

基于头颈背部关键点的奶牛跛行检测

张智荣^{1,2}, 褚燕华^{1,2*}, 王月明^{2,3}, 王丽颖^{1,2}, 申煜浩^{1,2}, 李鑫^{1,2}

(1. 内蒙古科技大学数智产业学院(网络安全学院), 包头 0140010; 2. 草原畜牧业人工智能内蒙古自治区工程研究中心, 包头 0140010; 3. 内蒙古科技大学自动化与电气工程学院, 包头 0140010)

摘要:为实现奶牛跛行的自动检测, 该研究鉴于跛行奶牛行走时, 头部起伏较大且背部弓起的特点, 提出了一种基于深度学习的奶牛头、颈、背部6个关键点的跛行检测算法。首先, 在通道旁固定摄像头, 采集奶牛行走的视频数据, 利用YOLOv8n-seg实例分割算法将奶牛从图像中识别出来; 其次, 使用DeepLabCut算法提取视频帧序列中奶牛的头、颈、肩、背部中心、腰和尾部6个关键点坐标, 对比分析实例分割得到的6种格式输入图像在MobileNet-V2和ResNet系列主干网络上的训练效果后, 最终选择按目标检测框裁剪后的分割结果图和ResNet-152作为DeepLabCut算法的最佳输入和最优主干网络; 最后, 对比分析4种时间序列模型和FN-BiLSTM模型在奶牛跛行检测中的表现。试验结果表明, FN-BiLSTM算法的性能最优, 在包含16头奶牛16段视频的测试集上跛行识别的准确率达到97.16%。研究表明, 该算法可为养殖场奶牛跛行检测提供技术支持。

关键词: 奶牛跛行; 深度学习; 实例分割; 关键点检测

doi: 10.11975/j.issn.1002-6819.202406078

中图分类号: TP391.4

文献标志码: A

文章编号: 1002-6819(2024)-21-0157-08

张智荣, 褚燕华, 王月明, 等. 基于头颈背部关键点的奶牛跛行检测[J]. 农业工程学报, 2024, 40(21): 157-164. doi: 10.11975/j.issn.1002-6819.202406078 <http://www.tcsae.org>

ZHANG Zhirong, CHU Yanhua, WANG Yueming, et al. Detecting cow lameness using the key points of head, neck, and back[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(21): 157-164. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202406078 <http://www.tcsae.org>

0 引言

跛行是全球奶牛面临的仅次于乳腺炎和不孕症的三大主要问题之一, 不仅影响奶牛的正常活动, 还会减少产奶量, 对养殖场的经济产生负面影响^[1-2]。随着奶牛养殖规模的不断扩大, 人工检测奶牛跛行的难度也越来越大, 因此, 自动检测奶牛跛行对保障牛只健康、提高生产效率和养殖场的经济效益至关重要。

传统的人工识别奶牛跛行效率低下, 劳动力强度大、成本高, 难以实时监测奶牛的健康状况^[3]。为了提高跛行检测效率, 一些研究人员采用接触式或非接触式传感器来判断奶牛是否跛行。THORUP等^[4]将试验设备放置在奶牛的四肢与脊背上, 通过获取上述奶牛部位的行走姿态进行步态估计, 从而建立识别奶牛是否为跛行的方法。杨丽娟等^[5]使用两个平行力板来测量牛跨过板时的力, 研究表明该系统可以识别跛行的动物并且可以识别受影响的肢体。苏力德等^[6]使用无线三维加速度传感器系统采集奶牛趾蹄加速度信号, 利用改进子序列动态时间规整算法进行跛行识别。这些方法可用于奶牛跛行检测, 但接触式传感器在奶牛日常活动中容易损坏或丢失,

给养殖场带来较大的经济负担; 而非接触式传感器则易受电磁干扰和噪声等环境因素的影响, 导致测量结果不稳定或不准确。

近年来, 随着计算机视觉和深度学习等技术的不断发展, 利用视频分析技术对奶牛进行跛行检测受到越来越多研究学者的关注^[7]。FLOWER等^[8]研究提出, 牛在行走或直立过程中, 背部弓形姿态的变化以及头部上下摆动的动作可以作为检测跛行的指标。宋怀波等^[9]通过分析头部、颈部及背部连接处的拟合直线斜率特性, 经过数据清洗后, 利用KNN分类算法实现了奶牛跛行检测, 正确率达到93.89%。POURSABERI等^[10]通过选取奶牛背部的3个关键点, 拟合圆以计算背部曲率, 并通过设定阈值对跛行进行了分类。康熙等^[11]通过选取肩部、背部和臀部3个关键点, 利用弓背曲率探究奶牛在热红外视频和可见光视频中的行走情况。结果显示, 热红外视频中的跛行检测精度为90.0%, Macro-F1为0.90, 表明热红外视频下的奶牛跛行特征提取与检测更为稳定, 不易受光线和环境变化的影响, 适用于跛行检测。JIANG等^[12]提出了一种单流长期光流卷积网络, 该网络利用光流信息反映奶牛行走时全身的运动信息, 从而提高了奶牛跛行识别的准确性, mAP达到了98.24%。ZHAO等^[13]通过DeepLabCut跟踪侧视步行视频中的身体关键点, 并使用机器学习方法来量化奶牛的蹄和头部的运动, 最终实现了跛行检测的准确率和召回率分别为89.2%和90.7%。上述研究在跛行检测方法和技术手段

收稿日期: 2024-06-12 修订日期: 2024-09-29

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

作者简介: 张智荣, 研究方向为计算机视觉。Email: 1923240971@qq.com

*通信作者: 褚燕华, 副教授, 硕士生导师, 研究方向为计算机视觉, 自然语言处理。Email: chuyanhua@imust.edu.cn

上都取得了一定的进展。

然而,在实际应用中存在2个问题:1)传统的帧差法和光流法依赖于像素变化来检测运动,易受背景变化、光线波动、阴影和栏杆遮挡等环境因素的干扰,难以准确区分奶牛与背景;2)少量关键点(如背部或头部)因遮挡或光线变化存在丢失,影响检测精度。

为了提高养殖场奶牛跛行检测准确性和效率,本文提出了一种基于深度学习的奶牛头、颈、背部6个关键点的跛行检测算法。首先,使用实例分割算法获取奶牛行走的坐标和像素区域,以减少光线变化和遮挡的影响,提高奶牛分割的精度;其次,设计RGB图、二值掩膜图、分割结果图、按目标检测框裁剪后的RGB图、二值掩膜图和分割结果图6种关键点检测数据集,对比分析6种数据集使用不同主干网络的关键点检测效果,选出关键点检测模型的最优输入和最优主干网络;然后,自动提取奶牛头、颈、肩、背部中心、腰和尾部6个关键点;最后,使用改进的FN-BiLSTM(filter noise bidirectional LSTM)模型实现奶牛跛行检测。本研究旨在帮助农场主或兽医及时发现奶牛的跛行情况,从而采取预防或治疗措施,为优化养殖场的跛行检测提供新的手段。

1 材料与方法

1.1 数据采集与数据集构建

1.1.1 数据采集

本次试验视频采集自内蒙古自治区呼和浩特市某奶牛养殖场。视频拍摄时间集中在11月下旬14:00至19:00。视频拍摄地点位于奶牛前往挤奶区的通道内,该通道设有栏杆。拍摄时将两台移动设备(型号为iPhone13和iPhone12)固定安装在通道一侧,垂直于奶牛行进方向,距离栏杆约4 m,距离地面高度约1.3 m。在整理数据集时剔除了存在应激反应加速前进、中途停止、或回望的奶牛行走视频。最终筛选出83头奶牛行走视频共160段,每个视频时长3~18 s。视频的帧率为30帧/s,分辨率为1920像素(水平)×1080像素(垂直)。

1.1.2 数据集构建

本研究构建了3个数据集:实例分割数据集、关键点检测数据集以及奶牛跛行检测数据集。

实例分割数据集是将视频分解为帧图像,并通过结构相似性算法^[14]剔除冗余图像,得到1025幅图像后用LabelMe标注整个奶牛轮廓。将数据集按8:1:1的比例划分为训练集、验证集和测试集。

关键点检测数据集包含6种格式的图像:RGB图、二值掩膜图、分割结果图,以及按目标检测框裁剪后的RGB图、二值掩膜图和分割结果图。具体的生成流程如下:首先,将视频输入实例分割模型,使用训练好的最佳权重进行预测,得到分割结果视频和包含目标奶牛掩膜信息的TXT文件。然后,通过Python脚本生成二值分割视频,并根据每帧中的目标检测框坐标裁剪出原始RGB视频、二值掩膜视频和分割结果视频中的奶牛,从而得到裁剪后的RGB帧图像、二值掩膜帧图像和分割结

果帧图像。最后,去除相似和奶牛不完整的帧,每种格式分别获得1400幅图像,在奶牛的头部、颈部、肩部、背部中心、腰部和尾部进行标注,形成图像与关键点对应的关系。最终,将各数据集的90%用于训练,10%用于测试。6种格式的图像示例如图1所示。

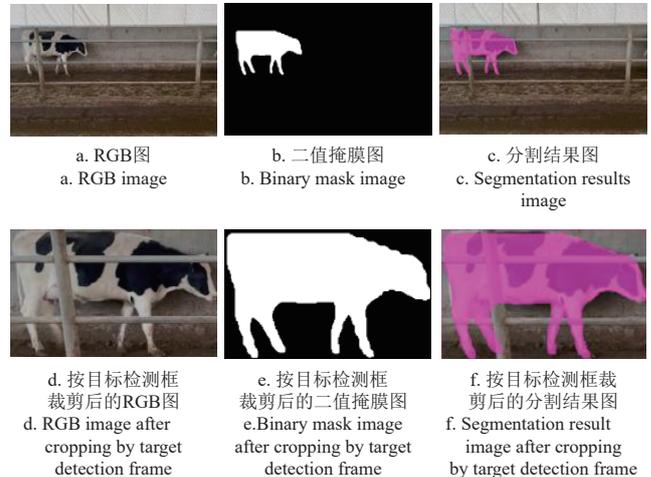


图1 6种格式的图像示例

Fig.1 Examples of the 6 image formats

奶牛跛行检测数据集包括32头牛,共80段视频,总帧数为20493帧。视频中健康奶牛和跛行奶牛的划分均由奶牛养殖场的两位兽医专家评估。其中,健康奶牛的行走视频为40段,共10575帧;跛行奶牛的行走视频为40段,共9918帧。将数据集按8:2的比例进行划分,64段视频用于训练,16段视频用于测试。为保持类别平衡,训练集和测试集中的正常牛和跛行牛视频数量相同。具体信息如表1所示。

表1 奶牛跛行检测数据集

Table 1 Cow lameness detection dataset

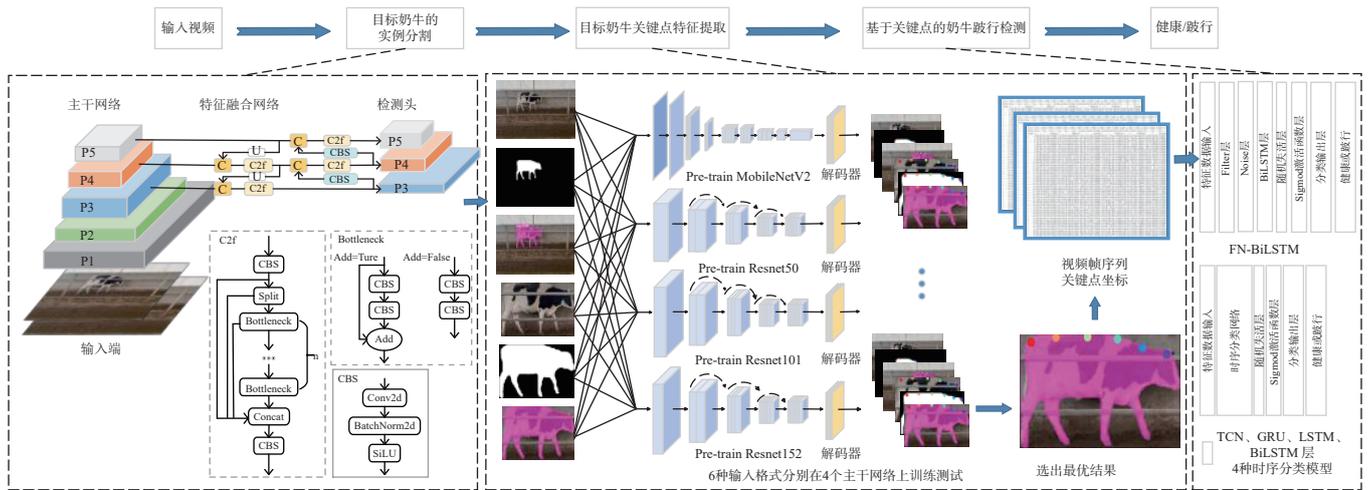
数据集 Dataset	健康奶牛 Heathy cow		跛行奶牛 Lameness cow		总计 Total	
	视频 数量 Video count	帧数 Frames	视频 数量 Video count	帧数 Frames	视频 数量 Video count	帧数 Frames
	训练集 Training set	32	8944	32	8038	64
测试集 Testing set	8	1631	8	1880	16	3511
总计 Total	40	10575	40	9918	80	20493

1.2 奶牛跛行检测方法

跛行奶牛往往通过头部低垂、行走时头部起伏大和背部弓起调整重心,以减轻疼痛和不适,减少受伤或疼痛部位的压力和负荷,相比之下,健康奶牛在行走时头部轻微摆动,背部保持平直,显示出正常的步态和身体平衡^[15-16]。

为了检测奶牛是否跛行,本文在奶牛去往挤奶区的通道旁设置固定摄像头,采集奶牛侧视步行的视频。总体技术路线如图2所示,首先,通过实例分割算法处理采集的视频,获取奶牛的坐标和像素区域;其次,对比6种格式的图像使用不同主干网络检测关键点的效果,选出关键点检测模型的最优输入和最优主干网络,自动提取奶牛的头、颈、肩、背部中心、腰和尾部6个关键

点特征；最后，将基于关键点的时间序列坐标数据送入 时序分类模型中，实现奶牛跛行检测。



注：主干网络 Backbone 中的 P1~P5 表示从浅层到深层提取的多尺度特征图，P1 表示细节丰富的特征，P5 表示全局语义信息更强的特征；C2f 模块由 CBS、特征分层 Split 与 n 个 Bottleneck 模块跨尺度张量拼接而成；CBS 是由二维卷积 Conv2d、二维批量归一化 BatchNorm2d 和 SiLU 激活函数组成的合成模块；Concat 表示张量拼接；U 表示 Upsample 用于上采样；Bottleneck 表示残差模块，包含两种结构：一种由 2 个 CBS 跨尺度张量相加，另一种不相加；检测头 Head 中的 P3~P5 用于多尺度物体的检测和分割。
 Note:P1-P5 in the backbone network denote multi-scale feature maps extracted from shallow to deep layers, P1 denotes detail-rich features and P5 denotes features with stronger global semantic information; the C2f module consists of the CBS, the feature layering Split with n Bottleneck modules spliced across the scale tensor; CBS is a synthesis module consisting of 2D convolution, 2D batch normalisation and SiLU activation functions; Concat denotes tensor splicing; U denotes Upsample for upsampling; Bottleneck denotes the residual module, which contains two structures: One consists of two CBS cross-scale tensors summed together, and the other not added together; P3-P5 in the detection head are used for the detection and segmentation of multi-scale objects.

图 2 总体技术路线图
 Fig.2 Overall technology roadmap

1.2.1 目标奶牛的实例分割

为了准确识别通道中运动奶牛的坐标位置和像素区域，试验选用了兼顾速度与精度的 YOLOv8n-seg 算法^[17]进行实例分割，YOLOv8n-seg 包括输入端、主干网络、特征融合网络和检测头 4 个部分。输入端负责将输入图片缩放到训练需要的尺寸；主干网络采用了 CSPDarknet 结构进行目标特征提取，将特征图分为两部分进行卷积操作和跳跃连接，既保持了网络深度又减少了参数数量和计算量，提高了网络效率。特征融合网络使用特征金字塔 PANet 进行多尺度融合，加强不同维度的特征；检测头部分将主干网络提取的特征进行处理，并输出目标的类别、位置边界框、以及实例分割的掩膜信息。

1.2.2 目标奶牛关键点特征提取

为自动提取奶牛头颈背部特征点位置，本文选取基于深度学习的 DeepLabCut 算法^[18]开展奶牛头、颈、肩、背部中心、腰、尾部 6 个关键点检测的研究。选择在 ImageNet 等大型数据集上经过预训练的 MobileNet-V2^[19]、ResNet-50、ResNet-101 和 ResNet-152^[20] 作为 DeepLabCut 算法的主干网络。ResNet 系列通过残差连接解决了深层网络的训练挑战，能够捕捉复杂特征，而 MobileNet-V2 则以其轻量级设计在资源受限环境中表现优异；此外，采用迁移学习的思想，利用预训练的卷积神经网络作为基础网络，能够加速训练过程并提升模型的准确性^[21-22]。本文在原有的基础网络上，去除 MobileNet-V2 和 ResNet 残差网络结构中的分类层，接入反卷积层进行上采样，在特定任务的微调过程中，网络权重在带有注释身体部位位置的帧组成的数据集上进行端到端的训练和调整，

网络通过学习奶牛各部位的特征，为特定部位的可能位置分配较高的概率，从而更准确地检测这些部位，对不相关区域分配较低的概率，以减少误检，使得网络能够更精准地识别和定位奶牛的关键部位，提升关键点检测的精度^[23-24]。

奶牛行走时头、颈、背等 6 个关键点的位置与其运动规律息息相关。精确定位这些关键点是检测奶牛跛行的关键。为了得到最佳的关键点提取模型，对比分析了图 1 中 6 种格式的图像在不同主干网络上提取关键点的效果，从而选择效果最佳的输入格式和网络模型作为 DeepLabCut 的输入和主干网络。

1.2.3 基于关键点的奶牛跛行检测

利用关键点提取网络获取到的最优权重，对奶牛跛行数据集中的所有视频进行处理，获取视频帧序列关键点坐标数据集。连续的视频帧序列关键点坐标数据可以有效地反应奶牛在行走时头颈背部的运动规律。

虽然每个视频都包含了奶牛的完整行走轨迹，但由于奶牛行走速度不同，每个视频的帧数不同，直接提取各视频帧序列关键点坐标作为训练样本输入，将面临序列长度不一致的问题。为了解决这个问题，将对关键点坐标序列进行采样处理。每段视频每隔 30 帧提取一组坐标，形成采样序列，再将连续 6 帧采样值组合为一个样本输入，当连续帧不足 6 帧时采用零值补齐。通过间隔采样和零值补齐，统一不同视频的关键点坐标序列长度，减小计算复杂度，并将组合后的样本作为批训练数据输入神经网络模型。

奶牛跛行行为与时间密切相关，时序卷积网络

(temporal convolutional network, TCN)^[25]、门控循环单元 (gated recurrent unit, GRU)^[26]、长短期记忆网络 (long short-term memory, LSTM)^[27]、双向长短期记忆网络 (bidirectional LSTM, BiLSTM)^[28]、FN-BiLSTM 是 5 种时间序列模型。TCN 利用一维卷积层来捕捉时间序列中的模式和特征。GRU 是一种递归神经网络, 通过引入门控机制来控制信息的流动。LSTM 是另一种改进的递归神经网络, 用于捕捉长时间依赖关系, 克服梯度消失和爆炸问题。BiLSTM 是 LSTM 的双向扩展版本, 能够同时考虑前后文信息, 通过两层相反方向的 LSTM 组成。FN-BiLSTM 在 BiLSTM 模型基础上进行了改进, 在输入层后增加了 Filter 层和 Noise 层, 使模型更加专注于关键数据特征, 避免过拟合并提高泛化能力, 然后使用 Bidirectional 函数将 LSTM 层包装在一起, 最后是输出层, 使用 Sigmoid 激活函数对行为识别结果进行分类并输出。

为了验证 FN-BiLSTM 模型中各改进点的有效性, 对其进行消融试验; 通过对比 GRU、TCN、LSTM、BiLSTM、FN-BiLSTM 5 种模型对奶牛跛行检测的效果, 选取最佳的模型进行跛行检测。

1.3 评价指标

本研究采用精确率、召回率、平均精度均值 (mean average precision, mAP) 作为评价奶牛实例分割的指标。

使用均方根误差 (root mean square error, RMSE) 作为奶牛头颈背部特征区域关键点检测的评价指标。用精确率、召回率、准确率作为奶牛跛行检测的评价指标。

2 结果与分析

为了实现养殖场通道下基于头颈背部关键点的奶牛跛行检测, 依次对目标奶牛实例分割、关键点特征提取和跛行检测 3 个阶段进行试验分析。本文训练模型所用的试验环境的操作系统为 Ubuntu 18.04.6 LTS、CPU 为 Intel(R)Core(TM)i9-9900K、GPU 为 NVIDIA GeForce GTX 3090Ti、显存为 24 GB、内存为 64 GB、算法框架为 PyTorch 1.13.1、TensorFlow 2.10.0。

2.1 目标奶牛的实例分割

试验使用 1.1.2 节中的实例分割数据集, 在 YOLOv8n-seg 网络上迭代 300 次。在整个训练过程中输入图片分辨率为 640 像素×640 像素, 批量大小为 16, 线程数为 4, 初始学习率为 0.01, 权重衰减系数为 0.0005。YOLOv8n-seg 模型对奶牛实例分割的精确率、召回率、mAP 分别为 99.97%、100%、99.50%。实例分割检测结果如图 3 所示, 可以看出目标奶牛的分割效果较好, 基本不受栏杆遮挡、光线变化、以及距离远近的影响, 从检测的效果上看, 用 YOLOv8n-seg 对奶牛个体进行实例分割是可行的。

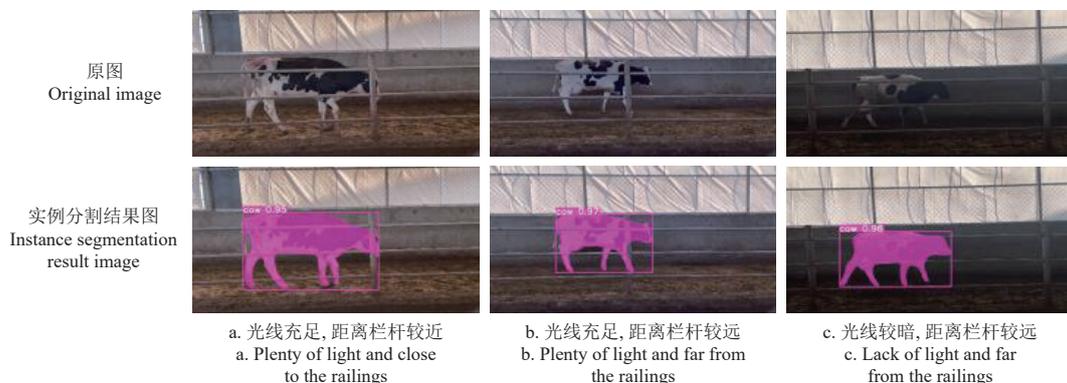


图 3 实例分割检测结果

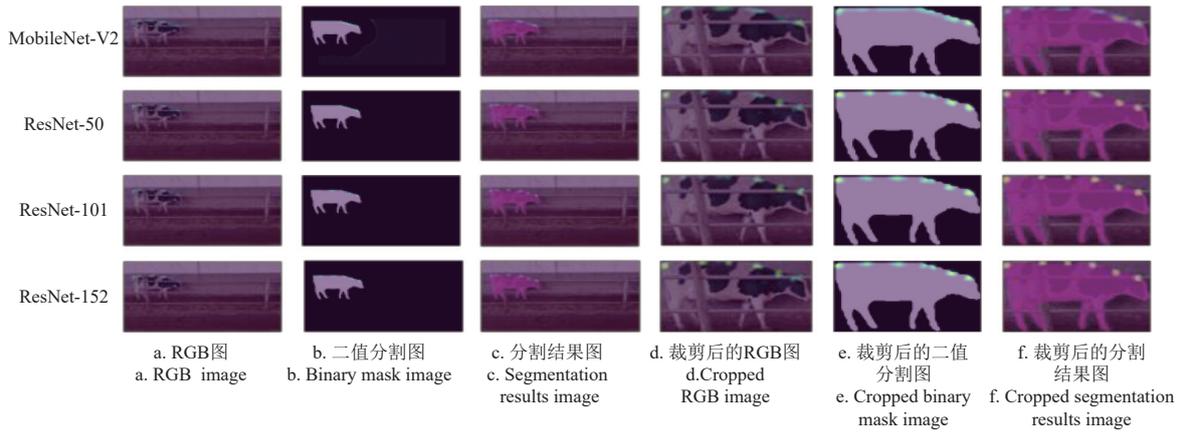
Fig.3 Instance segmentation detection results

2.2 目标奶牛关键点特征提取

准确识别关键点是确保奶牛跛行检测高准确率的基础。为了确定 DeepLabCut 算法的最佳输入和主干网络, 将 1.1.2 节中的 6 种格式的关键点检测数据集分别在 4 个主干网络 (MobileNet-V2、ResNet-50、ResNet-101、ResNet-152) 上训练迭代 200 000 次, 使用随机梯度下降作为模型优化器, 采用可变的学习率, 设置初始学习率为 0.05, 当训练迭代超过 10 000 次后, 学习率调整为 0.02。在 RGB 图像、二值掩膜图、分割结果图、按目标检测框裁剪后的 RGB 图像、二值掩膜图、分割结果图的测试集中各选取一张图像, 通过特征点的分布情况展示 4 个模型的检测效果。不同主干网络模型使用不同输入格式的特征点分布图如图 4 所示。

由图 4 可知, ResNet-152 作为主干网络时, 6 种输

入图像中的关键点特征较为清晰, ResNet-50 和 ResNet-101 表现相似, 而 MobileNet-V2 无法准确检测出 6 个关键点的具体位置。同时, 裁剪后的 RGB 图像 (图 4d)、二值掩膜图 (图 4e)、分割结果图 (图 4f) 中的特征点相较于 RGB 图像 (图 4a)、二值掩膜图 (图 4b)、分割结果图 (图 4c) 更清晰, 说明裁剪后的图像减少了背景干扰, 增强了模型对目标的关注, 从而提高关键点识别的准确性; 此外, 从图 4 中各特征点分布情况可以看出, 裁剪后的分割结果图 (图 4f) 相较于裁剪后的 RGB 图 (图 4d) 和裁剪后的二值掩膜图 (图 4e), 更能清晰地定位各关键点, 尤其是肩部、背部中心和腰部的关键点, 原因在于, 裁剪后的分割结果图 (图 4f) 通过保留目标奶牛的纹理信息并抑制不相关区域的干扰, 使得关键点检测更专注于目标区域, 从而更好地定位关键点。



注：图 d~f 分别为按目标检测框裁剪后的 RGB 图、按目标检测框裁剪后的二值掩膜图、按目标检测框裁剪后的分割结果图。
 Note: Fig.d-f are RGB image after cropping by target detection frame, Binary mask image after cropping by target detection frame, and Segmentation result image after cropping by target detection frame, respectively.

图 4 不同主干网络模型使用不同输入格式的特征点分布结果图

Fig.4 Feature point distribution results of different backbone network models using six input formats

为了更准确地分析结果，表 2 统计了不同主干网络模型使用不同输入格式的训练集均方根误差和测试集均方根误差。

表 2 不同主干网络使用不同输入格式的性能对比

Table 2 The performance comparison of different backbone networks using various input formats

主干网络 Backbone	图片类型 Image type	均方根误差 Root mean square error/像素	
		训练集 Training set	测试集 Testing set
MobileNet-V2	RGB 图	12.31	12.97
	二值掩膜图	19.01	20.80
	分割结果图	11.98	15.35
	裁剪后的 RGB 图	6.66	11.98
	裁剪后的二值掩膜图	9.26	16.03
	裁剪后的分割结果图	7.21	14.78
ResNet-50	RGB 图	6.72	9.34
	二值掩膜图	10.29	12.66
	分割结果图	5.71	9.37
	裁剪后的 RGB 图	4.98	8.52
	裁剪后的二值掩膜图	8.52	11.81
	裁剪后的分割结果图	2.28	6.17
ResNet-101	RGB 图	6.34	10.76
	二值掩膜图	8.27	9.35
	分割结果图	5.60	8.49
	裁剪后的 RGB 图	4.60	7.37
	裁剪后的二值掩膜图	8.01	11.05
	裁剪后的分割结果图	2.23	4.32
ResNet-152	RGB 图	6.25	10.75
	二值掩膜图	7.68	7.01
	分割结果图	5.52	9.00
	裁剪后的 RGB 图	3.59	6.03
	裁剪后的二值掩膜图	7.44	10.46
	裁剪后的分割结果图	2.04	4.28

由表 2 数据可知，使用 ResNet-152 为主干网络训练裁剪后的分割结果图时，训练集和测试集的均方根误差最小，分别为 2.04 像素，4.28 像素。观察表 2 中数据可知，ResNet 系列整体优于 MobileNet-V2，原因在于 ResNet 结构比 MobileNet 更复杂，具有更强的特征提取能力。随着 ResNet 卷积层数升高，训练集均方根误差逐渐减小，说明增加网络的深度，能够提取更多层次的特征，促进不同空间和语义尺度的学习，从而提升关键点检测效果。

2.3 基于关键点的奶牛跛行检测

为了评估引入 Filter 层与 Noise 层对模型性能的影响，对 FN-BiLSTM 模型进行了消融试验；其次，为了选择最优的跛行检测模型，设计了对比试验，比较了 TCN、GRU、LSTM、BiLSTM 和 FN-BiLSTM 5 种时间序列模型。所有模型均使用 1.1.2 节中的奶牛跛行检测数据集，训练迭代次数均为 300 次。

2.3.1 消融试验

改进的 FN-BiLSTM 模型是在 BiLSTM 基础上增加了 Filter 层和 Noise 层。为了探究这两个新增层对奶牛跛行检测性能的影响，利用控制变量法设计消融试验，在 BiLSTM 模型中逐步加入 Filter 层和 Noise 层，构建不同的模型，消融试验对比结果如表 3 所示。

表 3 消融试验对比结果

Table 3 Comparative results of ablation experiments			
模型 Model	滤波 Filter	噪音 Noise	准确率 Accuracy/%
BiLSTM	-	-	92.33
	√	-	94.66
	-	√	95.83
	√	√	97.16

注：“√”表示添加该模块，“-”表示不添加该模块。
 Note: "√" means add the module, "-" means do not add the module.

由表 3 可知，Filter-BiLSTM 比 BiLSTM 的准确率高 2.33 个百分点，滤波处理主要是对奶牛头、颈、背部关键点轨迹数据进行平滑处理，突出重要特征并抑制不重要特征，帮助模型更好地捕捉各关键点间的趋势，从而提高预测准确性。Noise-BiLSTM 比 BiLSTM 高 3.50 个百分点，因为添加噪声增加了数据的多样性和泛化能力，降低了过拟合风险。FN-BiLSTM 比 Filter-BiLSTM 高 2.50 个百分点，原因在于滤波后的数据更具鲁棒性，即使加入噪声后仍能保持稳定，使模型更容易学习数据的真实模式。

2.3.2 对比试验

为了验证本文提出的 FN-BiLSTM 跛行检测模型性能，分别在相同数据集和试验环境下对 GRU、TCN、

LSTM、BiLSTM 模型进行训练和测试。各模型在训练过程中的准确率和损失值变化曲线如图 5 所示。

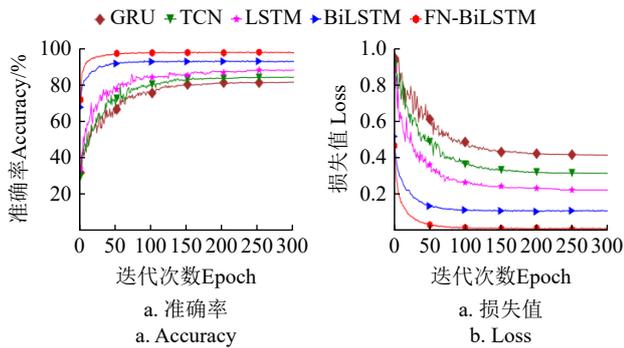


图 5 不同模型准确率与损失变化

Fig.5 Variation of accuracy and loss for different models

由图 5 可知,随着迭代次数的增加,5 种模型的准确率逐步提升,而损失值逐渐降低,最终趋于稳定。在相同的测试集上对训练好的模型进行评估,并根据分类结果计算模型的精确率、召回率和准确率。在检测奶牛跛行的试验中,不同模型的性能结果如表 4 所示。

表 4 不同模型的性能结果

模型 Model	精确率 Precision	召回率 Recall	准确率 Accuracy
FN-BiLSTM	95.71	99.04	97.16
GRU	81.97	79.37	80.00
TCN	82.19	83.49	81.83
LSTM	89.22	86.67	87.50
BiLSTM	91.64	93.67	92.33

由表 4 可知,改进后的 FN-BiLSTM 的精确率、召回率和准确率分别为 95.71%、99.04%、97.16%,比 BiLSTM 提高了 4.07、5.37、4.83 个百分点。FN-BiLSTM 模型的准确率分别比 GRU、TCN、LSTM 提高了 17.16、15.33、9.66 个百分点。分析原因可知,由于 GRU 的门控结构相对简单,对长期依赖的建模能力不如 LSTM 和 BiLSTM。TCN 在处理非平移不变特征方面能力较为有限,而单向 LSTM 对顺序较为敏感,当奶牛的行走方向相反时,关键点的顺序也会相反,导致模型表现不佳。相较之下,BiLSTM 能够双向处理序列数据,有效捕捉长期依赖关系。在此基础上,FN-BiLSTM 引入了 Filter 层和 Noise 层,更好地提升了模型对复杂特征和噪声的处理能力。因此,FN-BiLSTM 模型的综合性能更优,更适合用于奶牛跛行检测。

3 奶牛跛行检测模型对比试验

为了客观评估本文所提出的实例分割_关键点检测(头颈背)_FN-BiLSTM 模型在跛行检测中的性能,本研究使用相同数据集比较了 5 组试验的跛行检测效果。

试验 1 使用 JIANG 等^[29-30]提出的 CBPE_CORE_CBCE_Noise+BiLSTM 模型。该方法通过 CBPE 模块提取奶牛背部位置坐标,使用 CORE 模块获得奶牛的像素区域,再通过 CBCE 模块提取奶牛背部曲率,最后利用 Noise+BiLSTM 模块预测曲率数据,匹配跛行奶牛的曲

率特征,从而实现跛行检测。试验 2 将试验 1 中的 CORE 模型的首尾帧差分法替换为实例分割算法。试验 3 使用奶牛四蹄的关键点,试验 4 在获取奶牛运动特征时采用传统帧差法。试验 5 是本文提出的模型。试验结果如表 5 所示。

表 5 奶牛跛行检测模型对比

试验序号 Test No.	模型 Model	关键点个数 Number of key points	准确率 Accuracy/%
1	CBPE_CORE_CBCE_Noise+BiLSTM	3	75.00
2	CBPE_实例分割_CBCE_Noise+BiLSTM	3	87.50
3	实例分割_关键点检测(蹄部)_FN-BiLSTM	4	88.46
4	帧差法_关键点检测(头颈背)_FN-BiLSTM	6	63.18
5	实例分割_关键点检测(头颈背)_FN-BiLSTM	6	97.16

由表 5 可知,试验 2 的准确率比试验 1 提高了 12.5 个百分点,试验 5 比试验 4 提高了 33.98 个百分点,证明结合实例分割算法(试验 2、5)的跛行检测模型均表现出较高的准确性,由于帧差法易受背景和光照变化的影响,当光照充足或光线较暗,且奶牛距离前景栏杆较远时,奶牛背部容易受到背景水泥墙面与铁丝网的干扰,导致奶牛像素区域不完整,从而影响关键点的提取,致使准确率较低。在试验 2 中,通过 CBCE 模块结合 3 点共圆定理获取背部曲率数据,并根据奶牛的曲率特征判断其是否跛行。然而,轻度跛行奶牛的背部曲率与健康奶牛相似,导致检测模型难以区分二者,因此,仅依靠背部特征信息具有一定的局限性。此外,不同关键点的组合(如蹄部或头、颈、背部关键点)对准确率有一定的影响。蹄部关键点和头、颈、背部关键点均能提供重要的运动信息,但在 FN-BiLSTM 模型中,使用头、颈、背部关键点时,跛行检测的准确率达到 97.16%,表明这些关键点能够更稳定地捕捉奶牛的姿态变化,并提供更可靠的运动信息。试验 5 结合实例分割和多关键点检测提升了跛行检测的准确性。结果表明,在跛行检测中,结合实例分割与足够数量的关键点,是提升检测准确性的一种有效策略。

4 结论

针对传统的图像处理方式容易受光线变化和遮挡的影响,使用少量关键点时鲁棒性较差的问题,本文提出了一种基于头颈背部关键点的奶牛跛行检测算法。结论如下:

1) YOLOv8n-seg 实例分割模型在获取奶牛图像序列及其全身语义信息方面表现出色,尤其在养殖场环境下光线变化、栏杆遮挡、牛只与摄像头距离远近等不同条件下,在测试集上的精确率、召回率、mAP 分别达到了 99.97%、100%、99.50%。

2) 在关键点提取阶段,使用按目标检测框裁剪后的分割结果图作为输入,以 ResNet-152 网络为主干网络时,模型性能最佳,在训练集和测试集上的均方根误差分别

为 2.04 像素和 4.28 像素。

3) 在基于关键点的奶牛跛行检测试验中, 随机选取了 16 头奶牛 16 段奶牛行走的视频进行测试。在本文构建的 FN-BiLSTM 模型上, 跛行检测的准确率、精确率、召回率分别达到了 97.16%、95.71%、99.04%。

此外, 由于养殖场中跛行奶牛数量较少, 本研究的分类样本相对有限, 未来将通过扩充数据集来提升模型的泛化能力; 其次, 本文算法使用头、颈、背部特征时, 能有效检测跛行, 但未融合蹄部关键点, 未来研究将着重探索结合奶牛全身关键点的跛行识别方法, 以进一步提升算法性能。

[参 考 文 献]

- [1] GARVEY M. Lameness in dairy cow herds: Disease aetiology prevention and management[J]. *Dairy*, 2022, 3(1): 199-210.
- [2] TSOUSIS G, BOSCO C, PRAXITELOUS A. The negative impact of lameness on dairy cow reproduction[J]. *Reproduction in Domestic Animals*, 2022, 57: 33-39.
- [3] 韩书庆, 张晶, 程国栋, 等. 奶牛跛行自动识别技术研究现状与挑战[J]. *智慧农业 (中英文)*, 2020, 2(3): 21-36. HAN Shuqing, ZHANG Jing, CHENG Guodong, et al. Current state and challenges of automatic lameness detection in dairy cattle[J]. *Smart Agriculture*, 2020, 2(3): 21-36. (in Chinese with English abstract)
- [4] THORUP V M, MUNKSGAARD L, ROBERT P E, et al. Lameness detection via leg-mounted accelerometers on dairy cows on four commercial farms[J]. *Animal*, 2015, 9(10): 1704-1712.
- [5] 杨丽娟, 张永, 刘德环, 等. 基于压力分布测量系统的奶牛跛行早期识别[J]. *农业机械学报*, 2016, 47(S1): 426-432. YANG Lijuan, ZHANG Yong, LIU Dehuan, et al. Early recognition for dairy cow lameness based on pressure distribution measurement system[J]. *Transaction of the Chinese Society for Agricultural Machinery*, 2016, 47(S1): 426-432. (in Chinese with English abstract)
- [6] 苏力德, 张永, 王健, 等. 基于改进动态时间规整算法的奶牛步态分割方法[J]. *农业机械学报*, 2020, 51(7): 52-59. SU Lide, ZHANG Yong, WANG Jian, et al. Segmentation method of dairy cattle gait based on improved dynamic time warping algorithm[J]. *Transaction of the Chinese Society for Agricultural Machinery*, 2020, 51(7): 52-59. (in Chinese with English abstract)
- [7] 李前, 初梦苑, 康熙, 等. 基于计算机视觉的奶牛跛行识别技术研究进展[J]. *农业工程学报*, 2022, 38(15): 159-169. LI Qian, CHU Mengyuan, KANG Xi, et al. Research progress on lameness recognition technology in dairy cows using computer vision[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(15): 159-169. (in Chinese with English abstract)
- [8] FLOWER F C, SANDERSON D J, WEARY D M. Hoof pathologies influence kinematic measures of dairy cow gait[J]. *Journal of Dairy Science*, 2005, 88(9): 3166-3173.
- [9] 宋怀波, 姜波, 吴倩, 等. 基于头颈部轮廓拟合直线斜率特征的奶牛跛行检测方法[J]. *农业工程学报*, 2018, 34(15): 190-199. SONG Huaibo, JIANG Bo, WU Qian, et al. Detection of dairy cow lameness based on fitting line slope feature of head and neck outline[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2018, 34(15): 190-199. (in Chinese with English abstract)
- [10] POURSABERI A, BAHR C, PLUK A, et al. Real-time automatic lameness detection based on back posture extraction in dairy cattle: Shape analysis of cow with image processing techniques[J]. *Computers and Electronics in Agriculture*, 2010, 74(1): 110-119.
- [11] 康熙, 李树东, 张旭东, 等. 基于热红外视频的奶牛跛行运动特征提取与检测[J]. *农业工程学报*, 2021, 37(23): 169-178. KANG Xi, LI Shudong, ZHANG Xudong, et al. Features extraction and detection of cow lameness movement based on thermal infrared videos[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2021, 37(23): 169-178. (in Chinese with English abstract)
- [12] JIANG B, YIN X, SONG H. Single-stream long-term optical flow convolution network for action recognition of lameness dairy cow[J]. *Computers and Electronics in Agriculture*, 2020, 175: 105536.
- [13] ZHAO K, ZHANG M, JI J, et al. Automatic lameness scoring of dairy cows based on the analysis of head-and back-hoof linkage features using machine learning methods[J]. *Biosystems Engineering*, 2023, 230: 424-441.
- [14] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: From error visibility to structural similarity[J]. *IEEE Trans Image Process*, 2004, 13(4): 600-612.
- [15] SPRECHER D J, HOSTETLER D E, KANEENE J B. A lameness scoring system that uses posture and gait to predict dairy cattle reproductive performance[J]. *Theriogenology*, 1997, 47(6): 1179-1187.
- [16] O'Callaghan K A, CRIPPS P, DOWNHAM D Y, et al. Subjective and objective assessment of pain and discomfort due to lameness in dairy cattle[J]. *Journal of Applied Animal Welfare Science*, 2003, 12(4): 605-610.
- [17] JOCHER G, CHAURASIA A, QIU J. Ultralytics YOLOv8[EB/OL]. (2023-02-30) [2024-04-01] <https://github.com/ultralytics/ultralytics>.
- [18] MATHISA, MAMIDANNA P, CURY K M, et al. DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning[J]. *Nature Neuroscience*, 2018, 21(9): 1281-1289.
- [19] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 4510-4520.
- [20] WU Z, SHEN C, VAN DEN HENGEL A. Wider or deeper: Revisiting the resnet model for visual recognition[J]. *Pattern Recognition*, 2019, 90: 119-133.
- [21] NATH T, MATHISA A, CHEN A C, et al. Using DeepLabCut for 3D markerless pose estimation across species and behaviors[J]. *Nature Protocols*, 2019, 14(7): 2152-2176.
- [22] MATHISA A, BIASI T, SCHNEIDER S, et al. Pretraining boosts out-of-domain robustness for pose estimation[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Honolulu, HI, USA: IEEE, 2021: 1859-1868.
- [23] 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]. *Transaction of the Chinese Society for Agricultural Machinery*, 2023, 54(2): 249-255, 292. (in Chinese with English abstract)
- [24] NASIRI A, YODER J, ZHAO Y, et al. Pose estimation-based lameness recognition in broiler using CNN-LSTM network[J]. *Computers and Electronics in Agriculture*, 2022, 197: 106931.
- [25] LEA C, FLYM M D, VIDAL R, et al. Temporal convolutional networks for action segmentation and detection[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA: IEEE, 2017: 156-165.
- [26] DEY R, SALEM F M. Gate-variants of gated recurrent unit (GRU) neural networks[C]// *2017 IEEE 60th International Midwest Symposium on Circuits and Systems*. Boston, MA, USA: IEEE, 2017: 1597-1600.
- [27] VAN H G, MOSQUERA C, NAPOLES G. A review on the long short-term memory model[J]. *Artificial Intelligence Review*, 2020, 53(8): 5929-5955.
- [28] SHSRFUDDIN A A, TIHAMI M N, ISIAM M S. A deep recurrent neural network with bilstm model for sentiment classification[C]// *2018 International Conference on Bangla Speech and Language Processing*. Sylhet, Bangladesh: IEEE, 2018: 1-4.
- [29] JIANG B, WU Q, YIN X, et al. FLYOLOv3 deep learning for key parts of dairy cow body detection[J]. *Computers and Electronics in Agriculture*, 2019, 166: 104982.
- [30] JIANG B, SONG H, WANG H, et al. Dairy cow lameness detection using a back curvature feature[J]. *Computers and Electronics in Agriculture*, 2022, 194: 106729.

Detecting cow lameness using the key points of head, neck, and back

ZHANG Zhirong^{1,2}, CHU Yanhua^{1,2*}, WANG Yueming^{2,3}, WANG Liying^{1,2}, SHEN Yuhao^{1,2}, LI Xin^{1,2}

(1. *School of Digital and Intelligent Industry(School of Cyber Science and Technology), Inner Mongolia University of Science and Technology, Baotou 0140010, China;* 2. *Engineering Laboratory for Big Data Traceability in Grassland Animal Husbandry, Inner Mongolia Autonomous Region, Baotou 0140010, China;* 3. *School of Automation and Electrical Engineering, Inner Mongolia University of Science and Technology, Baotou 0140010, China*)

Abstract: Cow lameness has represented a significant challenge on the economic viability of dairy operations. The overall performance can increase the risk of health issues for affected cows, leading to reduced milk production. Consequently, cow lameness is crucial to maintain both the welfare of the herd and the profitability of dairy farms. Lameness typically shows the observable indicators during walking, such as a lowered head position, pronounced head movement, and an arched back, whereas, healthy cows demonstrate the minimal head movement, straight back, normal gait and body equilibrium. In this study, a deep learning-based algorithm was proposed to automatically detect the lameness in cows, according to these outstanding movement features. A systematic investigation was implemented to detect the cow lameness, thus tracking the movement patterns of six key anatomical points: the head, neck, shoulder, center of the back, loin, and tail. Firstly, two mobile devices were positioned adjacent to the passage, leading to the milking area. The video data was collected for 160 walking sequences from 83 cows. YOLOv8n-seg instance segmentation was employed to accurately identify the cows in the images, and then extract their coordinates and pixel regions. The computational efficiency and accuracy were improved to reduce the effects of light variations in the channel, background barbed wire fence boundaries, and foreground fence occlusion. Secondly, the six types of keypoint detection datasets were constructed after instance segmentation, including RGB images, binary mask images, segmentation images along with their cropped versions, according to the target detection frame. Four backbone networks, MobileNet-V2, ResNet-50, ResNet-101, and ResNet-152, were used to train and test these datasets. Segmented images that cropped by the detection frame were selected as the optimal input format, with ResNet-152 chosen as the best-performing backbone network. Then, the DeepLabCut algorithm was used to automatically extract the coordinates of six key points from the video sequences: the head, neck, shoulder, center of the back, waist, and tail, resulting in the creation of a lameness detection dataset. Lastly, a comparative analysis was performed to evaluate the performance of Temporal Convolutional Network (TCN), Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), Bidirectional LSTM, BiLSTM, and FN-BiLSTM models in the claudication detection. (Bidirectional LSTM, BiLSTM), and FN-BiLSTM models in the lameness detection. Ablation experiments were conducted on the FN-BiLSTM model, in order to verify the effects of the Filter and Noise layers on the lameness detection in cows. The results demonstrated that the FN-BiLSTM model was achieved in the optimal performance with 97.16% accuracy, 95.71% precision, and 99.04% recall for the lameness recognition on a test set of 16 videos from 16 cows. Moreover, the instance segmentation model exhibited the high efficacy to capture the image sequences of cows and their whole-body semantic information, even under variable illumination conditions and different bovine-to-camera distances. The precision, recall, and mAP of the test set reached 99.97%, 100%, and 99.5%, respectively. During the keypoint detection phase, the optimal performance was achieved, when utilizing the cropped segmentation maps as input, with ResNet-152 as the backbone network, resulting in the root mean square errors of 2.04 pixels and 4.28 pixels for the training and test sets, respectively. These findings can offer a valuable technical approach for the automated detection of cow lameness in the livestock industry. This finding has the potential to enhance the efficiency and animal welfare of dairy operations, thereby promoting the sustainable development of the livestock.

Keywords: cow lameness; deep learning; instance segmentation; key point detection