

许兴时, 王云飞, 邓红兴, 等. 基于跨模态共享特征学习的夜间牛脸识别方法 [J]. 华南农业大学学报, 2024, 45(5): 793-801.

XU Xingshi, WANG Yunfei, DENG Hongxing, et al. Nighttime cattle face recognition based on cross-modal shared feature learning[J]. Journal of South China Agricultural University, 2024, 45(5): 793-801.

# 基于跨模态共享特征学习的夜间牛脸识别方法

许兴时, 王云飞, 邓红兴, 宋怀波

(西北农林科技大学 机械与工程学院/农业农村部农业物联网重点实验室/陕西省农业信息

感知与智能服务重点实验室, 陕西 杨凌 712100)

**摘要:**【目的】解决夜间环境下牛只身份信息难以有效识别的问题, 以为牛只全天候监测提供技术基础。【方法】提出了一种基于跨模态共享特征学习的夜间牛脸识别方法。首先, 模型框架采用浅层双流结构, 有效提取不同模态的牛脸图像中的共享特征信息; 其次, 引入 Triplet 注意力机制, 跨维度地捕捉交互信息, 以增强牛只身份信息的提取; 最后, 通过嵌入扩展模块进一步挖掘跨模态身份信息的表征。【结果】本文提出的夜间牛脸识别模型在测试集上的平均精度均值、一阶累积匹配特征值 (CMC-1) 和五阶累积匹配特征值 (CMC-5) 分别为 90.68%、94.73% 和 97.82%, 相较于未进行跨模态训练的模型, 提高了 19.67、18.91 和 12.00 个百分点。【结论】本研究提出的模型为夜间牛只身份识别问题提供了可靠的解决方案, 为实现牛只全天候持续监测奠定了坚实的技术基础。

**关键词:** 牛; 身份识别; 异质面部识别; 跨模态; 注意力机制; 共享特征; 夜间

中图分类号: TP391.4; S823

文献标志码: A

文章编号: 1001-411X(2024)05-0793-09

## Nighttime cattle face recognition based on cross-modal shared feature learning

XU Xingshi, WANG Yunfei, DENG Hongxing, SONG Huaibo

(College of Mechanical and Electronic Engineering, Northwest A&F University/Key Laboratory of Agricultural

Internet of Things, Ministry of Agriculture and Rural Affairs/Shaanxi Key Laboratory of Agricultural

Information Perception and Intelligent Service, Yangling 712100, China)

**Abstract:** 【Objective】To address the challenge of effectively recognizing cattle identity in the nighttime, and lay the technical foundation for 24-hour monitoring of cattle. 【Method】A nighttime cattle face recognition method based on cross-modal shared feature learning was proposed. The model framework adopted a shallow dual-stream structure to effectively extract shared feature information from different modalities of cattle face images. Additionally, a triplet attention mechanism was introduced to capture intermodal interaction information across dimensions, enhancing the extraction of cattle identity information. Finally, an embedded extension module was utilized to further explore the representation of cross-modal identity information. 【Result】The

收稿日期: 2024-05-10 网络首发时间: 2024-07-15 09:30:49

首发网址: <https://link.cnki.net/urlid/44.1110.S.20240711.1454.007>

作者简介: 许兴时, 硕士研究生, 主要从事模式识别研究, E-mail: [xingshixu@nwfau.edu.cn](mailto:xingshixu@nwfau.edu.cn); 通信作者: 宋怀波, 教授, 博士, 主要从事精准养殖研究, E-mail: [songhuaibo@nwsuaf.edu.cn](mailto:songhuaibo@nwsuaf.edu.cn)

基金项目: 国家重点研发计划 (2023YFD1301800); 国家自然科学基金 (32272931); 陕西省农业重点核心技术项目 (2023NYGG005); 陕西省科技创新引导计划 (2022QFY11-02)

nighttime cattle face recognition model proposed in this article achieved a mean average precision, the first order cumulative matching eigenvalue (CMC-1) and the fifth order cumulative matching eigenvalue (CMC-5) of 90.68%, 94.73% and 97.82% on the test set, respectively. Compared to the model without cross-modality training, the three indexes improved by 19.67, 18.91 and 12.00 percentage points, respectively. 【Conclusion】 The proposed method provides a reliable solution for nighttime cattle identity recognition, laying a solid technical foundation for the application of continuous 24-hour monitoring of cattle.

**Key words:** Cattle; Identification; Heterogeneous face recognition; Cross-modality; Attention mechanism; Shared feature; Nighttime

实现准确的牛只身份识别是实现牛只精准养殖的先决条件<sup>[1-4]</sup>。在智慧养殖系统中,所有数据信息如体尺、体质量、体况、运动量、进食量等都必须与对应的牛只身份关联,以实现有效记录,从而支持个体级别的健康监测和养殖管理<sup>[5-7]</sup>。优秀的牛只身份识别方法已被证实具有提高牛只福利、减轻农民工作量、提升消费者信任、节省政府管理成本、降低银行和保险公司在保险索赔和抵押贷款业务中被欺骗的风险等一系列作用,有着广泛的直接和间接受益对象<sup>[8-10]</sup>。近年来,以计算机视觉技术为基础的非接触式身份识别方法成为了一种具有潜力的牛只身份解决方法。相较于基于射频识别 (Radio frequency identification, RFID) 电子耳标的方法,这类方法不仅能够避免给奶牛带来疼痛或感染风险,还消除了因耳标脱落造成识别失败的担忧<sup>[11-12]</sup>。考虑到牛脸特征具有身份唯一性且相对容易获取,许多研究人员致力于开发准确、高效的牛脸识别模型,以期实现牛只的非接触身份识别<sup>[13-17]</sup>。

然而,现有研究主要关注白天情况下的牛只身份识别,在照明较差或不可用的夜间,身份识别模型的准确度受到了严重影响<sup>[4]</sup>。但是,在夜间对牛只进行监测同样至关重要,因为牛在这段时间依然会进行觅食、反刍、爬跨等行为,部分行为甚至比白天更为活跃<sup>[18-19]</sup>。由于夜间的低光照使得 RGB 图像变得缺乏信息表达能力,现有的研究无法在黑暗条件下实现牛只身份的精准识别。因此,亟需开发一种在夜间能够有效进行牛脸识别的模型,以实现

对牛只 24 h 持续的监测。

一种直接的方法是在牛棚中增加照明,以使牛只身份识别模型在夜间能够准确地提取牛脸特征并实现精确识别。然而,这种方法可能会干扰牛只的正常昼夜节律,甚至导致牛只产生应激反应<sup>[20]</sup>。因此,在夜间实现牛脸识别应该借助不依赖可见光的成像装置。考虑到大多数监控摄像机能够在黑暗中自动从 RGB 模式切换到红外 (Infrared, IR) 模

式,且 IR 成像过程不会对牛只和牛场工人造成干扰,因此使用 IR 图像实现牛脸识别是一种可行的方法。在现实条件下,具有牛只身份标签的图像通常是 RGB 图像。然而,由于 IR 图像和 RGB 图像是异质的,两类图像特征之间具有一定的模态差异,若直接使用模型将 IR 图像在 RGB 图库进行检索,其识别精度不高。

本文提出了 RGB-IR 跨模态牛脸识别任务,这是一个在实际应用中切实存在但在智慧养殖领域缺乏研究的任务。该任务要求在仅有 RGB 注册集的条件下,模型能够对 1 幅 IR 牛脸图像准确检索,以确定牛只的身份。相比于传统的牛脸识别任务,RGB-IR 跨模态牛脸识别任务更加具有挑战性。它不仅要解决传统的牛脸识别任务中牛头遮挡、姿态改变等模态内变化,还要处理 RGB 图像和 IR 图像之间额外存在的跨模态差异。

因此,本文提出了一种基于跨模态共享特征学习的夜间牛脸识别方法。通过设计一个双流结构的特征提取网络,将任意模态的图像嵌入到相同的特征空间中,实现 RGB 与 IR 牛脸图像之间的跨模态匹配。其中,采用 Triplet 注意力机制来加强特征的跨维度交互以提升识别精度,利用嵌入扩展模块生成更多可靠嵌入以增强跨模态信息的表征能力。本文提出的方法旨在解决夜间牛脸识别难题,为实现全天候无间断的牛只精准养殖提供技术基础。

## 1 材料与方法

### 1.1 数据获取与数据集构建

本研究采集的图像源自陕西省咸阳市杨陵区某商业农场,拍摄日期为 2024 年 2 月 12 日至 25 日。研究以 92 头荷斯坦奶牛作为研究对象,分别在日间和夜间采集其面部的 RGB 数据和 IR 数据。RGB 数据采集使用尼康 D5600 相机进行视频录制,分辨率为 1 920 像素×1 080 像素,帧率为 60 帧/s。IR 数据采集则采用窄带 850 工业相机进行视

频录制, 辅以红外辅助灯补光, 分辨率和帧率与 RGB 相同。采集完成后, 将视频每 30 帧截取出 1 张图像以获得图像数据。由于截帧后的图像包含牛脸之外的背景噪声, 需要进一步处理以构建高质量的牛脸数据集。为节省人工成本和时间成本, 本研究使用 YOLOv5 训练了一个牛脸检测模型以辅助确定图像中的牛脸位置。模型检测结果经人工校正后, 牛脸图像被准确截取出来。截取获得的牛脸图像共 7019 幅, 其中 RGB 图像 3932 幅、IR 图像 3087 幅。所有图像的尺寸被统一调整为  $3\times384\times144$ 。最终, 由 60 头奶牛提供的共 4 572 幅图像构成训练集, 而由另外 32 头奶牛提供的共 2 447 幅图像组成测试集。训练集和测试集中牛只身份是相互正交的, 即测试集中出现的牛只在训练过程中不会出

现。表 1 展现了本研究构建的 RGB-IR 跨模态牛脸识别数据集的具体细节。

图 1a 和图 1b 分别展示了 RGB 和 IR 数据集的部分样本。通过观察可以发现, 不同模态的图像有着一定的特征差异, IR 图像相比于 RGB 图像, 丢失了除斑纹特征之外的许多其他特征, 缺乏丰富的纹理特征和结构表达。此外, 同一模态下同一个体不同图像之间也存在着因姿态、光照等因素变化带来的较大域内差异, 这进一步加剧了跨模态牛脸识别任务的困难程度。

1.2 夜间牛脸识别模型构建

1.2.1 模型总体结构 为了将来自不同模态牛脸图像的特征投影到共享特征空间中, 本研究构建了如图 2 所示的夜间牛脸识别模型。该模型以

表 1 RGB-IR 跨模态牛脸识别数据集具体细节  
Table 1 Overview of RGB-IR cross-modal cattle face recognition dataset

数据集 Dataset	牛只数量 Number of cattles	RGB图像数量 Number of RGB images	IR图像数量 Number of IR images	图像总数量 Total number of images
训练集 Training set	60	2 570	2 002	4 572
测试集 Test set	32	1 362	1 085	2 447

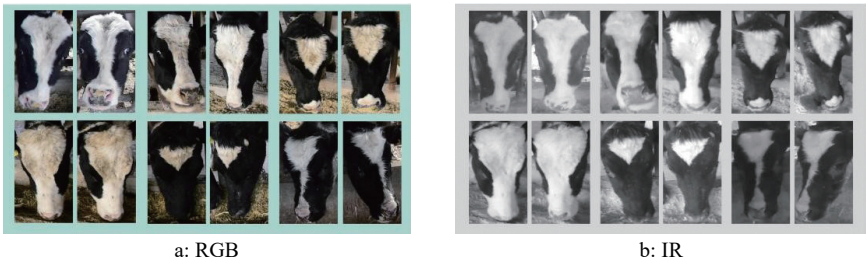
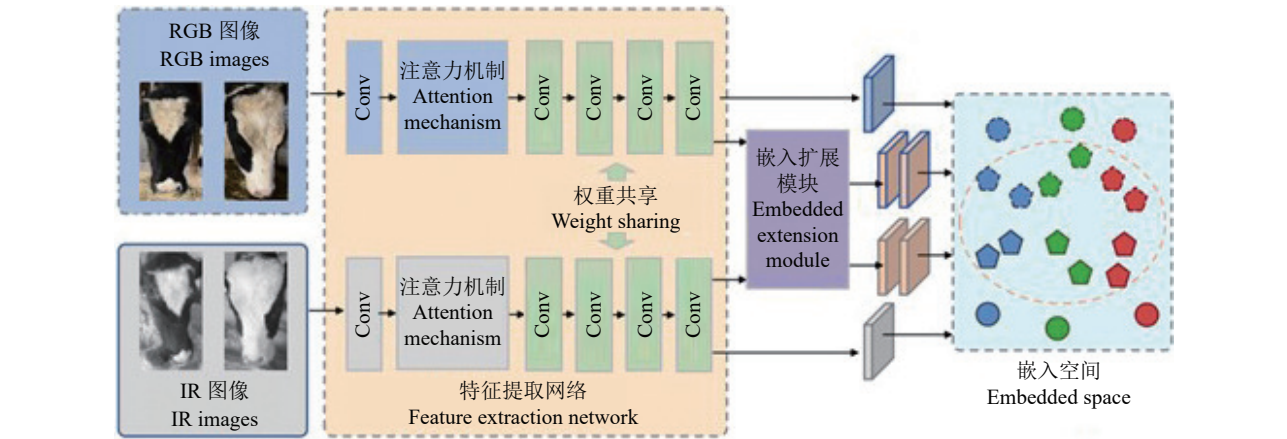


图 1 数据集的部分图像样本  
Fig. 1 Partial image samples in dataset



嵌入空间中圆形和五边形色块分别表示原始嵌入和扩展嵌入, 虚线和实线色块分别表示 RGB 图像和 IR 图像的嵌入, 色块的不同颜色表示不同身份的牛只个体

In the embedded space, circles and pentagons represent the original and extended embeddings respectively, dashed and solid color blocks represent the embeddings of RGB images and IR images respectively, different colors of the color blocks represent individual cattles of different identities

图 2 夜间牛脸识别模型  
Fig. 2 Night cattle face recognition model



ResNet50 作为主干特征提取网络用于提取牛脸图像中的身份特征信息。ResNet<sup>[21]</sup> 是一种经典的深度残差神经网络模型框架,它通过残差学习框架和跳跃连接解决了梯度消失问题,具有较深的网络结构,能够实现身份特征的稳定有效提取。表 2 展示了 ResNet50 的具体模型结构,它由 5 个阶段组成。其中,第 1 个阶段包括 1 个卷积核大小为 7×7、步幅为 2 的卷积操作,以及 1 个窗口大小为 3×3、步幅为 2 的最大值池化操作。其余的 4 个阶段分别由多个具有残差结构的瓶颈模块组成。

考虑到不同模态图像之间存在一定的属性差异,本研究将模型的低层结构部分设置为双流结构,以更好地捕获 2 个不同模态中的低级特征模式。而高层结构部分的网络参数则被设计为 2 个模态共享,旨在增强模型提取模态共享特征的能力。同时,本研究在双流结构的每个分支的第 1 个阶段后面都加入了 Triplet 注意力机制以增强模型性能,并在特征提取后利用嵌入扩展模块生成更多可靠嵌入以进一步提升识别精度。

1.2.2 Triplet 注意力机制 对具有身份信息特征进行充分提取是构建精准牛只识别模型的基础。本研究中,为了增强网络对牛只身份信息的关注,引

表 2 ResNet 模型结构  
Table 2 Model structure of ResNet

阶段 Stage	操作 Operation	重复次数 Stack number
1	Conv, 7×7, 64, stride 2 Max pool, 3×3, stride 2	1
2	Conv, 1×1, 64 Conv, 3×3, 64 Conv, 1×1, 256	3
3	Conv, 1×1, 128 Conv, 3×3, 128 Conv, 1×1, 128	4
4	Conv, 1×1, 256 Conv, 3×3, 256 Conv, 1×1, 1024	6
5	Conv, 1×1, 512 Conv, 3×3, 512 Conv, 1×1, 2048	3

入轻量化的 Triplet 注意力机制<sup>[22]</sup>,其利用三支结构捕捉跨维度交互来计算注意力权重。如图 3 所示,第 1 个分支获取空间维度  $H$  和  $W$  的交互关系,中部的分支负责计算通道维度  $C$  和空间维度  $W$  的注意力权重,底部的最后 1 个分支负责计算通道维度  $C$  和空间维度  $H$  的注意力权重。

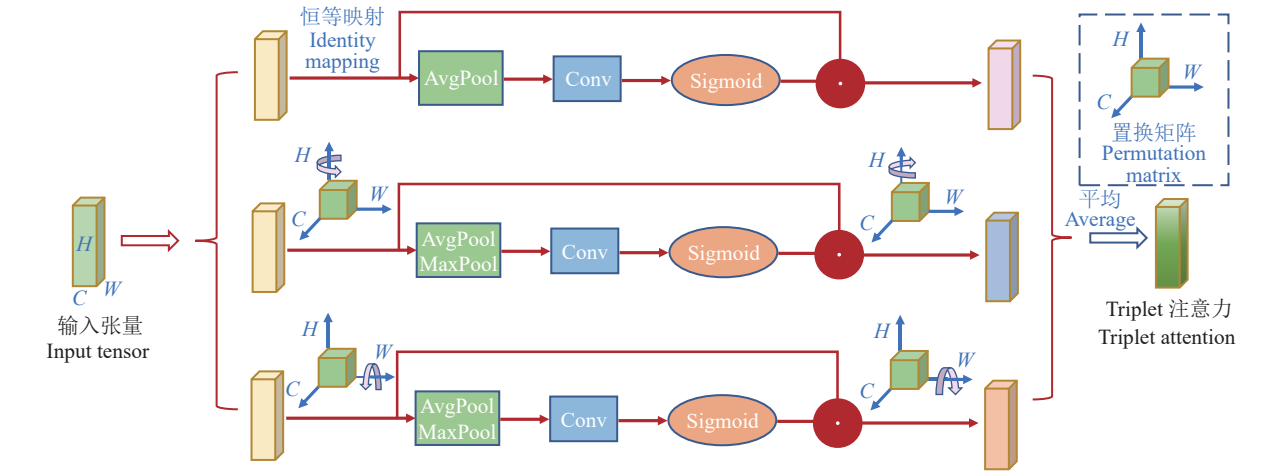


图 3 Triplet 注意力原理  
Fig. 3 Triplet attention schematic

如图 4 所示,当给定输入张量  $M$  的形状为  $C \times H \times W$  时,3 个分支具体实现如下:

在第 1 个分支中,输入张量  $M$ ,通过通道最大值池化和通道平均值池化得到张量  $M_1(2 \times H \times W)$ ,然后通过卷积、BN 层及 Sigmoid 非线性激活函数生成成为空间注意力权重,经残差连接与  $M$  相加得到形状为  $C \times H \times W$  的输出。

在第 2 个分支中,输入张量  $M$ ,沿  $H$  轴逆时针

旋转  $90^\circ$  得到旋转张量  $\hat{M}_2(W \times H \times C)$ ,再经过 Z-Pool 操作得到张量  $\hat{M}_2^*(2 \times H \times C)$ ,其中 Z-Pool 操作是在  $W$  维度进行平均值池化和最大值池化,在保留丰富特征的基础上缩小了深度。然后通过卷积、BN 层及 Sigmoid 激活函数生成注意力权重,经残差连接后,沿  $H$  轴顺时针旋转  $90^\circ$  保持和输入张量形状一致。

在第 3 个分支中,输入张量  $M$ ,沿  $W$  轴逆时针

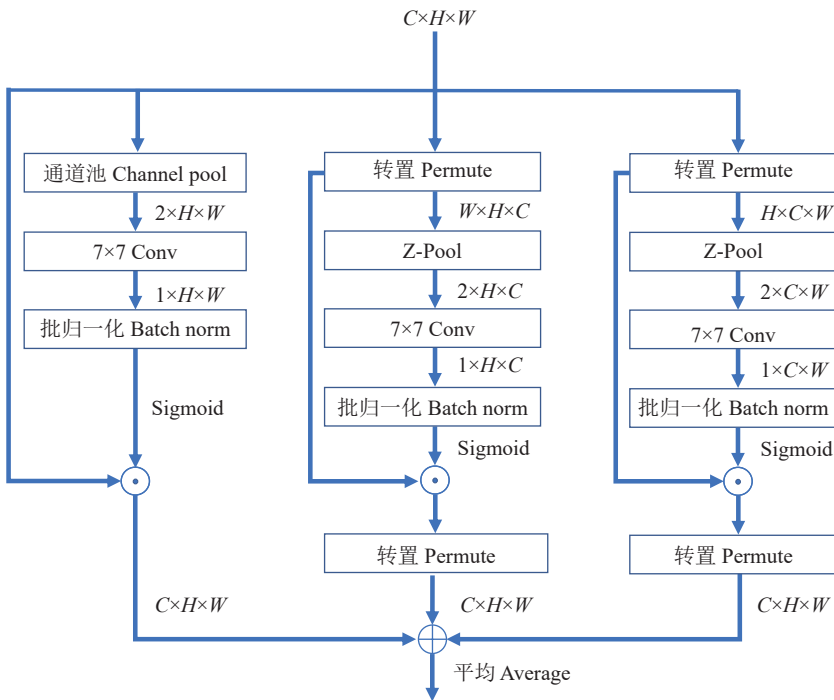


图4 Triplet 注意力结构

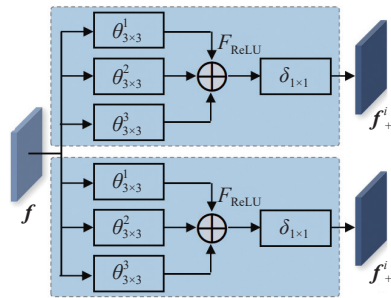
Fig. 4 Triplet attention architecture

旋转  $90^\circ$  得到旋转张量  $\hat{M}_3(H \times C \times W)$ , 再经过 Z-Pool 操作得到张量  $\hat{M}_3^*(2 \times C \times W)$ , 然后通过卷积、BN 层及 Sigmoid 激活函数生成注意力权重, 经残差连接后, 沿  $W$  轴顺时针旋转  $90^\circ$  保持和输入张量形状一致。

最后对 3 个分支输出的  $C \times H \times W$  维度特征进行相加取平均。

该方法通过跨维度信息交互, 解决了传统方法计算单一维度注意力权重时信息大量丢失等问题, 可以在学习任务中捕捉到更复杂的关联和依赖性。

**1.2.3 嵌入扩展模块** 出色的嵌入表达能够使识别模型具备更准确、更鲁棒的识别效果<sup>[23-24]</sup>。为使牛脸识别模型在样本有限的条件下充分挖掘跨模态的身份信息表征, 本研究提出的夜间牛脸识别模型在特征提取后, 利用嵌入扩展模块生成更多可靠嵌入以进一步提升识别精度。如图 5 所示, 提出的嵌入扩展模块由多个嵌入生成分支组成, 每个分支都能产生一个新的嵌入特征。在每个分支中, 特征图  $f$  首先经过 3 个不同扩张率的  $3 \times 3$  空洞卷积, 得到的 3 个特征图通道大小为特征图  $f$  通道大小的  $1/4$ 。接着, 将这些特征图相加融合, 并使用 ReLU 激活函数以增强模型的非线性表示能力。最后, 通过一个卷积核大小为  $1 \times 1$  的卷积层改变特征图的通道数量, 使生成的扩展特征图  $f_+$  与输入的原特征图  $f$  尺寸相匹配。因此, 第  $i$  个分支生成的扩展嵌入  $f_+^i$  可以表示为:



$f$  代表原始的嵌入特征,  $f_+^i$  代表第  $i$  个分支生成的扩展嵌入特征,  $\theta_{3 \times 3}^n(\cdot)$  代表扩张率为  $n$  的  $3 \times 3$  空洞卷积,  $F_{\text{ReLU}}(\cdot)$  代表非线性激活函数,  $\delta_{1 \times 1}(\cdot)$  代表  $1 \times 1$  卷积

$f$  represents the original embedded features,  $f_+^i$  represents the extended embedded features generated by the  $i$ -th branch,  $\theta_{3 \times 3}^n(\cdot)$  represents a  $3 \times 3$  dilated convolution with a dilation rate of  $n$ ,  $F_{\text{ReLU}}(\cdot)$  represents a nonlinear activation function, and  $\delta_{1 \times 1}(\cdot)$  represents a  $1 \times 1$  convolution

图5 嵌入扩展模块

Fig. 5 Embedding expansion module

$$f_+^i = \delta_{1 \times 1} \{ F_{\text{ReLU}}[\theta_{3 \times 3}^1(f) + \theta_{3 \times 3}^2(f) + \theta_{3 \times 3}^3(f)] \}, \quad (1)$$

式中,  $f$  代表原始的嵌入特征,  $f_+^i$  代表第  $i$  个分支生成的扩展嵌入特征,  $\theta_{3 \times 3}^n(\cdot)$  代表扩张率为  $n$  的  $3 \times 3$  空洞卷积,  $F_{\text{ReLU}}(\cdot)$  代表非线性激活函数,  $\delta_{1 \times 1}(\cdot)$  代表  $1 \times 1$  卷积。

嵌入扩展模块旨在生成更多嵌入以提高识别精度, 但现有操作未能有效地确保嵌入的多样性。因此, 本研究使用嵌入增强损失  $L_{\text{emb}}$  来保证生成的扩展嵌入尽可能多样化, 从而减少 RGB 和 IR 图像之间的模态差异。首先, 为促使模型学习不同的特征并挖掘不同的跨模态线索, 扩展嵌入与原始嵌入

之间应该具有一定距离。其次,扩展嵌入应该减少 RGB 和 IR 图像之间的模态差异。需要拉近由 RGB 模态生成的扩展嵌入与原始 IR 嵌入的距离,以及拉近由 IR 模态生成的扩展嵌入与原始 RGB 嵌入的距离。最后,考虑到过分拉近扩展嵌入与原始嵌入之间的距离可能导致不同类的嵌入变得更近。因此,为避免不同类之间的嵌入过于接近,需要确保各个嵌入的类内距离小于类间距离。基于以上要求,RGB 模态中生成的嵌入增强损失可表达为:

$$L(f_R, f_I, f_{R+}) = [D(f_I^a, f_{R+}^{i,a}) - D(f_R^a, f_{R+}^{i,a}) - D(f_R^a, f_{R+}^b)], \tag{2}$$

式中,  $D(\cdot, \cdot)$  表示 2 个嵌入之间的欧氏距离,  $f_R$  和  $f_I$  分别代表 RGB 模态和 IR 模态的原始嵌入特征,  $f_{R+}^i$  代表第  $i$  个分支的生成的扩展嵌入,  $a$  和  $b$  表示不同的牛只身份。

同样,对于由 IR 生成的扩展嵌入,应满足:

$$L(f_I, f_R, f_{I+}) = [D(f_R^a, f_{I+}^{i,a}) - D(f_I^a, f_{I+}^{i,a}) - D(f_I^a, f_{I+}^b)]. \tag{3}$$

因此,嵌入增强损失最终的计算公式为:

$$L_{emb} = L(f_R, f_I, f_{R+}) + L(f_I, f_R, f_{I+}). \tag{4}$$

为了确保多个不同分支生成的嵌入能够最大化地捕获不同的信息特征表征,研究设置正交损失强制不同分支生成的嵌入特征彼此最小化重叠元素。因此,本研究设置的分支正交损失为:

$$L_{branch} = \sum_{m=1}^{i-1} \sum_{n=m+1}^i (f_+^{mT} f_+^n), \tag{5}$$

式中,  $m$  和  $n$  分别代表嵌入扩展模块中的第  $m$  个和第  $n$  个分支。分支正交损失可以强制生成的嵌入学习更多信息的特征表示。

**1.2.4 综合损失函数** 同时使用交叉熵损失  $L_{CE}$  和三元组损失  $L_{Tri}$  能够促进牛只身份识别模型取得更好的识别效果<sup>[24]</sup>。为了获得令人满意的牛脸识别精度,本研究同样利用这 2 类损失促使模型学习具有判别性的特征。结合上文针对嵌入扩展模块提出的嵌入增强损失  $L_{emb}$  和分支正交损失  $L_{branch}$ , 本研究最终的损失函数公式如下:

$$Loss = L_{CE} + L_{Tri} + L_{emb} + L_{branch}. \tag{6}$$

## 2 试验平台参数及结果评价指标

### 2.1 试验环境参数

本文试验的操作系统为 Ubuntu 18.04.6 LTS, 它包含 Intel(R) Xeon(R) Gold 5217 CPU @ 3.00

GHz, 4 张 32 G NVIDIA Tesla V100 GPUs, 256 GB RAM, 6 TB 硬盘。本研究以 Python 3.8 作为编程语言,深度学习框架采用 PyTorch 1.7,加速环境为 Cuda 11.0,编程平台为 PyCharm。所有对比算法均在相同环境下运行。

### 2.2 模型训练参数

采用 COCO 数据集上的预训练权重作为 ResNet50 部分的初始化权重参数,其他部分采取 Kaiming 初始化。嵌入扩展模块的分支数设置为 2,优化器选用随机梯度下降法,初始学习率设为 0.01,学习率衰减因子为 0.003,衰减动量因子为 0.95。输入图像尺寸为  $3 \times 384 \times 144$ ,批处理大小为 4,最大迭代次数为 80 次。

### 2.3 模型结果评价指标

为验证该模型性能,本研究选用平均精度均值 (Mean average precision, mAP) 和  $k$  阶累积匹配特征值 (Cumulative matching characteristics, CMC- $k$ ) 作为模型识别效果的评价指标,选用参数量 (Parameters) 和计算量 (FLOPs) 作为模型复杂度的评价指标。mAP 衡量的是模型的平均检索性能; CMC- $k$  是指在模型测试结果中置信度最高的前  $k$  个检索结果中出现正确匹配的概率; Parameters 是指模型中所含的参数数量,决定着模型文件的大小和推断时所需要的内存量,可有效衡量算法的空间复杂度; 计算量是指模型在推断时的浮点运算次数,可有效衡量算法的时间复杂度。其中, mAP 的具体计算方式如式 (7)~(9) 所示。

$$P = \frac{TP}{TP + FP}, \tag{7}$$

$$R = \frac{TP}{TP + FN}, \tag{8}$$

$$mAP = \frac{1}{C} \sum_{c=1}^C \int_0^1 P(R) dR, \tag{9}$$

式中,  $P$  和  $R$  分别表示精确率和召回率,  $C$  表示检测类数,  $TP$  表示实际为正例且被划分为正例的个数,  $FP$  表示实际为负例但被划分为正例的个数,  $FN$  表示实际为正例但被划分为负例的个数。

## 3 结果与讨论

### 3.1 夜间牛脸识别模型结果

本研究利用测试集对提出的夜间牛脸识别模型的识别效果进行测试。测试集中的 IR 图像为问询集, RGB 图像为注册集,测试中对于问询集的任意 1 幅图像模型都会在注册集中匹配与其身份最接近的图像。图 6 为训练过程 Loss 值及 mAP 变

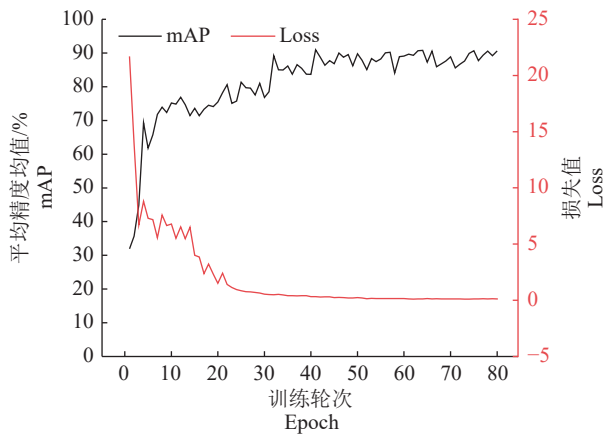


图 6 训练过程参数变化曲线

**Fig. 6 Change curve of parameter in training process**

化。模型训练初期, 学习率较高, 损失曲线在前 15 轮迅速下降, 随着迭代的进行, 损失曲线逐渐变缓并于 40 轮次左右达到收敛, 模型达到稳定状态; 类似地, mAP 值在前 15 轮迅速上升, 之后逐渐趋于平缓并在 87% 附近波动。

除了对提出模型进行测试外, 本研究还测试了

在训练过程中仅使用 RGB 图像而未使用 IR 图像进行跨模态训练的模型, 以突显夜间牛脸识别的挑战和必要性。为保证公平比较, 2 个模型均采用相同网络结构, 只是在训练过程中双流网络的 2 个分支输入不同。2 个模型的牛脸识别表现如表 3 所示。

本文提出模型的 mAP、CMC-1 和 CMC-5 分别为 90.68%、94.73% 和 97.82%。与未进行跨模态训练的模型相比, 本研究提出模型的 mAP 提高了 19.67 个百分点, CMC-1 提高了 18.91 个百分点, CMC-5 提高了 12.00 个百分点, 模型的识别效果大大提高, 更加适应夜间条件下的牛只身份识别任务。

3.2 消融试验

本研究提出的夜间牛脸识别模型基于浅层双流模型结构, 通过加入 Triplet 注意力机制优化模型身份特征提取能力, 通过增加嵌入扩展模块充分挖掘跨模态的身份信息表征。为验证各部分的性能表现, 使用控制变量法进行消融试验。在相同的数据集和训练参数下进行训练及测试, 结果如表 4 所示。

表 3 跨模态训练效果对比试验结果 <sup>1)</sup>			
Table 3 Comparative experimental result of cross-modal training effect			
模型 Model	mAP/%	CMC-1/%	CMC-5/%
未进行跨模态训练的模型 Model without cross-modal training	71.01	75.82	85.82
提出的模型 Proposed model	90.68	94.73	97.82

1) mAP: 平均精度均值; CMC-1: 一阶累积匹配特征值; CMC-5: 五阶累积匹配特征值

1) mAP: Mean average precision; CMC-1: Cumulative match characteristic at rank 1; CMC-5: Cumulative match characteristic at rank 5

表 4 各个模型消融试验结果 <sup>1)</sup>							
Table 4 Result of ablation experiment for each model							
模型结构 Model structure	Triplet注意力机制 Triplet attention mechanism	嵌入扩展模块 Embedding extension modules	mAP/%	CMC-1/%	CMC-5/%	Parameters/M	FLOPs/G
单流 Single-stream	√		77.23	84.55	91.64	9.18	4.73
			81.13	86.91	92.36	9.24	4.75
	√	√	79.00	85.09	92.36	9.18	4.73
		√	81.42	87.09	92.91	9.24	4.75
全双流 Full dual-stream	√		80.73	88.73	94.00	9.18	4.73
			84.88	87.45	95.09	9.24	4.75
	√	√	82.75	89.27	93.64	9.18	4.73
		√	87.96	91.64	96.17	9.24	4.75
浅层双流 Shallow dual-stream	√		81.86	89.45	94.73	9.18	4.73
			89.60	92.00	95.09	9.24	4.75
	√	√	86.19	89.64	96.55	9.18	4.73
		√	90.68	94.73	97.82	9.24	4.75

1) mAP: 平均精度均值; CMC-1: 一阶累积匹配特征值; CMC-5: 五阶累积匹配特征值; Parameters: 参数量; FLOPs: 浮点运算量

1) mAP: Mean average precision; CMC-1: Cumulative match characteristic at rank 1; CMC-5: Cumulative match characteristic at rank 5; Parameters: Number of parameters; FLOPs: Floating point operations



如表 4 所示,在未使用嵌入扩展模块和 Triplet 注意力机制的条件下,本研究采用的浅层双流结构相比于单流结构和全双流结构,模型识别效果明显提升。相比于单流结构,模型的 mAP、CMC-1 和 CMC-5 分别提高了 4.63、4.90 和 3.09 个百分点。相比于全双流结构,模型的 mAP、CMC-1 和 CMC-5 分别提高了 1.13、0.72 和 0.73 个百分点。这表明浅层双流结构能够兼顾不同模态之间的模态差异并实现模态共享特征的有效提取。

在使用浅层双流结构且未使用嵌入扩展模块的条件下,采用 Triplet 注意力机制后,模型的 Parameters 和 FLOPs 较采用 Triplet 注意力机制前分别仅增加了 0.06 M 和 0.02 G, mAP、CMC-1 和 CMC-5 分别提高了 7.74、2.55 和 0.36 个百分点。这表明引入的轻量化 Triplet 注意力机制,能够跨维度交互以抑制冗余信息,有效增强了网络对牛只身份信息的关注。

嵌入扩展模块通过在训练时生成更多可靠嵌入,实现有限样本条件下跨模态身份信息表征的充分挖掘。在采用 Triplet 注意力机制的浅层双流结构的基础上,模型使用嵌入扩展模块之后, mAP、CMC-1 和 CMC-5 较使用该模块前分别又提高了 1.08、2.73 和 2.73 个百分点。同时,由于测试时嵌入扩展模块并不参与牛脸图像的特征提取过程,所以并不会增加模型在推理时的 Parameters 和 FLOPs。

综上所述,本研究提出的夜间牛脸识别模型的各个改进部分都提高了牛只身份识别的准确性,每个部分对模型性能的提升都有不同的贡献,证明了各个优化操作的有效性。

## 4 结论

准确识别牛只身份是现代智慧畜牧的重要环节。然而,由于夜间光线条件的限制,现有的识别模型在这种环境下通常难以实现满意的识别效果。本研究针对夜间牛只身份识别难的问题,提出了一种基于跨模态共享特征学习的夜间牛脸识别方法。该方法实现了在仅有 RGB 注册集条件下利用 IR 牛脸图像进行准确的牛只身份识别。提出的模型采用浅层双流结构,有效将不同模态的牛脸图像特征投影到共享特征空间中。同时,利用 Triplet 注意力机制,该模型能够跨维度地捕捉交互信息,从而有效增强牛只身份信息的提取。此外,通过嵌入扩展模块,进一步挖掘跨模态身份信息的表征。试验结果显示,本研究提出模型的 mAP、CMC-1 和 CMC-5 分别为 90.68%、94.73% 和 97.82%。相较于未进

行跨模态训练的模型,本研究提出的模型在 mAP、CMC-1 和 CMC-5 分别提高了 19.67、18.91 和 12.00 个百分点,为夜间牛只身份识别任务提供了可靠的解决方案。

### 参考文献:

[1] 熊安然, 熊本海, 蒋林树. 奶牛数字化养殖技术研究进展[J]. 中国乳业, 2020, 11: 29-32.

[2] 杨亮, 王辉, 陈睿鹏, 等. 畜禽个体身份标识技术发展进程与展望[J]. 猪业科学, 2023, 40(9): 24-27.

[3] KAUR A, KUMAR M, JINDAL M K. Cattle identification with muzzle pattern using computer vision technology: A critical review and prospective[J]. *Soft Computing*, 2022, 26(10): 4771-4795.

[4] 许兴时, 王云飞, 华志新, 等. 融合 YOLOv5s 与通道剪枝算法的奶牛轻量化个体识别方法[J]. 农业工程学报, 2023, 39(15): 152-162.

[5] LI R, WEN Y, ZHANG S, et al. Automated measurement of beef cattle body size via key point detection and monocular depth estimation[J]. *Expert Systems with Applications*, 2024, 244: 123042.

[6] XU X, WANG Y, SHANG Y, et al. Few-shot cow identification via meta-learning[J]. *Information Processing in Agriculture*, 2024, 4: 1-11.

[7] 王政, 宋怀波, 王云飞, 等. 奶牛运动行为智能监测研究进展与技术趋势[J]. 智慧农业, 2022, 4(2): 36-52.

[8] HOSSAIN M, KABIR M, ZHENG L, et al. A systematic review of machine learning techniques for cattle identification: Datasets, methods and future directions[J]. *Artificial Intelligence in Agriculture*, 2022, 6: 138-155.

[9] MAHMUD M, ZAHID A, DAS A, et al. A systematic literature review on deep learning applications for precision cattle farming[J]. *Computers and Electronics in Agriculture*, 2021, 187: 106313.

[10] QIAO Y, KONG H, CLARK C, et al. Intelligent perception for cattle monitoring: A review for cattle identification, body condition score evaluation, and weight estimation[J]. *Computers and Electronics in Agriculture*, 2021, 185: 106143.

[11] HUANG X, HU Z, QIAO Y, et al. Deep learning-based cow tail detection and tracking for precision livestock farming[J]. *IEEE/ASME Transactions on Mechatronics*, 2023, 28(3): 1213-1221.

[12] FERREIRA R, BRESOLIN T, ROSA G, et al. Using dorsal surface for individual identification of dairy calves through 3D deep learning algorithms[J]. *Computers and Electronics in Agriculture*, 2022, 201: 107272.

[13] WENG Z, MENG F, LIU S, et al. Cattle face recognition based on a two-branch convolutional neural network[J]. *Computers and Electronics in Agriculture*, 2022, 196: 106871.

[14] LU Y, WENG Z, ZHENG Z, et al. Algorithm for cattle identification based on locating key area[J]. *Expert Sys-*



tems with Applications, 2023, 228: 120365.

[15] 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.

[16] BAKHSHAYESHI I, ERFANI E, TAGHIKHAH F, et al. An intelligence cattle reidentification system over transport by siamese neural networks and YOLO[J]. IEEE Internet of Things Journal, 2024, 11(2): 2351-2363.

[17] YANG L, XU X, ZHAO J, et al. Fusion of RetinaFace and improved FaceNet for individual cow identification in natural scenes[J/OL]. Information Processing in Agriculture, (2023-09-02) [2024-05-01]. <https://doi.org/10.1016/j.inpa.2023.09.001>.

[18] SENGER P L. The estrus detection problem: New concepts, technologies, and possibilities[J]. Journal of Dairy Science, 1994, 77(9): 2745-2753.

[19] WANG Z, HUA Z, WEN Y, et al. E-YOLO: Recognition of estrus cow based on improved YOLOv8n model[J]. Expert Systems with Applications, 2024, 238: 122212.

[20] CASEY T, PLAUT K. Circadian clocks and their integration with metabolic and reproductive systems: Our current understanding and its application to the management of dairy cows[J]. Journal of Animal Science, 2022, 100(10): 233.

[21] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 770-778.

[22] MISRA D, NALAMADA T, ARASANIPALAI A U, et al. Rotate to attend: Convolutional triplet attention module[C]//2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa, HI, USA: IEEE, 2021: 3139-3148.

[23] ZHANG Y, WANG H. Diverse embedding expansion network and low-light cross-modality benchmark for visible-infrared person re-identification[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE, 2023: 2153-2162.

[24] WANG Y, XU X, WANG Z, et al. ShuffleNet-Triplet: A lightweight RE-identification network for dairy cows in natural scenes[J]. Computers and Electronics in Agriculture, 2023, 205: 107632.

【责任编辑 庄 延】