

VanillaFaceNet: 一种高精度快速推理的牛脸识别方法

栾浩天^{1,3}, 齐咏生^{1,2,3*}, 刘利强^{1,2,3}, 王朝霞^{1,2,3}, 李永亭^{1,2,3}

(1. 内蒙古工业大学电力学院, 呼和浩特 010051; 2. 大规模储能技术教育部工程研究中心, 呼和浩特 010080;
3. 内蒙古自治区高等学校智慧能源技术与装备工程研究中心, 呼和浩特 010080)

摘要: 快速精确定牛只身份对于牛只活体贷款, 改善牛只骗保等问题具有重要意义。针对不同牛只面部差异小, FaceNet 网络层数深, 推理速度较慢, 模型分类精度不足等问题, 该研究提出了基于 FaceNet 的牛脸识别方法-VanillaFaceNet。该方法首先将 FaceNet 的主干特征提取网络替换为极简网络 VanillaNet-13 并提出动态激活和增强型线性变换的激活函数两种方法提高网络的非线性; 然后, 提出一种新的 DBCA(dual-branch coordinate attention) 注意力模块, 能够更好地反映不同牛只面部特征之间的差异, 从而提高网络的识别精度; 最后, 针对 triplet loss 仅能减小牛只类间差异的问题, 采用 center-triplet loss 联合监督来减少牛只类内差异, 从而提高了相同牛只身份比对的准确性。基于自建的牛脸数据集对该模型进行训练和测试, 试验结果表明, VanillaFaceNet 对牛只识别的准确率达到 88.21%, 每秒传输帧数为 26.23 帧。与 FaceNet、MobileFaceNet、CenterFace、CosFace 和 ArcFace 算法相比, 本文算法的识别准确率分别提高了 2.99、9.58、6.26、3.85 和 4.49 个百分点, 推理速度分别提升了 2.67、0.77、0.10、1.28 和 0.94 帧/s。该模型对牛只有较为优秀的识别效果, 适于在嵌入式设备上部署, 实现了牛只面部识别精度和推理速度之间的平衡。

关键词: 识别; 特征; 提取; 牛脸; FaceNet; 注意力机制

doi: 10.11975/j.issn.1002-6819.202401011

中图分类号: TP391

文献标志码: A

文章编号: 1002-6819(2024)-18-0120-12

栾浩天, 齐咏生, 刘利强, 等. VanillaFaceNet: 一种高精度快速推理的牛脸识别方法[J]. 农业工程学报, 2024, 40(18): 120-131. doi: 10.11975/j.issn.1002-6819.202401011 <http://www.tcsae.org>

LUAN Haotian, QI Yongsheng, LIU Liqiang, et al. VanillaFaceNet: A high-precision and rapid inference for bovine face recognition[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(18): 120-131. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202401011 <http://www.tcsae.org>

0 引言

牛只面部识别技术作为生物特征识别领域的核心技术, 近年来得到了广泛关注和深入研究^[1-2]。传统的牛只身份识别方法通常采用耳标、项圈等物理标签、以及射频识别芯片^[3]。然而, 这些基于标签的方法容易引起牛只应激反应, 影响动物福利。此外, 由于射频识别芯片的频率差异, 导致每个芯片的识别范围各不相同, 可能导致错误识别。为了解决这些问题, 研究者们近年来开始探索如何利用计算机视觉技术实现对动物身份的非接触式识别^[4-5]。在计算机视觉技术早期的发展中, 主要依赖于传统的图像处理方法, 如边缘检测^[6]、特征提取^[7]、模式匹配^[8]等。以往的研究中, KUMAR 等^[9-10]采用综合各种特征提取、特征降维以及分类方法, 进行了牛脸识别的研究, 并比较了这些传统方法在牛脸识别中

的效果。ZHAO 等^[11]提出了一种用于荷斯坦奶牛图像提取和身份识别的视觉系统, 利用 FAST、SIFT 以及 FLANN 等方法进行特征提取、描述和匹配, 其最高识别率达到 96.72%。

然而, 随着深度学习和计算机视觉领域的进步, 牛只面部识别技术不断演进, 逐渐采用基于深度学习的方法, 卷积神经网络的兴起展现出了强大的特征提取能力, 能够更准确地捕捉和识别牛只面部特征, 实现更高层次的识别性能。这种非接触式^[12-15]的识别方式不仅提高了准确性, 还减少了对动物的干扰, 为农牧领域带来了新的可能性^[16-18]。以此为背景, ULADZISLAU^[19]采用 YOLOv7 检测和提取图像中的牛口部区域, 然后采用 OSNet 体系结构, 实现了牛只鼻纹识别, 试验表明, 所提出的方法有 92.8% 的准确率, 但参数量较多, 达到了 12 万, 并且采用鼻纹作为特征进行识别容易出现鼻纹被饲料遮挡等问题。XU 等^[20]提出了一种新的牛脸识别框架 CattleFaceNet, 结合轻量级的 RetinaFace-mobilenet 和 ArcFace, 并比较了三种损失函数在牛脸识别中的应用, 研究结果显示, CattleFaceNet 识别率为 91.3%, 但每秒传输帧数 (frames per second, FPS) 仅有 24 帧/s, 推理速度较慢。另一方面, LI 等^[21]引入了短路连接的 BasicBlock 以及注意力机制, 在奶牛识别中实现了 97.95% 的准确率。GAO 等^[22]通过检测跟踪形成牛只个体轨迹, 并通过高

收稿日期: 2024-01-02 修订日期: 2024-06-23

基金项目: 国家自然科学基金项目 (62363029); 内蒙古科技计划项目 (2020GG0283, 2021GG0256); 内蒙古自然科学基金项目 (2022MS06018, 2024QN06020); 呼和浩特市高校院所协同创新项目 (XTCX2023-16); 自治区直属高校基本科研业务费项目 (ZTY2024024)

作者简介: 栾浩天, 研究方向为计算机视觉、机器学习。

Email: 806069933@qq.com

*通信作者: 齐咏生, 教授, 研究方向为机器人协同控制技术。

Email: qys@imut.edu.cn

斯混合模型进行牛只身份分类, 其研究结果显示 Top-1 的准确率为 57.0%, Top-4 的准确率为 76.9%。HU 等^[23]利用融合特征训练的支持向量机分类器对奶牛个体进行识别, 其在包含 93 头奶牛侧视图的数据集上实现了 98.36% 的识别准确率。ANDREW 等^[24]提出了一种在开放的畜群环境中通过头顶成像自动检测、定位和识别个体动物的方法, 通过卷积神经网络和深度度量学习技术实现对生物特征进行自动检测和识别, 当仅对一半的牛只进行训练时, 准确率达到 93.8%。JIANG 等^[25]提出了 FLYOLOv3 网络, 旨在实现对复杂场景中奶牛关键部位的有效检测。测试结果显示 FLYOLOv3 算法的准确率达到 99.18%。

尽管上述深度卷积神经网络在识别性能上表现出色, 但它们普遍对硬件和计算资源有较高的要求, 导致推理速度过慢, 从而限制了在实际需要快速响应的应用场景中的可行性。为了解决这一问题, 本文在 FaceNet^[26] 网络结构的基础上采用 VanillaNet-13^[27] 降低网络层数, 并结合双分支坐标注意力机制 (dual-branch coordinate attention, DBCA), 提出了一种高识别准确率和低时间复杂度的牛脸识别模型, 该模型旨在为非接触式牛只个体精准识别提供参考。本文的方法采用开集识别, 即测试集中的牛只对模型是未知的。本文的主要贡献如下:

(1) 在主流目标识别方法中, 选择了 FaceNet 算法作为基本模型。通过在 FaceNet 中使用 VanillaNet-13 作为骨干网络, 提高了算法的推理速度, 使得训练好的模型能够应用于处理器性能有限的平台, 并使用动态激活和增强型线性变换的激活函数两种方法, 进一步提升网络的非线性性。

(2) 针对不同牛只面部特征差异小, 模型识别精度低的问题, 设计了一种 DBCA 注意力机制。通过引入基于坐标方向的全局最大池化和全局平均池化, 有效增强了算法模型对牛只显著性特征的提取能力。在处理细微差异的不同牛只面部特征时, 利用空间位置信息提高模型对目标区域定位的准确性。

(3) 由于 triplet loss 仅关注不同牛只间的离散性而忽略相同牛只内部的紧凑性, 因此引入了 center loss^[28] 来学习每个牛只身份的类中心, 使得同一身份的牛只特征之间的距离变得更加紧凑, 从而提高相同牛只身份比对准确性。

最后, 为验证算法的可移植性, 将算法部署到嵌入式平台 Jetson AGX Xavier 进行了测试和验证。

1 数据集制作

本数据集于 2023 年 3 月在内蒙古呼和浩特市欧太牧场采集, 如图 1a 所示, 在自然条件下, 随机选取 280 只成年牛只进行视频拍摄, 每头牛的视频时长为 8~20 s, 分辨率为 1 920×1 080 像素。利用 Open CV 对采集的视频流进行分帧处理, 将同一头牛的图像整理到一个文件夹下, 并手动剔除相似图像, 得到 16 800 张图片。

为了提升模型的泛化性能, 采用 4 种方法对牛脸数

据集进行扩充, 包括随机翻转、随机旋转、改变对比度以及添加椒盐噪声。扩充后的图像总数达到 50 000 张。按照 8:1:1 的比例, 将所有牛只个体划分为训练集、验证集和测试集。其中, 训练集中包含 40 000 张图片; 验证集中包含 5 000 张图片; 测试集中包含 5 000 张图片。为了实现开集的牛脸识别, 在训练集、验证集和测试集中引入不同的牛只个体。如图 1b 所示, 在进行模型训练之前, 进行数据预处理, 将原始图像大小统一调整为 160×160 像素, 并通过在图像边缘处添加灰条进行填充, 以防止失真。

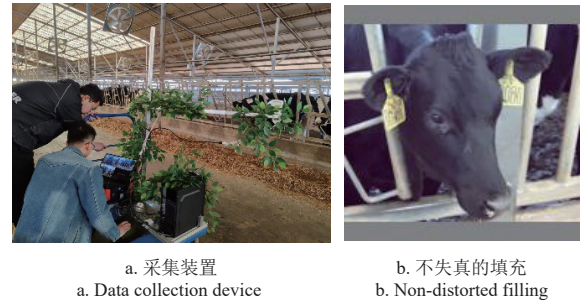


图 1 数据采集与数据预处理
Fig.1 Data collection and data preprocessing

2 牛脸识别模型

2.1 FaceNet 模型

FaceNet 模型结构图如图 2 所示。



注: Batch 表示输入的牛脸图像样本; DEEP ARCHITECTURE 表示深度卷积神经网络; L2 表示 L2 正则化; EMBEDDING 表示 L2 正则化后生成的特征向量; Triplet Loss 表示三元组损失函数。

Note: Batch represents the input cow face image samples; DEEP ARCHITECTURE represents the deep convolutional network; L2 represents L2 normalization; EMBEDDING represents the feature vectors generated after L2 normalization; Triplet Loss represents the triplet loss function.

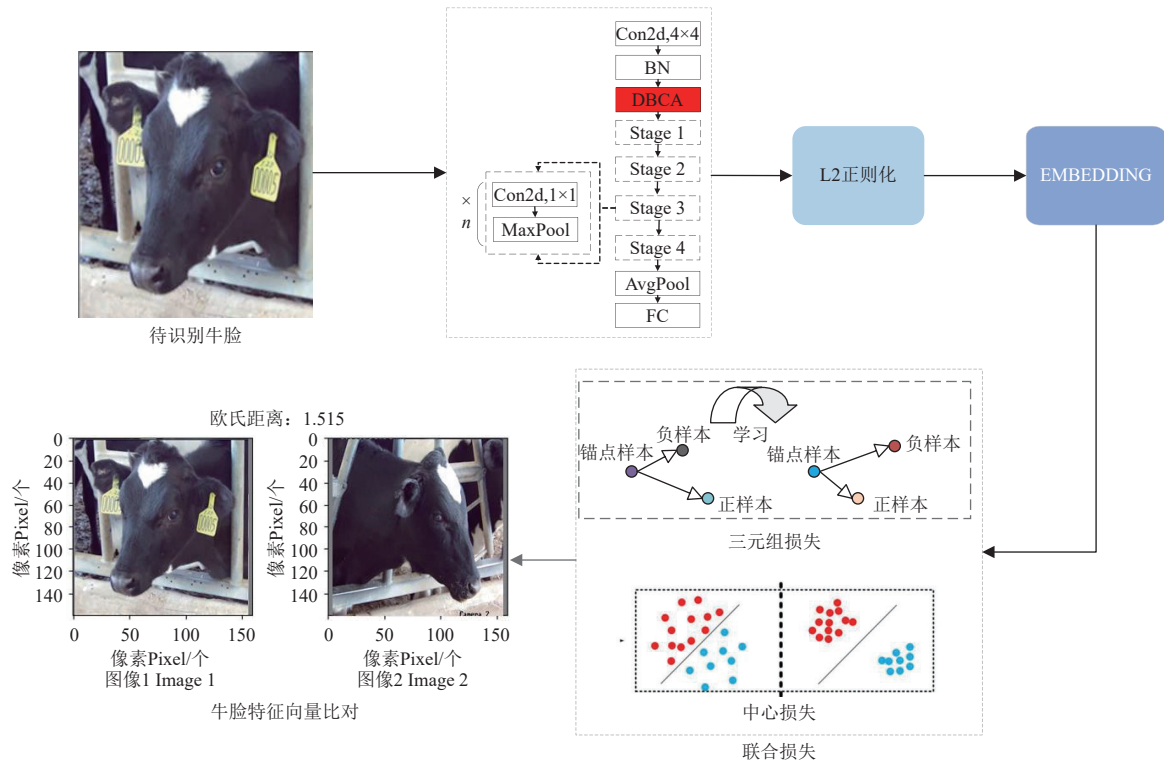
图 2 FaceNet 网络结构图
Fig.2 FaceNet network architecture diagram

将一张牛脸图像输入到 DEEP ARCHITECTURE 中, 经过深度卷积操作后生成一组牛脸数据, 之后对这组数据进行 L2 正则化处理 and embedding 操作, 将单个牛脸图像转换为在欧氏空间中的 128 维牛脸特征向量。通过比较两个特征向量之间的欧氏距离来度量两头牛之间的相似度。若两张牛脸图片对应的特征向量的欧氏距离小于设定的阈值, 则判定这两张牛脸图像属于同一头牛; 反之, 则认为它们不属于同一头牛。

2.2 基于 FaceNet 的牛脸识别模型

在本文中, 基于牛只面部图像的特征, 对 FaceNet 进行改进, 如图 3 所示, 改进算法包括三个部分: 1) 采用动态激活和增强型线性变换的激活函数增强 VanillaNet 主干网络的非线性能力; 2) 在主干网络的浅层加入一种 DBCA (dual-branch coordinate attention) 注意力机制, 增强算法对牛只显著性特征的提取; 3) 设计 triplet loss

和 center loss 联合监督训练以最大限度地提高牛只特征的类内紧凑性和类间可分性。



注: Conv2 d 表示普通卷积; BN 表示批量归一化; DBCA 表示本文提出的 DBCA 注意力机制; AvgPool 表示平均池化; MaxPool 表示最大池化; FC 表示全连接层; EMBEDDING 表示 L2 正则化后生成的特征向量。

Note: Conv2 d represents standard convolution; BN represents batch normalization; DBCA represents the DBCA attention mechanism proposed in this paper; AvgPool represents average pooling; MaxPool represents max pooling; FC represents fully connected layer; EMBEDDING represents the feature vectors generated after L2 regularization.

图3 VanillaFaceNet 整体架构图

Fig.3 Overall architecture diagram of VanillaFaceNet

2.2.1 主干特征提取网络

主干特征提取网络是 FaceNet 的主要组成部分, 输入图像尺寸为 $160 \times 160 \times 3$ 像素, 输出图像尺寸为 128×1 像素的特征向量。

FaceNet 最初的主干是 Inception-ResNetV1。鉴于在牛只特征提取中, 浅层信息的提取至关重要, 而多层卷积可能导致浅层信息的严重丢失, 另外, 多层卷积会降低算法推理速度, 特别是在算力受限的嵌入式设备上, 实时性表现较差。为了解决这些问题, 选择只有 13 层的 VanillaNet-13 作为牛只特征提取算法的主干网络。

VanillaNet-13 作为一种新的神经网络架构, 通过舍弃过多的深度、shortcut 等操作, 解决了模型复杂度的问题, 使其在资源有限的环境下表现出色。VanillaNet-13 的网络结构如表 1 所示。

由于 VanillaNet-13 网络结构简单, 导致了网络的非线性表达能力较差。为解决这一问题, 提出以下解决方案:

1) 动态激活

在网络训练过程中, 通过训练两个带有激活函数的卷积层, 两个卷积可合并到一个卷积中, 从而有效地减少推理时间, 其中激活函数随训练次数的增加逐渐减少为恒等映射。在训练期间充分利用激活函数的表达能力,

而在推理阶段又通过合并卷积层来提高网络推理速度:

$$A'(x) = (1 - \lambda)A(x) + \lambda x \quad (1)$$

式中 $A(x)$ 表示普通激活函数, λ 是一个超参数, 用来平衡修改后的激活函数 $A'(x)$ 的非线性。将当前训练的 epoch 和总的 epoch 分别表示为 e 和 E , 则 $\lambda = \frac{e}{E}$ 。在开始训练时, $e = \lambda = 0$, $A'(x) = A(x)$, 此时网络有着很强的非线性。训练过程中, λ 随着训练的迭代进行线性变化, 使得网络在训练过程中 λ 的变化更加平滑。当训练结束时, $e = E$, 即 $\lambda = 1$, $A'(x) = x$, 此时两个卷积中间不具有非线性激活的能力, 因此训练结束时两个卷积合并为一个卷积, 大大减少了网络的推理时间。

表 1 主干特征提取模块结构

Table 1 Backbone feature extraction module structure

模块结构 Module structure	输入 Input	VanillaNet-13
stem	160×160	$4 \times 4, 512, \text{stride } 4$
dual-branch coordinate attention(DBCA)	40×40	-
阶段 1 Stage1	40×40	$[1 \times 1, 1, 024] \times 2$ MaxPool 2×2
阶段 2 Stage2	20×20	$[1 \times 1, 2048] \times 2$ MaxPool 2×2
阶段 3 Stage3	10×10	$[1 \times 1, 4, 096] \times 7$ MaxPool 2×2
阶段 4 Stage4	5×5	$[1 \times 1, 4, 096] \times 1$
分类器 Classsifier	5×5	AvgPool 7×7 $1 \times 1, 1, 000$

2) 增强型线性变换的的激活函数

提高网络的非线性通常采用激活函数的串行堆叠。然而, 激活函数的串行堆叠导致并行计算能力过剩时的高延迟。因此, 本文转向采用激活函数的并行堆叠。对于神经网络中输入 x 的单个激活函数表示为 $A(x)$, 其并行堆叠可以表示为

$$A_s(x) = \sum_{i=1}^n a_i A(x + b_i) \quad (2)$$

式中 n 表示堆叠激活函数的数量, a_i 和 b_i 分别是每个激活的规模和偏置, 以避免简单的累积。通过并发堆叠, 能够显著增强激活函数的非线性。

此外, 现有激活函数过于简化了对输入像素空间无关的假设, 在独立假设下, 难以准确估计复杂环境和多目标的输入分布。激活函数中的空间不敏感是限制牛脸识别任务显著改善的主要障碍, 因此, 采用线性变换在激活函数中嵌入空间上下文信息, 以提高模型对复杂特征分布的拟合能力。通过聚合每个激活函数输入的邻域实现了对全局特征的学习。给定一个输入特征 $x \in R^{H \times W \times C}$, 其中 H 、 W 和 C 是高、宽和通道数, 考虑上下文信息和局部依赖关系的激活函数的广义公式为

$$A_s(x_{h,w,c}) = \sum_{i,j \in \{-n,n\}} a_{i,j,c} A(x_{i+h,j+w,c} + b_c) \quad (3)$$

式中 $a_{i,j,c}$ 和 b_c 分别代表激活函数的规模和偏置, x 代表神经网络的输入, 其中 $h \in \{1, 2, \dots, H\}$, $w \in \{1, 2, \dots, W\}$, $c \in \{1, 2, \dots, C\}$ 。通过引入极小的空间开销, 将激活函数扩展为二维激活函数。当 $n=0$ 时, 基于增强线性变换的激活函数 $A_s(x)$ 退化为普通激活函数 $A(x)$, 这一方法可以被视为对现有激活函数的一般扩展。

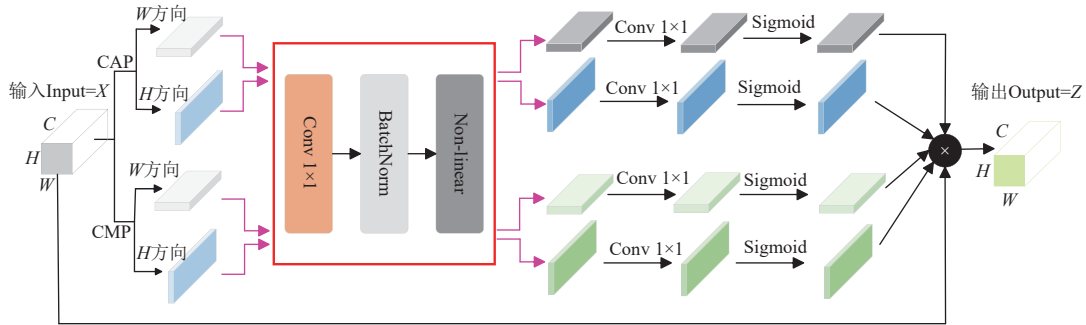
2.2.2 DBCA 注意力模块

注意机制通过关注重要特征并抑制不必要特征来增强现有网络的表示能力。在许多计算机视觉任务中, 如图像分类^[29], 图像分割^[30-31]和目标检测^[32-34], 注意力机制已被证明能够提升模型性能。

在基于卷积神经网络的注意力机制中, 第一步通常是对特征信息进行聚合, 其中最常用的操作是全局平均池化。然而, 全局平均池化存在两个主要缺点: 1) 全局信息是平均的, 导致忽略了显著特征; 2) 只能对一维信息(通道信息或空间信息)进行编码。这些限制可能影响模型对复杂场景和多层次特征的有效捕捉。因此, 在注意力机制的设计中, 需要综合考虑这些因素, 以更全面地提升网络的感知能力和性能表现。

受 CBAM (convolutional block attention module) ^[35] 注意力机制的启发, 提出一种新型的 DBCA 注意力机制。DBCA 注意力机制分别进行了基于坐标方向的平均池化和最大池化, 本文将分别称为坐标平均池化 (coordinate average pooling, CAP) 和坐标最大池化 (coordinate max pooling, CMP)。文中将全局平均池化和全局最大池化分解为一维特征编码, 并分别沿水平和垂直方向对每个通道的特征图进行编码, 从而生成与方向相关的特征图。通过引入了基于坐标方向的最大池化, DBCA 能够聚合牛脸特征的显著信息, 更好地反映牛脸特征之间的差异。

DBCA 模块如图 4 所示。生成注意图的计算过程包括坐标信息聚合、跨维交互和注意力生成三个步骤。



注: C 、 H 、 W 分别表示特征的通道数、高和宽; CAP 表示坐标平均池化、CMP 表示坐标最大池化; Conv 1×1 表示 1×1 卷积; BatchNorm 表示批量归一化; Non-linear 表示激活函数; Sigmoid 表示 Sigmoid 激活函数。

Note: C , H , and W represent the number of channels, height, and width of the features respectively; CAP represents Coordinate Average Pooling; CMP represents Coordinate Max Pooling; Conv 1×1 represents 1×1 convolution; BatchNorm represents batch normalization; Non-linear represents the activation function; Sigmoid represents the Sigmoid activation function.

图 4 DBCA 注意力机制结构图

Fig.4 DBCA (dual-branch coordinate attention) attention mechanism structure diagram

1) 坐标信息聚合

给定一个中间特征映射 X 作为输入特征图, 沿水平坐标和垂直坐标分别进行平均池化和最大池化生成聚合特征。CAP 在高度方向和宽度方向的聚合特征公式如下:

$$F_h^a = p_h^a(X) = \frac{1}{W} \sum_{j=1}^W X_j \quad (4)$$

$$F_w^a = p_w^a(X) = \frac{1}{H} \sum_{i=1}^H X_i \quad (5)$$

式中 X_j , X_i 是 DBCA 注意力机制的输入特征图, $p_h^a(\cdot)$ 和 $p_w^a(\cdot)$ 分别表示基于高度方向和基于宽度方向的 CAP, $F_h^a \in R^{H \times 1 \times C}$, $F_w^a \in R^{1 \times W \times C}$ 是对高度方向和宽度方向坐标信息进行编码得到的两组聚合特征。同理, 基于 CMP 在高度方向和宽度方向的聚合特征为

$$F_h^m = p_h^m(X) = \max_w(X) \quad (6)$$

$$F_w^m = p_w^m(X) = \max_h(X) \quad (7)$$

式中 $p_h^m(\cdot)$ 和 $p_w^m(\cdot)$ 分别是基于高度方向和宽度方向的 CMP, 将高度方向和宽度方向的坐标信息嵌入到两组聚合特征点 $F_h^m \in R^{H \times 1 \times C}$, $F_w^m \in R^{1 \times W \times C}$ 。

通过对不同方向的特征进行聚合, 使算法更加关注浅层信息的位置。在获取每个方向的权值后, 将其与两侧的输入结合, 使输出特征图的浅层信息得到更好的表达。由于不同牛只面部之间的差异可能非常微妙, 因此本文引入了最大池化的概念, 通过在两个坐标方向上对显著特征信息进行聚合和精确位置信息的编码来实现长期依赖建模, 有效避免了由于单一聚合操作导致的特征表示不足的问题, 这有助于区分相似的目标。相较于只使用平均池化, 最大池化能够更好地捕捉差异最大的像素。在坐标信息聚合的步骤中, CAP 和 CMP 分别沿两个坐标方向对特征进行聚合, 不仅保留了空间位置信息, 还嵌入了特征图的通道信息。

2) 跨维交互

在获得 4 个方向的注意力图后, 将所得特征分别进行合并, 这有助于捕获跨通道信息并保持精确的位置信息。之后并通过 1×1 的共享卷积层传递, 使得四组聚合特征的相互作用相对独立, 从而有效地避免了互相干扰的情况。 1×1 卷积块将由 F_h^a 、 F_w^a 、 F_h^m 、 F_w^m 共享, 因此跨维交互操作可表示为

$$(T_h^a, T_w^a) = \delta(V_0 * (F_h^a, F_w^a)) \quad (8)$$

$$(T_h^m, T_w^m) = \delta(V_0 * (F_h^m, F_w^m)) \quad (9)$$

式中 $*$ 代表卷积操作, (\cdot, \cdot) 表示 concat 操作, V_0 表示共享 1×1 卷积, δ 表示非线性操作。跨维卷积交互首先通过 concat 操作进行堆叠, 然后通过一个共享 1×1 卷积核 $V_0 \in R^{1 \times 1 \times C \times C / r}$ 进行降维, r 用来控制块的大小比例。本文对中间特征进行批量归一化和非线性激活操作, 并通过分裂方法最终得到 4 个特征变换后的输出: $T_h^a, T_w^a \in R^{H \times 1 \times C}$ 和 $T_h^m, T_w^m \in R^{1 \times W \times C}$ 。利用 4 组聚合特征进行特征变换, 避免了两个坐标方向信息的相互干扰。

3) 注意力生成

从式 (8)~(9) 中得到的经过转换的特征被用于生成注意力图。在此步骤之前, 聚合的特征是相互独立进行编码和交互的。在这一步中, 通过利用 1×1 卷积操作和 sigmoid 函数, 将这些特征转换为具有与输入相同数量的通道。整个流程描述如下:

$$Y_h^a = \sigma(V_1 * (T_h^a)) \quad (10)$$

$$Y_w^a = \sigma(V_1 * (T_w^a)) \quad (11)$$

$$Y_h^m = \sigma(V_1 * (T_h^m)) \quad (12)$$

$$Y_w^m = \sigma(V_1 * (T_w^m)) \quad (13)$$

式中 σ 表示 sigmoid 函数, V_1 表示 1×1 卷积。以上是

DBCA 块生成 4 组坐标关注的整个过程, DBCA 块的输出可以表示为

$$Z = (f_{co}(f_{co}(f_{co}(X, Y_h^a), Y_w^a), Y_h^m), Y_w^m)) \quad (14)$$

式中 $f_{co}(\cdot, \cdot)$ 为坐标乘法, Z 表示经过 DBCA 块重新标定的输出特征图。可以看出, 两组注意力图是并行计算的, 并同步校准特征图, 这种并行计算设计使得四组聚合特征共享跨维交互功能, 从而大大减少了参数量。

根据式 (16)~(17), 由 DBCA 块生成的注意力图以输入为条件, 双分支注意力图中一组包含平均信息, 另一组包含显著信息, 都是基于输入特征图进行编码的。由于每组信息的每个元素对应一个高度坐标或宽度坐标, 因此能够显著提高输入特征的识别程度, 这些元素反映了目标对象所在的行或列。最后, 将得到的注意力映射以互补的方式应用于输入特征。本文提取的 DBCA 注意力机制同时保留了全局特征的获取, 兼顾了局部特征的捕获, 增强了浅层信息的特征提取。整个流程使 DBCA 块能够实现对于牛只的精确识别。

DBCA 作为一种可以同时获取通道信息和位置信息的注意力机制, 可以很容易地嵌入到网络模型中, 以提高其性能。在输入二维数据时, 注意力机制通过不同维度的一维池化核将这些输入参数转换为一维输出, 后续所有计算都是一维运算。因此, 与整体神经网络模型相比, DBCA 不会增加太多的计算资源。同时, DBCA 使模型能够同时考虑每个空间位置周围的上下文信息, 避免了一些干扰信息, 使模型能够更准确地定位目标目标, 获得更有价值的特征信息。

2.2.3 损失函数

FaceNet 最初采用 triplet loss 损失函数来监督训练网络, 以获取具有较强类间判别力的深度特征。然而, 在实际应用中, 不同牛只之间的特征比对相对容易, 而相同牛只的身份对比则相对困难。这是因为不同身份牛只面部的特征区分度较为明显, 但相同牛只面部的特征未能紧凑地聚集在一起。

针对复杂分布的牛脸数据, 本文既追求数据在特征空间中的良好类间可分性, 更重要的是希望数据保持类内的紧凑性。因为同一头牛的类内变化很可能大于类间变化, 保持类内紧凑性对于样本判别至关重要。

为此, 本文提出一种基于 center loss 和 triplet loss 联合监督的算法, 以增强类内相似性。在本文的方法中, center loss 用于进行相同牛只特征的深度聚类, 通过学习每个类的类中心, 促使类内距离更加紧凑。通过巧妙结合 triplet loss 和 center loss 的优势, 有效地最小化了牛只特征的类内距离, 同时最大化牛只特征的类间距离, 从而提高了相同牛只身份比对的准确性。模型损失 L 为:

$$L = L_{triplet} + \lambda L_{center} \quad (15)$$

式中 $L_{triplet}$ 为 triplet loss, L_{center} 为 center loss, λ 取 0.005。联合损失函数的目标在于有效地将特征分布到各个类别, 而基于距离的惩罚旨在减少类别内的方差。这一设计有助于提高类内紧凑性, 同时增加了类间样本的可分离性。

triplet loss 增强了欧式空间中的类间可分性。triplet loss 的计算涉及选择多个图像实例, 其中每个实例包含一个锚点样本 (Anchor) 以及来自同一类的正样本 (Positive) 和来自不同类别的负样本 (Negative), 如图 5 所示。

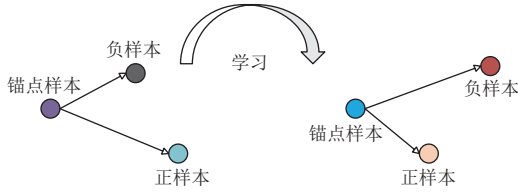


图 5 三元组损失函数
Fig.5 triplet loss function

然而, 由于 triplet loss 无法提供全局最优约束, 导致在某些情况下, 类间距离小于类内距离。triplet loss 的计算公式如下:

$$L_{triplet} = \sum_{i=1}^N [\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha, 0]_+ \quad (16)$$

式中 $\|\cdot\|$ 为欧氏距离, $\|f(x_i^a) - f(x_i^p)\|_2^2$ 表示锚点样本和正样本之间的欧氏距离, $\|f(x_i^a) - f(x_i^n)\|_2^2$ 表示锚点样本和负样本之间的欧氏距离。 α 为 triplet loss 余量, 本文将 α 设为 0.3。triplet loss 只考虑 $\|f(x_i^a) - f(x_i^p)\|_2^2$ 和 $\|f(x_i^a) - f(x_i^n)\|_2^2$ 的差值, 而忽略它们的绝对值。通过 center loss, 致力于在每个训练批次内最小化不同领域特征与其对应中心之间的距离。在每次训练迭代后, 每个类别中心 c_{y_j} 将按标签的功能更新。随着类别中心 c_{y_j} 的更新, 可以缩短同一类别的特征与批处理中相同牛只特征之间的距离。

center loss 的主要思想在于为每一类别的深度特征学习一个中心, 并通过对深度特征与其相应的类别中心之间的距离进行惩罚, 以补偿 triplet loss 的不足。这样的设计旨在强制每个数据的特征表示更接近特征空间中对应的类中心, 从而促使整体类级分布在特征空间中更加紧凑。对于大规模牛脸数据集中的频繁样本类, 即使图像来自同一类, 图像在视觉外观上也可能存在较大的差异。这可能导致频繁样本类在特征空间中相对较大的区域中扩散, 即锚点样本和正样本在特征空间中可以相距很远。为了处理这一问题并提高训练稳定性, 涉及类中心的 triplet loss 被引入。center loss 表示如下:

$$L_{center} = \frac{1}{2} \sum_{j=1}^B \|f_{i_j} - c_{y_j}\|_2^2 \quad (17)$$

式中 L_{center} 代表 center loss, f_{i_j} 代表所有特征, c_{y_j} 表示类别为 y_j 的所有特征中心。 B 是 batch size 的数量。更具体地说, 在训练过程中, 在每个类别上以 minibatch 为尺度进行统计, 得到该类别特征的中心, 目标是使得所有特征到其对应中心的距离尽可能小。

center loss 目标是通过通过对每个类别的样本与其对应类别样本中心之间的偏移进行惩罚, 从而使学习到的特征具有更好的泛化性和辨别能力。该方法旨在让同一类别的样本尽可能地聚合在一起, 提高特征的紧凑性和可分

离性。

采用联合监督方案时, 可以同时最大限度地提高样本的紧凑性和可分离性。此外, 所提出的两种算法都充分利用整个训练集进行分析学习类中心, 以捕获深度特征空间的全局信息。这一策略有助于减少类内变异, 并将特征分布限制在其类中心附近。通过这样的联合训练方案, 我们能够更好地优化特征表示, 提高模型性能。

2.3 评价指标

本文使用准确率 ACC (accuracy)、验证率 VR (validation rate)、每秒传输帧数 (frames per second, FPS)、延迟 (latency) 作为评价指标衡量模型的性能。ACC 和 VR 计算公式如下:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (18)$$

$$VR = \frac{TP}{TP + FN} \times 100\% \quad (19)$$

式中 TP、TN、FP、FN 分别表示模型的正确例、真反例、假正例、假反例。ACC 是指模型在总样本中正确判断同一牛脸样本与正确判断不同牛脸样本在总样本中所占的比例; VR 则表示模型在验证集中正确地将属于目标身份的牛脸识别为正例的比例, 即在所有同一牛脸样本中正确判断同一牛脸样本的比例; FPS 指每秒传输的图像帧数; Latency 指网络预测一张图片所需要的时间。

在进行模型测试时, 使用已经训练好的 VanillaFaceNet 网络处理正负牛脸图像对, 获取相应的特征向量。为度量两个向量之间的距离, 我们采用欧几里得距离, 并将其与最优阈值进行比较, 以确定图像对是否属于同一头牛。

为了确定最佳阈值, 设定阈值范围为 [0,2], 每 0.01 间隔取一个值作为一个阈值。对于每个阈值, 计算识别精度, 并选择使得识别精度最大的阈值作为最佳阈值。最终, 确定测试集的最佳阈值为 0.95。

在模型测试阶段, 特征提取模型得到两张未知牛只面部图像的特征向量。然后, 计算这两个特征向量之间的欧式距离。最后, 通过与最佳阈值 0.95 比较, 确定两张牛脸图像是否属于同一头牛。

3 试验与分析

本研究选择台式计算机作为模型的训练平台, 其中 CPU 采用了 Intel i7-11 700; GPU 为 NVIDIA RTX 3 090 显卡、运行内存 24 G; 操作系统为 Ubuntu 18.04; CUDA 版本 11.0; Cudnn8.1.0 加速; PyTorch 版本 1.11.0; Python 环境 3.6。

为了验证本文提出的改进 FaceNet 算法, 使用自制的牛脸数据集进行了试验。为了公平比较, 本文使用权重衰减为 0.0005 的 SGD 训练了 100 个 epoch 的所有模型。初始学习率设置为 0.001, 当连续 5 个 epoch 损失函数未减小时, 进行学习率衰减, 学习率最小为 $1e-6$, 批大小设置为 96。之后, 将训练好的模型移植到嵌入式主板 Jetson AGX Xavier 进行测试。

3.1 训练结果

VanillaFaceNet 模型在自制牛脸数据集的性能如图 6a 和图 6b 所示。在模型训练过程中, 本文使用欧式距离作为度量方式, 并保存准确率最高的模型权重。随着训练次数的增加, 损失值一开始呈下降趋势, 然后逐渐趋于稳定。在模型的前 60 个 epoch 中, 损失值波动较大; 鉴于本文采用了分阶段学习技术, 模型在 60 个 epoch 后的损失值逐渐趋于稳定。在第 5 个 epoch 和第 60 个 epoch 之间, 学习率下降到原来的十分之一, 从而加速了模型的收敛。

ROC 曲线用于展示不同阈值下的真正例率 (true positive rate) 与假正例率 (false positive rate) 之间的权衡关系。通过改变判定为正例的阈值, 在 ROC 曲线上得到不同的点, 从而形成曲线。理想情况下, ROC 曲线越靠近左上角, 说明模型在不同阈值下表现越好, 即真正例率高且假正例率低。如图 6b 所示:

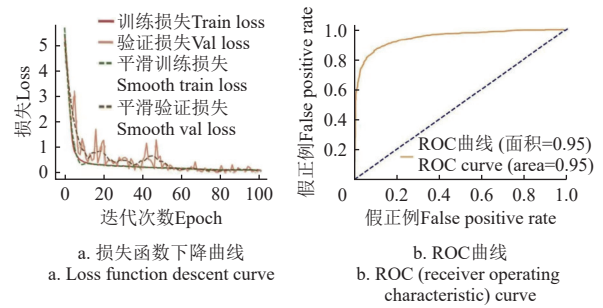


图 6 VanillaFaceNet 的性能曲线

Fig.6 Performance curve of VanillaFaceNet

3.2 消融试验

为验证本文算法改进的有效性, 首先在自制的牛脸数据集上进行消融试验。通过替换 FaceNet 的主干网络、去除 DBCA 注意力机制与 center loss 来比较算法对准确率的影响。消融试验结果如表 2 所示:

表 2 消融试验

Table 2 Ablation experiment

模型名称 Model name	准确率 Accuracy/%	验证率 Validation rate/%	每秒传输帧数 Frames per second FPS/(帧·s ⁻¹)	延迟 Latency/ms
InceptionResNetv1	85.22	82.59	23.56	42.44
InceptionResNetv1+CA	85.16	79.48	23.93	41.79
InceptionResNetv1+DBCA	85.63	83.01	23.90	41.84
InceptionResNetv1+ center loss	86.28	84.46	24.16	41.39
InceptionResNetv1+DBCA+center loss	86.54	84.84	24.24	41.25
VanillaNet	86.75	84.05	25.93	38.57
VanillaNet+CA	87.27	85.98	25.73	38.87
VanillaNet+DBCA	87.39	85.82	25.11	39.82
VanillaNet+center loss	87.60	86.90	25.89	38.62
VanillaNet+DBCA+center loss	88.21	87.41	26.23	28.12

由表 2 中消融试验结果可以看出, 主干网络替换为 VanillaNet 之后, 提高了 FaceNet 的推理速度; 在网络增加 DBCA 注意力机制和 center loss 后, 对于牛只面部识别的精度有了明显提高。综上, 与基准模型相比, 三种改进措施均提高了模型的识别准确率, 最终识别准确率达到 88.21%, 相比于原始 FaceNet 提高了 2.99 个百分点, FPS 达到了 26.23 帧/s, 相比于原始 FaceNet 提高了 2.67 帧/s, 验证本文算法的有效性和先进性。

3.3 算法对比试验

本文选择 MobileFaceNet^[36]、CenterFace^[37]、CosFace^[38] 和 ArcFace^[39] 经典识别算法进行对比试验, 进而验证所提算法的有效性。MobileFaceNet 采用轻量级的网络结构, 设计用于在资源受限的移动设备上执行人脸识别。CenterFace 是一种准确且高效的人脸检测和识别算法, 中心性对齐机制是该算法的一个重要特点, 它有助于提高检测准确性。CosFace 通过引入余弦相似度损失函数, 使得在特征空间中的类别之间的角度更大。这有助于提高模型对不同人脸之间的区分度。ArcFace 引入角度余弦相似度损失, 旨在优化人脸识别中类别间的分离度。与传统的 softmax 损失相比, ArcFace 在人脸识别任务中通常表现更好。各对比算法均与 VanillaFaceNet 的训练环境、训练世代均相同。得到各算法的识别准确率、模型大小和 FPS 如表 3 所示。

表 3 算法对比试验

Table 3 Algorithm comparison experiment

算法 Algorithm	准确率 Accuracy/%	模型大小 Model size/MB	FPS/(帧·s ⁻¹)
VanillaFaceNet	88.21	101.73	26.23
FaceNet	85.22	87.4	23.56
MobileFaceNet	78.63	4.8	25.46
CenterFace	81.95	21.8	26.13
CosFace	84.36	86.5	24.95
ArcFace	83.72	67.4	25.29

由表 3 可明显看出本文所提算法 VanillaFaceNet 的识别准确率最高, 达到了 88.21%, 与 MobileFaceNet、CenterFace、CosFace 和 ArcFace 算法相比, 本文算法的识别准确率分别提高了 9.58、6.26、3.85 和 4.49 个百分点。时间复杂度远低于其他算法, 算法性能优良。在算法精度方面: 加入改进的注意力机制模块和 center loss 可以有效提高牛脸识别精度, 在算法复杂度方面: 将 FaceNet 的主干网络替换为只有 13 层的 VanillaNet, 使得时间复杂度大幅减少, FPS 达到了 26.23 帧/s。

3.4 牛只之间欧氏距离对比试验

根据前面的试验, 表 3 显示了牛只个体识别模型在测试集上的识别准确率。表 4 给出了 10 只不同牛只 20 张图像的判断结果, 任意两张图像在测试集中构成一对图像。两张行列号相同的面部图像属于同一头牛。

研究表明, 对角线的值小于其他位置的值。这

表明, 同一头牛的两张面部图像之间的相似性小于属于不同牛只的两张面部图像之间的相似性。对于表 5 中的

10 头牛, 采用 0.95 的阈值可以有效判定两张未知的牛只面部图像是否来自同一头牛。

表 4 不同牛只图像的相似性比较

Table 4 Comparison of image similarity among different cattle



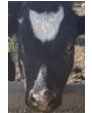


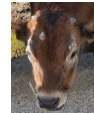








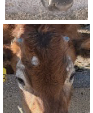

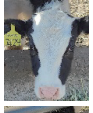

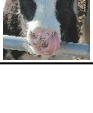
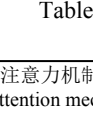
牛只图像 Cow image										
	0.84	1.45	1.48	1.51	1.40	1.42	1.50	1.36	1.54	1.44
	1.21	0.48	1.28	1.06	1.29	1.43	1.29	1.14	1.03	1.43
	1.48	0.96	0.74	1.14	0.96	1.15	1.08	1.20	1.04	1.37
	1.61	1.278	1.04	0.92	1.03	0.98	0.97	1.37	1.20	1.22
	1.56	1.102	0.99	1.04	0.35	1.00	0.99	1.30	1.14	1.27
	1.41	1.380	1.07	1.13	1.16	0.34	1.12	1.25	1.12	1.21
	1.46	1.211	0.97	1.00	1.25	1.12	0.67	1.22	1.10	1.02
	1.43	1.283	1.17	1.04	1.19	1.19	1.16	0.22	0.96	1.06
	1.56	1.080	1.12	1.05	0.96	1.13	1.04	1.10	0.74	1.02
	1.44	1.323	1.158	1.01	1.14	1.039	1.04	0.99	0.98	0.34

表 5 不同注意力机制对比试验

Table 5 Comparative experiment of different attention mechanisms

注意力机制名称 Attention mechanism name	准确率 Accuracy/%	验证率 Validation rate/%	模型大小 Model size/MB	FPS/ (帧·s ⁻¹)
Vanillanet+DBCA	88.21	87.41	101.73	26.23
Vanillanet+SE ^[40]	88.17	86.63	101.66	23.08
Vanillanet+CBAM	86.57	84.82	101.79	23.07
Vanillanet+ECA ^[41]	87.63	86.08	101.53	23.12
Vanillanet+CA ^[42]	87.27	85.98	101.73	26.23

3.5 不同注意力机制对比试验

插入不同注意力机制的网络在牛脸测试集上的结果如表 5 所示。与具有其他注意机制的网络相比, 具有

DBCA 块的网络具有最高的精度, 并且可以忽略内存和计算开销。此外, 这再次证实了本文提出的 DBCA 注意力机制可以比其他注意力机制更好地帮助网络提取特征。

3.6 特征可视化

为了更好地理解 VanillaFaceNet 与基线网络 FaceNet 对于相似牛脸的识别效果, 本文使用 Grad-CAM 来可视化算法对图像的注意力区域, 输入图像为编号 8 130、10 616、10 453、10 032、8 009 的 5 张相似的牛脸图像。

梯度加权类激活映射 (gradient-weighted class activation mapping, grad-CAM) 通过分析梯度的信息来捕获模型对图像中不同牛脸图像部分的关注程度, 从而生成一

一个好的视觉解释, 指示模型在做出预测时对图像的哪些部分有更强的关注, 可以帮助理解模型在牛脸识别任务中为何做出特定的预测。VanillaFaceNet 和基线网络 FaceNet 的 Grad-CAM 结果如图 7 所示, 红色表示该区域中存在高激活, 而蓝色表示预测类别的弱激活。可以

发现, 当 FaceNet 未得到改进时, 该模型较少关注牛只面部区域。在 FaceNet 上进行上述优化操作后, 可以发现 VanillaFaceNet 网络的活跃区域更大, 对于牛只面部的关注度更高。对于相似的牛脸, 提出的 VanillaFaceNet 具有更好的识别效果。

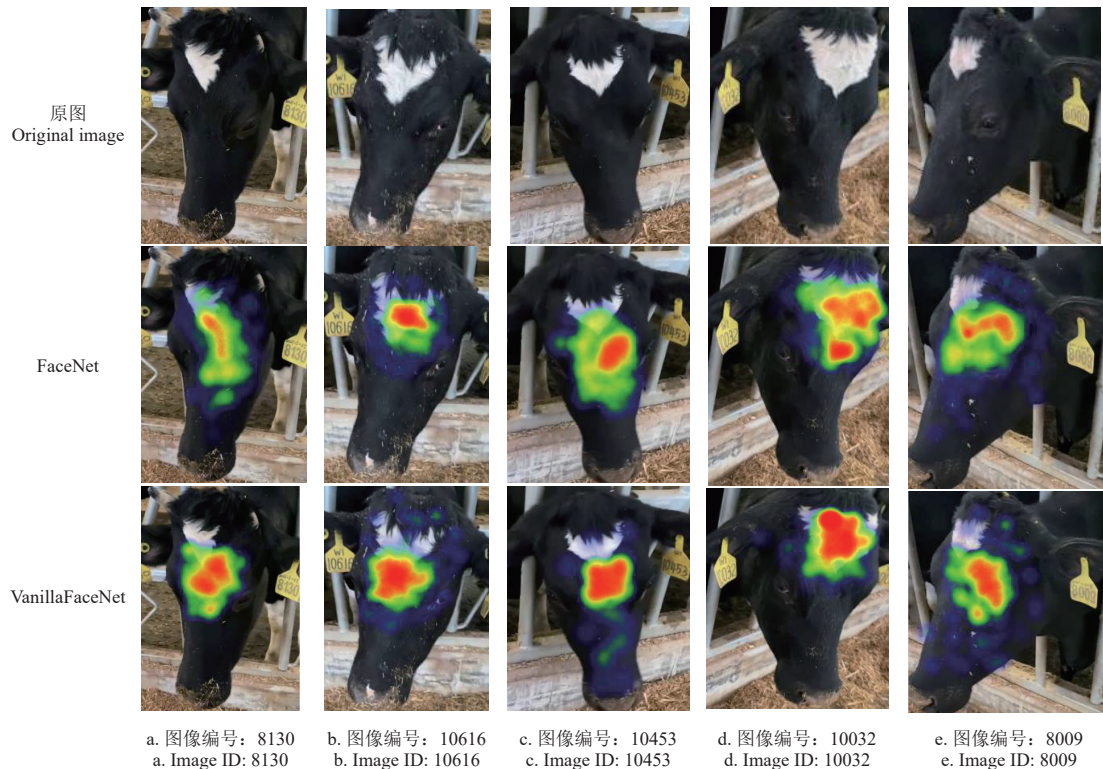


图 7 使用 FaceNet 和 VanillaFaceNet 的相似牛只面部热力图
Fig.7 Heatmaps of similar cattle faces using FaceNet and VanillaFaceNet

3.7 算法移植测试

为了保证算法在提升精度的同时满足移植要求, 将本文算法训练好的模型移植到嵌入式平台 Jetson AGX Xavier 上进行测试, 使用 RetinaFace 检测算法与本文的 VanillaFaceNet 结合搭建牛脸识别平台, 识别效果如图 8 所示。

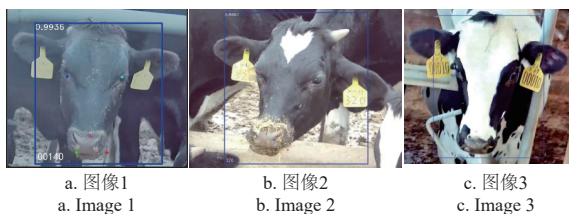


图 8 Jetson AGX Xavier 下牛脸识别结果

Fig.8 Recognition results of cattle faces on Jetson AGX Xavier

首先将牛只图像输入到 RetinaFace 网络中, 检测牛只面部在图片中的位置, 之后, 根据牛只面部的的位置进行裁剪, 得到牛只面部图像并输入到 VanillaFaceNet 进行牛只身份识别。图 8 左上角数字为 RetinaFace 检测是否为牛只面部的概率, 左下角数字为 VanillaFaceNet 网络识别出的牛只身份的编号, 通过与牛只耳边编号对比显示, 通过牛脸进行牛只身份识别均正确, 验证了

VanillaFaceNet 算法在 Jetson AGX Xavier 上的可行性。

4 结论

本文旨在开发一种精度高, 时间复杂度低的开集牛只识别模型, 并将其有效部署到计算有限的嵌入式设备中。本文算法具有以下特点:

- 1) 针对 FaceNet 原始主干特征提取网络层数多、算法复杂度高的问题, 将主干网络替换为只有 13 层 VanillaNet-13, 以提高网络的推理速度;
- 2) 为解决不同牛只面部差异小、难以区分的问题, 提出一种新的 DBCA 注意力机制, 通过加入全局最大池化来提取牛只显著特征, 从而提高网络识别精度;
- 3) 采用 center loss 和 triplet loss 联合训练, 以增强欧式空间中牛脸特征的类内紧凑性和类间可分性。

通过实际嵌入式系统测试表明, 虽然 VanillaFaceNet 网络的 Model size 较大, 但是网络层数仅有 13 层, 可以在资源有限的嵌入式平台中部署, 并对牛只识别的准确率达到 88.21%, 每秒传输帧数为 26.23 帧, 表现出准确率高和算法时间复杂度低的特点, 为非接触式开集牛脸识别提供了参考。未来将进一步探索更好的参数分配, 以实现更加轻量化、高性能的架构。

[参 考 文 献]

- [1] 何东健, 刘建敏, 熊虹婷等. 基于改进 YOLOv3 模型的挤奶奶牛个体识别方法[J]. 农业机械学报, 2020, 51(4): 250-260.
HE Dongjian, LIU Jianmin, XIONG Hongting, et al. An individual identification method for milking cows based on improved YOLOv3 model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(4): 250-260. (in Chinese with English abstract).
- [2] 邢永鑫, 孙游东, 王天一. 基于改进 SSD 算法对奶牛的个体识别[J]. 计算机工程与应用, 2022, 58(2): 208-214.
XING Yongxin, SUN Youdong, WANG Tianyi. Individual identification of cows based on improved SSD algorithm[J]. Computer Engineering and Application, 2022, 58(2): 208-214. (in Chinese with English abstract).
- [3] 何东建, 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.
HE Dongjian, LIU Dong, ZHAO Kaixuan. Research progress on intelligent perception and behaviour detection of animal information in precision animal husbandry[J]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(5): 231-244. (in Chinese with English abstract)
- [4] 张宏鸣, 周利香, 李永恒. 基于改进 MobileFaceNet 的羊脸识别方法[J]. 农业机械学报, 2022, 53(5): 267-274.
ZHANG Hongming, ZHOU Lixiang, LI Yongheng, et al. Sheep face recognition method based on improved mobilefaceNet[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(5): 267-274. (in Chinese with English abstract)
- [5] 齐咏生, 焦杰, 鲍腾飞, 等. 基于自适应注意力机制的复杂场景下牛脸检测算法[J]. 农业工程学报, 2023, 39(14): 173-183.
QI Yongsheng, JIAO Jie, BAO Tengfei, et al. Cattle face detection algorithm in complex scenes using adaptive attention mechanism[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(14): 173-183. (in Chinese with English abstract).
- [6] 黄胜, 冉浩杉. 基于语义信息的精细化边缘检测方法[J]. 计算机工程, 2022, 48(3): 204-210.
HUANG Sheng, RAN Haoshan. Refined edge detection method based on semantic information[J]. Computer Engineering, 2022, 48(3): 204-210. (in Chinese with English abstract)
- [7] 康熙, 李树东, 张旭东, 等. 基于热红外视频的奶牛跛行运动特征提取与检测[J]. 农业工程学报, 2021, 37(23): 169-178.
KANG Xi, LI Shudong, ZHANG Xudong, et al. Features extraction and detection of cow lameness movement based on thermal infrared videos[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(23): 169-178. (in Chinese with English abstract).
- [8] LI Y. A survey on edge intelligent video surveillance with deep reinforcement learning[J]. Journal of Network Intelligence, 2022, 7(1): 70-83.
- [9] KUMAR S, TIWARI S, SINGH S K. Face recognition for cattle[C]// Proceedings of 2015 Third International Conference on Image Information Processing(ICIIP). Wagnaghat, India: IEEE, 2015: 65-72.
- [10] KUMAR S, TIWARI S, SINGH S K. Face recognition of cattle: can it be done[J]. National Acad. Sci. India Section A Phys. Sci., 2016, 86(2): 137-148.
- [11] ZHAO K, JIN X, JI J, et al. Individual identification of Holstein dairy cows based on detecting and matching feature points in body images[J]. Biosystems Engineering, 2019, 181: 128-139.
- [12] LI J, CHEN W, ZHU Y, et al. Intelligent detection and behavior tracking under ammonia nitrogen stress[J]. Neurocomputing, 2023, 559: 126809.
- [13] XIONG H, XIAO Y, ZHAO H, et al. AD-YOLOv5: An object detection approach for key parts of sika deer based on deep learning[J]. Computers and Electronics in Agriculture, 2024, 217: 108610.
- [14] ISLAM M N, YODER J, NASIRI A, et al. Analysis of the drinking behavior of beef cattle using computer vision[J]. Animals, 2023, 13(18): 2984-2994.
- [15] NASIRI A, AMIRIVOJDAN A, ZHAO Y, et al. Estimating the feeding time of individual broilers via convolutional neural network and image processing[J]. Animals, 2023, 13(15): 2428-2439.
- [16] 张垚鑫, 朱荣光, 孟令峰, 等. 改进 ResNet18 网络模型的羊肉部位分类与移动端应用[J]. 农业工程学报, 2021, 37(18): 331-338.
ZHANG Yaixin, ZHU Rongguang, MENG Lingfeng, et al. Classification of mutton location on the animal using improved ResNet18 network model and mobile application[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(18): 331-338. (in Chinese with English abstract).
- [17] MENG F, LI J, ZHANG Y, et al. Transforming unmanned pineapple picking with spatio-temporal convolutional neural networks[J]. Computers and Electronics in Agriculture, 2023, 214: 108298.
- [18] WANG C, LI C, HAN Q, et al. A performance analysis of a litchi picking robot system for actively removing obstructions, using an artificial intelligence algorithm[J]. Agronomy, 2023, 13(11): 2795-2816.
- [19] ULADZISLAU S. 基于牛鼻纹识别的牛只身份鉴别方法[D]. 重庆: 重庆理工大学, 2023.
ULADZISLAU S. Recognition of Cattle using Cow Nose Image Pattern[D]. Chongqing: Chongqing University of Technology, 2023.

- [20] XU B, WANG W, GUO L, et al. CattleFaceNet: A cattle face identification approach based on RetinaFace and ArcFace loss[J]. *Computers and Electronics in Agriculture*, 2022, 193: 106675.
- [21] LI S, FU L, SUN Y, et al. Individual dairy cow identification based on lightweight convolutional neural network[J]. *Plos one*, 2021, 16(11): e0260510.
- [22] GAO J, BURGHARDT T, ANDREW W, et al. Towards self-supervision for video identification of individual holstein-friesian cattle: The Cows2021 dataset[J/OL]. 2021: 2105.01938[2021-5-5].<https://arxiv.org/pdf/2105.01938>.
- [23] HU H, DAI B, SHEN W, et al. Cow identification based on fusion of deep parts features[J]. *Biosystems engineering*, 2020, 192: 245-256.
- [24] ANDREW W, GAO J, MULLAN S, et al. Visual identification of individual Holstein-Friesian cattle via deep metric learning[J]. *Computers and Electronics in Agriculture*, 2021, 185: 106133.
- [25] JIANG B, WU Q, YIN X, et al. FLYOLOv3 deep learning for key parts of dairy cow body detection[J]. *Computers and Electronics in Agriculture*, 2019, 166: 104982.
- [26] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Santiago, Chile: IEEE, 2015: 815-823.
- [27] CHEN H, WANG Y, GUO J, et al. VanillaNet: the Power of Minimalism in Deep Learning[J/OL].2023: 2305.12972[2023-5-23].<https://arxiv.org/pdf/2305.12972>.
- [28] WEN Y, ZHANG K, LI Z, et al. A discriminative feature learning approach for deep face recognition[C]//Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14. Springer International Publishing, 2016: 499-515.
- [29] 白强, 高荣华, 赵春江, 等. 基于改进 YOLOV5s 网络的奶牛多尺度行为识别方法[J]. *农业工程学报*, 2022, 38(12): 163-172.
BAI Qiang, Gaoronghua, ZHAO Chunjiang, et al. Multi-scale behavior recognition method for dairy cows based on improved YOLOV5s network[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, 38(12): 163-172. (in Chinese with English abstract).
- [30] DING X, SHEN C, ZENG T, et al. SAB Net: A Semantic Attention Boosting Framework for Semantic Segmentation[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022: 1-13.
- [31] ZHU C, LI L, WU Y, et al. Saswot: Real-time semantic segmentation architecture search without training[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2024, 38(7): 7722-7730.
- [32] LOU H, DUAN X, GUO J, et al. DC-YOLOv8: small-size object detection algorithm based on camera sensor[J]. *Electronics*, 2023, 12(10): 2323.
- [33] WANG G, CHEN Y, AN P, et al. UAV-YOLOv8: a small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios[J]. *Sensors*, 2023, 23(16): 7190.
- [34] 胡广锐, 周建国, 陈超, 等. 融合轻量化网络与注意力机制的果园环境下苹果检测方法[J]. *农业工程学报*, 2022, 38(19): 131-142.
HU Guangrui, ZHOU Jianguo, CHEN Chao, et al. Fusion of the lightweight network and visual attention mechanism to detect apples in orchard environment[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, 38(19): 131-142. (in Chinese with English abstract).
- [35] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). Munich, Germany: Springer Link, 2018: 3-19.
- [36] CHEN S, LIU Y, GAO X, et al. MobileFaceNets: Efficient cnns for accurate real-time face verification on mobile devices[C]//Biometric Recognition: 13th Chinese Conference, CCBR 2018, Urumqi, China, August 11-12, 2018, Proceedings 13. Springer International Publishing, 2018: 428-438.
- [37] XU Y, YAN W, YANG G, et al. CenterFace: joint face detection and alignment using face as point[J]. *Scientific Programming*, 2020(1): 7845384.
- [38] WANG H, WANG Y, ZHOU Z, et al. Cosface: Large margin cosine loss for deep face recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City, USA: IEEE, 2018: 5265-5274.
- [39] DENG J, GUO J, XUE N, et al. Arcface: Additive angular margin loss for deep face recognition[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Long Beach, USA: arXiv, 2019: 4690-4699.
- [40] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City, USA: IEEE, 2018: 7132-7141.
- [41] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. arXiv, USA: IEEE, 2020: 11534-11542.
- [42] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. arXiv, USA: IEEE, 2021: 13713-13722.

VanillaFaceNet: A high-precision and rapid inference for bovine face recognition

LUAN Haotian^{1,3}, QI Yongsheng^{1,2,3*}, LIU Liqiang^{1,2,3}, WANG Zhaoxia^{1,2,3}, LI Yongting^{1,2,3}

(1. College of Electric Power, Inner Mongolia University of Technology, Hohhot 010051, China; 2. Large-scale Energy Storage Technology Engineering Research Center of Ministry of Education, Hohhot 010080, China; 3. Inner Mongolia Autonomous Region University Smart Energy Technology and Equipment Engineering Research Center, Hohhot 010080, China)

Abstract: Intelligent farming has been an ever-increasing trend in agricultural production, with the development of artificial intelligence (AI) and Internet of Things (IoT). Rapid and accurate identification of cattle identity is of great significance to prevent the insurance fraud for the live cattle loans in the cattle industry. Among them, computer vision can be expected for the cattle face recognition in the modernization transformation of the livestock industry. Smart devices and systems can also be integrated to achieve the intelligent cattle management, feeding, and disease prevention. However, the traditional identification (such as ear tags and collars) has limited the large-scale production in recent years, due to the small differences in facial features among different cattle, the deep layers of the FaceNet network, slow inference speeds, and insufficient classification accuracy. In this study, a cattle face recognition was proposed using FaceNet, called VanillaFaceNet. Firstly, the backbone feature extraction network of FaceNet was replaced with the latest simplified network. VanillaNet-13. Dynamic activation and enhanced linear transformation of activation functions were proposed to improve the non-linearity of the network. Specifically, dynamic activation was fully utilized the expressive power of activation functions during training when dynamically adjusting, in order to flexibly adapt the variations in data distribution at different stages of training. Dynamic activation was used to merge the convolutional layers during inference phase. The computational load was reduced to improve the inference speed of networks. The performance and efficiency of model were then enhanced during training and inference. Activation functions with linear transformations were significantly enhanced the non-linearity through parallel stacking. Multiple activation functions were stacked in parallel, thus enabling each layer to capture more complex features. Additionally, spatial context information was embedded within the activation functions. The spatial relationships among features were better utilized to fit the complex feature distributions. Non-linearity and integration of spatial context information were achieved in a more accurate and efficient model when processing complex data. Secondly, DBCA (Dual-Branch Coordinate Attention) module was added into the global maximum pooling. Global average pooling was used to aggregate significant features of cattle faces, in order better represent the differences among cattle facial features. Therefore, the accuracy network was improved to recognize the cattle. Finally, a center loss was introduced to train the network with the center-triplet loss joint supervision, because the triplet loss was only reduced the inter-class differences among cattle. The intra-class separability of cattle was improved to compactly aggregate the same category of cattle. Thus, the accuracy of comparisons was improved among the same identities of cattle. Cattle face videos were collected at the Otai Ranch in Hohhot, Inner Mongolia Autonomous Region. An image dataset was constructed to train and test the model for the cattle face recognition. The experimental results show that VanillaFaceNet was achieved an accuracy of 88.21% in the cattle recognition, with a frame rate of 26.23 frames per second (FPS). Compared with FaceNet, MobileFaceNet, CenterFace, CosFace, and ArcFace, the model was improved the recognition accuracy by 2.99, 9.58, 6.26, 3.85, and 4.49 percentage points, respectively, and the inference speed by 2.67, 0.77, 0.10, 1.28, and 0.94 frames/s, respectively. The recognition accuracy and speed were greatly improved to fully meet the requirements of the ranch for the accuracy and real-time performance of cattle recognition. The excellent performance was achieved in the cattle recognition, suitable for the deployment on embedded devices, such as Jetson AGX Xavier. A better balance was also gained between accuracy and inference speed of cattle facial recognition.

Keywords: recognition; feature; extraction; cow face; FaceNet; attention mechanism