

基于深度学习数据增强技术的肉牛行为识别 与养殖状态评估研究

李慢¹, 郜静茹, 赵永才, 代新雨, 刘雯秀, 简蕊, 关曼玉, 李广博

(淮北理工学院电子与信息工程学院, 安徽 淮北 235000)

摘 要: 肉牛行为的准确识别是智慧养殖中个体状态监测和精细化管理的重要基础。针对实际养殖场景中光照变化、行为姿态差异以及训练样本规模对识别性能影响不明确等问题, 本文以9类肉牛典型行为图像为研究对象, 构建基于深度学习骨干网络与组合数据增强的行为识别流程。数据集包含饮水、进食、争斗、舔舐、躺卧、爬跨、探究、站立和行走9类行为, 覆盖低光、正常光照、高光和黑暗4种光照条件, 共4500张图像。首先, 在统一训练条件下比较ResNet-50、DenseNet-121、EfficientNet-B0和Swin-T四种骨干网络。实验结果表明, EfficientNet-B0的测试准确率为 $98.37\% \pm 1.36\%$, 参数量为4.02 M, MACs为0.41 G; 其准确率较ResNet-50高0.07个百分点, 较DenseNet-121和Swin-T均高0.59个百分点, 同时具有四种网络中最低的参数量与计算量, 因此选其作为后续实验的固定骨干。进一步设置不同训练样本规模实验与数据增强消融实验, 以分析样本量和翻转、颜色抖动、旋转、随机擦除等增强操作的影响。实验二结果表明, 当每类训练样本由每类10张增至全量训练集(每类400张)时, 测试准确率由 $69.70\% \pm 0.76\%$ 提升至 $98.37\% \pm 1.36\%$, 训练样本规模对识别性能影响显著, 且随样本量增大边际收益总体呈递减趋势。数据增强消融实验表明, 逐项去除翻转、颜色抖动、旋转和随机擦除后, 测试准确率较完整增强基线($98.37\% \pm 1.36\%$)分别下降2.74、3.11、2.89和2.96个百分点, 验证了各增强操作对复杂光照与姿态条件下识别性能的贡献, 其中颜色抖动的作用最为显著。研究可为复杂光照条件下肉牛行为的自动识别与智能养殖监测提供参考。

关键词: 肉牛行为识别; EfficientNet-B0; 深度学习; 数据增强; 复杂光照

基金项目: 安徽省教育厅优秀青年教师培育项目(YQYB2026117); 安徽省智慧农业技术与装备重点实验室开放基金(AEC2026016; AEC2025009); 安徽省教育厅自然科学基金项目(重大)(2024AH040217); 省级质量工程-教学研究项目(2023jyxm1009; 2023xjzts117; 2023sdxx145); 基于人工智能下的家禽康养与疾病预警技术研究(S202413620055); 基于自监督学习与对抗增强的小样本信号的智能感知研究(2025SK019); 基于深度学习的生猪识别与康养研究(S202513620083)

DOI: doi.org/10.65436/hssj.v1i9.48

¹**作者简介:** 李慢(2005—), 女, 安徽池州人, 本科生。研究方向: 人工智能;

郜静茹(2005—), 女, 安徽阜阳人, 本科生。研究方向: 人工智能;

赵永才(2007—), 男, 安徽蚌埠人, 本科生。研究方向: 人工智能;

代新雨(2005—), 女, 安徽阜阳人, 本科生, 研究方向: 人工智能;

刘雯秀(2005—), 女, 安徽滁州人, 本科生。研究方向: 人工智能;

简蕊(2004—), 女, 安徽蚌埠人, 本科生。研究方向: 人工智能;

关曼玉(1998—), 女, 安徽淮北, 硕士, 助教。研究方向: 人工智能与大数据;

李广博(1997—), 男, 安徽淮北人, 讲师, 博士。研究方向: 人工智能、智能信号感知、模式识别、计算机视觉等。

通讯作者: 李广博

随着规模化和智能化养殖的发展,对肉牛个体行为进行持续、准确监测已成为智慧畜牧的重要研究内容。肉牛的站立、躺卧、行走、进食、饮水、争斗等行为能够反映其活动状态和养殖环境适应情况,对养殖管理、异常行为发现及生产决策具有参考价值。传统人工观察或视频回放方式工作量大、主观性较强,难以满足大规模养殖场景下连续、自动化监测的需求。近年来, Li 等构建了 CBVD-5 奶牛行为视频数据集,数据覆盖不同光照及夜间条件,为真实牧场环境下的行为识别研究提供了公开数据基础^[1]; Bai 等面向复杂养殖场景提出 X3DFast 双路径行为识别模型,并在包含遮挡、模糊、视角与光照变化的视频数据上验证了模型的有效性^[2]。上述研究表明,基于计算机视觉的非接触式行为识别具有较好的应用潜力,但复杂背景与环境变化仍会影响模型的稳定性和部署效率。

深度学习能够利用多层网络结构自动提取图像中的纹理、轮廓和高层语义特征,已成为动物行为图像识别的重要技术路线。ResNet 通过残差连接改善深层网络训练^[3], DenseNet 通过密集连接增强特征传播与复用^[4], EfficientNet 采用复合缩放思想平衡网络深度、宽度与输入分辨率^[5], Swin Transformer 则通过层次化结构与移位窗口自注意力实现视觉表征^[6]。这些典型视觉骨干在特征提取方式、参数规模和计算复杂度方面存在明显差异,因此有必要在统一实验条件下比较其对肉牛行为图像的适用性。同时,实际养殖环境中的复杂光照、姿态变化以及有限样本也可能影响模型的泛化性能。

基于上述问题,本文从骨干网络筛选、训练样本规模和数据增强三个方面开展肉牛行为识别实验。首先,在相同数据划分、训练参数和增强策略下,对 ResNet-50、DenseNet-121、EfficientNet-B0 和 Swin-T 进行对比;其次,以综合性能较优的 EfficientNet-B0 为固定骨干,设置不同训练样本规模以绘制学习曲线;最后,通过逐项移除翻转、颜色抖动、旋转和随机擦除操作开展数据增强消融实验,并结合混淆矩阵、t-SNE 和训练曲线对结果进行分析。

本文主要工作包括:(1)构建包含 9 类肉牛行为和 4 种光照条件的图像数据集,并采用分层抽样方式划分训练集、验证集和测试集;(2)完成四种典型骨干网络的统一对比,依据识别准确率、参数量和 MACs 确定后续实验骨干;(3)设计不同训练样本量实验和数据增强消融实验,以分析样本规模及各增强操作对肉牛行为识别性能的影响;(4)通过不同训练样本规模实验,揭示了样本量对肉牛行为识别性能的影响规律:测试准确率随样本量增大而整体提升,少样本阶段提升显著、后期边际收益递减,为实际场景中训练样本采集与标注成本的控制提供了依据;同时,数据增强消融实验表明四种增强操作均有显著贡献,去除任一操作准确率下降 2.74~3.11 个百分点,其中颜色抖动贡献最大,验证了组合增强策略对复杂光照场景的有效性。

一、研究方法

(一) 肉牛行为识别流程

本文识别流程由图像输入、数据预处理与增强、骨干网络特征提取、全连接分类和损失计算组成。输入图像经预处理后,在训练阶段按照设定的数据增强策略进行随机变换,再送入骨干网络提取特征;骨干输出的高维表示经 9 类全连接层映射得到各行类别的预测分数,训练过程中采用交叉熵损失函数进行参数优化。测试阶段关闭随机增强,仅执行必要的确定性预处理,并在固定测试集上计算识别准确率。

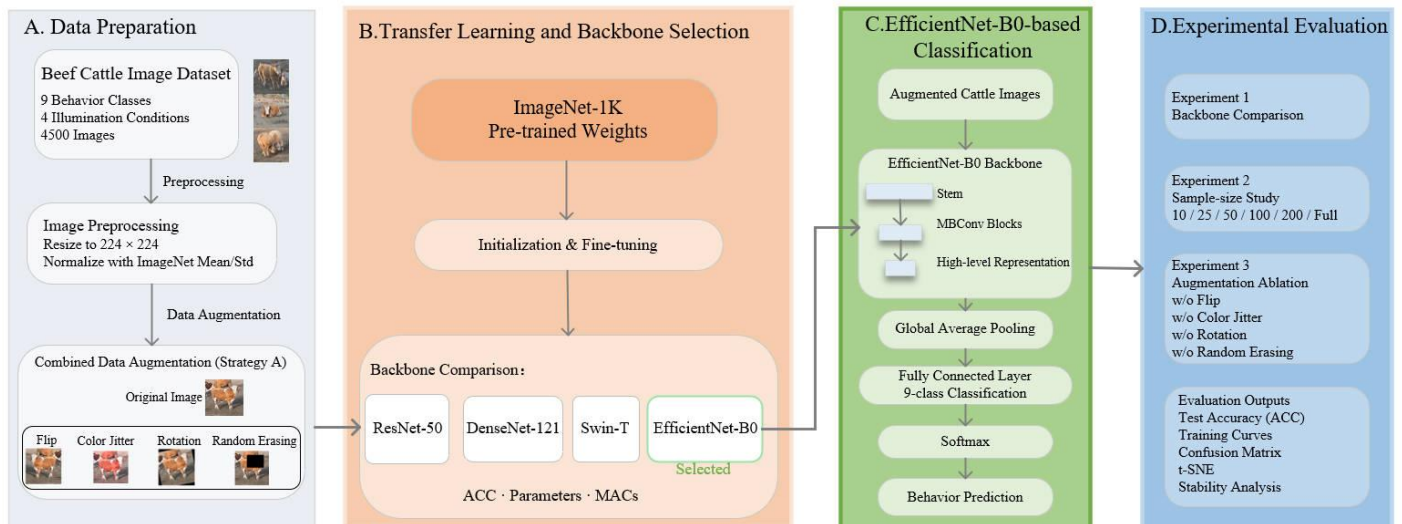


图 1 基于 EfficientNet-B0 与组合数据增强的肉牛行为识别框架

(二) 骨干网络与分类器

实验选取 ResNet-50、DenseNet-121、EfficientNet-B0 和 Swin-T 四种典型视觉骨干网络进行比较。其中，EfficientNet 采用复合缩放方法统一调整网络深度、宽度与输入分辨率，其缩放关系可表示为：

$$d = \alpha^\phi, w = \beta^\phi, r = \gamma^\phi, \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \quad (1)$$

式中， d 、 w 和 r 分别表示网络深度、宽度和输入分辨率； α 、 β 、 γ 为网络缩放系数，满足 $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$ ； ϕ 为复合缩放系数。

四种网络均采用 ImageNet-1K 预训练权重进行初始化，并将原有分类层替换为输出维度为 9 的全连接层。实验一仅改变骨干网络类型，其余数据划分、训练策略和增强组合保持一致，以保证对比的公平性。根据实验一结果，EfficientNet-B0 在测试准确率上取得最高均值，且参数量和 MACs 最低，因此后续样本量实验和数据增强消融实验固定采用 EfficientNet-B0。

(三) 组合数据增强策略

为增强模型对不同姿态、拍摄方向、光照变化和局部遮挡的适应能力，训练阶段采用由随机翻转、颜色抖动、随机旋转和随机擦除构成的策略 A。数据增强能够在不额外增加人工标注成本的情况下扩展训练分布，是缓解过拟合和提升视觉模型泛化能力的常用方法^[7]。其中，随机翻转用于模拟方向变化；颜色抖动用于改变亮度、对比度等图像属性；随机旋转用于扩展姿态角度分布；随机擦除通过遮挡局部区域降低模型对单一局部纹理的依赖。各增强操作的具体概率和参数范围为：随机水平翻转概率为 0.5；颜色抖动亮度、对比度与饱和度调整范围 ± 0.2 ，色相调整范围 ± 0.1 ；随机旋转角度范围 $\pm 15^\circ$ ；随机擦除概率为 0.25，擦除面积占图像 2%~20%。所有图像统一缩放至 224×224 ，并采用 ImageNet 均值与标准差进行归一化。实验三将在完整增强策略基础上逐项移除一种增强操作，以分析各操作的贡献。

(四) 损失函数与评价指标

模型训练采用 CrossEntropyLoss。设样本真实类别为 y ，模型对第 c 类的预测概率为 p_c ，则交叉熵损失可表示为：

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \ln(p_{i,c}) \quad (2)$$

式中， C 为类别数， y_c 为真实类别的 one-hot 编码， p_c 为模型对第 c 类的预测概率。

本文以测试集准确率 (Accuracy, ACC) 作为主要识别指标，并同时报告模型参数量 (M) 和乘加运算量 (MACs, G)，用于综合评估不同骨干网络的识别性能与计算复杂度。每个骨干网络分别使用随机种子 42、123 和 2024 进行 3 次独立实验，最终报告测试准确率的均值 $\pm$ 标准差。

$$ACC = \frac{1}{N} \sum_{i=1}^N I(\hat{y}_i = y_i) \quad (3)$$

式中， $N_{correct}$ 为预测正确的样本数， N_{total} 为测试集样本总数。

二、实验设计与结果分析

(一) 数据集与实验设置

本文使用肉牛行为图像数据集开展实验，共包含 9 类行为：饮水 (drinking)、进食 (eating)、争斗 (fighting)、舔舐 (licking)、躺卧 (lying)、爬跨 (mounting)、探究 (searching)、站立 (standing) 和行走 (walking)。数据覆盖低光、正常光照、高光和黑暗 4 种光照条件，每类随机选取 500 张图像，共 4500 张。采用按类别分层抽样 (stratified split) 方式按照 8:1:1 划分训练集、验证集和测试集，对应样本量分别为 3600、450 和 450 张。

原始视频素材由课题组于肉牛养殖牧场通过固定监控摄像头采集，并辅以部分手持拍摄；其中行为图像帧由作者本人对监控视频进行逐帧截取、筛选与整理，剔除模糊、重复及多目标遮挡等无效帧后，得到本文所用图像数据集。

表 1 数据集及划分设置

项目	设置
类别数	9 类
光照条件	低光、正常光照、高光、黑暗
每类样本	500 张
总样本	4500 张
划分方式	按类分层抽样（8：1：1）
训练集	3600 张
验证集	450 张
测试集	450 张

表 2 主要训练参数

项目	设置
预训练权重	ImageNet-1K
分类头	9 类全连接层
优化器	AdamW
初始学习率	1.00×10^{-4}
学习率调度	CosineAnnealingLR
Batch size	32
损失函数	CrossEntropyLoss
训练轮数	100 epochs
随机种子	42、123、2024
主要评价指标	测试集准确率（ACC）

实验在 Windows 操作系统下完成，Python 环境为 Anaconda（yolov5gpu），PyTorch 版本为 2.6.0+cu126，torchvision 版本为 0.21.0+cu126，CUDA 版本为 12.6，GPU 为 NVIDIA GeForce RTX 4060 Laptop GPU。模型训练采用 AdamW 优化器，初始学习率设为 1.00×10^{-4} ，学习率调度器为 CosineAnnealingLR，Batch size 为 32，训练 100 epochs。其他软件版本包括 timm 1.0.29、scikit-learn 1.6.1、matplotlib 3.9.4、pandas 2.2.3。

（二）实验一：不同骨干网络对比

实验一以骨干网络为唯一变量，对 ResNet-50、DenseNet-121、EfficientNet-B0 和 Swin-T 进行比较。四种网络均使用策略 A（翻转+颜色抖动+旋转+随机擦除）进行训练，数据集、训练轮数和随机种子保持一致。表 3 给出了三个随机种子下的测试准确率、均值±标准差以及模型复杂度。

表 3 不同骨干网络识别性能对比

骨干网络	Seed 42	Seed 123	Seed 2024	均值±标准差	参数量/M	MACs/G
EfficientNet-B0	100.00%	96.67%	98.44%	98.37 ± 1.36%	4.02	0.41
ResNet-50	99.56%	97.56%	97.78%	98.30 ± 0.90%	23.53	4.13

骨干网络	Seed 42	Seed 123	Seed 2024	均值±标准差	参数量/M	MACs/G
DenseNet-121	99.56%	96.00%	97.78%	97.78 ± 1.45%	6.96	2.90
Swin-T	100.00%	95.78%	97.56%	97.78 ± 1.73%	27.53	3.13

由表 3 可知, EfficientNet-B0 的平均测试准确率为 98.37%±1.36%, 在四种骨干网络中数值最高, 较 ResNet-50 高 0.07 个百分点, 较 DenseNet-121 和 Swin-T 均高 0.59 个百分点。与此同时, EfficientNet-B0 的参数量仅为 4.02 M, MACs 为 0.41 G, 均为四种网络中的最低值。与 ResNet-50 相比, 其参数量减少约 82.92%, MACs 减少约 90.07%; 与 DenseNet-121 相比, 参数量减少约 42.24%, MACs 减少约 85.86%; 与 Swin-T 相比, 参数量减少约 85.40%, MACs 减少约 86.90%。从波动性看, ResNet-50 的标准差最低 (0.90%), EfficientNet-B0 的标准差为 1.36%, 低于 DenseNet-121 和 Swin-T。综合识别精度和模型复杂度, 本文选择 EfficientNet-B0 作为实验二和实验三的稳定骨干网络。

图 4、图 5 进一步从训练过程、误分类结构与特征分布三个角度对四种骨干网络进行可视化分析 (图 5 以随机种子 42 为例)。图 4 给出了四种骨干网络的验证准确率与验证损失曲线, 四种网络均在前 20 个 epoch 内快速收敛, 验证准确率随后趋于稳定, 其中 EfficientNet-B0 与 ResNet-50 的收敛速度和最终验证准确率整体接近, DenseNet-121 与 Swin-T 的验证损失波动相对略大, 与表 3 中各网络标准差的结果基本一致。图 5 (e) ~ (h) 为四种骨干网络在测试集上的混淆矩阵 (按行归一化), 各网络主对角线数值均接近 1, 表明 9 类行为均能得到有效识别; 其中 EfficientNet-B0 与 Swin-T 在 seed 42 下实现完全正确分类, ResNet-50 和 DenseNet-121 仅存在极少量误分类, 且误分类主要发生在探究、站立、行走等姿态相似的行为之间, 如 ResNet-50 将 2% 的探究样本分别误分为站立和行走, DenseNet-121 将 2% 的舔舐样本误分为探究、并将 2% 的探究样本误分为行走。图 5 (a) ~ (d) 的 t-SNE 特征分布显示, 四种骨干网络提取的特征在二维空间中均形成与行为类别对应的独立聚类簇, 同类样本聚集紧密、异类样本分离清晰, 其中 EfficientNet-B0 的类间边界最为分明, 从特征分布角度进一步印证了其最优的识别性能。

(三) 实验二: 不同训练样本规模实验

实验二固定骨干网络为 EfficientNet-B0, 并保持策略 A 不变, 通过改变每类训练样本数量分析模型的学习曲线。当前实验计划设置 10-shot、25-shot、50-shot、100-shot、200-shot 以及全量基线。验证集和测试集保持固定不变, 各样本量均使用随机种子 42、123 和 2024 进行独立实验。其中全量基线采用与实验一一致的完整训练集 (共 3600 张, 即每类 400 张), 其结果与实验一 EfficientNet-B0 全量设置保持一致, 作为性能上限对照。

表 4 不同训练样本规模实验结果

训练样本规模	训练样本总量	测试 ACC (均值±标准差)
10-shot	90	69.70 ± 0.76%
25-shot	225	80.44 ± 1.37%
50-shot	450	85.93 ± 0.52%
100-shot	900	90.74 ± 0.86%
200-shot	1800	93.19 ± 1.21%
全量基线	3600	98.37 ± 1.36%

由表 4, 图 2 可知, 随着训练样本规模的增大, 模型测试准确率呈明显上升趋势。当每类训练样本仅为 10 张时, 测试准确率为 69.70%±0.76%, 较全量基线低 28.67 个百分点; 当样本量增至每类 25 张时, 准确率提升至 80.44%±1.37%, 为各阶段中提升幅度最大 (较 10-shot 提升 10.74 个百分点); 50-shot、100-shot 和 200-shot 对应的准确率分别为 85.93%±0.52%、90.74%±0.86%和 93.19%±1.21%, 较全量基线分别低 12.44、7.63 和 5.18 个百分点。总体来看, 准确率在样本量较小时随样本量增加快速提升, 100-shot 之后增幅明显收窄 (100-shot 至 200-shot 仅提升 2.45 个百分点), 而由 200-shot 增至全量时仍提升 5.18 个百分点, 表明在该数据集上性能尚未完全饱和,

但边际收益总体呈递减趋势。从波动性看,各样本量下 3 个随机种子的标准差介于 0.52%~1.37%之间,其中 50-shot 时波动最小(0.52%), 25-shot 时波动最大(1.37%),说明样本量较少时模型性能对随机初始化仍较为敏感。上述结果表明,当标注成本受限时,每类 100~200 张训练样本即可获得接近全量的识别性能,可为实际养殖场景中的数据采集规模与标注策略提供参考。

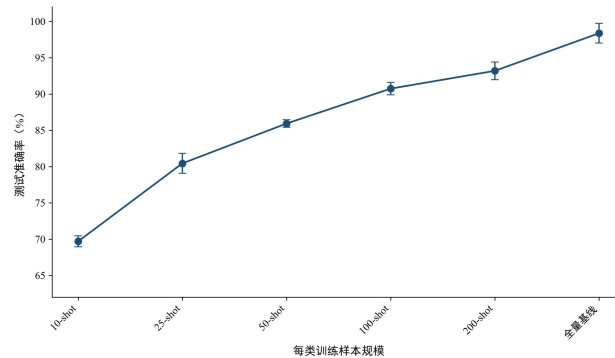


图2 不同训练样本规模下的学习曲线

(四) 实验三：数据增强消融实验

实验三固定 EfficientNet-B0 与全量训练设置,以“翻转+颜色抖动+旋转+随机擦除”为完整增强基线,在此基础上依次去除翻转、颜色抖动、旋转和随机擦除,比较各增强组合的测试准确率变化。除增强组合外,其余训练参数保持一致,每种新增组合使用 3 个随机种子重复训练。

表5 数据增强消融实验结果

设置	翻转	颜色抖动	旋转	随机擦除	测试 ACC
完整增强	√	√	√	√	98.37 ± 1.36%
去翻转	×	√	√	√	95.63 ± 0.55%
去颜色抖动	√	×	√	√	95.26 ± 0.58%
去旋转	√	√	×	√	95.48 ± 0.28%
去随机擦除	√	√	√	×	95.41 ± 0.46%

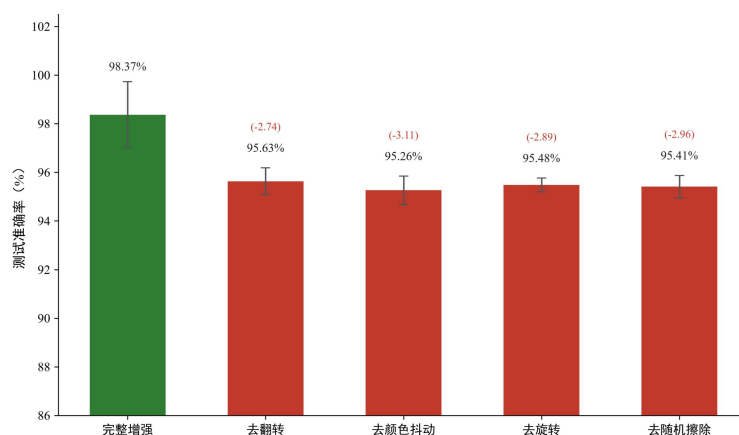


图3 不同增强组合的测试准确率对比及各增强操作贡献度（绿色为完整增强基线，红色为留一设置，柱上数值为均值准确率，红色数值为较基线下降幅度，误差棒为 3 次重复的标准差）

由表 5 可知,完整增强基线的测试准确率为 98.37%±1.36%;依次去除翻转、颜色抖动、旋转和随机擦除后,准确率分别降至 95.63%±0.55%、95.26%±0.58%、95.48%±0.28%和 95.41%±0.46%,较基线分别下降 2.74、3.11、

2.89 和 2.96 个百分点，各去除设置的标准差均小于基线。四种增强操作对识别性能均具有正向贡献，其中去除颜色抖动带来的性能下降最大，表明颜色抖动通过亮度、对比度与饱和度扰动，对低光、高光及黑暗等光照差异的适应作用最为关键；去除翻转的下降幅度相对最小，但仍达到 2.74 个百分点，说明水平翻转提供的姿态方向多样性同样不可或缺。上述结果验证了由随机翻转、颜色抖动、随机旋转和随机擦除构成的组合增强策略能够有效提升模型在复杂光照与姿态条件下的泛化能力。

为进一步直观展示各增强操作的贡献差异，图 3 给出了完整增强基线与四种留一设置的测试准确率对比。由图 3 可见，完整增强基线（绿色柱）以 98.37% 显著高于各留一设置（红色柱），其中去除颜色抖动后的准确率最低（95.26%），对应 3.11 个百分点的最大降幅；去除翻转后的降幅相对最小（2.74 个百分点），但仍表明水平翻转对姿态方向多样性具有不可替代的作用。各留一设置的误差棒均明显小于基线，反映出去除某一增强后模型在不同随机种子下的表现趋于一致，也从稳定性角度印证了组合增强策略通过引入多样化变换提升了模型的泛化上限。

(五) 可视化与稳定性分析

本节结合训练/验证 Accuracy 与 Loss 曲线、混淆矩阵和 t-SNE 特征分布对模型进行可视化与稳定性分析。混淆矩阵用于观察 9 类行为之间的具体误分类关系；t-SNE 用于将高维特征映射到二维空间，分析不同类别的类内聚合与类间分离情况；稳定性分析基于不同随机种子的重复实验结果进行。由图 4、图 5 可见，四种骨干网络在训练过程中均能稳定收敛，误分类主要发生在探究、站立、行走等姿态相近的行为之间，各类别特征在二维空间中类内聚合紧密、类间分离清晰；各网络 3 次独立实验的测试准确率标准差介于 0.90%~1.73% 之间，其中 EfficientNet-B0 为 1.36%，整体波动可控，表明模型在固定实验设置下具有较好的重复性与稳定性。

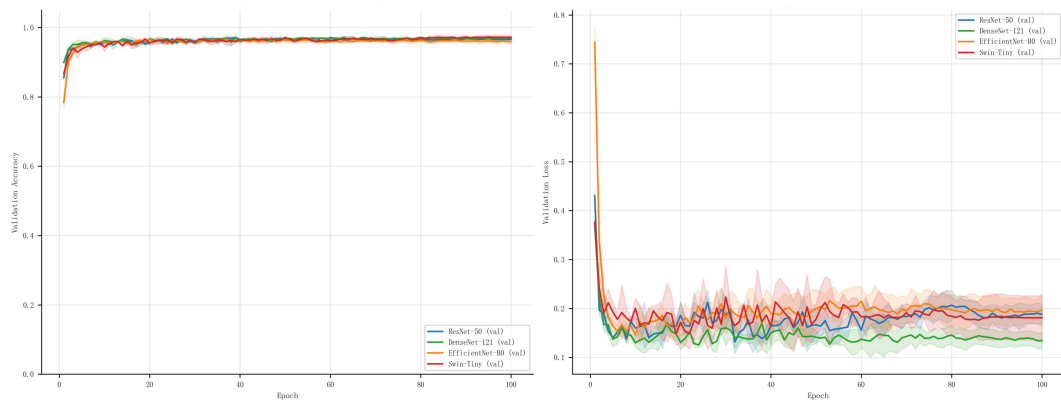


图 4 四种骨干网络训练过程曲线（验证准确率与验证损失）

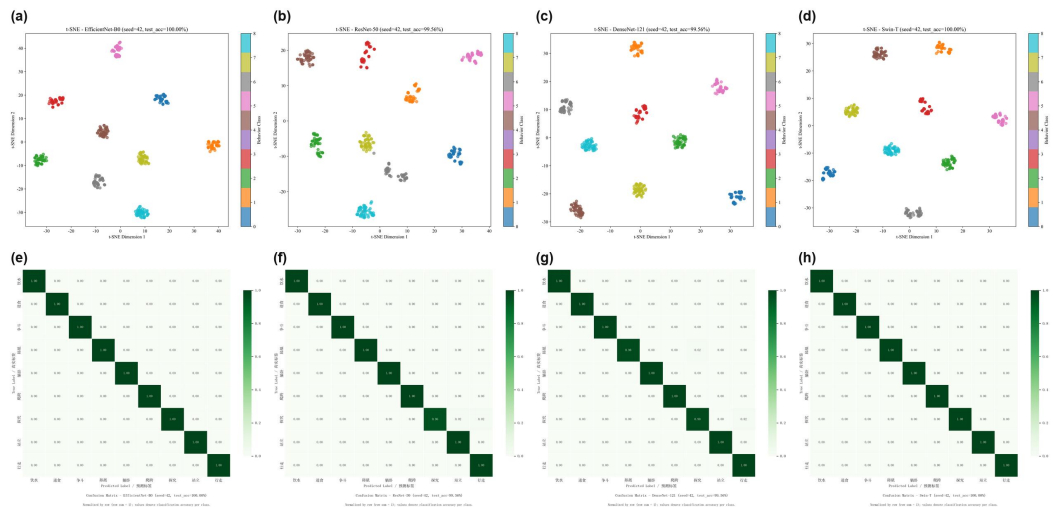


图 5 四种骨干网络测试集特征可视化 (seed 42): (a) ~ (d) t-SNE 特征分布, (e) ~ (h) 混淆矩阵

三、结论

本文针对复杂光照条件下肉牛行为识别问题，构建了包含 9 类行为、4 种光照条件的图像数据集，并提出基于 EfficientNet-B0 与组合数据增强的识别方案。在统一实验设置下，EfficientNet-B0 以 $98.37\% \pm 1.36\%$ 的平均测试准确率取得最优识别性能，同时仅需 4.02 M 参数量和 0.41 G MACs，在识别精度与计算复杂度之间实现良好平衡；训练样本规模实验表明，测试准确率随样本量增大整体提升，少样本阶段提升显著而后期边际收益递减；数据增强消融实验验证了翻转、颜色抖动、旋转和随机擦除四种操作均有显著贡献，其中颜色抖动作用最大；可视化分析进一步证实各类行为特征类内聚合紧密、类间分离清晰，模型具有良好稳定性。

当前研究主要基于单帧图像进行行为分类，尚未建模连续视频中的时序关联与个体跟踪信息；且数据采集自单一养殖场，跨场景、跨品种的泛化能力有待进一步验证。未来工作将引入时序建模、目标检测与个体跟踪方法以提升复杂养殖场景下的连续监测能力，并探索轻量化结构与量化部署，支撑移动端与边缘设备的实时应用。

参考文献：

- [1] LI K, FAN D, WU H, et al. A new dataset for video-based cow behavior recognition[J]. Scientific Reports, 2024, 14:18702.
- [2] BAI Q, GAO R, WANG R, et al. X3DFast model for classifying dairy cow behaviors based on a two-pathway architecture[J]. Scientific Reports, 2023, 13:20519.
- [3] 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.
- [4] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017:4700-4708.
- [5] TAN M, LE Q V. EfficientNet: Rethinking model scaling for convolutional neural networks[C]//Proceedings of the 36th International Conference on Machine Learning. PMLR, 2019, 97:6105-6114.
- [6] LIU Z, LIN Y, CAO Y, et al. Swin Transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021:10012-10022.
- [7] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6:60.

Research on Cattle Behavior Recognition and Farming Status Evaluation Based on Deep Learning Data Augmentation Technology

LI Man, GAO Jingru, ZHAO Yongcai, DAI Xinyu, LIU Wenxiu, JIAN Rui, GUAN Manyu, LI Guangbo

(College of Electronic and Information Engineering, Huaibei Institute of Technology, Huaibei 235000, China)

Abstract: Accurate recognition of beef cattle behavior is an important basis for individual status monitoring and refined management in intelligent livestock farming. This study investigates nine typical cattle behaviors under four illumination conditions and develops a recognition pipeline based on deep visual backbones and data augmentation. The dataset contains 4,500 images of drinking, eating, fighting, licking, lying, mounting, searching, standing, and walking behaviors. Under identical training settings, ResNet-50, DenseNet-121, EfficientNet-B0, and Swin-T were compared. EfficientNet-B0 achieved an average test accuracy of $98.37\% \pm 1.36\%$ with 4.02 M parameters and 0.41 G MACs, providing the best numerical accuracy and the lowest model complexity among the four backbones. Therefore, EfficientNet-B0 was selected as the fixed backbone for subsequent experiments. Further experiments are designed to investigate different training sample sizes and the contributions of flipping, color jitter, rotation, and random erasing. In Experiment 2, the test accuracy rose from $69.70\% \pm 0.76\%$ with 10 training samples per class to $98.37\% \pm 1.36\%$ with the full training set, showing that the training sample size significantly affects recognition performance, with the largest gains in the low-sample regime and diminishing marginal returns as the sample size grows. In Experiment 3, removing flipping, color jitter, rotation, or random erasing from the full augmentation baseline decreased the test accuracy by 2.74, 3.11, 2.89, and 2.96 percentage points, respectively, confirming that all four operations contribute to recognition under complex illumination and posture variations, with color jitter making the most significant contribution. The study provides a reference for automatic cattle behavior recognition under complex illumination conditions.

Keywords: Beef cattle behavior recognition; EfficientNet-B0; Deep learning; Data augmentation; Complex illumination