

基于坐标注意力机制和 YOLO v5s 模型的山羊脸部检测方法

郭阳阳^{1,2} 洪文浩¹ 丁屹¹ 黄小平^{1,2}

(1. 安徽大学互联网学院, 合肥 230039;

2. 农业生态大数据分析与应用技术国家地方联合工程研究中心, 合肥 230039)

摘要: 山羊的脸部检测对羊场的智能化管理有着重要的意义。针对实际饲养环境中,羊群存在多角度、分布随机、灵活多变、羊脸检测难度大的问题,以 YOLO v5s 为基础目标检测网络,提出了一种结合坐标信息的山羊脸部检测模型。首先,通过移动设备获取舍内、舍外、单头以及多头山羊的图像并构建数据集。其次,在 YOLO v5s 的主干网络融入坐标注意力机制,以充分利用目标的位置信息,提高遮挡区域、小目标、多视角样本的检测精度。试验结果表明,改进 YOLO v5s 模型的检测精确率为 95.6%,召回率为 83.0%,mAP_{0.5} 为 90.2%,帧速率为 69 f/s,模型内存占用量为 13.2 MB;与 YOLO v5s 模型相比,检测精度提高 1.3 个百分点,模型所占内存空间减少 1.2 MB;且模型的整体性能远优于 Faster R-CNN、YOLO v4、YOLO v5s 模型。此外,本文构建了不同光照和相机抖动的数据集,来进一步验证本文方法的可行性。改进后的模型可快速有效地对复杂场景下山羊的脸部进行精准检测及定位,为动物精细化养殖时目标检测识别提供了检测思路和技术支持。

关键词: 羊脸检测; YOLO v5s; 坐标注意力机制; 精准畜牧

中图分类号: TP391.4 文献标识码: A 文章编号: 1000-1298(2023)07-0313-09

OSID:



Goat Face Detection Method by Combining Coordinate Attention Mechanism and YOLO v5s Model

GUO Yangyang^{1,2} HONG Wenhao¹ DING Yi¹ HUANG Xiaoping^{1,2}

(1. School of Internet, Anhui University, Hefei 230039, China

2. National Engineering Research Center for Agro-Ecological Big Data Analysis and Application, Hefei 230039, China)

Abstract: Animal face detection is of great significance to the intelligent management of animal farm. At present, goats have the characteristics of multi angle, random distribution and flexibility in the actual feeding environment, which greatly increases the difficulty of goat face detection. Therefore, a goat face detection model combined with coordinate information was proposed based on YOLO v5s target detection network. Firstly, indoor, outdoor, single and multiple goat images were obtained by using mobile devices to build sample data sets. Secondly, coordinate attention mechanism (CA) was integrated into the backbone network of YOLO v5s to make full use of target position information and improve the target recognition accuracy in the occluded area, small target and multi view sample images. The proposed YOLO v5s - CA based approach achieved a precision of 95.6%, a recall of 83.0%, an mAP_{0.5} of 90.2%, a frame rate of 69 f/s and a model size of 13.2 MB. Compared with that of the original YOLO v5s model, the detection precision of YOLO v5s - CA was increased by 1.3 percentage points, and the memory space was reduced by 1.2 MB. And the overall performance of the YOLO v5s - CA was better than that of the Faster R-CNN, YOLO v4 and YOLO v5s. Experimental results showed that the proposed YOLO v5s - CA approach can improve the detection precision of occluding and small targets by introducing target coordinate information. In addition, datasets with different lighting and camera shake were simulated and constructed to further verify the feasibility of the proposed method. Overall, the proposed deep learning-based goat face detection approach can quickly and effectively detect and locate goat faces in complex scenes, providing detection ideas and technical support for target detection and recognition in intelligent animal farm.

Key words: goat face detection; YOLO v5s; coordinate attention mechanism; precision animal farm

收稿日期: 2022-11-13 修回日期: 2023-02-07

基金项目: 安徽省自然科学基金项目(1908085QF284)和安徽省教育厅自然科学基金项目(KJ2021A0024)

作者简介: 郭阳阳(1991—),男,讲师,博士,主要从事畜禽智能化监测技术研究,E-mail: guoyangyang113529@ahu.edu.cn

通信作者: 黄小平(1984—),男,讲师,博士,主要从事畜牧养殖信息化研究,E-mail: hxpjng@mail.ustc.edu.cn

0 引言

随着居民生活水平的不断提高和消费主体的不断转变,羊肉的需求也日益增加,但当下肉羊养殖方式较为传统,产业化进程较缓慢^[1]。现代化农业的发展,促进了畜牧业向规模化、智能化、现代化方向发展,而信息化是重要标志^[2],其中动物的目标区域精准检测对牧场信息化具有重要意义^[3-4]。在目标检测基础上可进一步挖掘动物的深层次信息(体况、行为、疾病等),对实现动物生长信息和健康状况监测提供了数据支撑。

传统的动物检测方法常通过电子耳标、项圈等集成传感器设备给定动物标号来实现动物个体检测与识别。但随着时间延长,传感器设备会出现脱落、标签模糊等现象,从而造成目标信息丢失、误检测等情况^[5],此外,长时间的佩戴会导致牲畜感染或患上某些疾病,影响动物福利^[6-7]。而计算机视觉作为一种非接触式无创伤的目标检测方法在智慧畜牧业中得到广泛应用。计算机视觉技术主要从图像或视频中提取生物视觉相关特征(例如,形状、颜色、纹理等),在此基础上通过机器学习等算法实现动物区域检测,但检测效果依赖于相关特征提取的是否充分^[8-10]。

深度学习技术的发展,使其特征的提取能力提高,有效解决外界因素带来的干扰,满足现代化畜牧业的需求^[11-12]。目前,深度学习技术已被应用到动物目标检测中,其中基于面部特征的目标检测是热门研究点之一^[13-14]。秦兴等^[15]在双线性卷积神经网络(Bilinear-CNN)的基础上提出了一种非侵入式的猪面部检测识别模型,猪面部检测识别准确率达到95.73%。XU等^[16]提出了一种孪生密集胶囊网络(Siamese DB capsule network, SDBCN)对牛脸图像进行了检测,检测准确率达到93%。张宏鸣等^[17]将通道注意力算法(Efficient channel coalesce spatial information attention, ECCSA)融入到MobileFaceNet的深度卷积层和残差层中,并采用余

弦退火进行动态学习率调优,构建了羊脸检测识别模型(ECCSA-MFC)。试验结果表明,该方法在人工筛选的数据集中准确率达到96.73%。SONG等^[18]提出了一种基于YOLO v3模型修剪的羊脸检测方法(YOLO v3-P)。通过使用羊脸数据集对YOLO v3中的锚框进行聚类,mAP从95.3%增加到96.4%,模型尺寸也缩小到原始模型尺寸的1/4。上述研究表明,高质量面部数据集对面部检测识别模型的性能提升有至关重要的作用,但并没有提升模型在更加复杂多变的现实环境中的检测性能。此外,羊只通常遍体纯色,羊头灵活,给脸部检测识别增加了一定难度;羊养殖环境较为复杂,遮挡、重叠、光线干扰等会造成漏检、误检问题的发生,在保证所构建模型检测准确率的前提下,也应满足牧场实时性的需求。

本文以黄淮山羊为研究对象,构建具有多样性的数据集(包含舍内、舍外、单头以及多头的山羊图像)。基于YOLO v5s检测模型,提出一种引入坐标注意机制的山羊面部检测识别方法。即在YOLO v5s模型的主干网络中引入坐标注意力机制,以充分利用山羊脸部的位置信息,从而提高遮挡、小目标、多视角样本的识别准确率和快速性,改善整体模型的性能。为进一步验证本研究方法的鲁棒性和应用性,构建模拟相机抖动和不同光照(白天、夜间)下的数据集,并对模型进行测试。

1 材料与方法

1.1 羊脸数据采集

山羊图像采集于安徽省合肥博大牧业科技开发有限公司。在实际饲养环境下,通过手机拍摄并获取自然状态下的黄淮山羊图像。每幅图像中包含1~20只山羊不等。图像格式为jpg格式。数据集中包含不同角度、不同场景(舍内、舍外)、遮挡、单头以及多头等纯白色山羊,以充分模拟羊场的复杂性。采集的部分山羊图像如图1所示。



图1 采集的山羊图像

Fig.1 Collected images of goats

1.2 数据集构建

为了检测本研究采用模型的泛化性,从拍摄的图像中选取 400 幅作为样本数据集,其中存在遮挡且多头羊只的样本量约为 300。数据集按 4:1 分配为训练集与测试集。使用 Labelme 图像标注工具对羊脸进行标记,并将得到的 json 文件转化为 txt 形式,制作 YOLO v5s 模型所需格式数据集。YOLO v5s 的作者在其源代码中添加了大量的数据增强策略,本研究基

于 Mosaic 数据增强方式^[19],选择图像色度、饱和度、亮度随机调整,以及图像翻转、错切、缩放、平移等数据增强方法,并将由 4 幅随机图像按照一定的概率转换成单幅图像作为每次的输入图像,从而扩大数据的多样性。最终样本数据集扩充为 4 000 幅。从图 2 中可知,与单幅图像的增强变换结果相比(图 2c ~ 2i),Mosaic 图像增强(图 2b)不仅扩充了样本数量,也进一步丰富了多场景多尺度信息。



图 2 增强图像示例

Fig. 2 Example of augmented images

2 羊脸检测识别方法

2.1 YOLO v5s 网络结构

YOLO v5 属于轻量级模型,运行速度快,可满足牧场实时准确的需求,且模型内存占用量小,保证了该模型可移植到移动设备中,应用更加便携灵活。YOLO v5 具有 YOLO v5s、YOLO v5m、YOLO v5l、YOLO v5x 共 4 个模型,网络宽度和深度依次增大。本研究选择 YOLO v5s 作为基础模型。

YOLO v5s 网络架构由 Input、Backbone、Neck、Output 组成,各部分具体组成如下。

Input:采用 Mosaic 数据增强方式^[19],对图像进行缩放、翻转、马赛克、色彩变化等操作,并自动计算数据集的最佳锚框值。

Backbone:由 CSPNet^[20] 模块和 SPPF^[21] 模块组成。CSPNet 将基础层的特征映射成两部分,并通过局部跨层结构融合,获取不同层更丰富的特征图。SPPF 模块进行最大池化操作并进行张量拼接。

Neck:由 PANet^[22] 模块组成。PANet 自顶向下将高层特征图中的语义信息与不同层 CSP 模块的输出特征进行有效聚合,再采用自底向上路径聚合

结构聚合浅层特征。

Output:采用 GIOU_Loss^[23] 为损失函数,增加了相交尺度的衡量,采用非最大抑制算法去除损失函数中贡献值较小的预测结果,生成边界框和预测类别。

2.2 引入坐标注意力机制的 YOLO v5 检测模型

针对实际饲养环境下羊只多角度、多尺度等特点,且为了进一步在山羊图像中精确地检测出脸部区域,本研究在 YOLO v5s 的基础上引入坐标注意力机制(Coordinate attention,CA),即 YOLO v5s-CA 模型,使模型关注于目标区域位置信息,以提高对遮挡和小目标区域的检测精度。改进后的山羊脸部检测网络结构图如图 3 所示,在主干网络每个 C3 模块后加入坐标注意力机制模块。

注意力机制作为提升卷积神经网络模型性能的一种有效方法,得到了广泛的应用^[24-26]。其中,坐标注意力(Coordinate attention,CA)^[27] 是一种新的高效注意力机制,其结构图如图 4 所示。其通过精确的位置信息对通道关系和长期依赖性进行编码,第 1 步将坐标信息嵌入通道注意力中,对输入 X ,采用尺寸 $(H,1)$ 和 $(1,W)$ 的池化核沿着水平坐标方向

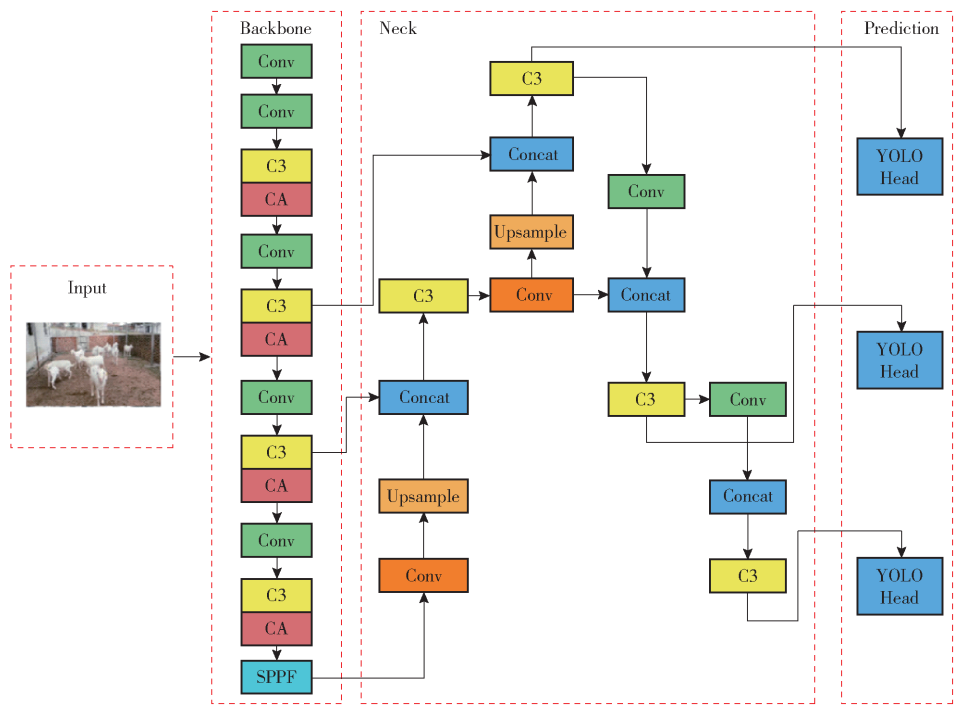


图 3 改进的 YOLO v5s 结构
Fig. 3 Improved YOLO v5s structure

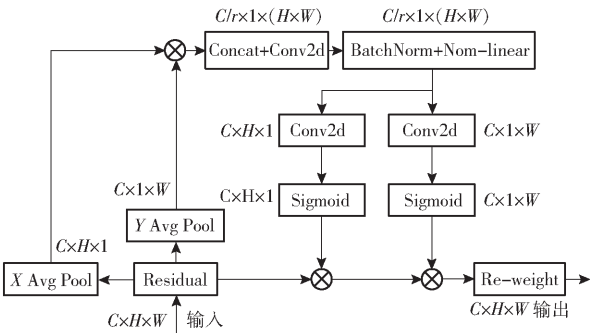


图 4 坐标注意力模块结构
Fig. 4 CA model structure

和竖直坐标方向对每个通道进行编码, H 、 W 分别为输入特征的高度和宽度,因此,高度为 h 的第 c 个通道的输出表达式为

$$z_c^h(h) = \frac{1}{W} \sum_{i=0}^{W-1} x_c(h,i) \tag{1}$$

式中 $z_c^h(h)$ ——高度为 h 的第 c 个通道的输出
 $x_c(h,i)$ ——第 i 行特征向量

同理,宽度为 w 的第 c 个通道的输出表达式为

$$z_c^w(w) = \frac{1}{H} \sum_{j=0}^{H-1} x_c(j,w) \tag{2}$$

式中 $z_c^w(w)$ ——宽度为 w 的第 c 个通道的输出
 $x_c(j,w)$ ——第 j 列特征向量

利用一维全局池化操作将 2 个空间方向进行特征聚合,使得注意力模块捕捉到沿着一个空间方向的相关关系,并保存沿着另一个空间方向的位置信息,这有助于网络更准确地获取感兴趣的目标。第 2 步生成坐标注意力,首先将两个空间方向的输

出张量进行拼接,使用 1×1 卷积进行 F_1 变换操作,表达式为

$$f = \delta(F_1([z^h, z^w])) \tag{3}$$

式中 f ——在水平方向和竖直方向上编码空间信息的中间特征图

δ ——非线性激活函数

F_1 —— 1×1 卷积变换

z^h 、 z^w ——编码输出

然后将 f 切分为 2 个单独的张量,再分别利用 1×1 卷积变换成 X 相同的通道数,最后利用 Sigmoid 激活函数得到注意力权重矩阵为

$$g^h = \sigma(F_h(f^h)) \tag{4}$$

$$g^w = \sigma(F_w(f^w)) \tag{5}$$

式中 f^h 、 f^w —— f 在水平方向和竖直方向的切分张量
 F_h 、 F_w —— 1×1 卷积变换
 g^h 、 g^w ——水平、竖直方向的注意力权重矩阵
 σ ——Sigmoid 激活函数

最终输出量 $y_c(i,j)$ 为

$$y_c(i,j) = x_c(i,j) g_c^h(i) g_c^w(j) \tag{6}$$

式中 $x_c(i,j)$ ——输入特征图
 $g_c^h(i)$ 、 $g_c^w(j)$ ——水平方向和竖直方向的注意力图

最后获取的 $y_c(i,j)$ 增强了目标位置信息的敏感度,通过自顶向下将改进后获取的各层 $y_c(i,j)$ 进行聚合,采用非最大抑制算法去除贡献值较小的预测结果,生成边界框和预测类别,实现羊脸区域检测。

3 试验平台及模型评价指标

3.1 试验平台与对比模型

本研究试验平台配置:操作系统为 Windows 10; CPU 为 Intel (R) Core (TM) i5 - 10200H 2.4 GHz; GPU 为 NVIDIA GeForce RTX 1650 Ti; CPU 显存为 4 GB; 内存为 16 GB; 加速环境为 CUDAv10.2; Pytorch 版本为 1.9.1; Python 版本为 3.9.0。

为了验证 YOLO v5s - CA 模型的检测速率和轻量化,将 YOLO v5s 和 YOLO v5s - CA 主干网络替换成轻量级网络 MobileNet v3,即构建 YOLO v5s - MobileNet v3 和 YOLO v5s - MobileNet v3 - CA 模型,并与 YOLO v5 - CA 模型进行对比分析。此外,为了验证本研究方法的整体性能,选择 Faster R - CNN、YOLO v4 和 YOLO v5s 模型作为对比模型。

3.2 评价指标

采用 $mAP_{0.5}$ 、 $mAP_{0.5:0.95}$ 、精确率、召回率、帧速率和内存占用量作为模型评价指标,精确率 (Precision, P) 表示预测正样本中为正样本的比例,召回率 (Recall, R) 表示所有正样本中预测正确的

比例。 $mAP_{0.5}$ 表示交并比 (IOU) 为 0.5 时平均精度均值, $mAP_{0.5:0.95}$ 表示 IOU 从 0.5 到 0.95、间隔为 0.05 时平均精度均值。帧速率表示目标网络每秒可以检测的图像数。

4 结果与讨论

4.1 不同模型检测速度与轻量化结果对比

在构建的数据集基础上,将 YOLO v5s - CA 与 YOLO v5s、YOLO v5s - MobileNet v3 和 YOLO v5s - MobileNet v3 - CA 模型进行对比研究。由表 1 可知,将主干网络替换成轻量级网络 MobileNet v3 后,模型内存占用量明显下降,检测速率有所提升,但检测精度低于原模型 (YOLO v5s 94.3%, YOLO v5s - CA 95.6%)。YOLO v5s - CA 的帧速率为 69 f/s,模型内存占用量为 13.2 MB,可满足轻量化实时性检测的需求。此外,在 YOLO v5s 主干网络融入坐标注意力机制 (CA) 后,较 YOLO v5s 模型的准确率提升 1.3 个百分点,召回率下降 1.1 个百分点, $mAP_{0.5}$ 提升 0.8 个百分点,帧速率下降 8 f/s。可见, YOLO v5s - CA 模型融合坐标注意力机制牺牲了较小的帧速率但提升了检测精度。

表 1 轻量级网络对比
Tab. 1 Comparison of lightweight networks

模型	精确率/ %	召回率/ %	$mAP_{0.5}$ / %	$mAP_{0.5:0.95}$ / %	帧速率/ ($f \cdot s^{-1}$)	内存占用量/ MB
YOLO v5s	94.3	84.1	89.4	55.3	77	14.4
YOLO v5s - MobileNet v3	86.3	73.8	79.5	41.3	84	7.4
YOLO v5s - MobileNet v3 - CA	77.9	60.6	67.2	33.4	78	7.3
YOLO v5s - CA	95.6	83.0	90.2	52.5	69	13.2

以 YOLO v5s 为基础网络的模型采用 Mosaic 数据增强方式,图像输入尺寸为 640 像素 $\times$ 640 像素,批量大小为 8,最大迭代次数为 300,使用随机梯度下降 (Stochastic gradient descent, SGD) 作为优化器,权重衰减为 0.000 5,动量为 0.937,初始学习率设置为 0.01,训练过程中采用余弦退火算法进行学习率更新。训练过程中各模型验证集损失值及 $mAP_{0.5}$ 变化如图 5 所示。

可以看出, YOLO v5s - CA 与 YOLO v5s 基础网络收敛速度相近,且二者收敛速度明显优于将主干网络替换成 MobileNet v3 的模型,在 $mAP_{0.5}$ 方面, YOLO v5s - CA 最终超过了 YOLO v5s,表现最优,约 100 个迭代周期后趋于平稳,而基于 MobileNet v3 模型的效果明显不佳。综上可知,轻量级网络 MobileNet v3 对于模型性能提升效果不大,在帧速率略高的情况下,检测精度反而降低,由此可见,

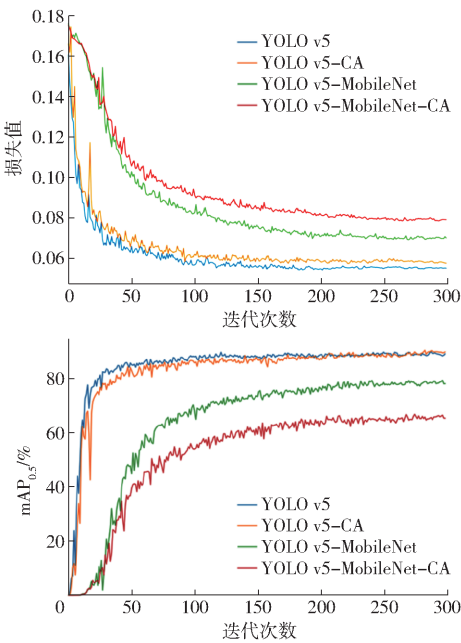


图 5 训练过程中损失值和 $mAP_{0.5}$ 变化曲线

Fig. 5 Changes in loss and $mAP_{0.5}$ during training

YOLO v5s-CA 的整体性能在检测精度和检测速度方面均取得了不错的效果。

4.2 不同模型对山羊脸部检测结果的对比

为了验证本文方法的有效性,与 Faster R-CNN、YOLO v4 和 YOLO v5s 模型在测试数据集上进行了检测对比分析,结果如表 2 所示。由表 2 可知,YOLO v5s-CA 与 Faster R-CNN 相比,其检测精确率提高 17.3 个百分点,召回率提高 0.1 个百分点,mAP_{0.5}提高 5.4 个百分点,帧速率是 Faster R-CNN 的 8.63 倍;与 YOLO v4 相比,其检测精确率提高 27.5 个百分点,召回率提高 10.7 个百分点,mAP_{0.5}提高 21.0 个百分点,帧速率是 YOLO v4 的 2.03 倍。可以看出,YOLO v5s-CA 相比 Faster R-CNN 和

YOLO v4,其检测精确率及性能均更优,且在检测速度上有明显的提升。此外,YOLO v5s-CA 和 YOLO v5s 均取得了较优的检测结果,本研究对 YOLO v5s-CA 和 YOLO v5s 的检测性能结合检测结果图像进一步对比分析,检测结果如图 6 所示。

表 2 不同模型检测结果
Tab.2 Different model experiment results

模型	精确率/	召回率/	mAP _{0.5} /	帧速率/	内存占用
	%	%	%	(f·s ⁻¹)	量/MB
Faster R-CNN	78.3	82.9	84.8	8	108.2
YOLO v4	68.1	72.3	69.2	34	243.9
YOLO v5s	94.3	84.1	89.4	77	14.4
YOLO v5s-CA	95.6	83.0	90.2	69	13.2

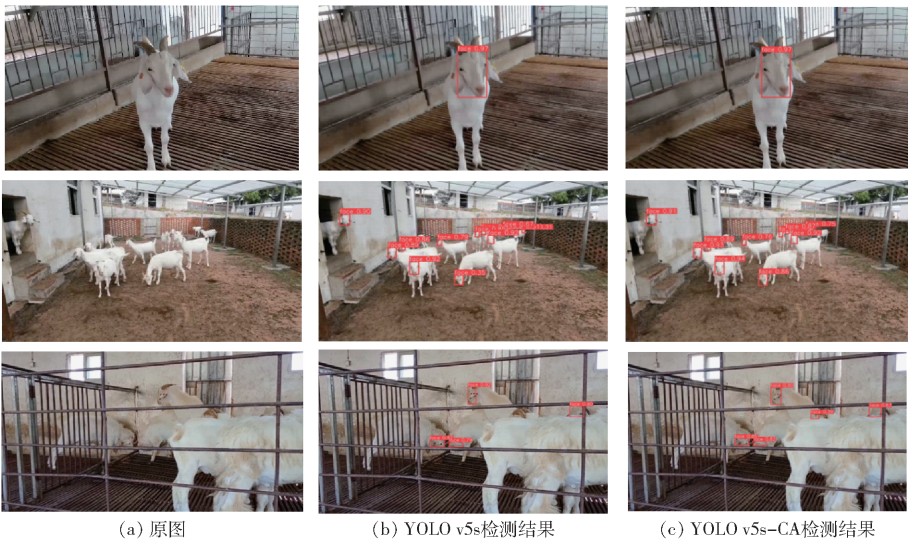


图 6 YOLO v5s 和 YOLO v5s-CA 的检测结果对比

Fig.6 Comparison of detection effects between YOLO v5s and YOLO v5s-CA

从图 6 中可以看出,YOLO v5s 和 YOLO v5s-CA 模型在单只羊的脸部检测效果均较好,没有较大差异。而对于多只羊图像中的羊脸检测,由于图像中存在头部区域重叠和小目标,YOLO v5s 模型会重复检测或漏检;YOLO v5s-CA 融入坐标注意力机制,更好地利用了坐标信息从而更精确地获取感兴趣的特征,在检测多目标重叠或小目标情况下效果有所提升。

4.3 不同光照和模糊化下山羊脸部检测结果对比

为了适应不同光照以及摄像机抖动带来的检测干扰,选择 100 幅图像分别进行 Gamma ($\gamma = 0.5 \sim 0.7$; $\gamma = 1.5 \sim 3.5$) 亮度调整和采用 PIL 库中的 BLUP 进行模糊处理,构建数据集来模拟不同光照和相机抖动的情况。在光照增强、光照减弱、模糊 3 种数据集上,对已构建的 YOLO v5s-CA 模型进行可行性分析,结果如表 3 所示。

从表 3 中可以看出,在光照增强和减弱情况下,YOLO v5s-CA 的检测精度略高于 YOLO v5s,但在

表 3 不同光照和模糊条件下的检测结果

Tab.3 Different lighting and blur conditions

experiment results				%
模型	处理条件	精确率	召回率	mAP _{0.5}
YOLO v5s	光照增强	91.6	83.1	88.2
	光照减弱	92.3	83.6	89.0
	模糊化	88.7	78.1	83.3
YOLO v5s-CA	光照增强	92.9	82.8	88.4
	光照减弱	93.0	83.8	88.7
	模糊化	87.1	74.5	81.5

模糊化中略低于 YOLO v5s。模糊化后可能导致目标区域边界不明显,从而获取的位置信息略有偏差,但目前硬设备已较好地解决了抖动问题。图 7 为不同情况下的检测结果。可以看出,对于单只山羊,模型均得到了较好的检测结果。而对于多只山羊,在不同光照情况下,YOLO v5s 会出现漏检而 YOLO v5s-CA 的检测结果有所改善。在模糊化情况下,两个模型检测结果相差不大。

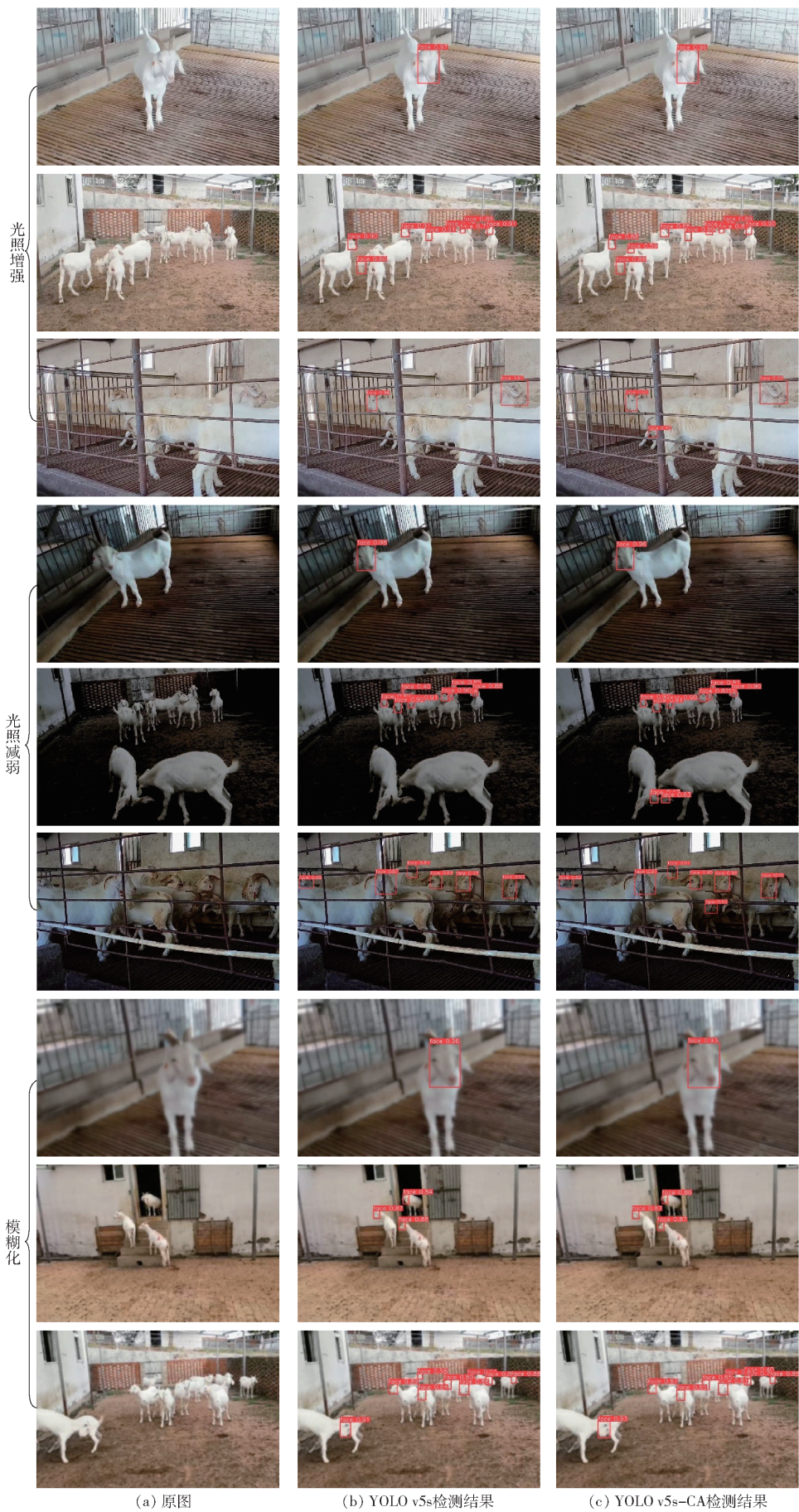


图 7 YOLO v5s 和 YOLO v5s-CA 在不同光照和模糊条件下检测结果对比

Fig. 7 Comparison of detection effects between YOLO v5s and YOLO v5s-CA under different lighting and blur conditions

4.4 讨论

4.4.1 数据集对检测结果的影响

模型检测效果主要依赖于脸部特征,故脸部特征的有效获取直接影响检测的性能。YOLO v5s-CA 在引入坐标注意力机制基础上更加充分地获取脸部区域特征。数据集中包含多角度、多尺度以及多目标等纯白色山羊图像。纯白色山羊不具有明显的花纹和斑块等特征,给山羊脸部检测增加了一定的难度。羊头部存在多角度的现象,当头部扭转、低头造成的头部特别是脸部特征的大量丢失,该情况下不能很好地检测出目标区域;此外,遮挡(外物遮挡、种间遮挡)造成脸部信息的丢失时检测效果较差。但相对于 YOLO v5s 模型,YOLO v5s-CA 在该数据上的检测效果有所提升。

4.4.2 模型应用性分析与展望

本研究中的 YOLO v5s-CA 模型内存占用量为 13.2 MB,帧速率为 69 f/s,检测精确率为 95.6%,故可移植到移动平台实现便携灵活操作。此外,在检测到羊群的脸部区域的基础上,可以作为个数统计的辅助手段;其次,羊群中存在攻击、爬跨等社会行为,这也是生产者比较关注的行为。攻击行为存在安全隐患;爬跨行为对应发情行为,这对于生产者都是重要的信息。而这些行为可通过脸部区域与身体区域的相对位置关系来进行判断,例如攻击行为(图 8 黄色框对应攻击行为):羊头低下准备攻击,头部位置相对较低;爬跨行为(图 8 红色框对应爬跨行为):头部区域远高于身体区域。下一步可进一步检测获取身体区域来实现社会行为的识别,但对于密集性羊群、多角度、多尺度的数据集中,关联同一只羊的脸部区域与身体区域仍具有较大的挑战。

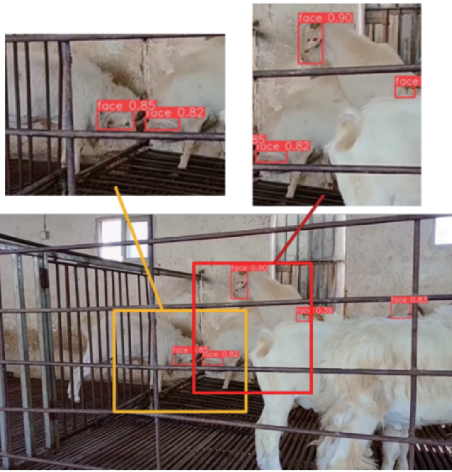


图 8 羊群社会行为示例
Fig.8 Example of social behaviors

5 结论

- (1)在 YOLO v5s 的主干网络中加入坐标注意力机制,充分获取目标区的位置信息,提高了对重叠区域和小目标区域的检测性能。改进 YOLO v5s 模型的检测精度为 95.6%,召回率为 83.0%,mAP_{0.5}为 90.2%,帧速率为 69 f/s,模型内存占用量为 13.2 MB。
- (2)将 YOLO v5s 中的主干网络替换为 MobileNet v3 进行轻量级网络对比分析,结果表明轻量级网络对于模型的整体性能提升不大,其检测速度有所提升但检测精度明显下降,而 YOLO v5s-CA 在检测精确率和检测速度上均较优。
- (3)改进 YOLO v5s 模型的检测精确率分别比 Faster R-CNN、YOLO v4 和 YOLO v5s 高 17.3、27.5、1.3 个百分点,可以有效改善漏检、误检问题。但在羊群过于拥挤、头部区域重叠严重、头部扭转等造成脸部信息大量丢失时,检测效果较差,仍有待提高。

参 考 文 献

[1] 丁丽娜. 中国羊肉市场供求现状及未来趋势研究[D]. 北京:中国农业大学,2014.
DING Li'na. Research on the current situation and future trend of mutton supply and demand in China[D]. Beijing: China Agricultural University, 2014. (in Chinese)

[2] 滕光辉. 畜禽设施精细养殖中信息感知与环境调控综述[J]. 智慧农业, 2019, 1(3): 1-12.
TENG Guanghui. Information sensing and environment control of precision facility livestock and poultry farming[J]. Smart Agriculture, 2019, 1(3): 1-12. (in Chinese)

[3] 何东健, 刘冬, 赵凯旋. 精准畜牧业中动物信息智能感知与行为检测研究进展[J]. 农业机械学报, 2016, 47(5): 231-244.
HE Dongjian, LIU Dong, ZHAO Kaixuan. Review of perceiving animal information and behavior in precision livestock farming [J]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(5): 231-244. (in Chinese)

[4] 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.

[5] BABOT D, HERNÁNDEZ-JOVER M, CAJA G, et al. Comparison of visual and electronic identification devices in pigs: on-farm performances[J]. Journal of Animal Science, 2006, 84(9): 2575-2581.

[6] FOSGATE G T, ADESIYUN A A, HIRD D W. Ear-tag retention and identification methods for extensively managed water buffalo (*Bubalus bubalis*) in Trinidad[J]. Preventive Veterinary Medicine, 2006, 73(4): 287-296.

- [7] CARNÉ S, CAJA G, ROJAS-OLIVARES M A, et al. Readability of visual and electronic leg tags versus rumen boluses and electronic ear tags for the permanent identification of dairy goats[J]. *Journal of Dairy Science*, 2010, 93(11): 5157–5166.
- [8] ANDREW W, GREATWOOD C, BURGHARDT T. Visual localisation and individual identification of holstein friesian cattle via deep learning[C]//*Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017: 2850–2859.
- [9] KUMAR S, SINGH S K, SINGH R, et al. Recognition of cattle using face images[J/OL]. *Animal Biometrics*, 2017: 79–110. DOI:10.1007/978-981-10-7956-6_3.
- [10] GUO Y, AGGREY S E, OLADEINDE A, et al. A machine vision-based method optimized for restoring broiler chicken images occluded by feeding and drinking equipment[J]. *Animals*, 2021, 11(1): 123.
- [11] QIAO Y, SU D, KONG H, et al. Individual cattle identification using a deep learning based framework[J]. *IFAC-PapersOnLine*, 2019, 52(30): 318–323.
- [12] LI G, HUANG Y, CHEN Z, et al. Practices and applications of convolutional neural network-based computer vision systems in animal farming: a review[J]. *Sensors*, 2021, 21(4): 1492.
- [13] XU B, WANG W, GUO L, et al. CattleFaceNet: a cattle face identification approach based on RetinaFace and ArcFace loss[J]. *Computers and Electronics in Agriculture*, 2022, 193: 106675.
- [14] 杨蜀秦, 刘杨启航, 王振, 等. 基于融合坐标信息的改进 YOLO V4 模型识别奶牛面部[J]. *农业工程学报*, 2021, 37(15): 129–135.
YANG Shuqin, LIU Yangqihang, WANG Zhen, et al. Improved YOLO V4 model for face recognition of dairy cow by fusing coordinate information[J]. *Transactions of the CSAE*, 2021, 37(15): 129–135. (in Chinese)
- [15] 秦兴, 宋各方. 基于双线性卷积神经网络的猪脸识别算法[J]. *杭州电子科技大学学报(自然科学版)*, 2019, 39(2): 12–17.
QIN Xing, SONG Gefang. Pig face recognition algorithm based on bilinear convolution neural network[J]. *Journal of Hangzhou Dianzi University (Natural Sciences)*, 2019, 39(2): 12–17. (in Chinese)
- [16] XU F, GAO J, PAN X. Cow face recognition for a small sample based on Siamese DB Capsule Network[J]. *IEEE Access*, 2022, 10: 63189–63198.
- [17] 张宏鸣, 周利香, 李永恒, 等. 基于改进 MobileFaceNet 的山羊识别方法[J]. *农业机械学报*, 2022, 53(5): 267–274.
ZHANG Hongming, ZHOU Lixiang, LI Yongheng, et al. Sheep face recognition method based on improved MobileFaceNet[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(5): 267–274. (in Chinese)
- [18] SONG S, LIU T, WANG H, et al. Using pruning-based YOLOv3 deep learning algorithm for accurate detection of sheep face[J]. *Animals*, 2022, 12(11): 1465.
- [19] PARK J, WOO S, LEE J Y, et al. BAM: bottleneck attention module[J]. *arXiv preprint arXiv:1807.06514*, 2018.
- [20] LIU W, LIAO S, HASAN I. Center and scale prediction: a box-free approach for object detection[J]. *arXiv preprint arXiv:1904.02948*, 2019.
- [21] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904–1916.
- [22] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 8759–8768.
- [23] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 658–666.
- [24] 李文涛, 张岩, 莫锦秋, 等. 基于改进 YOLOv3-tiny 的田间行人与农机障碍物检测[J]. *农业机械学报*, 2020, 51(增刊1): 1–8.
LI Wentao, ZHANG Yan, MO Jinqiu, et al. Detection of pedestrian and agricultural vehicles in field based on improved YOLOv3-tiny[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(Supp.1): 1–8. (in Chinese)
- [25] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141.
- [26] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//*Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 3–19.
- [27] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 13713–13722.