

文章编号: 2095-2163(2023)06-0148-06

中图分类号: TP391.41

文献标志码: A

结合注意力机制 GoogLeNet 的羊只个体识别

李章辉, 王天一, 李远征

(贵州大学 大数据与信息工程学院, 贵阳 550025)

摘要: 随着人工智能技术的快速发展, 针对羊只的智能化养殖技术需求也日益渐增。为解决羊只个体识别工作量大、工作效率低和人工依赖性强等问题, 本文提出了结合注意力机制的 GoogLeNet 对羊只进行个体识别。模型引入通道注意力机制 SE-Net (Squeeze-and-Excitation Networks) 模块, 将通道内有用信号进行筛选放大, 减弱噪声的影响。在训练过程中, 对学习率、优化器参数进行了调整。实验结果表明, 改进网络与原始网络相比, 在羊个体识别任务上具有更高的准确度以及较好的泛化性。

关键词: 智能化养殖; 羊个体识别; 注意力机制; 人工智能

Recognition of sheep individual based on GoogLeNet combined with attention mechanism

LI Zhanghui, WANG Tianyi, LI Yuanzheng

(College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China)

【Abstract】 With the rapid development of artificial intelligence technology, the demand for intelligent sheep breeding technology is also increasing. In order to solve the problems of large workload, low work efficiency and strong artificial dependence on sheep individual identification, this paper proposes to use GoogLeNet combined with attention mechanism to identify sheep individual. This model introduces the channel attention mechanism in SE-Net (Squeeze-and-Excitation Networks) to screen and amplify the useful signals in the channel to reduce the influence of noise. During the training process, the learning rate and optimizer parameters are adjusted. The experimental results show that compared with the original network, the improved network has higher accuracy and better generalization in the sheep individual recognition task.

【Key words】 intelligent farming; sheep individual identification; attention mechanism; artificial intelligence

0 引言

随着智能化养殖^[1]的快速发展, 智能化养殖场建设和畜牧养殖精准化逐渐成为高效管理的理念。高效的养殖场管理需要精准化到个体的身份信息, 因此收集和识别个体信息变得尤为重要。在智能化养殖过程中, 为每只羊建立个体档案有利于对养殖场精准化饲养和实时监控每只羊的身体健康信息提供了前提条件。传统的养殖场对动物的个体识别主要靠人工识别, 但在大型养殖场, 仅靠人工无法做到有效的个体识别。由于智能化养殖具有效率高、识别准确以及人力成本低等优势, 因此智能化养殖正在逐步取代传统化养殖。

智能化养殖减少了人工在大规模养殖中管理不

当等问题, 涉及到个体识别^[2]、目标检测^[3]等领域。在动物个体识别领域, 韩丁等^[4]提出了将空间变换网络引入 VGGNet 网络, 对羊面部表情数据集的实验。宋一凡等^[5]对比和总结了以牛只面部、躯干和口鼻 3 个部位特征的非接触式机器视觉^[6]识别方法的研究成果。马娜等^[7]设计了卷积层-池化层-卷积层-池化层 2 层卷积神经网络^[8]模型, 对猪个体身份进行识别。拉毛杰等^[9]使用卷积神经网络改进了畜牧业动物图像的识别准确率。在动物目标检测领域, 魏斌等^[10]基于深度学习模型的羊脸检测效果较为理想, 而羊脸识别工作在正面羊脸上取得了较高的准确率。张宏鸣等^[11]提出通过添加长短距离语义增强模块进行多尺度融合, 结合 Mudeep 重识别模型实现了肉牛多目标跟踪。张宏鸣等^[12]提

基金项目: 贵州省科学技术基金(ZK[2021]304); 贵州省科技支撑计划([2021]176)。

作者简介: 李章辉(1997-), 男, 硕士研究生, 主要研究方向: 图像处理; 王天一(1989-), 男, 博士, 副教授, 主要研究方向: 量子通信、深度学习、大数据与人工智能; 李远征(1997-), 男, 硕士研究生, 主要研究方向: 图像处理。

通讯作者: 王天一 Email: tywang@gzu.edu.cn

收稿日期: 2022-05-13

出了一种基于 YOLOv3 模型和卷积神经网络的多目标肉牛进食行为识别方法,为养殖场中对动物行为非接触式检测提供了新的途径。何东健^[13]提出基于最大连通区域的目标循环搜索环境建模、目标检测算法,提高了识别犊牛动作的正确率(如:躺、站、走、跳等)。

本文在 GoogLeNet 网络中引入注意力机制模块组成 SE-GoogLeNet,将有用信号进行筛选放大,对 dropout 参数进行调整。在训练过程中,调整学习率大小以及对优化器参数进行优化,提高了对复杂背景下羊的识别准确率。实验表明,改进后的 SE-GoogLeNet 相比改进前的模型,验证集精度提高了 2.2%,并利用采集羊图像对训练好的模型进行了评价。

1 SE-GoogLeNet

1.1 GoogLeNet

在一些典型的深度学习网络中,大都存在着网络训练收敛慢、训练时间长、容易出现梯度消失和梯度爆炸等现象。为避免上述问题发生,在 GoogLeNet 网络中提出了 Inception 模块,网络深度有 22 层,结构如图 1 所示。GoogLeNet 网络采取全局均值池化策略,代替全连接层以减少参数,并增加了两个辅助分类器帮助训练,将 Inception (4a) 和 Inception (4d) 模块的输出用作分类,并按一个较小的权重(0.3)加到最终分类结果中,不仅达到了模型融合的效果,同时给网络增加了反向传播的梯度信号,也提供了额外的正则化,有益于网络的训练。

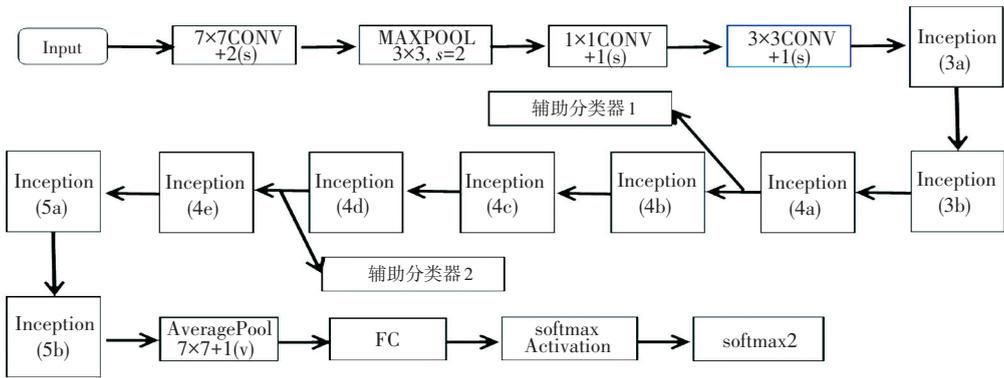


图 1 GoogLeNet 网络模型图

Fig. 1 GoogLeNet network model architecture

Inception 模块中存在滤波器,且滤波器的输出合并构成下一层的输入。具体来说,Inception 模块利用 1×1 卷积核降维,再利用 3×3、5×5 卷积核进行卷积操作,将输入的特征矩阵分别与这 4 个分支运算并得到 4 个输出,输出在深度上拼接得到最终的输出(使用 Padding 填充让 4 个分支输出能在宽度和高度上保持一致)。使用不同大小卷积核融合了

不同尺度的特征信息,增加了网络的宽度以及网络对尺度的适应性,使得卷积网络在特征提取过程中得到了不同大小的感受野,使网络在训练过程中对不同大小个体有了更强的识别能力。如图 2 所示,Inception 模块结构增加了网络的宽度,可以达到更好的识别效果。将网络由深变宽,避免了网络过深出现梯度消失和模型退化的问题。

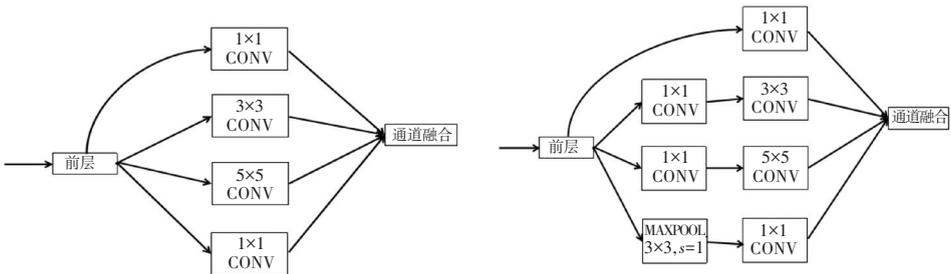


图 2 Inception 模块

Fig. 2 Inception module

相比于传统的 Inception 模块,GoogLeNet 网络提出的 Inception 模块在分支 2、3、4 中添加了 1×1 的卷积层,这 3 个 1×1 的卷积核是为了对通道方向上进行降维,减少参数量。如图 3 所示,以 Inception

(3a) 模块中 1×1 卷积后连接 5×5 卷积分支为例计算参数量,使用 1×1 卷积操作的模型参数仅为未使用 1×1 卷积模型参数的 12.9%。

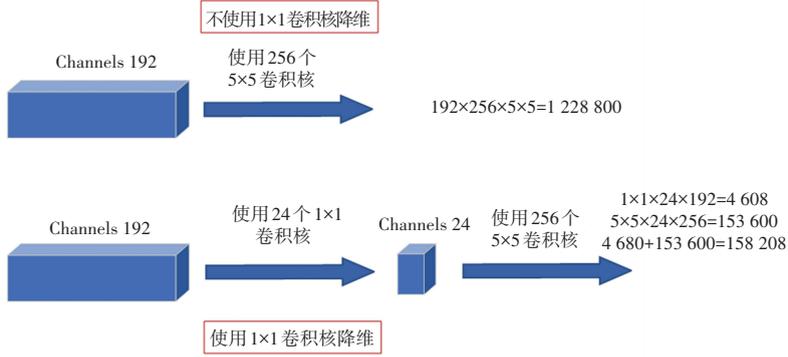


图 3 Inception 模块参数量对比

Fig. 3 Inception module parameter comparison

1.2 改进 SE-GoogLeNet

虽然 GoogLeNet 网络层数比较深,有较强的特征提取能力,但本文研究的是在养殖场环境下的羊个体识别,在识别过程中会遭受到不同程度噪声的干扰,这些噪声的干扰也会随着网络的训练进行传递,对识别个体精度造成不利的影响。

本文提出的改进 GoogLeNet 网络结构如图 4 所示。该网络在主分类器末的平均池化层后引入了注意力机制模块,以改变每个通道的权重,将有用的特征进行放大,去除池化后学习到冗余特征信息。为了防止过拟合现象的发生,将主网络中随机失活(dropout)层随机失活神经元由 40%增大至 70%。

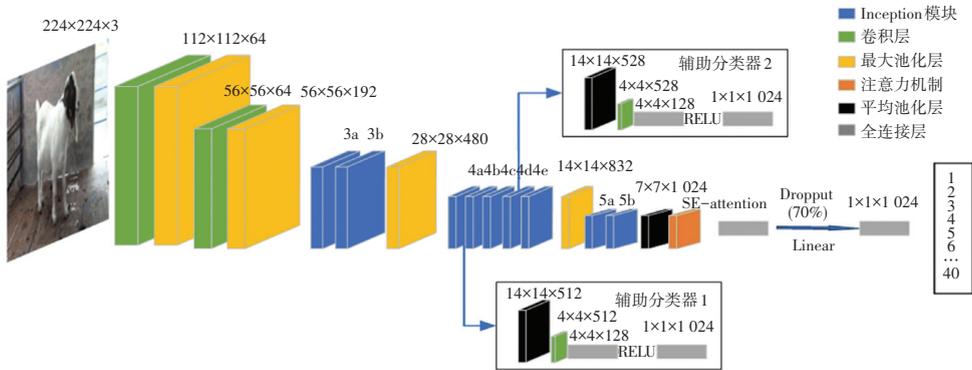


图 4 SE-GoogLeNet 网络结构图

Fig. 4 SE-GoogLeNet network structure diagram

注意力机制结构如图 5 所示,该机制是对输入进来的特征层进行压缩操作,每一层进行平均池化操作得到本层的参数。池化操作是在水平和垂直两个维度进行,针对于 C 个通道得到 C 个平均池化后的输出,接着进行两次全连接层,对输出结果进行一个 Sigmoid 激活函数处理,将输出的值固定在 0~1 之间,得到每一个通道的权值,将此权值乘以输入进

来的每个特征层,得到最终输出。压缩操作函数公式如式(1)所示:

$$Z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (1)$$

其中, Z_c 表示输出特征层; F_{sq} 表示压缩操作函数; H, W 表示特征图的宽和高; i, j 表示特征图上的坐标位置。

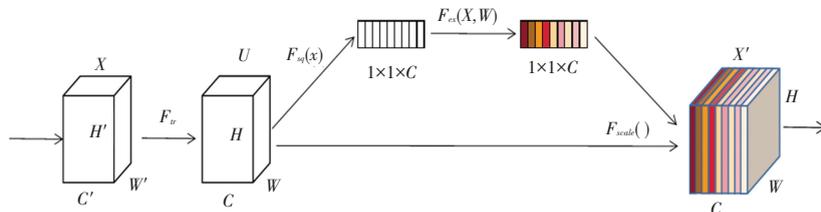


图 5 注意力机制结构图

Fig. 5 Structure diagram of attention mechanism

由于本文采集到的数据集较少,易发生过拟合现象,因此对主分类器中的随机失活(dropout)层随机失活从 40%增大至 70%。在网络训练过程中,从中随机丢弃一些隐藏单元,相当于在一个大的模型中随机抽取出一个小的模型。由于随机失活的过程是动态的,每次训练过程中随机失活的神经元都有可能是不相同的。也就是说只训练当前子模型使用过的参数,而对于未使用过的参数不进行更新。这些抽取出来的子模型共享参数,最终相当于训练了很多的子模型。在进行预测时,综合所有子模型的预测结果,很好的解决了过拟合的问题。具体公式如式(2)所示:

$$r_j^{(l)} \sim \text{Bernoulli}(p) \quad (2)$$

其中, $\text{Bernoulli}(p)$ 是指伯努利分布。以一个概率为 p 的伯努利分布,随机生成与节点数相同的 0、1 值, p 就是每个神经元被保留下的概率。

$$\begin{aligned} \tilde{y}^{(l)} &= r^{(l)} * y^{(l)} \\ z_i^{(l+1)} &= W_i^{(l+1)} \tilde{y}^{(l)} + b_i^{(l+1)} \\ y_i^{(l+1)} &= f(z_i^{(l+1)}) \end{aligned} \quad (3)$$

在经过伯努利分布计算后得出概率 p 后,对输入的神经元重新计算进行随机失活。式 3 中, $Z_i^{(l+1)}$ 是第 l 层的输入; $y_i^{(l+1)}$ 是第 l 层的输出; $W_i^{(l+1)}$ 是第 l 层的权重矩阵; f 是激活函数。

针对多分类问题,主分类器和副分类器的交叉



图 6 羊部分图片数据集

Fig. 6 Examples of sheep data set

由于在网络训练中需要大量的数据来达到更好的训练效果,提高网络检测的精度,但由于采集到的图片数量有限,将对抽取的 600 张图片进行数据增强,增加数据数量,防止在网络训练中出现过拟合的现象。本文采用的数据增强主要是对图片进行多个角度的旋转、对比度的调整、亮度的调整、随机平移等方式对源图片进行数据扩充,数据扩充为原始图

熵损失函数表达式如式(4)所示:

$$L = \frac{1}{N} \sum_i L_i = - \frac{1}{N} \sum_i \sum_{c=1}^M y_{ic} \log(p_{ic}) \quad (4)$$

其中, M 表示类别的数量; y_{ic} 表示符号函数(0 或 1),若样本 i 的真实类别等于 c 取 1,否则取 0; P_{ic} 表示观测样本, i 属于类别 c 的预测概率。

SE-GoogLeNet 网络采用一个主分类器和两个辅助分类器进行输出,用定义好的损失函数分别求得 3 个预测分类器的预测与真实标签之间的损失,将主分类器的损失值与两个副分类器损失值的 0.3 倍相加得到总损失值。总损失值计算如式(5)所示:

$$\text{Loss} = \text{Loss0} + \text{Loss1} * 0.3 + \text{Loss2} * 0.3 \quad (5)$$

其中, Loss 表示总损失; Loss0 表示主分类器损失; Loss1 、 Loss2 表示副分类器损失。

2 实验结果与分析

2.1 实验数据

本文中数据均来自贵州省某羊养殖场的拍摄视频,分别对 40 只羊进行了视频拍摄(分辨率为 1 920×1 080)。经过 Adobe Premiere Pro 软件进行视频抽帧,对视频中羊的全身、头部等突出的个体特征进行图像筛选,并将不同成长期的羊只混合在一起,共计抽取 600 张样本图像。部分图像示例如图 6 所示。

片的 7 倍数量,共计 4 200 张羊的图像。

本文在训练过程中将数据按照 8 : 2 的比例划分为训练集和验证集。考虑到每张照片尺寸大小的原因,使用 transforms 工具包将所有的样本图像进行随机裁剪,设置大小为 224×224 像素,并将其进行均值和方差都为 0.5 的标准化处理,以便高效的对网络进行训练。处理后训练集和验证集的数据量见表 1。

表1 数据集统计

Tab. 1 Statistics of dataset

羊个体数	数据总量	训练集数量	验证集数量
40	4 200	3 360	840

2.2 实验环境

本次实验是在 Python3.9、Pytorch1.11 的环境下进行,采用 RTX2070 加速训练,GPU 显存为 8 GB,CPU:AMD Ryzen 5 3600,CPU 基频为 3.49 GHZ。

2.3 超参数设置

在深度学习网络训练中,超参数选择的好坏不仅和模型自身有关,还和软硬件的配置相关。本实验在进行超参数调整之后,利用改进的 SE-GoogLeNet 模型进行训练时,采用批量训练方式对训练集进行随机打乱,验证集保持不变,Batch_size 的大小设置为 16,训练周期为 60。采用 Adam 优化器训练,eps 值设为 $1e-5$,初始学习率为 0.000 08。

2.4 结果分析

为了了解不同优化算法^[14]对 SE-GoogLeNet 网络模型性能的影响,本文选择 Adam、SGD、RMSprop 3 种算法在相同的实验环境下进行训练,结果见表 2。由表 2 中数据可见,Adam 的训练时长最短,且精度最高,相比 RMSprop 算法、SGD 算法高出了 1.4%、29.9%。因此,本文采用 Adam 优化算法进行训练。

表 2 不同优化算法下的模型精度

Tab. 2 Model accuracy under different optimization algorithms

优化算法	训练时间:/h	验证集准确率/%
SGD	2.1	64.6
Adam	1.6	94.5
RMSprop	1.8	93.1

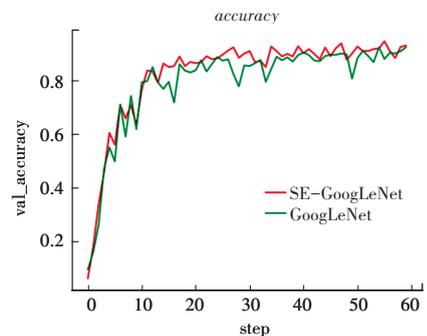
在采集的羊图像数据集上进行训练,保证了每个模型在相同实验环境下完成训练。在训练过程中,每个训练周期后都会对模型的精确度进行验证并保存数据。引入注意力机制后,模型的识别准确率得到了较大的提升,但是模型的大小增长非常小,进一步显示出模型性能的优越性。改进后的网络与 3 种经典分类网络在羊只数据集上的准确率和模型占用内存容量对比见表 3。

表 3 模型准确率和模型占用内存容量对比

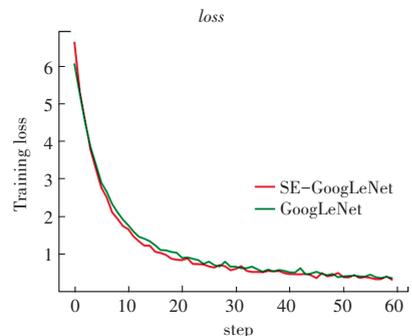
Tab. 3 Comparison of model accuracy and model memory capacity

模型	验证集准确率/%	内存容量/MB
SE-GoogLeNet	94.5	40
GoogLeNet	92.3	39.8
AlexNet	90.1	55.9
VGG16	91.8	512

本文将数据集在 GoogLeNet、改进后 SE-GoogLeNet 上的验证集准确率的曲线展示如图 7 所示。



(a) 验证集准确率对比图



(b) 损失值曲线对比图

图 7 改进前与改进后 GoogLeNet 网络在羊图像上的验证集精确度与损失值曲线对比图

Fig. 7 Comparison of validation set accuracy and loss value curves of GoogLeNet network on sheep images before and after improvement

由图 7 可见不同模型之间的曲线趋势,SE-GoogLeNet 代表本文改进后的模型曲线。模型学习羊只身体特征,不断提高对个体的分类能力。本文构建的基于注意力机制的 SE-GoogLeNet 模型准确率更高,分类效果更好(如图 7(a))。其主要原因是:一方面在于 Inception 模块将不同大小个体学习到的特征信息进行多尺度融合(Multiscale fusion),Inception 模块有不同尺寸大小的卷积核进行提取特征,并将这些特征在深度上进行融合,进一步提高了模型的特征提取能力;另一方面在于加入注意力机制将有用信息的权重增强,削弱了噪声的干扰,因而提高了模型最终的分类能力。

Loss 损失函数在模型训练中是一个重要参数,可以通过其下降趋势反映训练效果的好坏。在训练网络过程中,不断地将数据传入到网络中进行迭代,经过正向传播后得出一个预测值,将预测值与实际标签做出对比得到一个损失值,将得到的损失值进行反向传播来修正权重的误差,使得损失函数不断地减小,依次循环,训练精度将会不断的得到提高。

由不同模型在训练时的损失函数对比(图 7(b))可以看出, $Loss$ 值在第 55 个周期时已经收敛, 达到了模型精度最高且泛化性能最好点的值。

利用 SE-GoogLeNet 网络模型训练得到的权重

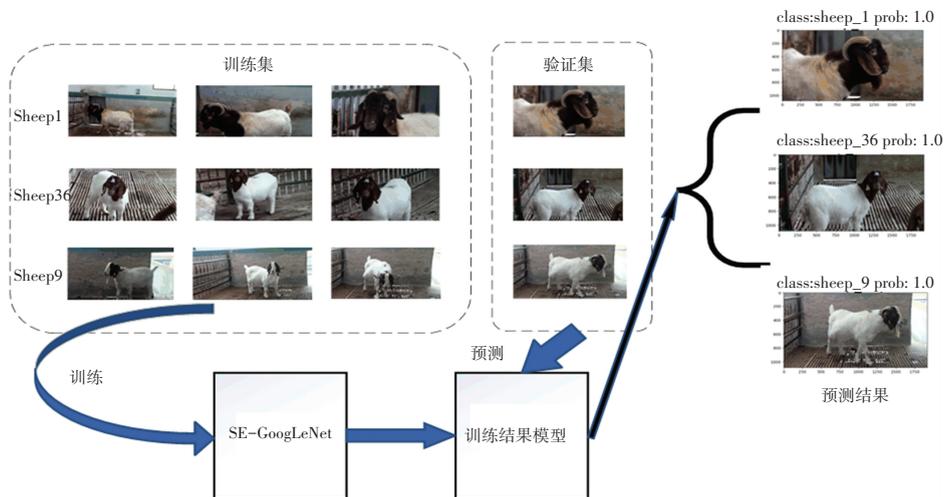


图 8 羊只预测结果

Fig. 8 Prediction result of sheep

3 结束语

本文针对目前大规模羊场养殖中存在的羊个体识别的效率低、人工依赖性较大等问题, 提出了基于结合注意力机制的 GoogLeNet 网络改进的 SE-GoogLeNet 模型, 在网络中引入注意力机制模块用来减少噪声影响, 提高了模型的精度; 同时, 在相同实验环境下研究了不同优化算法对模型性能的影响。实验结果表明, 改进后的 SE-GoogLeNet 模型在对 40 只羊的验证集准确率达到 94.5%, 相对于原 GoogLeNet 模型精确率上升了 2.2%, 改进后的 SE-GoogLeNet 模型相比于其它主流算法具有较高的准确率, 在智能化养殖中将会有很好的前景。

参考文献

- [1] 王怀禹. 我国养猪智能化发展研究与应用进展[J]. 猪业科学, 2021, 38(9): 38-41.
- [2] 何钦, 李根, 严永煜, 等. 基于图像深度学习的牛个体的识别与统计[J]. 电子测试, 2021(21): 68-71, 28.

对随机选取的 3 只羊图像进行预测, 预测结果均达到了 100% 准确率, 说明改进后的 SE-GoogLeNet 能较好的对羊只个体识别, 羊只预测结果如图 8 所示。

- [3] 更桑吉, 安见才让. 畜牧业动物图像目标检测改进研究[J]. 计算机时代, 2021(10): 20-22, 27.
- [4] 韩丁, 王斌, 王亮, 等. 基于改进 VGGNet 的羊个体疼痛识别方法[J]. 农业机械学报, 2022(6): 311-317.
- [5] 宋一凡, 王娟, 李建丽, 等. 精准化养殖模式下牛只个体识别方法综述[J]. 黑龙江畜牧兽医, 2021(22): 48-53, 148-149.
- [6] 王少华, 何东健, 刘冬. 基于机器视觉的奶牛发情行为自动识别方法[J]. 农业机械学报, 2020, 51(4): 241-249.
- [7] 马娜, 徐苗. 基于卷积神经网络的猪只个体身份识别研究[J]. 计算机时代, 2022(4): 51-54.
- [8] 季长清, 高志勇, 秦静, 等. 基于卷积神经网络的图像分类算法综述[J]. 计算机应用, 2022, 42(4): 1044-1049.
- [9] 拉毛杰, 安见才让. 基于卷积神经网络的畜牧业动物图像识别研究[J]. 软件, 2020, 41(8): 43-45.
- [10] 魏斌, Masum Billah, 王美丽, 等. 基于深度学习的羊脸检测与识别方法[J]. 家畜生态学报, 2022, 43(3): 47-50.
- [11] 张宏鸣, 汪润, 董佩杰, 等. 基于 DeepSORT 算法的肉牛多目标跟踪方法[J]. 农业机械学报, 2021, 52(4): 248-256.
- [12] 张宏鸣, 武杰, 李永恒, 等. 多目标肉牛进食行为识别方法研究[J]. 农业机械学报, 2020, 51(10): 259-267.
- [13] 何东健, 孟凡昌, 等. 基于视频分析的犏牛基本行为识别[J]. 农业机械学报, 2016, 47(9): 294-300.
- [14] 李明, 来国红, 常晏鸣, 等. 深度学习算法中不同优化器的性能分析[J]. 信息技术与信息化, 2022(3): 206-209.

(上接第 147 页)

- [7] CADAR C, DUNBAR D, ENGLER D R. Klee: unassisted and automatic generation of high-coverage tests for complex systems programs[C]//OSDI. 2008: 209-224.
- [8] Shay Artzi, Julian Dolby, Frank Tip. Fault Localization for Dynamic Web Applications[J]. IEEE Transactions on software engineering, 2012, 38(2): 314-335.
- [9] ZHENG Yan, WANG Zan, FANG Xiangyu. Localizing multiple

soft-ware faults Based on evolution algorithm[J]. The Journal of Systems and Software, 2018, 139: 107-123.

- [10] DE MOURA L, BJØRNER N. Z3: An efficient SMT solver[C]//Tools and Algorithms for the Construction and Analysis of Systems: 14th International Conference, TACAS 2008, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2008, Budapest, Hungary, March 29-April 6, 2008. Proceedings 14. Springer Berlin Heidelberg, 2008: 337-340.