

# Swin-BiCAS:一种基于稀疏注意力与 空间-通道融合的牛脸识别方法\*

刘禹辰<sup>1</sup>,赵婵婵<sup>2</sup>,许思轩<sup>1</sup>,晓艳<sup>3</sup>,石宝<sup>4</sup>

(1. 内蒙古工业大学智能科学与技术学院,呼和浩特 010080; 2. 内蒙古工业大学通识教育学院,呼和浩特 010080; 3. 内蒙古师范大学经济管理学院,呼和浩特 010022; 4. 内蒙古工业大学信息工程学院,呼和浩特 010051)

**摘要:**随着养牛业逐步向精细化、智能化方向转型,如何实现快速而准确的牛只个体身份识别,已成为当前畜牧管理中的关键技术瓶颈。传统接触式识别方法存在成本高、易篡改等缺陷,而现有非接触式视觉识别技术对牛脸的特征提取能力不足,在复杂养殖环境下的鲁棒性较差。本文提出了一种新的牛脸识别方法(Swin-BiCAS):首先,在Swin Transformer中引入Bi-former Block,通过稀疏化的双层路由注意力机制,有效弥补传统窗口注意力机制在建模长距离依赖方面的不足,在降低计算复杂度的同时提升细粒度特征提取能力;其次,提出一种特征融合模块CASmerge,通过协同优化空间感知与通道关系来增强特征表达能力,使模型能够同时保持对空间细节的敏感性和对通道间相关性的建模能力。实验结果达到了99.1%的准确率,证明了所提方法的有效性。

**关键词:**个体识别; Swin Transformer; 注意力机制; 牛脸识别; 特征融合

**中图分类号:**TP391; S823; S818.9 **文献标志码:**A

畜牧业是现代农业的一个重要领域,对国民经济发展和居民生活保障具有战略意义。精准畜牧业(Precision livestock farming, PLF)<sup>[1]</sup>是现代农业智能化发展的核心方向。当前智慧畜牧产业快速发展,数字化精细管控需求迫切,牛只的个体识别问题成为畜牧业数字化管理的关键技术瓶颈。作为PLF的基础环节,牛的个体识别对于生长监测、健康管理、行为分析及产品溯源具有重要意义<sup>[2]</sup>,实现牛的个体识别是在牛场实施PLF、提高畜牧生产效率的基本前提。此外,优秀的牛只识别方法被证明可产生多重增益效应,如在动物福利层面实现精准健康管理<sup>[3]</sup>,在运营层面降低人力成本,在市场端增强消费者对产品的信任度,在监管层面优化政府资源分配效率,在金融领域有效抑制保险欺诈和信贷风险<sup>[4]</sup>等。因此,对牛只识别技术准确性、鲁棒性的探索和需求急剧增加,相关研究已成为当前畜牧业智能化转型的重要研究方向。

可靠的个体标识是实现牛只精准识别的关键前提。根据标识符的来源,现有识别方法可分为人工标识和生物特征识别两大类。人工标识方法通过外源性标记实现个体区分,主要包括耳缺刻、耳标、烙印、热烙标记和射频识别(RFID)技术等<sup>[5]</sup>。然而,这类方法在大规模养殖场景的自动化应用中存在显著局限性。近年来,随着畜牧业的快速发展和规模化养殖的推进,RFID技术成为当前应用最

\* 收稿日期:2025-10-01; 修回日期:2026-03-19

**基金项目:**内蒙古自治区科技计划项目(2025YFHH0083);内蒙古自然科学基金项目(2025MS06039, 2023LHMS07016, 2025QN06005)

**作者简介:**刘禹辰(2001—),女,河南郑州人,2023级硕士研究生。E-mail: 15038215423@163.com

**通信作者:**石宝(1982—),男(蒙古族),内蒙古兴安盟人,教授,博士。主要从事数字图像处理研究。E-mail: kshibao@163.com

广泛的牛只识别手段<sup>[6]</sup>,但是在实际应用中存在识别距离受限、信号碰撞以及标签重复使用、伪造或丢失等问题,导致其人工维护成本居高不下<sup>[7]</sup>。这些缺陷使得RFID技术在智能养殖和食品溯源等场景中的应用效率降低且经济性不足。此外,上述方法普遍具有侵入性或需强制佩戴设备,可能诱发牲畜应激反应,不符合现代动物福利标准<sup>[8]</sup>。

牛只个体识别技术的研究已从传统人工方法逐步发展为基于生物特征和深度学习的智能化方法,非侵入式牛只识别技术研究取得了显著进展<sup>[9]</sup>。早期研究中,Baranov等<sup>[10]</sup>通过实验验证了牛鼻纹图案在牛只个体识别中的可行性。Allen等<sup>[11]</sup>提出基于虹膜图像的识别系统,在受控环境下取得了较高的识别准确率。然而,由于采集设备成本较高且对拍摄条件要求严格,该方法在实际牧场环境中难以推广。Gaber等<sup>[12]</sup>提出的基于韦伯局部描述符(WLD)特征和Adaboost分类器的口鼻识别方法,达到99.5%的准确率。随着深度学习的发展,基于卷积神经网络和Transformer结构的牛脸识别方法逐渐成为研究热点。Noviyanto等<sup>[13]</sup>利用DenseNet和ResNet等网络实现高精度牛只识别。Yao等<sup>[14]</sup>通过卷积神经网络实现奶牛面部识别任务。Kumar等<sup>[15]</sup>引入了基于支持向量机(SVM)和模式识别的奶牛个体识别系统。杨蜀秦等<sup>[16]</sup>构建了一种融合坐标注意力模块的YOLOV4优化检测框架,该模型适配复杂场景,对各类毛色奶牛面部识别具有稳定效果。Lu等<sup>[17]</sup>提出定位关键区域(Locate key area, LKA)模块,通过类激活图精确定位关键区域,增强了特征判别力。Wang等<sup>[18]</sup>利用三元组损失增强个体判别能力。这些研究为基于视觉的牛只识别提供了重要技术基础,但在复杂环境下仍面临光照变化、姿态变化和跨品种差异等挑战。

近年来,牛脸识别算法不断涌现,识别精度已有显著提升。但在自然牧场场景下,现有研究仍面临诸多挑战,如复杂天气易造成图像光照不均、特征模糊,同时牛在自由活动姿态与拍摄角度的多变会导致图像纹理、轮廓差异明显<sup>[19]</sup>。

针对上述问题,以Swin Transformer作为主干网络,提出了一种新的混合注意力Swin框架(Swin-BiCAS),引入Biformer Block的动态稀疏注意力机制,在降低计算冗余的同时聚焦关键面部区域,减少噪声干扰。同时,提出CASmerge模块,嵌入坐标注意力机制,通过坐标分解增强空间位置感知,提升姿态鲁棒性,并通过SaE模块动态调整通道权重,增强跨品种特征判别力。这些改进形成协同增强效应,使模型能够同时捕捉牛脸的全局结构和局部细微特征。

综上所述,本文主要贡献如下:

①提出一种多模态增强的牛脸识别框架,构建轻量化但高判别力的牛脸特征表示模型,能够在复杂光照、姿态变化和部分遮挡等非受控条件下实现高精度个体识别,为智能化畜牧管理提供可靠的技术支持;②引入Biformer Block,通过稀疏化双层路由注意力机制(Bi-level routing attention, BRA)与动态键值对筛选策略,在保证计算效率的同时,显著提升模型对牛脸关键区域的长距离依赖建模能力,解决传统窗口注意力感受野受限的问题;③提出一个新的CASmerge模块,通过协同优化空间感知与通道关系,有效提升牛脸特征表达的丰富性和判别性,实现更鲁棒、更准确的牛脸识别。

## 1 Swin-BiCAS牛脸识别模型结构设计

图1展示了Swin-BiCAS的基本架构,并展示了以Swin Transformer作为主干网络的方法。将Swin Transformer作为主干网络,在Stage 2后,嵌入Biformer Block,将Stage 2的输出输入Stage 3和Biformer Block,通过双层路由注意力机制构建动态稀疏注意力路径,增强局部空间关系建模能力。然后,将Stage 3和Biformer Block的输出输入CASmerge模块进行特征融合,通过协同优化空间感知与通道关系,提升关键特征的表达能力。最后,将CASmerge的输出输入Stage 4和第二个Biformer Block,将二者的输出再次融合,通过第二个CASmerge模块得到最终的深度特征。

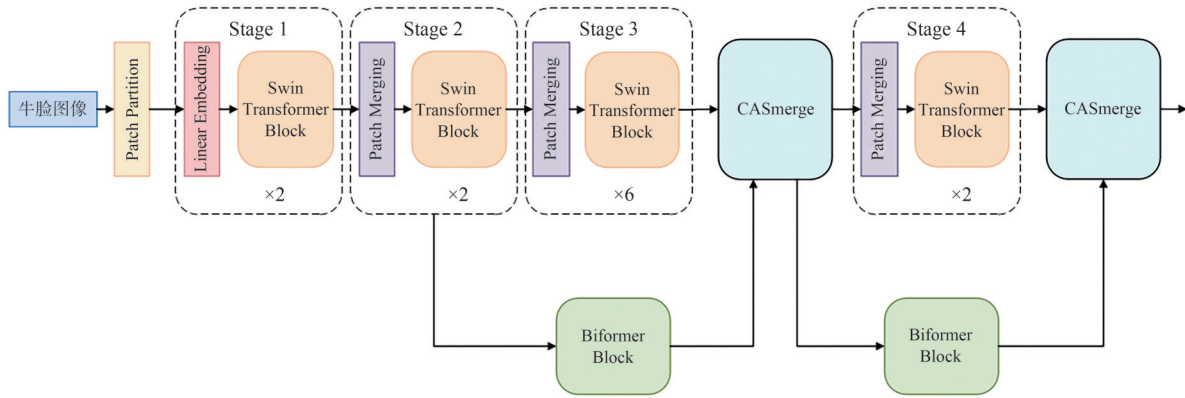


图1 Swin-BiCAS模型整体架构

Fig. 1 Overall architecture of the Swin-BiCAS model

### 1.1 Swin Transformer

与 Vision Transformer 不同, Swin Transformer 采用分层窗口计算 (Hierarchical window-based self-attention) 机制。Swin Transformer 将图像划分为固定大小且不重叠的局部窗口, 在每个窗口间进行独立计算, 这些窗口通过补丁嵌入转换为固定维度的向量表示。引入窗口移位策略来增强跨窗口信息交互, 引入层级特征下采样机制来构建多尺度特征表示。窗口的注意力公式为

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}} + B\right)\mathbf{V},$$

$$\mathbf{Q} = \mathbf{X}\mathbf{W}^q, \mathbf{K} = \mathbf{X}\mathbf{W}^k, \mathbf{V} = \mathbf{X}\mathbf{W}^v,$$

其中,  $\mathbf{Q}, \mathbf{K}, \mathbf{V}$  分别表示查询 (Query)、键 (Key)、值 (Value) 矩阵,  $\mathbf{W}^q, \mathbf{W}^k, \mathbf{W}^v \in \mathbb{R}^{C \times C}$  为投影权重,  $C$  为通道数,  $d$  表示注意力维度,  $B$  表示相对位置偏置。

### 1.2 Biformer Block

Biformer Block 结构如图 2 所示, 首先对输入图像进行  $3 \times 3$  卷积隐式编码相对位置信息, 使用双层路由注意力 (BRA) 动态筛选重要区域, 仅对少数相关 token 计算注意力, 避免全局计算。然后, 通过 2 层扩展比为  $e$  的多层感知器 (MLP) 对每个位置的 token 进行非线性变换。

BRA 机制通过区域划分、路由筛选、细粒度注意力 3 步构建动态稀疏注意力机制。给定一个 2D 特征图  $X$ , 将  $X$  划分为  $S \times S$  个非重叠区域, 每个区域包括  $HW/S^2$  个特征向量, 其中,  $H$  和  $W$  分别为特征图的高和宽。使用共享的线性层生成  $\mathbf{Q}, \mathbf{K}, \mathbf{V}$  矩阵。

区域到区域路由 (Region-to-region routing) 筛选每个区域需要交互的 Top- $k$  个相关区域。首先, 对每个区域的  $\mathbf{Q}$  和  $\mathbf{K}$  进行平均池化, 得到区域级表征  $\mathbf{Q}^r = \text{AvgPool}(\mathbf{Q}), \mathbf{K}^r = \text{AvgPool}(\mathbf{K}), \mathbf{Q}^r, \mathbf{K}^r \in \mathbb{R}^{S^2 \times C}$ , 然后, 计算区域间相关性矩阵  $\mathbf{A}^r$ , 公式为  $\mathbf{A}^r = \mathbf{Q}^r (\mathbf{K}^r)^T$ , 其中,

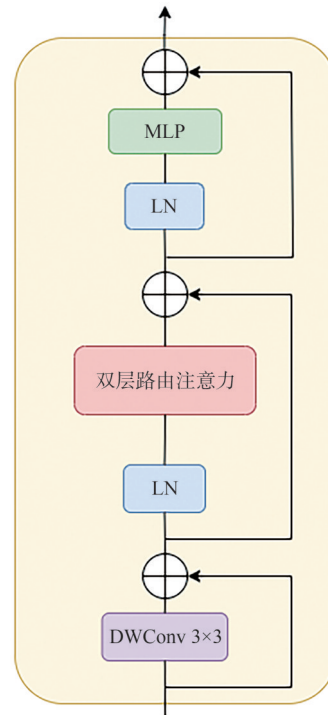


图2 Biformer Block结构

Fig. 2 Structure of the Biformer Block

$A^r(i, j)$ 表示区域 $i$ 与区域 $j$ 的语义关联强度。最后对每个区域保留最相关的 $k$ 个相邻区域,生成路由索引矩阵 $I^r \in \mathbb{N}^{S^2 \times k}$ :

$$I^r = \text{topIndex}(A^r).$$

令牌到令牌注意力(Token-to-token attention)仅对路由筛选后的区域进行细粒度注意力计算,根据 $I^r$ 从 $K$ 和 $V$ 中收集相关区域的令牌:

$$K^g = \text{gather}(K, I^r), V^g = \text{gather}(V, I^r),$$

其中, $K^g, V^g \in \mathbb{R}^{S^2 \times k \frac{HW}{S^2} \times C}$ 是收集到的键值张量。再将注意力应用于收集到的键值对:

$$O = \text{Attention}(Q, K^g, V^g) + L(V),$$

其中, $L(V)$ 为可选的局部上下文增强项。

BRA的计算由3部分组成:线性投影、区域到区域路由和令牌到令牌注意力,总的计算量可以表示为三者计算的总和:

$$\begin{aligned} F_{\text{lops}} &= F_{\text{lopsproj}} + F_{\text{lopsrouting}} + F_{\text{lopsattn}} \\ &= 3HWC^2 + 2(S^2)^2C + 2HWk \frac{HW}{S^2}C \\ &= 3HWC^2 + C \left( 2S^4 + \frac{k(HW)^2}{S^2} + \frac{k(HW)^2}{S^2} \right) \\ &\geq 3HWC^2 + 3C \left( 2S^4 \frac{k(HW)^2}{S^2} \frac{k(HW)^2}{S^2} \right)^{\frac{1}{3}} \\ &= 3HWC^2 + 3Ck^{\frac{2}{3}}(2HW)^{\frac{4}{3}}. \end{aligned}$$

与窗口注意力相比,BRA通过动态路由自动学习交互模式,更具灵活性。

### 1.3 CASmerge 模块

为增强特征表达能力,设计CASmerge特征融合模块,如图3所示。该模块由坐标注意力模块与SaE通道注意力模块组成,通过联合建模空间位置信息与通道相关性,实现特征的自适应重标定。

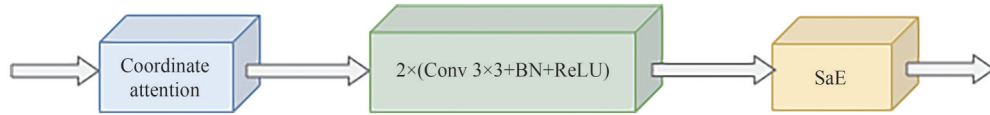


图3 CASmerge模块结构

Fig. 3 Structure of the CASmerge module

#### 1.3.1 坐标注意力模块

坐标注意力模块分为两个核心步骤:坐标信息嵌入和坐标注意力生成,同时建模通道关系和远程依赖关系,保留精确的位置信息,其结构如图4所示。

在坐标信息嵌入阶段,为避免传统通道注意力在空间信息表达上的不足,坐标注意力机制通过方向分解的特征编码方式在保持通道建模能力的同时引入位置信息。该机制将全局池化分解为两个方向的一维特征编码:对输入特征图 $X$ 的每个通道 $c$ 和高度 $h$ ,沿高度方向进行一维平均池化,核大小为 $(H, 1)$ ,生成特征图为

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i).$$

类似地,宽度为 $w$ 的第 $c$ 个通道的输出为

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, w).$$

上述两个变换分别沿着两个空间方向聚合特征,产生一对方向感知特征映射,可以避免位置信息的丢失,能够更准确地定位目标区域。

在坐标注意力生成阶段,基于嵌入的坐标信息,通过轻量级变换生成注意力权重。给定生成的聚合特征映射 $z^h$ 和 $z^w$ ,对二者进行拼接操作,通过共享的 $1 \times 1$ 卷积 $F_1$ 和非线性激活生成中间特征 $f$ 为

$$f = \delta(F_1([z^h, z^w])), f \in \mathbb{R}^{C/r \times (H+W)},$$

其中, $[\cdot, \cdot]$ 表示沿空间维度的拼接操作, $\delta$ 为非线性激活函数, $r$ 是控制块大小的缩减比。然后,将 $f$ 沿空间维度进行拆分,分别用 $1 \times 1$ 卷积 $F_h$ 和 $F_w$ 变换,并通过Sigmoid激活得到注意力权重 $g^h$ 和 $g^w$ ,公式为

$$g^h = \sigma(F_h(f^h)), g^w = \sigma(F_w(f^w)),$$

将 $g^h$ 和 $g^w$ 广播至输入尺寸,与原始特征元素逐项相乘,实现方向感知型特征调制,公式为

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j).$$

### 1.3.2 SaE 模块

图5展示了SaE模块结构。SaE模块通过多路径全连接结构建模通道间的复杂关系,利用Sigmoid函数生成通道权重,对特征图进行自适应重标定,通过增加路径多样性来提升特征表示能力,同时通过共享的reduction保持计算效率。公式如下:

$$y_b^c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_{b,c,i,j}, y \in \mathbb{R}^{B \times C},$$

$$y_k = \text{ReLU}(W_k \cdot y), W_k \in \mathbb{R}^{(C/r) \times C}, k = 1, 2, 3, 4,$$

$$y_{\text{concat}} = [y_1, y_2, y_3, y_4] \in \mathbb{R}^{B \times (4C/r)},$$

$$s = \sigma(W_c \cdot y_{\text{concat}}),$$

$$x_{\text{out}} = x \odot s_{\text{expand}},$$

其中, $x_{\text{out}}$ 表示SaE模块的输入特征, $y_b^c$ 表示第 $b$ 个样本、第 $c$ 个通道的池化结果, $y_k$ 表示第 $k$ 个路径的输出, $W_k$ 表示可学习权重矩阵, $r$ 表示压缩比, $W_c$ 表示融合权重矩阵, $s$ 表示通道注意力权重。

SaE模块的计算可归纳为

$$x_{\text{out}} = x \odot \sigma(W_c \cdot [\text{ReLU}(W_1 \cdot y), \text{ReLU}(W_2 \cdot y), \text{ReLU}(W_3 \cdot y), \text{ReLU}(W_4 \cdot y)]).$$

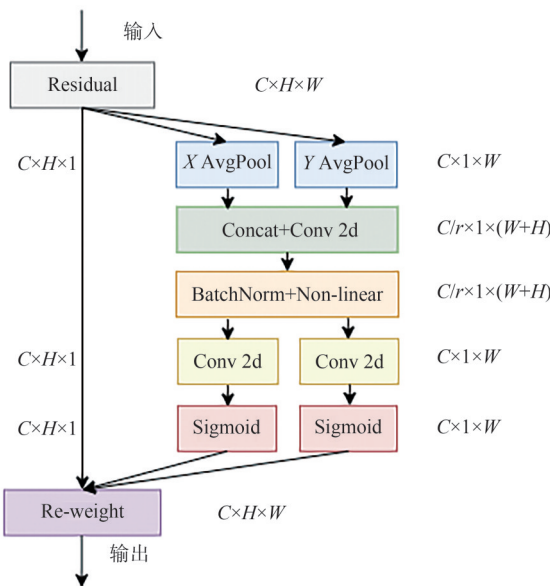


图4 坐标注意力模块结构

Fig. 4 Structure of the coordinate attention module

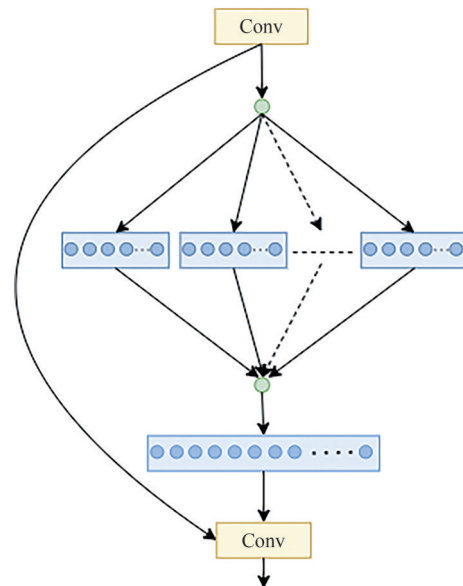


图5 SaE模块结构

Fig. 5 Structure of the SaE module

## 2 实验

### 2.1 数据集

选取131头荷斯坦奶牛和西门塔尔肉牛作为研究对象,数据采集时间为2024年10月2日至2024年10月6日,采集地点为内蒙古自治区兴安盟科尔沁右翼中旗的两个养牛场。数据采集通过视频录制的方式进行,使用手机进入养牛场近距离采集牛只视频,在复杂的农场环境中捕捉牛的高分辨率图像,分辨率设置为1920像素×1080像素,视频帧率为30 fps。录制时从不同的角度和距离完成每头牛的数据采集。

利用视频帧提取技术将拍摄的视频分解为图像,为了保证数据的多样性,采用结构相似性指数(Structural-similarity index measure, SSIM)算法将相似性过高的图像数据进行剔除,SSIM的计算公式为

$$S_{SIM,x,y} = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)},$$

其中, $x$ 和 $y$ 表示两张图像, $\mu_x$ 和 $\mu_y$ 表示各图像的均值, $\sigma_x$ 和 $\sigma_y$ 是标准差, $\sigma_{xy}$ 表示协方差, $c_1$ 和 $c_2$ 为平滑参数。

进行视频帧提取后,图像中含有除牛脸以外的背景噪声,需要进一步处理以构建牛脸数据集。使用YOLOv8(You only look once version 8)来训练牛脸检测模型,以协助确定图像内牛脸所在位置。对检测结果进行人工复核后,牛脸图像被准确裁剪。裁剪后的牛脸图像像素尺寸调整为112×112。为提升模型训练稳定性与特征一致性,在输入网络前对图像进行统一预处理,包括直方图归一化增强对比度、像素值归一化处理以及随机水平翻转、亮度扰动等数据增强策略,以提高模型对光照变化与姿态变化的鲁棒性。最后,从131头牛中共获得了10327张牛脸图像,并且每头牛的牛脸图像不少于30张,其中一部分图像如图6所示。

数据集包含荷斯坦奶牛与西门塔尔牛两个品种,二者在面部纹理与毛色分布上存在明显差异。通过在混合品种数据上进行训练与测试,能够有效评估模型的跨品种泛化能力。将牛脸数据集按照7:2:1分为训练集、验证集和测试集,不同的数据集之间不存在交集。

### 2.2 评估指标

为了综合评价模型的性能,使用精确率(Precision,  $P_{rec}$ )、召回率(Recall,  $R_{ec}$ )、准确率(Accuracy,  $A_{cc}$ )和F1分数(F1-score,  $F_1$ )作为评价指标,公式为

$$P_{rec} = \frac{T_p}{T_p + F_p}, R_{ec} = \frac{T_p}{T_p + F_N},$$

$$A_{cc} = \frac{T_p + T_N}{T_p + F_N + F_p + T_N}, F_1 = \frac{2T_p}{2T_p + F_N + F_p},$$

其中, $T_p$ 和 $T_N$ 表示正确分类的正例和负例的数量, $F_N$ 和 $F_p$ 表示错误分类的正例和负例的数量。

### 2.3 实验设置

实验基于PyTorch框架实现,借助16 GB RAM的NVIDIA Tesla P100 GPU完成加速计算,使用Adam优化器,输入图像像素尺寸设置为112×112,批大小设置为32,初始学习率设置为0.0001。

### 2.4 实验结果

图7展示了Swin-BiCAS模型在训练集和验证集上的准确率情况。如图7所示,对于训练集,在早期阶段准确率快速增长,当训练轮数增加至约20个epoch时,模型准确率逐渐达到较高水平,并接近峰值状态。随着迭代的进行,准确率的增长趋势逐渐放缓,趋于稳定。这表明Swin-BiCAS模型具有良好的收敛性能,在训练过程中未出现明显的过拟合或振荡现象。图8为Swin-BiCAS模型预测的归一化混淆矩阵,可以看出,大多数类别在对角线上均呈现出较高的数值,说明模型能够较为准确地识别出不同个体牛只的面部特征。图9是在数据集中任意选出5头牛的可视化结果,可以看出,热

力值高的地方大多集中在面部区域,这表明模型在进行分类识别时重点关注牛面部区域,能够有效提取并利用具有区分度的核心特征进行身份判定。

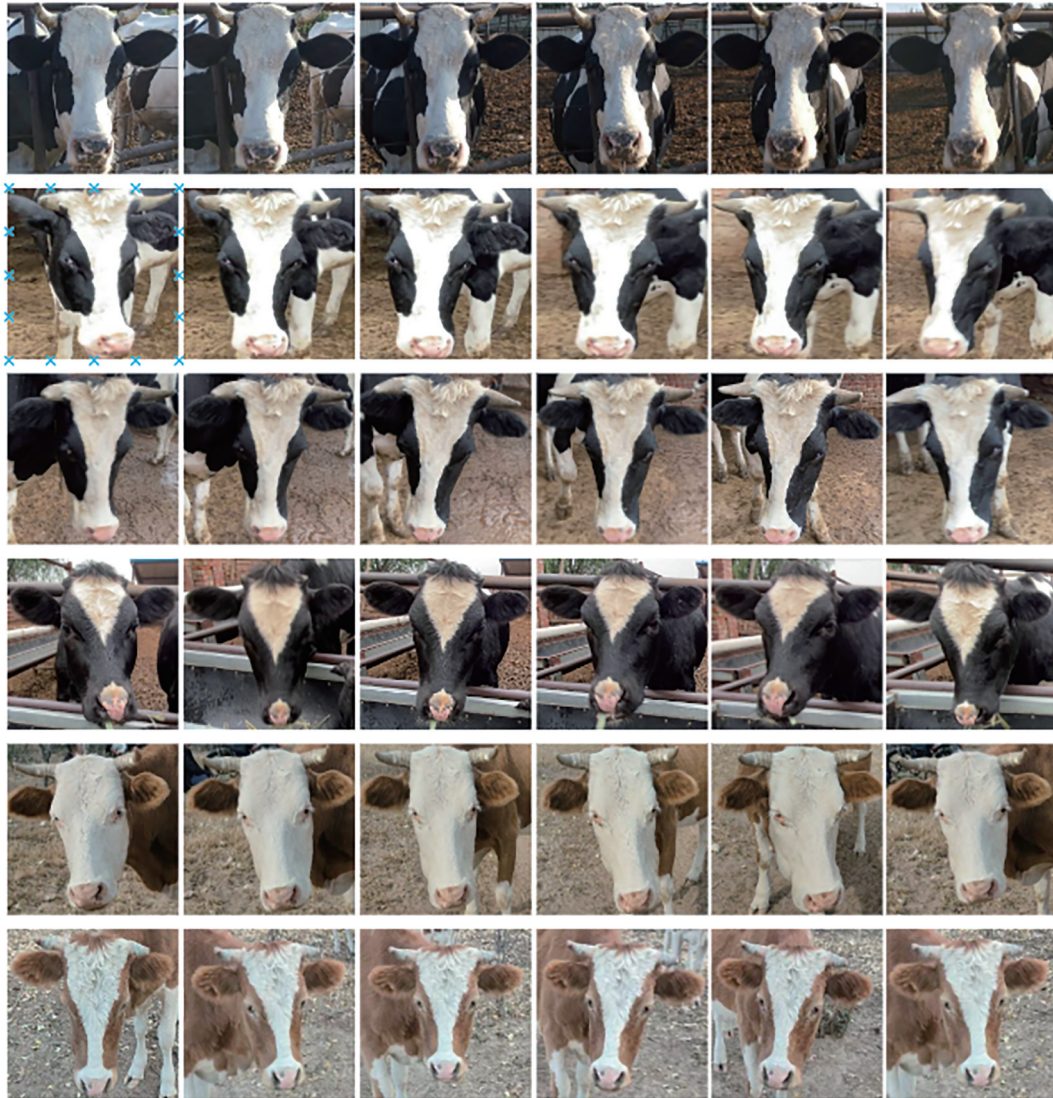


图6 牛脸数据集的部分示例

Fig. 6 Partial examples from the cow face dataset

## 2.5 对比实验结果

为了全面评估 Swin-BiCAS 模型的性能,选择 MaxViT<sup>[20]</sup>、MetaFormer<sup>[21]</sup>、Next-ViT<sup>[22]</sup>、ViT<sup>[23]</sup>、MLP-Mixer<sup>[24]</sup>和 ResNet50<sup>[25]</sup>进行对比实验,结果如表 1 所示。Swin-BiCAS 模型在各项评价指标上均取得最佳性能,其中准确率达到 99.1%,显著优于其他视觉 Transformer 模型和卷积神经网络模型。与传统 CNN 模型 ResNet50 相比,本文模型的准确率与  $F_1$  值分别提升了 4.4 个百分点和 5.4 个百分点。这表明基于自注意力机制的特征建模方式能够更有效地捕获牛脸区域的全局结构信息,提高个体区分能力。与 ViT 和 MLP-Mixer 等全局建模结构相比, Swin-BiCAS 模型表现更优,说明分层窗口注意力结合稀疏路由机制能够在保持计算效率的同时增强局部与全局特征协同建模能力。

## 2.6 消融实验

为深入探究各项改进措施对牛脸识别模型准确率的具体贡献,设计消融实验,在相同的实验环境下对添加的模块分别进行探索,实验结果如表 2 所示。

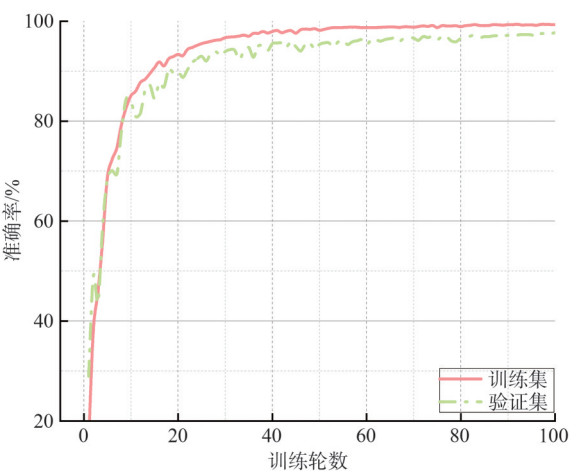


图 7 Swin-BiCAS 模型在训练集与验证集上的  
准确率

Fig. 7 Accuracy of the Swin-BiCAS model on the  
training and validation sets

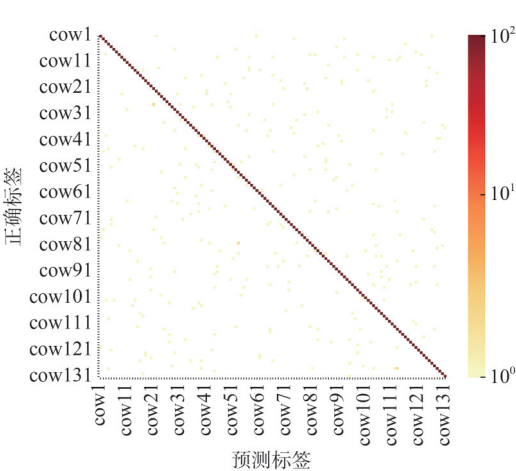


图 8 Swin-BiCAS 模型在测试集上的  
归一化混淆矩阵

Fig. 8 Normalized confusion matrix of the  
Swin-BiCAS model on the test set



图 9 可视化结果

Fig. 9 Visualization results

表 1 Swin-BiCAS 与主流基线模型的性能对比

| Table 1 Performance comparison of Swin-BiCAS with mainstream baseline models |           |          |          | %     |
|------------------------------------------------------------------------------|-----------|----------|----------|-------|
| 模型                                                                           | $P_{rec}$ | $R_{ec}$ | $A_{cc}$ | $F_1$ |
| MaxViT                                                                       | 93.1      | 92.8     | 94.4     | 92.9  |
| MetaFormer                                                                   | 94.4      | 93.2     | 95.1     | 93.8  |
| Next-ViT                                                                     | 96.8      | 95.6     | 97.2     | 96.2  |
| ViT                                                                          | 90.4      | 90.6     | 91.9     | 90.5  |
| MLP-Mixer                                                                    | 97.3      | 96.9     | 97.8     | 97.1  |
| ResNet50                                                                     | 93.6      | 93.2     | 94.7     | 93.4  |
| Swin-BiCAS(本文模型)                                                             | 98.7      | 98.9     | 99.1     | 98.8  |

从实验结果可以看出,在基准 SwinT 模型基础上引入 Biformer Block 后, $F_1$  值提高 0.2 个百分点,准确率提高 0.3 个百分点,表明稀疏路由注意力机制能够增强关键区域的长距离依赖建模能力,从而提升特征表达能力。当仅引入 CASmerge 模块时,模型性能提升更为显著, $F_1$  值提高 1.2 个百分

点,准确率提高 1.5 个百分点,说明空间位置信息建模与通道关系重标定能够有效增强牛脸细粒度特征表达能力,提高不同个体之间的判别能力。当同时引入 Biformer Block 与 CASmerge 模块时,模型性能获得最大提升,准确率较基准模型提升 2.8 个百分点。

表 2 消融实验结果  
Table 2 Ablation study results

| 模型               | Biformer Block | CASmerge | $F_1/\%$ | $A_{cc}/\%$ |
|------------------|----------------|----------|----------|-------------|
| SwinT            | ×              | ×        | 95.5     | 96.3        |
| SwinT+Biformer   | √              | ×        | 95.7     | 96.6        |
| SwinT+CASmerge   | ×              | √        | 96.7     | 97.8        |
| Swin-BiCAS(本文模型) | √              | √        | 98.8     | 99.1        |

3 结论与展望

针对复杂养殖环境下牛脸识别鲁棒性不足的问题,提出了一种基于改进 Swin Transformer 的牛脸识别方法,通过引入稀疏路由注意力机制,有效解决了传统窗口注意力感受野受限的问题,实现关键区域的长距离依赖建模;同时设计 CASmerge 特征融合模块,通过空间位置感知增强与通道关系重标定协同优化,提高了特征表达的判别能力,能够在自然场景下实现准确的牛脸识别。此外,跨品种数据实验表明模型具备良好的泛化能力,能够适应不同牛只面部特征差异。实验结果表明,该方法可以准确识别牛只身份,优于现有算法。

未来的研究工作将使用更多的品种来扩展样本数据,以提高模型跨品种的泛化能力。针对牛只生长过程中面部形态变化,可以引入时序建模,提升长期身份一致性判断。同时,可以结合边缘计算设备优化模型部署性能,推动其在智慧牧场中的实际应用。

参考文献:

[1] Mahmud M S, Zahid A, Das A K, et al. A systematic literature review on deep learning applications for precision cattle farming[J]. Computers and Electronics in Agriculture, 2021, 187: 106313.

[2] Kawagoe Y, Zin T T, Kobayashi I. Individual identification of cow using image processing techniques[C]//2022 IEEE 4th Global Conference on Life Sciences and Technologies(LifeTech). Osaka, Japan: IEEE, 2022: 570-571.

[3] Yin M S, Ma R Q, Luo H L, et al. Non-contact sensing technology enables precision livestock farming in smart farms[J]. Computers and Electronics in Agriculture, 2023, 212: 108171.

[4] 刘晋维, 李富忠, 陈新文, 等. 基于孪生网络和度量学习的牛脸检测与识别模型[J]. 计算机技术与发展, 2025, 35(9): 175-181.

[5] Zhang M J, Wang X P, Feng H H, et al. Wearable internet of things enabled precision livestock farming in smart farms: A review of technical solutions for precise perception, biocompatibility, and sustainability monitoring[J]. Journal of Cleaner Production, 2021, 312: 127712.

[6] 杨梅, 赵建敏. 基于三元损失的保险业务牛脸识别系统的设计[J]. 光电子·激光, 2022, 33(8): 831-839.

[7] Xu X S, Wang Y F, Hua Z X, et al. Light-weight recognition network for dairy cows based on the fusion of YOLOv5s and channel pruning algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39

- (15):153-163.
- [8] 刘和鑫,尚文婧,赵翔宇,等. SpaMGCN:基于多视图图神经网络识别空间域的方法[J]. 内蒙古大学学报(自然科学版),2025,56(5):545-557.
- [9] 宋静,苟学义,梅佳慧,等. 基于SPEI的多尺度法分析干旱气候对黄河流域内蒙古段植被覆盖的影响[J]. 内蒙古大学学报(自然科学版),2025,56(5):463-475.
- [10] Baranov A, Graml R, Pirchner F, et al. Breed differences and intra-breed genetic variability of dermatoglyphic pattern of cattle[J]. Journal of Animal Breeding and Genetics, 1993, 110(1/2/3/4/5/6):385-392.
- [11] Allen A, Golden B, Taylor M, et al. Evaluation of retinal imaging technology for the biometric identification of bovine animals in Northern Ireland[J]. Livestock Science, 2008, 116(1/2/3):42-52.
- [12] Gaber T, Tharwat A, Hassanien A E, et al. Biometric cattle identification approach based on Weber's local descriptor and AdaBoost classifier[J]. Computers and Electronics in Agriculture, 2016, 122:55-66.
- [13] Noviyanto A, Arymurthy A M. Beef cattle identification based on muzzle pattern using a matching refinement technique in the SIFT method[J]. Computers and Electronics in Agriculture, 2013, 99:77-84.
- [14] Yao L Y, Hu Z, Liu C X, et al. Cow face detection and recognition based on automatic feature extraction algorithm [C]//Proceedings of the ACM Turing Celebration Conference-China. New York, USA:ACM, 2019:1-5.
- [15] Kumar S, Singh S K. Cattle recognition: A new frontier in visual animal biometrics research[J]. Proceedings of the National Academy of Sciences, India Section A:Physical Sciences, 2019, 90(4):689-708.
- [16] 杨蜀秦,刘杨启航,王振,等. 基于融合坐标信息的改进YOLO V4模型识别奶牛面部[J]. 农业工程学报, 2021, 37(15):129-135.
- [17] Lu Y Z, Weng Z, Zheng Z Q, et al. Algorithm for cattle identification based on locating key area[J]. Expert Systems with Applications, 2023, 228(11):120365.
- [18] Wang Y F, Xu X S, 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.
- [19] 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.
- [20] Tu Z Z, Talebi H, Zhang H, et al. MaxViT: Multi-axis vision transformer[DB/OL]. (2022-09-09)[2025-05-29]. <https://arxiv.org/abs/2204.01697>.
- [21] Yu W H, Luo M, Zhou P, et al. MetaFormer is actually what you need for vision[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA:IEEE, 2022:10809-10819.
- [22] Li J S, Xia X, Li W, et al. Next-ViT: Next generation vision transformer for efficient deployment in realistic industrial scenarios[DB/OL]. (2022-08-16)[2025-05-29]. <https://arxiv.org/abs/2207.05501>.
- [23] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[DB/OL]. (2021-06-03)[2025-05-29]. <https://doi.org/10.48550/arXiv.2010.11929>.
- [24] Tolstikhin I, Houlsby N, Kolesnikov A, et al. MLP-Mixer: An all-MLP architecture for vision[DB/OL]. (2021-06-11)[2025-05-29]. <https://doi.org/10.48550/arXiv.2105.01601>.

- [25] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA:IEEE, 2016:770-778.

(责任编辑 黄宝琦)

## Swin-BiCAS: A Cattle Face Recognition Method Based on Sparse Attention and Spatial-Channel Fusion

LIU Yuchen<sup>1</sup>, ZHAO Chanchan<sup>2</sup>, XU Sixuan<sup>1</sup>, XIAO Yan<sup>3</sup>, SHI Bao<sup>4</sup>

(1. *College of Intelligent Science and Technology, Inner Mongolia University of Technology, Hohhot 010080, China;*

2. *School of General Education, Inner Mongolia University of Technology, Hohhot 010080, China;*

3. *School of Economics and Management, Inner Mongolia Normal University, Hohhot 010022, China;*

4. *School of Information Engineering, Inner Mongolia University of Technology, Hohhot 010051, China*)

**Abstract:** As cattle farming gradually transitions toward refined and intelligent management, achieving rapid and accurate individual identification of cattle has become a critical technological bottleneck in modern livestock management. Traditional contact-based identification methods suffer from high costs and susceptibility to tampering, while existing non-contact visual recognition technologies exhibit limited feature extraction capabilities for cattle faces and lack robustness in complex farming environments. To address these challenges, a novel cattle face recognition method named Swin-BiCAS is proposed. Specifically, a BiFormer Block is integrated into the Swin Transformer framework, which employs a sparsified Bi-Level Routing Attention mechanism to effectively compensate for the limitations of traditional window-based attention in modeling long-range dependencies. This design enhances fine-grained feature extraction while reducing computational complexity. Furthermore, a feature fusion module named CAS-merge is proposed to enhance feature representation by jointly optimizing spatial awareness and inter-channel relationships. This enables the model to maintain sensitivity to spatial details while effectively modeling channel-wise correlations. Experimental results demonstrate the effectiveness of the proposed method, achieving an accuracy of 99.1%.

**Key words:** individual identification; Swin Transformer; attention mechanism; cow face recognition; feature fusion