



甘霖惠,杜佳磊,麻晓丽,余有信,朱文博,刘宇,王步钰. 基于RDS-Mask R-CNN的绵羊姿态自动检测方法研究[J]. 中国农业大学学报,2026,31(02):172-182.
GAN Linhui, DU Jialei, MA Xiaoli, YU Youxin, ZHU Wenbo, LIU Yu, WANG Buyu. Research on automated sheep posture detection based on RDS-Mask R-CNN[J]. *Journal of China Agricultural University*, 2026, 31(02): 172-182.
DOI: 10.11841/j.issn.1007-4333.2026.02.15

基于RDS-Mask R-CNN的绵羊姿态自动检测方法研究

甘霖惠 杜佳磊 麻晓丽 余有信 朱文博 刘宇 王步钰*

(内蒙古农业大学 计算机与信息工程学院, 呼和浩特 010011)

摘要 绵羊的姿态与其健康及福利密切相关。随着智能化畜牧业需求的增长,自动、准确地检测绵羊姿态尤为重要。本研究提出基于Mask R-CNN基准网络的新型RDS-Mask R-CNN绵羊姿态检测算法,以Res2Net101作为特征提取网络,同时引入可变形卷积(Deformable convolution network, DCN),以更精准捕捉绵羊在不同位置的姿态特征,并运用软非极大值抑制(Soft non-maximum suppression, Soft NMS)算法实现重叠实例目标的准确分割。结果表明:1)目标检测框架算法对比:与该领域最经典的YOLOv3和Faster R-CNN相比,改进的算法在平均精度均值(Mean average precision, mAP)上分别提升了16.68%和8.64%;2)不同改进策略的算法对比:改进算法相较于基准网络,边界框平均精度均值(Bounding box mean average precision, Bbox mAP)提高6.21%,分割平均精度均值(Segmentation mean average precision, Segm mAP)提高6.61%,分别达到87.34%和81.50%;3)相较于Mask R-CNN,改进模型在识别绵羊站立与躺卧姿态时边界框平均精度(Bounding box average precision, Bbox AP)分别提高了6.84%和5.58%,分割平均精度(Segmentation average precision, Segm AP)分别提高了7.25%和5.17%;4)模型可解释性可视化结果表明RDS-Mask R-CNN能精准捕获绵羊站立和躺卧姿态关键部位深度特征,表明模型自动检测可行且具有可解释性。综上,本研究提出的RDS-Mask R-CNN算法,有效提升了绵羊姿态检测的精准度,为智慧养殖提供了技术支撑。

关键词 绵羊姿态识别; RDS-Mask R-CNN; 可变形卷积

中图分类号 S826 文章编号 1007-4333(2026)02-0172-11 文献标志码 A

Research on automated sheep posture detection based on RDS-Mask R-CNN

GAN Linhui, DU Jialei, MA Xiaoli, YU Youxin, ZHU Wenbo, LIU Yu, WANG Buyu*

(School of Computer and Information Engineering, Inner Mongolia Agricultural University, Hohhot 010011, China)

Abstract The posture of sheep is closely linked to their health and welfare. With the growing demand for smart livestock farming, automatic and accurate detection of sheep posture has become increasingly important. This study proposes a novel RDS-Mask R-CNN algorithm for sheep posture detection based on the Mask R-CNN baseline. Res2Net101 is employed as the feature extraction network, and the deformable convolution network (DCN) is introduced to more accurately capture posture features of sheep in different positions, while the soft non-maximum suppression (Soft NMS) algorithm is used to achieve accurate segmentation of overlapping instances. The results show that: 1) Comparison of object detection framework algorithms: Compared with the classic YOLOv3 and Faster R-CNN in this field, the proposed algorithm has improved mean average precision

收稿日期: 2025-05-15; 接受日期: 2025-07-26

基金项目: 国家乳业技术创新中心项目(2023-JSGG-4); 内蒙古自治区“英才兴蒙”工程团队项目(2025TYL10); 内蒙古农业大学青年教师科研能力提升计划(BR220148)

第一作者: 甘霖惠(ORCID:0009-0003-0154-6494), 硕士研究生, E-mail: ganlh@emails.imau.edu.cn

通讯作者: 王步钰(ORCID:0000-0002-5002-1219), 教授, 主要从事数字技术在畜牧应用领域的研究, E-mail: bywang05@163.com

(mAP) by 16.68% and 8.64%, respectively. 2) Comparison of algorithms with different improvement strategies: The proposed algorithm increases the bounding box mean average precision (Bbox mAP) by 6.21% and segmentation mean average precision (Segm mAP) by 6.61% compared to the baseline, reaching 87.34% and 81.50%, respectively. 3) Compared with Mask R-CNN, the proposed model improved bounding box average precision (Bbox AP) by 6.84% and 5.58%, respectively, and segmentation average precision (Segm AP) by 7.25% and 5.17%, respectively, in recognizing the standing and lying postures of sheep. 4) Model interpretability visualization: This indicates that RDS-Mask R-CNN can accurately capture the deep features of key body parts in sheep's standing and lying postures, demonstrating the feasibility and interpretability of automatic detection by the model. In conclusion, the proposed RDS-Mask R-CNN algorithm effectively improves the accuracy of sheep posture detection, providing technical support for smart livestock farming.

Keywords sheep posture recognition; RDS-Mask R-CNN; deformable convolution network

畜牧业与国家粮食安全、居民生活水平密切相关,是农业的重要支柱^[1]。作为五大牧区之一,内蒙古的畜牧业对其经济发展意义重大,其中养羊业作为核心支柱,成为经济发展的关键驱动力^[2]。现代养羊业正逐步向智慧养殖转型,绵羊作为规模化集约化养殖的核心畜种,其站立和躺卧姿态检测成为智慧生产管理的基础,通过高精度检测,可以为健康监测(如分析躺卧时长辅助蹄病预警)和繁殖管理(如发情期站立姿态特征识别)提供数据支撑,推动智慧养殖实践。同时,绵羊姿态检测技术能克服人工观察的偏差,为实现连续监测提供技术基础,也为养羊业从经验向数据转型提供了技术支撑。

传统检测方法如在动物颈部^[3]、腿部^[4]等部位佩戴传感器,部署成本高、易损坏,也易引发动物的应激反应,因此,非接触、无应激的绵羊姿态检测成为研究热点^[5]。近年卷积神经网络推动计算机视觉和深度学习技术革新,成为图像识别^[6]、智慧农业^[7]等领域的核心驱动力。深度卷积神经网络在畜牧业生产中优势显著,其在检测上呈现三大亮点:即时响应的实时性、高效的处理能力以及连续工作特性。凭借这些特性,深度卷积神经网络能够精准执行目标识别^[8]、姿态检测^[9]和目标跟踪^[10]等任务。在基于计算机视觉的非接触式绵羊的姿态检测中,现有研究已实现较高的准确率。例如,Yu等^[11]提出结合EfficientNet-B0轻量级网络和SENet注意力机制的YOLOv3目标检测模型,以识别母羊的发情行为,平均准确率达99.78%;Jiang等^[12]采用YOLOv4目标检测模型,对山羊的采食、饮水、活动和非活动行为进行识别,平均准确率超96.86%;Gu等^[13]基于深度学习构建了两阶段绵羊姿态检测方法,分类阶段所有行为的分类准确率超94%;Deng等^[14]采用YOLOv3算法检测绵羊的站立、进食和躺卧3种姿

态,平均精度达到92.47%,这些研究证明了深度卷积神经网络在姿态检测中的有效性,为后续研究奠定了重要基础。然而,现有研究仍存在一些局限:1)数据集多样性不足。如Yu等^[11]针对母羊发情行为、Jiang等^[12]针对山羊行为的识别,均未考虑不同羊品种,品种间的差异(如毛色、躯体比例)可能导致模型学到的特征不同。2)环境存在一定局限性。如Gu等^[13]的研究虽实现了高分类准确率,但数据来源局限于养殖圈舍内,未涉及户外环境。3)遮挡情况处理不足。如Deng等^[14]的研究虽有92.47%的平均精度,但其明确指出,当个体重叠或被障碍物遮挡时,被遮挡个体的检测结果不准确,反映出模型对不完整目标特征的提取能力不足,难以适应真实规模化集约化养殖中的复杂干扰。

综上,本研究将计算机视觉与深度学习技术相结合,提出了一种基于RDS-Mask R-CNN的绵羊姿态检测新方法。具体而言,在目标检测算法Mask R-CNN的基础上,通过三方面改进实现优化:首先将特征提取网络ResNet替换为Res2Net,更有效地提取多尺度特征;其次,引入可变形卷积,动态调整卷积核形状,适应目标形态变化,更好地处理目标重叠现象,提升复杂场景下的检测能力;最后,采用Soft NMS算法,在目标重叠时能更有效地保留有效边界框,减少因重叠导致的误识别,进一步提高检测准确性。本研究构建了涵盖不同品系与养殖环境的绵羊姿态数据集,旨在提升复杂养殖场景下的绵羊姿态检测精度,为绵羊的规模化、集约化的智慧养殖提供技术支持。

1 材料与方法

1.1 数据采集与预处理

在2022年10月8日—22日期间,从呼和浩特市

2个集约化示范羊场采集了连续15 d的视频数据。包含4个品种共124只绵羊,其中一个室外半开放羊场(由圈舍和露天运动场构成,数据从运动场采集)饲养黑头杜泊羊,另一个非全封闭室内羊场混养杜蒙羊、杜泊羊、萨福克羊和陶赛特羊,采集的数据为后续构建姿态数据集提供支撑。

本研究利用云边端架构传输数据,绵羊视频数据采集流程见图1。羊舍布设科达球形云台摄像机,支持1 920像素×1 080像素全高清画质,以25帧/秒的速率实时采集视频,并采用H.264编码

技术进行数据压缩。通过调节摄像机角度,使其拍摄范围能够全面覆盖绵羊日常活动区域。将羊场机房内的台式PC配置为边缘服务器,利用以太网实现摄像机与服务器的数据交互。基于RTSP协议,边缘服务器借助FFMPEG软件实时捕获可见光通道画面,将获取的视频流同步存储至硬盘,文件名为“采集起始时刻_摄像机编号”,10 min自动生成新的MP4文件。边缘服务器通过Rsync协议向存储服务器增量传输文件后,自动清理文件释放空间。

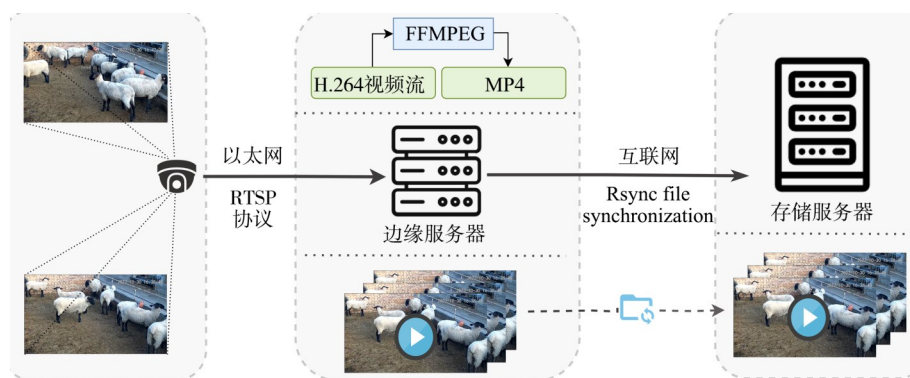


图1 绵羊视频数据采集流程图

Fig. 1 Flowchart of sheep video data acquisition

为构建用于模型训练、验证和测试所需的数据集,利用FFMPEG软件对采集的视频进行自动化抽帧与专家选帧。研究发现,相邻帧中绵羊姿态数据存在显著重叠现象,为防止样本重复削弱模型泛化效果,以3 s为周期进行图像采样,人工筛选后共得到8 720张涵盖不同光照和遮挡情况的绵羊图像,筛选遵循图像质量、姿态代表性、场景多样性与羊只分布标准,由2名专家独立初筛并协商确定最终图像集,以降低主观偏差。

1.2 数据集构建

构建的数据集涵盖“Stand”(站立)和“Lay”(躺卧)两类姿态。站立状态表现为四肢稳定支撑地面,躯干保持直立;躺卧状态则呈现躯体与地面大面积接触,四肢自然伸展或微屈的特征。标注由两名经专业培训的标注员(标注前评估Kappa系数=0.85)采用Labelme^[15]工具对8 720张图像进行多边形轮廓标注(不标注被遮挡率超过90%的绵羊),为每只羊分配对应的姿态标签,并生成JSON文件(记录图像路径和大小、标注形状、绵羊位置及类别信息),标注分歧通过讨论与仲裁确保标注一致,绵羊

姿态标注样例图如图2所示,数据集按7:2:1比例随机划分为训练集、验证集和测试集,三者图像互不重复以确保模型的泛化能力。

数据集的构建:建立用于存放原始图像的JPEGImages目录和存储标注信息的Annotation目录,随后调用Labelme官方提供的Python脚本,将标注数据转换为COCO^[16]格式,最后将图像与COCO格式标注文件分别归类至对应目录。最终,训练集有6 704张图像(19 989个“Stand”,19 566个“Lay”),验证集有1 344张图像(5 600个“Stand”,5 460个“Lay”),测试集有672张图像(3 000个“Stand”,2 706个“Lay”)。如图3所示,图片数量(a)与绵羊数量(b)数据集分布图。

数据未采用增强技术(如旋转、翻转或色彩扰动)的原因如下:(1)数据集已涵盖真实场景关键变量:光照(采集时段亮度波动)、遮挡(羊群互遮及设施遮挡)、多品系(4个绵羊品种)、姿态多角度(自然肢体伸展),真实场景的多样性有效抑制了过拟合风险,在一定程度上模拟了数据增强的效果。(2)为确保算法改进贡献的可归因性,排除增强技术对模



图 2 绵羊姿态标注样例图

Fig. 2 Sample diagram of sheep postures annotation

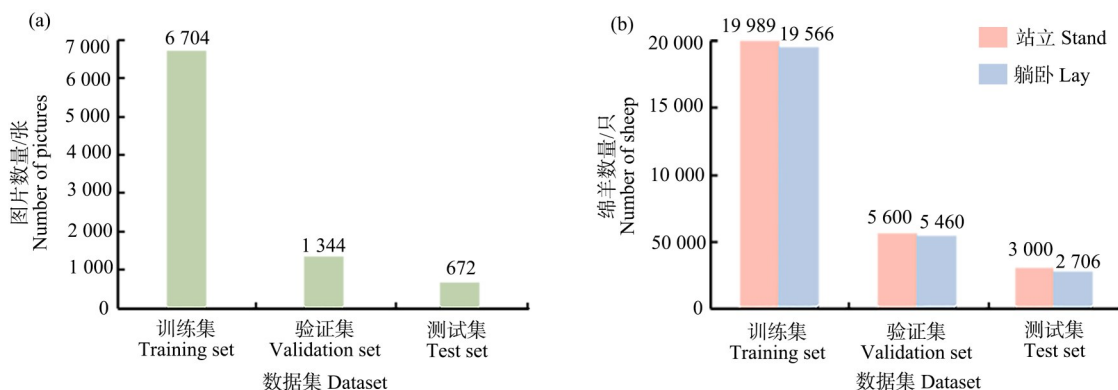


图 3 图片数量(a)与绵羊数量(b)数据集分布图

Fig. 3 Distribution of the number of pictures (a) and the number of sheep (b) in dataset

型能力的干扰,训练全程使用原始数据。

1.3 多目标绵羊姿态检测模型

在真实养殖场景中,实现多目标绵羊姿态的精准检测,面临着多重挑战,不仅是羊只之间会相互遮挡,羊场内的栅栏、柱子等固定设施也会遮挡羊只,同时,不同时段的自然光照变化,导致采集的图像亮度不

均、噪声增加,此外,绵羊在图像中的空间跨度较大,远处绵羊的细节模糊,且不同角度的绵羊外形有差异,加大了特征提取的难度。针对上述难题,本研究在 Mask R-CNN 基础上,融入了 Res2Net101 网络、可变形卷积 DCN 和 Soft NMS 算法,构建了 RDS-Mask R-CNN 模型,模型结构图如图 4。

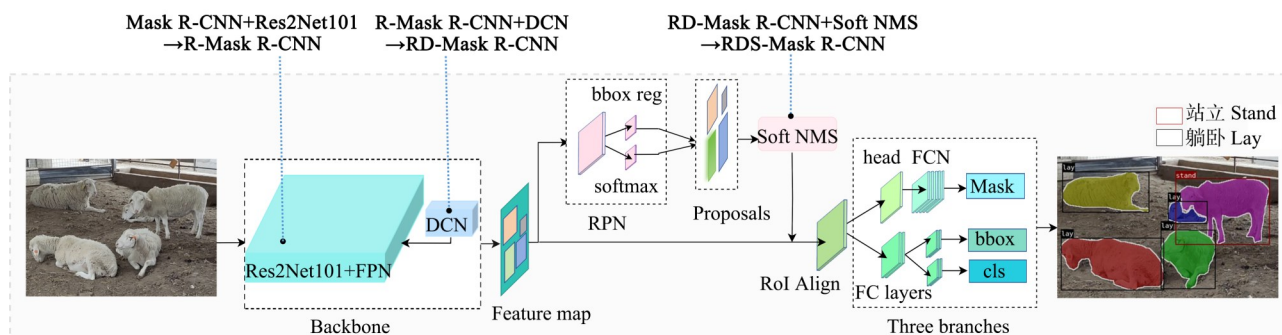


图 4 RDS-Mask R-CNN 模型结构图

Fig. 4 Structural diagram of RDS-Mask R-CNN model

1.3.1 特征提取网络改进

Mask R-CNN的ResNet101残差连接机制仅支持单一尺度特征在残差块内局部传递,使得网络在处理深层特征时,难以充分融合多尺度信息,最终导致高层语义特征的表达能力不足。杨巧宁等^[17]研究表明,Res2Net网络相较于传统的ResNet网络,能够更有效地提取多尺度特征;进一步地,Li等^[18]在复杂交通场景(存在目标密集遮挡与尺度变化)的消融试验表明:使用Res2Net101网络的模型的遮挡目标检测mAP为94.6%,显著优于使用ResNet101网络的mAP(89.6%),表明了它在处理尺度变化与遮挡问题的有效性。因此,本研究选用Res2Net101网络^[19]替代原有结构,该网络突破ResNet101的局限,通过设计多层级残差连接结构,实现特征在不同尺度下的并行传输与交叉融合,有效增强了特征提取的丰富性与准确性。网络的输出表达如式(1)。

$$y_i = \begin{cases} x_i & i=1; \\ K_i(x_i) & i=2; \\ K_i(x_i + y_{i-1}) & 2 < i \leq s \end{cases} \quad (1)$$

式中: y_i 为第*i*个输出特征; x_i 为第*i*个输入特征子集; $i \in \{1, 2, \dots, s\}$,其中*i*为特征组的序号,*s*为Res2Net模块中特征组的数量; K_i 为第*i*个 3×3 卷积算子。

如图5所示,ResNet残差结构(a)与Res2Net残差结构(b),与传统ResNet的单路径卷积不同,Res2Net将卷积操作划分为多个并行分支,并在后续分支中融入前一分支的特征输出。这使得感受野不断扩大,例如 y_2 的感受野为 3×3 , y_3 为 5×5 , y_4 为 7×7 ,从而能够捕捉到图像中更大范围的上下文信息,有助于模型更好地理解不同尺度的特征。最终,所有分支的输出通过 1×1 卷积整合到一起,获得更加丰富的多尺度特征。综上,Res2Net101与Mask R-CNN融合构建的R-Mask R-CNN,能有效提升绵羊多尺度特征捕捉性能。

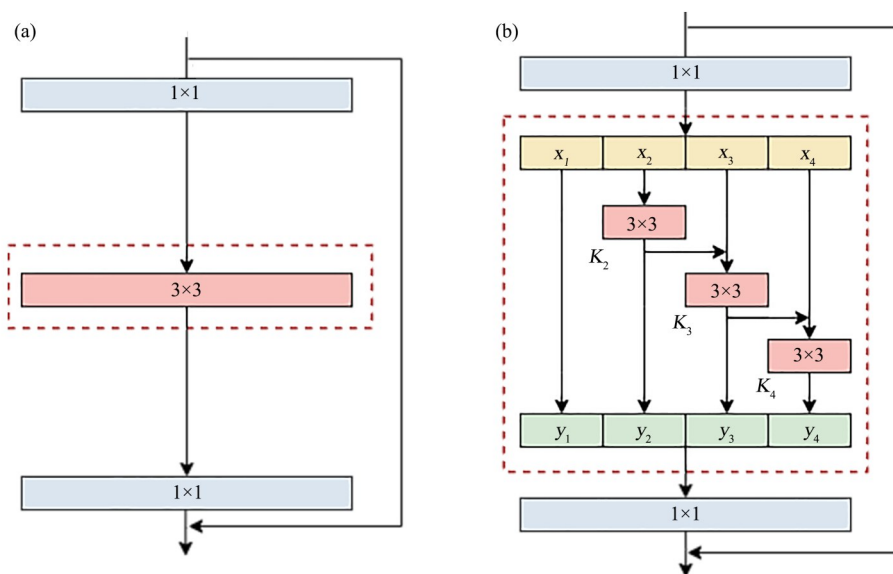


图5 ResNet残差结构(a)与Res2Net残差结构(b)

Fig. 5 ResNet residual structure (a) vs Res2Net residual structure (b)

1.3.2 卷积方式改进

传统卷积核因固定形态与尺寸的局限性,难以精准匹配姿态多变、轮廓复杂的绵羊目标。所以本研究对Res2Net101网络架构针对性优化,将Res_3至Res_5层中的 3×3 标准卷积单元替换为DCN^[20],从而实现绵羊形态与姿态信息的高精度提取,如图6所示,添加可变形卷积的Res2Net101网络结构图,即对30个残差模块进行了改造。改造后,

Res2Net101通过DCN中的偏移预测网络灵活调节卷积核采样位,以适配绵羊不规则形态。经上述改进,生成了RD-Mask R-CNN模型。

1.3.3 非极大值抑制算法改进

在集约化养殖环境中,绵羊群体密集导致目标遮挡频繁,容易导致边界框高度重叠。传统非极大值抑制(Non-maximum suppression, NMS)算法采用固定阈值策略,当重叠度超过阈值时,会舍弃与

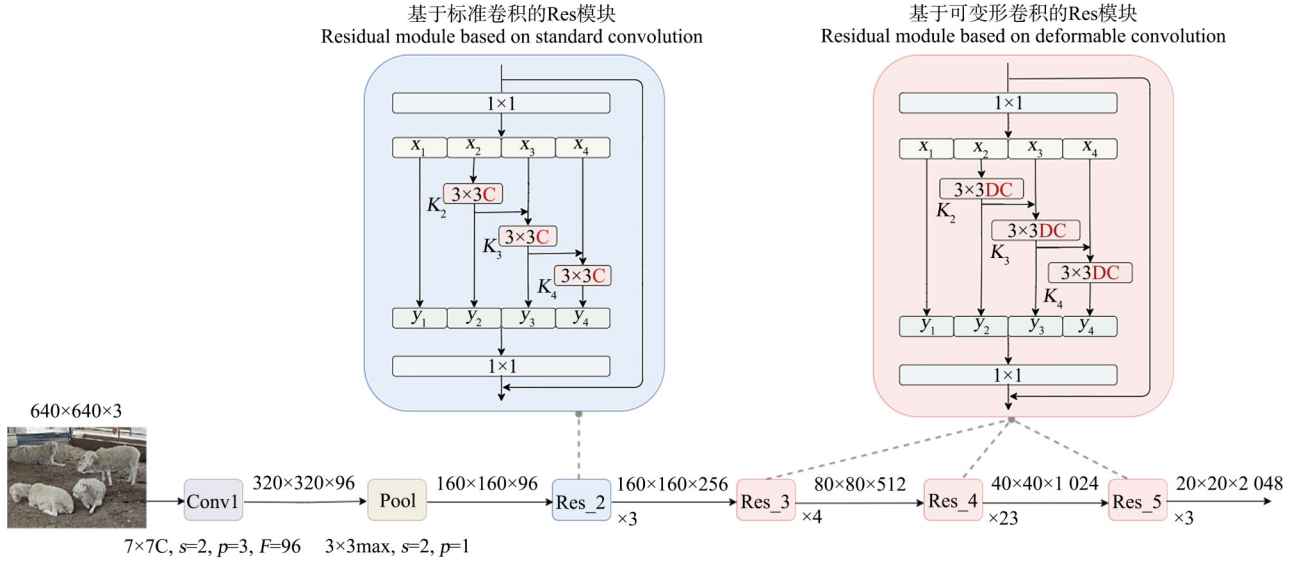


图6 添加可变形卷积的Res2Net101网络结构图

Fig. 6 Structural diagram of Res2Net101 network with deformable convolutions

最高得分框重叠的其他边界框,容易产生绵羊漏检。针对该问题,在RD-Mask R-CNN模型中引入Soft NMS算法,生成了RDS-Mask R-CNN模型。Soft NMS算法包括线性衰减与高斯衰减2种惩罚衰减函数。因线性衰减函数是不连续的,容易出现得分断层,所以使用高斯衰减函数(式(2)),当边界框与最高分框M存在大面积重叠时,Soft NMS算法不会立即将其分数置零,而是以平滑的方式衰减边界框得分,并通过参数 $\sigma=0.3$ 控制衰减速度,通过量化边界框间的交并比(IoU),生成衰减系数来调整非极大置信度框的分数,当框间重叠度较高时,衰减系数同步增大,实现对冗余框分数的阶梯式抑制,从而保留潜在目标信息,增强算法在复杂养殖场景下的稳定性和准确性。

$$S_i = S_i e^{-\frac{\text{IoU}(M, b_i)^2}{\sigma}}, \forall b_i \notin D \quad (2)$$

式中: S_i 为第*i*个边界框的得分;IoU(M, b_i)为边界框*M*与第*i*个边界框 b_i 的交并比; σ 为高斯函数的参数; D 为最终保留的边界框集合。

1.4 试验环境配置

在Ubuntu 20.04操作系统上进行试验,硬件环境包括Intel Xeon Gold 6139 (2.30GHz) CPU、256GB内存、NVIDIA GeForce RTX 3090显卡(24GB显存)和4TB硬盘。软件环境使用Anaconda 23.7.3、Python 3.10、PyTorch 2.0、

CUDA 11.7、CUDNN 8.5和MMDetection 2.28.2。训练采用SGD优化器,关键超参数设置如下:训练轮次50,批处理大小24,初始学习率0.02,动量0.9,权重衰减系数0.0001。超参数设置适配任务需求:动量0.9为目标检测任务中SGD的经典设置,通过平滑梯度更新抑制因样本姿态差异导致的训练波动,加速收敛;权重衰减系数0.0001用于缓解训练数据中光照变化、目标遮挡引发的过拟合,增强泛化能力;批处理大小24与训练轮次50的组合,兼顾RTX 3090的显存限制(24GB)与数据集规模(8720张图像),在保证训练效率的同时维持收敛稳定性。

1.5 评价指标

本研究采用精确率(Precision)、召回率(Recall)、F1分数(F1-Score)、平均精度(Average precision, AP)和mAP评估模型性能。在计算mAP时,采用了IoU为0.50:0.95的范围作为评价依据。

2 结果与分析

2.1 目标检测框架对比分析

为验证RDS-Mask R-CNN模型对自然场景下多目标绵羊的姿态识别效果,本研究将其与YOLOv3^[21]、Faster R-CNN^[22]以及Mask R-CNN目标检测算法进行了对比试验。结果如表1所示,

RDS-Mask R-CNN 的 mAP 达到 87.34%，相比 YOLOv3、Faster R-CNN 和 Mask R-CNN 分别提升了 16.68%、8.64% 和 6.21%，同时 Precision、Recall 和 F1 分数相较于 3 种目标检测算法均有提升。这说明改进的模型在检测精确率和召回率上均具备较好效果。

表 1 不同目标检测模型的性能评价
Table 1 Performance evaluation of different object detection models

模型 Model	平均精度均值/% mAP	精确率/% Precision	召回率/% Recall	F1 分数/% F1-Score
YOLOv3	70.66	95.67	96.83	96.25
Faster R-CNN	78.70	97.51	98.68	98.09
Mask R-CNN	81.13	97.87	98.92	98.39
RDS-Mask R-CNN	87.34	98.21	99.27	98.74

注：加粗数字表示该列指标中的最优值。下同。
Note: The bold numbers indicate the optimal values in the corresponding column indicators. The same as belows.

综上所述，RDS-Mask R-CNN 在自然生产环境中识别多目标绵羊姿态具有良好的适用性和稳定性。

2.2 不同改进策略的算法对比分析

本研究通过算法对比试验，系统验证了模型改进的有效性和每次改进对模型性能的贡献，结果如表 2 所示。

相较于 Mask R-CNN，R-Mask R-CNN 采用 Res2Net101 网络后，虽然 Recall 下降 0.09%，但 Bbox mAP 和 Segm mAP 分别提升 2.80% 和 2.13%，综合性能更优。这得益于 Res2Net101 的多级残差连接，扩大了网络感知范围并增强了特征表达，使模型更精准地捕捉绵羊多尺度特征。在此基础上，将可变形卷积添加到 Res2Net101 最后 3 个阶段的残差模块。结果显示，RD-Mask R-CNN 较 R-

Mask R-CNN 在各项指标上均实现提升，其中 Bbox mAP 和 Segm mAP 分别提升 2.82% 和 4.34%，借助 DCN 灵活调节卷积核采样位点的机制，显著提升了模型应对绵羊姿态变化的能力。将 RD-Mask R-CNN 的 NMS 替换为 Soft NMS 后，性能进一步提升：Bbox mAP 从 86.75% 增至 87.34%，Segm mAP 从 81.36% 增至 81.50%，同时 Recall 和 F1 分数分别提高 0.29% 和 0.15%。Soft NMS 相较于 NMS 有更柔和的重叠框抑制策略，保留了重叠目标的信息，提升了模型对多目标绵羊姿态的识别准确性。

综上，通过对 3 种策略优势的整合，RDS-Mask R-CNN 模型相较于 Mask R-CNN 模型在各项指标都有提升，其中在 Bbox mAP 和 Segm mAP 上分别累计提升了 6.21% 和 6.61%，验证了改进的有效性。

表 2 不同改进策略的算法对姿态识别的性能评价
Table 2 Performance evaluation of algorithms with different improvement strategies for postures recognition

算法 Algorithm	边界框平均精度均值/% Bbox mAP	分割平均精度均值/% Segm mAP	精确率/% Precision	召回率/% Recall	F1 分数/% F1-Score
Mask R-CNN	81.13	74.89	97.87	98.92	98.39
R-Mask R-CNN	83.93	77.02	98.10	98.83	98.46
RD-Mask R-CNN	86.75	81.36	98.21	98.98	98.59
RDS-Mask R-CNN	87.34	81.50	98.21	99.27	98.74

2.3 目标检测试验分析

为探究 RDS-Mask R-CNN 模型对绵羊不同姿态的检测与分割效能，将其与原始 Mask R-CNN 模型进行横向对比试验，验证改进模型在目标识别精

度上的提升，具体数据见表 3，相较于 Mask R-CNN，RDS-Mask R-CNN 对绵羊站立姿态的 Bbox AP 提高 6.84%，Segm AP 提高 7.25%；对躺卧姿态的 Bbox AP 提高 5.58%，Segm AP 提高 5.17%，站

立和躺卧的 Precision、Recall 也有小幅度提升。躺卧姿态的 Bbox AP 和 Segm AP 均高于站立姿态,分析其原因是:躺卧姿态下绵羊与地面接触面积大,轮廓特征显著且稳定;而站立姿态因肢体伸展、角度变化导致形态特征复杂度较高。

如图 7 所示,边界框(a)与分割(b)的精确率与召回率曲线,展示了 IoU 为 0.5:0.95 范围内绵羊站

立和躺卧姿态的 P-R 曲线。曲线向右上方凸起程度越大表明模型的性能越好,由图 7 可知,绵羊站立和躺卧姿态的 P-R 曲线在图中占据了较大的面积,反映出模型识别这 2 种姿态时,能够兼顾高精确率与高召回率,其中绵羊躺卧姿态因具备清晰可辨的形态特征,使得 RDS-Mask R-CNN 算法在处理该类目标时,能够更精准地识别。

表 3 基于不同目标检测算法的绵羊姿态识别性能评价

Table 3 Performance evaluation of sheep postures recognition based on different object detection algorithms

模型 Model	边界框平均精度/% Bbox AP		分割平均精度/% Segm AP		精确率/% Precision		召回率/% Recall	
	Stand	Lay	Stand	Lay	Stand	Lay	Stand	Lay
Mask R-CNN	77.90	84.35	70.80	79.79	97.01	98.74	98.29	99.55
RDS-Mask R-CNN	84.74	89.93	78.05	84.96	97.54	98.88	98.89	99.65

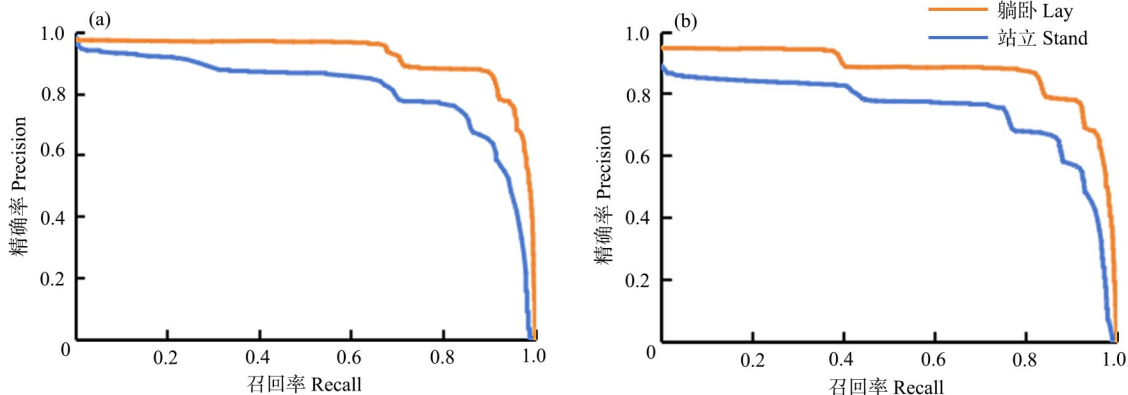


图 7 边界框(a)与分割(b)的精确率与召回率曲线

Fig. 7 Precision-recall curves for bounding box detection (a) and segmentation (b)

如图 8 所示,不同光照(a)和不同遮挡(b)的检测结果样例图,展示了 RDS-Mask R-CNN 在复杂养殖环境中的鲁棒性表现:如图 8(a)所示,强光、阴影及昏暗光照下,模型均能准确检测绵羊姿态,原因如下:绵羊白色毛发在阴影或昏暗环境中与背景色差明显;Res2Net101 融合浅层细节(如纹理边缘)与深层语义特征(如整体轮廓),缓解光照导致的局部模糊及昏暗环境下的小目标退化;DCN 动态调整感受野适应绵羊形态,提升了鲁棒性。如图 8(b)所示,不同遮挡情况下的识别效果:无论绵羊数量多少、遮挡程度轻重(包括羊只互遮或设施遮挡),模型均能较好实现目标分割与姿态识别,这得益于 Res2Net101 与 DCN 的协同作用,增强了模型对深度特征的挖掘

和复杂区域的关注。此外,Soft NMS 通过抑制冗余候选框,使模型能更好地捕捉重叠目标。

综上,RDS-Mask R-CNN 融合 Res2Net101、DCN 和 Soft NMS 的优势,对站立和躺卧姿态的检测与分割均有良好效果,且在光照不均与复杂遮挡挑战下,仍能实现多目标绵羊姿态的精确识别。

2.4 模型可解释性可视化分析

为实现对绵羊姿态识别模型的行为学可解释性分析,从测试集中各随机选取 3 张站立与躺卧的图像,运用 EigenCAM 可视化技术,得到 EigenCAM 图,对各姿态的关键特征区域进行深度剖析。图 9 展示了 RDS-Mask R-CNN 模型对绵羊站立和躺卧姿态的可视化结果。

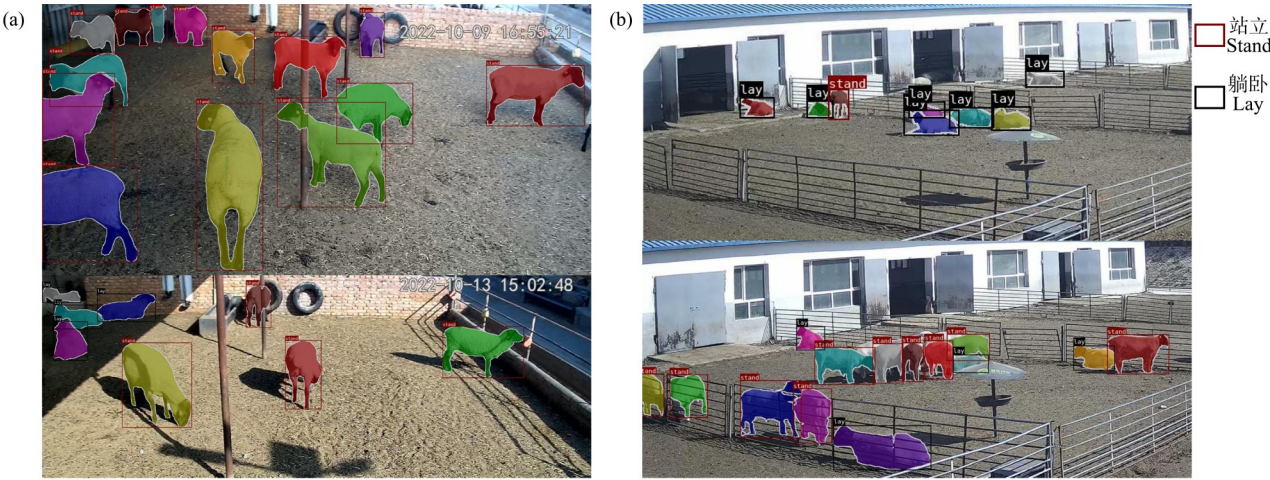


图8 不同光照(a)和不同遮挡(b)的检测结果样例图

Fig. 8 Example images of detection results for varying illumination (a) and varying occlusion (b)

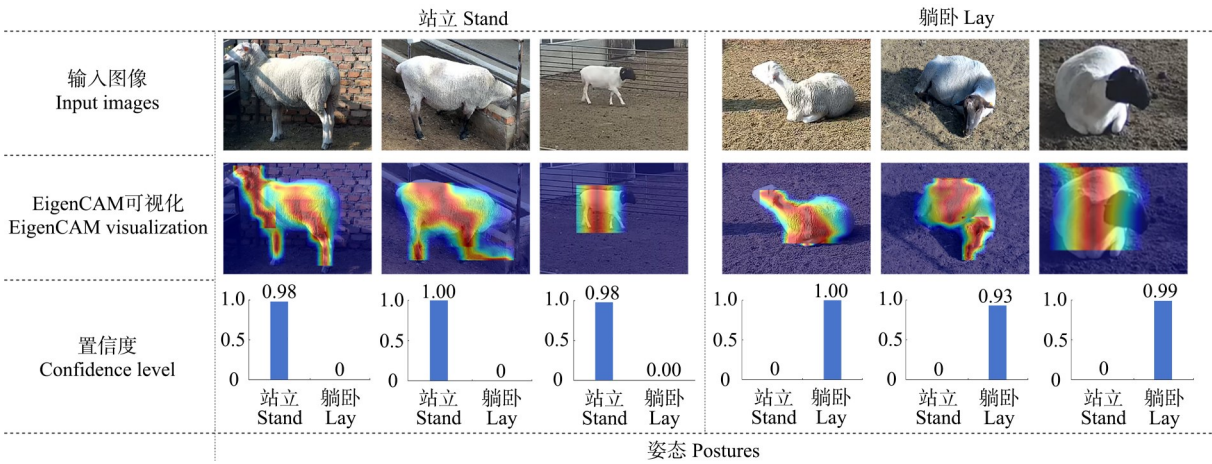


图9 绵羊站立和躺卧姿态的可视化结果

Fig. 9 Visualization results of the standing and lying postures of sheep

对绵羊站立姿态的可视化结果显示,在站立状态下不同姿态的绵羊,模型能够准确定位目标,有效过滤环境干扰,重点挖掘头部、躯干及四肢的深度特征。结合动物行为机理,这些部位是判别绵羊站立姿态的核心要素,而EigenCAM图中这些区域的高亮显示,直观验证了模型对关键特征的精准识别能力,同时置信度表明模型能够准确地识别出绵羊的站立姿态。对绵羊躺卧姿态的可视化结果可知,当绵羊在躺卧姿态下有角度和姿态差异时,模型也能够精准聚焦于与地面接触面积明显增加的头部和躯干部位,有效捕捉其深度特征,置信度表明模型能够准确识别绵羊的躺卧姿态。综上,研究表明RDS-Mask R-CNN在姿态检测任务中兼具实践可行性与理论可解释性。

3 讨论

尽管绵羊姿态检测已取得了良好进展^[11-13],但在绵羊品种多样、环境不同(涉及室内与室外)的养殖场中应用该技术,对模型的泛化能力与鲁棒性提出了更高要求。Deng等^[14]在其研究中指出,遮挡会导致被遮挡个体的检测结果不准确。本研究构建了涵盖多品种与不同环境的数据集,基于该数据集训练的RDS-Mask R-CNN模型在Recall(99.27%)与mAP(87.34%)上均表现优异,其与YOLOv3、Faster R-CNN经典模型的对比结果,进一步证实了模型在精度和鲁棒性上的优势,为解决类似的检测任务提供了可参考的思路。

针对养殖环境中的目标遮挡与姿态多变问题,

本研究在Mask R-CNN框架内集成了Res2Net101、DCN与Soft NMS。消融试验揭示了性能提升的内在原因:Res2Net101通过多尺度特征融合,显著增强了对绵羊整体轮廓及局部细节的捕捉能力,这种对多尺度信息的融合能力,是模型能够适应不同品种体型和复杂姿态的前提。DCN则通过自适应调整卷积核采样点,从根本上提升了对不规则形体及部分遮挡特征的提取灵活性,应对了姿态多变的问题。而Soft NMS采用的高斯衰减策略,通过柔和处理重叠边界框,有效缓解了遮挡场景中的漏检,从而在保持高召回率的同时,也避免了因误删目标重叠边界框而导致的精确率下降。

对不同姿态的识别结果发现躺卧姿态的Bbox AP(89.93%)与Segm AP(84.96%)均高于站立姿态的Bbox AP(84.74%)和Segm AP(78.05%)。这一差异主要源于2种姿态自身的差异性:躺卧姿态下绵羊躯体轮廓稳定、与地面接触面积大,特征显著且一致;而站立姿态因其四肢伸展角度与躯干空间关系复杂,其特征变异度更大,识别难度更高。EigenCAM可视化结果提供了直观证据:在躺卧姿态中,模型关注区域集中且稳定于躯干与头部;而在站立姿态中,模型需分散注意力以覆盖四肢。从养殖场景实际意义来看,高精度的躺卧识别可为“异常躺卧时长分析”功能提供可靠输入。

当前研究仍有一定的局限性:模型仅适用于绵羊场景,尚未实现跨物种姿态检测的通用模型;未系统应用数据增强技术,模型在极端光照、罕见视角等边界场景下的泛化能力可能受限。未来研究将围绕三方面开展:一是构建跨物种姿态识别的通用模型;二是结合自监督学习与数据增强提升模型在边界场景的泛化能力;三是融合RGB与热红外多模态数据,获得更丰富的特征,提高检测精度,最终实现高普适性与强鲁棒性的动物姿态检测技术。

4 结 论

本研究为实现绵羊姿态自动、精准检测,提出了基于RDS-Mask R-CNN的检测模型,主要结论如下:

(1)数据集构建:从内蒙古农业大学2个国家级示范养殖场,采集124只绵羊连续15 d、每天24 h视频数据,经预处理与标注,构建覆盖不同品系和养殖环境下的姿态样本数据集。

(2)算法框架设计:设计了基于RDS-Mask R-CNN的多目标检测与分割算法框架,用Res2Net101网络解决多尺度特征提取不足问题;引入可变形卷积模块提升对不规则形状的拟合能力;采用Soft NMS算法减少绵羊漏检问题,实现对重叠实例目标的精确分割。

(3)试验结果:RDS-Mask R-CNN模型的mAP达87.34%,比YOLOv3、Faster R-CNN和Mask R-CNN分别提高16.68%、8.64%和6.21%;不同改进策略的算法对比试验证实了改进的有效性;在绵羊不同姿态检测方面,模型表现优异,站立和躺卧姿态的Bbox AP分别为84.74%和89.93%,Segm AP分别为78.05%和84.96%,体现了模型对不同姿态的适应性;EigenCAM可视化结果表明,模型能自动聚焦姿态判别的关键区域。

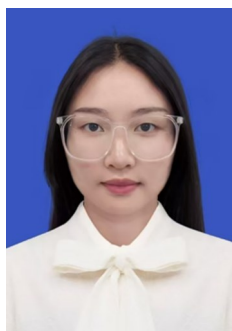
综上,本研究提出的RDS-Mask R-CNN模型能够在自然生产环境中自动且精准地识别多目标绵羊姿态,为实现绵羊的智慧养殖提供了技术支持。

参考文献 References

- [1] 邓小明,王文月,王小龙,葛毅强. 以科技创新引领支撑畜牧业现代化发展的实践思路[J]. 中国畜牧杂志, 2024, 60(3): 282-285
Deng X M, Wang W Y, Wang X L, Ge Y Q. Practical thinking of supporting the modernization development of animal husbandry with scientific and technological innovation[J]. *Chinese Journal of Animal Science*, 2024, 60(3): 282-285 (in Chinese)
- [2] 常倩,王士权,李秉龙. 畜牧业纵向协作特征及其影响因素分析:来自内蒙古养羊户的经验证据[J]. 中国农业大学学报, 2016, 21(7): 152-160
Chang Q, Wang S Q, Li B L. The characteristics and influencing factors of vertical coordination in animal husbandry: An evidence from sheep and goat farmers in Inner Mongolia[J]. *Journal of China Agricultural University*, 2016, 21(7): 152-160 (in Chinese)
- [3] Vázquez Diosdado J A, Barker Z E, Hodges H R, Amory J R, Croft D P, Bell N J, Codling E A. Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system[J]. *Animal Biotelemetry*, 2015, 3(1): 15
- [4] Tran D N, Nguyen T N, Khanh P C P, Tran D T. An IoT-based design using accelerometers in animal behavior recognition systems[J]. *IEEE Sensors Journal*, 2022, 22(18): 17515-17528
- [5] 张玉,齐景伟,金晓,斯日古楞,高静,曹琪娜,阿丽玛,曹晓波. 信息技术在放牧绵羊福利中应用的研究进展[J]. 家畜生态学报, 2018, 39(3): 8-11
Zhang Y, Qi J W, Jin X, Si R, Gao J, Cao Q N, A L M, Cao X B. Research progress on utilization of information technology in grazing sheep welfare[J]. *Journal of Domestic Animal Ecology*, 2018, 39(3): 8-11 (in Chinese)
- [6] Hu J Y, Fei R R, Du F, Chang P J, Zhang J S. Convolutional neural network incorporating misclassification information for image recognition[J]. *Soft Computing*, 2024, 28(2): 1009-1021
- [7] Jiao L, Dong S F, Zhang S Y, Xie C J, Wang H Q. AF-RCNN: An anchor-free convolutional neural network for multi-categories agricultural pest detection[J]. *Computers and Electronics in Agriculture*, 2020, 174: 105522

- [8] Hansen M F, Smith M L, Smith L N, Salter M G, Baxter E M, Farish M, Grieve B. Towards on-farm pig face recognition using convolutional neural networks[J]. *Computers in Industry*, 2018, 98: 145-152
- [9] 刘璠瑛, 曹珏, 郭彬彬, 陈慧杰, 戴子淳, 龚长万. 基于改进YOLO v5复杂场景下肉鹅姿态的检测算法研究[J]. *南京农业大学学报*, 2023, 46(3): 606-614
Liu Y Y, Cao X, Guo B B, Chen H J, Dai Z C, Gong C W. Research on detection algorithm about the posture of meat goose in complex scene based on improved YOLO v5[J]. *Journal of Nanjing Agricultural University*, 2023, 46(3): 606-614 (in Chinese)
- [10] 涂淑琴, 刘晓龙, 梁云, 张宇, 黄磊, 汤寅杰. 基于改进DeepSORT的群养生猪行为识别与跟踪方法[J]. *农业机械学报*, 2022, 53(8): 345-352
Tu S Q, Liu X L, Liang Y, Zhang Y, Huang L, Tang Y J. Behavior recognition and tracking method of group-housed pigs based on improved DeepSORT algorithm [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(8): 345-352 (in Chinese)
- [11] Yu L H, Pu Y H, Cen H L, Li J B, Liu S Y, Nie J, Ge J B, Lv L Z, Li Y L, Xu Y L, Guo J J, Zhao H X, Wang K. A lightweight neural network-based method for detecting estrus behavior in ewes[J]. *Agriculture*, 2022, 12(8): 1207
- [12] Jiang M, Rao Y, Zhang J Y, Shen Y M. Automatic behavior recognition of group-housed goats using deep learning[J]. *Computers and Electronics in Agriculture*, 2020, 177: 105706
- [13] Gu Z S, Zhang H Y, He Z Q, Niu K. A two-stage recognition method based on deep learning for sheep behavior[J]. *Computers and Electronics in Agriculture*, 2023, 212: 108143
- [14] Deng X F, Yan X L, Hou Y M, Wu H, Feng C R, Chen L Y, Bi M X, Shao Y. Detection of behaviour and posture of sheep based on YOLOv3[J]. *INMATEH Agricultural Engineering*, 2021: 457-466
- [15] Torralba A, Russell B C, Yuen J. LabelMe: Online image annotation and applications[J]. *Proceedings of the IEEE*, 2010, 98(8): 1467-1484
- [16] Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick C L. Microsoft COCO: Common objects in context [C]. In: *Computer Vision - ECCV 2014*. Cham: Springer, 2014: 740-755
- [17] 杨巧宁, 蒋思, 纪晓东, 杨秀慧. 基于多尺度特征提取的单目图像深度估计[J]. *北京化工大学学报:自然科学版*, 2023, 50(1): 97-106
Yang Q N, Jiang S, Ji X D, Yang X H. Monocular image depth estimation based on multi-scale feature extraction[J]. *Journal of Beijing University of Chemical Technology : Natural Science Edition*, 2023, 50(1): 97-106 (in Chinese)
- [18] Li C J, Qu Z, Wang S Y. PerspectiveNet: An object detection method with adaptive perspective box network based on density-aware [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(5): 5419-5429
- [19] Gao S H, Cheng M M, Zhao K, Zhang X Y, Yang M H, Torr P. Res2Net: A new multi-scale backbone architecture[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(2): 652-662
- [20] Zhu X Z, Hu H, Lin S, Dai J F. Deformable ConvNets V2: More deformable, better results[C]. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 15-20, 2019, Long Beach, CA, USA. IEEE, 2020: 9300-9308
- [21] Yang H B, Liu P, Hu Y Z, Fu J N. Research on underwater object recognition based on YOLOv3[J]. *Microsystem Technologies*, 2021, 27(4): 1837-1844
- [22] 刘啸虎, 肖德琴, 刘又夫, 王春桃, 周敏. 基于Faster R-CNN和时序统计的肉鸭行为节律分析[J]. *中国家禽*, 2023, 45(11): 95-104
Liu X H, Xiao D Q, Liu Y F, Wang C T, Zhou M. Analysis on rhythmic behavior of meat ducks based on Faster R-CNN and time-series statistics[J]. *China Poultry*, 2023, 45(11): 95-104 (in Chinese)

责任编辑: 秦梅



第一作者简介: 甘霖惠, 硕士研究生, 现就读于内蒙古农业大学计算机与信息工程学院计算机科学与技术专业, 主要研究领域为数字技术在畜牧业中的应用研发, 在校期间获得内蒙古农业大学校级二等奖学金。



通讯作者简介: 王步钺, 教授, 研究生导师。中国计算机学会杰出会员, 中国计算机学会信息系统执行委员, 中国计算机学会数字农业分会执行委员, 智慧畜牧自治区高校重点实验室副主任, 主要从事数字技术在畜牧应用领域的研发工作。