

采用改进 CenterNet 模型检测群养生猪目标

房俊龙, 胡宇航, 戴百生^{*}, 吴志东

(1. 东北农业大学电气与信息学院, 哈尔滨 150030; 2. 农业农村部生猪养殖设施工程重点实验室, 哈尔滨 150030)

摘 要: 为实现对群养环境下生猪个体目标快速精准的检测, 该研究提出了一种针对群养生猪的改进型目标检测网络 MF-CenterNet (MobileNet-FPN-CenterNet) 模型, 为确保目标检测的精确度, 该模型首先以无锚式的 CenterNet 为基础结构, 通过引入轻量级的 MobileNet 网络作为模型特征提取网络, 以降低模型大小和提高检测速度, 同时加入特征金字塔结构 FPN (Feature Pyramid Networks) 以提高模型特征提取能力, 在保证模型轻量化、实时性的同时, 提高遮挡目标和小目标的检测精度。以某商业猪场群养生猪录制视频作为数据源, 采集视频帧 1 683 张, 经图像增强后共得到 6 732 张图像。试验结果表明, MF-CenterNet 模型大小仅为 21 MB, 满足边缘计算端的部署, 同时对生猪目标检测平均精确度达到 94.30%, 检测速度达到 69 帧/s, 相较于 Faster-RCNN、SSD、YOLOv3、YOLOv4 目标检测网络模型, 检测精度分别提高了 6.39、4.46、6.01、2.74 个百分点, 检测速度分别提高了 54、47、45、43 帧/s, 相关结果表明了该研究所提出的改进型的轻量级 MF-CenterNet 模型, 能够在满足目标检测实时性的同时提高对群养生猪的检测精度, 为生产现场端的群养生猪行为实时检测与分析提供了有效方法。

关键词: 计算机视觉; 图像处理; 群养生猪; 目标检测; MobileNet; FPN; CenterNet

doi: 10.11975/j.issn.1002-6819.2021.16.017

中图分类号: TP391

文献标志码: A

文章编号: 1002-6819(2021)-16-0136-09

房俊龙, 胡宇航, 戴百生, 等. 采用改进 CenterNet 模型检测群养生猪目标[J]. 农业工程学报, 2021, 37(16): 136-144.

doi: 10.11975/j.issn.1002-6819.2021.16.017 <http://www.tcsae.org>

Fang Junlong, Hu Yuhang, Dai Baisheng, et al. Detection of group-housed pigs based on improved CenterNet model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(16): 136-144. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2021.16.017 <http://www.tcsae.org>

0 引 言

目前在中国生猪养殖过程中, 对生猪的质量体态、冷热应激、饮食排泄、攻击咬尾等状态和行为的监测主要以人工监测为主, 这使得生猪养殖管理成本增高, 同时还会带来生猪的应激反应, 更严重的会出现人畜交叉感染的风险^[1]。随着计算机视觉技术的发展, 使得利用该技术对群养生猪进行非接触式、低应激的健康监测成为可能, 对生猪个体目标进行检测是实现生猪自动化监测与智能化分析的前提^[1-3]。快速而又准确的生猪目标检测, 对实现精准、个性化的生猪健康智能监测具有重要意义。

近年来, 已有研究人员开始利用计算机视觉技术对生猪个体目标进行检测, 所采用的方法主要包括传统目标检测方法和基于卷积神经网络 (Convolutional Neural

Networks, CNN) 的目标检测方法。杨心等^[4]通过粘连猪体分割和基于广义 Hough 聚类的方法对生猪头、尾进行检测, 检测平均正确率达到 87.28%。马丽等^[5]经过图像处理获取猪只轮廓图, 提出联立猪只外接矩形高宽比和低频傅里叶系数构建猪只侧视图的特征向量, 利用马氏距离判别法进行检测, 检测准确率达到 91.7%。谢双云等^[6]采用 GMM 和图像粒化方法, 依据图像序列的前景图分析粒子运动属性, 融合前景图和粒化图, 最终得到精确的猪只目标检测。然而, 传统检测方法在复杂条件下检测精度不高、检测速度慢, 随着深度学习特别是 CNN 模型在图像识别和目标检测领域的成功应用, 基于 CNN 的目标检测方法也逐渐被应用到群养生猪的检测中。基于 CNN 的目标检测可分为两阶段 (Two-Stage) 和单阶段 (One-Stage) 两大类。基于 Two-Stage 的模型主要包括 R-CNN 系列^[7-8] (R-CNN, Fast R-CNN, Faster R-CNN 等), 基于单阶段的模型主要包含 YOLO (You Only Look Once) 系列^[9-12]和 SSD (Single Shot MultiBox Detector) 系列^[13]。在生猪个体检测领域, Riekert 等^[14]利用 2D 摄像机和 Faster R-CNN 检测模型对生猪的位置和姿态进行检测, 平均检测精度达到 80.2%。Yang 等^[15]通过 Faster R-CNN 模型对生猪个体以及头部位置进行检测, 再通过关联算法检测生猪进食行为, 检测准确度达到 95%以上, 召回率达到 80%以上。Cowton 等^[16]利用 Faster R-CNN 模型对生猪位置进行检测, 精度达到 90.1%, 满足试验预期效果。王婷婷等^[17]通过采用 Faster R-CNN 模型对有无遮

收稿日期: 2021-03-10 修订日期: 2021-07-21

基金项目: 国家重点研发计划项目 (2018YFD0500704); 国家自然科学基金青年科学基金项目 (31902210); 黑龙江省高校青年创新人才培养计划项目 (UNPYSCT-2018142); 黑龙江省科学基金青年科学基金项目 (QC2018074); 东农学者计划“青年才俊”项目 (18QC23); 农业部生猪养殖设施工程重点实验室开放课题 (SK201707); 财政部和农业农村部: 国家现代农业产业技术体系资助

作者简介: 房俊龙, 博士, 教授, 博士生导师。研究方向为智能检测与控制。

Email: junlongfang@126.com

^{*}通信作者: 戴百生, 副教授。研究方向为视觉智能计算、智慧畜牧。

Email: bsdai@neau.edu.cn

挡的生猪分别进行检测, 进而通过改进 Faster R-CNN 模型完成对生猪质量的评估。基于 Two-Stage 的生猪目标检测方法检测精度高但检测速度相对较慢, 难以满足实时检测要求, 且模型体量较大, 使得向嵌入式平台移植带来挑战。谢涛^[18]提出了一种基于稀疏化深度可分离网络结构的快速 SSD 生猪目标检测模型, 可以快速而精准的检测生猪姿态, 满足实时检测要求。苏恒强等^[19]针对图像光照强度、生猪种类颜色、障碍物遮挡等干扰, 提出了基于 YOLOv3 的生猪目标检测方法。燕红文等^[20]提出了将特征金字塔注意力信息融入 Tiny-YOLO 特征提取过程中, 对群养生猪进行了高精度的目标检测。Hanse 等^[21]使用 YOLO 对生猪姿态进行检测并对身体部位(头、身、臀)进行训练, 进而检测出生猪头部位置。李泊等^[22]通过 YOLOv3 检测生猪头尾目标, 并采取阈值分割与前景椭圆拟合推理缺失部位, 生猪头尾检测精度达到 96.22%。上述基于 One-Stage 的模型能够在检测精度和检测速度达到一定的平衡, 但其检测精度和检测速度仍然有待提高。

特别地, 上述目标检测模型主要采用先验框的设定, 未能较好解决模型检测时图像中目标尺寸大小以及目标相互遮挡等因素对检测结果的影响, 同时模型检测实时性不高、体积较大、需要算力较高, 难以满足向移动式终端或嵌入式设备进行移植的需求, 给实际场景的应用带来挑战。

为使得生猪目标检测模型更加易于向计算资源有限且性能较低的边缘计算硬件平台上进行部署, 在保证实时性检测的同时尽量提高检测精度, 本文有针对性的提出了一种基于 MF-CenterNet (MobileNet-FPN-CenterNet) 模型的群养生猪目标检测方法, 该方法在保证检测实时性、可移植于边缘计算端的同时, 对存在不同目标尺寸、相互遮挡等因素下的生猪个体都有较好的检测能力。

1 材料与方法

1.1 数据收集

本文所研究的生猪视频采集于黑龙江省齐齐哈尔市建华区双合村金城养殖场, 为了试验结果的普遍性, 分别于 2020 年 7 月 20 号(天气晴, 光照强)和 2020 年 9 月 27 号(多云, 光照弱)进行 2 次采集, 选取了大白、长白 2 个品种的猪作为采集对象, 猪场单个猪舍面积(长×宽)为 7 m×5 m, 每个猪舍生猪数量 10~25 头, 选取其中 4 舍日龄 80 d、质量平均约 41 kg 的育肥期的群养生猪共计 85 头作为试验对象。本文采用手持设备对不同光照和不同数量的生猪进行视频采集, 共拍摄得到时长为 20~40 min 的 10 段视频, 每个猪舍选取 2 段时长超过 30 min 的视频作为原始视频。原始视频格式为 MP4, 分辨率为(帧宽度×帧高度)1 280 像素×720 像素, 帧速率 29.9 帧/s, 本文所采集数据中, 包含不同头数生猪的监控画面统计结果, 如表 1 所示。

1.2 数据预处理

1.2.1 视频帧抽取

通过对所采集视频数据进行预处理, 以得到群养生

猪目标检测图像数据集。具体对采集到的视频进行视频帧抽取, 为防止数据重复冗余, 本文每隔 20 s 抽取一张视频帧图片, 在抽取过程中由于生猪运动会在图片中出现运动模糊和目标重影现象, 删去这些目标不清晰的图片, 再对其进行随机排序, 共获得包含 25 563 只生猪个体目标的 1 683 张群养生猪目标检测原始图像数据集。

表 1 不同生猪头数所对应的图像张数

Table 1 The number of images corresponding to different number of pigs

图像中生猪头数 The number of pigs in the picture	6~9	10~13	14~17	18~21	21~25
图像张数 The number of images	923	1 738	2 764	912	395

1.2.2 数据增强

为了增加可供模型训练的数据量, 提高模型的泛化能力和鲁棒性, 本文对生猪原始图像数据进行旋转(90°, 180°, 270°)、翻转、增加噪音等数据增强操作, 共获得 6 732 张图像, 增强后的图像如图 1 所示。

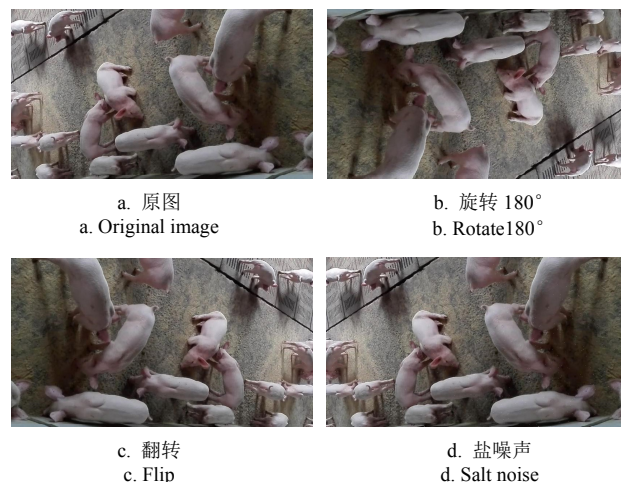


图 1 图像增强示例图

Fig.1 Sample images of image enhancement

1.2.3 图像标注

目标检测模型在训练和测试过程中, 需要提供群养生猪目标的真实位置信息, 本文使用 labelImg 图像标注工具(<https://github.com/tzutalin/labelImg>)对增强后的共计 6 732 张图像中的生猪目标进行标注, 并将标注好的信息按照通用的 PASCAL VOC 格式保存为与图像相对应且同名的 xml 格式文件, 该文件中记录了图片中每个目标的类别名称、矩形框位置等信息。

1.2.4 数据集划分

经上述整理后, 共获得 6 732 张标注过的图像及其对应的标注文件, 本文按照 6:2:2 比例将数据集划分为训练集、验证集、测试集, 训练集包含 4 038 张图像, 验证集包含 1 347 张图像, 测试集包含 1 347 张图像。训练集和验证集用于模型训练, 测试集用于模型检测结果验证。

2 群养生猪目标检测模型

2.1 CenterNet 网络模型

CenterNet 是由 Zhou 等^[23]近些年提出的无锚点

(Anchor-Free) 目标检测方法^[23], 其在遮挡目标检测方面相比于采用先验框设定的目标检测网络具有更好的检测效果, 且在目标检测过程中仅需提取每个目标的中心点, 无需进行后处理, 相比于传统目标检测网络提高了检测速度。CenterNet 使用残差网络 (ResNet50)^[24] 作为主干特征提取网络, 残差网络结构由 Conv Block 和 Identity Block 组成, 并通过转置卷积 (Transpose Convolution) 进行上采样, 其具体网络框架如图 2 所示。

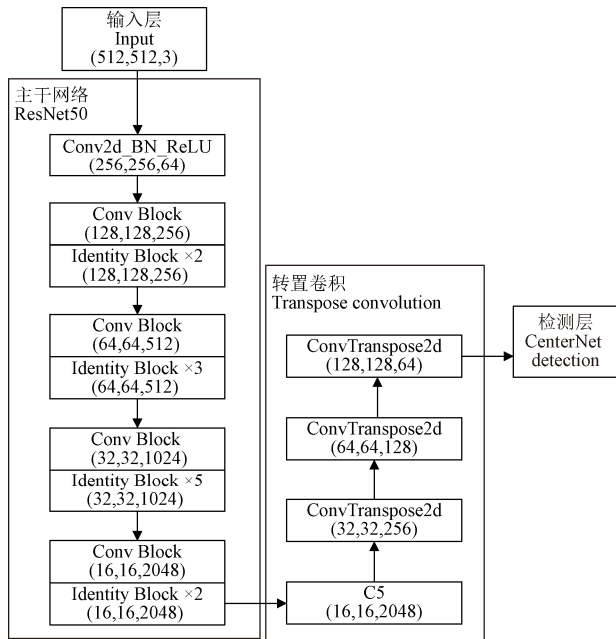


图 2 CenterNet 网络结构图

Fig.2 CenterNet network structure diagram

CenterNet 网络在数据输入后, 首先进行一次卷积、批标准化 (Batch Normalization, BN)、激活函数 (ReLU) 和最大池化; 其后, 经历 4 组残差结构, 每组残差结构分别由改变网络的维度的 Conv Block 和加深网络的 Identity Block 组成, 每组的 Block 个数分别为 3、4、6、3。在经过第 4 次 Conv Block 和 Identity Block 的堆叠之后, 完成了网络的特征初步提取, 获得维度为 (16, 16, 2048) 的有效特征层; 随后, 对获取到的有效特征层利用转置卷积进行 3 次上采样, 获得了高分辨率特征图, 最后使用 CenterNet 检测层进行结果预测。

2.2 MF-CenterNet 网络模型

尽管 CenterNet 在检测精度和检测速度上较现有目标检测算法均有不同的提升, 且对存在遮挡情况下的目标检测具有一定的鲁棒性, 但其仍不能满足低计算资源平台上的实时检测要求, 且由于模型体积过大难以部署在边缘计算端, 另外, 其对小目标检测的效果也不理想^[23]。

为进一步提高该目标检测模型的检测速度和对具有低计算资源的边缘计算端的可移植性, 以及对真实养殖环境中易出现的小目标的检测能力, 基于文献[25], 本文提出一种轻量级的群养生猪目标检测网络, 其在 CenterNet 目标检测网络的基础上, 首先改进其主干特征提取网络, 在保证检测精度的同时降低了模型总参数量,

提高了模型检测速度。考虑到 MobileNet 的基本单元是深度可分离卷积, 相较于传统卷积极大的减少了计算量及模型参数量, 为此, 本文使用轻量型的 MobileNet 系列网络^[26-28]对 CenterNet 的主干特征提取网络进行改进。由于 MobileNet 网络具有不同的实现版本, 其具体网络实现结构的选择将在 3.1 节进行具体讨论。

此外, 为进一步加强模型特征提取能力, 本文进一步使用特征金字塔结构 (Feature Pyramid Networks, FPN) 来融合低分辨率的高层语义信息和高分率的低层特征信息, 以提高该模型对群养生猪目标检测的精度, 尤其是具有较小尺寸的生猪目标的检测精度。FPN 的网络结构如图 3 所示。

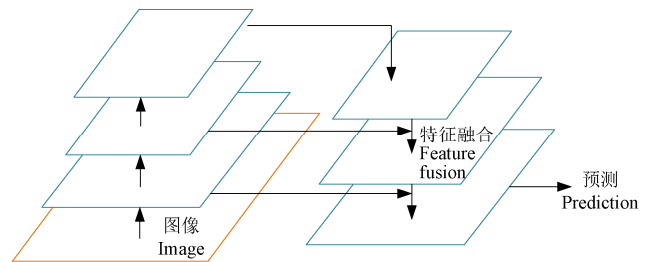


图 3 FPN 网络结构图

Fig.3 FPN network structure diagram

本文所使用的 FPN 是在一个自顶向下连接的跳跃式传递 (Skip Connection) 的网络结构上仅使用最后一层进行检测, 图像经过多次上采样并进行特征融合, 仅使用最后一步生成的特征进行检测。

为便于后文讨论, 本文将融合了轻量型 MobileNet 网络和 FPN 结构的 CenterNet, 记作 MF-CenterNet (MobileNet-FPN-CenterNet), 其整体网络框架如图 4 所示。

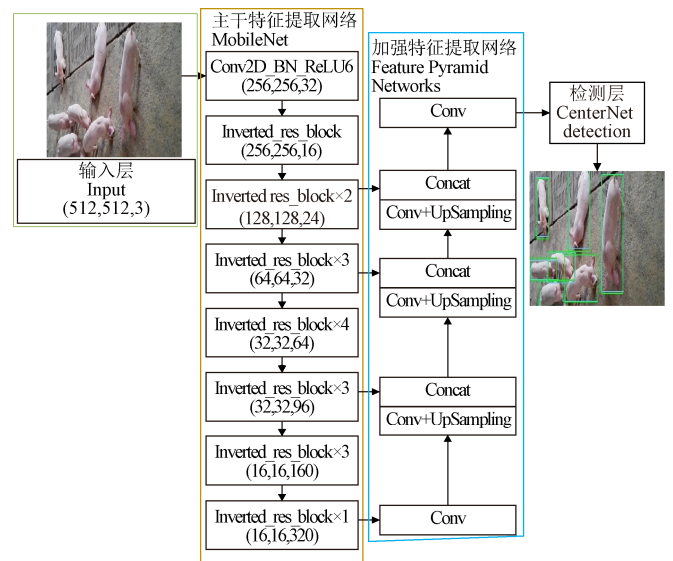


图 4 MF-CenterNet 网络结构图

Fig.4 MF-CenterNet network structure diagram

考虑到 CenterNet 属于 Anchor-Free 型的目标检测方法, 其检测层与传统 Anchor-based 型目标检测方法有所不同, 其主要包括 3 个预测分支, 即热力图预测、中心

点预测和尺寸（宽高）预测，各预测示意图如图 5 所示。

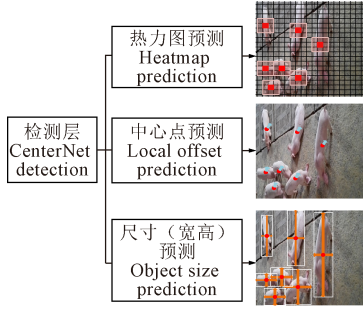


图 5 CenterNet 检测层示意图
Fig.5 CenterNet detection layer diagram

基于 CenterNet 的群养生猪目标检测模型就是将图像视为由无数关键点组成，通过关键点估计即热力图峰值确定生猪个体的中心点，再根据中心点位置回归得出目标宽高，最后确定生猪目标，其检测过程主要包含以下 3 个部分（三者为并行结构）：

1) 热力图（Heatmap）预测

CenterNet 检测层先对输入的特征图进行 3×3 的卷积，进行特征整合，再通过 BN 及 ReLU，最后利用 1×1 的卷积，将特征层的通道数调整为 num_classes。热力图预测的结果会判断每一个特征点是否存在对应的物体，以及物体种类和概率。

热力图的 Loss 整体思想和损失函数（Focal Loss）类似，对于容易分类的样本，可以适当减少其训练比例。热力图损失值（ L_k ）的计算公式如下：

$$L_k = \frac{1}{N_{\text{yyc}}} \sum_{\text{yyc}} \begin{cases} \left(1 - \hat{Y}_{\text{yyc}}\right)^{\alpha} \ln\left(\hat{Y}_{\text{yyc}}\right) & , Y_{\text{yyc}}=1 \\ \left(1 - \hat{Y}_{\text{yyc}}\right)^{\beta} \left(\hat{Y}_{\text{yyc}}\right)^{\alpha} \ln\left(1 - \hat{Y}_{\text{yyc}}\right) & , \text{otherwise} \end{cases} \quad (1)$$

式中 Y_{yyc} 为高斯核， α 、 β 是 Focal Loss 的超参数， N 是图像中关键点个数。

2) 中心点（Reg）预测

CenterNet 检测层先对输入的特征图进行 3×3 的卷积，进行特征整合，再通过 BN 及 ReLU，最后利用 1×1 的卷积，将特征层的通道数调整为 2。中心点预测结果可以对特征点的 x 轴、 y 轴坐标进行调整，获得物体中心的 x 轴、 y 轴坐标，即每一个特征点上对应的物体 x 轴、 y 轴坐标相对于同一个特征点 x 轴、 y 轴坐标的偏移情况。中心点偏移损失值（ L_{off} ）公式如下：

$$L_{\text{off}} = \frac{1}{N} \sum_p \left| \hat{O}_{\tilde{p}} - \left(\frac{p}{R} - \tilde{p} \right) \right| \quad (2)$$

式中 R 为输出的步长（Stride）， p 是真实关键点的中心点坐标， \tilde{p} 是下采样后低分辨率的中心点坐标， $\hat{O}_{\tilde{p}}$ 是对某一个关键点位置 \tilde{p} 的偏移预测。

3) 宽高（WH）预测

CenterNet 检测层先对输入的特征图进行 3×3 的卷积，进行特征整合，再通过 BN 及 ReLU，最后利用 1×1

的卷积，将特征层的通道数调整为 2。宽高预测结果直接回归获得对应物体框的宽高。目标宽高损失值（ L_{size} ）公式如下：

$$L_{\text{size}} = \frac{1}{N} \sum_{k=1}^N \left| \hat{S}_{p_k} - S_k \right| \quad (3)$$

式中 p_k 为目标 k 的包围框（bounding box）的中心位置， S_k 为目标尺寸。故总的损失值（Loss）为

$$\text{Loss} = L_k + \lambda_{\text{size}} L_{\text{size}} + \lambda_{\text{off}} L_{\text{off}} \quad (4)$$

式中 λ_{off} 、 λ_{size} 分别为 Reg 预测和 WH 预测的系数，根据文献[23]， λ_{off} 、 λ_{size} 分别取 0.1、0.2。

2.3 模型评价与训练

2.3.1 试验平台

本文试验所用的模型训练环境为 64 位 Windows10 操作系统，硬件配置 CPU 为 Intel(R) 12-Core(TM) i7-10750H，主频 2.60 GHz，显卡为 NVIDIA GeForce GTX 1660Ti GPU，内存为 16GB。试验采用的 Python 编程语言，并在 Torch 1.2.0 进行网络搭建、训练和测试，并采用迁移学习思想，使用 mobilenet_weights.pth 对轻量型骨干网络参数进行初始化。

2.3.2 评价指标

为验证所提目标检测算法的有效性，本文试验主要采用以下 6 个评价指标：准确率 P （Precision）、召回率 R （Recall）、平均精确率 mAP（Mean Average Precision）、平衡分数 $F1$ （F1-Score）、检测速度 FPS（Frame Per Second）、模型大小 Model Size。

准确率是度量目标检测模型对某一类别目标检测结果的准确性，即模型检测出的真实生猪目标数量占所有检测出目标数量的比例，其计算公式为

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (5)$$

式中 TP 为图像中生猪目标被正确检测为生猪的个数，FP 为图像中非生猪目标被检测为生猪的个数。

召回率是衡量一个目标检测模型把所有的检测目标都找到的能力，即模型正确检测出的生猪目标数量占所有真实生猪目标数量的比例，其计算公式为

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (6)$$

式中 FN 为图像中生猪目标被检测成其他种类物体的个数。

精确率是 PR 曲线与坐标轴所围成的面积，平均精确率是多类别 AP 的平均值即 P 对 R 的积分。

$$\text{mAP} = \text{AP} = \int_0^1 P(R) dR \quad (7)$$

平衡分数又称为 $F1$ 分数（F1-Score），它同时兼顾了分类模型的准确率和召回率，被定义为精确率和召回率的调和平均数。

$$F1 = 2 \times \frac{P \cdot R}{P + R} \quad (8)$$

检测速度 FPS 又称每秒帧率是每秒内目标检测模型可以处理的图片数量，即检测速度越快越能实现实时检

测, 其计算公式如下:

$$\text{FPS} = \frac{M}{S} \quad (9)$$

式中 M 为图像的总张数, S 为目标检测过程中处理所有图片所用的时间, s。

模型大小 (Model Size) 是目标检测网络训练所生成模型的空间利用率, 即模型大小由网络的基本结构决定。

2.3.3 模型训练参数设置

为了在训练模型时进行迁移学习, 本试验将网络的输入图像全部转化为 512×512 (像素) 尺寸大小, 同时为加快训练速度、防止训练初期权值被破坏, 采用冻结训练的方式, 前 50 次迭代 (Epoch) 进行冻结训练, 每 32 张图像作为一个 Batch Size, 50 次 Epoch 训练完成后解冻, 解冻后每 8 张图像作为一个 Batch Size, 每完成一个 Epoch 更新一次权值并保存, 权值衰减速率设为 0.5, 冻结时初始学习率设置为 0.001, 解冻后初始学习率设置为 0.000 1。

3 结果与分析

3.1 MobileNet 模型选择

MobileNet 系列网络当前主要有 3 个版本 (v1、v2、v3), 为验证不同版本模型作为特征提取主干网络对生猪目标检测的影响, 本文将分别采用 MobileNetv1、v2、v3 对 CenterNet 主干网络进行替换, 并对比集成不同版本 MobileNet 后模型性能与原始 CenterNet 性能的差异。

值得注意的是, MobileNet 系列网络的核心是通过引入深度可分离卷积来降低网络的参数以达到模型轻量化的目的, 但在进行深度卷积的过程中, 高、低维度信息转换会造成一定的特征信息损失, 特别是在 MobileNetv1 中, 再经过 ReLU 函数激活后, 特征信息的损失会进一步加大^[26]。为此, MobileNetv2 引入倒置残差 (Inverted Residuals) 和线性激活函数来改善网络中特征信息的损失问题^[27], 在性能上有所提升。MobileNetv3 则进一步引入神经结构搜索 NAS (Neural Architecture Search) 技术来进行网络参数的优化, 以提高模型的特征提取能力^[28], 并提出了面向高资源应用场景的 MobileNetV3-Large 和面向低资源应用场景的 MobileNetV3-small 两个版本。为了对比不同版本的 MobileNet 应用于生猪目标检测任务中的效果, 以及考虑本文重点关注面向边缘计算端的生猪目标检测模型, 本文通过试验重点对比分析集成 MobileNetv1、MobileNetv2 以及 MobileNetv3-small 的目标检测性能。图 6 给出了集成不同版本 MobileNet (v1、v2、v3) 的 MobileNet-CenterNet 在训练集上的损失值变化曲线, 横坐标为迭代次数, 纵坐标为模型训练损失值。

从图 6 不同版本 MobileNet 进行模型训练时的 Loss 值 (损失) 曲线可以看出, MobileNetv2-CenterNet 在训练时模型收敛更快, 且在模型损失曲线达到收敛时, 模型预测结果与真实目标位置所产生的误差也更小, 因此,

相较于其他版本, MobileNetv2-CenterNet 更容易在生猪目标检测数据集上进行训练。为了进一步验证, 集成不同版本 MobileNet 的 CenterNet 模型对生猪目标的检测性能, 本文接下来对相关模型在测试集上的性能表现进行对比分析, 其试验结果如表 2 所示。

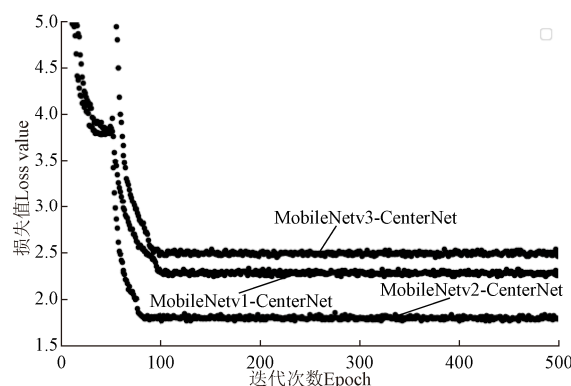


图 6 集成不同版本 MobileNet 模型的 Loss 值变化对比
Fig.6 Integrate the comparison of loss value change curves of different versions of MobileNet models

从对比试验结果表 2 可以看出, CenterNet 检测精度最高, 达到 93.67%, 检测速度为 27 帧/s 低于本文所采集视频的真实帧速率 29.9 帧/s, 模型总体参数量过大, 算法实时性和移植能力有待提高。在使用 MobileNet 对 CenterNet 主干网络进行改进后, MobileNetv1-CenterNet 的检测速度最快, 达到了 78 帧/s, 相比于原 CenterNet 提高 51 帧/s, MobileNetv3-CenterNet 模型的总体参数量约为 6 000 000 个, 相比于原 CenterNet 降低了 27 000 000 个。在检测精度方面, 主干网络改进后相比原 CenterNet 均有不同下降, 其中 MobileNetv2-CenterNet 检测精度为 93.37%, 在所有改进后的模型中为最高, 且与原 CenterNet 精度较为接近, 但 MobileNetv2-CenterNet 检测速度达到 62 帧/s 比原 CenterNet 提高了 35 帧/s, 总体参数量为 8 000 000 个降低了 25 000 000 个。

3.2 MF-CenterNet 与不同 CNN 网络检测性能对比

在利用 MobileNetv2 对 CenterNet 主干网络进行改进的基础上, 进一步使用 FPN 替换原有的转置卷积, 以加强网络特征提取能力, 达到提高模型检测精度的目的, 也即本文所提出的 MF-CenterNet 模型。为验证使用 FPN 的效果, 本文对使用 FPN 改进 MobileNetv2-CenterNet 前后的网络结果进行对比, 并与其他常见用于目标检测的 CNN 模型进行比较, 其结果如表 3 所示。

由表 3 可见, 本文所提出的 MF-CenterNet 目标检测模型对真实场景下群养生猪目标检测精度最高, 其检测精度达到 94.30%, 较没有集成 FPN 的 MobileNetv2-CenterNet 检测精度提高了 0.93 个百分点, 比 Faster-RCNN、SSD、YOLOv3、YOLOv4、CenterNet 目标检测模型分别提高了 6.39、4.46、6.01、2.74、0.63 个百分点。与此同时, 检测速度达到 69 帧/s, 比改进 FPN 之前 MobileNetv2-CenterNet 快了 7 帧/s, 与 Faster-RCNN、SSD、YOLOv3、YOLOv4 相比, 检测速度分别提高了 54、47、45、43 帧/s。

表 2 原 CenterNet 与集成不同版本 MobileNet 后的试验结果对比
Table 2 Comparison of trial results between original CenterNet and integrated versions of MobileNet

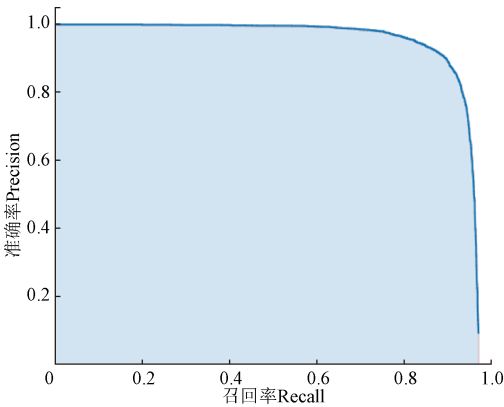
检测模型 Detection model	主干网络 Backbone	准确率 Precision/%	召回率 Recall/%	平均精度 Mean average precision/%	平衡分数 F1-Score	检测速度 Detection speed/ (帧·s ⁻¹)	总参数量 Total params
CenterNet	ResNet50	97.12	76.83	93.67	0.86	27	33 000 000
MobileNet-CenterNet	MobileNetv1	97.16	58.67	90.70	0.73	78	9 000 000
	MobileNetv2	95.55	79.39	93.37	0.87	62	8 000 000
	MobileNetv3	97.72	34.03	87.03	0.50	53	6 000 000

表 3 不同生猪目标检测模型性能对比
Table 3 Performance comparison of different models for pig detection

检测模型 Detection model	主干网络 Backbone	模型大小 Model size/MB	平均精度 Mean average precision/%	检测速度 Detection speed/ (帧·s ⁻¹)
Faster-RCNN	VGG 16	270	87.91	15
SSD	VGG 16	91	89.84	22
YOLOv3	Darknet53	235	88.29	24
YOLOv4	CSPDarknet53	244	92.56	26
CenterNet	Resnet50	125	93.67	27
MobileNetv2-CenterNet	MobileNetv2	32	93.37	62
MF- CenterNet	MobileNetv2	21	94.30	69

模型大小上，MF-CenterNet 仅为 21MB，相比于 Faster-RCNN、SSD、YOLOv3、YOLOv4 分别降低了 249、70、214、223 MB，较 MobileNetv2-CenterNet 减小了 104 MB。因此集成 FPN 后的目标检测模型 MF-CenterNet 有效的提高了生猪检测效果。本节所做试验证明了使用 FPN 结构替换原转置卷积，不仅可以提高一定的检测精度，而且进一步提高了检测精速度、降低了模型大小。为了更直观展示本文所提出模型在不同召回率下所对应的准确率，将 MF-CenterNet 目标检测网络的精确度结果可视化为 PR 曲线，曲线与横纵坐标所围成的面积即为精确度值，MF-CenterNet 目标检

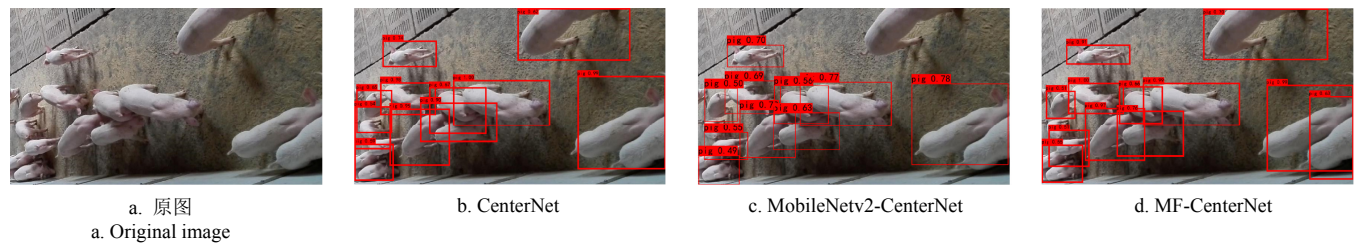
测网络 PR 曲线如图 7 所示。



注：阴影部分的面积为 MF-CenterNet 模型的精确度值。
Note: The shaded area is the accuracy value of the MF-CenterNet model.

图 7 MF-CenterNet 模型的 PR 曲线
Fig.7 PR curve of MF-CenterNet model

为更直观体现 MF-CenterNet 目标检测模型的检测结果，本节将 MF-CenterNet 模型和原始 CenterNet 模型以及只改进主干网络的 MobileNetv2-CenterNet 模型的生猪目标检测结果进行对比，如图 8 所示，图中 CenterNet 及 MobileNetv2-CenterNet 均有未被检测出的生猪目标，由此可见，MF-CenterNet 模型有效提高了群养生猪的目标检测效果，尤其是对遮挡目标和小目标均有很好的检测，满足对真实场景下群养生猪目标检测的需求。



注：红色框为 MF-CenterNet 模型的生猪目标检测框。
Note: The red box is the pig detection frame of MF-CenterNet model.

图 8 MF-CenterNet 与改进前模型检测效果对比

Fig.8 Comparison of detection effect between MF-CenterNet and the model before improvement

3.3 MF-CenterNet 与现有生猪目标检测方法对比

表 4 给出了本文所提 MF-CenterNet 模型在本文测试集中的生猪目标检测结果以及现有生猪目标检测方法在其文献中报告的检测结果。其中，文献[17]是通过采用 BING 算法提取图像候选区域，通过改进 CNN 模型对候选框分类，最后利用非极大值抑制算法剔除冗余候选框，实现对生猪目标的检测。文献[20]是将深度为 3 的特征金

字塔注意力模块 FPA（即 FPA-3）与 Tiny-YOLOv3 相结合，对不同场景下的多目标生猪个体进行检测。文献[21]则通过使用稀疏化可分离卷积（S-DSC）对 SSD 进行改进，训练时自主采样困难样本并精简模型预测网络实现生猪目标的快速检测。
由表 3 可以看出，本文所提出的 MF-CenterNet 模型尽管在检测精度上并未达到最高，相比于其他文献中所

提方法在平均精度上低 1~2 个百分点, 考虑可能存在的原因是, 一方面, 本文所用图像均为斜侧方位进行拍摄, 生猪相互遮挡情况较其他文献中使用的数据较为明显, 提高了对生猪目标检测的难度; 另一方面, 本文试验所用的图像中, 生猪个数平均在 13 头以上, 而文献[17, 20—21]所使用的图像中生猪个数平均在 6~7 头, 这也提高了目标检测的难度。与文献[17]及文献[21]的网络相比本文所改进的网络更加轻量化, 满足实时检测的需求。与文献[20]相比, 本文所选用的基础网络 CenterNet 相比于 YOLOv3 具有更好的检测精度, 再通过特征金字塔结构进行优化, 进一步提高检测效果。虽然本文所提出的改进方案有效地提高了原始模型的检测效果, 但在目前生猪检测领域并未处于最前端行列, 尤其是检测精度相对较低。虽然本文所改进的 MF-CenterNet 检测精度相对较低, 但在检测速度上, 本文所改进的 MF-CenterNet 模型提升较为显著, 分别较现有工作提高了 67、57、21.5 帧/s。同时在模型大小上, 与现有文献中的参数相比, 本文所提的 MF-CenterNet 模型也具有较为突出的优势。

表 4 MF-CenterNet 与现有检测方法结果对比

Table 4 Comparison between MF-CenterNet and existing detection methods

方法 Methods	单栏生猪平均 头数 Average number of pigs in a single column	图像 数据集 Image data set	检测网络 Detection network	平均 精度 Average precision/ %	检测速度 Detection speed / (帧·s ⁻¹)
文献[17] Literature[17]	7	8 000	改进 CNN 算法	96	1.4
文献[20] Literature[20]	6	4 104	FPA-Tiny-YOLO	85.85	47.5
文献[21] Literature[21]	6	3 000	融合困难样本的 SDSC-SSD	96.21	12
本文 This paper	13	6 732	MF-CenterNet	94.30	69

综上, 本文提出的轻量型 MF-CenterNet 群养生猪目标检测模型在保证模型轻量化、检测实时性的同时, 在检测精度上也达到很好的效果, 满足对真实养殖环境下群养生猪的实时、精准检测。

4 结 论

本文提出了一种轻量型的 MF-CenterNet 群养生猪目标检测模型, 在满足实时检测和模型轻量化的同时, 有效提高了群养条件下相互遮挡和较小尺寸生猪目标的检测能力, 在群养生猪出现拥挤或遮挡的场景下, 平均精度达到 94.30%, 目标检测速度达到 69 帧/s, 模型大小仅为 21 MB。与经典网络模型 Faster-RCNN、SSD、YOLOv3、YOLOv4 相比, 平均精度分别提高了 6.39、4.46、6.01、2.74 个百分点, 检测速度增加了 54、47、45、43 帧/s, 模型大小减少了 249、70、214、223 MB。通过采用 MobileNetv2 改进 CenterNet 的主干网络, 并结合 FPN 加强模型特征提取能力, 在保证模型检测实时性和轻量化的同时增加了平均精度, 与原网络 CenterNet 相比平均精

度提高了 0.63 个百分点。本文工作为实现生猪养殖过程中实时精准的数量盘点提供了方法, 此外, 通过与生猪个体质量估测、饮食排泄频次统计、攻击咬尾行为识别等健康监测模型相结合, 为规模化养殖场实现精准、个性化的生猪健康监测提供了思路。接下来, 本文将进一步研究改进加强特征提取网络、减少网络结构层、增加注意力机制等方法, 在保证实时检测及网络轻量化的同时提高检测精度, 根据应用需求与资源限制, 构建精确度与速度相平衡的轻量级网络。

[参 考 文 献]

- [1] 华利忠, 冯志新, 张永强, 等. 以史为鉴, 浅谈中国非洲猪瘟的防控与净化[J]. 中国动物传染病学报, 2019, 27(2): 96-104.
- [2] 刘杏, 王凤雪, 温永俊, 等. 猪繁殖与呼吸综合征病毒和猪圆环病毒 2 型混合感染的流行病学调查[J]. 吉林农业大学学报, 2016, 38(4): 456-459.
- [3] 任鹏举, 李鹏, 张秋雨, 等. 新型猪瘟疫苗的研究进展[J]. 中国畜牧兽医, 2018, 45(7): 1958-1964.
- [4] 杨心, 朱伟兴. 基于广义 Hough 聚类的粘连猪头尾识别[J]. 江苏农业科学, 2018, 46(9): 230-235.
- [5] 马丽, 纪滨, 刘宏申, 等. 单只猪轮廓图的侧视图识别[J]. 农业工程学报, 2013, 29(10): 168-174.
- [6] 谢双云, 王芳, 田建艳, 等. 融合高斯混合建模和图像粒化的猪只目标检测[J]. 黑龙江畜牧兽医, 2016(1): 29-32.

- Xie Shuangyun, Wang Fang, Tian Jianyan, et al. Pig target detection based on Gaussian mixture modeling and image granulation[J]. Heilongjiang Animal Husbandry and veterinary, 2016(1): 29-32. (in Chinese with English abstract)
- [7] Shaoqing R, Kaiming H, Ross G, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6): 1137-1149.
- [8] Girshick R. Fast R-CNN[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
- [9] 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(CVPR). IEEE Computer Society: Piscataway, NJ. 2016: 779-788.
- [10] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 6517-6525.
- [11] Redmon J, Farhadi A. Yolov3: An incremental improvement[J/OL]. [2019-07-10]. USA: 2018. <https://arxiv.org/abs/1804.02767>
- [12] Bochkovskiy A, Wang C Y, Liao H. YOLOv4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv: 2004.10934, 2020.
- [13] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision(CVPR). IEEE Computer Society: Piscataway, NJ. 2016: 21-37.
- [14] Riekert M, Klein A, Adrion F, et al. Automatically detecting pig position and posture by 2D camera imaging and deep learning[J]. Computers and Electronics in Agriculture, 2020, 174: 105391.
- [15] Yang Q, Xiao D, Lin S. Feeding behavior recognition for group-housed pigs with the Faster R-CNN[J]. Computers and Electronics in Agriculture, 2018, 155: 453-460.
- [16] Cowton J, Kyriazakis I, Bacardit J. Automated individual pig localisation, tracking and behaviour metric extraction using deep learning[J]. IEEE Access, 2019, 7: 108049-108060.
- [17] 王婷婷. 改进的 Faster R-CNN 在目标参数测量方面的研究与应用[D]. 哈尔滨: 哈尔滨工程大学, 2018.
- Wang Tingting. Research and Application of Improved Fast r-cnn in Target Parameter Measurement[D]. Harbin: Harbin Engineering University, 2018. (in Chinese with English abstract)
- [18] 谢涛. 基于深度学习的生猪行走行为视频检测与分析[D]. 太原: 太原理工大学, 2019.
- Xie Tao. Video Detection and Analysis of Pig Walking Behavior Based on Deep Learning[D]. Taiyuan: Taiyuan University of technology, 2019. (in Chinese with English abstract)
- [19] 苏恒强, 郑笃强. 基于深度学习技术生猪图像目标检测算法的应用研究[J]. 吉林农业大学学报, 2020, 5: 1-8.
- Su Hengqiang, Zheng Duqiang. Application research on pig image target detection algorithm based on deep learning technology[J]. Journal of Jilin Agricultural University, 2020, 5: 1-8. (in Chinese with English abstract)
- [20] 燕红文, 刘振宇, 崔清亮, 等. 基于特征金字塔注意力与深度卷积网络的多目标生猪检测[J]. 农业工程学报, 2020, 36(11): 193-202.
- Yan Hongwen, Liu Zhenyu, Cui Qingliang, et al. Multi target pig detection based on feature pyramid attention and deep convolution network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(11): 193-202. (in Chinese with English abstract)
- [21] Hanse A, Wonseok C, Sunhwa P, et al. Image processing and deep learning techniques for fast pig's posture determining and head removal[J]. KIPS Transactions on Software and Data Engineering, 2019, 8(11): 457-464.
- [22] 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]. 农业机械学报, 2020, 51(7): 44-51.
- Li Bo, Shen Mingxia, Liu Longshen, et al. Head and tail discrimination method of group pig based on YOLO v3 and graph structure model[J]. Acta mechanization agricola Sinica, 2020, 51(7): 44-51. (in Chinese with English abstract)
- [23] Zhou X, Wang D, Krhenbühl P. Objects as points[J]. arXiv preprint arXiv, 2019, 1904: 07850
- [24] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas: IEEE, 2016: 770-778.
- [25] Xu Y, Yan W, Sun H, et al. CenterFace: Joint face detection and alignment using face as point[J]. Scientific Programming, 2020, 1314-1324.
- [26] Howard A, Zhu M, Chen B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii: IEEE, 2017, 1704-1712.
- [27] Sandler, Howard A, Zhu M, et al. MobileNetV2: Inverted residuals and linear bottlenecks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018, 4510-4520.
- [28] Howard A, Sandler M, Chu G, et al. Searching for MobileNetV3[C]. // Proceedings of the IEEE International Conference on Computer Vision, 2019, 140.

Detection of group-housed pigs based on improved CenterNet model

Fang Junlong, Hu Yuhang, Dai Baisheng^{*}, Wu Zhidong

(1. School of Electrical Engineering and Information, Northeast Agricultural University, Harbin 150030, China; 2. Key Laboratory of Pig Breeding Facilities Engineering, Ministry of Agriculture and Rural Affairs, Harbin 150030, China)

Abstract: Rapid and accurate detection of pigs has been critical to intelligent monitoring of health status within a group-housed breeding environment on large-scale farms. However, a large number of parameters make it difficult to achieve real-time performance in edge computing platforms for practical production. In this study, an improved CenterNet model (named MF-CenterNet) was proposed to detect pigs in group-housed breeding conditions, in order to improve the real-time performance of detection and the accuracy of localizing pigs with body occluded and small body size. An anchor-free CenterNet was also used to ensure the accuracy of detection, especially for the pig with body occluded. A lightweight MobileNet network was first introduced into the CenterNet (instead of ResNet50), as the backbone network of feature extraction for the smaller model size and higher detection speed. In addition, the feature pyramid structure (FPN) was then added to improve the ability of feature extraction for small pig objects. As such, the CenterNet was integrated with the MobileNet and FPN, named MF-CenterNet (i.e., MobileNet-FPN-CenterNet, MF-CenterNet). An image dataset of a private pig was collected to evaluate the performance of MF-CenterNet. All images were then captured from Jincheng Farm, Qiqihar City, Heilongjiang Province, China. Specifically, 1683 video frames were extracted from the video recording of pigs collected in the commercial pig farm, and 6732 images were obtained with the operation of the data argument. All pig objects within the images were then labeled with the labeling tool. The experimental results show that the size of the MF-CenterNet model was only 21MB, which satisfied the deployment of the model to an edge computing platform. The mean average precision (mAP) of pig detection was 94.30%, and the detection speed was up to 69 frames/s. The model of CenterNet integrated with MobileNetv2 achieved the best performance, in terms of accuracy, speed, and model size, where different versions of Mobile Net were combined. The CenterNet model integrated with the MobileNetv2 and FPN (MF-CenterNet) further improved the detection performance, indicating more robust in detecting the pig objects with small body size and body occluded. The improved MF-CenterNet greatly increased the mAP by 0.63 percentage points, and the speed by 42 frames/s, while the size of the model was reduced by 104 MB, compared with the original CenterNet. Furthermore, the mAP detection was improved by 6.39, 4.46, 6.01, and 2.74 percentage points, while, the detection speed was improved by 54, 47, 45, and 43 frames/s, respectively, compared with the common CNN-based object detection models, including Faster RCNN, SSD, YOLOV3, and YOLOV4 model. Consequently, the MF-CenterNet achieved the state-of-the-art mAP performance, higher detection speed, and the deployability of the model in a substantial manner. Therefore, this lightweight object detection model, MF-CenterNet, can meet the requirements of real-time, rapid, and high accuracy of detection on the group-housed pigs. The finding can also be expected to serve as a new way for real-time detection and prerequisite model in the behavior analysis of pigs during modern intensive production.

Keywords: computer vision; image processing; group-housed pig; object detection; mobileNet; FPN; centerNet