

文章编号: 2095-2163(2023)05-0175-07

中图分类号: TP391.4

文献标志码: A

基于改进 YOLOv4 的羊只检测方法研究

李远征, 李章辉, 王天一

(贵州大学 大数据与信息工程学院, 贵阳 550025)

摘要: 羊只自动检测是大规模智能化羊养殖的基础。针对养殖场环境中存在围栏遮挡以及目标相互遮挡导致检测方法效率低、精度差和易漏检等问题, 本文提出了一种基于改进 YOLOv4 的养殖场环境下羊只检测方法。采用轻量级网络 ShuffleNet V2 作为主干特征提取网络, 使用深度可分离卷积替换普通卷积, 在提升检测精度的同时将网络轻量化; 引入注意力机制增强特征提取能力; 在预测阶段使用 DIoU-NMS 提高检测精度。改进方法的检测精度达 93.57%, 检测速度达 60 frame/s, 参数量降低至 41.13MB, 能够有效提升养殖场环境下羊只检测的精度与速度。

关键词: 羊只检测; YOLOv4; 轻量化网络; 注意力机制; 非极大值抑制

Sheep detection method in farm environment based on improved YOLOv4

LI Yuanzheng, LI Zhanghui, WANG Tianyi

(College of Big Data and Information Engineering, Guizhou University, Guiyang, 550025, China)

[Abstract] Automatic detection of sheep is the basis for large-scale intelligent sheep breeding. Aiming at the problems of low efficiency, poor accuracy and easy leakage of detection methods due to fence occlusion in the farm environment and mutual occlusion of targets, this paper proposed a sheep detection method based on improved YOLOv4 in the farm environment. The lightweight network ShuffleNet V2 was used as the backbone feature extraction network, and the ordinary convolution was replaced by depthwise separable convolution, which improved the detection accuracy and lightened the network. Introduction of attention mechanisms to enhance feature extraction; DIoU-NMS was used during the prediction phase to improve detection accuracy. The detection accuracy of the improved method is 93.57%, the detection speed is 60 frame/s, and the parameter amount is reduced to 41.13 MB, which can effectively improve the accuracy and speed of sheep detection in the farm environment.

[Key words] sheep detection; YOLOv4; lightweight network; attention mechanism; non maximum suppression

0 引言

随着对羊肉等羊产品需求量的增长, 羊只的养殖正在从传统的家庭散式养殖向规模化、信息化养殖方式转变。在新的养殖模式下, 对羊只的养殖管理技术也有了新的需求, 信息化、智能化养殖手段日益受到养殖人员以及客户的青睐^[1-2]。羊只自动检测是信息化、智能化养殖的基础。在传统的规模化羊只养殖产业中, 通常采用在羊躯干部位手工标记或在羊耳处打上耳标, 通过扫码或 RFID 的方式进行羊只检测。手工标记不仅费时费力, 且经常出现标记被污染的现象, 且难以掌握所有羊只的数量及位置。此外, 打耳标不仅会给羊造成不可逆的损伤,

且需工作人员使用感应设备在羊群附近才能进行检测。基于深度学习的目标检测技术已经发展的较为成熟^[3], 通过养殖场中的监控设备非常容易获得大量丰富的羊只个体信息, 利用目标检测算法提取图像中羊的特征, 训练羊只目标检测模型, 从而实现羊只自动检测^[4], 可大大降低养殖成本, 提升工作效率。于意^[5]结合通道注意力模块 SENet 和 GIoU 损失函数的 YOLOv3 算法, 实现了奶山羊的检测。但是, 由于养殖环境背景复杂、个体之间相互遮挡, 因此检测效率仍然有待提升。为提高养殖场环境下羊只检测效率, 本文建立羊只图像数据集, 在 YOLOv4^[6]算法的基础上加以改进, 降低了模型的运算复杂度, 减少羊只相互遮挡造成的漏检, 提高了检

基金项目: 贵州省科学技术基金(黔科合基础-ZK[2021]一般 304, 黔科合基础[2020]1Y254)。

作者简介: 李远征(1997-), 男, 硕士研究生, 主要研究方向: 深度学习、图像处理; 李章辉(1997-), 男, 硕士研究生, 主要研究方向: 深度学习、图像处理; 王天一(1989-), 男, 博士, 副教授, 主要研究方向: 量子通信、图像处理、计算机视觉。

通讯作者: 王天一 Email: tywang@gzu.edu.cn

收稿日期: 2022-05-19

测精度。

1 YOLOv4 目标检测算法改进

1.1 改进 YOLOv4 网络结构

与通用数据集相比,养殖场环境中的羊只图像存在背景复杂、目标旋转、相互遮挡以及图像模糊的问题。同时,使用原始网络进行训练检测时,网络训练时间较长,且需要的显存较大,对聚集在一起的羊只进行检测时容易出现漏检现象,影响检测的精度。

针对上述问题,为提升目标检测算法在养殖场环境下羊只检测的性能,本文在 YOLOv4 算法基础上,对主干特征提取网络、注意力机制和非极大值抑制 3 个部分做出改进,改进后的算法网络结构如图 1 所示。改进的 YOLOv4 使用 ShuffleNet V2^[7] 替代

CSPDarkNet53 作为网络的特征提取部分,使用大量的逐通道卷积 (Depthwise Convolution, DWConv) 降低网络计算量,并使用深度可分离卷积^[8] 替换路径聚合网络 (Path Aggregation Network, PANet) 模块中的普通卷积,实现 YOLOv4 的轻量化;在特征提取网络中引入注意力机制,利用基于卷积块的注意模块 (Convolutional Block Attention Module, CBAM)^[9] 中的通道注意力机制和空间注意力机制,自适应地提升对羊只特征的关注度,减弱对无关特征的关注度,提升检测的精度;在后处理部分改进非极大值抑制 (Non - Maximum Suppression, NMS),使用 DIoU-NMS^[10] 替代传统 NMS,利用预测框之间的距离信息,保留被误删的预测框,减少漏检现象的发生。

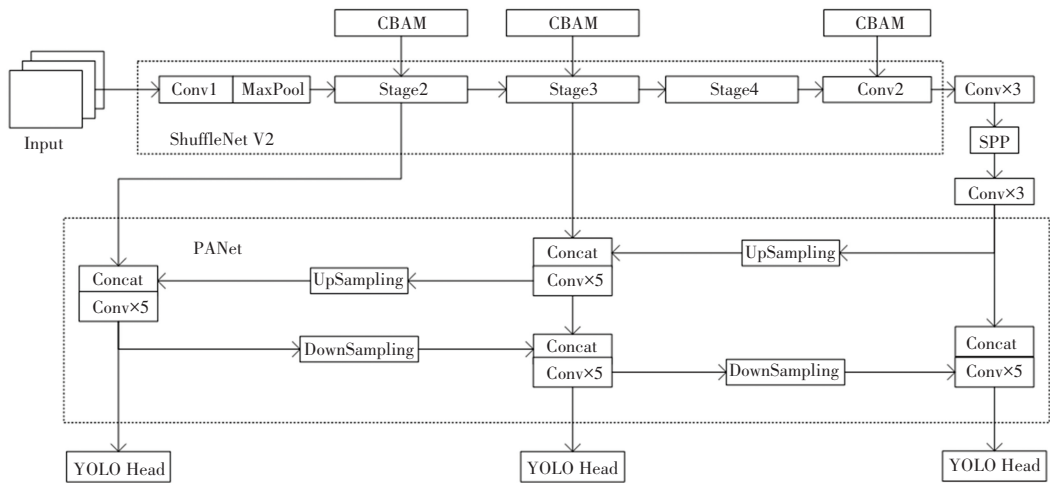


图 1 改进 YOLOv4 目标检测网络结构

Fig. 1 Improved YOLOv4 object detection network structure

1.2 ShuffleNet V2 特征提取网络

针对养殖场环境中羊只检测这一任务,其解决方法应当能够适用于各种规模的养殖企业,因而降低神经网络的计算成本以及对硬件设备的要求,是羊只自动检测算法走向实用的重要方向。更少的参数有助于将网络集成到移动便携设备,降低算法对软硬件的要求,增加易用性。本文在不牺牲 YOLOv4 网络检测性能的前提下,对该网络进行轻量化处理,使用 ShuffleNet V2 作为特征提取网络,并将 PANet 模块的普通卷积替换为深度可分离卷积,节省了 YOLOv4 算法在养殖场环境下过剩的性能,将参数量降低到了原来的 1/6 左右,同时网络的检测精度以及检测速度也得到了提升。

1.2.1 深度可分离卷积

深度可分离卷积就是先进行 DW 卷积再进行点卷积,即 1×1 常规卷积,可融合各个特征通道之间

的信息,改变输出通道数。常规卷积操作如图 2 所示,卷积核个数为 N ,每个卷积核的通道数为 M 。DW 卷积操作如图 3 所示,卷积核个数为 M ,每个卷积核的通道数为 1,仅对输入特征矩阵的一个通道进行卷积运算。常规卷积与 DW 卷积的计算量如下:

$$C_1 = D_K \times D_K \times M \times N \times D_F \times D_F \quad (1)$$

$$C_2 = D_K \times D_K \times M \times D_F \times D_F \quad (2)$$

式中: C_1 为常规卷积计算量, C_2 为逐通道卷积计算量, D_K 为卷积核尺寸, D_F 为输出图像尺寸, M 为输入通道数, N 为输出通道数。

显然, DW 卷积的计算量仅为常规卷积的 $1/N$ 。与普通卷积相比, DW 卷积具有较少的参数和较低的运算成本,对减轻网络负担起到了一定的作用。

然而,虽然 DW 卷积操作的计算量大大减少,但

通道之间没有了信息交互,降低了网络特征提取的性能。为了达到特征交互的目的, ShuffleNet V2 使用 Channel Shuffle 增强了两个分支之间的信息交流,在降低参数和计算复杂度的同时改善了模型效果。

26×26 和 1024×13×13 这 3 种不同尺寸的特征作为 PANet 的输入,通过对不同尺寸的特征融合来加强感受视野,得到最显著的特征。同时,为进一步降低网络计算量,将 PANet 结构中的 3×3 常规卷积替换为深度可分离卷积。

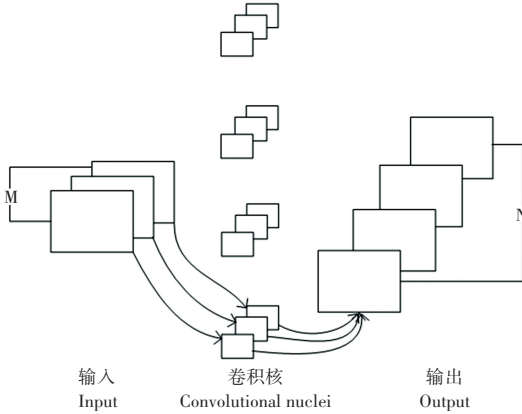


图 2 常规卷积操作

Fig. 2 Regular convolution operations

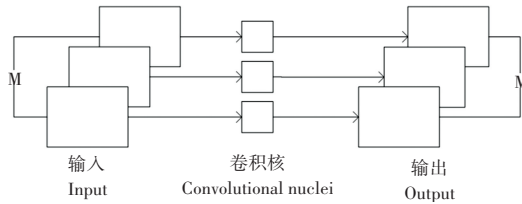


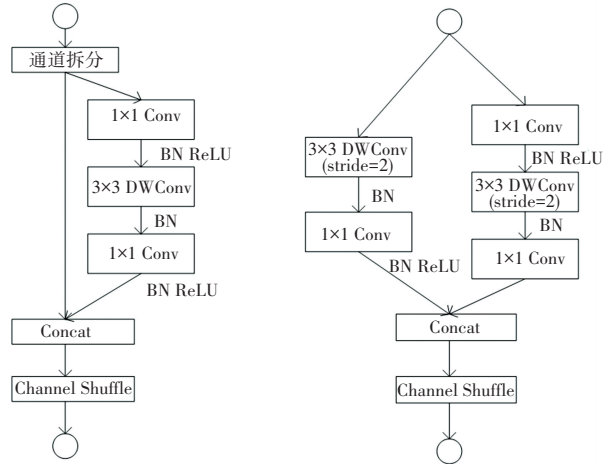
图 3 逐通道卷积操作

Fig. 3 Depthwise convolution operation

1.2.2 ShuffleNet V2 单元结构

ShuffleNet V2 单元结构如图 4 所示。对于特征提取单元(图 4(a)),在每个单元开始时, X 通道的输入经通道被拆分为两部分,一个分支不进行操作;另一个分支经过三层卷积后,通道数与特征尺寸不变,两个分支进行拼接,通道总数仍为 X , 之后的 Channel Shuffle 增强了两个分支之间的信息交流。在下采样单元(图 4(b)),对 X 通道的输入特征同时处理两次,一个分支经过两层卷积层后通道数不变,特征尺寸减半;另一个分支经过三层卷积后通道数不变,特征尺寸减半,两个分支拼接后通道总数为 $2X$, 特征尺寸减半。ShuffleNet V2 单元通过使用 1×1 的卷积和 3×3 的 DW 卷积,极大的降低了网络的计算量,同时利用 Channel Shuffle 保证了提取到的特征质量。

ShuffleNet V2 网络中的 Stage 通过图 4 中模块 (b) 实现下采样,再利用多个模块 (a) 堆叠,实现了对目标特征的高效提取。本文使用的模型为 ShuffleNet v2, Stage2 中模块 (a) 堆叠 3 次, Stage3 堆叠 7 次, Stage4 堆叠 3 次。当输入为 $3 \times 416 \times 416$ 时, Stage2、Stage3 和 Conv 层提取到的 $116 \times 52 \times 52$ 、 $232 \times$



(a) 特征提取单元

(b) 下采样单元

图 4 ShuffleNet V2 单元

Fig. 4 ShuffleNet V2 unit

1.3 注意力机制

近年来,图标检测技术研究发展迅速,但仍存在特征提取不足、识别精度低等问题。注意力机制可以从特征图像中提取出更多的关键信息,抑制无用信息,从而有效地提高神经网络模型的识别精度。由于养殖场中存在大量围栏、地板等设施,且图像质量易受到光照的影响。因此,获取的羊只图像中包含大量的无关信息及噪声。而这些无关特征在神经网络训练时也会被反复地提取并用来预测,影响了网络对羊只的检测精度和普适性。因此,本文在神经网络进行特征提取时,利用 CBAM 的通道注意力机制(Channel Attention Module)和空间注意力机制(Spatial Attention Module),赋予对预测正确结果比较重要的通道(或空间)一个较高的权重,对预测结果较小影响或有负面影响的通道(或空间)赋予一个较低的权重。通过注意力机制对网络特征的自适应调节,增强提取到的羊只特征,抑制提取到的无关背景、亮度和噪声等特征,在增加少量参数和计算量的同时提高检测的精度。

添加注意力模块后的网络结构如图 5 所示。CBAM 通过依次使用通道和空间注意力模块重新校准特征图的权重,对 ShuffleNet V2 提取到的 3 种不同尺度的特征进行有选择地增强,并抑制无用特征,可以有效的抑制图像中的噪声。

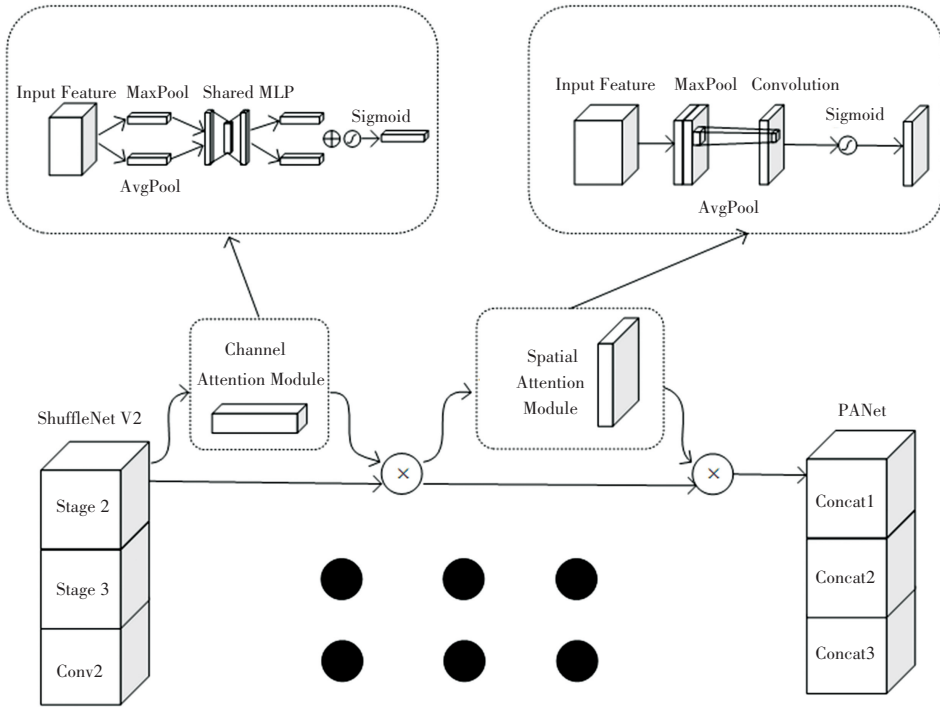


图5 添加CBAM模块的网络结构

Fig. 5 Network structure of adding CBAM module

设输入特征先通过通道注意力模块生成一个一维通道注意力权重,然后通过空间注意力模块生成一个二维空间注意力权重,通道注意力模块和空间注意力模块均使用平均池化和最大池化操作提取特征。前者将得到的特征经过共享多层感知机(Multilayer Perceptron, MPL)后对元素求和,后者则对得到的特征进行卷积,最后二者通过 Sigmoid 函数归一化得到通道和空间注意力权重。

1.4 改进非极大值抑制

由于养殖场环境中羊只数量较多,无论在休息、饮食、行走等状态下,羊只之间均存在不同程度的遮挡情况。当羊只之间相互遮挡程度严重时,就极易发生漏检现象。因此,需要对目标检测生成的预测框进行后处理,以提升检测精度。

在目标检测任务中,交并比(Intersection over Union, IoU)用来衡量两个预测框的相似程度,其计算公式为

$$U_{IoU} = \frac{(A \cap B)}{(A \cup B)} \quad (3)$$

式中: U_{IoU} 为 IoU 值, A 、 B 为两个不同的预测框。

非极大值抑制(Non-Maximum Suppression, NMS)是一种常用的后处理方法。在对 YOLO-Head 获得的预测框进行处理时,首先取出置信度最高的

检测框,然后和其它检测框逐个计算对应的 IoU 值,并将 IoU 超过 NMS 阈值的框全部抑制掉,通过这种方式来删除同一个目标多余的检测框。可以看出,在经典 NMS 算法中, IoU 是唯一考量的因素,但 IoU 只考虑了预测框的重叠区域,在遮挡的情况下往往会导致被遮挡的目标预测框被抑制,造成漏检的错误。由于养殖场环境下羊只个体之间存在大量不同程度的相互遮挡,因此本文使用 DIoU-NMS 解决上述问题。

$DIoU$ 在 IoU 的基础上引入了一个与两个边界框距离相关的惩罚项 R , 该惩罚项定义为

$$R_{DIoU} = \frac{\rho^2(a, b)}{c^2} \quad (4)$$

式中: R_{DIoU} 表示惩罚项值, c 表示包含 A 、 B 两个预测框的最小包络框对角线长度, a 、 b 表示预测框 A 、 B 的中心点坐标, ρ 表示计算二者的欧式距离(图 6 中的 d)。

在大规模检测中,不仅需要考虑预测框与真实框的重叠区域,还要考虑二者中心点之间的距离,则 DIoU-NMS 可定义为

$$s_i = \begin{cases} s_i, U_{IoU} - R_{DIoU}(A, B) < \varepsilon \\ 0, U_{IoU} - R_{DIoU}(A, B) \geq \varepsilon \end{cases} \quad (5)$$

式中: s_i 表示 B 的得分值, ε 为非极大值抑制的阈值。

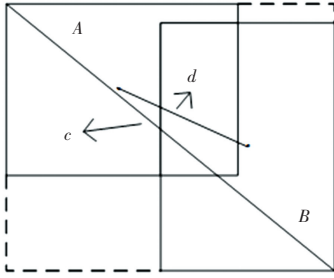


图 6 DIoU 计算参数

Fig. 6 DIoU calculation parameters

使用 DIoU-NMS 时,同时考虑重叠区域和两个边界框中心点之间的距离,增加了其在目标相互遮挡情况下的检测性能。当羊只之间距离较近时,二者对应的预测框重叠面积会较大,即 IoU 值较大,此时 NMS 将认为二者为同一只羊的预测框,仅保留置信度最高的预测框,此时两只羊的预测框仅保留下了一个,从而发生漏检现象。而 DIoU-NMS 在抑制其他预测框时会考虑到预测框之间的位置信息,当两个预测框的中心点存在一定距离时,将认为二者之间是不同羊只的预测框,将其保留下来,避免了漏检现象的发生。

2 实验结果与分析

2.1 数据集与实验环境

实验所用的羊只图像来自于贵州省某养殖场内监控视频。所用羊只彩色图像共 400 幅,为视频中抽帧截取获得,时间涵盖了一天内的各个时段。使用图像标注工具 Labellmg,将羊只图像标注为 VOC 格式数据集,图 7 为羊只标注图像。由于养殖场环境复杂,羊的动作姿态丰富,通过对数据进行旋转、添加噪声和改变图像亮度等操作,将数据集扩充为 4 000 张。随机选取其中 3 200 张作为训练集,800 张作为测试集。



图 7 羊只标注图像

Fig. 7 Sheep labeling picture

实验环境为: Ubuntu18.04、python3.7、Pytorch1.

2.0、显卡型号为 RTX 2080Ti、CPU 为 i7-9700K。

2.2 评价指标

为验证检测网络的性能,使用 AP (Average Precision) 和 FPS (Frame Per Second) 作为主要评价指标。AP 的计算公式如下:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (6)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (7)$$

$$P_{AP} = \int_0^1 P(r) dr \quad (8)$$

式中: P 为查准率 (Precision), R 为查全率 (Recall), N_{TP} 为网络将羊只目标检测正确的数量, N_{FP} 为将非羊只目标误检为羊只目标的数量, N_{FN} 为网络未检测出的羊只数量, $P(r)$ 为查准率 - 查全率 (Precision - Recall) 曲线, P_{AP} 为 AP 值。

2.3 实验结果

实验将输入图像尺寸设置为 412×412 , 使用迁移学习技术,利用网络在大型数据集中训练得到的结果作为预训练权重。使用冻结模型主干网络方式进行训练,冻结训练轮数 (Epoch) 设为 50, Batch size 设为 16, 初始学习率设为 0.001, 总训练轮数设为 300, Batch size 设为 8, 初始学习率设为 0.000 1; 使用余弦退火方法动态调整学习率,使用 Mosaic 对数据进行增强。

图 8 为本文算法训练时的损失函数曲线。由此可以看出,在前 50 轮损失波动较大,第 50 轮解冻后,使用更小的学习率,曲线波动逐渐平缓,损失函数逐渐收敛。使用训练 300 轮得到的权重进行测试, AP 达到了 93.57%。

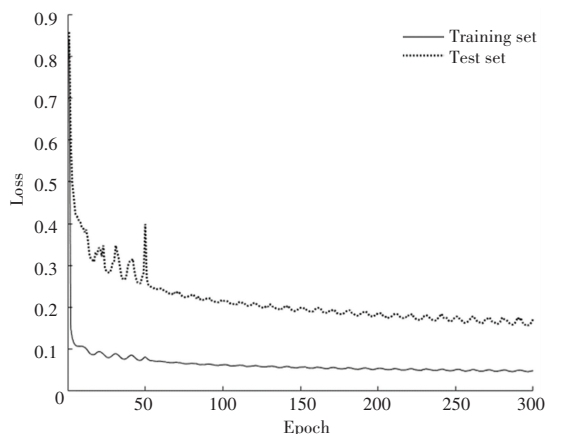


图 8 损失函数

Fig. 8 Loss function

2.4 消融实验

为验证本文改进模块效果,对网络改进部分进

行了对比实验,结果见表1。可以看出,使用轻量化特征提取网络及轻量化模块深度可分离卷积后,模型参数减少了84.01%,*AP*提升了1.42%,*FPS*增加了16;引入CBAM后,在增加1.1%的参数量的情况下提升了0.03%的检测精度,*FPS*减小了1;引入DIoU-NMS后,模型参数量不变,*AP*提升了0.18%,*FPS*减小了6。本文各项改进均对*AP*有不同程度的提升,在轻量化模块对参数量的大幅优化下,引入注意力机制对模型参数量有微量提升,DIoU-NMS在满足实时性要求的前提下提升了检测精度。

表1 不同改进方式测试集上性能比较

Tab. 1 Performance comparison on different improvements on test sets

model	<i>AP</i> / %	<i>FPS</i>	参数大小/MB
YOLOv4	91.87	57	243.90
lightweight	93.29	73	39.00
CBAM	91.90	56	246.53
DIoU-NMS	92.05	51	243.90
Lightweight + CBAM	93.44	67	41.13
Lightweight +DIoU-NMS	93.39	63	39.00
CBAM +DIoU-NMS	92.07	50	246.53
Lightweight + CBAM +DIoU-NMS	93.57	60	41.13

2.5 检测模型对比实验

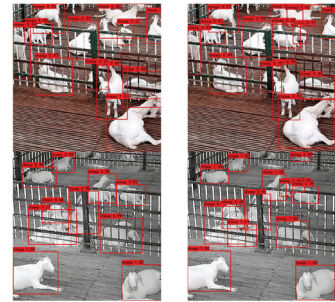
使用经典网络Faster-RCNN^[11]、SSD^[12]、YOLOv3^[13]、YOLOv4、YOLOv5s^[14]算法在本文数据集上进行训练测试,训练轮数设为100,将*AP*、*FPS*与模型参数量作为主要评价指标,得到的预测结果见表2。本文提出算法*AP*达到了91.95%,参数大小为41.13 MB,*FPS*达到了60。虽然*FPS*相较于原有YOLOv4网络有所提升,但在检测精度得到提升的同时,参数大小相比YOLOv3和YOLOv4下降了80%左右;Faster-RCNN作为两阶段目标检测经典算法,其参数大小明显高于其它网络模型,*FPS*也仅有17,难以满足实时检测的要求;SSD虽然参数大小及*FPS*与本文算法相近,但*AP*低于本文算法;YOLOv3与YOLOv5虽然*FPS*高于本文算法,但参数量较大且*AP*低于本文算法。

表2 经典检测算法性能比较

Tab. 2 Classical detection algorithm performance comparison

model	<i>AP</i> / %	<i>FPS</i>	参数大小/MB
Faster-RCNN	83.24	17	521.43
SSD	75.75	87	90.07
YOLOv3	72.80	83	234.69
YOLOv4	88.99	57	243.90
YOLOv5s	78.17	108	177.88
ours	91.95	60	41.13

图9为改进前后羊只目标检测结果的对比。图9(a)为使用YOLOv4训练100轮得到的权重预测得到的结果;在相同条件下,图9(b)是本文改进YOLOv4算法获得的结果。通过对比可以看出,原YOLOv4算法对羊只检测的结果存在定位不准确以及漏检的情况;而改进的YOLOv4检测出的羊只目标定位更加准确,正确检测到的羊只数量更多。



(a) YOLOv4 (b) 本文算法

图9 改进前后检测结果对比

Fig. 9 Improve the comparison of before and after detection results

3 结束语

针对养殖场环境下对羊只进行实时检测比较困难,羊只之间相互遮挡导致漏检率较高等问题,本文提出基于YOLOv4算法的改进方案。首先降低算法对软硬件的要求,在网络参数量减少2/3左右的同时,提升了检测的精度;其次,在特征提取过程中引入注意力机制,提升检测精度;最后,使用DIoU-NMS引入预测框之间的位置信息,提升了网络的检测精度。实验结果表明,本文算法在自制羊只数据集上取得了良好的表现,与原YOLOv4方法相比,参数量明显减少,检测精度得到提升。

在后续的工作中,将继续从实际场景下的应用角度出发,不断研究和改进相关算法,尝试融合疾病检测、个体识别等任务,设计出更优良的羊只自动检测模型。

参考文献

- [1] 王旺虎,郭子记,李锦旗. 智慧养殖在肉羊产业中的应用[J]. 畜牧兽医杂志,2021,40(4):81-84.
- [2] CAPPAL M G, PICCIAU M, NIEDDU G, et al. Long term performance of RFID technology in the large scale identification of small ruminants through electronic ceramic boluses: Implications for animal welfare and regulation compliance[J]. Small Ruminant Research, 2014, 117(2-3): 169-175.
- [3] 姚群力,胡显,雷宏. 深度卷积神经网络在目标检测中的研究进展[J]. 计算机工程与应用,2018,54(17):1-9.
- [4] 王栋. 基于CNN的奶山羊监控视频中目标检测算法研究[D]. 咸阳:西北农林科技大学,2018:27-46.