

文章编号: 2095-2163(2022)01-0084-06

中图分类号: TP39

文献标志码: A

融合视觉上下文与跨通道信息的生猪脸部轻量化检测模型

戴百生^{1,2}, 李润泽¹, 张逸轩¹, 刘洪贵^{2,3}, 刘润泽³

(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; 3 东北农业大学 动物科学技术学院, 哈尔滨 150030)

摘要: 生猪脸部的检测,是实现生猪个体身份、情绪状态和行为模式智能化自动识别与分析的关键环节。针对生猪在真实养殖环境中脸部区域通常存在遮挡、视角多变以及模型部署端资源有限等问题,提出了一种融合视觉上下文与跨通道信息的生猪脸部轻量化检测模型。本文采用 ShuffleNet v2 作为检测模型的特征提取模块;设计 Nonlocal-PAN 特征融合模块,融合不同尺度特征图中的视觉上下文信息和跨通道信息;利用 NanoDet 检测头模块来完成对生猪脸部目标的检测。实验结果表明,该模型在 *IoU* 为 0.50、0.95、0.50:0.95 水平下,分别获得了 96.98%、81.16% 和 67.44% 的平均精度,且模型大小保持在 8.2 MB 的轻量化水平,具有良好的边缘端移植性和较高精度,为生猪养殖智能化监测与分析提供了新技术。

关键词: 生猪脸部检测; ShuffleNet; 视觉上下文; 跨通道; 智慧畜牧

Lightweight model of pig face detection by fusing vision contextual and cross-channel information

DAI Baisheng^{1,2}, LI Runze¹, ZHANG Yixuan¹, LIU Honggui^{2,3}, LIU Runze³

(1 College 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;

3 College of Animal Science and Technology, Northeast Agricultural University, Harbin 150030, China)

【Abstract】 Pig face detection is one of the most important prerequisite step for pig individual identification, emotion status and behaviors recognition in the field of smart pig breeding. To solve the problem of face occlusion and variety about visual angle in real senses of breeding, a lightweight model of pig face detection by fusing vision contextual and cross-channel information is proposed in this paper. Firstly, a lightweight feature extraction network, i.e., ShuffleNet v2, is used as backbone network of the proposed model. Secondly, a nonlocal-PAN module is designed for fusing vision contextual information and cross-channel information of extracted feature maps with different scales. Finally, a lightweight NanoDet head is adopted as the object detection module based on those aggregated features. Experimental results show that the proposed model has achieved an accuracy of 96.98%, 81.16%, and 67.44% at *IoU* level of 0.50, 0.95 and 0.50:0.95 respectively with model size of 8.2 MB. The proposed model has an excellent ability of deployment in edge computing hardware and superior detection performance, which can be applied effectively in the field of smart pig breeding.

【Key words】 Pig face detection; ShuffleNet; Vision context; Cross-channel; Intelligent livestock framing

0 引言

随着生猪养殖规模化程度的日益提高,养殖标准化、精细化以及养殖场管理现代化,已成为生猪养殖行业未来发展的必然趋势。实现生猪的精准养殖与规模化猪场的现代化管理,亟需对生猪的个体身份、健康状态、行为模式进行智能化的感知与理解。而生猪脸部正是区分其个体信息、反应其情绪状态

与行为意图的重要区域^[1]。利用计算机视觉技术对生猪脸部进行非接触式快速、准确的自动检测,是实现生猪现代化养殖与管理的重要手段。

目前,国内外已有一些关于生猪脸部检测与识别方面的研究。Wada 等人^[2]通过手工定位并分割出生猪脸部等区域,借助人脸识别领域经典的 Eigenfaces 方法,实现对生猪个体的身份识别。秦兴等人^[3]同样采用手工定位的方式,在提取出生猪脸部图像后,通过训练双线性卷积神经网络(Bilinear-

基金项目: 黑龙江省普通本科高等学校青年创新人才培养计划(UNPYSCT-2018142)。

作者简介: 戴百生(1986-),男,博士,副教授,主要研究方向:智慧畜牧与数字农业;李润泽(2000-),男,本科生,主要研究方向:计算机视觉;张逸轩(2001-),男,本科生,主要研究方向:计算机视觉;刘洪贵(1979-),男,博士,副教授,主要研究方向:动物行为与福利学;刘润泽(1996-),男,博士研究生,主要研究方向:动物福利与健康养殖。

通讯作者: 戴百生 Email:bsdai@neau.edu.cn

收稿日期: 2021-10-16

CNN)对其个体身份进行识别。为了提高生猪脸部识别的实用性,Marsot 等人^[4]则采用 Haar 级联分类器,对生猪脸部进行自动检测,针对检测出的生猪脸部图像区域,训练 CNN 模型完成对生猪个体身份的识别。燕红文等人^[5]为了提高生猪脸部定位精度,进一步将注意力机制与 Tiny-YOLO 模型相结合,在实现生猪脸部检测的同时,识别生猪脸部的不同姿态。Hansen 等人^[6]则利用 Mask-RCNN 模型,对生猪脸部区域进行检测与分割,并结合生猪眼部区域特征,对生猪情绪状态进行监测,以判断生猪是否处于焦虑状态。

虽然上述研究在生猪脸部检测方面取得了一定成功,但多数方法仍不能较好地处理复杂场景下的生猪脸部检测,难以满足不同养殖条件下生猪的实时监测需求。特别是在真实饲养环境中,生猪脸部区域通常会被养殖设施和其他个体遮挡,导致在视频画面中脸部区域显示不完整,或由于生猪自身的运动,造成脸部姿态的旋转、扭曲等视角变化的多样性。这些因素都会给生猪脸部的自动检测算法在真实应用场景下带来不利影响。

为了提高生猪目标检测的精度、速度和实用性,本文首先构建了包含限位栏和群养两种真实饲养环境下的生猪脸部检测数据集,用于生猪脸部目标检测算法的训练与测试;其次,本文基于 NanoDet 架构^[7],提出了一种融合视觉上下文与跨通道信息的生猪脸部轻量化检测。在应用轻量化特征提取骨干网络的同时,设计了 Nonlocal-PAN 特征融合模块,来融合图像内部的上下文信息与跨通道信息,并采用 Anchor-Free(无锚点)目标检测结构,对生猪脸部进行检测,以提高对存在遮挡、视角变化等复杂情况下的生猪目标检测的准确性、实时性和实用性。

1 试验数据

1.1 数据采集

本文构建的生猪脸部检测图像数据,采集自东北农业大学阿城实验实习基地猪场。为获取不同养殖环境下的生猪脸部图像数据,于 2020 年 7 月~10 月,分别选取了该猪场限位栏饲养条件下的妊娠猪舍和群养条件下的保育猪舍进行视频和图像数据采集。针对限位栏下的妊娠母猪,通过海康威视监控摄像头,固定角度拍摄 5 个栏位妊娠母猪共计 5 个批次的视频场景数据;针对群养环境下的保育猪,利用 VIVO 手机,非定点移动拍摄 12 个群养栏中约 120 头保育猪图像场景数据。采集到的限位栏和群养条件下的

生猪场景数据分别如图 1(a)、(b)所示。



(a) 妊娠母猪限位栏饲养场景



(b) 保育猪群养场景

图 1 不同养殖环境下生猪场景数据示例

Fig. 1 Illustrations of pigs in different breeding condition

1.2 数据整理

针对限位栏下妊娠母猪视频数据,对其进行视频帧解析处理,每隔 5 s 进行视频帧的抽取,获得分辨率为 2 560×1 440 像素的图像数据;针对群养环境下保育猪图像数据,将图像尺寸统一调整分辨率为 768×1 024 像素。分别从两种不同养殖条件下的生猪图像中,各随机选取 2 000 幅,共同构成生猪脸部检测数据集,共计 4 000 幅图像数据。对数据集中所有图像,利用 labelImg 图像标注工具^[8],对图像中不同姿态、视角和大小的生猪脸部区域进行标注,并按照 Microsoft COCO 目标检测数据集^[9]的标注格式,存储生猪脸部目标区域的标注信息。

为了训练和验证本文所提生猪脸部检测模型,按照 9:1 的比例将数据集划分为训练集和测试集。其中,训练集 3 600 幅,包含 11 570 个生猪脸部目标框;测试集 400 幅,包含 1 346 个生猪脸部目标框。

2 检测模型

本文提出的生猪脸部检测模型总体结构如图 2 所示,主要由特征提取模块(Backbone 网络)、特征融合模块(Neck 网络)及目标检测模块(Head 网络)组成。特征提取模块采用轻量化的 ShuffleNet v2 模型,提取待检测图像特征;特征融合模块设计了 Nonlocal-PAN 结构,融合图像视觉上下文及跨通道信息;目标检测模块采用轻量化的 NanoDet 目标检测头,实现对生猪脸部目标的检测。

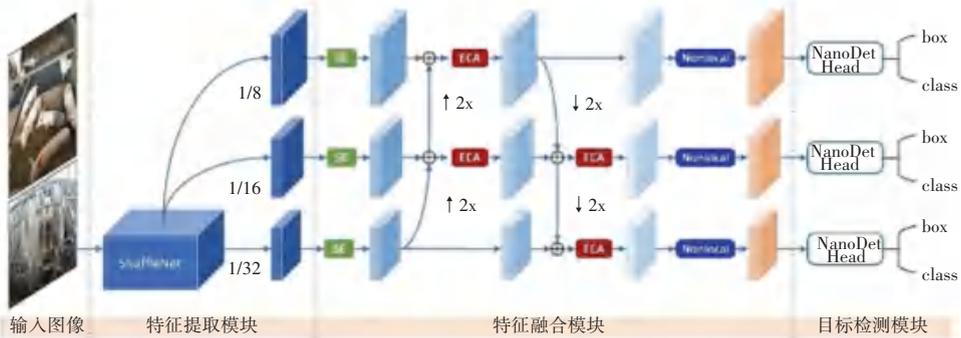


图2 生猪脸部目标检测网络结构

Fig. 2 Network structure of proposed pig face detection model

2.1 特征提取模块

为了使生猪脸部检测模型在保证检测精度的同时具有易在边缘计算端部署的轻量化结构,以便于该模型在生产现场中进行实际应用,本文采用轻量化的 ShuffleNet v2 网络^[10]作为检测模型的特征提取模块(即 Backbone 网络)。该网络提出一种通道分割(Channel Split)和通道随机混合(Channel Shuffle)机制。通过对通道分组并进行重新组合,到达混合信息的效果,实现在减少模型参数的同时提高网络特征提取性能,该网络的基本结构如图3所示。

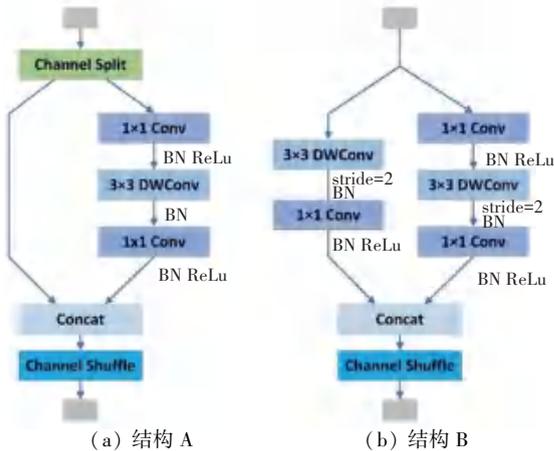


图3 ShuffleNet 基本结构

Fig. 3 Basic structure of ShuffleNet

该结构改进了自残差网络(ResNet)的残差结构^[11]。其中,图3(a)结构通过通道分割操作,将特征图的通道分成两组,左侧分支直接映射,右侧分支通过深度可分离卷积操作,确保输入通道和输出通道保持一致,以提高网络效率,再将左右两分支输出进行拼接后,使用通道随机混合操作,得到该结构的最终输出。图3(b)则通过不进行通道分割的量分支结构,实现通道数量的加倍,用以网络的降采样。整个 ShuffleNet v2 即通过这些基本结构堆叠而成,在轻量化模型参数的同时,提高模型计算效率,并利用特征重用的思想提高模型特征提取能力。为了提

高对不同尺寸生猪脸部目标检测的性能,本文将 ShuffleNet v2 网络中体现不同尺度特征的第2、3、4阶段的输出作为后续特征融合模块的输入。

2.2 特征融合模块

为了提高对存在遮挡和视角多变的生猪脸部目标的检测性能,需要综合利用生猪躯干、环境设施等视觉上下文信息,以及生猪脸部不同视角表现出的不同模式特征,来辅助定位复杂场景下的生猪脸部区域。本文提出一种 Nonlocal-PAN 结构,对提取到的特征进行上下文信息融合和通道注意力的加权,该网络的具体结构如图2中特征融合模块所示。受 YOLO-Nano 模型^[12]启发,本文先在现有 PAN 网络^[13]不同尺度的横向连接(lateral connection)中加入传统的 SE 通道注意力块^[14],来突出不同视角下生猪脸部表现出来的不同模式特征,进而可在各尺度特征融合输出后,加入轻量化的 ECA 通道注意力块^[15]。ECA 内部结构如图4所示。

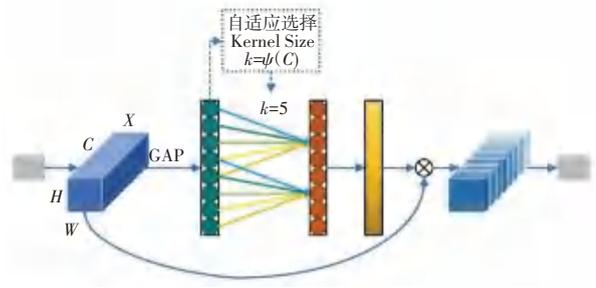


图4 ECA 注意力块结构

Fig. 4 Structure of ECA attention block

通过自适应选择一维卷积核大小,实现局部跨通道交互,进一步提高特征融合能力;并在最终融合特征图上利用 Nonlocal 块^[16]来建模不同尺度特征图上视觉上下文信息,提高网络对存在遮挡的生猪脸部特征提取能力,Nonlocal 块结构如图5所示,其计算过程如下:

$$z_i = W_z y_i + x_i \quad (1)$$

$$y_i = \frac{1}{C(x)} \sum_j f(x_i, x_j) g(x_j) \quad (2)$$

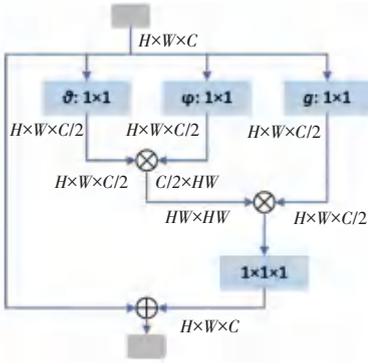


图 5 Nonlocal 块结构

Fig. 5 Structure of nonlocal block

式中, x 为输入特征图; i, j 是位置索引; y 为与输入特征图 x 大小一致的经过非局部操作的输出; f 为计算任意两位置间的特征相似性度量函数; g 为任意位置 j 处的特征表示函数; $C(x)$ 为规范化因子; W_z 为权重矩阵; $g(x_j)$ 可通过 $W_g x_j$ 形式计算获得。在本文工作中, 相似性度量函数 f 选用如下形式进行计算:

$$f(x_i, x_j) = e^{\vartheta(x_i) \cdot T\varphi(x_j)}$$

式中, $\vartheta(x_i)$ 和 $\varphi(x_j)$ 可分别通过 $W_\vartheta x_i$ 和 $W_\varphi x_j$ 计算获得。

2.3 目标检测模块

基于前述特征融合模块输出的 3 个不同尺度上的特征图, 采用轻量化的 NanoDet 检测模块^[7] 完成对生猪脸部区域的检测。该模块主要通过 ATSS^[17] 机制, 自动选择正负样本进行训练, 并引入 Generalized focal loss^[18] 和 GIoU loss^[19], 对生猪脸部区域进行类别预测和位置回归, 模型总体损失函数计算方式如下:

$$L = \frac{1}{N_{pos}} \sum_z L_Q + \frac{1}{N_{pos}} \sum_z 1_{\{C_z^* > 0\}} (\lambda_0 L_B + \lambda_1 L_D) \quad (3)$$

式中, L_Q 为衡量定位质量的损失函数 (Quality Focal Loss), 用来综合考虑类别预测得分与定位质量预测得分; L_D 为衡量包围框分布的损失函数 (Distribution Focal Loss), 以加速预测包围框聚焦到目标位置邻近区域的分布中; L_B 为包围框回归损失函数, 即 GIoU 损失函数; N_{pos} 为正样本个数; z 为特征图中位置索引; $1_{\{C_z^* > 0\}}$ 为示性函数; λ_0, λ_1 为权衡权重。

3 实验结果与分析

3.1 模型训练

本文使用 Pytorch 框架进行实验, 采用随机梯度下降方法 (SGD) 作为模型训练的优化器, 批处理样本数 (Batch size) 为 96, 迭代次数 (Epoch) 为 150, 初

始学习率 (Learning rate) 为 0.14, 并根据 Epoch 适时调整学习率, 调整因子设置为 0.1, 网络输入图像尺寸统一调整为 512×512 像素。

图 6 给出了训练集上的模型损失函数曲线, 从中可以看出, 在生猪脸部目标检测训练集上, 本文所采用的损失函数均能够有效收敛。

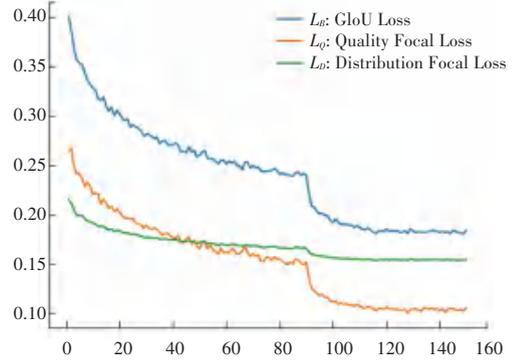


图 6 训练集上不同损失函数收敛曲线

Fig. 6 Convergence curves of different model loss on training set

3.2 评价指标

考虑本文主要针对单一类别的生猪脸部目标进行检测, 故采用该单一类别在不同交并比 (IoU) 水平下的平均精度 (Average Precision, AP)、平均召回率 (Average Recall, AR) 以及模型大小作为评价本文生猪脸部检测模型的性能指标, 其中:

$$AP = \int_0^1 p(r) dr \quad (4)$$

$$AR = 2 \int_{0.5}^1 r(o) do = \frac{2}{n} \sum_{i=1}^n \max(IoU(g t_i) - 0.5, 0) \quad (5)$$

式中, p 为精度; r 为召回率; o 为预测框与真实目标框 (gt) 间的交并比 (IoU) 值。

3.3 结果分析

表 1 给出了本文检测模型与轻量化 NanoDet 检测模型在生猪脸部测试集上的检测性能对比结果。在 400 张测试图像共计 1 346 个生猪脸部目标的测试数据上, 本文模型在 IoU 为 0.5、0.75 和 0.50:0.95 水平下, 分别获得了 96.98%、81.16% 和 67.44% 的平均精度, 并取得了最大检测目标个数为 100 水平下 72.82% 的平均召回率。相较于 NanoDet 模型, 本文模型在模型大小仅提升 0.5 MB 的前提下, 在各指标上均有一定的提高。在增加 Nonlocal-PAN 特征融合模块输入与输出特征通道数 (取 $C = 192$) 时, 模型检测性能可以得到进一步提升, 但随之模型大小也有相对提升。

表1 本文模型与轻量化 NanoDet 模型对生猪脸部检测性能的对比

Tab. 1 Comparison of pig face detection performance between proposed model and NanoDet model

方法	AP		AR		模型大小/ MB
	@ 0.50/ %	@ 0.75/ %	@ 0.50;0.95/ %	@ 100/ %	
NanoDet	96.22	77.81	65.85	71.37	7.7
本文模型	96.98	81.16	67.44	72.82	8.2
本文模型 ($C = 192$)	96.74	81.84	68.21	73.40	11.8

表2 给出本文模型和文献[5]中使用的 Tiny-YOLO 模型在生猪脸部检测时的不同性能对比。相较于文献[5]所用数据集,本文数据中存在遮挡和视角变化情况更为突出,且在模型大小更小的情况下,本文模型仍然取得了较为优势的生猪脸部检测性能。

表2 不同生猪脸部检测模型性能对比

Tab. 2 Comparison of different pig face detection models

方法	mAP / %	测试图像 数量	生猪脸部 数量	模型大小/ MB
改进的 Tiny-YOLO ^[5]	82.38	420	2 106	≥ 20
本文模型	96.98	400	1 346	8.2

本文模型在两种不同养殖环境下生猪脸部目标检测结果如图7所示。从图中可以看出,对于存在遮挡、视角变化的生猪脸部,本文模型在不同养殖环境下都能较为有效的检测出其位置。



图7 不同养殖环境下生猪脸部目标检测结果

Fig. 7 Pig face detection results of different breeding condition

4 结束语

本文提出了一种轻量化的生猪脸部目标检测模型,通过在特征融合网络 PAN 结构中引入 Nonlocal 上下文建模与通道注意力机制,融合了图像中视觉上下文信息与跨通道信息,提高了对存在遮挡与视角变化的生猪脸部目标的检测性能。与现有轻量化目标检测模型相比,在不同 IoU 水平下,本文所提模型的平均精度均有所提升,且与现有生猪脸部检测模型相比,本文所提模型在具有更小模型大小的同时,提升了模型检测的性能。实验结果表明本文模型具有良好的边缘计算端支持特性和实用性。未来工作将改进损失函数的设计,进一步提高对密集遮挡环境下的生猪脸部目标检测性能。

参考文献

- [1] CAMERLINK I, COULANGE E, FARISH M, et al. Facial expression as a potential measure of both intent and emotion [J]. Scientific Reports, 2018, 8(1): 17602.
- [2] WADA N, SHINYA M, SHIRAISHI M. Pig face recognition using eigenspace method [J]. ITE Transactions on Media Technology & Applications, 2013, 1(4): 328-332.
- [3] 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.
- [4] MARSOT M, MEI J, SHAN X, et al. An adaptive pig face recognition approach using convolutional neural networks [J]. Computers and Electronics in Agriculture, 2020, 173: 105386.
- [5] 燕红文, 刘振宇, 崔清亮等. 基于改进 Tiny-YOLO 模型的群养生猪脸部姿态检测 [J]. 农业工程学报, 2019, 35(18): 169-179.
- [6] HANSEN M F, SMITH M L, SMITH L N, et al. Towards on-farm pig face recognition using convolutional neural networks [J]. Computers in Industry, 2018, 98: 145-152.
- [7] RANGILYU. NanoDet [EB/OL]. <https://github.com/RangiLyu/nanodet>, 2021.
- [8] TZUTALIN. labelImg [EB/OL]. <https://github.com/tzutalin/labelImg>, 2021.
- [9] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context [C] // 2014 European Conference on Computer Vision. Springer, Cham, 2014: 740-755.
- [10] MA N, ZHANG X, ZHENG H T, et al. ShuffleNet V2: practical guidelines for efficient cnn architecture design [C] // 2018 European Conference on Computer Vision. Springer, Cham, 2018: 122-138.
- [11] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [12] YJH0410. YOLO - Nano [EB/OL]. <https://github.com/yjh0410/YOLO-Nano>, 2021.
- [13] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C] // 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.

(下转第94页)