

# 群养猪侵略性行为的深度学习识别方法

高 云<sup>1,2</sup>, 陈 斌<sup>1</sup>, 廖慧敏<sup>1</sup>, 雷明刚<sup>2,3</sup>, 黎 煊<sup>1,2</sup>, 李 静<sup>1</sup>, 罗俊杰<sup>1</sup>

(1. 华中农业大学工学院, 武汉 430070; 2. 生猪健康养殖协同创新中心, 武汉 430070;

3. 华中农业大学动物科技学院动物医学院, 武汉 430070)

**摘 要:** 为了解决因传统机器视觉和图像处理方法的局限性以及复杂的猪体姿态和猪舍环境导致对群养猪侵略性行为识别的有效性、准确率较低的问题, 该文基于深度学习的方法, 提出使用 3D CONV 的群养猪侵略性行为识别算法-3DConvNet。分 3 个批次采集 18 头 9.6 kg 左右的大白仔猪视频图像, 选用第一批次中包含 28 d 内各个时段的撕咬、撞击、追逐、踩踏 4 大类, 咬耳、咬尾、咬身、头撞头、头撞身、追逐以及踩踏 7 小类侵略性行为以及吃食、饮水、休息等非侵略性行为共计 740 段 (27 114 帧) 视频作为训练集和验证集, 训练集和验证集比例为 3:1。结果表明, 3D ConvNet 网络模型在训练集上的识别准确度达 96.78%, 在验证集上识别准确度达 95.70%。该文算法模型对于不同训练集批次的猪只以及不良照明条件下依然能准确识别侵略性行为, 算法模型泛化性能良好。与 C3D 模型进行对比, 该文提出的网络模型准确率高出 43.27 个百分点, 单帧图像处理时间为 0.50 s, 可满足实时检测的要求。研究结果可为猪场养殖环境中针对猪只侵略性行为检测提供参考。

**关键词:** 卷积神经网络; 机器视觉; 模型; 行为识别; 侵略性行为; 深度学习; 群养猪

doi: 10.11975/j.issn.1002-6819.2019.23.024

中图分类号: TP391.41

文献标志码: A

文章编号: 1002-6819(2019)-23-0192-09

高 云, 陈 斌, 廖慧敏, 雷明刚, 黎 煊, 李 静, 罗俊杰. 群养猪侵略性行为的深度学习识别方法[J]. 农业工程学报, 2019, 35(23): 192—200. doi: 10.11975/j.issn.1002-6819.2019.23.024 http://www.tcsae.org

Gao Yun, Chen Bin, Liao Huimin, Lei Minggang, Li Xuan, Li Jing, Luo Junjie. Recognition method for aggressive behavior of group pigs based on deep learning[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(23): 192—200. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2019.23.024 http://www.tcsae.org

## 0 引 言<sup>1</sup>

集约化的养猪中侵略性行为 (包括打斗, 追逐等) 经常在猪群中发生, 侵略性的行为易对猪只身体造成伤害, 在恶劣的猪舍环境中引起猪只感染, 严重时导致死亡, 对猪场造成损失<sup>[1-3]</sup>。侵略性行为的发生还会影响猪舍内的食物配给, 猪群体系中弱势等级的猪只缺失食物和饮水进给, 导致猪只生长缓慢, 影响母猪繁殖力, 造成严重的经济损失<sup>[4-7]</sup>。目前猪场对于猪只侵略性行为的监测识别主要是依靠人工观察记录, 这在集约化养殖的猪场内会产生高昂的人工成本, 且针对侵略性行为的观察记录会不可避免的造成大量的漏检, 也无法保证集约化的猪场内对于侵略性行为监测的实时性, 准确性和高效性。集约化环境下的群养猪侵略性行为的自动检测识别, 是提高猪只福利及经济效益的重要基础<sup>[8-11]</sup>。

目前, 国内外已有许多研究者针对猪只侵略性行为的检测识别做了相应的研究。Oczak 等利用图像处理技术和人工神经网络针对猪舍内发生的高、中程度的侵略性行为的分类问题进行了研究, 实现了高、中程度的侵略性行为的分类, 但是需要对图像两帧的像素变化进行计

算提取特征, 会产生大量的计算, 无法对大批量数据的处理保持高效性<sup>[12]</sup>。Viazzi 等提出的采用线性判别分析 (linear discriminant analysis, LDA) 对猪只运动历史图像中提取的特征进行分类来识别侵略行为, 该方法需要先获取所有猪只的历史运动图像然后从中提取相关运动强度特征, 同样导致了当样本量很大时产生大量的计算代价, 且 LDA 对于样本分类过分依赖均值信息, 实际猪舍环境内猪只只会发生很多运动强度很小的侵略性行为, 如咬尾、咬耳等, 其泛化能力还有待验证<sup>[13]</sup>。Chen 等使用层次聚类算法提取猪只的加速度特征, 用于识别侵略性行为。猪只的加速度是侵略性行为的重要特征, 但是在该方法中作者仅关注了视频帧中最先开始发生侵略性行为的猪只, 丢弃了其余非侵略性的猪只, 这导致其余猪只的信息被完全抛弃, 而这部分信息是有可能发生侵略性行为的<sup>[14]</sup>。Jonguk 等使用支持向量机 (support vector machine, SVM) 处理运动猪只速度有关的 5 个特征, 实现了对侵略性行为发生与否的识别, 虽然该研究实现了较高的识别准确度, 但是仅是针对追逐和敲击两种侵略性行为, 且提取猪只速度的特征需进行额外计算, 也存在一定计算代价问题, 难以实现实时检测<sup>[15]</sup>。以上学者关于猪只侵略性行为识别的研究都是基于图像处理技术提取猪只图像中的某一特征, 再结合机器学习等手段对特征进行处理。但在实际应用中, 传统的图像处理技术需要额外提取特征, 存在效率低、工作量大的问题。由于不同种类猪只个体差异性大, 随着猪只质量的增长,

收稿日期: 2019-07-10 修订日期: 2019-10-29

基金项目: “十三五”国家重点研发计划项目 (2016YFD0500506); 中央高校自主创新基金 (2662018JC003, 2662018JC010, 2662017JC028)

作者简介: 高 云, 副教授, 博士, 主要从事农业智能检测与控制方面的研究。Email: angelclouder@mail.hzau.edu.cn

猪只非刚性的躯体也会随着发生变化,故提取的特征可能不具有普适性。此外由于猪舍内猪体粘连、遮挡、不良照明条件以及猪只的侵略性行为复杂等原因,传统的方法难以在集约化养殖的猪场内做到对于猪只侵略性行为的实时、高效的检测。

近几年深度学习相比于传统方法在图像和视觉领域展现了其强大的优越性,深度学习通过对低维特征到高维特征的提取学习,能够做到对绝大部分场景下各类任务进行检测识别<sup>[16-20]</sup>。正是在其他领域展现了其强大的学习泛化能力,深度学习也在猪只行为检测方面得到了大量的应用。Yang 等基于 Fast R-CNN 检测猪只进食行为<sup>[21]</sup>。Yang 等使用全卷积神经网络的方法分割母猪和仔猪,使用母猪几何特征和仔猪的动态计算乳房区域以及提取对应的空间信息,再从视频帧中提取运动强度和占领指数以识别母猪母性护理行为<sup>[22]</sup>。杨秋妹等使用卷积神经网络针对个体猪只饮水行为做出相应的研究<sup>[23]</sup>。Zheng 等使用 Faster R-CNN 来对母猪的站立、躺卧等行为做出识别<sup>[24]</sup>。深度学习在猪只的简单行为上均展现了优异的性能,但目前针对存在多头猪只状态交互的较高级的侵略性行为研究还比较少见<sup>[8]</sup>。

本文采用深度学习的方法,搭建 3D 卷积神经网络模型,并将其用于对群养猪中侵略性行为的识别,避免了传统图像处理方法中复杂、繁琐的特征选择、处理等问题。通过对网络模型进行训练,得到了一个端到端的,有效识别侵略性行为的模型,并通过对不同批次猪只、视频长短、不良照明条件的试验,验证算法的泛化性与可行性。

## 1 材料与方法

### 1.1 猪只侵略性行为的定义

猪只的侵略性行为涉及到群养猪中多头猪只的状态交互,是一个复杂,渐进的行为。在发生侵略性行为的初期,猪只通过鼻子嗅闻、轻推等方式进行初步试探,随后侵略性行为逐渐加剧,往往伴随着更激烈的挤压,咬,撞击等行为<sup>[2]</sup>。在侵略性行为最剧烈的时候,会出现咬耳朵,咬身体等现象。撕咬通常持续时间长,往往会造成皮肤损伤,创口等<sup>[1,6,22]</sup>。故在此项研究中定义的侵略性行为是撕咬,挤压,撞击,追逐,如表 1 所示。

表 1 侵略性行为的定义

Table 1 Definition of aggressive behavior

侵略种类 Type of aggressive	侵略性行为名称 Name of aggressive behavior	行为描述 Behavior description
撕咬 Bite	咬耳朵 Bite the ear	咬另外一只猪只的耳朵
	咬尾 Bite the tail	咬另外一只猪只的尾巴
	咬身体 Bite the body	咬另外一只猪只的身体部位,包含嘴部,颈部等
撞击 Knock	头对头撞击 Head to head knock	猪只用头部撞击另一只猪的头部
	头对身撞击 Head to body knock	猪只用头部撞击另一只猪的身体部位
追逐 Chase	追逐 Chase	在发生任一侵略性事件后导致的追逐
踩踏 Tread	踩踏 Tread	猪只用脚踏踩其他猪只的头部、身体等

### 1.2 数据采集

#### 1.2.1 试验条件

试验数据采集分别于 2018 年 3 月 12 日—4 月 9 日,2018 年 4 月 19 日—5 月 16 日以及 2018 年 6 月 9 日—7 月 6 日在湖北省武汉市华中农业大学试验猪场内进行 3 个批次的数据采集。3 个批次氨气浓度分别为<3.80, 15.18, 37.95 mg/m<sup>3</sup>。其中<3.80 mg/m<sup>3</sup>组为模拟仔猪舍在通风良好的条件下的氨气浓度。因猪舍内总有粪便,会产生一定的氨气,因此将试验环境的氨气控制到<3.80 mg/m<sup>3</sup>作为对照。试验对象为 18 头 9.6 kg 左右的大白保育猪,猪只外观颜色均匀。18 头猪养殖在课题组前期研究设计的环境多变量控制养殖箱中<sup>[25]</sup>,养殖箱内部空间长、宽、高为 2、1.5、2 m,底部粪槽深度设计为 0.45 m,养殖箱内部设有进食槽、饮水槽。试验中温度控制在 27.0~27.4 ℃,相对湿度控制在 50%~70%。为保证猪只正常生活习性,每天 8:00 与 17:00 喂食两次,饲喂模式一致。箱内采用自动控制的 LED 灯照明,照明时间设定为 7:00 至 18:00,其余时间 LED 灯关闭。

#### 1.2.2 视频采集

试验通过 Kinect V2 摄像头采集 RGB 视频,位于养殖箱的上部。镜头距养殖箱内部地板高度约为 1.8 m,采用顶视角度的方式,可以获取整个养殖箱内部 6 头猪只的全部信息,不会出现漏拍猪只现象。摄像头连接一台便携式笔记本电脑,将采集的养殖箱内部 6 头猪只的顶视角彩色视频存储在 SEAGATE 移动硬盘上,出于存储成本的考虑,录制视频帧率为 5 帧/s,存储像素为 1 920×1 080,存储为 AVI 格式。数据采集平台及养殖箱如图 1 所示。

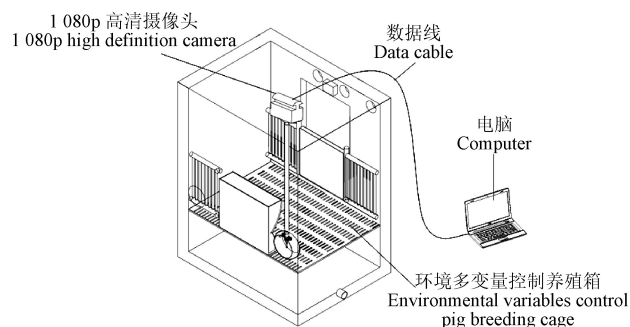


图 1 数据采集平台

Fig.1 Data acquisition platform

#### 1.2.3 视频数据预处理与标注

为了训练和评估侵略行为识别模型的性能,需要将获取的数据进行标注。通过人工查阅采集的约 900 h 的视频数据,对视频进行处理,分为侵略性行为和 2 种类别,视频长度取决于猪只侵略性行为持续的时间。根据前文中侵略性行为的定义,为区分侵略性行为与非侵略行为,仅将至少持续 5 帧的侵略行为进行标注。在一个侵略性行为发生之后的 5 s 内若再有侵略性行为发生,则将其归为同一次侵略性行为,即设置侵略性行为间隔为 5 s。将录制的视频中存在掉帧现象的视频段舍弃。最终定义侵略性行为视频时长最少为 3 s。

通过在 Python 3.6 上编辑的程序代码,对视频进行标注。将标注的数据集随机划分为训练集,验证集以及测试集,并获取对应的训练集、验证集以及测试集文档目录,网络模型将会依赖这个目录实现对数据集的读取。

#### 1.2.4 数据集介绍

根据表 1 中定义的侵略性行为,对视频图像进行了剪辑分段处理。在群养猪侵略性行为发生过程中,常涉及到多种侵略性行为同时发生,或是一个行为结束后另一个行为的再次发生,故视频中至少包含一种所定义的侵略性行为,经过细致的分类后最终得到的 3 批数据中对应的侵略性行为发生的次数统计如表 2 所示。

表 2 各类侵略性行为统计  
Table 2 Various aggressive behaviors counting

行为分类 Behavior classes	第一批数据中 次数 Number of times in the first batch	第二批数据中次数 Number of times in the second batch	第三批数据中次数 Number of times in the third batch
咬耳朵 Bite ear	364	91	146
咬尾 Bite tail	120	37	32
咬身体 Bite body	268	76	107
头对头撞击 Head to head hit	161	64	79
头对身撞击 Head to body hit	149	58	68
追逐 Chase	104	37	53
踩踏 Trample	50	32	29

根据表 2 可以看出,在所有的行为类别中咬耳和咬身体行为出现的次数最多,在 3 期试验完成后均发现多数猪只的耳部和部分猪只身上都有伤口,无疑侵略性行为对猪只的健康和福利造成了严重的影响。

为了训练和评估所搭建的网络,将采集的第一批在 2018 年 3 月 12 日—4 月 9 日,在  $37.95 \text{ mg/m}^3$  的氨气浓度下采集的试验数据按照 60%、20%、20% 的比例作为网络的训练集、验证集与测试集。后面两批试验数据(2018 年 4 月 19 日—5 月 16 日,  $15.18 \text{ mg/m}^3$ ; 2018 年 6 月 9 日—7 月 6 日,  $<3.80 \text{ mg/m}^3$ ) 中选取一部分侵略性行为与非侵略性行为制作为测试集,数据集详情如表 3 所示。

表 3 数据集划分  
Table 3 Dataset partition

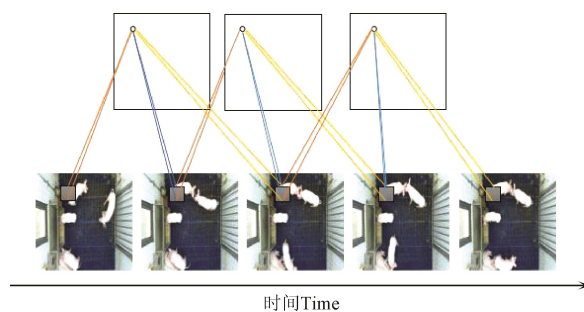
数据集类别 Dataset category	行为分类 Behavior classes	数量(段/帧数) Number (epoch/frames)	获取时间 Acquisition time
训练集 Train dataset	侵略性行为 Aggressive	285/10 556	2018 年 3 月 12 日— 4 月 9 日
	非侵略性行为 None-Aggressive	270/9 364	
验证集 Validation dataset	侵略性行为 Aggressive	95/3 518	
	非侵略性行为 None-Aggressive	90/3 676	
测试集一 Test dataset 1	侵略性行为 Aggressive	175/6 230	2018 年 4 月 19 日— 5 月 16 日
	非侵略性行为 None-Aggressive	170/6 770	
测试集二 Test dataset 2	侵略性行为 Aggressive	161/4 862	
	非侵略性行为 None-Aggressive	160/4 860	
测试集三 Test dataset 3	侵略性行为 Aggressive	220/6 727	2018 年 6 月 9 日— 7 月 6 日
	非侵略性行为 None-Aggressive	180/6 098	

### 1.3 侵略性行为识别算法

#### 1.3.1 2D 与 3D CONV 网络模型

现有的采用深度学习方法针对猪群行为的研究,通常是使用常规的 2D 卷积核(2D CONV)搭建卷积神经网络。2D CONV 是针对单张图片进行卷积操作,提取的是图片的空间特征。针对猪只的分割,识别,行为检测取得了很好的成果<sup>[21-24,26]</sup>。但是针对猪只的侵略性行为识别而言,仅通过单张图片进行识别并不准确,侵略性行为是一个随时间进行的一个完整的行为,如果只在一帧图像上得出结果对侵略性行为做出判断,这样将会丢失了侵略性行为在时间维度上的运动信息,导致很高的错误识别率,难以做出有效的判断。所以针对猪只侵略性行为的识别,需要结合时间和空间维度上的信息。

3D 卷积核(3D CONV)是一种在 2D CONV 中加入了时间维度信息的卷积,如图 2 所示是 3D CONV 对于视频帧进行卷积运算时的操作,在对图像中的特征进行学习的卷积运算过程中,3D CONV 多了在时间维度上的运算,其卷积核大小为  $d \times d \times k$ 。卷积核在对当前帧图像进行卷积操作时,还会在时间维度上对接下来  $k-1$  帧图像进行相应的卷积操作,即提取了时间序列上  $k$  帧图像的信息。 $d, k$  取决于网络所定义的卷积核的尺寸。3D CONV 使得卷积提取的特征融合了在时间维度上的附近时间域上的信息,保留了运动信息,为模型提取时间和空间上的特征奠定了基础。3D CONV 搭建的卷积神经网络依然具有 2D CONV 的局部连接,权值共享,多层次结构等优点,这使得使用 3D CONV 的卷积神经网络具有很强大的学习能力<sup>[27]</sup>。



注:点、线代表均代表卷积运算中的计算过程。

Note: The points and lines represent the calculation process in the convolution operation.

图 2 3D 卷积计算时间维度为 3 的运算

Fig.2 Process of 3D convolution operation with the time dimension is three

#### 1.3.2 3D CONV 侵略性行为识别网络模型

本文在 C3D<sup>[28]</sup>网络结构的基础上,对 C3D 网络结构进行重新构建和优化,通过比较不同网络层数和卷积核大小对网络模型准确度的影响,确定最佳的识别群养猪侵略性行为的网络参数和模型。

如图 3 所示,最终提出的模型为 3D CONVNet,在 C3D 的基础上做了 4 项改进:

1) C3D 网络只有 8 层,对于卷积神经网络而言,网络的层数越深所能提取到的特征越多,所学习到的特征也越来越抽象。针对本文的侵略性行为识别的任务而言,

需要低级和高级抽象的特征相结合才能对任务做出准确识别。所以在充分考虑计算代价和模型性能的情况下, 将网络层数加深到了 19 层。

2) 3D CONV 与 2D CONV 相比, 在时间维度上多了一个深度为  $k$  大小的卷积运算, 导致运算量剧增, 故需要大量数据进行训练得到一个较优的模型。在数据量不充足的情况下, 网络通常会遭受过拟合的风险。所以在网络结构中加入了 Dropout<sup>[29]</sup>, 这不仅可以有效的防止过拟合, 还显著减少了计算代价, 可以更容易地去添加卷积层数以学习更有意义的特征, 还增强了网络的鲁棒性。

3) 在网络结构中添加了 Batch Normalization (BN) 层<sup>[30]</sup>, 这是为了保证各层的参数数据分布不会随着卷积操作的进行发生剧烈的变化, 网络在一个不变的数据分布范围内更容易学习到有用的特征。此外, 可有效避免在深层卷积神经网络中可能出现的梯度消失问题, 还可提高网络训练速度。

4) 在网络中采用了多尺度特征融合的方法, 多尺度特征融合在最新的目标检测算法 SSD 和 YOLO v3 中都得到了应用<sup>[31-33]</sup>, SSD 和 YOLO v3 都是目前在目标检测方面最先进的算法。在网络中融合不同尺度的特征是提高模型性能的重要手段, 采用的特征融合让模型充分利用到了各个阶段所提取的时间和空间上的特征, 在学习更抽象、更高级的特征时, 仍然结合了低层的信息。这有效利用了各层卷积核所提取特征的侧重点不同的特性。多尺度特征融合的引入, 让网络融合更多的低层信息, 对于网络模型待解决的涉及到低层与高层特征相关联的识别任务而言, 起到了关键作用。

如图 3a 所示, 所提出的侵略性行为识别网络由前置特征提取网络, 中间特征融合提取网络以及最后的输出网络三部分组成。

前置特征提取网络由 3 个卷积块组成, 前置特征提取网络及卷积块的构成如图 3b 所示。前置特征提取网络在网络进行更进一步的特征提取和特征融合之前, 提取出一些有效的特征, 减少有可能存在的噪声, 减少一些无效信息对于模型性能的影响。前置特征提取网络第一个卷积块卷积核大小为  $3 \times 3 \times 1$ , 第二、三卷积块大小均为  $3 \times 3 \times 3$ , 卷积块输出通道数逐渐增多, 分别为 16, 32, 64。为了在进行特征融合之前, 保留更多的有效特征信息, 所以在第一个卷积操作时, 并不对时间序列上的运动信息进行采集, 更多的利用当前帧的信息, 且采用的 Max-pooling 在前置特征提取网络的步长为 (2,2,1), 这会让网络更多的保留当前所提取的特征图信息, 而不是和后序时序特征融合, 让网络保留更多的当前帧图像的特征。卷积层后增加了 Batch Normalization 层、Relu 激活层, 以及 Max-pooling 层。

多尺度特征融合由特征融合提取网络完成, 如图 3c 所示。在这个特征融合提取阶段, 主干网络上设置了 3 个卷积阶段, 网络仍然会继续提取更深层次的特征。为了避免一些无效的特征被多次计算, 导致计算代价的提高和网络模型性能的下降, 仅在一个卷积阶段完成后进行特征融合。在每个卷积阶段内设置了 5 层卷积, 卷

积核的大小均为  $3 \times 3 \times 3$ , 卷积通道数依次增长到 64, 128, 256, 512, 卷积步长均为 (1, 1, 1)。在跨越式特征融合里, 因为特征的通道数并不相同, 所以并不能直接进行融合, 在特征传递的连接中设置了尺寸为  $1 \times 1 \times 1$  的卷积核, 以保证特征融合时的通道数一致。

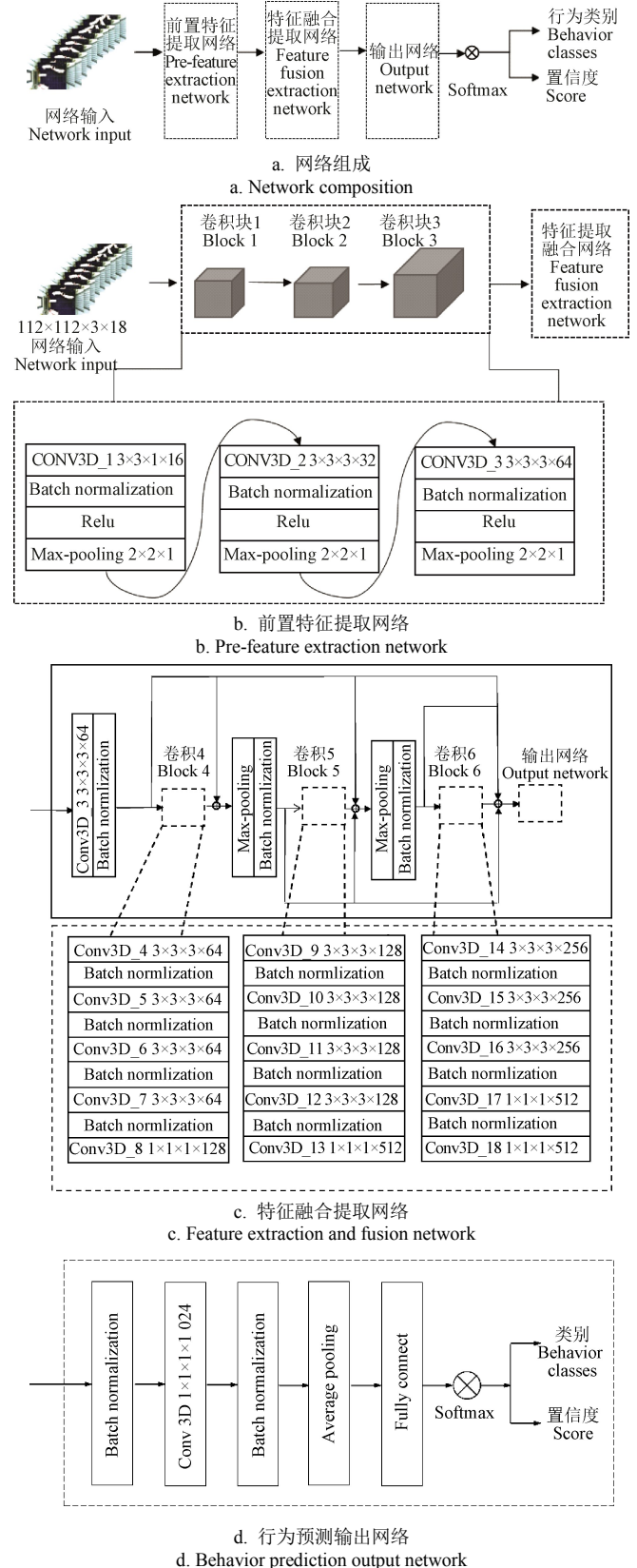


图 3 网络结构

Fig.3 Network structure



输出网络如图 3d 所示, 在做最终的输出预测之前, 网络对融合的特征进行了一次卷积操作, 卷积核大小为  $1 \times 1 \times 1$ , 步长为  $(1, 1, 1)$ , 并将网络的通道数提升到 1 024, 这有利于网络充分整合时间和空间上的信息。在连接到全连接层之前对特征图进行平均池化处理, 然后经全连接层后输入到 Softmax 完成最终的类别预测和置信度计算。

网络的任务是识别侵略性行为与非侵略性行为, 属于一个二分类问题。网络采用 Adam 梯度下降法的反向传播更新优化模型, 网络的损失函数采用 categorical\_crossentropy loss, 如式 (1) 所示。

$$L = -(y \lg \hat{y} + (1 - y) \lg(1 - \hat{y})) \quad (1)$$

式中  $L$  为损失函数;  $y$  为真实的样本标签, 0 或 1;  $\hat{y}$  为模型预测的样本标签。

### 1.3.3 训练参数设置

网络采用的激活函数均为整流线性单元 (rectified linear unit, relu), 采用的优化算法为 Adam 梯度下降法, batch\_size 设定为 32, momentum 为 0.9, 设置迭代次数为 20, 基础学习率为 0.005, Dropout 失活率为 0.5, 使用 L2 正则化函数, 正则化权重衰减系数 weight\_decay 为 0.005。

### 1.4 评价指标

为了全面、合理的对网络模型性能做出评价, 采用准确率 (Accuracy), 查准率 (Precision), 召回率 (Recall) 和 F1 值 4 个指标来评估模型的性能, 如式 (2) ~ (6) 所示。

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$F1 = \frac{2 \cdot TP}{\text{样本总数} + TP - TN} \quad (5)$$

式中 TP 是正确识别侵略性行为的个数, TN 是正确识别非侵略性行为的个数, FP 是将非侵略性行为识别位侵略性行为的个数, FN 是将侵略性行为识别为非侵略性行为的个数。

### 1.5 试验流程

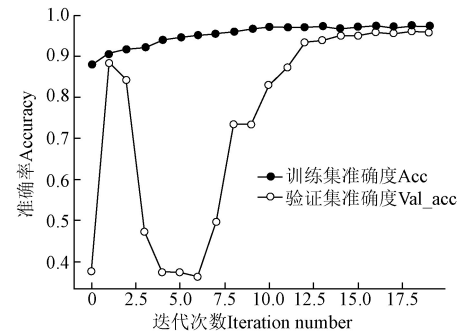
群养猪侵略性行为的识别检测试验步骤如下。

1) 从采集的数据中提取出包含侵略性行为的视频片段, 制作出训练集, 验证集和测试集; 2) 搭建群养猪侵略性行为识别检测网络; 3) 将制作好的训练集作为网络的输入进行训练; 4) 使用验证集对网络模型进行测试, 得到 loss 函数以及网络识别准确度; 5) 根据得到的网络识别准确度作为评价指标, 调整网络参数, 如学习率, Batch\_size, weight\_decay 等; 6) 调整参数后再次训练网络, 重复 3) ~ 5), 直至 Loss 曲线收敛, 训练集与验证集识别准确度相近。

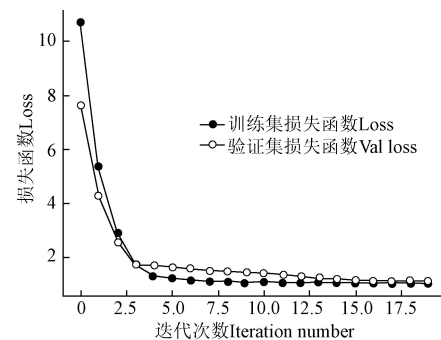
## 2 结果与分析

### 2.1 模型性能分析

在试验平台上, 使用搭建的 3D CONVNet 对训练集数据集进行 20 次迭代的训练, 同时也采用验证集对网络模型训练状况做一个初步的评价。网络在包含 380 个侵略性行为片段 (14 074 帧), 360 个非侵略性行为片段 (13 040 帧) 的数据集上的训练集和验证集识别准确度和模型 loss 曲线如图 4 所示。



a. 模型准确度曲线  
a. Model accuracy curve



b. 模型 loss 曲线  
b. Model loss curve

图 4 模型训练曲线  
Fig.4 Model training curve

由图 4 可以看出, 网络存在一个明显的训练优化过程, 随着迭代次数不断增加, 训练集和验证集的分类误差 Loss 曲线逐渐降低, 迭代至 15 次时, Loss 曲线趋近于收敛。模型在训练集上的准确度达到了 96.78%, 在验证集上的准确度也由刚最初的 37% 逐渐上涨至 95.70%。从第 10 次迭代过后, 训练集和验证集的准确度的差距逐渐减小, 最终两者的准确度差在一个良好范围内, 在迭代 20 次过后, Loss 和识别准确度基本不再变化。识别准确度曲线以及 Loss 曲线展现了一个良好的深度学习模型的训练过程, 模型逐渐学习到正确识别群养猪侵略性行为的特征, 达到了一个较好的训练效果, 且模型没有陷入过拟合和局部最优。

模型对群养猪侵略性行为有较好的识别性能。如图 5 所示, 是随机抽取的模型对图像帧中侵略性行为和侵略性行为的识别效果, 图 5a 是猪只在进食状态中, 系统判定为非侵略性行为 (Nor), 置信度为 0.967; 图 5b 为发生进食行为之外又发生撕咬行为, 系统判定为侵略性

行为（Attack），置信度为 0.821。

对于追逐、撞击、踩踏、咬身体等侵略性行为，模型均具有良好的表现。模型对于不同环境状态下的侵略性行为识别有很好的鲁棒性，如在群养的 6 头猪只中，侵略性行为仅发生在 2 头猪只中，其余 4 头猪只处于进食，饮水等非侵略性行为下，模型可准确识别出侵略性行为的发生；在一次侵略性行为发生后的几秒钟之后，猪群中再次发生侵略性行为，模型也可准确识别出侵略性行为；在涉及到多头猪只的侵略性行为亦可准确识别。此外，模型还可以对当前行为做出判断的同时给出一个置信度分数，这更有利于网络模型对于识别是否为侵略性行为提供依据和可解释性。

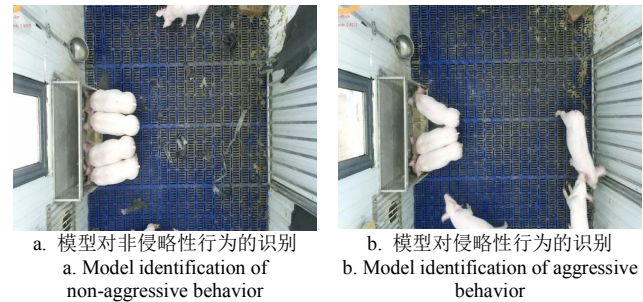


图 5 模型识别结果  
Fig.5 Model identification result

2.3 实际效果的检验

为进一步验证模型性能，需要在实际猪舍环境下检测算法的有效性以及泛化性能。本节试验以相同养殖环境下不同猪只、不同视频段时长、不良照明条件等因素，对模型性能进行试验评估。

2.3.1 不同批次猪只对模型性能的影响

在实际的群猪养殖环境中，不同批次的同类猪只虽总体上没有明显区别，但是不同猪只受到环境等外界因素影响仍存在一定差异，如形体，身体部位轮廓，行为习惯等。针对不同批次猪只，使用了测试集一、测试集二以及测试集三对模型做出评价，表 4 是模型在 3 个测试集上的测试结果的混淆矩阵。

表 4 模型在测试集上结果的混淆矩阵  
Table 4 Confusion matrix of model's result on test set

数据集 Dataset	行为类别 Behavior classes	分类结果 The result of classification		总计 Total
		侵略性行为 Aggressive	非侵略性行为 None-aggressive	
		侵略性行为 Aggressive	非侵略性行为 None-aggressive	
测试集一 Test dataset 1	侵略性行为 Aggressive	165	10	175
	非侵略性行为 None-aggressive	13	157	170
测试集二 Test dataset 2	侵略性行为 Aggressive	144	17	161
	非侵略性行为 None-aggressive	9	151	160
测试集三 Test dataset 3	侵略性行为 Aggressive	189	31	220
	非侵略性行为 None-aggressive	5	175	180
总计 Total	侵略性行为 Aggressive	498	58	556
	非侵略性行为 None-aggressive	27	483	510

从表 3 中统计出在全部测试集的 1 066 个视频片段中，总共 981 段（侵略性+非侵略性）视频被正确分类，85 段（侵略性+非侵略性）视频分类错误。准确率为 92.03%，侵略性行为的查准率指标为 94.86%，侵略性行为召回率指标为 89.57%，调和 Recall 与 Precision 的 F1 值为 92.14%。测试集一在 3 个测试集中取得了最优的表现性能，侵略性行为识别准确率为 94.29%。这是由于测试集一与训练集是来源于同一批次猪只，故模型对该批次猪只数据拟合得更好。测试集二、三与训练集虽来源于不同批次的猪只，但是通过统计计算得到测试集二、三的准确度指标分别为 89.44%，85.91%。这充分说明了模型的泛化性能良好，在针对同一养殖环境下的不同猪只，该模型依旧可以以较高的准确地识别侵略性行为的发生与否。

2.3.2 视频长短对模型性能的影响

因为侵略性行为发生的动因、种类以及猪只个数差异等原因，导致发生侵略性行为的时间长短不一。为分析侵略性行为持续时长对模型识别侵略性行为的性能的影响，将测试集二、测试集三中的侵略性行为视频按时间段进行分类，不同时间段的分布及分类测试结果如图 6 所示。

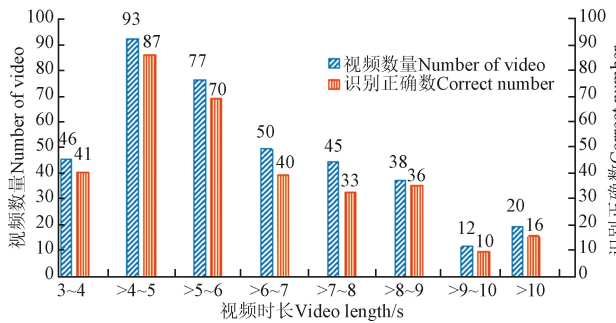


图 6 不同侵略性行为持续时长的测试结果  
Fig.6 Test result of different aggressive behavior duration

从图 6 中统计计算得出，视频段时长主要分布在 >4~7 s 内（220 段）。这一时长段的视频在测试集总数中的占比达到了 57.74%，且识别准确度达到了 89.55%，其中又以 >4~5 s 内的视频段量最多，占比达到了 24.41%，识别准确度达到了 93.55%。3~4 s 内的视频与 >6~7 s 内的视频占比基本一致，且识别准确度均达到 80 % 以上。但是从 6~7 s 及 7~8 s 的视频中，视频段占比呈逐渐减小的趋势，识别准确度也是基本逐渐降低的趋势，在 7 s 及以上的视频段中仅取得了 73.33% 的识别准确度。

对于视频误识别的原因，主要有三点，一是在侵略性行为发生的过程中，受侵略猪只的身体部位特征被侵略性猪只所遮挡，如耳朵，尾部等部位，模型没有检测到受侵害部位，所以无法判断是否发生侵略性行为；二是侵略性行为过于温和，帧与帧之间基本无明显变化，模型无法捕获时间维度上的运动信息，做出了错误的识别结果；三是对于较长的视频，3D CONVNet 对视频的逐帧处理时间过长，容易丢失时间维度上的运动信息，导致时间维度上的运动信息对模型进行识别的作用有

限,网络过分依赖于空间维度上的信息,导致了长视频的较高的误识别率。

### 2.3.3 不良照明条件对模型性能的影响

试验中,根据猪只作息习惯提供照明的 LED 灯在 7:00—18:00 开启,其余时间仅通过养殖箱侧壁窗口的自然光照明,但在实际的观察中发现有部分侵略性行为发生在照明条件不佳的情况下。为了评估模型在不良照明条件下的侵略性行为识别的表现,在测试集中将不良照明条件下的视频段进行筛选,将这批数据单独进行测试,最终得到的结果如表 5 所示。

表 5 不良照明条件视频识别结果  
Table 5 Result of poor-lighting condition video

行为类别 Behavior Classes	总计 Total	正确识别 Correct classification	错误识别 Wrong classification	准确率 Accuracy/%
侵略性行为 Aggressive	43	34	9	79.07
非侵略性行为 None-aggressive	27	27	0	100
总计 Total	70	61	9	87.14

由表 5 可以看出,模型对于不良照明条件下侵略性行为识别准确率依然达到了 79.07%。在实际的猪只养殖过程中,不良照明现象是普遍存在的,而本文所提出的侵略性行为检测模型,对于猪只侵略性行为识别依然可以在不良照明条件下取得较优的结果,体现了模型对于光照条件的良好的适应性,更能促进模型在实际的集约化养殖猪场内运用。

### 2.4 模型参数设置

本文提出的 3D ConvNet 特征融合提取网络中卷积核大小为  $3 \times 3 \times 3$ ,卷积块内部层数为 5,该网络参数设置是对特征融合提取网络的不同网络参数进行试验验证的基础上,确定的最优的识别侵略性行为的网络参数。通过对比试验发现,当卷积核大小为  $3 \times 3 \times 3$ ,卷积块内部卷积层数为 5 时,网络取得了 95.70% 的识别准确度。卷积核大小为  $3 \times 3 \times 1$  时,识别准确度仅为 49.22%。卷积块内部卷积层数设置为 9 层时,模型待训练参数增加至  $7\,401 \times 10^3$ ,识别准确度仅为 63.67%。

### 2.5 不同模型对比分析

将本文网络模型与 C3D 模型以及其他基于 C3D 网络的改进模型的检测识别效果进行对比。采用的训练集与验证集均为前文所述的数据集。训练参数的设置均保持一致。4 个网络模型在验证集上检测识别性能如表 6 所示。

C3D 网络模型在本文的数据的验证集上仅取得了 52.23% 的识别准确度。在将 C3D 网络的卷积层加深至 19 层后,得到 C3D\_1 模型,识别准确度仅提升至 64.58%,在实际训练中 C3D\_1 模型花费了大量的训练时间。在 C3D\_1 的基础上,在网络中增加了 BN 层得到 C3D\_2 (BN) 模型,在同样的数据集上,C3D\_2 模型随仅取得了少量的准确度提升,达到 65.63%,但是在实际训练中加入 BN 层加速了网络收敛速度,网络仅运行了 5 个 epoch 后 loss 函数便趋近于收敛。3D ConvNet 模型与 C3D 模型相比,在验证集上的准确度上提升了 43.47 个百分点,与

同样主干网络结构的 C3D\_2 模型以及 C3D\_1 模型相比,3D ConvNet 在引入了多尺度特征融合过后,识别准确度提高至 95.70%,且实际训练模型至收敛的时间大大缩短。

表 6 不同模型的性能比较

Table 6 Comparison of performances of different recognition networks

模型 Model	准确率 Accuracy/%	模型待训练 参数数量 Model training parameter size/ $10^3$	单帧图像平均 识别时间 Single frame average recognition time/s
C3D	52.23	78 003	2.3
C3D_1 (19 层)	64.58	116 603	3.1
C3D_2 (BN)	65.63	116 616	3.0
3D ConvNet	95.70	1 741	0.5

对照各算法的单帧图像平均识别时间,3D ConvNet 在 C3D 网络的基础上同时加深了网络的宽度和深度后,由于 Dropout 以及 Batch Normalization 层的加入,使网络模型待训练参数数量减小至  $1\,741 \times 10^3$ ,单帧图像平均识别时间是 0.5 s,为所有对比网络中最短,且比次短时间 C3D 模型的 2.3 s 减少了 1.8 s,大大提升了识别效率。

针对本研究的猪只侵略性行为识别的目的,在集约化养殖猪场内,要对群养猪的侵略性行为做到实时监测,模型大小和单帧图像的平均识别时间极其重要,模型过大不易加载和运行,且在移动端占用内存严重,检测识别时间过长,无法实现实时检测。本文提出的网络模型训练过后的大小仅为 76.3 MB,在移动端上的移植不会受到限制。且在 CPU 端 (Intel (R) Core (TM) i5-7500) 的单帧图像检测时间为 0.50 s,基本满足集约化群养猪侵略性行为实时检测的要求。

## 3 结 论

本文基于深度学习研究了对群养猪侵略性行为进行识别的网络模型,该网络模型在识别测试集的 1 066 个视频片段中取得较好的效果,具体结论如下:

1) 基于 C3D 网络提出了一种用于对群养猪侵略性行为进行识别的 3D ConvNet 网络模型,网络结构在宽度和深度两个方面进行改进。在深度上加深卷积层的数量,并添加 BN 层和 Dropout 层;在宽度上,在网络模型中设置多尺度特征融合,实现了对侵略性行为发生与否的判断以及置信度的输出。

2) 3D ConvNet 网络模型在测试集上取得了 92.03% 的准确率,在测试集的 1 066 个片段中,总共 981 段(侵略性+非侵略性)视频被正确分类。侵略性行为的查准率指标为 94.86%,侵略性行为召回率指标为 89.57%,调和 Recall 与 Precision 的 F1 值为 92.14%。且在相同环境的不同批次猪只的测试集上以及在照明不良条件下表现出良好的泛化能力。

3) 与 C3D 网络,C3D\_1 (19 层)网络和 C3D\_2 (BN)网络相对比,在相同训练集和验证集的条件下,3D ConvNet 在验证集上的识别准确率超过 C3D 网络,C3D\_1 (19 层)和 C3D\_2 (BN)网络,达到 95.70%,

在处理单帧图像识别速度最快仅需 0.5 s。网络模型在高准确度的同时提高了图像检测时间, 具有良好的有效性和实时性。

结果说明基于 3D 卷积核的群养猪侵略性行为识别网络的模型是稳定有效的。该算法为群养猪侵略性行为识别提供方法和思路, 为后续针对集约化养殖环境下的猪只行为自动监测识别打下了基础。

### [参 考 文 献]

- [1] Turner S P, Farnworth M J, White I M S, et al. The accumulation of skin lesions and their use as a predictor of individual aggressiveness in pigs[J]. *Applied Animal Behaviour Science*, 2006, 96(3/4): 245—259.
- [2] Kongsted, Grete A. Stress and fear as possible mediators of reproduction problems in group housed sows: a review[J]. *Acta Agriculturae Scandinavica, Section A-Animal Science*, 2004, 54(2): 58—66.
- [3] 朱志谦. 工厂化养猪对猪行为及性能的影响与对策[J]. *畜牧与兽医*, 2007(12): 40—41.
- [4] Verdon M, Hansen C F, Rault J L, et al. Effects of group housing on sow welfare: A review[J]. *Journal of Animal Science*, 2015, 93(5): 1999.
- [5] 施正香, 李保明, 张晓颖, 等. 集约化饲养环境下仔猪行为的研究[J]. *农业工程学报*, 2004, 20(2): 220—225.
- [6] Shi Zhengxiang, Li Baoming, Zhang Xiaoying, et al. Behaviour of weaning piglets under intensive farm environment[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2004, 20(2): 220—225. (in Chinese with English abstract)
- [7] 张振玲, Rachel S E Peden, Simon P Turner, 等. 猪混群攻击行为研究进展[J]. *猪业科学*, 2018, 35(12): 34—37.
- [8] 杨飞云, 曾雅琼, 冯泽, 等. 畜禽养殖环境调控与智能养殖装备技术研究进展[J]. *中国科学院院刊*, 2019, 34(2): 163—173.
- [9] 何东健, 刘冬, 赵凯旋. 精准畜牧业中动物信息智能感知与行为检测研究进展[J]. *农业机械学报*, 2016, 47(5): 231—244.
- [10] He Dongjian, Liu Dong, Zhao Kaixuan. Review of perceiving animal information and behavior in precision livestock farming[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2016, 47(5): 231—244. (in Chinese with English abstract)
- [11] 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.
- [12] Ma Li, Ji Bin, Liu Hongshen, et al. Differentiating profile based on single pig contour[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2013, 29(10): 168—174. (in Chinese with English abstract)
- [13] 张萌, 钟南, 刘莹莹. 基于生猪外形特征图像的瘦肉率估测方法[J]. *农业工程学报*, 2017, 33(12): 308—314.
- [14] Zhang Meng, Zhong Nan, Liu Yingying. Estimation method of pig lean meat percentage based on image of pig shape characteristics[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2017, 33(12): 308—314. (in Chinese with English abstract)
- [15] 刘龙申, 沈明霞, 柏广宇, 等. 基于机器视觉的母猪分娩检测方法研究[J]. *农业机械学报*, 2014, 45(3): 237—242.
- [16] Liu Longshen, Shen Mingxia, Bo Guangyu, et al. Sows parturition detection method based on machine vision[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2014, 45(3): 237—242. (in Chinese with English abstract)
- [17] Oczak M, Viazzi S, Ismayilova G, et al. Classification of aggressive behaviour in pigs by activity index and multilayer feed forward neural network[J]. *Biosystems Engineering*, 2014, 119: 89—97.
- [18] Viazzi S, Ismayilova G, Oczak M, et al. Image feature extraction for classification of aggressive interactions among pigs[J]. *Computers and Electronics in Agriculture*, 2014, 104: 57—62.
- [19] Chen C, Zhu W, Ma C, et al. Image motion feature extraction for recognition of aggressive behaviors among group-housed pigs[J]. *Computers and Electronics in Agriculture*, 2017, 142: 380—387.
- [20] Jonguk L, Long J, Daihee P, et al. Automatic recognition of aggressive behavior in pigs using a kinect depth sensor[J]. *Sensors*, 2016, 16(5): 631—641.
- [21] 孙钰, 周焱, 袁明帅, 等. 基于深度学习的森林害虫无人机实时监测方法[J]. *农业工程学报*, 2018, 34(21): 74—81.
- [22] Sun Yu, Zhou Yan, Yuan Mingshuai, et al. UAV real-time monitoring for forest pest based on deep learning[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2018, 34(21): 74—81. (in Chinese with English abstract)
- [23] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// *International Conference on Neural Information Processing Systems*. Curran Associates Inc. 2012: 1097—1105.
- [24] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39(6): 1137—1149.
- [25] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[C]// *European Conference on Computer Vision*. Springer, Cham, 2014: 818—833.
- [26] Zhang Z, Fidler S, Urtasun R. Instance-Level Segmentation for Autonomous Driving with Deep Densely Connected MRFs[C]// *Computer Vision & Pattern Recognition*. 2016.
- [27] Yang Qiumei, Xiao Deqin, Lin Sicong. Feeding behavior recognition for group-housed pigs with the Faster R-CNN[J]. *Computers and Electronics in Agriculture*, 2018, 144: 453—460.
- [28] Yang Aqing, Huang Huasheng, Zheng Chan. High-accuracy image segmentation for lactating sows using a fully convolutional network[J]. *Biosystems Engineering*, 2018, 176: 36—47.
- [29] 杨秋妹, 肖德琴, 张根兴. 猪只饮水行为机器视觉自动识别[J]. *农业机械学报*, 2018, 49(6): 232—238.
- [30] Yang Qiumei, Xiao Deqin, Zhang Genxin. Automatic pig drinking behavior recognition with machine vision[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2018, 49(6): 232—238. (in Chinese with English abstract)
- [31] Zheng Chan, Zhu Xunmu, Yang Xiaofan. Automatic recognition of lactating sow postures from depth images by deep learning detector[J]. *Computers and Electronics in Agriculture*, 2018, 147: 51—63.
- [32] 高云, 陈震撼, 王瑜, 等. 多环境参数控制的猪养殖箱设计及箱内气流场分析[J]. *农业工程学报*, 2019, 35(2): 203—212.
- [33] Gao Yun, Chen Zhenhan, Wang Yu, et al. Design for pig breeding chamber under multiple environment variable control and analysis of internal flow field[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2019, 35(2): 203—212. (in Chinese with English abstract)
- [34] 高云, 郭继亮, 黎焯, 等. 基于深度学习的群猪图像实例分割方法[J]. *农业机械学报*, 2019, 50(4): 179—187.
- [35] Gao Yun, Guo Jiliang, Li Xuan, et al. Instance-level



- segmentation method for group pig images based on deep learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(4): 179–187. (in Chinese with English abstract)
- [27] Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning[M]. 北京: 人民邮电出版社, 2016.
- [28] Tran D, Bourdev L, Fergus R, et al. Learning spatiotemporal features with 3D convolutional networks[C]//2015 IEEE International on Computer Vision and Pattern Recognition, 2015: 4694–4702.
- [29] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929–1958.
- [30] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]// International Conference on International Conference on Machine Learning. JMLR.org, 2015.
- [31] Zhang Z, Zhang X, Peng C, et al. ExFuse: Enhancing feature fusion for semantic segmentation[C]// European Conference on Computer Vision. Springer, Cham, 2018.
- [32] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot MultiBox detector[C]// European Conference on Computer Vision. 2016.
- [33] Joseph Redmon, Farhadi Ali. YOLOv3: An incremental improvement[EB/OL].[2018-04-08].https://arxiv.org/pdf/1804.02767.pdf.

## Recognition method for aggressive behavior of group pigs based on deep learning

Gao Yun<sup>1,2</sup>, Chen Bin<sup>1</sup>, Liao Huimin<sup>1</sup>, Lei Minggang<sup>2,3</sup>, Li Xuan<sup>1,2</sup>, Li Jing<sup>1</sup>, Luo Junjie<sup>1</sup>

(1. College of Engineering, Huazhong Agricultural University, Wuhan 430070, China; 2. Cooperative Innovation Center for Sustainable Pig Production, Wuhan 430070, China; 3. College of Animal Science and Technology, College of Animal Medicine, Huazhong Agricultural University, Wuhan 430070, China)

**Abstract:** Pigs like to fight with each other to form a hierarchy relationship in groups. Aggressive behaviors, mostly fighting, are frequently found in intensive pig raising facilities. Strong aggressive behaviors can cause other pigs lack of food and water, growing slowly, wounds, sick and even dead in serious situation. This considerably reduces health and welfare of pigs and further decreases economic benefits of pig industries. Monitoring and recognizing aggressive behaviors among pig group is the first step to manage the aggressive behaviors in group pigs effectively. Traditional human recording method is time-consuming and labor-intensive. This method can't be used 24 hours a day, 7 days a week. Machine vision technique brings an automatic monitoring method to solve this problem. In this paper, we introduced a new method for aggressive behaviors monitoring based on deep learning. The experiments were held under controlled environments, which were achieved in an environment-controlled chamber designed previously. The details of the chamber were depicted in a published paper written by our research group. Nursery pigs were fed under three different concentration levels of NH<sub>3</sub> gas, which were <3.80 mg/m<sup>3</sup>, 15.18 mg/m<sup>3</sup>, 37.95 mg/m<sup>3</sup>, with a suitable temperature of around 27 °C and the comfortable humidity between 50%-70%. Each nursery group had six pigs and were weight around 9.6 kg. During each 28 days' experiment of three concentration levels of NH<sub>3</sub>, videos were taken from the top of the chamber. An end-to-end network, named 3D CONVNet, was proposed for aggressive behavior recognition of group pigs in this paper, which based on a C3D network and built with 3D convolution kernels. The network structure of the 3D CONVNet was improved in both width and depth dimensions. The number of main convolutional layers was increased to 19, extra batch normalization and dropout layers were added to deepen the network. Furthermore, the multi-scale feature fusion method was introduced to widen the network. This improvement had bettered the performance of the algorithm considerably. To train the 3D CONVNet, 380 aggressive (14 074 frames) and 360 none-aggressive videos (13 040 frames) were chosen from experimental videos recording in experiments of two concentration levels. These videos were randomly divided into training set and validation set, and the ratio of each set is 3:1. Another 556 aggressive videos and 510 none-aggressive videos from the three experimental batches were chosen to build the testing set. There was no overlap among training set, validation set, and testing set. Results showed a total of 981 videos, including aggressive and non-aggressive behaviors, was correctly recognized from the whole 1066 testing videos. The precision of the 3D CONVNet was proved to be 92.03% on testing set. Among them, the precision, recall rate and F1-Score for aggressive behaviors were 94.86%, 89.57%, and 92.14%, respectively. The precision for different NH<sub>3</sub> concentration experimental levels were 94.29%, 89.44%, and 85.91%, respectively, which showed the generalization performance of the 3D CONVNet. With the similar heat environments, the 3D CONVNet also showed the good performances under different illumination condition. The comparison with C3D, C3D\_1 (19 layers) and C3D\_2 (BN) networks resulted in 95.7% on validation set, 43.27 percent higher than the C3D network. The recognition on single image using the 3D CONVNet was only 0.5 s, which was much faster than the other three networks. Therefore, the 3D CONVNet was effective and robust in aggressive behavior recognition among group pigs. The algorithm provides a new method and technique for aggressive behavior auto-monitoring of group pigs and helps improve establishment of auto-monitoring system in pig farms and manage level of pig industry.

**Keywords:** convolutional neural network; machine vision; models; behavior recognition; aggressive behavior; deep learning; group pigs