智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

# 卧床失能病人面部表情识别方法研究

朱博文, 吴永明, 刘洋

引用本文:
朱博文, 吴永明, 刘洋. 卧床失能病人面部表情识别方法研究[J]. 智能系统学报, 2026, 21(4): 952-962.
ZHU Bowen, WU Yongming, LIU Yang. Research on facial expression recognition method for bedridden patients[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(4): 952-962.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202510015

DOI: 10.11992/tis.202510015

# 卧床失能病人面部表情识别方法研究

朱博文, 吴永明, 刘洋
(广东工业大学 机电工程学院, 广东 广州 510006)

**摘　要:** 针对卧床失能病人因面部僵硬、目光呆滞、表情钝化等因素, 导致传统人脸识别方法难以准确有效捕捉病人复杂面部表情特征与变化的问题, 研究一种基于 YOLOv11 的卧床失能病人面部表情识别方法。该方法针对性设计高效上卷积模块解决特征捕捉适配性不足问题、轻量特征融合模块强化多尺度特征复用、C2PSA-CAA(C2 position-sensitive attention-context-aware attention)检测头模块提升复杂遮挡场景鲁棒性, 通过三级模块协同优化检测准确性。对本文方法进行验证, 实验结果显示: 在公开数据集 RAF-DB 上, 平均精度均值 mAP50 和 mAP50-95 分别达到 87.83% 和 87.83%, 与 YOLOv11 相比, 分别提高 8.46% 和 8.48%, 参数量仅为 $2.8\times10^6$; 在自建卧床失能病人数据集上, mAP50 和 mAP50-95 分别达到 87.21% 和 56.52%, 与 YOLOv11 相比, 分别提高 3.21% 和 2.67%。实验结果表明, 本文方法可更准确地实现卧床失能病人的面部表情识别, 有助于实现卧床失能病人的智能监护。

**关键词:** 卧床失能病人; 面部表情识别; YOLOv11; 特征融合; 深度学习; 注意力机制; 特征融合; 图像分类; RAF-DB 数据集
**中图分类号:** TP391.4　　**文献标志码:** A　　**文章编号:** 1673−4785(2026)04−0952−11

**中文引用格式: 朱博文, 吴永明, 刘洋. 卧床失能病人面部表情识别方法研究 [J]. 智能系统学报, 2026, 21(4): 952–962.**
**英文引用格式: ZHU Bowen, WU Yongming, LIU Yang. Research on facial expression recognition method for bedridden patients[J]. CAAI transactions on intelligent systems, 2026, 21(4): 952–962.**

# Research on facial expression recognition method for bedridden patients

ZHU Bowen, WU Yongming, LIU Yang
(School of Electromechanical Engineering, Guangdong University of Technology, Guangzhou 510006, China)

**Abstract**: To address the challenge that conventional facial recognition methods fail to accurately capture complex facial expression features and dynamics in bedridden patients with disabilities—who often present with facial stiffness, dull gaze, and flattened affect—this study proposes a YOLOv11-based facial expression recognition method specifically tailored for this patient population. This method is specifically designed with three hierarchical modules to collaboratively optimize detection accuracy: the efficient up-convolution block to address the deficiency of adaptability in feature capture, the lightweight feature fusion module to enhance multi-scale feature reuse, and the C2 position-sensitive attention-context-aware attention detection head module to improve robustness in complex occlusion scenarios. Evaluation of the proposed method demonstrated that on the public RAF-DB dataset, it achieved mean average precision (mAP) values of 87.83% for mAP50 and 87.83% for mAP50-95, representing improvements of 8.46% and 8.48% respectively, compared to YOLOv11, with a parameter count of only $2.8\times10^6$. On a self-constructed dataset of bedridden disabled patients, the method achieved mAP50 and mAP50-95 values of 87.21% and 56.52%, corresponding to increases of 3.21% and 2.67% over YOLOv11.The experimental results indicate that the proposed method enables more accurate facial expression recognition in bedridden disabled patients with disabilities, thereby contributing to intelligent care monitoring for this population.

**Keywords**: bedridden patients; facial expression recognition; YOLOv11; feature fusion; deep learning; attention mechanism; feature fusion; image classification; RAF-DB dataset

随着人口老龄化进程的加速, 卧床失能病人群体规模持续扩大。卧床失能病人由于身体机能受限, 难以通过常规方式表达自身的生理不适和

**收稿日期:** 2025−10−15.
**通信作者:** 吴永明. E-mail: ymwu@gdut.edu.cn.

心理状态，使护理工作面临诸多挑战[1]。应用机器视觉技术，准确识别卧床失能病人的面部表情变化，及时预警病人的身体变化和需求，可以为卧床失能病人提供精准且及时的护理服务，优化护理工作效率[2]。

卧床失能病人的面部表情与普通人的差异主要由生理限制、心理状态、药物作用、并发症及社交互动等多种因素共同导致。生理上可能因肌肉萎缩、神经损伤或慢性疼痛表现为表情僵硬、不对称或反应迟缓。心理层面因抑郁、焦虑或长期社交隔离易出现目光呆滞、笑容减少或表情钝化[3]。长期使用镇静剂、肌松药等可能引发眼神涣散或“面具脸”，抗精神病药物会导致异常面部动作。这些因素叠加可能使卧床失能病人的面部表情的协调性、丰富性和自然程度显著降低，还可能存在头部微颤、身体姿势固定导致面部姿态变化范围有限等状况，给病人的面部表情识别带来了全新挑战。且卧床失能病人所处环境特殊，常伴有光照不均、病床周边医疗器械和遮挡物较多等问题[4]，现有人脸视觉识别技术难以获得理想效果，亟待开发具有针对性的解决方案。

针对面部表情识别的研究，在通用场景下已取得一定成果。Ghimire 等[5] 采用基于几何特征的识别方式，通过提取面部器官的位置、形状等几何参数来判断表情。辛静[6] 采用基于外观特征的方法，利用图像的灰度、颜色等信息进行识别。随着深度学习兴起，卷积神经网络（convolutional neural networks, CNN）在面部表情识别领域崭露头角，其强大的特征自动提取能力，相较于传统方法显著提升了识别准确率[7-8]。黎克迅等[9] 提出的模型，采用 Transformer 架构，结合自适应模块和损失函数，能更好整合信息、减少光照干扰，在低光人脸数据集上性能更优，尤其极端低光下精度提升显著。Ma 等[10] 将结合 YOLO(you only look once) 技术的高效检测与 Mamba 的线性复杂度优势，无需预处理直接输入含背景的原始图像，实现面部表情的检测与分类。Chen 等[11] 提出的动态特征补偿方案，在处理极微弱表情时仍需依赖复杂计算，难以适配资源受限场景。

综上所述，现有主流面部表情识别方法在卧床失能病人场景中存在三大核心局限：1）特征捕捉适配性不足，传统几何特征法依赖面部器官的规则形态与明显位移，而卧床失能病人因肌肉萎缩导致表情幅度缩小，关键特征点位移微弱，该类方法易出现特征误判；2）场景鲁棒性欠缺，主流面部表情识别模型虽优化了低光环境适应能力，但未考虑病床周边医疗器械遮挡，如氧气管、输液架遮挡面部、头部微颤导致的模糊等特殊干扰；3）样本适配性缺失，现有公开数据集以健康人群表情为主，其表情丰富度、动态范围与卧床失能病人存在显著差异，导致模型迁移至目标场景时出现“域偏移”。

针对上述三大局限，本文通过模块化设计实现精准解决：针对特征捕捉适配性不足，设计高效上卷积模块 (efficient up-convolution block, EUCB)，强化微弱表情细节的提取与表征，解决卧床失能病人表情幅度小、特征点位移微弱的问题；针对场景鲁棒性欠缺，构建 C2PSA-CAA(C2 position-sensitive attention-context-aware attention) 检测头模块，融合位置敏感注意力与上下文感知机制，增强对遮挡区域的特征聚焦，适配医疗场景的复杂干扰；针对样本适配性缺失，一方面自主构建卧床失能病人专用表情数据集，另一方面通过通用反向瓶颈（universal inverted bottleneck, C2f-UIB）模块轻量特征融合模块提升模型对小样本、异质样本的特征复用能力，缓解域偏移问题。在此基础上，本文对 YOLOv11 算法进行改进，提出了一种针对卧床失能病人面部表情识别算法 SAFE(sparse weak-face augmented feature enhanced)-YOLO，可准确有效地捕捉病人复杂的面部表情特征与变化。

## 1　数据集的采样与处理

### 1.1　数据采集

本研究的数据来源主要分为公开数据集以及自建数据集。因为没有可用的卧床失能病人面部表情识别的数据集，故自建一个数据集。自建数据集聚焦于公开可用的网络开源资源平台，包括但不限于部分医疗研究机构公开的演示素材库、志愿者在严格匿名和知情同意前提下贡献的公开图片集、特定针对失能病人群体关怀的技术演示数据库。通过精心筛选与主题高度相关的关键词，例如“卧床”“失能老人”“长期护理”“康复表情”等，严格遵守各平台的数据使用条款，收集了原始图像样本共计 1 996 张。经过数据增强方法扩展至 19 956 张，数据集分为训练集样本 14 393 张，测试集样本 5 563 张。公开数据集 RAF-DB[12] 是专门应用于情感和表情识别的人脸表情数据库，表情分类为愤怒、厌恶、恐惧、快乐、中性、悲伤和惊讶，大约包含 30 000 张照片，主要包括 12 271 张训练数据集以及 3 068 张测试数据集。

### 1.2　匿名化处理

鉴于自建数据集涉及敏感的个人生物识别信息且研究对象为特殊群体，确保数据隐私保护和符合伦理规范至关重要。为实现人脸面部表情的匿名化处理，基于 dlib（Davis King's LIBrary）[13] 中的人脸识别功能构建了一个面部特征点检测与区域模糊系统。该方法首先利用预训练的 68 点人脸特征点检测模型定位面部双眼、眉毛、鼻子与嘴巴等关键区域，通过凸包计算与形态学膨胀构建特征区域掩码；进而对不同区域实施分级高斯模糊：对非人脸背景进行高强度模糊，整个人脸区域进行中度模糊，而对面部表情关键区域则采用较低强度模糊，在保护身份信息的同时尽可能保留表情结构特征。最终输出图像统一调整至 640 像素×640 像素分辨率，以保持数据尺度一致性。

### 1.3　数据预处理

为提升模型的泛化能力和鲁棒性，并弥补自建数据库的初始数据量相对有限性，在图像匿名化处理的基础上进行数据预处理工作。采用多样化的图像变换技术对已匿名化的原始样本 1 996 张进行大规模扩增，核心操作包含 3 部分。1）①几何变换类增强，随机水平翻转、±15 °内小角度随机旋转、±10% 范围内的平移与缩放，模拟病人头部轻微晃动的真实场景；②像素调节类增强，随机亮度、对比度、饱和度调整，添加高斯噪声，适配医疗场景中光照不均、设备成像噪声等问题；③色域变换类增强，随机灰度化，提升模型对不同成像设备的适应性。2）标准化：所有增强后的图像均被统一缩放至 640 像素×640 像素固定尺寸，并将像素值线性归一化到 [0,1] 或标准正态分布范围，以加速模型收敛并提升训练稳定性。3）质量筛选与校验：在整个预处理过程中，同步进行质量筛选，剔除任何因匿名化或增强过程导致面部关键区域严重扭曲、模糊过度以至于表情语义完全丢失，确保最终数据集中所有图像均清晰传达了可识别的情感表达状态。将失能病人面部表情分为积极、中立和消极三大类，如图 1 所示。

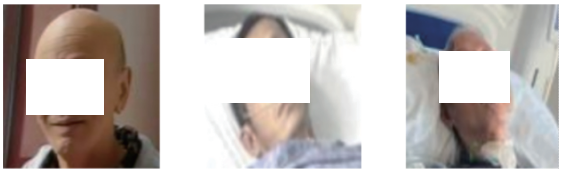


**图 1　自建数据库的表情分类**

**Fig. 1　Expression classification in the self-built database**

本分类主要基于临床护理的实际需求设计：卧床失能病人因生理、心理、药物因素导致表情钝化，难以表现出普通人的精细化情绪，临床护理中核心需快速判断病人的情绪倾向，为护理干预提供初步依据；精细化情绪的识别需结合生理信号、肢体动作等多模态信息，本研究为单模态视觉识别，故先实现粗粒度分类，后续将结合多模态数据开展细粒度情绪识别研究。

## 2　卧床失能人面部表情识别模型

### 2.1　改进 YOLOv11 卧床失能病人面部表情识别模型

YOLOv11 是一种端到端深度学习模型，可直接在原始图像中检测、定位目标并进行分类，其核心架构基于 CNN，通过多尺度预测和网络分割技术高效学习物体特征，兼顾检测精度与实时性。YOLOv11 模型由 4 个模块组成：输入端（Input）负责图像预处理、骨干网络（Bankbone）提取深层特征、颈部网络（Neck）融合多尺度信息、检测头（Head）输出目标的位置与类别。

YOLOv11 作为最新一代 YOLO 系列模型，在通用目标检测领域表现出显著优势：1）检测效率突出，采用深度可分离卷积与动态锚框匹配机制，推理速度得到大幅度提示，适合资源受限的护理设备部署；2）多尺度特征融合能力较强，通过颈部网络的自适应特征金字塔结构，可有效捕捉不同尺寸目标；3）轻量化设计均衡，基础版本模型大小仅 5.22 MB，便于嵌入式设备集成。但将其直接应用于卧床失能病人表情识别时，仍存在固有局限：1）上采样模块采用简单插值法，导致微弱表情细节，如五官边缘、肌肉纹理丢失，边缘特征模糊，需要增强局部细节的提取能力；2）特征融合模块对低冗余特征的筛选能力不足，难以区分面部生理缺陷与动态表情特征；3）检测头缺乏针对局部遮挡的注意力机制，对病床场景中高频出现的局部遮挡样本识别准确率仅为 82% 左右，无法满足护理需求。

为解决这些问题，设计一种在 YOLOv11 模型基础上改进的卧床失能病人面部表情识别模型，称为 SAFE-YOLO 模型，具体改进措施有：1）采用 EUCB[14]，通过提升空间分辨率增强了细节识别能力；2）将 UIB 模块[15] 与 Mobile MQA（mobile multi-query attention）模块相结合形成轻量化卷积模块 C2f-UIB，C2f-UIB 模块通过优化特征融合过程提高了模型精度；3）将注意力机制模块 (context-aware attention, CAA)[16] 与 C2PSA 模块相结合形成检测头模块 C2PSA-CAA，该模块通过增强特征提取能力改善了模型对复杂医疗场景的适应性。SAFE-YOLO 的网络结构如图 2 所示。

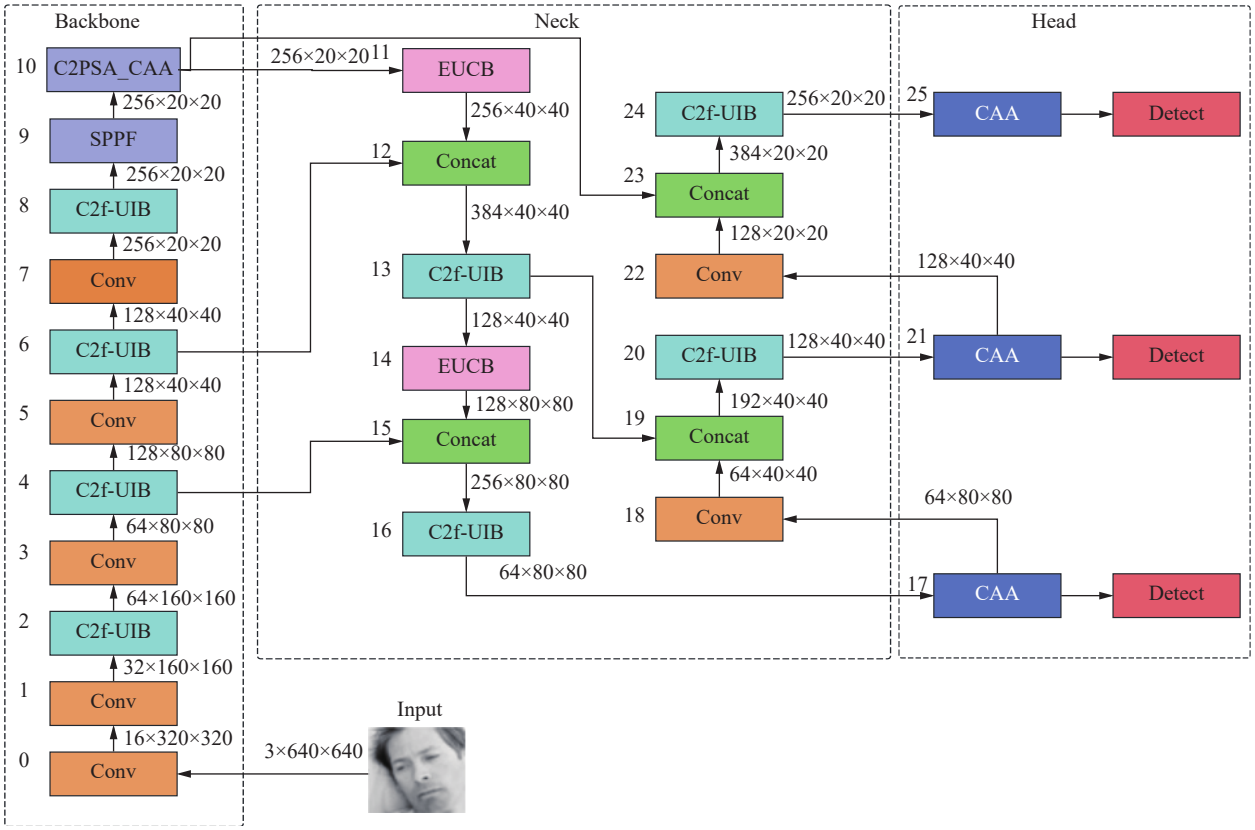


**图 2　SAFE-YOLO 的网络结构**

**Fig. 2　Network architecture of SAFE-YOLO**

## 2.2　上采样模块 EUCB 的设计

在卧床失能病人面部表情的识别中，面部表情细节难以分辨，且因背景复杂容易混淆或忽略，导致误判或漏判，不满足医疗环境下的严谨性要求。YOLOv11 的上采样模块采用 nn.Upsample 模块，该模块无额外参数，减轻了模型的计算量及内存占用，但在自适应训练时不能调整特征，特别是采用的最近邻近插值会导致特征图的边缘锯齿化，双线性插值会导致高频细节模糊，影响检测精度。为解决此问题，采用 EUCB，通过结合多尺度深度可分离卷积（depthwise separable convolution）、通道混洗（channel shuffle）和点卷积（pointwise convolution），实现特征图的空间分辨率提升。EUCB 摒弃传统参数密集的转置卷积（transposed convolution），采用双线性插值与深度可分离卷积组合。理论上，深度卷积的计算复杂度为 $O(D_K \cdot D_K \cdot M \cdot D_F \cdot D_F)$，而标准卷积为 $O(D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F)$，（其中 $M$ 为输入通道数，$N$ 为输出通道数），当 $N \gg M$ 时，计算量可降低近 $1/N$ 倍，同时通过通道混洗恢复跨通道信息交互，避免特征解耦导致的表达能力下降。EUCB 结构如图 3 所示。

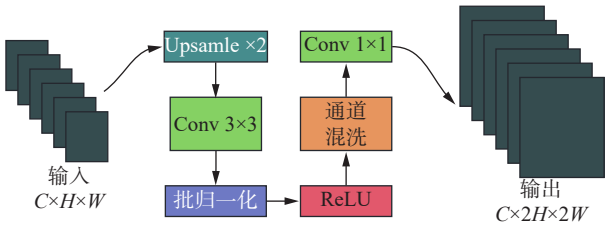


**图 3　EUCB 结构**

**Fig. 3　Structure of EUCB**

EUCB 模块的通过 4 阶段实现高效特征上采样：首先采用双线性插值将特征图分辨率扩大两倍，当给定输入特征图 $\boldsymbol{X}_{\text{up}} \in \mathbb{R}^{C\times 2H\times W}$，通过上采样操作将特征图尺度扩展至原来的两倍 $\boldsymbol{X}_{\text{up}} \in \mathbb{R}^{C\times 2H\times 2W}$：

$$\boldsymbol{X}_{\text{up}} = \text{Upsamle}(\boldsymbol{X}) \in \mathbb{R}^{C\times 2H\times 2W}$$

然后使用 3×3 深度卷积，每组通道独立卷积，进行空间特征提取，接着进行批归一化 (batch normalization, BN) 操作和 ReLU 激活，减少参数量和计算量，计算复杂度仅为标准卷积的 $1/C$：

$$\boldsymbol{Y} = \text{ReLU}(\text{BN}(\text{DWConv}(\boldsymbol{X}_{\text{up}}))$$

其次，引入通道混洗操作以促进跨组信息交互，增强特征融合能力。该操作通过打破分组卷积（group convolution）所带来的通道隔离，使不同

分组内的特征信息得以交互。其核心过程如下：将输入张量 $\boldsymbol{Y}$ 重塑为 $[B\ G\ M\ H\ W]$ 的 5 维形式，其中 $G$ 为分组数，$M=C/G$ 为每组包含的通道数；对分组 $G$ 与组内通道 $M$ 所在的维度进行转置，变换维度顺序为 [0,2,1,3,4]；再次将张量重塑回原始的 4 维形状 $[B\ C\ H\ W]$。此过程可形式化地表述为

$$\boldsymbol{Y}_{\text{shuffle}} = \mathcal{R}(\mathcal{P}(\mathcal{R}(Y,[B\ G\ M\ H\ W]),[0,2,1,3,4]),[B\ C\ H\ W])$$

式中：$\mathcal{R}$表示 reshape 对通道进行重塑，$\mathcal{P}$表现 transpose 转置操作。

最后进行 1×1 逐点卷积动态调整通道权重，形成最终输出特征，点卷积支持通道维度的灵活调整，适配不同任务需求：

$$\boldsymbol{Z} = \text{PWConv}(\boldsymbol{Y}_{\text{shuffle}})$$

### 2.3　C2f-UIB 特征融合机制

YOLOv11 中的 C3K2 模块相较于前的 YOLO 中的 C3 模块使用了 2×2 卷积代替了 3×3 卷积，显著减少了计算量并提升梯度多样性，但对复杂场景检测能力下降，深层特征的抽象能力下降，因此设计了一种 C2f-UIB 模块，在轻量化的情况下，保证检测精度得到提升。在深度卷积神经网络设计中，特征重用和计算效率是提升模型性能的关键因素。C2f-UIB 架构如图 4 所示。

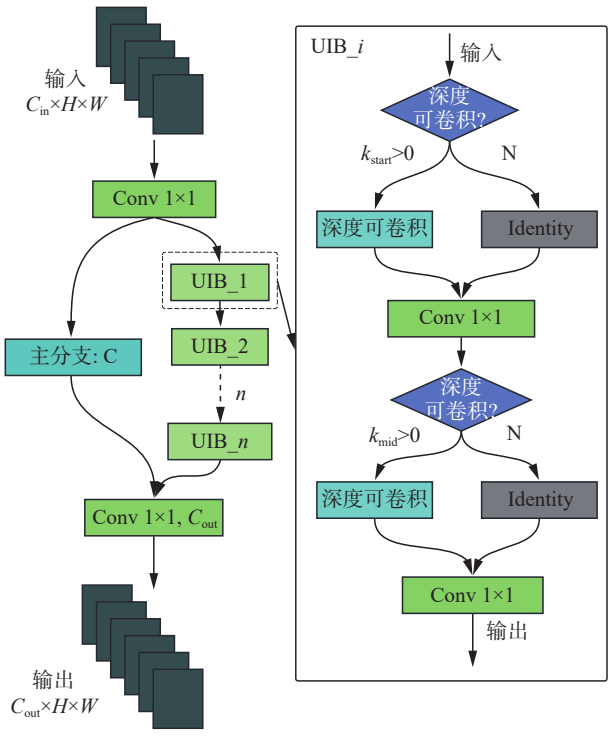


**图 4　C2f-UIB 架构**

**Fig. 4　C2f-UIB architecture**

C2f-UIB 模块结合了跨阶段部分（cross stage partial, CSP）连接模块与 UIB 模块，CSP 结构的优势在于通过跨阶段特征分流与融合，将输入特征分为主分支与次级分支，主分支保留原始特征流以维持梯度连续性，次级分支通过卷积变换提取差异化特征，可有效缓解深度网络中的梯度消失问题，这一特性适配卧床失能病人表情特征的低信噪比特点，能在过滤背景噪声的同时，保留核心表情特征。

UIB 融合了前馈网络[17]、反转瓶颈[18] 以及 ConvNext（convolutional network next generation）[19]，采用窄通道→宽通道→窄通道的设计，在中间层激活高维表征以捕捉微弱表情细节，最终通过 1×1 卷积压缩回原始维度，解决表情特征在跨尺度融合中出现的语义不一致问题，尤其适合整合眼部、嘴部等分散的局部表情特征。UIB 模块如图 5 所示。

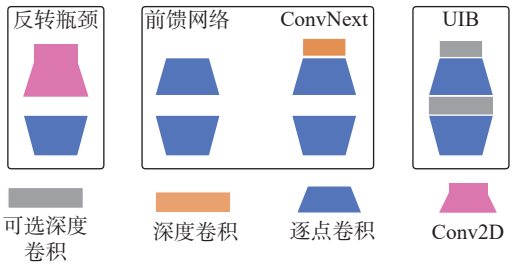


**图 5　UIB 模块**

**Fig. 5　UIB module**

C2f-UIB 模块由特征分割层、多级瓶颈层和特征融合层构成，将输入特征沿通道维度分割为两路，仅对一路进行非线性 UIB 变换，另一路直连保留梯度通路。这种“部分更新”策略大幅减少重复计算。具体实现过程如下：

给定输入特征图 $\boldsymbol{X}\in\mathbb{R}^{C\times H\times W}$。首先采用 1×1 卷积将通道数扩展为 $2C$，随后沿通道维度分割为两个子张量：

$$\boldsymbol{X}_{\text{expand}} = \text{Conv}_{1\times1}(\boldsymbol{X}) \in \mathbb{R}^{2C\times H\times W}$$

$$\boldsymbol{X}_{\text{min}}, \boldsymbol{X}_{\text{side}} = \text{Split}\left(\boldsymbol{X}_{\text{expand}}\right), \quad \boldsymbol{X}_{\text{main}} \in \mathbb{R}^{C\times H\times W}$$

式中：$C=\lfloor 0.5eC_{\text{out}}\rfloor$，$e$ 为通道扩展系数。

然后进行多级瓶颈处理，每个次级分支通过 $n$ 个级联的 UIB 进行特征转换，每个 UIB 单元通过 4 步操作实现空间–通道联合优化：

1）初始深度卷积：当配置参数初始深度卷积 $k_{\text{start}}>0$ 时，采用深度可分离卷积提取局部特征其中 $k_{\text{start}}=\{3,5\}$：

$$\boldsymbol{X}'=\text{DWConv}(\boldsymbol{X},k_{\text{start}},s')$$

$$s' = \begin{cases} 1, & \text{middle\_downsample=True} \\ 2, & \text{其他} \end{cases}$$

其中middle_downsample表示是否进行下采样。

2）通道扩展投影：使用 1×1 卷积将通道扩展至$C_{\exp}=\text{make_divisible}(C\cdot r,8)$，其中 $r$ 为扩展率：

$$\boldsymbol{X}_{\exp}=\text{ReLU6}\left(\text{BN}\left(\text{Conv}_{1\times1}\left(\boldsymbol{X}'\right)\right)\right)$$

3）中间深度卷积：通过配置参数 $k_{\text{mid}}=\{3,5\}$，根据下采样需求调整步长：

$$\boldsymbol{X}_{\text{mid}}=\text{DWConv}(\boldsymbol{X}_{\exp},k_{\text{start}},s'')$$

$$s'=\begin{cases}2, & \text{middle\_downsample=True}\\ 1, & \text{其他}\end{cases}$$

4）通道压缩投影：通过 1×1 卷积恢复通道维度，抑制冗余特征：

$$\boldsymbol{Y}=\text{BN}\left(\text{Conv}_{1\times1}\left(\boldsymbol{X}_{\text{mid}}\right)\right)\in\mathbb{R}^{C\times H''\times W''}$$

5）最后进行跨阶段特征融合，将主分支与所有的 UIB 输出拼接后压缩通道：

$$\boldsymbol{Y}_{\text{out}}=\text{Conv}_{1\times1}\left(\text{Concat}\left(\boldsymbol{X}_{\text{main}},\boldsymbol{Y}_1,\boldsymbol{Y}_2,\cdots,\boldsymbol{Y}_n\right)\right)$$

### 2.4　C2PSA-CAA 模块的设计

C2PSA-CAA 模块通过特征分流与多级注意力增强的协同机制实现高效特征优化，其核心架构围绕 4 个关键参数构建：通道分割比例（$e$）、扩展卷积、位置敏感注意力块（position-sensitive attention, PSA）及堆叠次数（$n$）。各参数通过级联作用形成闭环优化空间。输入特征图首先经过扩展卷积将通道数扩展至原始两倍，随后通过通道分流（Split）操作按比例 $e$ 解耦为基路径与增强路径，基路径保留原始特征信息，增强路径则通过 $n$ 次级联的 PSABlock 进行多尺度注意力精炼，通过通道分流仅对部分通道施加注意力计算，其余通道保持原始流，此外，CAA 模块采用水平/垂直一维卷积替代二维卷积，将空间上下文建模的参数量从 $O(HW)$ 降至 $O(H+W)$。整体结构如图 6 所示，采用分治策略进行通道分割和注意力增强处理。

CAA 模块的运算过程可表示为

$$\boldsymbol{A}(x)=\sigma(\text{Conv}_2(\text{VC}(\text{HC}(\text{Conv}_1(\text{AP}(x))))))$$

其中：$\sigma$ 表示 Sigmoid 激活函数，AP 为采用 7×7 核的均值池化，VC 是核尺寸为 (1,hkernel) 的水平卷积，HC 是核尺寸为 (vkernel,1) 的垂直卷积。

C2PSA_CAA 模块的前向传播过程可形式化为

$$a,b=\text{Split}(\text{Conv}_{1\times1}(x))$$

$$b'=\text{PSABlock}^n(b)$$

$$y=\text{Conv}_{1\times1}(\text{Concat}(a,b'))$$

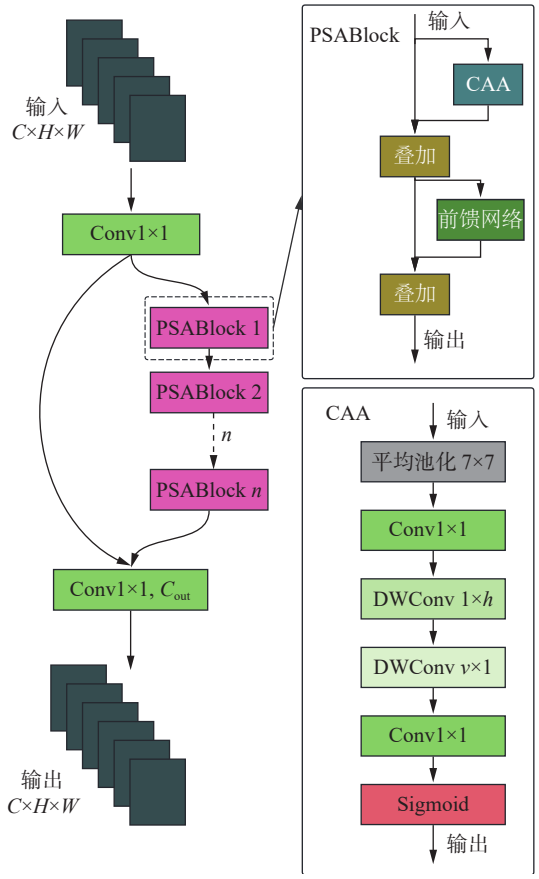


**图 6　C2PSA-CAA 结构**

**Fig. 6　C2PSA-CAA architecture**

## 3　实验结果与分析

### 3.1　实验环境配置

实验在 PyTorch 框架下进行训练，在各个数据集上输入的图片像素为 640 像素×640 像素，迭代周期（epoch）为 300，使用 SGD 优化器使初始学习率为 0.01，权重衰减系数 0.000 5，早停设置为 100，避免出现过拟合以及节省训练时间。实验环境具体配置如表 1 所示。

**表 1　实验环境配置**

**Table 1　Experimental environment configuration**

| 配置名称 | 环境参数 |
|---|---|
| CPU | 13th Gen Intel(R) core(T)i9-13900HX |
| GPU | NVIDIA GeForce RTX 4060 Laptop GPU |
| 内存 | 16 GB |
| Python | 3.9.21 |
| CUDA | 12.8 |
| CuDNN | 8.5.0 |
| PyTorch | 11.7 |

### 3.2　在不同模型下的对比

为全面评估所提算法的性能，将改进的

SAFE-YOLO 模型与 YOLOv5、YOLOv8、YOLOv9、YOLOv10 及 YOLOv11 等多个主流 YOLO 系列模型，在相同的 RAF-DB 面部表情数据集及训练环境下进行了综合对比，结果如表 2 所示。其精确率为 90.09%，$F_1$ 得分为 79.13%，平均精度精度 mAP50 为 87.83%、mAP50-95 为 87.83%，在所有对比模型中均位列第一，综合检测性能最佳。同时，其召回率 79.13% 也处于第二高水平，而模型大小 5.68 MB 控制得当，仅在原始轻量版基础上略有增加。这凸显 SAFE-YOLO 在显著提升检测精度的同时，成功维持了模型的高效率，实现了精度与计算资源的优异平衡。

**表 2　不同模型在 RAF-DB 数据集下的实验结果**

**Table 2　Experimental results of different models on the RAF-DB dataset**

| 模型 | 精确度/% | 召回率/% | $F_1$得分/% | 模型大小/MB | 平均精度均值 mAP50/% | 多尺度精度均值 mAP50-95/% |
|---|---|---|---|---|---|---|
| YOLOV5 | 74.83 | 72.22 | 73.16 | 5.03 | 77.62 | 77.62 |
| YOLOV8 | 80.55 | 77.09 | 78.66 | 5.97 | 82.95 | 92.93 |
| YOLOV9 | 80.75 | **79.33** | 79.92 | **4.43** | 84.12 | 84.11 |
| YOLOV10 | 83.32 | 74.35 | 78.10 | 5.49 | 81.67 | 81.33 |
| YOLOv11 | 74.62 | 72.15 | 73.41 | 5.22 | 79.37 | 79.35 |
| SAFE-YOLO | **90.09** | 79.13 | **84.16** | 5.68 | **87.83** | **87.83** |

注：加粗表示结果在本列中最好。

不同模型在自建数据集下的实验结果如表 3 所示。在自建数据集上的对比实验表明，SAFE-YOLO 模型在多项关键指标上均显著优于其他模型。其精确率达到 87.45%，召回率为 82.45%，$F_1$ 得分 84.87%，平均精度均值 (mAP50) 和多尺度精度均值 (mAP50-95) 分别达到 87.21% 和 56.52%，同时模型大小控制在 5.68MB。与 YOLOv11 相比，SAFE-YOLO 在精确率、召回率和 $F_1$ 得分分别提升 4.11、1.81 和 3.05 个百分点，mAP50 与 mAP50-95 分别提升 3.21 和 2.67 个百分点，而模型大小仅增加 0.46MB。结果表明，SAFE-YOLO 在保持模型轻量化的同时，显著提升了检测精度与综合性能。

**表 3　不同模型在自建数据集下的实验结果**

**Table 3　Comparison of different models on the self-built dataset**

| 模型 | 精确度/% | 召回率/% | $F_1$得分/% | 模型大小/MB | 平均精度均值 mAP50/% | 多尺度精度均值 mAP50-95/% |
|---|---|---|---|---|---|---|
| YOLOV5 | 83.51 | 81.57 | 82.40 | 5.03 | 83.74 | 52.73 |
| YOLOV6 | 82.66 | 80.81 | 81.84 | 8.29 | 84.17 | 52.72 |
| YOLOV8 | 83.51 | 83.33 | 83.42 | 5.97 | 84.50 | 53.10 |
| YOLOV9 | 84.66 | 80.43 | 82.49 | **4.43** | 84.44 | 54.04 |
| YOLOV10 | 77.78 | **83.14** | 80.37 | 5.49 | 84.41 | 54.12 |
| YOLOv11 | 83.34 | 80.64 | 81.82 | 5.22 | 84.00 | 53.85 |
| SAFE-YOLO | **87.45** | 82.45 | **84.87** | 5.68 | **87.21** | **56.52** |

注：加粗表示结果在本列中最好。

为验证 SAFE-YOLO 模型的优势，本文在 RAF-DB 数据集中进行性能评估，与多种先进方法进行对比，包括 MA-Net(multi-scale and local attention network)、PASM、AMP-Net(adaptive multilayer perceptual attention network)、PAT-ResNet-101(probabilistic attribute tree residual network)、MTAC(multi-task assisted correction)、EDGL-FLP (enhanced discriminative global-local feature learning with priority)、DFF(dual-branch feature fusion)、HANet(hierarchical attention network) 和 GA-MDA (granularity-aware multi-dimensional adaptive attention network) 等，结果如表 4 所示。可以看出，SAFE-YOLO 取得了最低的参数量 $2.8\times10^6$，分别比 GA-MDA、MTAC、EDGL-FLP 和 HANet 减少

$3.72\times10^7$、$8.19\times10^7$、$6.81\times10^7$ 和 $6.68\times10^7$; 其参数量仅为 DFF 模型的 12.6%。尽管 SAFE-YOLO 的精确度较 CEPrompt(cross-modal emotion-aware prompting) 低 2.43 百分点, 但其参数量较当前主流方法降低 1~2 数量级, 在模型轻量化方面具有显著优势。实验表明, SAFE-YOLO 能以极低的计算成本实现实时推理, 为资源受限场景下的表情识别任务提供了高效解决方案。

**表 4　RAF-DB 数据集上实验结果对比**

**Table 4　Comparison of experimental results on the RAF-DB dataset**

| 方法 | 参数量/$10^6$ | 精确度/% |
|---|---|---|
| AA-ViT(2024)[20] | 107.44 | 90.66 |
| pACNN(2018)[21] | 138.34 | 83.05 |
| MA-Net(2021)[22] | 50.54 | 88.40 |
| LA-Net(2021)[23] | 15.03 | 87.00 |
| PASM(3 rounds)(2022)[24] | 120.00 | 88.68 |
| AMP-Net(2022)[25] | 105.67 | 89.25 |
| PAT-ResNet-101(2023)[26] | 49.41 | 88.43 |
| MTAC(SwinTrans)(2023)[27] | 84.70 | 90.52 |
| EDGL-FLP (best)(2023)[28] | 70.89 | 89.90 |
| DFF(2024)[29] | 22.23 | 89.60 |
| HANet(2024)[30] | 69.63 | 91.92 |
| CEPrompt(2024)[31] | 86.19 | **92.43** |
| Poster++(2025)[32] | 71.80 | 92.05 |
| GA-MDA(2025)[33] | 40.00 | 92.01 |
| 双流特征交叉融合(2025)[34] | 52.60 | 91.82 |
| SAFE-YOLO | **2.80** | 90.09 |

注: 加粗表示结果在本列中最好。

### 3.3　特征图的可视化测试

借助可视化热力图 CAM(lass activation mapping)[35] 以及其变体 Grad-CAM[36]、CAM++, 对集成各类机制后的空间网络特征图实施了可视化分析。数据集部分样本的可视化热力图及识别结果如图 7 所示, 通过热力图对比了 YOLOv11 模型与 SAFE-YOLO 模型在面部表情识别任务中的特征响应机制。图 7 中系统展示了 3 种基本情绪类别的原型面部图像, 以及 2 种模型分别针对这些图像生成的热力图。热力图采用渐变色谱, 直观呈现模型推理过程中对面部不同区域像素的关注强度。从对比结果来看, SAFE-YOLO 模型生成的热力图普遍表现出更为聚焦的特征响应: 其高激活区域集中分布于眼部、嘴部肌肉等关键表情特征部位, 激活边界相对清晰, 背景及非相关区域的激活程度显著降低。相比之下, YOLOv11 模型的热力图呈现出更弥散的特征响应模式: 高激活区域虽大致覆盖相关面部区域, 但激活值在面部更大范围内呈梯度分布, 且部分图像中背景及面部次要区域存在明显的中等激活甚至少量高激活现象, 表现出更强的激活伪影在识别性能方面, YOLOv11 模型存在将中立类别误识别为消极类别的情况, 而 SAFE-YOLO 模型不仅能实现正确识别, 且在积极和消极样本的识别任务中, 置信度均高于 YOLOv11 模型。综上, SAFE-YOLO 模型在聚焦关键特征区域方面展现出显著优势, 有效降低了噪声干扰。

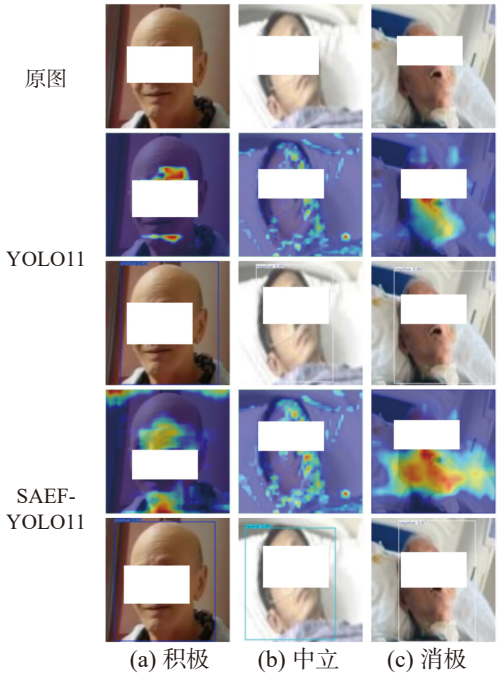


**图 7　数据集部分样本的可视化热力图及识别结果**

**Fig. 7　Visualization heatmaps and recognition results of partial dataset samples**

表 5 给出了基于 YOLOv11 架构 RAF-DB 数据集上的消融实验结果, 通过逐步激活 EUCB、C2f-UIB 和 CSP2A-CAA 等 3 个创新模块评估其性能贡献。实验以基准模型 YOLOv11 为参照, 模块单独激活时显示: CSP2A-CAA 模块显著提升召回率至 78.54% 并达成最优单模块 $F_1$ 值 80.10%, 而 EUCB 模块在均衡性上表现突出, $F_1$ 值为 77.05%; 值得注意的是 C2f-U_IB 模块虽产生最高单模块精度达到 82.62% 却导致召回率异常降至 72.05%, 揭示其可能引发正负样本判别失衡。组合实验中, EUCB 与 CSP2A-CAA 的协同效应最为显著, 相较基准模型实现 $F_1$ 提升近 10 百分点, 该组合在召回率上已超越三模块集成方案。三模块

全集成时精度达到 90.09% 与召回率同时达到 79.13%, mAP50 与 mAP50-95 同时达到 87.83%。在模型效率方面, 所有改进方案均保持轻量化特性, 模型大小为 5.13~6.34 MB, 尤其全模块集成模型大小仅增大 8.8% 即获得 mAP50 8.46 百分点的提升, 验证了架构改进的有效性。

**表 5　RAF-DB 数据集上消融实验结果**

**Table 5　Results of ablation study on the RAF-DB dataset**

| YOLOv11 | EUCB | C2f-UIB | CSP2A-CAA | 精确率/% | 召回率/% | $F_1$得分/% | 模型大小/MB | 参数量/$10^6$ | 平均精度均值 mAP50/% | 多尺度精度均值 mAP50-95/% |
|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | □ | □ | □ | 74.62 | 72.15 | 73.41 | 5.22 | 2.911 | 79.37 | 79.35 |
| ☑ | ☑ | □ | □ | 76.95 | 77.34 | 77.05 | 6.00 | 2.998 | 81.44 | 81.44 |
| ☑ | □ | ☑ | □ | 82.62 | 72.05 | 76.96 | **5.13** | 2.546 | 81.64 | 81.63 |
| ☑ | □ | □ | ☑ | 82.02 | 78.54 | 80.10 | 5.60 | 3.081 | 84.36 | 84.36 |
| ☑ | ☑ | □ | ☑ | 86.86 | **80.10** | 83.40 | 6.34 | 3.167 | 86.67 | 86.66 |
| ☑ | □ | ☑ | ☑ | 83.88 | 82.17 | 82.88 | 5.51 | 2.716 | 86.47 | 86.47 |
| ☑ | ☑ | ☑ | □ | 79.39 | 73.48 | 76.30 | 5.30 | **2.633** | 80.20 | 80.19 |
| ☑ | ☑ | ☑ | ☑ | **90.09** | 79.13 | **84.16** | 5.68 | 2.802 | **87.83** | **87.83** |

注: 加粗表示结果在本列中最好。

## 4　结束语

针对卧床失能病人面部表情识别受光照不均、遮挡多、表情特征复杂等问题, 提出基于 YOLOv11 改进的 SAFE-YOLO 算法, 通过 EUCB 上采样模块、C2f-UIB 特征融合模块及 C2PSA-CAA 检测头模块的创新设计, 在公开数据集 RAF-DB 和自建卧床失能病人面部表情数据集上均实现性能提升, 其中 RAF-DB 数据集上 mAP50 和 mAP50-95 较 YOLOv11 分别提高 8.46 和 8.48 百分点, 自建数据集上分别提高 3.21 和 2.67 百分点, 且模型大小保持 5.68MB 的轻量化优势, 参数量 $2.8\times10^6$, 兼顾识别精度与工程部署效率; 该算法不仅有效解决了卧床失能场景下表情识别的核心痛点, 为护理人员及时察觉病人需求提供技术支撑, 也为医疗场景下细粒度识别任务提供参考范式。后续可通过联合医疗机构扩大数据采集范围、结合视频序列分析捕捉动态表情,开展疼痛、焦虑等细粒度临床情绪的识别研究, 推动智慧护理从“被动响应”向“主动预判”发展。

## 参考文献:


[1] GAYA-MOREY F X, BUADES-RUBIO J M, PALANQUE P, et al. Deep learning-based facial expression recognition for the elderly: a systematic review[EB/OL]. (2025−02−04)[2025−10−15]. https://arxiv.org/abs/2502.02618.

[2] 王信, 汪友生. 基于深度学习与传统机器学习的人脸表情识别综述[J]. 应用科技, 2018(1): 65−72.
WANG Xin, WANG Yousheng. Facial expression recognition based on deep learning and traditional machine learning[J]. Applied science and technology, 2018(1): 65−72.

[3] 陈微, 祁郑晴, 李雪, 等. 失能老人居家护理需求的研究进展[J]. 当代护士 (中旬刊), 2022, 29(11): 1−5.
CHEN Wei, QI Zhengqing, LI Xue, et al. Research progress on home care needs of disabled elderly people[J]. Modern nurse, 2022, 29(11): 1−5.

[4] 李钦云, 宋岳涛. 老年长期照护的国内外现状和展望[J]. 实用老年医学, 2023, 37(1): 83−86.
LI Qinyun, SONG Yuetao. Present situation and prospect of long-term care for the elderly at home and abroad[J]. Practical geriatrics, 2023, 37(1): 83−86.

[5] GHIMIRE D, LEE J. Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines[J]. Sensors, 2013, 13(6): 7714−7734.

[6] 辛静. 基于帧间灰度差的动态表情识别[D]. 天津: 天津大学, 2009.
XIN Jing. Dynamic facial expression recognition based on the gray difference of frames[D]. Tianjin: Tianjin University, 2009.

[7] AVANIJA J, MADHAVI K R, SUNITHA G, et al. Facial expression recognition using convolutional neural network[C]//2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition. Hy-

derabad: IEEE, 2022: 1−7.

[8] JIANG Qiqi, PENG Xiwei, CHEN Hanyu, et al. Facial expression recognition based on residual network[C]// 2022 41st Chinese Control Conference. Hefei: IEEE, 2022: 7000−7006.

[9] 黎克迅, 高治军. LFSepNet: 融合 Transformer 的照明和面部特征解耦人脸识别方法[J]. 计算机工程与应用, 2026, 62(4): 201−209.
LI Kexun, GAO Zhijun. LFSepNet: face recognition method with illumination and facial feature decoupling based on transformer[J]. Computer engineering and applications, 2026, 62(4): 201−209.

[10] MA Hui, LEI Sen, CELIK T, et al. FER-YOLO-mamba: facial expression detection and classification based on selective state space[EB/OL]. (2024−05−03)[2025−10−15]. https://arxiv.org/abs/2405.01828.

[11] CHEN Yangbo, PENG Chunyan. Occlusion-aware facial expression recognition with dynamic feature weighting and block loss optimization[C]//Proceedings of International Conference on Image, Vision and Intelligent Systems 2025. Singapore: Springer, 2026: 197−207.

[12] LI Shan, DENG Weihong, DU Junping. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2584−2593.

[13] SHARMA S, SHANMUGASUNDARAM K, RAMASAMY S K. FAREC: CNN based efficient face recognition technique using Dlib[C]//2016 International Conference on Advanced Communication Control and Computing Technologies. Ramanathapuram: IEEE, 2017: 192−195.

[14] RAHMAN M M, MUNIR M, MARCULESCU R. EMCAD: efficient multi-scale convolutional attention decoding for medical image segmentation[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 11769−11779.

[15] QIN Danfeng, LEICHNER C, DELAKIS M, et al. MobileNetV4: universal models for the mobile ecosystem[M]// Computer Vision–ECCV 2024. Cham: Springer Nature Switzerland, 2024: 78−96.

[16] CAI Xinhao, LAI Qiuxia, WANG Yuwei, et al. Poly kernel inception network for remote sensing detection[C]// 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 27706−27716.

[17] ASHISH V, NOAM S, NIKI P, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Auckland: NIPS, 2017: 6000–6010.

[18] SANDLER M, HOWARD A, ZHU Menglong, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 4510−4520.

[19] LIU Zhuang, MAO Hanzi, WU Chaoyuan, et al. A ConvNet for the 2020s[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 11966−11976.

[20] MOU Xiangwei, XIE Liuping, SONG Yongfu, et al. Dynamic occlusion-aware facial expression recognition guided by AA-ViT[J]. Electronics, 2026, 15(4): 764.

[21] LI Yong, ZENG Jiabei, SHAN Shiguang, et al. Occlusion aware facial expression recognition using CNN with attention mechanism[J]. IEEE transactions on image processing, 2019, 28(5): 2439−2450.

[22] ZHAO Zengqun, LIU Qingshan, WANG Shanmin. Learning deep global multi-scale and local attention features for facial expression recognition in the wild[J]. IEEE transactions on image processing, 2021, 30: 6544−6556.

[23] MA Hui, CELIK T, LI Hengchao. Lightweight attention convolutional neural network through network slimming for robust facial expression recognition[J]. Signal, image and video processing, 2021, 15(7): 1507−1515.

[24] LIU Ping, LIN Yuewei, MENG Zibo, et al. Point adversarial self-mining: a simple method for facial expression recognition[J]. IEEE transactions on cybernetics, 2022, 52(12): 12649−12660.

[25] LIU Hanwei, CAI Huiling, LIN Qingcheng, et al. Adaptive multilayer perceptual attention network for facial expression recognition[J]. IEEE transactions on circuits and systems for video technology, 2022, 32(9): 6253−6266.

[26] CAI Jie, MENG Zibo, KHAN A S, et al. Probabilistic attribute tree structured convolutional neural networks for facial expression recognition in the wild[J]. IEEE transactions on affective computing, 2023, 14(3): 1927−1941.

[27] LIU Yang, ZHANG Xingming, KAUTTONEN J, et al. Uncertain facial expression recognition via multi-task assisted correction[J]. IEEE transactions on multimedia, 2024, 26: 2531−2543.

[28] ZHANG Ziyang, TIAN Xiang, ZHANG Yuan, et al. Enhanced discriminative global-local feature learning with priority for facial expression recognition[J]. Information sciences, 2023, 630: 370−384.

[29] 侯海燕, 谭玉枚, 宋树祥, 等. 头部姿态鲁棒的面部表情识别[J]. 广西师范大学学报 (自然科学版), 2024, 42(6): 126−137.

HOU Haiyan, TAN Yumei, SONG Shuxiang, et al. Head pose-robust facial expression recognition[J]. Journal of Guangxi normal university (natural science edition), 2024, 42(6): 126−137.

[30] TAO Huanjie, DUAN Qianyue. Hierarchical attention network with progressive feature fusion for facial expression recognition[J]. Neural networks, 2024, 170: 337−348.

[31] ZHOU Haoliang, HUANG Shucheng, ZHANG Feifei, et al. CEPrompt: cross-modal emotion-aware prompting for facial expression recognition[J]. IEEE transactions on circuits and systems for video technology, 2024, 34(11): 11886−11899.

[32] MAO Jiawei, XU Rui, YIN Xuesong, et al. POSTER++: a simpler and stronger facial expression recognition network[J]. Pattern recognition, 2025, 157: 110951.

[33] 黎豐玮, 谭玉枚, 宋树祥, 等. 基于注意力引导的遮挡感知面部表情识别[J/OL]. 广西师范大学学报 (自然科学版) (2025−07−28)[2025−10−15]. https://doi.org/10.16088/j.issn.1001-6600.2024120301.

LI Fengwei, TAN Yumei, SONG Shuxiang, et al. Haiyingoocclusion-aware facial expression recognition based on attention guidance[J/OL]. Journal of Guangxi Normal University(natural science edition)(2025−07−28)[2025−10−15]. https://doi.org/10.16088/j.issn.1001-6600.2024120301.

[34] 党宏社, 孟饶辰, 高宛蓉. 基于双流特征交叉融合 Efficient Transformer 的人脸表情识别[J]. 计算机工程与应用, 2025, 61(15): 251−257.

DANG Hongshe, MENG Raochen, GAO Wanrong. Facial expression recognition based on dual-stream feature cross-fusion efficient transformer[J]. Computer engineering and applications, 2025, 61(15): 251−257.

[35] ZHOU Bolei, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discriminative localization[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 2921−2929.

[36] CHATTOPADHAY A, SARKAR A, HOWLADER P, et al. Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks[C]//2018 IEEE Winter Conference on Applications of Computer Vision. Lake Tahoe: IEEE, 2018: 839−847.

## 作者简介：

朱博文，硕士研究生，主要研究方向为机器视觉、智能制造。E-mail: 1643928725@qq.com。

吴永明，博士，教授，主要研究方向为智能制造、绿色制造。主持广东省产学研重点项目 4 项，参与科技部国际合作项目 1 项和省部级科研项目多项，获广东省科技进步二等奖 1 项，获国家专利授权 3 项，发表学术论文 60 余篇。E-mail: ymwu@gdut.edu.cn。

刘洋，硕士研究生，主要研究方向为机器视觉、智能制造。E-mail: 2051248370@qq.com。