[文章编号] 1003－4684(2025)02-0100-07

基于 MobileNetV2 的轻量级羊身份识别模型

郭依蓓，吴颖丹，邵洋琳，徐久红，李 炎
（湖北工业大学理学院，湖北 武汉 430068）

[摘 要] 针对目前基于深度学习的羊身份识别模型中存在的参数量多、计算量大，难以真正应用于羊场的问题，提出一个基于 MobileNetV2 进行改进的融合注意力机制和优化激活函数的轻量级羊身份识别网络模型 LAAMNet。为使模型轻量化的同时进一步提高识别精度，选取了轻量级网络 MobileNetV2，并在其瓶颈结构的基础上，修改了激活函数，同时增加了轻量级注意力机制。实验结果表明，与 MobileNetV2 相比，LAAMNet 不仅模型参数量减少了约 40%，而且对羊个体身份的识别准确率达到 98.78%，提高了 0.18 个百分点。该模型较好地兼顾了识别准确率和复杂度，为实际羊场部署提供了思路。

[关键词] 羊身份识别；MobileNetV2；轻量级；注意力机制；激活函数
[中图分类号] TP391 [文献标识码] A

随着我国国内居民对羊肉的需求持续增长[1]，为增强羊肉的供给能力，提高养殖效益，羊养殖行业正在加快转型，羊场养殖由传统的养殖模式逐步向现代化的智慧养殖方式转变[2-4]。而羊场智慧养殖的基础就是对羊个体的身份进行精准识别[5]。目前羊身份识别方法有传统的识别和基于深度学习两类方法，传统的识别包括：戴耳标、射频识别等，不仅成本较大，而且受众多因素的影响，缺乏稳定性[6-8]。随着图像处理和深度学习技术的发展，基于动物图像和深度学习网络模型的动物身份识别技术也由此兴起。具体地，王荣等[9]基于多尺度卷积神经网络实现了复杂多变环境下猪个体识别；秦兴等[10]利用双线性卷积神经网络构建猪面部识别模型，实现对猪身份精准识别；谢碧森等基于迁移学习对 VGG16 进行微调，实现对家猪身份识别[11]；吴梦茹等[12]基于 GoogleNet、ResNet、SE-ResNet、DenseNet 四种不同的深度学习网络建立了猪脸识别模型，对比发现基于 DenseNet 网络的猪脸模型识别准确率最高；姚礼垚等[13]使用 PnasNet5 实现了对奶牛身份的识别；黄俊华等[14]利用一种改进简化后的 VGG16 卷积神经网络对奶牛牛脸进行识别。基于羊脸图像和深度学习的羊身份识别也有了一定的研究进展，张宏鸣等[15]使用改进 MobileFaceNet 实现羊只个体身份识别；魏斌等[16]使用 VGGFace 模型对正面羊脸识别取得了较高准确率；尚诚等[17]利用残差网络 ResNet18 联合三元组损失函数和交叉熵损失函数实现了对奶山羊的识别。目前基于深度学习的羊身份识别模型主要关注于提高识别的准确率，因而模型的参数量和计算量往往过大，无法部署到实际羊场的监控设备。针对轻量级羊身份识别模型的研究比较少。周利香[18]、杨家林[19]等基于改进的轻量级网络构建轻量级羊身份识别模型，不仅缩减了模型大小，而且提高了模型对羊脸识别的准确率。

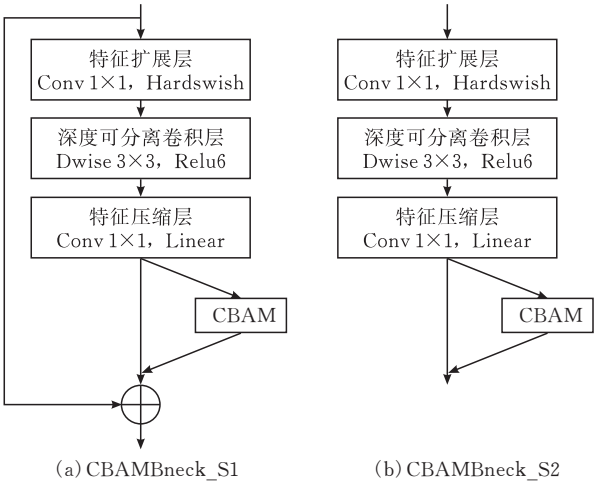
因此，本研究提出一个基于 MobileNetV2 的轻量级羊身份识别模型 LAAMNet，旨在构建一个规模较小但精度较高的轻量级羊身份识别模型。在 MobileNetV2 瓶颈结构的基础上，修改了激活函数，并添加了轻量级的通道和空间注意力模块，构建了性能更优的特征提取模块。既能减少信息损失，又能提高网络的特征提取能力，从而使模型的识别能力更强。实验结果显示，改进后的轻量级羊身份识别模型 LAAMNet 规模更小，识别准确率更高。

1 融合注意力机制瓶颈结构 CBAM-Bneck

轻量级羊身份识别模型 LAAMNet 在借鉴 MobileNetV2[20]瓶颈结构的基础上，将特征扩展层

[收稿日期] 2023－07－29
[基金项目] 湖北省社会科学基金(19Q062)
[第一作者] 郭依蓓(1995－)，女，山西临汾人，湖北工业大学硕士研究生，研究方向为遥感影像的高精度对地目标定位理论与方法。
[通信作者] 吴颖丹(1982－)，女，湖北武汉人，湖北工业大学副教授，研究方向为遥感影像的高精度对地目标定位理论与方法、高分辨率卫星遥感影像空间信息提取方法。

的 ReLU6 激活函数替换为 Hardswish 函数，并融合了轻量级 convolutional block attention module (CBAM)[21]注意力模块，构建羊脸特征提取瓶颈结构 CBAMBneck，CBAMBneck 根据不同步长，分为两种：CBAMBneck_S1 和 CBAMBneck_S2，其结构如图 1 所示。



(a) CBAMBneck_S1　　(b) CBAMBneck_S2

图 1　融合注意力机制瓶颈结构 CBAMBneck 结构

CBAMBneck_S1 和 CBAMBneck_S2 的不同在于：1)CBAMBneck_S1 结构中的深度可分离卷积层中，特征提取步长 stride 为 1，而 CBAMBneck_S2 中步长 stride 为 2；2)由于 CBAMBneck_S1 结构的输入特征矩阵与输出特征矩阵尺寸形状相同，增加了 shortcut 连接，避免因网络层数增加造成的梯度消失。

MobileNetV2 使用 ReLU6 激活函数，如

$$\mathrm{ReLU6}(x)=\begin{cases}0(x<0)\\x(0\leqslant x\leqslant 6)\\6(x>6)\end{cases}$$

但 ReLU6 在训练时较“脆弱”，输入为负时，会出现神经元永久性“死亡”，网络无法更新，导致信息损失。因此，将特征扩展层中的 ReLU6 激活函数替换为 Hardswish 激活函数，如

$$\mathrm{Hardswish}(x)=\begin{cases}0(x\leqslant -3)\\x(x+3)/6(-3<x<3)\\x(x\geqslant 3)\end{cases}\tag{2}$$

Hardswish 函数在输入为负时，使神经元不会像 ReLU6 那样出现“死亡”，仍能发挥作用，可以减少信息损失。而且计算量较小，在轻量级网络中具有出色的泛化能力和表达能力。

针对 MobileNetV2 存在的特征提取单一、感兴趣区域分散、微小特征提取较难等不足，在 MobileNetV2 瓶颈结构的基础上，引入 CBAM 模块，以提升模型识别准确度。CBAM 的优势在于：1)它同时包含了通道注意力和空间注意力，不仅可以跨通道捕获信息，还可以从空间方向捕获信息。通过嵌入 CBAM 模块，模型能够提取到更多关键有效信息，忽略次要无效信息，提高识别准确度；2)它非常轻量化，只增加少量计算量，不会给模型带来过多负担。CBAM 模块具体操作可以分为通道注意力和空间注意力两个步骤，其原理如图 2 所示。

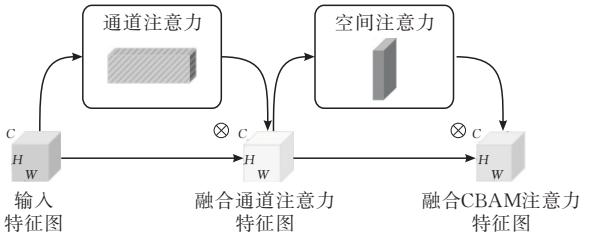


图 2　CBAM 模块原理

通道注意力机制的原理如图 3 所示。

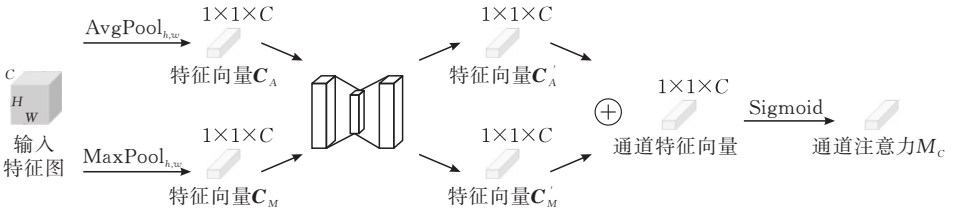


图 3　通道注意力机制原理

假设输入到 CBAM 注意力模块的特征图 F 的尺寸为 $H\times W\times C$，对每个通道所对应的 $H\times W$ 空间特征分别进行平均池化和最大池化运算，得到两个尺寸为 $1\times 1\times C$ 的特征向量 $\boldsymbol{C}_A$、$\boldsymbol{C}_M$；$\boldsymbol{C}_A$、$\boldsymbol{C}_M$ 经过多层感知机计算，得到相应的特征向量 $\boldsymbol{C}'_A$、$\boldsymbol{C}'_M$；$\boldsymbol{C}'_A$ 和 $\boldsymbol{C}'_M$ 进行逐元素相加，得到尺寸为 $1\times 1\times \boldsymbol{C}$ 的通道特征向量；通道特征向量再进行 Sigmoid 运算，得到通道注意力映射 M_C，计算过程如：

$$M_c=\sigma(\mathrm{MLP}(\mathrm{Avgpool}(F))+\mathrm{MLP}(\mathrm{Maxpool}(F)))$$

其中，F 为输入特征图，AvgPool、MaxPool 分别表示通道上特征图的平均池化操作和最大池化操作，MLP 为多层感知机前向计算函数，σ 为 Sigmoid 函数。特征图 F 与通道注意力映射 M_C 按元素相乘，得到经过通道注意力增强后的特征图 F_C，计算过程如下：

$$F_c=M_c\otimes F$$

式中，$\otimes$ 表示逐元素相乘操作。

空间注意力机制的原理如图 4 所示。

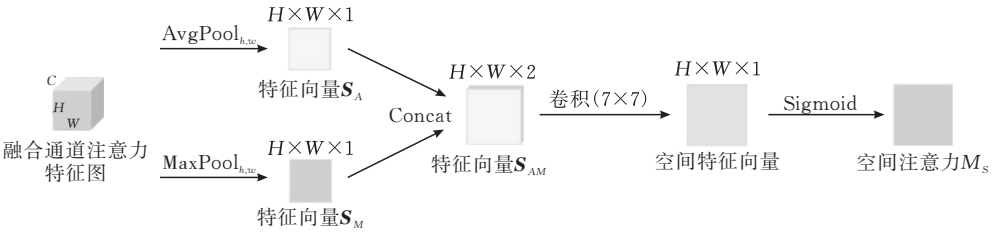


图 4 空间注意力机制原理

假设经过通道注意力增强后的特征图 F_C 进一步进行空间注意力增强，假设特征图 F_C 的尺寸为 $H\times W\times C$，对每个 1×1 空间所对应的所有通道特征分别进行平均池化和最大池化运算，得到两个尺寸为 $H\times W\times 1$ 的特征向量 S_A、S_M；将 S_A、S_M 拼接起来生成一个尺寸为 $H\times W\times 2$ 的特征向量 S_{AM}；S_{AM} 进行一个 7×7 卷积运算，得到尺寸为 $H\times W\times 1$ 的空间特征向量；空间特征向量再进行 Sigmoid 运算，得到最终的空间注意力映射：

$$M_s=\sigma(f_{7\times 7}(f_c(\mathrm{AvgPool}(F_C),\mathrm{MaxPool}(F_C))))$$

其中，f_c 为特征图通道拼接操作，$f_{7\times 7}$ 为 7×7 卷积运算。特征图 F_C 与空间注意力映射 M_s 按元素相乘得到特征图：

$$F_{s2}=M_s\otimes F_c$$

式中，$\otimes$ 表示逐元素相乘操作。特征图 F_{s2} 即为输入特征图 F 经过某 CBAMBneck_S2 模块处理后所得到的新特征图。

2 轻量级识别模型 LAAMNet

本研究以新的瓶颈结构 CBAMBneck 为基础模块，参照 MobileNetV2 网络的搭建思路，建立了一个轻量级的羊身份识别模型 LAAMNet。LAAMNet 的内部参数如表 1 所示，网络结构如图 5 所示。

表 1 LAAMNet 网络内部参数

Input	Operator	t	c	n	s
$224^2\times 3$	conv2d	—	32	1	2
$112^2\times 32$	CBAMBneck	1	16	1	1
$112^2\times 16$	CBAMBneck	6	24	2	2
$56^2\times 24$	CBAMBneck	6	32	3	2
$28^2\times 32$	CBAMBneck	6	64	4	2
$14^2\times 64$	CBAMBneck	6	96	3	1
$14^2\times 96$	CBAMBneck	6	160	3	2
$7^2\times 160$	CBAMBneck	6	320	1	1
$7^2\times 320$	conv2d，1×1	—	1280	1	1
$7^2\times 1280$	Avgpool 7×7	—	—	1	—
$1^2\times 1280$	conv2d，1×1	—	k	—	

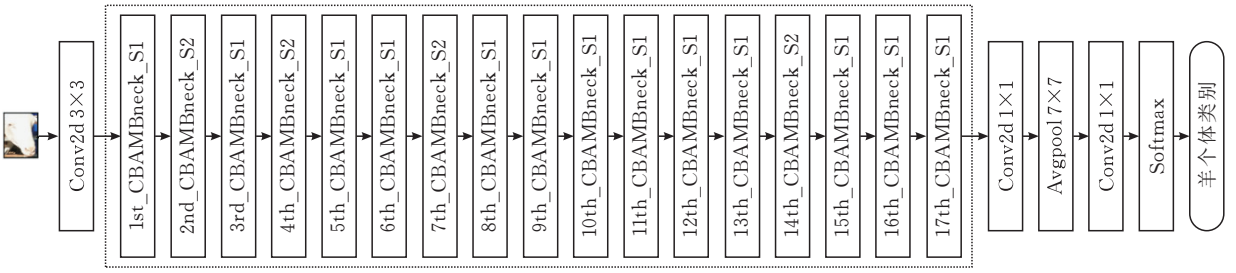


图 5 LAAMNet 网络结构

3 实验分析

3.1 羊脸数据集

本研究共选取 20 只羊，每次只选取一只羊，分别从距羊 1 m、1.5 m、2 m 的距离，水平 180°地移动摄像头，采集每只羊从左侧脸到正脸、再到右侧脸的羊脸视频数据；通过 PyCharm 平台将羊脸视频数据转换为 JPG 格式的羊脸图片数据；将得到的羊脸图片每隔 10 帧选取一张进行筛选；计算相邻两张图片的结构相似度指数(SSIM)，当 SSIM 的值大于 0.8 时，删除其中一张图片，只保留另一张图片，当 SSIM 的值小于或等于 0.8 时，两张图片都保留；采用随机裁剪和随机水平翻转进行数据增强，最终每只羊收集到 100 张图片，总共 2000 张羊脸图片；将 2000 张羊脸图片划分为正脸和侧脸，并分别从中抽取 9/10 的图片组成训练集，剩余的 1/10 的图片组成测试集，训练集和测试集的数据比例是 9∶1。部分羊脸数据见图 6。



图 6 部分羊脸数据集

3.2 实验环境

为了加快模型的训练速度，优化训练结果，本研

究使用了迁移学习技术，将 MobileNetV2 在 ImageNet 数据集上训练得到的预训练权重作为训练 LAAMNet 模型的初始化权重参数，在如表 2 所示规格的系统上进行训练和测试。

表 2　试验环境

Hardware	Software
CPU：Intel Xeon Silver 4116 CPU @2.1GHz	Ubuntu18.0.4
RAM：64.0 GB	CUDA10.2
GPU：NVIDIA Corporation GP102	PyTorch 1.8.1

模型训练参数：批量大小设置为 32，迭代次数为 200，初始学习率为 0.01，使用动量设置为 0.9 的随机梯度下降法，输入图像尺寸 224×224。

3.3　评价指标

为了客观地对改进模型的性能进行评价，本研究选取以下指标作为模型的综合评价指标。

1）准确率（Accuracy）：预测正确的样本占总样本的百分比

$$\text{Accuracy}=\frac{(T_P+T_N)}{(T_P+T_N+F_P+F_N)}\times 100\%$$

2）召回率（Recall）：实际为正的样本中被预测为正样本的概率

$$\text{Recall}=\frac{T_P}{(T_P+F_N)}\times 100\%$$

3）精确率（Precision）：所有预测为正的样本中实际为正样本的概率，如

$$\text{Precision}=\frac{T_P}{(T_P+F_P)}\times 100\%$$

式中：T_P 表示预测为正样本，实际也为正样本；T_N 表示预测为负样本，实际也为负样本；F_P 表示预测为正样本，实际为负样本；F_N 表示预测为负样本，实际为正样本。

3.4　实验结果

为了证明本研究采用的 Hardswish 激活函数策略相较于其他激活函数的优越性，比较了瓶颈结构使用不同激活函数的效果，在测试集上的试验结果如表 3 所示。

表 3　激活函数性能对比

model	Activation Function	Top-1 Accuracy/%	mean Recall/%	mean Precision/%
MobileNetV2	ReLU6	98.60	98.60	98.64
	Leaky ReLU	98.33	98.33	98.34
	ELU	97.50	97.50	97.53
	Mish	98.17	98.17	98.20
	Hardswish	98.67	98.67	98.69
	Hardswishstrategy	98.78	98.78	98.79

Hardswish 表示瓶颈结构特征扩展层和深度可分离卷积层均使用 Hardswish 激活函数；Hardswishstrategy 表示仅特征扩展层使用 Hardswish 激活函数，深度可分离卷积层仍使用 ReLU6 激活函数。

从表 3 可以看出，相比于 MobileNetV2 网络使用的激活函数 ReLU6，使用激活函数 Leaky ReLU、ELU、Mish 均会使测试集的准确率、召回率和精确率有一定程度的下降，而使用 Hardswish 作为激活函数则可使效果更好，这正是因为 Hardswish 激活函数能够有效避免神经元出现“死亡”，减少了信息损失。当瓶颈结构的特征扩展层和深度可分离卷积层均使用 Hardswish 激活函数时，准确率、召回率和精确率分别提升了 0.07、0.07 和 0.05 个百分点。而当使用 Hardswishstrategy 激活函数策略时，识别效果最好，测试集的准确率、召回率和精确率均提升最大，均达到了 0.18、0.18 和 0.15 个百分点。这是因为深度可分离卷积层到特征压缩层是从高维空间映射到低维空间，此时 ReLU6 激活函数不仅可以保持信息完整，不破坏信息表达，而且能将所需的复杂性引入到网络中，比其他激活函数更有效，故而效果更好。

可见，仅将 MobileNetV2 瓶颈结构特征扩展层的激活函数改成 Hardswish 激活函数，保留深度可分离卷积层的 ReLU6 激活函数不变，在不增加模型参数量的前提下，使模型的性能提升最大。

本研究为了使 CBAM 注意力机制发挥最大效能，探究了将 CBAM 模块嵌入到 MobileNetV2 瓶颈结构不同位置的表现，试验结果如表 4 所示。

从表 4 可以看出，与 MobileNetV2 瓶颈结构相比，CBAM_123 和 CBAM_23 的试验结果最差，不仅参数量和计算量最大，而且识别效果也最差，识别准确率明显大幅下降，模型已无法进行正确识别；CBAM_12、CBAM_13 和 CBAM_2 的识别准确率均略低于 MobileNetV2 瓶颈结构的识别准确率，模型的识别能力下降；只有 CBAM_1 和 CBAM_3 的识别准确率高于 MobileNetV2 瓶颈结构的识别准确率，模型的识别性能提升；而 CBAM_3 的参数量和计算量均少于 CBAM_1 的，且与 MobileNetV2 瓶颈结构均相差不大，因此，性能最好。与 MobileNetV2 瓶颈结构相比，CBAM_3 在测试集上的准确率提升了 0.29 个百分点，召回率提升了 0.29 个百分点，精确率提升了 0.25 个百分点，而模型的参数大

表 4 CBAM 注意力模块嵌入方式性能对比

model	CBAM introduction method	Top-1 Accuracy/%	mean Recall/%	mean Precision/%	Params size/MB	Total GFLOPs/M
MobileNetV2	—	98.60	98.60	98.64	8.58	652.465
	CBAM_1	98.83	98.83	98.87	17.19	671.379
	CBAM_2	98.44	98.44	98.47	17.19	666.051
	CBAM_3	98.89	98.89	98.89	9.02	658.296
	CBAM_12	98.06	98.06	98.07	25.81	684.965
	CBAM_13	96.83	96.83	96.90	17.63	677.211
	CBAM_23	7.67	7.67	7.70	17.63	671.883
	CBAM_123	5.00	5.00	0.25	26.25	690.797

“—”表示不嵌入 CBAM 模块，即 MobileNetV2 网络的瓶颈结构；CBAM_1 表示将 CBAM 模块嵌入到特征扩展层后面；CBAM_2 表示将 CBAM 模块嵌入到深度可分离卷积层后面；CBAM_3 表示将 CBAM 模块嵌入到特征压缩层后面；CBAM_12 表示同时在特征扩展层后面和深度可分离卷积层后面嵌入 CBAM 模块；CBAM_13 表示同时在特征扩展层后面和特征压缩层后面嵌入 CBAM 模块；CBAM_23 表示同时在深度可分离卷积层后面和特征压缩层后面嵌入 CBAM 模块；CBAM_123 表示同时在特征扩展层后面、深度可分离卷积层后面和特征压缩层后面嵌入 CBAM 模块。

小仅增加了 0.44 MB，总 GFLOPs 仅增加了 5.831 M。也就是说，注意力模块添加的数量和位置对模型的识别准确度均有一定的影响，并不是添加的注意力模块数量越多，识别准确率就越高；并且，将注意力模块添加在不同的位置也会对识别结果产生不同的影响。

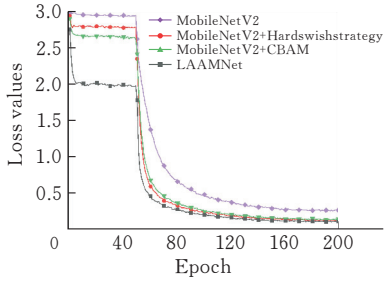
可见，在 MobileNetV2 瓶颈结构的特征压缩层后面嵌入 CBAM 模块，不仅使模型的识别效果最好，且增加的参数量最少，更有利于构建轻量级的羊脸识别模型。

为证明本研究对 MobileNetV2 网络进行的改进确实能够提升模型的性能，进行消融试验。消融试验结果如表 5 所示，迭代过程中，模型在训练集上的损失值和验证集上的准确率变化如图 7 所示。

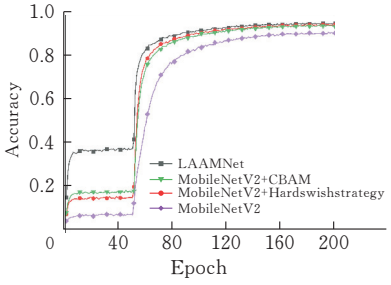
表 5 消融试验结果

model	Hardswish strategy	CBAM	Top-1 Accuracy/%	mean Recall/%	mean Precision/%	Params size/MB	Total GFLOPs/M
MobileNetV2	—	—	98.60	98.60	98.64	8.58	652.465
	√	—	98.78	98.78	98.79	8.58	652.465
	—	√	98.89	98.89	98.89	9.02	658.296
LAAMNet	√	√	99.00	99.00	99.01	9.02	658.296

“—”表示在改进 MobileNetV2 时不使用该策略，“√”表示在改进 MobileNetV2 时使用该策略。



(a)损失曲线



(b)准确度曲线

图 7 改进策略的模型迭代曲线

从表 5 可以看出，仅使用 Hardswishstrategy 激活函数策略将 MobileNetV2 瓶颈结构的激活函数替换后，模型的参数量和总的 GFLOPs 均保持不变，但准确率和召回率均提升了 0.18 个百分点，精确率提升了 0.15 个百分点；仅在 MobileNetV2 瓶颈结构中嵌入 CBAM 模块后，模型的准确率和召回率均提升了 0.29 个百分点，精确率提升了 0.25 个百分点，而参数量仅增加了 0.44MB，总的 GFLOPs 仅增加了 5.831M；使用 Hardswishstrategy 激活函数策略并嵌入 CBAM 模块，构建新的瓶颈结构 CBAM-Bneck 搭建模型 LAAMNet 后，模型的识别效果最好，准确率、召回率和精确率提升幅度最大，分别提升了 0.40、0.40 和 0.37 个百分点。由此说明，融合了两种改进措施的模型的性能优于只单独采纳一种改进措施的模型性能；并且，两种改进措施能够进行有效的融合，从而进一步提升了模型的识别性能。本研究所采取的改进措施是非常有效而且必要的。

从图 7 可以看出，训练过程中，在 MobileNetV2 瓶颈结构采用 Hardswishstrategy 激活函数策略或者嵌入 CBAM 注意力机制模块后，损失值均有一定程度的下降，识别准确率均逐渐上升；而使用 Hardswishstrategy 激活函数策略以及嵌入 CBAM 注意力机制模块后，损失值减少最多，在验证集上的识别准确率也提高最多。说明本研究综合上述两种改进措施所搭建的模型 LAAMNet 具有较好的性能，进一步证明了本研究所提出的一系列方法的有效性。

为了充分评估本研究搭建的羊身份识别模型 LAAMNet 的性能，与经典的深度学习网络模型 ResNet50，VGG16 进行对比试验，测试结果如表 6 所示。

表 6　各模型性能对比

model	Params size/MB	Total GFLOPs	Top-1 Accuracy/%	mean Recall/%	mean Precision/%
ResNet50	89.83	8.263 G	98.70	98.60	98.73
VGG16	512.48	30.933 G	98.10	98.10	98.14
LAAMNet	9.02	658.296 M	99.00	99.00	99.01

从表 6 可以看出，相比于 ResNet50 和 VGG16，LAAMNet 的参数规模和总的 GFLOPs 均有相当大幅度的缩减。LAAMNet 的参数规模约为 ResNet50 的 1/10，约为 VGG16 的 1/50；总的 GFLOPs 较 ResNet50 减少了约 92%，较 VGG16 减少了约 98%。说明本研究所建模型在模型规模方面相较于经典的深度模型，改进突出。并且，LAAMNet 的准确率、召回率和精确率均更高，相较于经典深度模型，其识别准确性更好。与 ResNet50 相比，LAAMNet 的准确率、召回率和精确率分别提高了 0.3、0.4 和 0.28 个百分点；与 VGG16 相比，LAAMNet 的准确率和召回率均提高了 0.9 个百分点，精确率提高了 0.87 个百分点。可见，本研究搭建的羊身份识别模型 LAAMNet 不仅更加轻量化，而且识别性能更优。

4　结论

本研究在轻量级网络 MobileNetV2 的基础上，将其瓶颈结构特征扩展层的 ReLU6 激活函数替换为 Hardswish 函数，并融合了轻量级 Convolutional Block Attention Module(CBAM)注意力模块，构建了性能更优的特征提取模块，既减少了网络的信息损失，又使网络更关注有效的关键信息，忽略无效信息，提高了模型的识别精度。实验结果表明，改进后的轻量级 LAAMNet 模型不仅规模更小，精度更高，验证了本研究改进方法的有效性。

后续对羊身份识别技术的研究还可在本文方法的基础上，进行如下三方面思考：

1)本研究致力于平衡模型识别精度和模型规模大小，后续可考虑兼顾模型识别精度、模型规模大小和模型推理速度，使模型在保持轻量化的基础上具有较快的推理速度，并能够进行准确识别。

2)本研究所建模型对数据集中已知的羊个体能够进行精确识别，准确度达到 99%，而对模型未曾接触过的个体是否也能够进行正确识别或者能够识别为“未知”类别值得探究和思考。

3)本研究提出的方法能够对羊个体进行准确识别，后续可继续探究该方法是否能够正确识别其他家畜，诸如猪、牛等。

[参　考　文　献]

[1] 陈琼. 城乡居民肉类消费研究[D].北京：中国农业科学院，2010.

[2] 刘月，李珍，邵丽玮，等.家庭羊场与规模羊场养殖成本构成比较及关键收益因素分析[J].北方牧业，2019，580(12)：20-21.

[3] 刘立安，白刚，张勇，等.智慧养殖在羊养殖产业中的应用效果[J].农业工程技术，2022，42(35)：68+70.

[4] 王旺虎，郭子记，李锦旗.智慧养殖在肉羊产业中的应用[J].畜牧兽医杂志，2021，40(04)：81-84.

[5] 张世龙. 基于深度学习的阿尔巴斯绒山羊面部识别研究[D].呼和浩特：内蒙古大学，2022.

[6] 邹小昱，尤佩华，刘思瑶，等. 种羊场 RFID 信息化管理系统研发[C]// 中国畜牧业协会.2013 中国羊业进展.当代畜牧杂志社，2013：203-206.

[7] 郑奎. 基于 RFID 与 ZigBee 的羊养殖信息管理嵌入式手持终端设计与实现[D].南京：南京农业大学，2013.

[8] 张颂. 羊面部图像自动采集系统的设计[D].呼和浩特：内蒙古农业大学，2022.

[9] 王荣，史再峰，高荣华，等.多变环境下基于多尺度卷积网络的猪个体识别[J].江西农业大学学报，2020，42(02)：391-400.

[10] 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(02)：12-17.

[11] 谢碧森，段清，刘俊晖，等.基于迁移学习的家猪图像识别研究[J].软件导刊，2020，19(07)：36-40.

[12] 吴梦茹，谢秋菊，李欣，等.基于深度学习的猪脸识别方法研究[J].智慧农业导刊，2021，1(09)：1-8.

[13] 姚礼垚. 基于深度学习的牛脸检测与精准识别[D].广

州:华南农业大学,2019.
[14] 黄俊华,田壮.基于改进 VGG 卷积神经网络的奶牛牛脸识别研究[J].湖北农机化,2019(13):126.
[15] 张宏鸣,周利香,李永恒,等.基于改进 MobileFaceNet 的羊脸识别方法[J].农业机械学报,2022,53(05):267-274.
[16] 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(03):47-50.
[17] 尚诚. 基于深度学习的羊只个体身份识别研究[D].杨凌:西北农林科技大学,2022.
[18] 周利香. 基于轻量化神经网络的羊脸识别方法研究[D].杨凌:西北农林科技大学,2022.
[19] 杨家林. 基于注意力机制的轻量级羊脸识别方法研究与实现[D].杨凌:西北农林科技大学,2022.
[20] SANDLER M , HOWARD A, ZHU , ML, et al . Mo-bileNetV2: inverted residuals and linear bottlenecks [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Salt Lake City, UT:IEEE,2018:4510-4520.
[21] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[J]. Lecture Notes in Computer Science, 2018,11211: 3-19.

A Lightweight Sheep Recognition Model Based on Improved MobileNetV2

GUO Yibei, WU Yingdan, SHAO Yanglin, XU Jiuhong, LI Yan
(*School of Science, Hubei Univ. of Tech.,Wuhan* 430068,*China*)

Abstract: In response to the problem that the sheep recognition model has a large number of parameters and is computationally intensive, which makes it difficult to be applied to actual sheep farms, a lightweight sheep recognition model LAAMNet is proposed based on MobileNetV2 with ancombiation of attention mechanism and optimized activation function. To improve the recognition accuracy, the activation function of MobileNetV2 bottleneck was modified and a lightweight attention mechanism was added to it; to lighten the model, both the expansion factor and the number of bottleneck of the MobileNetV2 were reduced. The experimental results show that compared with MobileNetV2, LAAMNet not only reduces the number of model parameters by about 40%, but also achieves an accuracy of 98.78% for the identification of individual sheep, an improvement of 0.18 percentage points. The model provides a better balance of recognition accuracy and complexity and provides ideas for the deployment in real sheep farms.

Keywords: sheep recognition; MobileNetV2; lightweight; Attention mechanism; Activation function

[责任编校:闫 品]