

基于改进 YOLO v7 - Pose 的牛脸关键点检测与姿态识别方法

黄小平^{1,2} 侯现坤¹ 郭阳阳^{1,2} 郑寰宇¹ 豆子豪¹ 刘梦艺¹ 赵晋陵^{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: 奶牛脸部关键点检测在牧场智能化中发挥着重要的作用, 它可以帮助实现牛脸识别、牛脸对齐、头部动作检测与行为识别等。针对目前奶牛养殖环境中存在牛脸遮挡、弱光照等问题, 提出了一种改进的 YOLO v7 - Pose 网络模型的算法, 可用于牛脸关键点检测和头部姿态识别。首先通过网络相机在牧场获取奶牛脸部图像并构建数据集。其次, 在 YOLO v7 - Pose 网络模型中引入 SPPFCSPCL 结构, 以提高奶牛脸部关键点的特征提取能力; 将关键点检测的损失函数 OKS 替换为 WingLoss 损失函数, 增加奶牛脸部关键点的检测精度; 最后, 使用 L1 范数对改进的模型剪枝, 使改进后的网络模型参数量降低。试验结果表明: 改进模型 YOLO v7 - SCLWL - Pose 检测牛脸关键点较原模型 AP 提升 5 个百分点, AP0.5 提升 2.7 个百分点, 改进后模型内存占用量仅为 106.7 MB, 减少 33.6%。将本文关键点检测用于姿态识别, 试验结果对抬头和低头等动作的识别准确率达到了 95.5% 和 86.5%。本研究为牧场奶牛行为识别提供了支撑技术。

关键词: 牛脸检测; 关键点检测; YOLO v7 - Pose; 姿态识别

中图分类号: TP391.4 文献标识码: A 文章编号: 1000-1298(2024)11-0084-09

OSID:



Cow Face Keypoint Detection and Pose Recognition Based on Improved YOLO v7 - Pose

HUANG Xiaoping^{1,2} HOU Xiankun¹ GUO Yangyang^{1,2} ZHENG Huanyu¹ DOU Zihao¹

LIU Mengyi¹ ZHAO Jinling^{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: Facial keypoint detection in dairy cows plays a crucial role in the automation of cow farms. It aids in cow face recognition, face alignment, head movement detection, and behavior recognition. In view of the problems of cow face occlusion and weak light in the current dairy farming environment, an improved algorithm of YOLO v7 - Pose network model was proposed, which can be used for keypoint detection and head pose recognition of cow face. Firstly, dairy cow facial images were collected from cow farms by using network cameras and a dataset was constructed. Secondly, the SPPFCSPCL structure was integrated into the YOLO v7 - Pose network model to enhance its feature extraction capabilities for cow facial keypoints. The WingLoss loss function replaced the OKS loss function for keypoint detection, thereby improving the accuracy of cow facial keypoint detection. Finally, L1 regularization was applied to prune the improved model, reducing the number of network parameters. The experimental results showed that the cow face keypoint detection of improved model YOLO v7 - SCLWL - Pose was improved by 5 percentage points and AP0.5 was improved by 2.7 percentage points compared with the original model AP, and the memory occupation of the improved model was only 106.7 MB, which was reduced by 33.6%. The keypoint detection was applied to pose recognition, and the experimental results showed that the recognition accuracy of the motions of looking up and looking down reached 95.5% and 86.5%. This research can provide support technology for behavior recognition in dairy cows on farms.

Key words: cow face detection; keypoint detection; YOLO v7 - Pose; pose recognition

收稿日期: 2024-08-19 修回日期: 2024-09-23

基金项目: 安徽省自然科学基金项目(2308085MC103)和安徽省教育厅高校项目(KJ2021A0024)

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

通信作者: 赵晋陵(1981—),男,教授,博士,主要从事农业信息化研究,E-mail: zhaopl@ahu.edu.cn

0 引言

随着奶牛规模化养殖的快速发展以及存栏规模的不断提高,牧场对信息化的需求越来越大。同时为了降低管理成本并提升动物福利,牧场机械化、智能化应用逐渐普及^[1]。奶牛脸部关键点检测是姿态识别和行为识别的关键,它可协助牧场作精准管理。牛脸部关键点检测需要对牛脸的某些特征点进行检测和定位,它是奶牛头部姿态估计的重要组成部分,也是奶牛采食行为分析、健康状况预警、疾病检测等重要支撑技术,因此牛脸部关键点检测具有重要的意义。

为了实现奶牛脸部关键点检测,传统的视觉方法通过对牛脸进行定位获取图像的纹理特征,如方向梯度直方图 (Histogram of oriented gradient, HOG)^[2]、局部二值化 (Local binary pattern, LBP)^[3]、尺度不变性特征转换 (Scale invariant feature transform, SIFT)^[4],最后通过机器学习算法检测出图像中的牛脸关键点。文献[5]通过自适应级联检测器定位牛脸位置,使用局部二值化和监督式梯度下降法提取牛脸轮廓。这些算法的缺点是实时性差,且轮廓关键点提取准确率取决于牛脸特征是否充分提取。

近年来深度学习的快速发展,使算法无需专门为特征提取建模,通过深度学习模型实现奶牛脸部的目标检测和关键点检测。这些算法可分为:端到端和非端到端检测模型。端到端的代表模型主要有 YOLO v5 - Pose^[6]、YOLO v7 - Pose^[7]、YOLO v8 - Pose^[8],可以实现端到端的目标检测和关键点检测。非端到端检测模型主要有 YOLO v5^[9]、YOLO v7^[10]、RetinaFace^[11]等,它们先检测出牛脸位置,再通过 HRNet^[12]、RTmPose^[13]等模型检测出牛脸部关键点。

非端到端的检测模型实时性差、算法复杂、不利于移动嵌入式应用场景^[14-16]。现有端到端算法^[17-19]检测出了动物脸部及关键点,但是没有进一步对脸部做姿态识别研究。

现有模型算法对牛及其他动物脸部关键点检测的研究工作较少,且大部分研究把动物关键点检测和目标检测分开。在遮挡、重叠和光线干扰等复杂的自然环境中,算法模型容易出现漏检、误检等问题。目前缺少端到端算法模型对牛脸部关键点检测,尤其缺乏牛脸部关键点检测的相关应用,例如姿态识别等。

本文提出一种改进的 YOLO v7 - Pose 端到端网络模型,并在复杂场景下对牛脸关键点检测和姿态识别进行研究。以荷斯坦奶牛为对象,构建不同光照强度、遮挡、单目标以及多目标等自然应用场景下的牛脸部图像,在 YOLO v7 - Pose 模型基础上构建端到端的奶牛脸部关键点检测模型。用 SPPFCSPCL 空间金字塔模块来代替 SPPCSPC 模块,以提高网络模型的特征提取能力;采用 WingLoss^[20] 损失函数代替对象关键点相似度 (Object keypoint similarity, OKS)^[21] 损失函数,以增加牛脸关键点检测精度;考虑后期在移动设备上移植,使用 L1 范数的过滤器剪枝算法^[22],减少模型参数量以实现模型轻量化。

1 试验材料

图像视频采集于安徽省六安市华好生态养殖有限公司与江苏省淮安市盱眙卫岗牧业有限公司。在牛场的自然饲养环境下,通过网络相机拍摄牛场的荷斯坦奶牛,视频每帧包含 1 ~ 7 头牛不等,分辨率为 1 920 像素 × 1 080 像素,存储格式为 MP4。将视频转成图像,并对图像进行筛选,最终获得 846 幅 jpg 格式图像。数据集中包含不同光照、遮挡、模糊、不同角度、单目标以及多目标等条件下的奶牛图像数据,充分展示了牛场的复杂性,图像样本如图 1 所示。为提高模型的泛化能力和鲁棒性,对图像进行上下翻转,加入噪声和随机点等,使数据集扩充为 2 538 幅,数据按照比例 8:2 随机划分为训练集和测试集,其中训练集样本为 2 031 幅,测试集样本为 507 幅。

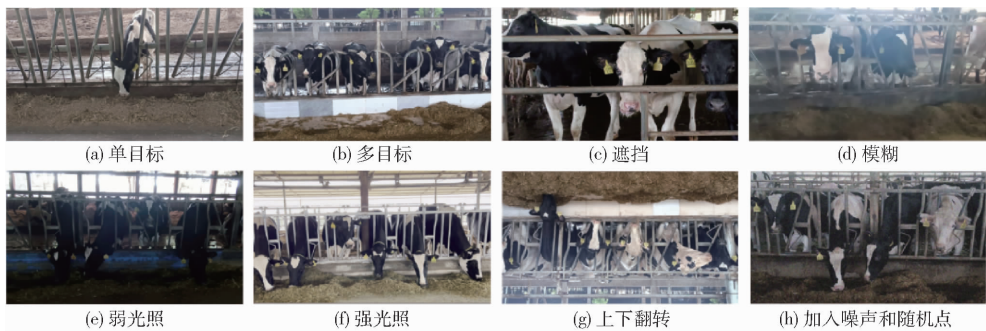


图 1 奶牛原始图像以及增强图像

Fig. 1 Original and enhanced images of cows

使用 Labelme 图像标注工具对图像中的牛脸部区域及关键点进行标注,目标框标注的范围为奶牛的面部。本文标注的奶牛脸部有较多特征,参照文献[23],选取左右眼睛、左右嘴角以及鼻尖中心作为检测关键点。最后将标注得到的 json 文件转为 txt 格式,以用于改进 YOLO v7 - Pose 的模型训练。

2 牛脸姿态估计方法

2.1 YOLO v7 - Pose 模型及其改进的基本思路

YOLO v7 - Pose 是端到端的关键点检测算法,具有创新的多分支堆叠结构,使其具有充分的特征提取能力。YOLO v7 - Pose 模型由 Input 输入端、Backbone 骨干网络、Head 特征提取网络、Prediction 预测层 4 部分组成。原始的 YOLO - Pose 主要针对人体 17 个关键点进行检测,本文仅对奶牛脸部 5 个关键点进行检测,因此对原有的

YOLO v7 - Pose 模型进行改进以适应本文应用场景。首先,将 SPPCSPC 模块重构为 SPPFCSPCL 模块,以增强网络模型对牛脸特征的提取能力。其次,采用 WingLoss 损失函数替代 OKS 损失函数,从而减小离群点的影响,提升牛脸关键点检测的精度。最后,考虑后期移动或嵌入式设备移植,本文使用 L1 范数的滤波器剪枝算法,从而减少模型参数量,实现模型轻量化。改进后网络模型如图 2 所示。

2.2 重构 SPPCSPC 模块

YOLO v7 - Pose 骨干网络生成的特征图被输入到 SPPCSPC 模块。SPPCSPC 模块由空间金字塔池化结构(Spatial pyramid pooling, SPP)^[24]、跨阶段局部连接结构(Cross stage partial connection, CSP)^[25]和 1 个卷积层组成。如图 3 所示,SPP 结构汇集无处理、5 × 5、9 × 9 和 13 × 13 等多尺度的感受野,可以帮助模型提升特征提取能力;CSP 通过跨连接方

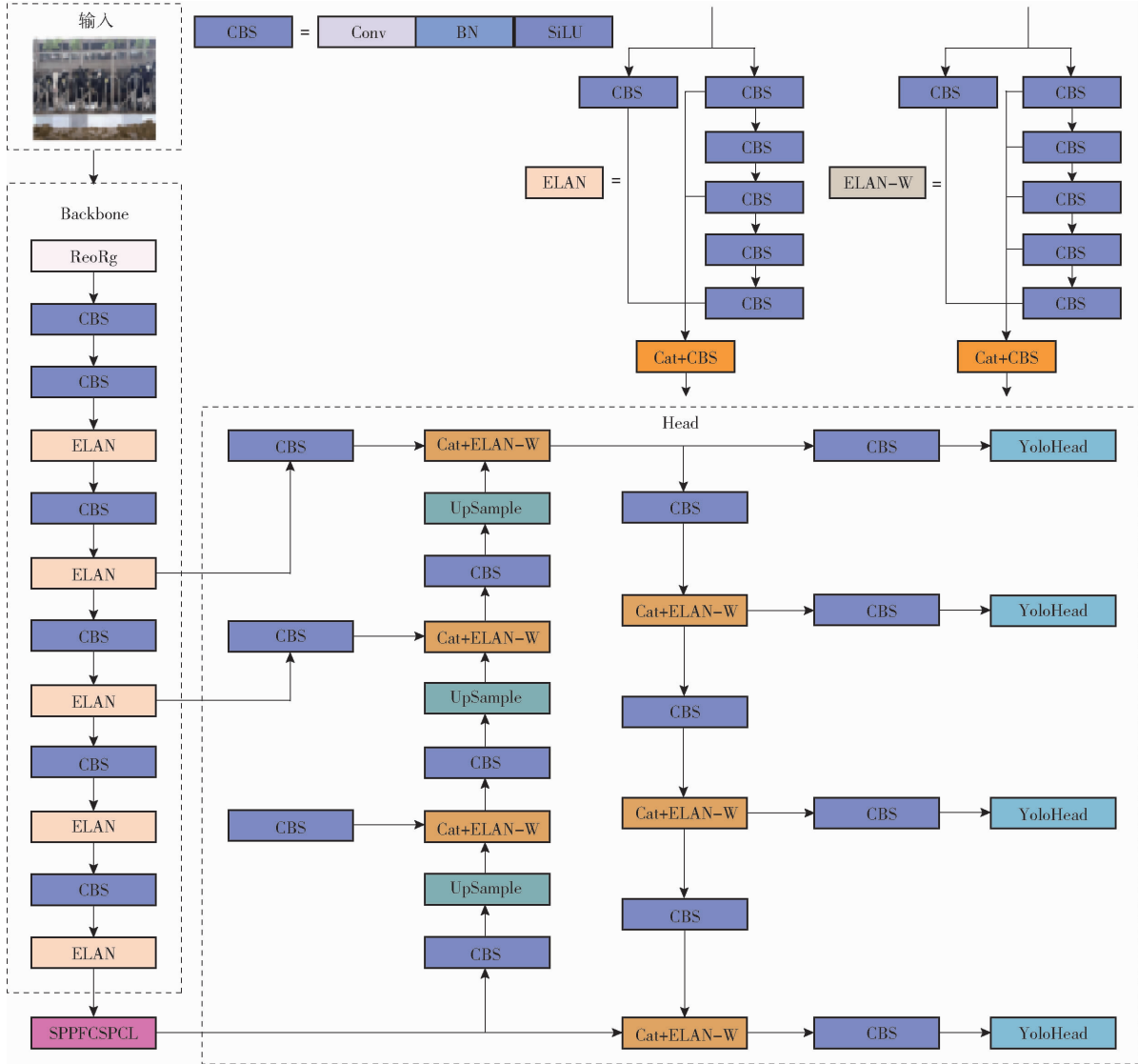


图 2 YOLO v7 - SCLWL - Pose 网络结构

Fig. 2 YOLO v7 - SCLWL - Pose network structure

法,使每个阶段之间共享信息,减少网络的计算量; CBS 为卷积层,用于进一步的特征提取。但是 SPPCSPC 结构中的并联结构会使检测速度变慢,且它对小目标的关注度不够深入,不利于本文牛脸部的关键点检测。为了提升检测速度,把并联结构改为 5×5 串联结构,提升了检测速度并保证了模块

5×5 、 9×9 和 13×13 感受野不变。为了提升牛脸关键点检测的精度,把串联的 5×5 最大池化层改为 3×3 的最大池化层,模块的感受野变为 3×3 、 5×5 和 7×7 ,这样更有利于在 $20 \times 20 \times 1\,024$ 的特征图像上对牛脸关键点进行特征提取。重构后的 SPPFCSPCL 模块如图 3 所示。

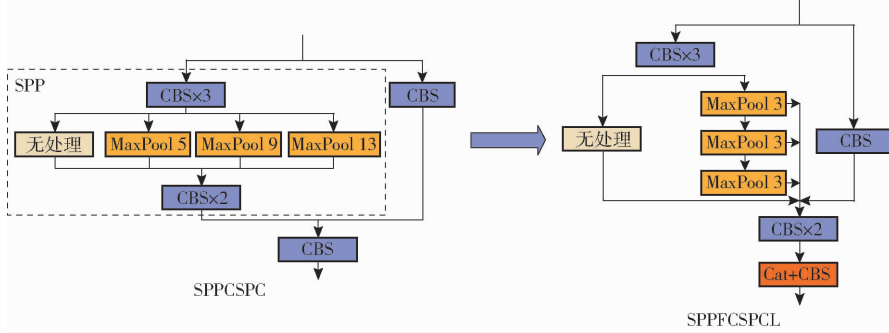


图 3 重构 SPPCSPC 为 SPPFCSPCL 模块

Fig. 3 Reconstruct SPPCSPC as SPPFCSPCL module

2.3 引入 WingLoss 损失函数

YOLO v7 - Pose 是基于回归的检测模型,使用完整的交并比 (Complete - IoU, CIOU)^[26] 损失作为检测框的损失函数,使用 OKS 损失作为关键点损失函数。OKS 是一种用于衡量关键点预测值与真实值之间相似度的指标。计算每个关键点的 OKS 时,要先构建关键点预测值与真实值的特征向量,然后使用欧氏距离来计算两个特征向量的相似度,损失函数计算公式为

$$OKS_p = \frac{\sum_i \exp\left(\frac{-d_{pi}^2}{2s_p^2\sigma_i^2}\right)\delta}{\sum_i v_{pi}} \quad (v_{pi} > 0) \quad (1)$$

式中 d_{pi} ——第 p 个目标第 i 个关键点预测值与真实值间的欧氏距离

v_{pi} ——第 p 个目标第 i 个关键点的可见性指数

s_p ——第 p 个目标边界框面积

σ_i ——第 i 个关键点标注值与实际值间的标准偏差

OKS 损失函数对离群点、异常值较为敏感,因为 OKS 是使用欧氏距离计算特征向量相似度。原模型 YOLO v7 - Pose 对人体 17 个关键点检测,本文仅对牛脸 5 个关键点进行检测,因此检测点数量较原模型减少,会导致后期关键点的损失相对检测框的损失较小。为了解决 OKS 损失函数在本文牛脸关键点检测时存在的问题,引入 WingLoss^[20] 损失函数,公式为

$$Wing(x) = \begin{cases} w \ln\left(1 + \frac{|x|}{\varepsilon}\right) & (|x| < w) \\ |x| - C & (|x| \geq w) \end{cases} \quad (2)$$

其中 $C = w - w \ln(1 + w/\varepsilon)$

其中, ε 是一个很小的数值用于约束非线性区域的曲率, C 是个常数,可以平滑连接线性和非线性的部分, w 用于将非线性部分限制在 $[-w, w]$ 区间内。

当 x 的绝对值小于 w 时,使用 $\ln x$ 函数,其梯度为 $1/x$,该部分可增强小误差的影响。为避免过度补偿小误差,在错误的方向梯度产生大的更新,需要在处理小的损失时对梯度进行限制,采用两个参数 w 和 ε 来控制梯度。当 x 绝对值大于等于 w 时, WingLoss 损失对于大的误差使用 L1 范数,使大误差能够快速恢复。

2.4 基于 L1 范数的通道剪枝

随着深度学习模型的深度增加,模型计算量和参数存储量会不断提高。为了降低模型计算量和参数量,删除训练模型中权重较小的过滤器核及其连接的特征图,以及特征图对应的卷积核,从而提高计算效率。如图 4 所示,对输入通道为 n_i 、宽度和高度为 w_i 和 h_i 的特征图 X_i 使用 n_{i+1} 个过滤器进行卷积,计算每个过滤器权重的 L1 范数来判断过滤器在每层卷积中的重要程度,然后对过滤器权重的 L1 范数进行排序,修剪掉 m 个权重较小的过滤器及其对应的特征图。由修剪比例确定 m ,对应修剪的特征图的下一个过滤器中的卷积核也被移除 (深色部分),修剪后重新训练模型以便恢复目标检测的精度。基于 L1 范数剪枝^[23] 可以修剪和压缩各个层的权重,使模型在检测速度和精度中达到最优。

2.5 基于关键点的牛脸姿态识别

以奶牛鼻子中心点为原点建立世界坐标系,获取牛脸左右眼睛、左右嘴巴的坐标。通过算法模型预测得到各关键点的对应检测坐标,结合相机内参,

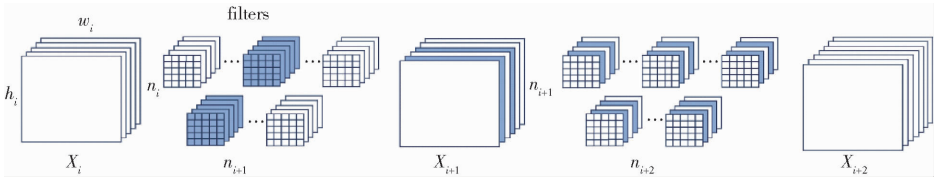


图 4 基于 L1 范数的剪枝过程
Fig.4 Pruning process based on L1 norm

可以得到旋转矩阵,再通过旋转矩阵计算绕 X 轴的俯仰角 (Pitch)、绕 Y 轴的偏转角 (Yaw) 和绕 Z 轴的滚动角 (Roll)^[27],如图 5a 所示。

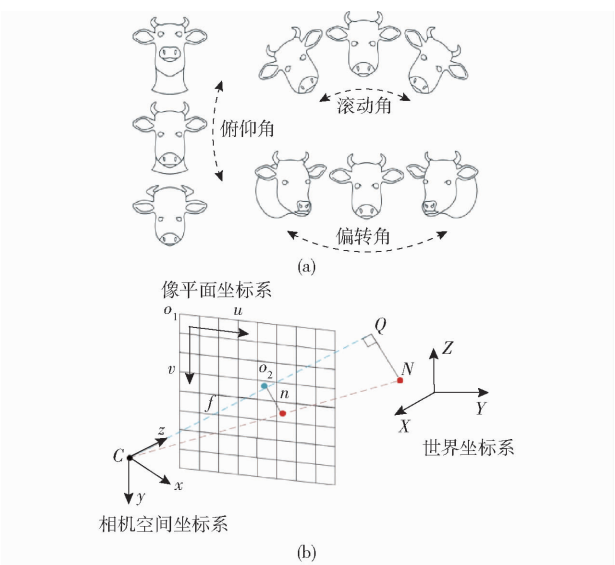


图 5 牛脸姿态的表示及坐标变换
Fig.5 Representation and coordinate transformation of cow face pose

如图 5b 所示, N 为牛脸关键点,相机坐标系下点 N 在像平面坐标系下的投影点为点 n ,点 o_1 是像平面坐标系下的原点,点 o_2 是相机坐标系的原点在像平面坐标系下的投影点。首先通过改进的模型预测出关键点位置,结合牛脸关键点在世界坐标系的三维坐标,利用张氏标定^[28]计算出相机内参和外参。最后通过外参可以得到旋转矩阵,根据旋转矩阵计算俯仰角 θ_x 、偏转角 θ_y 和滚动角 θ_z ,旋转矩阵及各姿态角度计算公式为

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{3}$$

$$\theta_x = \text{atan2}(r_{32}, r_{33}) \tag{4}$$

$$\theta_y = \text{atan2}(-r_{31}, \sqrt{r_{32}^2 + r_{33}^2}) \tag{5}$$

$$\theta_z = \text{atan2}(r_{21} + r_{11}) \tag{6}$$

3 试验及参数设置

3.1 试验设置

本试验在 Ubuntu 20.4 操作系统进行,CPU 为

Intel Core i7-9700,GPU 为 NVIDIA RTX3080 显卡,内存容量为 12 GB。模型训练采用 Pytorch 深度学习框架,版本为 1.12.1,Python 版本 3.10.6。批量大小参照模型占用显存大小进行设定,训练 500 个周期,学习率为 0.01,初始动量设为 0.937。模型训练时,YOLO v5-Pose、YOLO v7-Pose、YOLO v8-Pose 使用 OKS 损失函数。

3.2 评价指标

为评价模型对牛脸目标检测的性能,采用平均精度均值 (Mean average precision, mAP)、精度和召回率作为评价指标。mAP@0.5 表示使用 0.5 作为交并比 (Intersection over union, IOU) 阈值时平均精度均值, mAP@0.5:0.95 表示 IOU 阈值从 0.5 到 0.95 之间,步长为 0.05 时平均精度均值。

为评价关键点检测的性能,使用以下指标:浮点运算数 (Giga floating-point operations per second, GFLOPS)、内存占用量、平均精度 (Average precision, AP) 和平均召回率 (Average recall, AR)。测试时,计算 OKS 不需要增大相似度。AP0.5 表示当 OKS 阈值设置为 0.5 的平均精度。AR 表示 OKS 阈值 0.5 到 0.95 之间,每隔 0.05 平均召回率均值, AR0.5 为当 OKS 阈值设置为 0.5 时平均召回率。

4 试验结果与分析

4.1 不同模型下检测结果对比

在本文牛脸检测的应用场景中,为了验证模型 YOLO v7w6-Pose 比其它模型更具有精度优势,本试验选用 YOLO v5s6-Pose、YOLO v8n-Pose、YOLO v8l-Pose 共 3 个模型作对比。在测试数据集上牛脸检测结果如表 1 所示,牛脸关键点检测结果如表 2 所示。由表 1 可以看出,YOLO v7w6-Pose、YOLO v5s6-Pose、YOLO v8l-Pose 和 YOLO v8n-Pose 这 4 个模型的精度、召回率、mAP@0.5 都在 99% 以上,从宏观上看牛脸部检测效果都比较好。由表 2 可以看出,YOLO v7w6-Pose 的 AP、AP0.5、AR 和 AR0.5 值相比于其它几个模型具有明显优势,但是内存占用量是最高的,因此选用 YOLO v7w6-Pose 为基础模型,同时对其作轻量化改进。

表 1 不同模型牛脸检测指标对比

Tab.1 Comparison of detection indexes of cow face in different models

模型	精度	召回率	mAP@ 0. 5	mAP@ 0. 5 ; 0. 95
YOLO v5s6 - Pose	99. 3	99. 8	99. 7	84. 7
YOLO v8l - Pose	99. 8	100	99. 5	87. 0
YOLO v8n - Pose	99. 8	99. 9	99. 5	87. 2
YOLO v7w6 - Pose	99. 2	99. 9	99. 8	86. 0

表 2 不同模型关键点检测指标对比

Tab.2 Comparison of keypoint detection indexes of different models

模型	浮点运算数	内存占用量/MB	AP/ %	AP0. 5/ %	AR/ %	AR0. 5/ %
YOLO v5s6 - Pose	1.990×10^{10}	30. 4	46. 4	88. 0	55. 9	91. 2
YOLO v8l - Pose	1.685×10^{11}	89. 4	49. 5	89. 5	60. 8	93. 3
YOLO v8n - Pose	8.300×10^9	6. 5	36. 3	83. 0	47. 0	88. 0
YOLO v7w6 - Pose	1.009×10^{11}	160. 6	51. 7	89. 8	62. 7	94. 0

4.2 消融试验

为验证 SPPFCSPCL 模块和 WingLoss 损失函数对模型 YOLO v7 - Pose 的贡献,在数据集上进行消融试验。表 3 为 YOLO v7w6 - Pose、SPPFCSPCL 模块和 WingLoss 损失函数的不同组合方案得到的试验结果。由表 3 可知,在 YOLO v7w6 - Pose 引入 SPPFCSPCL 模块,其 AP、AP0. 5、AR 和 AR0. 5 分别提升 0. 7、1. 8、0. 6、0. 7 个百分点。试验结果表明,引入 SPPFCSPCL 模块有利于牛脸的特征提取,同时有利于模型精度提高。将 OKS 损失函数改为 WingLoss 损失函数后,其 AP、AP0. 5、AR 和 AR0. 5 分别提升 3、1. 5、3、0. 6 个百分点。试验结果表明,WingLoss 损失函数在训练的后期能够增强小误差的影响,从而提升关键点检测精度。同时引入 SPPFCSPCL 模块和 WingLoss 损失函数,检测精度进一步提升,与原始 YOLO v7w6 - Pose 相比 AP、AP0. 5、AR 和 AR0. 5 分别提升 3. 2、1. 6、2. 7、0. 6 个百分点。

表 4 模型轻量化消融试验结果

Tab.4 Model lightweight ablation experiment

模型	浮点运算数	内存占用量/MB	AP/ %	AP0. 5/ %	AR/ %	AR0. 5/ %
Our - YOLO v7	1.009×10^{11}	160. 6	54. 9	91. 4	65. 4	94. 6
Our - YOLO v7 + Ghost	5.880×10^{10}	99. 5	47. 9	88. 5	59. 7	92. 1
Our - YOLO v7 + Slim/1. 5	6.670×10^{10}	106. 7	55. 9	92. 9	66. 3	95. 2
Our - YOLO v7 + Slim/2. 0	5.010×10^{10}	80. 3	53. 8	91. 4	64. 1	94. 1
Our - YOLO v7 + Lamp/1. 5	6.670×10^{10}	106. 7	55. 9	92. 0	65. 8	94. 6
Our - YOLO v7 + Lamp/2. 0	5.010×10^{10}	80. 3	55. 1	91. 9	64. 7	94. 5
Our - YOLO v7 + Random/1. 5	6.670×10^{10}	106. 7	56. 1	90. 9	67. 0	94. 6
Our - YOLO v7 + Random/2. 0	5.010×10^{10}	80. 3	54. 6	92. 5	65. 3	95. 1
Our - YOLO v7 + L1/1. 5	6.670×10^{10}	106. 7	56. 7	92. 5	66. 5	95. 2
Our - YOLO v7 + L1/2. 0	5.010×10^{10}	80. 3	54. 9	91. 2	65. 6	94. 4

表 3 消融试验结果

Tab.3 Ablation results

YOLO v7 - Pose	SPPFCSPCL	WingLoss	AP	AP0. 5	AR	AR0. 5
✓			51. 7	89. 8	62. 7	94. 0
✓	✓		52. 4	91. 6	63. 3	94. 7
✓		✓	54. 7	91. 3	65. 7	94. 6
✓	✓	✓	54. 9	91. 4	65. 4	94. 6

注:✓表示该模型、模块或损失函数被使用。

4.3 轻量化试验

为了验证改进后的模型轻量化的效果,将使用 Ghost^[29] 卷积和 L1 范数^[22]、Slim (Network slimming)^[30]、Lamp (Layer-adaptive magnitude-based pruning)^[31]、Random^[32] 剪枝进行对比,剪枝前后权重参数量比例分别设置为 1. 5 和 2. 0。

Ghost 卷积使用普通卷积生成一部分特征图,普通卷积生成的特征图通过线性运算得到另外一部分特征图,从而减少模型的计算量和参数量。L1 剪枝是计算过滤器权重的 L1 范数,然后裁剪掉较小的权重及其连接的特征图,以及特征图对应的卷积核;Random 剪枝是随机修剪参数;Lamp 剪枝是通过修剪 Lamp 分数较小的连接;Slim 剪枝是通过 BN 层中的缩放因子 γ 来判断通道重要性,保留 γ 值较大的通道。

在表 4 中,将引入 SPPFCSPCL 模块和 WingLoss 损失函数的改进算法简称为 Our - YOLO v7。由表 4 可知,改进后模型中的 Conv 卷积换成 GhostConv 卷积,模型内存占用量减少 38%,但是 AP 和 AR 均有所下降。通过试验发现,对改进后的模型使用 L1 剪枝算法,剪枝前后权重参数量与剪枝后权重参数量比例设置为 1. 5,能减少模型过度拟合,使模型精度和轻量化达到平衡。L1 剪枝后模型精度比剪枝前的模型 AP 提升 1. 8 个百分点,AR 提升 1. 3 个百分点,模型内存占用量减少 33. 6%,浮点运算数减少 33. 9%。权重参数量比例设置为 2. 0 时相比于 1. 5,AP、AP0. 5、

AR、AR0.5 都有所下降。通过对比试验,最终选定剪枝前后权重参数量比例为 1.5。试验结果表明通过删除 L1 范数权重较低的滤波器及其连接的特征图,以及特征图对应的卷积核,能够降低计算成本,同时不会导致稀疏连接问题。

4.4 不同环境下检测结果

为了验证改进模型对奶牛脸部及关键点检测效果,分别对比在正常、弱光照、遮挡、模糊环境下

YOLO v7 - Pose 和 YOLO v7 - SCLWL - Pose 的检测性能。各种环境下样本检测结果如图 6 所示,原 YOLO v7 - Pose 算法检测奶牛关键点时出现漏检现象,改进后的模型 YOLO v7 - SCLWL - Pose 避免了漏检问题,同时牛脸检测的置信度有所提升。可见,YOLO v7 - Pose 引入 SPPFCSPCL 模块以及 WingLoss 损失函数并使用 L1 剪枝后,可以更好地对牛脸特征进行提取。



图 6 YOLO v7 - Pose、YOLO v7 - SCLWL - Pose 在不同光照和模糊条件下的检测效果
Fig.6 Detection effects of YOLO v7 - Pose, YOLO v7 - SCLWL - Pose under different conditions

为了更加直观地体现模型的关注区域,基于公理的梯度加权类激活映射 (Axiom-based grad-cam, XGrad - CAM) ^[33] 对牛脸重要特征区域绘制热力图进行可视化展示。在模型的预测层前加入 XGrad - CAM,直观地展示了在不同环境下的牛脸部注意力分布图,如图 7 所示。YOLO v7 - Pose 注意力分布比较分散,改进后模型 YOLO v7 - SCLWL - Pose 注意力分布更加集中,对脸部特征点的定位更加精准,这有利于牛脸关键点检测和姿态识别。

4.5 姿态识别结果

图 8 展示了算法对牛脸姿态的识别结果,其中 P 和 Y 分别表示俯仰角和偏转角,在强光照、弱光照和模糊环境下,算法均能计算出奶牛俯仰角和偏转角。以图 8a 强光照条件为例,从左到右奶牛头部俯仰情况呈从低到高再到低趋势,对应的俯仰角 (Pitch) 分别为 - 58.03°、- 21.45°、12.44°、- 26.22°和 - 62.80°,这也反映了牛头俯

仰情况。通过牛脸检测框的中心点坐标可以辅助判断奶牛是否低头,为进一步的行为识别提供技术支撑。偏转角 (Yaw) 表示奶牛头部左右偏转角度,可用于辅助判断奶牛是否处于大角度转动的行为。如果奶牛没有达到足够的饱腹感,它们可能表现出左右摇头、舔其它牛或犊的身体等非正常的口腔行为 ^[34 - 35]。

奶牛头部正常情况下约为 - 45°,因此将 Pitch 的阈值设置为 - 45°,小于 - 45°认为低头,大于 - 45°认为抬头。将 Yaw 的阈值设置为 0°,小于 0°代表向左转向,反之代表向右转向。选择 137 幅图像进行测试,分别统计每幅图像中奶牛的 Pitch 值和 Yaw 值,奶牛头部姿态估计试验结果如表 5 所示。从表 5 中可以看出,模型对抬头、低头、左右转动的姿态识别精度 AP 均大于 86%,AR 均在 90%左右。试验结果表明,模型对奶牛姿态识别具有可行性。

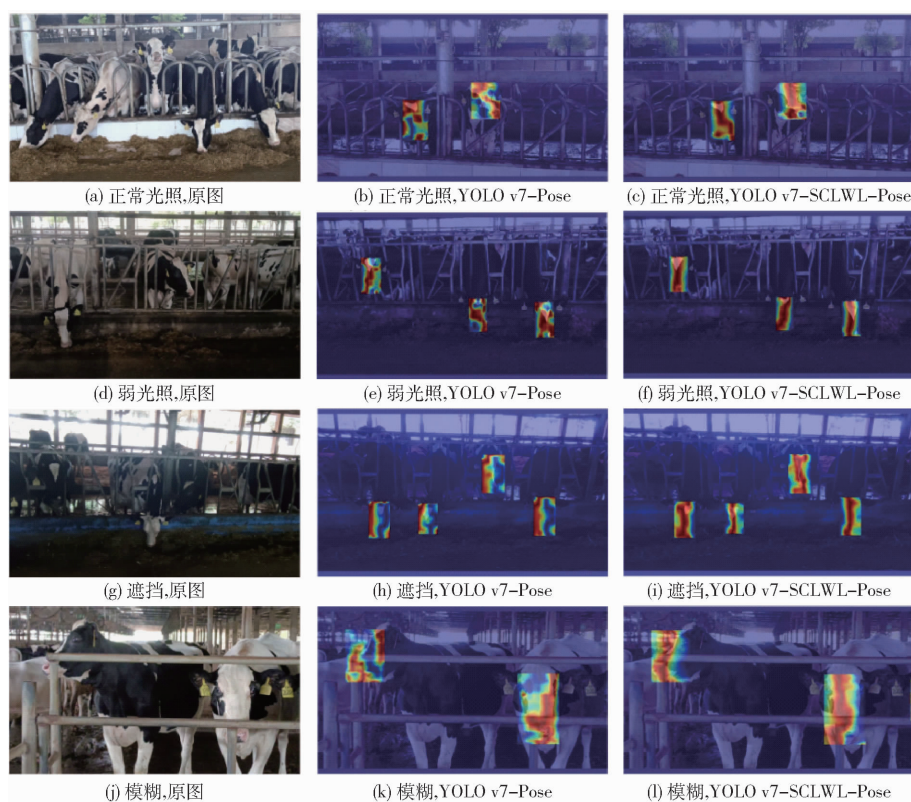


图 7 YOLO v7 - Pose、YOLO v7 - SCLWL - Pose 在不同光照和模糊条件下的热力图
Fig. 7 Thermal maps of YOLO v7 - Pose, YOLO v7 - SCLWL - Pose under different conditions

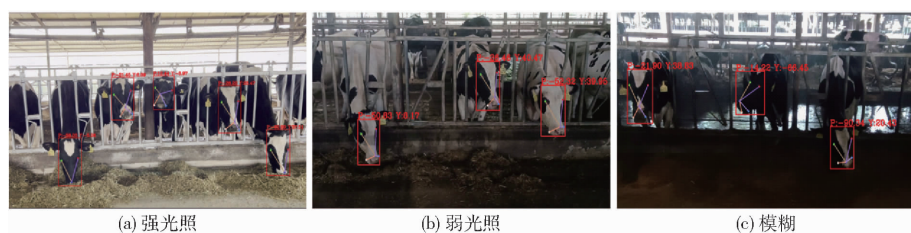


图 8 牛脸姿态估计结果

Fig. 8 Cow face pose estimation results

表5 奶牛头部姿态估计试验结果

Tab.5 Experimental results of head pose estimation of cows

姿态	AP	AR
抬头	95.5	91.2
低头	86.5	93.0
左转头	87.5	93.9
右转头	94.0	87.7

5 结论

(1)在 YOLO v7 - Pose 的模型中重构 SPPCSPC 模块为 SPPFCSPCL,使最大池化的并联结构变为串联结构,提高了计算速度并降低感受野,更有利于牛脸特征的提取。使用 WingLoss 函数以提高训练过程中应对小范围误差的能力,避免 OKS 损失函数造成的离群点干扰。引入 SPPFCSPCL 模块以及

WingLoss 损失函数后关键点检测 AP、AP0.5、AR 和 AR0.5 分别提升 3.2、1.6、2.7、0.6 个百分点。

(2) 为了将模型轻量化, 试验中使用 GhostConv 卷积以及 L1、Lamp、Random、Slim 剪枝算法, 试验结果表明, 使用 L1 剪枝可减少模型参数和计算量, 并提升了牛脸目标检测精度和关键点检测精度。基于 L1 剪枝后的模型对采食中的牛脸部的关键点坐标进行预测, 结合相机内参等, 计算出牛头俯仰角、偏转角, 最后识别出牛脸姿态。

(3)改进的 YOLO v7-Pose 模型经过 L1 剪枝之后,相较于原模型,关键点检测的 AP、AP0.5、AR 和 AR0.5 分别提升 5、2.7、3.8、1.2 个百分点。YOLO v7-SCLWL-Pose 可以有效地提升弱光照、模糊、遮挡等情况下牛脸检测效果。试验验证了牛脸姿态识别的可行性,为奶生采食行为分析提供了技术支撑。

参考文献

- [1] 何东健,宋子琪.基于热红外成像与骨架树模型的奶牛眼温自动检测[J].农业机械学报,2021,52(3):243-250.
HE Dongjian, SONG Ziqi. Automatic detection of dairy cow's eye temperature based on thermal infrared imaging technology and skeleton tree model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(3): 243-250. (in Chinese)
- [2] WANG Z, ZHANG J, DUAN C, et al. Vehicle detection in severe weather based on pseudo-visual search and HOG-LBP feature fusion[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2022, 236(7): 1607-1618.
- [3] NIU B, GAO Z, GUO B. Facial expression recognition with LBP and ORB features[J]. Computational Intelligence and Neuroscience, 2021(1): 8828245.
- [4] 夏晓华,赵倩,向华涛,等.多聚焦图像离焦模糊区域的SIFT特征提取[J].光学精密工程,2023,31(24):3630-3639.
XIA Xiaohua, ZHAO Qian, XIANG Huatao, et al. SIFT feature extraction method for the defocused blurred area of multi-focus images[J]. Optics and Precision Engineering, 2023, 31(24): 3630-3639. (in Chinese)
- [5] 蔡骋,宋肖肖,何进荣.基于计算机视觉的牛脸轮廓提取算法及实现[J].农业工程学报,2017,33(11):171-177.
CAI Cheng, SONG Xiaoxiao, HE Jinrong. Algorithm and realization for cattle face contour extraction based on computer vision[J]. Transactions of the CSAE, 2017, 33(11): 171-177. (in Chinese)
- [6] MAJI D, NAGORI S, MATHEW M, et al. YOLO-pose: enhancing YOLO for multi person pose estimation using object keypoint similarity loss[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 2637-2646.
- [7] DONG C, TANG Y, ZHANG L. MDA-YOLO Person: a 2D human pose estimation model based on YOLO detection framework[J]. Cluster Computing, 2024, 27: 12323-12340.
- [8] 刘莫尘,褚镇源,崔明诗,等.基于改进YOLOv8-Pose的红熟期草莓识别和果柄检测[J].农业机械学报,2023,54(增刊2):244-251.
LIU Mochen, CHU Zhenyuan, CUI Mingshi, et al. Red ripe strawberry recognition and stem detection based on improved YOLO v8-Pose[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(Supp. 2): 244-251. (in Chinese)
- [9] 黄小平,冯涛,郭阳阳,等.基于改进YOLOv5s的轻量级奶牛体况评分方法[J].农业机械学报,2023,54(6):287-296.
HUANG Xiaoping, FENG Tao, GUO Yangyang, et al. Lightweight dairy cow body condition scoring method based on improved YOLO v5s[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(6): 287-296. (in Chinese)
- [10] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLO v7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [11] CHOI E, BAHADORI M T, SUN J, et al. Retain: an interpretable predictive model for healthcare using reverse time attention mechanism[J]. Advances in Neural Information Processing systems, 2016, 29: 3512-3520.
- [12] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5693-5703.
- [13] JIANG T, LU P, ZHANG L, et al. RtmPose: real-time multi-person pose estimation based on mmpose[J]. arXiv Preprint arXiv:2303.07399, 2023.
- [14] 齐咏生,焦杰,鲍腾飞,等.基于自适应注意力机制的复杂场景下牛脸检测算法[J].农业工程学报,2023,39(14):173-183.
QI Yongsheng, JIAO Jie, BAO Tengfei, et al. Cattle face detection algorithm in complex scenes using adaptive attention mechanism[J]. Transactions of the CSAE, 2023, 39(14): 173-183. (in Chinese)
- [15] 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.
- [16] 郭阳阳,洪文浩,丁屹,等.基于坐标注意力机制和YOLOv5s模型的山羊脸部检测方法[J].农业机械学报,2023,54(7):313-321.
GUO Yangyang, HONG Wenhao, DING Yi, et al. Goat face detection method by combining coordinate attention mechanism and YOLO v5s model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(7): 313-321. (in Chinese)
- [17] 宋怀波,华志新,马宝玲,等.基于SimCC-ShuffleNetV2的轻量化奶牛关键点检测方法[J].农业机械学报,2023,54(10):275-281,363.
SONG Huaibo, HUA Zhixin, MA Baoling, et al. Lightweight keypoint detection method of dairy cow based on SimCC-ShuffleNetV2[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(10): 275-281, 363. (in Chinese)
- [18] LI X, SUN K, FAN H, et al. Real-time cattle pose estimation based on improved RTMPose[J]. Agriculture, 2023, 13(10): 1938.
- [19] 黄志杰,徐爱俊,周素茵,等.融合重参数化和注意力机制的猪脸关键点检测方法[J].农业工程学报,2023,39(12):141-149.
HUANG Zhijie, XU Aijun, ZHOU Suyin, et al. Keypoint detection method for pig face fusing reparameterization and attention mechanisms[J]. Transactions of the CSAE, 2023, 39(12): 141-149. (in Chinese)
- [20] FENG Z H, KITTLER J, AWAIS M, et al. Wing loss for robust facial landmark localisation with convolutional neural networks[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2235-2245.

of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2818 – 2826.

- [23] GE W. Deep metric learning with hierarchical triplet loss[C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 269 – 285.
- [24] 赵晋陵, 詹媛媛, 王娟, 等. 基于 SE – UNet 的冬小麦种植区域提取方法[J]. 农业机械学报, 2022, 53(9): 189 – 196. ZHAO Jinling, ZHAN Yuanyuan, WANG Juan, et al. SE – UNet-based extraction of winter planting areas[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(9): 189 – 196. (in Chinese)
- [25] 汪亚航, 宋晓宁, 吴小俊. 结合混合池化的双流人脸活体检测网络[J]. 中国图象图形学报, 2020, 25(7): 1408 – 1420. WANG Yahang, SONG Xiaoning, WU Xiaojun. Two-stream face spoofing detection network combined with hybrid pooling[J]. Journal of Image and Graphics, 2020, 25(7): 1408 – 1420. (in Chinese)
- [26] JAYARAJ R, LOKESH S. Automatic image annotation using adaptive convolutional deep learning model[J]. Intelligent Automation & Soft Computing, 2023, 36(1): 481 – 497.
- [27] PENG S, HUANG H, CHEN W, et al. More trainable inception-ResNet for face recognition[J]. Neurocomputing, 2020, 411: 9 – 19.
- [28] MANDAL B, OKEUKWU A, THEIS Y. Masked face recognition using Resnet – 50[J]. arXiv Preprint, arXiv:2104.08997, 2021.
- [29] 张宏鸣, 周利香, 李永恒, 等. 基于改进 MobileFaceNet 的羊脸识别方法[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)

(上接第 92 页)

- [21] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C] // 13th European Conference. Springer International Publishing, 2014: 740 – 755.
- [22] LI H, KADAV A, DURDANOVIC I, et al. Pruning filters for efficient convnets[C] // 5th International Conference on Learning Representations, 2017: 149804.
- [23] YANG L, XU X, ZHAO J, et al. Fusion of RetinaFace and improved FaceNet for individual cow identification in natural scenes[J/OL]. Information Processing in Agriculture, 2023. <http://doi.org/10.1016/j.inpa.2023.09.001>.
- [24] 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.
- [25] LIU W, LIAO S, REN W, et al. High-level semantic feature detection: a new perspective for pedestrian detection[C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5187 – 5196.
- [26] ZHENG Zhaohui, WANG Ping, LIU Wei, et al. Distance – IoU loss: faster and better learning for bounding box regression [C] // 34th AAAI Conference on Artificial Intelligence, 2020: 12993 – 13000.
- [27] 王海鹏, 王正良, 许威威, 等. 基于三维模型的 Android 手机端人脸姿态实时估计系统[J]. 计算机应用, 2015, 35(8): 2321 – 2326. WANG Haipeng, WANG Zhengliang, XU Weiwei, et al. Real-time face pose estimation system based on 3D face model on Android mobile platform[J]. Journal of Computer Applications, 2015, 35(8): 2321 – 2326. (in Chinese)
- [28] 陈科尹, 邹湘军, 关卓怀, 等. 基于混合蛙跳优化的采摘机器人相机标定方法[J]. 农业机械学报, 2019, 50(1): 23 – 34. CHEN Keyin, ZOU Xiangjun, GUAN Zhuohuai, et al. Camera calibration method of picking robot based on shuffled frog leaping optimization[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(1): 23 – 34. (in Chinese)
- [29] HAN K, WANG Y, TIAN Q, et al. Ghostnet: more features from cheap operations[C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580 – 1589.
- [30] LIU Z, LI J, SHEN Z, et al. Learning efficient convolutional networks through network slimming[C] // 16th IEEE International Conference on Computer Vision, 2017: 2736 – 2744.
- [31] LEE J, PARK S, MO S, et al. Layer-adaptive sparsity for the magnitude-based pruning[J]. arXiv Preprint arXiv:2010.07611, 2020.
- [32] MITTAL D, BHARDWAJ S, KHAPRA M M, et al. Recovering from random pruning: on the plasticity of deep convolutional neural networks[C] // IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018: 848 – 857.
- [33] FU R, HU Q, DONG X, et al. Axiom-based grad-cam: towards accurate visualization and explanation of CNNs[J]. arXiv Preprint arXiv:2008.02312, 2020.
- [34] 杨慧, 张思源, 彭容, 等. 荷斯坦犊牛反刍行为的发育规律及其与采食之间的相关性研究[J]. 动物营养学报, 2022, 34(3): 1614 – 1622. YANG Hui, ZHANG Siyuan, PENG Rong, et al. Research on developmental regularity of ruminating behavior in Holstein calves and its correlation with feed intake[J]. Chinese Journal of Animal Nutrition, 2022, 34(3): 1614 – 1622. (in Chinese)
- [35] GRELET C, DRIES V V, LEBLOIS J, et al. Identification of chronic stress biomarkers in dairy cows[J]. Animal, 2022, 16(5): 100502.