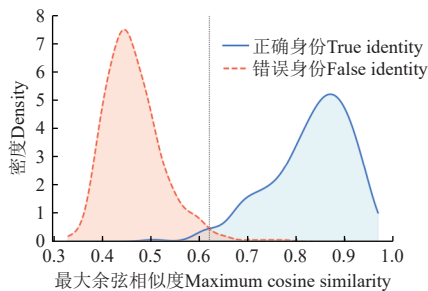


图 7 奶牛身份核密度估计图
Fig.7 Kernel density estimation plot of cow identity

另选 400 头奶牛作为测试多帧识别方法的样本，其中 300 头奶牛身份已存入数据库，其余 100 头奶牛身份未存入数据库；在采集到的数据中，平均每头奶牛通过测量区域的时间约为 5 s，每头奶牛采集 5 帧图像，分别进行单帧识别和多帧识别试验，结果如表 3 所示。

表 3 多帧识别结果

Table 3 Results of Multi-frame Identification

识别方法 Identification methods	阈值 Threshold	已入库牛识别率 Identification rate of registered cows/%	未入库牛识别率 Identification rate of unregistered cows/%
单帧识别	0.570 4	97.3	80.0
	0.620 4	95.3	91.0
	0.670 4	90.7	97.0
	0.570 4	98.7	85.0
多帧识别	0.620 4	97.7	94.0
	0.670 4	94.3	98.0

由试验可知，随着阈值的提高，模型对已入库牛的识别率逐渐下降，对未入库牛的识别率显著上升，这是因为随着阈值的增加，模型对奶牛是否入库的判断标准更加严格，导致更多识别成功但相似度较低的奶牛被错误地判断为未入库；而当阈值较低时，更多奶牛被识别为已入库，进而导致未入库奶牛的识别率下降。为了确保模型在保证对已入库奶牛较高识别率的同时尽可能减少对未入库牛的误判，选择核密度估计正确身份和错误身份曲线的交点作为入库验证阈值。

由表 3 可知，对已入库奶牛单帧识别的识别率达 95.3%，相较于测试集中的 Rank@1 结果有所下降，其原因有两点：在身份匹配过程中，有时正确身份最高相似

度低于某些错误身份的相似度，导致奶牛被错误识别为其他身份；模型在初次识别就匹配到正确身份，但由于相似度低于入库验证阈值，导致该奶牛被判断为未入库。通过对比不同日期下同一奶牛的背部图像发现，造成这两种问题的根本原因在于：不同日期的奶牛挤奶时站姿有差异，影响了匹配效果；不同日期的奶牛背部图案发生变化，如背部可能沾染污迹，或是工作人员喷涂记号导致前后出现差异，进而影响识别准确性。为了减少这些现象对识别结果的影响，通过多帧识别方法增强模型的鲁棒性，提高识别精度，减少误判的风险。不同日期的奶牛背部图像对比如图 8 所示。

由表 3 可知，在相同阈值下，多帧识别方法对已入库奶牛和未入库奶牛的识别率都高于单帧识别方法；在阈值 0.620 4 条件下，该方法对已入库牛的识别率为 97.7%，未入库牛的识别率为 94.0%，相较于单帧识别结果分别提高了 2.4 和 3.0 个百分点，证明多帧识别方法在实际应用场景下更加准确可靠。

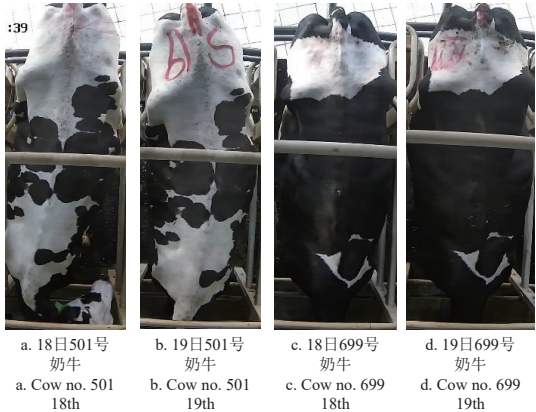


图 8 不同日期奶牛背部图像对比
Fig.8 Comparison of back images of the same cow across different dates

3 结 论

本研究针对智慧养殖场对于挤奶场景下奶牛无接触身份识别的需求，提出一种基于背部特征的奶牛身份识别方法，实现对挤奶场景下奶牛的身份识别，主要工作与研究结论如下：

1) 采集了总时长超过 30 h 的挤奶场景下奶牛视频数据，构建了包含 793 张图片、多种挤奶情况的奶牛目标检测数据集；同时构建了包含 3 145 头奶牛、52 834 张牛背图片的奶牛身份识别数据集，在身份识别数据集中，训练集和验证集共包含 2 202 头牛，37 731 张图片；测试集则包含 943 头牛及 15 103 张图片，这些数据集为挤奶场景下奶牛身份识别方法研究提供了有力支持。

2) 设计了轻量化目标检测模型 YOLOv8-DW，对比了其 YOLOv8 各版本在奶牛目标检测方面的性能。试验表明，YOLOv8-DW 的平均精度均值 mAP 达 99.3%，处理速度为 169.49 帧/s，参数量仅为 2.60 M，适合作为奶牛实时身份识别的目标检测模型。

3) 对比了 HRNet、EfficientNet、Swin Transformer、

Swin Transformer V2 和 ConvNeXt 这 5 种骨干网络在测试集的表现, 并对测试集中 79 头纯黑色背部奶牛进行试验对比。其中 HRNet 在测试集上的处理速度达到了 120.12 帧/s, 平均精度均值 mAP 达到了 99.76%, 在纯黑色背部奶牛的试验中也表现最好, 平均精度均值 mAP 达 98.88%, 证明 HRNet 适合作为奶牛身份识别的特征提取骨干网络。

4) 利用核密度估计计算入库验证阈值, 测试了在不同阈值下, 单帧识别和多帧识别方法对不同日期下奶牛的身份识别表现, 通过试验对比选择正确身份与错误身份曲线交点作为入库验证阈值, 多帧识别方法在 400 头奶牛样本中对已入库牛的识别率达 97.7%, 未入库牛的识别率达 94.0%, 结果表明, 多帧识别方法在实际应用场景下有效提高了识别准确性和鲁棒性。

【参 考 文 献】

- [1] QIAO Y, KONG H, CLARK C, et al. Intelligent perception for cattle monitoring: A review for cattle identification, body condition score evaluation, and weight estimation[J]. *Computers and Electronics in Agriculture*, 2021, 185: 106143.
- [2] REINEMANN D J, VAN DEN BORNE B H P, HOGEVEEN H, et al. Effects of flow-controlled vacuum on milking performance and teat condition in a rotary milking parlor[J]. *Journal of Dairy Science*, 2021, 104(6): 6820-6831.
- [3] 许兴时, 王云飞, 华志新, 等. 融合 YOLOv5s 与通道剪枝算法的奶牛轻量化个体识别方法[J]. *农业工程学报*, 2023, 39(15): 152-162.
XU Xingshi, WANG Yunfei, HUA Zhixin, et al. Light-weight recognition network for dairy cows based on the fusion of YOLOv5s and channel pruning algorithm[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2023, 39(15): 152-162. (in Chinese with English abstract)
- [4] MARTIN M S, KLEINHENZ M D, EDWARDS- ALLAWAY L N, et al. The effect of breed, sex, and oral meloxicam administration on pain biomarkers following hot-iron branding in Hereford and Angus calves[J]. *Journal of Animal Science*, 2022, 100(3): skac038.
- [5] AWAD A I. From classical methods to animal biometrics: A review on cattle identification and tracking[J]. *Computers and Electronics in Agriculture*, 2016, 123: 423-435.
- [6] LOMAX S, WITENDEN E, WINDSOR P, et al. Effect of topical vapocoolant spray on perioperative pain response of unweaned calves to ear tagging and ear notching[J]. *Veterinary Anaesthesia and Analgesia*, 2017, 44(1): 163-172.
- [7] STEAGALL P V, BUSTAMANTE H, JOHNSON C B, et al. Pain management in farm animals: Focus on cattle, sheep and pigs[J]. *Animals*, 2021, 11(6): 1483.
- [8] PEREIRA E, ARAÚJO Í, SILVA L F, et al. RFID technology for animal tracking: A survey[J]. *IEEE Journal of Radio Frequency Identification*, 2023, 7: 609-620.
- [9] SIMANUNGKALIT G, BARWICK J, COWLEY F, et al. Use of an ear-tag accelerometer and a radio-frequency identification (RFID) system for monitoring the licking behaviour in grazing cattle[J]. *Applied Animal Behaviour Science*, 2021, 244: 105491.
- [10] LESO L, BECCIOLINI V, ROSSI G, et al. Validation of a commercial collar-based sensor for monitoring eating and ruminating behaviour of dairy cows[J]. *Animals*, 2021, 11(10): 2852.
- [11] BEZEN R, EDAN Y, HALACHMI I. Computer vision system for measuring individual cow feed intake using RGB-D camera and deep learning algorithms[J]. *Computers and Electronics in Agriculture*, 2020, 172: 105345.
- [12] HITELMAN A, EDAN Y, GODO A, et al. Biometric identification of sheep via a machine-vision system[J]. *Computers and Electronics in Agriculture*, 2022, 194: 106713.
- [13] LARREGUI J I, CAZZATO D, CASTRO S M. An image processing pipeline to segment iris for unconstrained cow identification system[J]. *Open Computer Science*, 2019, 9(1): 145-159.
- [14] SAYGILI A, CIHAN P, ERMUTLU C Ş, et al. CattNIS: Novel identification system of cattle with retinal images based on feature matching method[J]. *Computers and Electronics in Agriculture*, 2024, 221: 108963.
- [15] KAUR A, KUMAR M, JINDAL M K. Cattle identification system: A comparative analysis of SIFT, SURF and ORB feature descriptors[J]. *Multimedia Tools and Applications*, 2023, 82(18): 27391-27413.
- [16] DIWAN T, ANIRUDH G, TEMBHURNE J V. Object detection using YOLO: Challenges, architectural successors, datasets and applications[J]. *Multimedia Tools and Applications*, 2023, 82(6): 9243-9275.
- [17] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//Proceedings of the European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016: 21-37.
- [18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [19] TAN M, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]//Proceedings of the International Conference on Machine Learning. Long Beach, CA, USA: PMLR, 2019: 6105-6114.
- [20] HUO B, LI C, ZHANG J, et al. SAFF-SSD: Self-attention combined feature fusion-based SSD for small object detection in remote sensing[J]. *Remote Sensing*, 2023, 15(12): 3027.
- [21] LI G, ERICKSON G E, XIONG Y. Individual beef cattle identification using muzzle images and deep learning techniques[J]. *Animals*, 2022, 12(11): 1453.
- [22] 刘世锋, 常蕊, 李斌等. 基于脸部 RGB-D 图像的牛只个体识别方法[J]. *农业机械学报*, 2023, 54(S1): 260-266.
LIU Shifeng, CHANG Rui, LI Bin, et al. Individual identification of cattle based on RGB-D images[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(S1): 260-266. (in Chinese with English abstract)
- [23] XU X, DENG H, WANG Y, et al. Boosting cattle face recognition under uncontrolled scenes by embedding enhancement and optimization[J]. *Applied Soft Computing*, 2024, 164: 111951.
- [24] WANG J, ZHANG X, GAO G, et al. Open pose mask R-CNN network for individual cattle recognition[J]. *IEEE Access*, 2023, 11: 113752-113768.
- [25] ANDREW W, GAO J, MULLAN S, et al. Visual identification of individual Holstein-Friesian cattle via deep metric learning[J]. *Computers and Electronics in Agriculture*, 2021, 185: 106133.

- [26] FU L, LI S, KONG S, et al. Lightweight individual cow identification based on Ghost combined with attention mechanism[J]. *Plos One*, 2022, 17(10): e0275435.
- [27] WANG R, GAO R, LI Q, et al. An ultra-lightweight method for individual identification of cow-back pattern images in an open image set[J]. *Expert Systems with Applications*, 2024, 249: 123529.
- [28] XIAO J, SI Y, XIE M, et al. A novel and convenient lying cow identification method based on YOLOx and CowbodyNet: A study with applications in a barn[J]. *Computers and Electronics in Agriculture*, 2024, 225: 109287.
- [29] BAKHSHAYESHI I, ERFANI E, TAGHIKHAH F R, et al. An intelligence cattle re-identification system over transport by siamese neural networks and YOLO[J]. *IEEE Internet of Things Journal*, 2023, 11(2): 2351-2363.
- [30] WANG J, SUN K, CHENG T, et al. Deep high-resolution representation learning for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 43(10): 3349-3364.
- [31] LIU Z, LIN Y, CAO Y, et al. Swin Transformer: Hierarchical vision Transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada: IEEE, 2021: 10012-10022.
- [32] LIU Z, HU H, LIN Y, et al. Swin Transformer V2: Scaling up capacity and resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: IEEE, 2022: 12009-12019.
- [33] LIU Z, MAO H, WU C Y, et al. A convnet for the 2020s[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: IEEE, 2022: 11976-11986.
- [34] WANG Z, HUA Z, WEN Y, et al. E-YOLO: Recognition of estrus cow based on improved YOLOv8n model[J]. *Expert Systems with Applications*, 2024, 238: 122212.

Cow recognition method in milking scene based on back features

LI Zehao^{1,2}, WANG Yueming^{2,3*}, PU Chaoyi^{2,3}

(1. School of Digital and Intelligent Industry(School of Cyber Science and Technology), Inner Mongolia University of Science and Technology, Baotou 014010, China; 2. Inner Mongolia Autonomous Region Engineering Research Center for Artificial Intelligence in Grassland Animal Husbandry, Baotou 014010, China; 3. School of Automation and Electrical Engineering, Inner Mongolia University of Science and Technology, Baotou 014010, China)

Abstract: Accurate recognition of dairy cow identity serves as the cornerstone for intelligent farming systems, particularly in milking scenarios where precise and rapid identification significantly enhances management efficiency while ensuring animal welfare. Traditional recognition techniques, such as branding and ear tagging, have been shown to potentially cause harm to the animals and negatively impact their well-being, and have consequently been gradually phased out with the advancement of technological developments. Although Radio Frequency Identification (RFID) technology remains widely adopted, limitations persist regarding device fragility and animal discomfort. In contrast, contactless recognition technology based on computer vision proved to have advantages such as low cost and high efficiency, making them more suitable for application in milking scene. To improve management efficiency and reduce operational costs, a real-time recognition method based on back features of dairy cows was proposed. This method involved capturing images of cows in milking scene using cameras to identify them, thereby reducing the probability of missed and false detections and avoiding harm to animal welfare. Videos of milking operations were collected, which were then decomposed into images, resulting in an object detection dataset containing 793 annotated images. Lightweight architectural improvements to YOLOv8 resulted in the YOLOv8-DW network. Comparative performance evaluations against multiple YOLOv8 variants confirmed the superiority of YOLOv8-DW in target detection and image extraction. The YOLOv8-DW model achieved remarkable results, with a precision of 98.9%, a recall of 96.4%, and a mean average precision of 99.3%, outperforming the different versions of the YOLOv8 network. The implementation of the YOLOv8-DW model in the NVIDIA GeForce RTX 3090 24 GB GPU environment demonstrated its exceptional efficiency, reaching a detection rate of 169.49 frames per second. This significant improvement in inference speed can be attributed to the lightweight nature of the YOLOv8-DW network, which maintains high accuracy while substantially reducing the computational complexity compared to the original YOLOv8 model. Milking video analysis using YOLOv8-DW facilitated the creation of an identity recognition dataset containing 3 145 unique cow identities across 52 834 images. This dataset was used to train and compare five backbone networks: HRNet, EfficientNet, ConvNeXt, Swin Transformer, and Swin Transformer V2. Five backbone networks underwent comprehensive evaluation through three testing protocols: standard test sets, temporally varied milking videos, and a specialized dataset of 79 solid black back cows. After careful analysis, HRNet network was selected as the most suitable model for the application. The HRNet network demonstrated exceptional performance in extracting the distinctive back features of the target cows and calculating the cosine similarity between the target cows and the cows in the database. On the test set, the HRNet model achieved a mean average precision of 99.76% and Rank@1 accuracy of 1, while maintaining an impressive image processing speed of 120.12 frames per second. Kernel density estimation established recognition thresholds through analysis of similarity score distributions between correct and incorrect identifications. Multi-frame detection combined with weighted voting based on this threshold achieved final recognition rates of 97.7% for registered cows and 94.0% for unregistered cows. This method achieved real-time identity recognition of dairy cows in milking scenarios, significantly improving management efficiency and reducing farming costs, while maximizing the protection of animal welfare. This provides important technical support for the intelligent farming of dairy cows.

Keywords: dairy cow; YOLOv8; identity recognition; back features; milking scene