

基于 ResNet 系列网络的肉牛姿态识别实现与教学研究

郜静茹¹, 赵永才¹, 李慢¹, 关曼玉¹, 李广博^{1*}

(1. 淮北理工学院电子与信息工程学院, 安徽 淮北 235000)

摘 要: 精准识别肉牛姿态行为是规模化智能养殖的核心基础, 可有效实现肉牛健康状态监测与养殖行为智能化管控。当前肉牛养殖姿态识别缺乏标准化数据集与简洁可靠的基线模型, 且多数研究未探究无数据增强场景下网络深度对识别精度的影响, 制约了算法落地应用。本文构建包含饮水、采食等 9 类肉牛常见姿态的图像数据集, 各类别均选取 500 张样本并按 8:1:1 比例划分数据集。研究以 ResNet50 为骨干网络搭建分类模型, 仅采用缩放与归一化基础预处理, 搭配 Adam 优化器与余弦退火学习率策略完成模型训练。模型测试集总体准确率达 92.00%, 宏平均 F1 值为 91.86%, 饮水、躺卧、爬跨姿态识别效果优异, 仅站立、行走姿态识别难度较高。实验表明, ResNet18、ResNet34、ResNet50 识别精度随网络深度递增显著提升, 测试集最高准确率可达 92.00%。本研究明确了无数据增强条件下网络深度对肉牛姿态识别的提升作用, 构建的数据集与实验基准可为复杂养殖场景下的识别算法优化提供重要支撑。

关键词: 肉牛姿态识别; 深度学习; ResNet

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研究项目 (重大) (2024AH040217); 省级质量工程-教学研究项目 (2023jyxm1009; 2023xjzlt117; 2023sdx145); 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.v1i9.69

规模化智能养殖场景中, 肉牛姿态行为识别是实现牛只健康监测、发情检测与精细化饲养管理的重要支撑, 各类行为特征可直接反映肉牛健康与生产状态^[6]。传统人工观察方式主观性强、效率低、无法全天候监测, 难以满足规模化养殖需求。基于深度学习的机器视觉识别技术具备非接触、可连续作业、客观性高的优势, 可有效解决传统监测方式的痛点, 是肉牛行为智能感知的主流技术方向。

目前肉牛行为识别研究主要分为目标检测与图像分类两大路线, 相关研究已构建多类别肉牛姿态数据集, 并针对光照干扰等难点开展优化, 验证了算法的应用可行性。但现有研究多聚焦于网络结构与检测框架的改进, 忽略了骨干网络深度对识别性能的影响, 且缺少无数据增强受限条件下的对照实验, 缺乏标准化基线模型与系统的消融对比基准, 不利于算法的进一步优化迭代。

针对上述问题, 本文构建包含 9 类肉牛典型姿态、覆盖多光照场景的均衡图像数据集, 并按 8:1:1 比例划分样本。本文以 ResNet 系列网络为基础搭建分类模型, 仅采用缩放、归一化基础预处理, 在无数据增强的条件下完成模型训练, 并通过单一变量对照实验, 探究不同网络深度对肉牛姿态识别效果的影响。

本文主要贡献如下: 构建了场景丰富、样本均衡的肉牛九类姿态数据集, 为相关研究提供数据基础; 在极简预处理条件下搭建了高精度 ResNet50 姿态识别基线模型; 通过系统性消融实验明确了网络深度对识别精度与模型参数量的影响规律, 可为复杂养殖场景下肉牛姿态识别算法的改进与落地提供参考。

¹ **作者简介:** 郜静茹 (2007—), 女, 安徽阜阳人, 本科生。研究方向为人工智能;

赵永才 (2007—), 男, 安徽蚌埠人, 本科生。研究方向: 人工智能;

李 慢 (2005—), 女, 安徽池州人, 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

一、相关工作

(一) 畜禽行为识别研究进展

深度学习技术已广泛用于牛、猪、羊、鸡等畜禽的典型行为识别,与人工观察相比,基于深度学习的识别方法具有客观、高效、可连续运行等优势,已成为畜禽行为智能感知的主流技术路线^[1]。在牛只行为识别领域,基于视频分析的奶牛智能监测研究取得了较多进展:视频分析技术具备无接触、低成本优势,奶牛行为识别可辅助早期疾病发现;注意力机制与混合卷积结构可提升识别准确率,轻量化模块则可降低模型计算复杂度^[2]。上述研究普遍表明,行为识别有助于牛只健康监测与精细化管理,模型精度与计算复杂度之间的平衡是工程应用中必须考虑的问题^[1-2]。

(二) 肉牛行为识别方法

针对肉牛这一研究对象,已有研究者构建了包含站立、躺卧、爬跨、争斗、舔舐、采食、饮水、行走、探寻 9 类行为的肉牛行为数据集,并提出 YOLOv8n_BiF_DSC 模型,通过动态蛇形卷积与注意力机制增强特征提取能力,改进后平均精度高于基础 YOLOv8n 模型^[3]。该研究还指出,同一姿态在不同角度与光照条件下会呈现不同特征表达,光照与背景干扰是特征提取的主要挑战。另有研究针对夜间低光照场景,采集肉牛站立、躺卧、进食、饮水 4 类行为,提出基于改进 YOLOv11n 的低光照图像增强识别方法,在夜间条件下取得了较高的检测精度^[4]。这些工作表明,肉牛行为识别已从可行性验证走向复杂场景适配阶段,光照鲁棒性、类别粒度与计算效率成为研究重点。

(三) ResNet 残差网络

ResNet 通过引入残差学习机制,在深层网络中建立了从输入到输出的恒等捷径连接,有效缓解了网络加深带来的梯度消失与训练退化问题^[4]。ResNet18 与 ResNet34 采用基础残差单元,ResNet50 采用三层瓶颈残差单元,在保持计算开销相对可控的同时显著提升了特征表达能力,在 ImageNet 等大规模图像分类任务上取得了当时最优的结果^[5]。由于结构简洁、性能稳定且易于部署,ResNet 系列至今仍是图像分类与特征提取任务中最常用的骨干网络之一,适合作为姿态识别研究的基线模型。

二、方法

(一) ResNet 骨干网络结构

本文采用 ResNet 系列作为姿态分类的骨干网络。ResNet 通过残差单元在非线性变换之外引入恒等映射,使网络可以学习相对于输入的残差,从而在加深网络层数时保持梯度有效传播。其中,ResNet34 由基础残差单元堆叠而成,ResNet50 采用三层瓶颈残差单元(Bottleneck),在保证特征表达能力的同时控制计算开销^[5]。两个网络的输入均为 $224 \times 224 \times 3$ 的图像,经初始卷积层与最大池化后依次通过 4 个残差阶段,随后经全局平均池化输出特征向量,并接入含 9 个输出节点的全连接分类层,对应 9 类肉牛姿态。模型采用随机初始化,不使用 ImageNet 预训练权重,以保证消融对比的公平性。

(二) 训练策略

实验统一采用如下训练配置:损失函数为交叉熵损失;优化器为 Adam,初始学习率 1×10^{-4} ,权重衰减系数 1×10^{-4} ;学习率调度采用余弦退火策略,总退火周期与训练轮数一致;批大小 32;训练轮数 100;固定随机种子 42,以保证数据加载、参数初始化与训练过程的不确定性。训练过程开启断点续训,每个轮次结束后保存最近检查点,并在验证集准确率刷新时自动保存最优模型;训练日志以 CSV 格式逐轮记录训练损失、训练准确率、验证损失、验证准确率与学习率。

三、实验

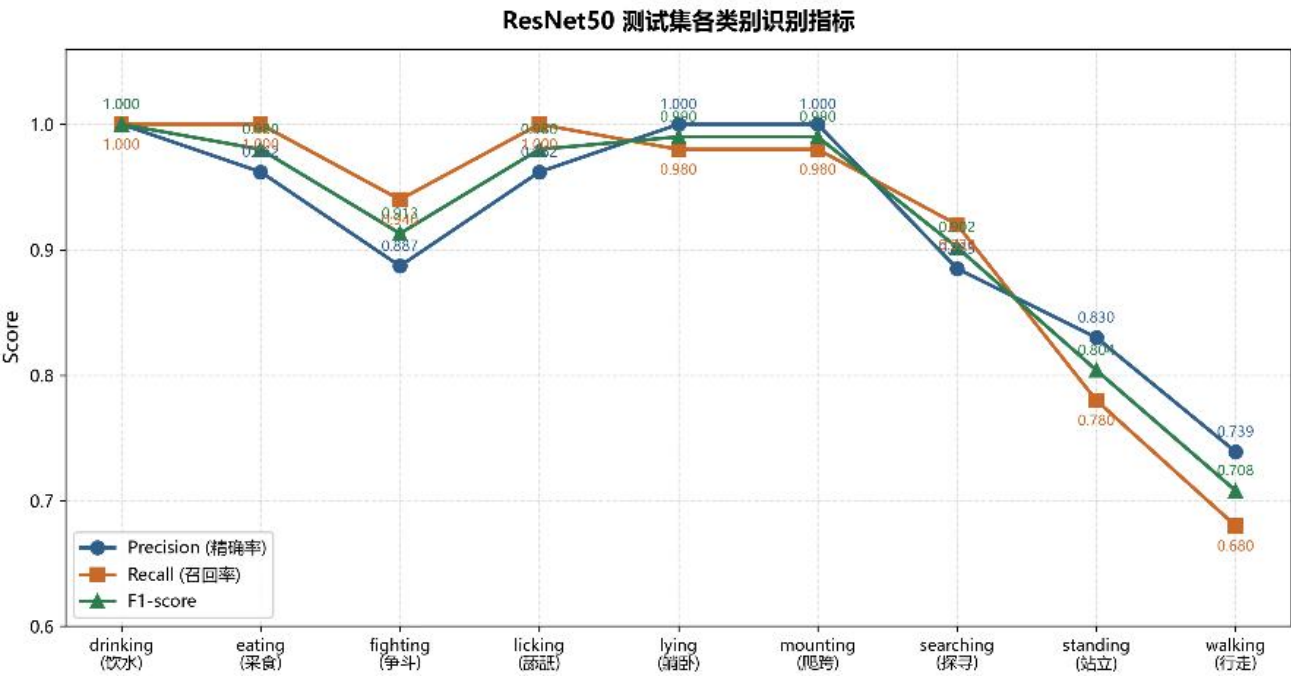
(一) 实验环境与评价指标

实验硬件环境为 NVIDIA GeForce RTX 3050（6 GB 显存），软件环境为 Python 3.13、PyTorch 2.11.0（CUDA 12.8）与 Torchvision 0.26.0，所有训练与推理均基于 GPU 3050 完成。评价指标包括测试集总体准确率（ACC）、每类精确率（Precision）、召回率（Recall）、F1 值、宏平均 F1 值，以及衡量模型复杂度的参数量（Params）与乘加运算次数（MACs，即 FLOPs 的一半）。

(二) 实验一：ResNet50 基线结果

实验一以 ResNet50 为骨干网络完成 100 轮训练，总耗时约 83.1 分钟。训练过程中，训练损失从第 1 轮的 1.68 快速下降，训练准确率稳步上升，至约第 55 轮后训练准确率达到 100%；验证损失整体呈下降趋势但在训练后期存在一定波动，验证准确率在 80.00%至 93.00%之间波动，最优验证准确率为 92.67%，出现在第 58 轮。测试集上的总体准确率为 92.00%。ResNet50 在测试集上各姿态类别的识别指标如表 1 所示。

表 1 ResNet50 在测试集上的每类识别指标



(三) 深度消融对比实验

在固定数据集、划分比例、预处理方式、训练参数与实验环境不变的前提下，仅更换骨干网络，完成 ResNet34 的 100 轮的完整训练，并与实验一的 ResNet50 结果统一在测试集上测评，得到两组模型的性能对比如表 2 所示。

表 2 两种 ResNet 深度模型的性能对比

模型	参数量 (M)	MACs(G)	最优验证准确率	测试集准确率	宏平均 F1	训练耗时 (分钟)
ResNet34	21.29	3.68	0.9000	0.8622	0.8593	58.0 分钟
ResNet50	23.53	4.13	0.9267	0.9200	0.9186	83.1 分钟

1.从精度维度看，随着网络深度加深，识别性能逐步提升：ResNet34 与 ResNet50 的测试集准确率分别为 86.22% 与 92.00%，宏平均 F1 值分别为 85.93%与 91.86%。ResNet50 较 ResNet34 提升 5.78 个百分点，提升幅度明显。

2.从复杂度维度看，参数量由 ResNet34 的 21.29 M 增至 ResNet50 的 23.53 M，MACs 由 3.68G 增至 4.13 G，ResNet50 较 ResNet34 的参数量与计算量均约为后者的 1.1 倍，但仍处于轻量级范围，单次前向推理开销可接受。

3.从计算开销看，两个模型的 FLOPs 分别为 7.36 G 与 8.26 G，ResNet50 在单卡 RTX 3050 上的单轮训练耗时

约 50 秒，100 轮总耗时 83.1 分钟，训练与推理开销均在可接受范围内；在计算资源受限的部署场景中同样具有实用价值。综合精度与复杂度两个维度，ResNet50 在本任务中的性价比最高，为最优模型。

四、可视化分析

为直观分析模型的识别行为，对 ResNet50 的训练过程、分类结果与特征分布进行了可视化。训练过程中的损失曲线与准确率曲线如组图 1 所示，测试集混淆矩阵，t-SNE 图，各类别样本预测置信度分布小提琴图，以及三模型的验证集曲线对比与主要指标对比如组图 2 所示。

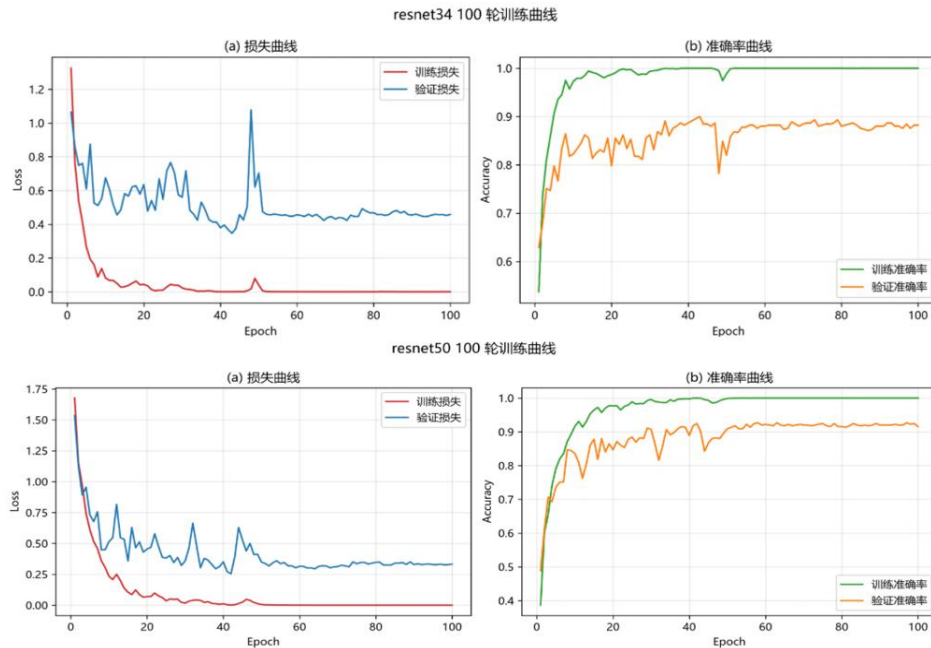


图 1 ResNet 34、50 的损失曲线与准确率曲线

从混淆矩阵可以看出，绝大多数样本位于对角线，其中饮水、采食、舔舐、躺卧等类别几乎没有混淆；主要误判集中于站立与行走之间的相互混淆，以及行走被误判为站立或探寻的情况，说明这两类姿态在图像特征上具有较高的相似性，是后续优化的重点方向。t-SNE 可视化结果显示，各姿态类别的特征在二维空间中整体呈分簇分布，类别间边界基本清晰；其中饮水、躺卧、爬跨等类别的特征簇紧凑且相互分离，而站立与行走类别的特征簇存在部分交叠，与混淆矩阵的结论相互印证。

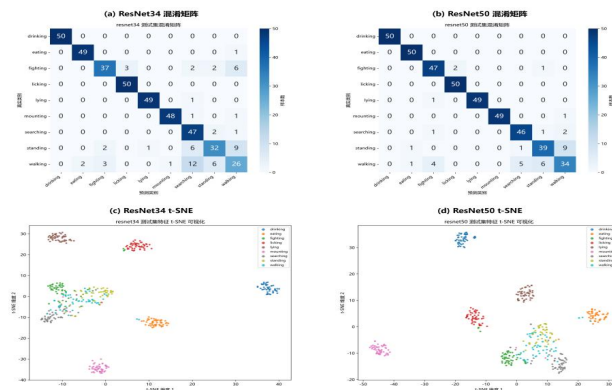


图 2 ResNet34、50 的混淆矩阵与 t-SNE 可视化

小提琴图 3、图 4 显示，模型对大多数类别样本给出了较高的预测置信度，置信度分布集中在 0.9 以上；

站立与行走类别的置信度分布相对分散，低置信度样本比例更高，反映了这两类样本的分类不确定性较大。

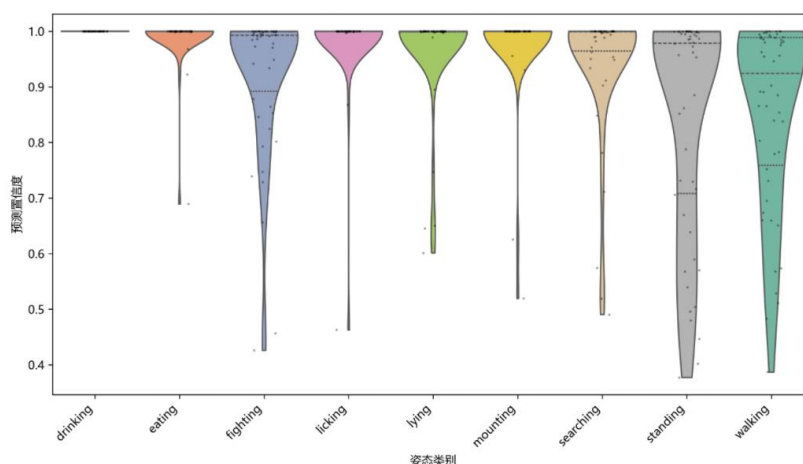


图 3 ResNet34 的小提琴图

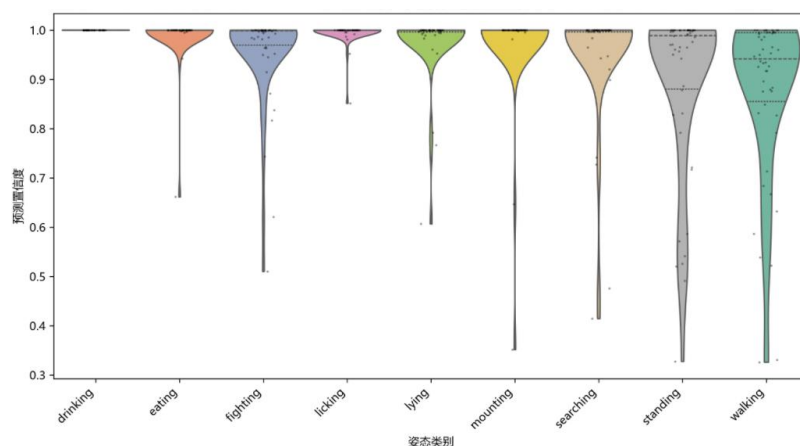


图 4 ResNet50 的小提琴图

五、结论

本文围绕肉牛姿态识别任务，构建了包含 9 类姿态、每类 500 张样本且覆盖多种光照场景的图像数据集，在仅缩放与归一化、不使用数据增强的预处理条件下，以 ResNet50 为骨干网络完成了 100 轮训练并取得 92.00% 的测试集准确率，其中饮水、躺卧、爬跨等类别的 F1 值达到 99.00% 以上。通过固定其余实验条件的深度消融对比发现，ResNet34 与 ResNet50 的测试集准确率分别为 86.22% 与 92.00%，加深网络深度能够有效提升姿态识别精度，ResNet50 在精度与复杂度的综合权衡中表现最优。混淆矩阵、t-SNE 与置信度分布分析表明，站立与行走类别是主要误判来源，后续可针对该类别的特征相似性开展专项优化。本文结果为肉牛姿态识别提供了可靠的基线与消融基准，也为后续引入数据增强、预训练初始化或轻量化结构等改进提供了对照基础。

参考文献：

- [1] 朱芷芄, 王海峰, 李斌, 等. 深度学习在畜禽典型行为识别中的研究进展[J]. 中国农业科技导报, 2024, 26(10).
- [2] 宋怀波, 王云飞, 邓洪兴, 等. 基于视频分析的规模化奶牛智能监测技术研究进展[J]. 华南农业大学学报, 2024, 45(5): 649-660.
- [3] LI Guangbo, SHI Guolong, ZHU Changjie. Dynamic Serpentine Convolution with Attention Mechanism Enhancement for Beef Cattle Behavior Recognition[J]. Animals, 2024, 14(3): 466.
- [4] 刘星宇, 王芳, 任力生. 基于改进 YOLOv11n 的低光照图像增强下夜间肉牛行为识别方法[J]. 农业工程学报, 2025, 41(19).
- [5] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, 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.

[6] LI G B, SUN J Y, GUAN M Y, et al. A new method for non-destructive identification and tracking of multi-object behaviors in beef cattle based on deep learning[J]. *Animals*, 2024, 14(17): 2464

Research on a Beef Cattle Posture Recognition Method Based on ResNet Deep Learning Networks

Gao Jingru¹, Zhao Yongcai¹, Li Man¹, Guan Manyu¹, Li Guangbo^{1*}

¹*School of Electronics and Information Engineering, Huaibei Institute of Technology, Huaibei, Anhui 235000, China*

Abstract: Accurate recognition of beef cattle postures and behaviors serves as the core foundation for large-scale intelligent livestock farming, enabling effective monitoring of cattle health status and intelligent management of farming behaviors. Currently, there is a lack of standardized datasets and simple, reliable baseline models for beef cattle posture recognition; furthermore, most existing studies have not investigated how network depth affects recognition accuracy in scenarios without data augmentation, which hinders the practical deployment of such algorithms. This study constructs an image dataset encompassing nine common beef cattle postures—including drinking and feeding—with 500 samples selected for each category, which are then divided into training, validation, and test sets in a ratio of 8:1:1. A classification model is developed using ResNet50 as the backbone network, employing only basic preprocessing (scaling and normalization) and training the model using the Adam optimizer combined with a cosine decay learning rate strategy. The model achieves an overall accuracy of 92.00% on the test set and a macro-F1 score of 91.86%; it demonstrates excellent recognition performance for drinking, lying, and climbing/leaping postures, whereas recognizing standing and walking postures presents greater challenges. Experiments demonstrate that the recognition accuracy of ResNet18, ResNet34, and ResNet50 increases significantly with increasing network depth., with the highest accuracy on the test set reaching 92.00%. This study clarifies the enhancing role of network depth in beef cattle posture recognition under conditions without data augmentation; the constructed dataset and experimental benchmarks can provide valuable support for optimizing recognition algorithms in complex livestock farming scenarios.

Keywords: beef cattle posture recognition; deep learning; ResNet;