

doi:10.6041/j.issn.1000-1298.2021.03.028

基于YOLOv4的猪只饮食行为检测方法

李菊霞¹ 李艳文¹ 牛帆¹ 李荣¹ 张韬¹ 景冰²

(1. 山西农业大学信息科学与工程学院, 太谷 030801; 2. 山西财贸职业技术学院计算机系, 太原 030031)

摘要: 针对猪舍环境下猪只饮食行为自动化检测程度较低的问题, 提出了一种基于YOLOv4的猪只饮食行为检测模型。基于多时间段、多视角和不同程度遮挡的猪只饮食图像, 建立了猪只饮食行为图像数据库, 利用YOLOv4深度学习网络的深层次特征提取、高精度检测分类特性, 对猪只饮食行为进行检测。结果表明, 基于YOLOv4的猪只饮食行为检测模型在不同视角、不同遮挡程度以及不同光照下均能准确预测猪只的饮食行为, 在测试集中平均检测精度(mAP)达到95.5%, 分别高于YOLOv3、Tiny-YOLOv4模型2.8、3.6个百分点, 比Faster R-CNN模型高1.5个百分点, 比RetinaNet、SSD模型高5.9、5个百分点。本文方法可为智能养猪与科学管理提供技术支撑。

关键词: 猪只饮食行为; 目标检测; YOLOv4; 图像处理

中图分类号: TP391.41 文献标识码: A 文章编号: 1000-1298(2021)03-0251-06

OSID:



Pig Diet Behavior Detection Method Based on YOLOv4

LI Juxia¹ LI Yanwen¹ NIU Fan¹ LI Rong¹ ZHANG Tao¹ JING Bing²

(1. College of Information Science and Engineering, Shanxi Agricultural University, Taigu 030801, China

2. Computer Department, Shanxi Vocational College of Finance and Trade, Taiyuan 030031, China)

Abstract: It is of vital significance to detect pig's eating and drinking behavior by using intelligent method, and analyze the law of eating and drinking water, which plays an important role in early warning of pig disease and maintaining pig welfare. Pig diet behavior detection model based on YOLOv4 was proposed. Aiming at the pig diet image with multi time period, multi view angle and different degrees of occlusion, the database of pig eating behavior image was established. The in-depth feature extraction and high-precision detection classification characteristics of YOLOv4 deep learning network were used to accurately detect pig eating behavior. The results from the whole experiments showed that the model based on YOLOv4 can accurately predict the diet behavior of pigs in different angles of view, different degrees of occlusion and different illuminations. The average detection accuracy (mAP) was 95.5%, which was 2.8 percentage points and 3.6 percentage points higher than that of the same series of YOLOv3 and Tiny-YOLOv4 models, 1.5 percentage points higher than that of Faster R-CNN model, 5.9 percentage points higher than that of RetinaNet model and 5 percentage points higher than that of SSD model. This method can accurately predict the occurrence of pig eating behavior and provide targeted and adaptive technical support for pig intelligent breeding and management.

Key words: pig diet behavior; target detection; YOLOv4; image processing

0 引言

猪只饮食行为是判断其健康状况的重要依据, 科学的猪只饮食管理有利于猪只健康。实时、连续、准确地监测猪只进食、饮水状况, 对预防猪只疾病、

提高养殖福利具有重要意义^[1]。

目前, 常见的牲畜进食、饮水行为检测方法涉及无线射频技术和机器视觉技术。无线射频技术需要给猪只佩戴耳标类传感器, 虽然能记录各项参数, 但传感器容易被动物咬食而损坏^[2]; 机器视觉技术具

收稿日期: 2020-09-16 修回日期: 2020-12-29

基金项目: 国家自然科学基金面上项目(31772651)和山西省重点研发计划专项(农业)(201803D221028-7)

作者简介: 李菊霞(1977—), 女, 副教授, 主要从事农业信息化与图像处理研究, E-mail: lijxsn@126.com

通信作者: 李艳文(1987—), 男, 讲师, 主要从事机器学习研究, E-mail: lywlive@163.com

有无接触、低成本等优点,在猪只养殖过程中得到大量应用。随着卷积神经网络(Convolutional neural network, CNN)的发展,Mask R-CNN^[3]、Faster R-CNN^[4]、YOLO、SSD^[5]等用于目标检测的深度学习模型在猪只行为检测和猪只姿态检测等方面得到广泛应用^[6-10]。相较于其他模型,基于回归思想的YOLO系列^[11-13]模型在确保检测精度的情况下提升了检测速度,适用于规模化生产环境,已被应用于苹果^[14]、鸡^[15]、奶牛^[16]等目标检测。

本文基于不同视角、不同遮挡程度以及不同光照强度的猪只饮食图像,建立猪只饮食行为的图像数据库,采用YOLOv4^[17]模型对猪只饮食行为进行检测,将猪只进食、饮水行为图像数据输入到深度学习网络中,利用YOLOv4网络的深层次特征提取、高精度检测分类特性,对猪只饮食行为进行准确检测。

1 数据采集与预处理

1.1 数据采集

实验数据采集于山西省临汾市汾西县猪只培育基地,选取4只成年公猪放入约3 m × 2.2 m × 2.0 m圈养环境中。于2020年6月25日起,上、下午以及夜间分别拍摄时长3 h视频,连续拍摄7 d。采用Allied Vision Technologies的Manta G-282C型相机移动拍摄。拍摄时,猪只距离镜头0.5~3 m不等,在不同视角、不同遮挡程度和不同位置情况下拍摄视频。

1.2 数据预处理

将采集到的数据进行筛选、增强和标注。

采用KMPlayer软件进行视频截取,将视频按1 f/s进行JPG格式的图像提取,将得到的图像进行筛选。为了增加模型训练数据的丰富度,所选样本包含不同时间段、不同光照强度和不同拍摄视角的样本,均匀选出2 000幅图像作为训练和测试样本数据。将样本尺寸统一调整为608像素×608像素,以减少运算量,提高模型训练速度。

为了避免本文所选6种模型在训练过程中出现过拟合现象,通过cutout、cutmix、mixup等图像增强技术扩展样本图像,如图1所示,使正样本和负样本数量达到2:1。在后续YOLOv4模型训练过程中并

未使用模型自带的cutmix数据增强方法。

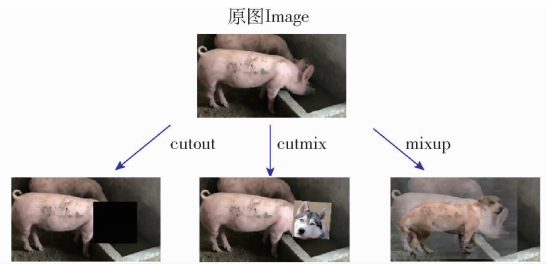


图1 数据增强

Fig. 1 Data augmentation

对样本图像增强后,使用LabelImg图像标注工具对猪只的饮食行为进行标注,如图2所示。使用PASCAL VOC^[18]格式的XML文件注释。

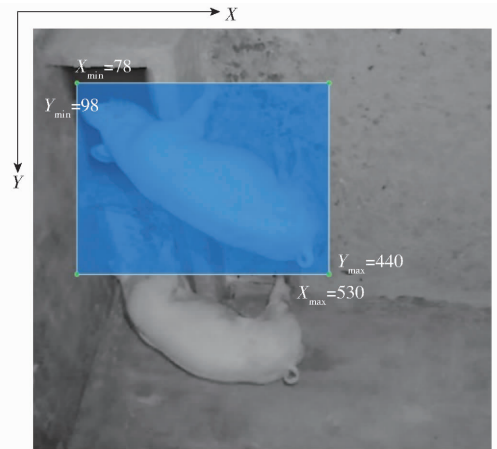


图2 数据标注

Fig. 2 Data annotation

2 YOLOv4 算法

2.1 YOLOv4 网络架构

对于检测模型来说,分类最优的主干网络不一定是合适的,适用于目标检测的主干网络需满足:①输入高分辨率图像,提高小目标的检测准确率。②更多的网络层,提高感受域来适应输入的增加。③更多的参数,提高检测多尺寸目标的能力。对比3种主干网络的参数,如表1所示,CSPDarknet53的参数量和传输速度优于其他主干网络,因此选择CSPDarknet53作为YOLOv4主干网络。

YOLOv4在CSPDarknet53中引入了空间金字塔池化模块(Spatial pyramid pooling, SPP)^[19],能够显著地改善感受域尺寸,将最重要的上下位特征提取

表1 用于图像分类的神经网络参数对比

Tab. 1 Parameters comparison of neural networks for image classification

主干网络	输入图像分辨率/ (像素×像素)	感受域/ (像素×像素)	参数量/ MB	输出层均值/ kB	传输速度/ (f·s ⁻¹)
CSPResNet50	512 × 512	425 × 425	20.6	1 058	62
CSPDarknet53	512 × 512	725 × 725	27.6	950	66
EfficientNet-B3	512 × 512	1 311 × 1 311	12.0	668	26

出来,网络处理速度没有明显下降。此外, YOLOv4 还使用路径聚合网络(Path aggregation network, PAN)^[20] 替换特征金字塔网络(Feature pyramid network, FPN)^[21] 进行多通道特征融合。如图 3 所示, FPN 下方

特征信息需要经过多层网络才能到达金字塔上层,途中有可能会丢失信息。PAN 网络结构中加入了自底向上的路径增强,可避免信息丢失问题,经过特征图拼接后获得的信息既有底层特征也有语义特征。

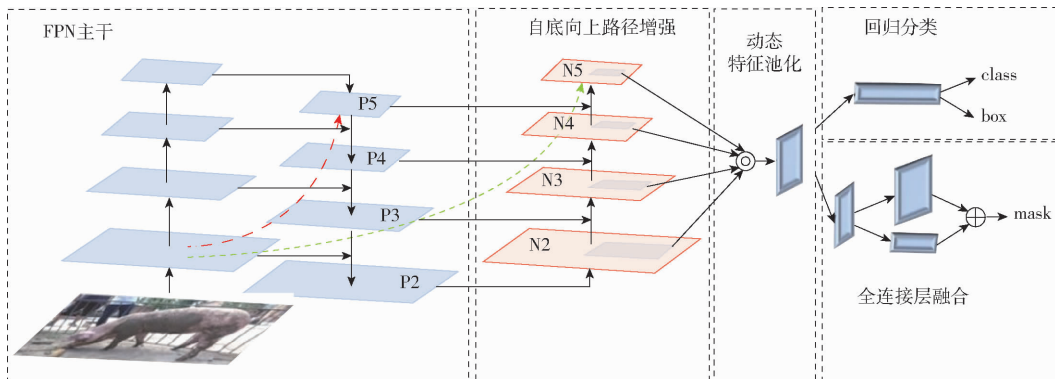


图 3 PAN 网络

Fig. 3 PAN network

最终,选择了 CSPDarknet53 作为主干网络, PAN 路径聚合网络作为颈部, YOLOv3 作为头部, 组成了 YOLOv4 整体结构。

2.2 YOLOv4 组件

2.2.1 空间注意力模块

空间注意力模块对输入特征图进行最大池化(Max - Pooling)和平均池化(Average - Pooling)操作,从而得到两个特征图集合。将其结果输入卷积层,之后再由一个 Sigmoid 函数创建出空间注意力。YOLOv4 对空间注意力模块(SAM)进行简单修改,将空间注意力机制(Spatial - wise attention)修改为点注意力机制(Point - wise attention),并将 PAN 中的快捷方式连接替换为串联,如图 4 所示。

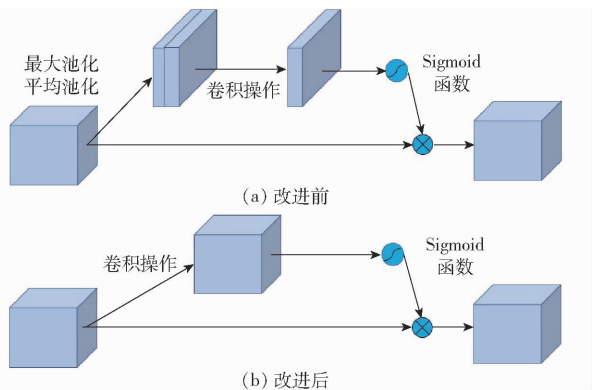


图 4 改进前后的 SAM

Fig. 4 Modified SAM

2.2.2 Box 回归函数

YOLOv4 使用全局交并比(Complete - IoU, CIoU)^[22] 作为 box 回归函数, CIoU 可以在矩形框(Bounding - box, Bbox)回归问题上获得更好的收敛速度和精度。CIoU 的惩罚项 R_{CIoU} 为

$$R_{CIoU} = \frac{\rho^2(b, b^{gt})}{c^2} + av \tag{1}$$

其中
$$a = \frac{v}{1 - I_{oU} + v} \tag{2}$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \tag{3}$$

- 式中
- ρ ——两个中心点间的欧氏距离
 - b ——预测框中心点
 - b^{gt} ——真实框中心点
 - c ——两个中心点最小外接矩的对角线长度
 - a ——trade - off 的参数
 - v ——长宽比的相似性
 - I_{oU} ——预测框与真实框的交并比
 - w^{gt} ——真实框宽度
 - h^{gt} ——真实框高度
 - w ——预测框宽度
 - h ——预测框高度

CIoU 损失函数 L_{CIoU} 定义为

$$L_{CIoU} = 1 - I_{oU} + \frac{\rho^2(b, b^{gt})}{c^2} + av \tag{4}$$

CIoU 损失函数的梯度类似于距离交并比(DIoU)损失函数的梯度,但还要考虑 v 的梯度,当长宽在 $[0, 1]$ 之间时, $w^2 + h^2$ 通常很小,会导致梯度爆炸,因此将 $\frac{1}{w^2 + h^2}$ 替换成 1。

2.3 评价指标

在机器学习中的目标检测领域,精确率-召回率(Precision - Recall, P - R)曲线用于衡量目标检测算法的性能。此外,用于衡量目标检测算法性能的常用指标还有平均检测精度(mAP)。检测精度(AP)指 P - R 曲线下方的区域面积, mAP 指同一模型对

猪只进食行为和饮水行为的 AP 平均值。

本文采用 P-R 曲线以及 mAP 作为评价指标来衡量模型对猪只饮食行为的检测性能。

3 实验与结果分析

图像数据集共包括 3 000 幅图像。进食、饮水行为各 1 250 幅,其他行为(排泄、休息等)共 500 幅。从该数据集中随机选择进食、饮水行为图像各 1 050 幅,其他行为数据 400 幅,共 2 500 幅图像作为训练集,其余 500 幅图像作为测试集。猪只行为检测模型在 NVIDIA GTX 1650 GPU 上进行训练和测试。

3.1 模型训练

YOLOv4 在训练之前需要设置相关参数,本实验将批处理尺寸设置为 4,最大迭代数设置为 4 000,动量为 0.949,权重衰减正则系数为 5×10^{-4} 。本文采用学习率机制,初始学习率为 1×10^{-3} ,一旦学习停滞,模型学习率通常将以 2~10 倍的速率衰减,衰减次数设置为最大迭代数的 80% 和 90%。即模型迭代到 3 200 次时学习率衰减至原学习率的 1/10,迭代到 3 600 次时,学习率会在前一个学习率基础上继续衰减至其 1/10。

3.2 结果分析

本文以相同的实验数据分别使用 YOLOv4、YOLOv3、Tiny-YOLOv4、Faster R-CNN、RetinaNet

和 SSD 模型进行训练,在相同测试集下进行测试对比。图 5 为部分测试结果,紫色框表示进食行为,绿色框表示饮水行为。图 5a 为侧视光照充足情况下猪只进食行为的检测结果,图 5b 为侧视光照充足情况下猪只饮水行为的检测结果;图 5c 为俯视光照不足情况下猪只进食、饮水行为同时发生的检测结果;图 5d 为俯视光照不足情况下遮挡状态猪只发生进食行为的检测结果。

由图 5a、5b 可知,6 种模型在光照充足的情况下均可以正确检测出猪只进食、饮水行为。在猪只进食行为检测结果中,YOLOv4 检测结果的置信度(预测类别为真实类别的概率)最高。在检测猪只饮水行为的结果中,6 种模型取得的置信度相差不大。由图 5c 可知,在光照不足以及进食行为、饮水行为同时发生的情况下,YOLOv3、Tiny-YOLOv4 以及 Faster R-CNN 3 种模型只能检测出饮水行为,而 YOLOv4、RetinaNet、SSD 模型对饮水行为和进食行为均能做出正确检测。由图 5d 可知,YOLOv4 可以正确检测出被遮挡状态的猪只进食行为,而其他 5 种模型均未能检测出被遮挡情况下的猪只进食行为。

由表 2 可知,YOLOv4 检测模型的 mAP 值高于 YOLOv3 模型 2.8 个百分点,高于 Tiny-YOLOv4 模型 3.6 个百分点,高于 Faster R-CNN 模型 1.5 个百分点,高于 RetinaNet 模型 5.9 个百分点,高于 SSD

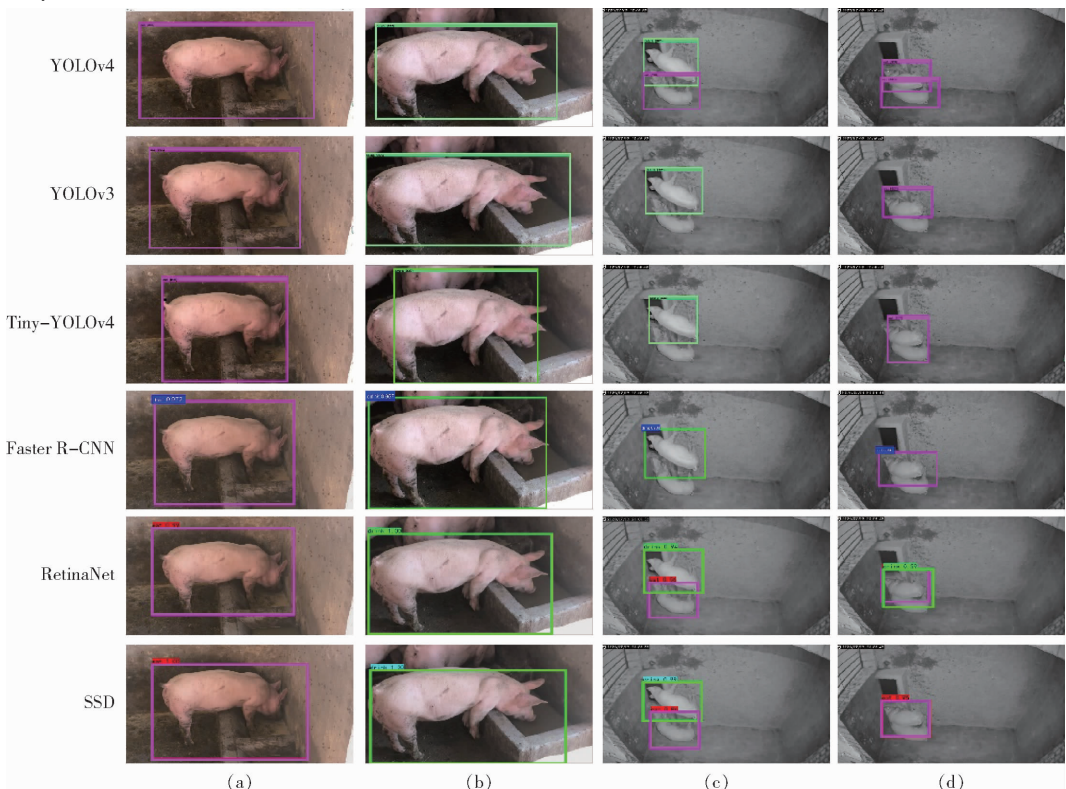


图 5 部分测试结果

Fig. 5 Partial forecast results

表2 不同模型对测试集识别的AP值和mAP值

Tab.2 AP and mAP of test set by different models

模型	%		
	AP		mAP
	进食行为	饮水行为	
YOLOv4	94.5	96.4	95.5
YOLOv3	90.3	95.1	92.7
Tiny - YOLOv4	90.9	92.9	91.9
Faster R - CNN	93.4	94.5	94.0
RetinaNet	90.7	88.6	89.6
SSD	90.8	90.3	90.5

模型5个百分点。YOLOv4对于猪只进食行为的检测精度比同系列的YOLOv3、Tiny - YOLOv4分别高4.2、3.6个百分点,对于饮水行为的检测精度则比YOLOv3、Tiny - YOLOv4高1.3、3.5个百分点。对比Faster R - CNN, YOLOv4模型对进食行为和饮水行为的检测精度分别提升了1.1、1.9个百分点,对比RetinaNet提升了3.8、7.8个百分点,对比SSD提升了3.7、6.1个百分点。

图6为YOLOv4、YOLOv3、Tiny - YOLOv4、SSD、Faster R - CNN、RetinaNet模型对猪只饮食行为检测的P - R曲线。由表2可知,6种模型对于进食和饮水两种行为的AP值均在90%左右,P - R曲线基本覆盖了整个坐标系。相较于其他5种模型的P - R曲线,YOLOv4模型的P - R曲线均位于最上方,优于其他5种模型。

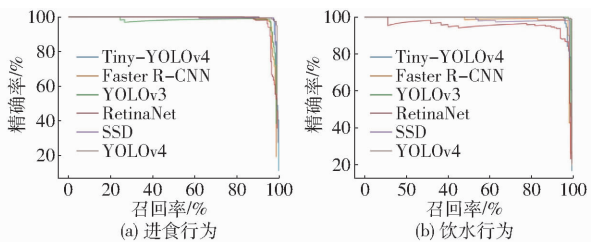


图6 P - R曲线

Fig.6 P - R curves

除上述实验外,本文还选取了768幅被遮挡状态猪只饮食图像,利用上述6种模型对其进行猪只饮食行为检测,实验结果如表3所示。

由表3可知,在被遮挡状态下,YOLOv4模型对猪只进食行为的检测精度达到85.8%,高于同系列的YOLOv3、Tiny - TOLOv4模型52.5、42.2个百分点,高于Faster R - CNN、RetinaNet、SSD模型24.3、44.8、31.1个百分点。YOLOv4模型对于猪只饮水

表3 被遮挡状态测试集在不同模型下的AP值和mAP值

Tab.3 AP and mAP of adhesive state test set by

模型	%		
	AP		mAP
	进食行为	饮水行为	
YOLOv4	85.8	93.8	89.8
YOLOv3	33.3	69.5	51.4
Tiny - TOLOv4	43.6	88.9	70.8
Faster R - CNN	61.5	82.8	72.2
RetinaNet	41.0	30.2	35.7
SSD	54.7	89.2	72.0

行为的检测精度达到93.8%,高于同系列的YOLOv3、Tiny - TOLOv4模型24.3、4.9个百分点,高于Faster R - CNN、RetinaNet、SSD模型11、63.6、4.6个百分点。YOLOv4模型的mAP值高于YOLOv3模型38.4个百分点,高于Tiny - YOLOv4模型19个百分点,高于Faster R - CNN模型17.6个百分点,高于RetinaNet模型54.1个百分点,高于SSD模型17.8个百分点。相较于表2的检测精度,被遮挡状态下的进食、饮水检测精度都有明显下降。这是因为在被遮挡状态下,位于检测设备前方的猪只会遮挡后方猪只的部分行为特征,从而致使检测精度下降。在本实验环境中,饮水槽位于进食槽前方,发生饮水行为的猪只会遮挡处于进食状态的猪只,所以表3中的猪只进食行为检测精度下降幅度远大于饮水行为。实验结果表明,YOLOv4模型在被遮挡状态下对猪只饮食行为的检测精度远优于其他5种模型。

4 结论

(1) YOLOv4模型在侧视、俯视、被遮挡状态以及光照不足的情况下均能实时有效地检测猪只饮食行为,检测效果优于同系列YOLOv3、Tiny - YOLOv4模型,以及主流检测模型Faster R - CNN、RetinaNet和SSD。

(2) YOLOv4模型在本文测试集中平均检测精度达到95.5%,分别高于YOLOv3、Tiny - YOLOv4、Faster R - CNN、RetinaNet、SSD模型2.8、3.6、1.5、5.9、5个百分点。

(3) YOLOv4模型对各测试集的检测精度(AP)、平均检测精度(mAP)以及P - R曲线均优于YOLOv3、Tiny - YOLOv4、SSD、Faster R - CNN和RetinaNet模型。

参 考 文 献

- [1] 薛佳俐,杨曙明. 生猪福利养殖系统的研究进展[J]. 农产品质量与安全, 2020(2): 64 - 69, 89. XUE Jiali, YANG Shuming. Research progress on welfare breeding system for pigs: a review [J]. Quality and Safety of Agro-products, 2020(2): 64 - 69, 89. (in Chinese)
- [2] 杨秋妹,肖德琴,张根兴. 猪只饮水行为机器视觉自动识别[J/OL]. 农业机械学报, 2018, 49(6): 232 - 238.

- YANG Qiumei, XIAO Deqin, ZHANG Genxing. Automatic pig drinking behavior recognition with machine vision [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(6): 232–238. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20180627&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2018.06.027. (in Chinese)
- [3] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961–2969.
- [4] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems, 2015: 91–99.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//European Conference on Computer Vision. Springer, Cham, 2016: 21–37.
- [6] 李丹,张凯锋,李行健,等.基于Mask R-CNN的猪只爬跨行为识别[J/OL].农业机械学报,2019,50(增刊):261–266,275. LI Dan, ZHANG Kaifeng, LI Xingjian, et al. Mounting behavior recognition for pigs based on Mask R-CNN [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(Supp.): 261–266, 275. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=2019s041&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2019.S0.041. (in Chinese)
- [7] 谢涛,王芳,田建艳,等.基于快速SSD的猪群舒适度监测[J].畜牧与兽医,2019,51(12):40–45. XIE Tao, WANG Fang, TIAN Jianyan, et al. Pig comfort monitoring based on fast SSD[J]. Animal Husbandry and Veterinary Medicine, 2019, 51(12): 40–45. (in Chinese)
- [8] 李泊,沈明霞,刘龙申,等.基于YOLO v3与图结构模型的群养猪只头尾辨别方法[J/OL].农业机械学报,2020,51(7):44–51. LI Bo, SHEN Mingxia, LIU Longshen, et al. Head and tail identification method for group-housed pigs based on YOLO v3 and pictorial structure model[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(7): 44–51. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20200706&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2020.07.006. (in Chinese)
- [9] 薛月菊,朱勋沐,郑婵,等.基于改进Faster R-CNN识别深度视频图像哺乳母猪姿态[J].农业工程学报,2018,34(9):189–196. XUE Yueju, ZHU Xunmu, ZHENG Chan, et al. Lactating sowpostures recognition from depth image of videos based on improved Faster R-CNN[J]. Transactions of the CSAE, 2018, 34(9): 189–196. (in Chinese)
- [10] 嵇杨培,杨颖,刘刚.基于可见光谱和YOLOv2的生猪饮食行为识别[J].光谱学与光谱分析,2020,40(5):1588–1594. JI Yangpei, YANG Ying, LIU Gang. Recognition of pig eating and drinking behavior based on visible spectrum and YOLOv2 [J]. Spectroscopy and Spectral Analysis, 2020, 40(5): 1588–1594. (in Chinese)
- [11] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society: Piscataway, 2016: 779–788.
- [12] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263–7271.
- [13] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv preprint arXiv, 2018:1804.02767.
- [14] 夏雪,孙琦鑫,侍啸,等.基于轻量级无锚点深度卷积神经网络的树上苹果检测模型[J].智慧农业,2020,2(1):99–110. XIA Xue, SUN Qixin, SHI Xiao, et al. Apple detection model based on lightweight anchor-free deep convolutional neural network[J]. Smart Agriculture, 2020, 2(1): 99–110. (in Chinese)
- [15] WANG J, WANG N, LI L, et al. Real-time behavior detection and judgment of egg breeders based on YOLO v3[J]. Neural Computing and Applications, 2020, 32(10): 5471–5481.
- [16] 何东健,刘建敏,熊虹婷,等.基于改进YOLO v3模型的挤奶奶牛个体识别方法[J/OL].农业机械学报,2020,51(4):250–260. HE Dongjian, LIU Jianmin, XIONG Hongting, et al. Individual identification of dairy cows based on improved YOLO v3 [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(4): 250–260. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20200429&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2020.04.029. (in Chinese)
- [17] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[J]. arXiv preprint arXiv, 2020:2004.10934.
- [18] EVERINGHAM M, GOOL L V, CHRISTOPHER K L, et al. The pascal visual object classes (VOC) challenge [J]. International Journal of Computer Vision, 2010, 88:303–308.
- [19] 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.
- [20] 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.
- [21] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117–2125.
- [22] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//AAAI, 2020: 12993–13000.