

增强FROM的遮挡人脸识别网络

张建川^{1,2}, 万学俊³, 潘文林^{4,5}, 赵周洲³, 蒋 作^{4,5}

- (1. 云南民族大学 电气信息工程学院, 云南 昆明 650500; 2. 云南省无人自主系统重点实验室, 云南 昆明 650504;
3. 玉溪市公安局 科技信息化支队, 云南 玉溪 653100; 4. 云南民族大学 数学与计算机科学学院;
5. 云南省教育厅 智慧交管工程研究中心, 云南 昆明 650500)

摘要: 针对目前主流人脸识别技术在非约束条件场景下性能显著下降的问题, 提出一种增强FROM的遮挡人脸识别网络(EFROM)。首先, 为了更好地利用不同尺度的特征信息, 将SE注意力引入到主干网络LResnet50E_IR之后, 为不同通道特征图赋予不同的权重, 再构建特征金字塔, 提出一种多级特征融合模块MFPF以融合多尺度特征图, 增强模型对不同尺度、形状遮挡的处理能力。其次, 根据人脸是否被遮挡生成特征掩码, 使用特征掩码对人脸特征中被损坏的特征进行抑制, 并利用剩下的人脸特征进行人脸识别。最后, 为保证特征掩码的正确性, 改进了遮挡损失函数以监督特征掩码的生成。理论分析与实验结果表明, EFROM在LFW数据集上的准确率达到99.44%, 与遮挡人脸识别算法(FROM)相比, 准确率提升了2.8%, 万分之一容忍度下正确接受率提升了23.43%。EFROM能够在保证准确进行1:1人脸验证的同时, 拉开不同身份人脸之间的距离, 在非约束条件下具有更好的鲁棒性。

关键词: 遮挡人脸识别; 深度特征抑制; 注意力机制; 特征掩码; 特征融合

DOI: 10.11907/rjdk.251138

扫描二维码阅读全文: 

中图分类号: TP183

文献标识码: A

文章编号: 1672-7800(2026)005-0186-10

Enhanced FROM for Occluded Face Recognition Network

ZHANG Jianchuan^{1,2}, WAN Xunjun³, PAN Wenlin^{4,5}, ZHAO Zhouzhou³, JIANG Zuo^{4,5}

- (1. School of Electrical and Information Technology, Yunnan Minzu University, Kunming 650500, China; 2. Yunnan Key Laboratory of Unmanned Autonomous Systems, Kunming 650504, China; 3. Detachment of Science and Technology Informatization, Yuxi Public Security Bureau, Yuxi 653100, China; 4. School of Mathematics and Computer Science, Yunnan Minzu University; 5. Intelligent Traffic Management Engineering Research Center, Education Department of Yunnan Province, Kunming 650500, China)

Abstract: To address the significant performance decline of current mainstream facial recognition technologies in unconstrained scenarios, we propose an Enhanced FROM for Occluded Face Recognition Network (EFROM). First, to better utilize feature information at different scales, we introduce Squeeze-and-Excitation (SE) attention following the backbone network (LResnet50E_IR) to assign different weights to different channel feature maps. We then construct a feature pyramid and propose a Multi-scale Feature Pyramid Fusion (MFPF) module to fuse multi-scale feature maps, enhancing the model's ability to handle occlusions of varying scales and shapes. Next, we generate feature masks based on whether the face is occluded, using these masks to suppress the damaged features in the facial representation, while leveraging the remaining facial features for recognition. Finally, to ensure the correctness of the feature masks, we improve the occlusion loss function to supervise the generation of the feature masks. Theoretical analyses and experimental results demonstrate that EFROM achieves an accuracy of 99.44% on the LFW dataset, and compared to the occluded face recognition algorithm (FROM) on the LFW occlusion dataset, it shows an accuracy improvement of 2.8%, along with a 23.43% increase in the true acceptance rate at a 1 in 10 000 tolerance level. EFROM effectively maintains accuracy in 1:1 face verification while increasing the distance between different identities, providing better robustness in unconstrained conditions.

Key Words: blocking facial recognition; deep feature suppression; attention mechanism; feature masking; feature fusion

收稿日期: 2025-04-15

基金项目: 国家自然科学基金项目(62462069); 云南省自然科学基金项目(62362071)

作者简介: 张建川(1998-), 男, 云南民族大学电气信息工程学院硕士研究生, 研究方向为计算机视觉; 万学俊(1984-), 男, 玉溪市公安局科技信息化支队副支队长, 研究方向为图像超分辨率重建、大数据和人工智能; 潘文林(1972-), 男, 博士, CCF高级会员, 云南民族大学数学与计算机科学学院教授、硕士生导师, 研究方向为信息建模、软件工程、数据工程等; 赵周洲(1987-), 男, 玉溪市公安局科技信息化支队副支队长、警务技术三级主管, 研究方向为视频监控系统建设、视频结构化和人像识别算法; 蒋作(1979-), 男, 博士, 云南民族大学数学与计算机科学学院副教授、硕士生导师, 研究方向为软件过程、深度学习。

0 引言

目前,人脸识别技术已广泛应用于安全监控、设备解锁和移动支付等领域。现有算法虽在约束条件下表现良好,但在遭遇遮挡(如口罩、眼镜)或光线、姿态变化等非约束场景时,识别性能显著下降。由于现实环境中遮挡具有随机性和区域不确定性,可能出现在人脸任意位置。因此,如何在复杂条件下仍能实现具有鲁棒性的人脸识别成为关键挑战。针对该问题,当前主流的解决方案主要分为两类:基于传统方法的识别系统和基于深度学习的识别模型。在传统方法中,基于子空间的方法和稀疏表示法是最具代表性的技术^[1-2]。这些方法各自具有优缺点,但在面对复杂的现实场景时,仍需作进一步的研究与创新。

近年来,基于深度学习的方法发展迅速。在遮挡人脸识别中,该方法可分为两类:基于深度人脸特征抑制的方法和基于深度人脸特征恢复的方法^[3-5]。Qiu等^[6]提出FROM(End2End Occluded Face Recognition by Masking Corrupted Features)遮挡人脸识别网络,用浅层特征生成特征掩码作用于深层特征以实现深度特征抑制,并引入一个遮挡模式损失函数监督深度特征抑制,可以较好地处理遮挡情况,但只能处理矩形遮挡。Li等^[7]提出一种特征融合模块,使用注意力机制与图结构以合并来自不同图像来源的特征,但模型的计算量和参数量很大。Xu等^[8]使用Haar小波变换将特征图分解为低频和高频部分,通过注意力机制将高频成分与低频成分结合,帮助模型关注重要的特征区域,增强模型的表达能力。然而,其在进行下采样,尤其是处理高频信息时,重要的细节和特征可能会丢失。Ni等^[9]提出一个上下文引导模块,在特征提取阶段嵌入上下文信息,并在空间上重建特征以提高表达能力,从而提升语义分割精度。但该方法需要为图像中的每个像素分配一个标签,计算量很大。Rahman等^[10]提出一种高效上采样卷积块EUCB(Efficient Uncertainty Calibration Block),在内部使用深度可分离卷积提取空间特征与通道特征,但其中的逐通道卷积与逐点卷积之间的联系较少,导致模型学习到的特征不够泛化^[11]。

针对以上问题,本文提出一种抑制遮挡特征的人脸识别网络,具体内容包括:

(1)将通道注意力机制引入到LResNet50E_IR的残差块之后,加强网络的特征提取能力^[12]。

(2)提出多尺度特征融合模块,结合小波变换下采样与EEUCB(Enhance Efficient Uncertainty Calibration Block)上采样,通过双向跨层连接实现自顶向下与自底向上的多层级特征交互,并引入可学习的权重分配机制,增强模型对复杂人脸特征的捕获能力。

(3)设计网格化遮挡模式损失函数,将人脸图像划分为网格单元,通过组合网格表征遮挡区域,将相似遮挡区

域归并为同一模式类别,并采用交叉熵损失监督特征掩码生成。

(4)针对现有遮挡数据集规模小、模式单一的问题,基于Casia-Webface构建包含随机遮挡位置、多样化遮挡物的大规模合成数据集。实验结果表明,该方法在自建数据集和LFW(Labeled Faces in the Wild)基准测试中,识别精度与正确接受率均优于主流模型FROM、SphereFace^[13]、CosFace^[14]、PDSN^[15]、AdaFace^[16]、MobileFaceNets^[17]、ArcFace^[18]等。

1 人脸48关键点对齐图像增强

在实际场景中,图像通常包含多张人脸,并伴有侧头、低头等复杂姿态,严重影响识别性能。通过特征点对齐技术可有效改善此类问题,提升识别精度。具体流程包括:①关键点定位:基于人脸检测结果,自动定位眼、眉、鼻、嘴角等关键特征点;②仿射变换对齐:通过仿射变换将特征点映射至标准模板位置,实现姿态归一化;③标准化输出:生成112×96像素的标准人脸图像。图像预处理结果如图1所示。



Fig. 1 Image preprocessing result

图1 图像预处理结果

2 本文方法

2.1 整体结构

为提升遮挡情况下的人脸识别效果,本文构建一个既能识别无遮挡人脸,又能识别有遮挡人脸的基于深度特征抑制的人脸识别网络。算法的结构框图如图2所示。

本文提出的基于深度特征抑制的遮挡人脸识别网络主要步骤为:①人脸48关键点对齐与裁剪,减少背景的影响

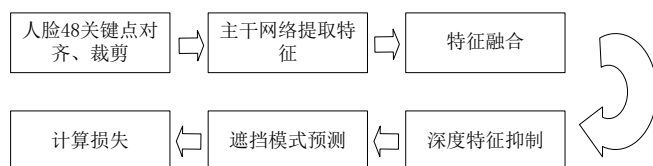


Fig. 2 Algorithm structure block diagram

图2 算法结构框图

响;②主干网络 LResNet50E_IR 进行特征提取;③通道注意力机制结合特征金字塔进行特征融合;④将提取的特征输入深度特征抑制模块,得到抑制之后的特征;⑤使用特

征掩码预测遮挡模式;⑥使用预测的遮挡模式和人脸特征计算损失。

网络的整体结构如图3所示。

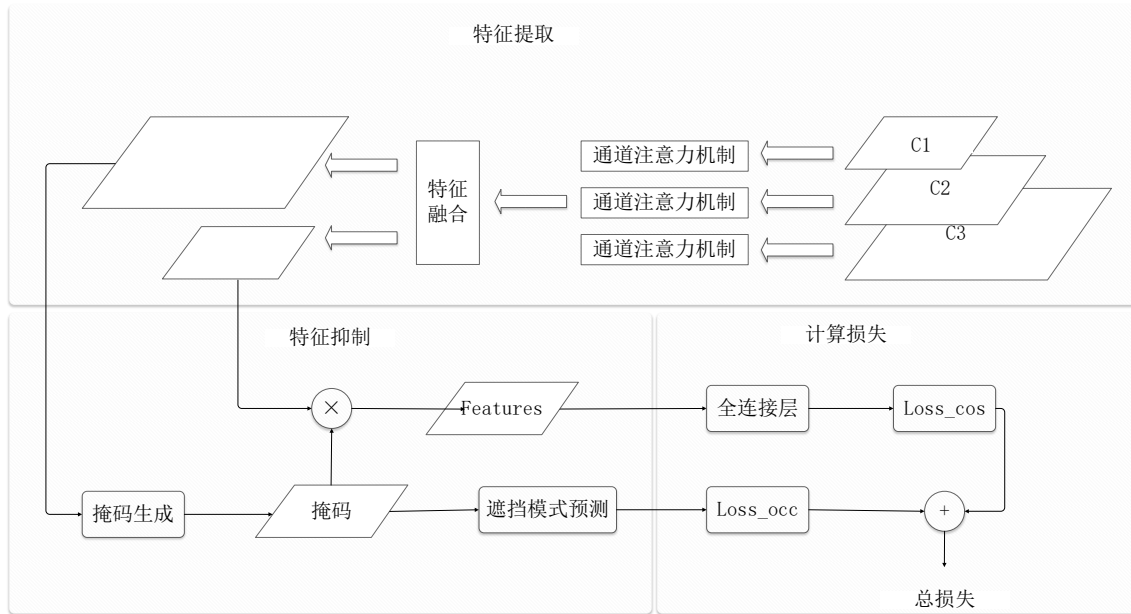


Fig. 3 Global network structure

图3 整体网络结构

在训练流程中,对齐后的人脸图像首先输入 LResNet50E_IR 主干网络,提取多尺度特征图 C1-C3。通过通道注意力机制筛选关键通道特征后,特征融合模块将全局上下文信息与局部细节相结合,生成深层特征 X1 和浅层特征 X2。基于 X2 预测的特征掩码与 X1 进行通道加权,抑制遮挡区域干扰得到优化特征。同时该掩码受交叉熵损失约束,以提升遮挡模式的预测精度。最后采用 Cos-Face 损失函数对优化特征进行度量学习,驱动模型学习更具判别性的人脸表征。各模块的具体设计将在后续详细展开。

2.2 特征提取网络结构

LResNet50E_IR 网络由 5 个 stage 构成,后 4 个 stage 内部包含一个 BlockIR-DS 残差块与多个 BlockIR 残差块。两种残差块的结构如图4、图5所示。最终的 LResnet50E_IR 主干网络结构如图6所示。

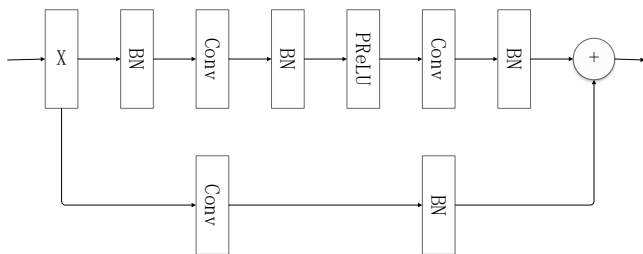


Fig. 4 BlockIR-DS residual structure

图4 BlockIR-DS 残差结构

2.3 通道注意力机制

本文使用融入了 SE (Squeeze-and-Excitation) 通道注意力机制的 LResnet50-IR 网络作为主干网络提取人脸特

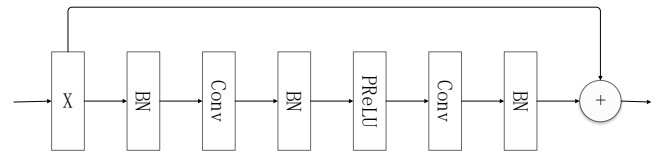


Fig. 5 BlockIR residual structure

图5 BlockIR 残差结构

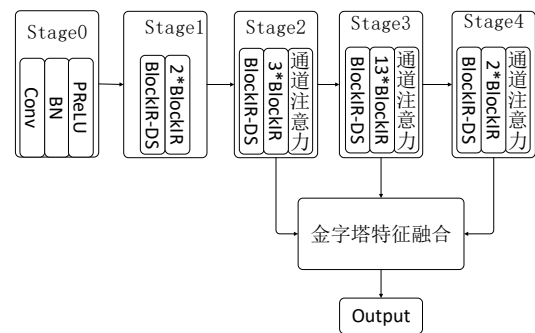


Fig. 6 Feature extraction network structure

图6 特征提取网络结构

征。在每个 Stage 后使用 SE 注意力机制,从而自适应地增强重要的通道特征,抑制次要通道特征,并且不会显著增加计算成本。通道注意力机制结构如图7所示。

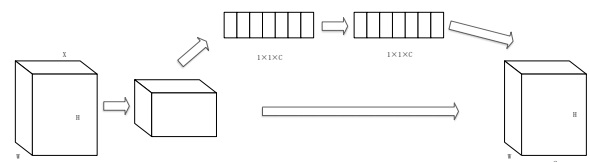


Fig. 7 Structure of channel attention mechanism

图7 通道注意力机制结构

通道注意力机制计算流程如下:

假设 $F_{\text{trans}}: X \rightarrow U$, 其中 $X \in \mathbb{R}^{H' \times W' \times C'}$, $U \in \mathbb{R}^{H \times W \times C}$, $U = [u_1, u_2, \dots, u_C]$, $K = [k_1, k_2, \dots, k_C]$, 每一个元素 k_i 为过滤器, 于是 $u_i = v_i \otimes X$, 其中 $\otimes$ 代表卷积操作。Squeeze 阶段如公式(1)所示, Excitation 阶段如公式(2)所示。

$$z_i = F_{\text{squeeze}}(u_i) = \frac{1}{H_2 * W_2} \sum_x \sum_y u_i(x, y) \quad (1)$$

$$s = F_{\text{ex}}(z, W) = \sigma_{\text{sigmoid}}(W_2 \sigma_{\text{Relu}}(W_1 z)) \quad (2)$$

式中, $z \in \mathbb{R}^C$, $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$, $W_2 \in \mathbb{R}^{\frac{C}{r} \times C}$, $\tilde{X}_C = s \odot u_C$; z_i 表示通道 i 的全局特征向量; F_{squeeze} 表示 Squeeze 操作, 用于从输入特征图中提取全局信息; u_i 表示第 i 个通道的特征图; H_2, W_2 表示特征图的高度和宽度; x, y 表示特征图上的坐标; s 表示通道的重要性分数; F_{ex} 表示 Excitation 操作, 用于生成通道的重要性分数; W_1, W_2 是线性变换矩阵, 用于调整全局特征向量的维度; σ_{sigmoid} 是 ReLU 激活函数; σ_{Relu} 是 Sigmoid 激活函数, 用于将输出值限制在 0~1 范围内; $\tilde{X}_C$ 是经过通道注意力机制处理后的特征图。

2.4 多级特征融合

2.4.1 模块整体架构

本文提出一种专用于融合人脸特征, 基于特征金字塔改进的 MFPP (Multi-scale Feature Pyramid Fusion) 模块, 包含双向融合路径, 最终输出具备空间、频域双模态的特征表示, 分别服务于人脸识别 (X2) 和遮挡检测 (X1)。MFPP 模块结构如图 8 所示。

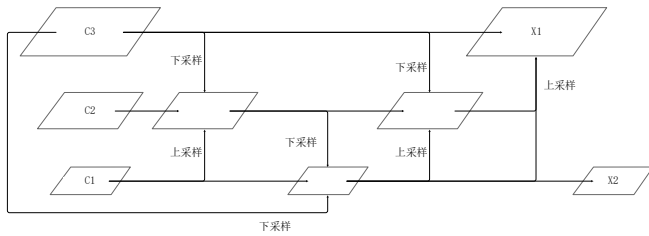


Fig. 8 Structure of feature fusion module

图 8 特征融合模块结构

2.4.2 模块对比

本文的 MFPP 模块相比 FPN^[19] (Feature Pyramid Net-

works)、PANet^[20] (Path Aggregation Network), 分别在融合方向、特征交互方式、频域处理、遮挡适应性等方面进行了一定改进, 如表 1 所示。

Table 1 Comparison of feature fusion modules

表 1 特征融合模块比较

改进方面	FPN	PANet	本文 MFPP	核心创新
融合方向	单顶自顶向下	双向浅层增强	任务驱动双向	X1, X2 差异化输出
遮挡适应性	固定尺度	多尺度检测	遮挡模式预测	
频域处理	无	无	小波+EEUCB	上采样结构改进
特征交互方式	逐元素相加	通道拼接	动态门控融合	三级融合+权重学习

2.4.3 关键技术

MFPP 模块继承了特征金字塔的多尺度特性, 但其针对遮挡人脸识别场景作出了 3 项关键改进: ①动态任务解耦: 深度分离识别特征流 (X2) 与遮挡特征流 (X1); ②频域增强设计: 引入小波变换捕捉多分辨率频域特征; ③可逆特征恢复: 通过 EEUCB 模块补偿下采样信息损失。

(1) MFPP 特征融合模块。为更好地利用深层、中层与浅层特征, 本文设计了一种特征融合机制 MFPP, 在每次进行三级特征融合时自适应地学习每级特征的重要程度, 然后通过小波变换下采样提取频域信息, 并使用 EEUCB 上采样弥补下采样丢失的部分信息, 最后通过动态门控融合 3 张不同层级的特征图。特征图对应的权重如公式 (3) 所示。

$$W = \sigma([C_1, C_2, C_3]) \quad (3)$$

其中, W 为权重, σ 为 sigmoid 函数, C 为待融合的特征图, EEUCB 表示上采样, HWT 表示小波变换下采样。三级特征图融合输出如公式 (4) 所示。

$$C_i' = W_{i-1} * \text{EEUCB}(C_{i-1}) + W_i * C_i + W_{i+1} * \text{HWT}(C_{i+1}) \quad (4)$$

MFPP、FPN 与 PANet 聚焦遮挡的能力可视化结果比较如图 9 所示。其中, FPN 无法找到遮挡位置, PANet 会将人脸位置误判为遮挡, 而 MFPP 能够准确分割人脸与遮挡。

(2) 小波一卷积协同下采样。小波变换可以在不同分辨率和尺度上提取图像特征, 这种多尺度特性能够有效地捕捉遮挡物体在不同尺度下的特征, 并且小波变换下采样可以在下采样时尽量保留更多重要信息。小波变换下采样结构如图 10 所示。

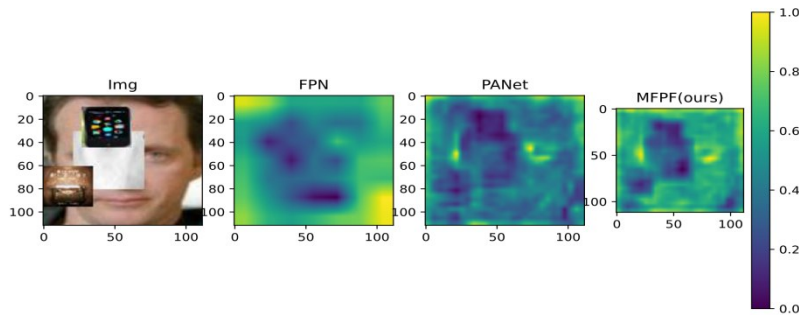


Fig. 9 Comparison of feature fusion effects

图 9 特征融合效果比较

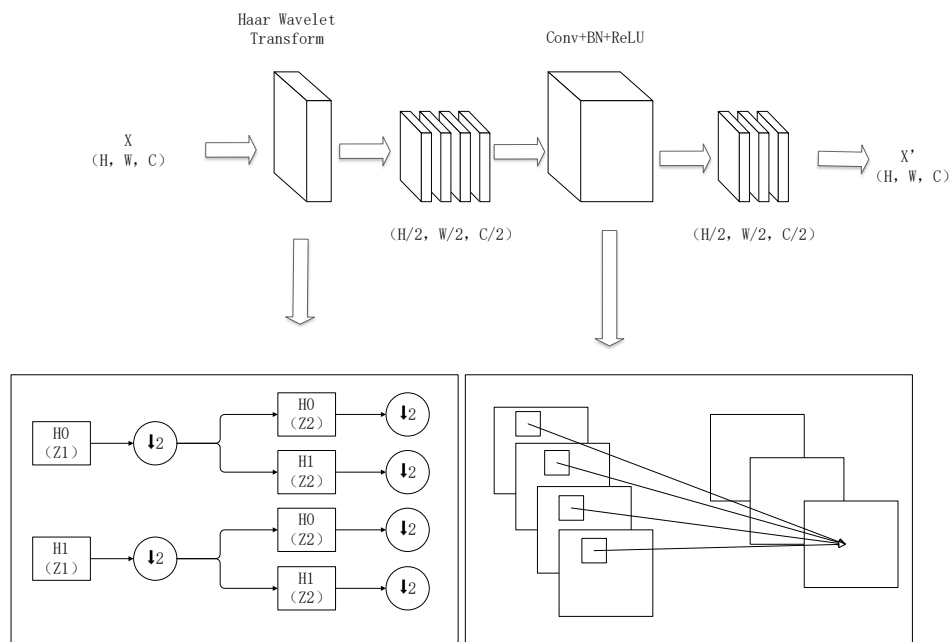


Fig. 10 Wavelet transform down-sampling structure

图 10 小波变换下采样结构

(3)EEUCB上采样模块。由于EUCB上采样使用的常规反卷积会导致棋盘效应,因此本文在其基础上引入空洞深度卷积,以增强EUCB内部逐深度卷积结果与逐点卷积结果之间的联系^[21]。两者的结构对比如图11所示。

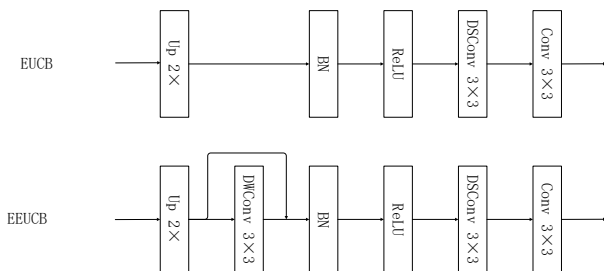


Fig. 11 Comparison of the up-sampled structure

图 11 上采样结构对比

EEUCB与EUCB可视化效果对比如图12所示。引入的空洞卷积增强了局部上下文感知能力,使得模型分割遮挡人脸的能力得到提升。

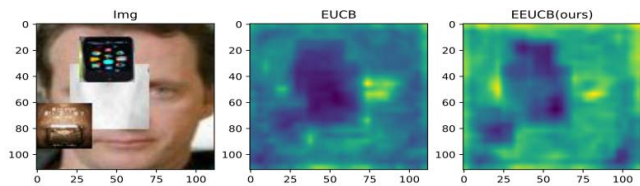


Fig. 12 Comparison of up-sampling effects

图 12 上采样效果对比

2.5 深度特征抑制

人脸图像受到遮挡影响后,主干网络提取的人脸特征里包含了损坏特征,这些损坏特征会干扰人脸识别。因此,提升遮挡人脸识别性能需要尽量减少损坏特征造成的影响。深度特征抑制的目的是从有遮挡的原图特征中准

确获取未被遮挡的特征,其实现方式可以看成一种注意力机制。

2.5.1 掩码生成

掩码生成模块由 Conv-PreLU-BN-Conv-Sigmoid 结构构成,卷积核大小为 3×3 ,步长为 2,从而让掩码与深层特征 X_2 的维度保持一致。最终可生成一个与主干网络提取的人脸特征图维度相同的特征掩码,将其与特征图相乘,从而得到抑制遮挡特征后的特征图。

考虑到模型可能将强光、反光等现象理解为遮挡,而这些现象并没有完全破坏对应位置的人脸特征,所以调整掩码的取值范围为 $0 \sim 1$ 之间,而不是非 0 即 1,以增强模型处理强光、反光现象的能力。将特征融合结果 X_1 输入到掩码生成模块,输出的掩码 $mask$ 维度与 X_2 一致, $mask$ 与 X_2 相乘可达到特征抑制的效果。

2.5.2 遮挡模式

遮挡对人脸识别的影响主要与遮挡物的位置密切相关,相同位置的遮挡无论物体类型如何,对人脸特征的破坏程度具有相似性。因此,可通过位置模式分类实现遮挡分析。为降低计算复杂度,本文采用网格划分策略,将输入图像均等分割为 16 个网格单元(见图 13),以网格为基本单位进行遮挡模式分类。遮挡模式由网格坐标乘积定义遮挡范围,并通过数值编码区分具体类型:单网格遮挡对应 16 种模式,3 个网格单元大小的遮挡有 16 种组合变体(横向与纵向各 8 种)。综合不同尺度的网格组合,最终构建出包含 101 种遮挡模式的分类体系,有效平衡了计算效率与遮挡表征的细粒度需求。

传统基于矩形区域的遮挡表征方法存在几何表征局限,难以适配现实场景中的不规则遮挡形态。为此,本文提出动态网格组合分类法:通过高密度网格划分(10×10 单

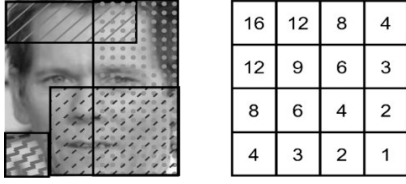


Fig. 13 Occlusion pattern classification diagram

图13 遮挡模式分类示意图

元)构建二进制遮挡编码矩阵(0/1 表征网格遮挡状态),理论上可组合出 2^{100} 种遮挡模式,实现对任意几何形状遮挡的渐进式逼近。该方法在计算资源约束下进行折衷优化,既通过网格细粒度提升覆盖真实遮挡的拓扑适应性(见图14),又通过二进制位运算降低模式匹配的复杂度。最终形成的非规则遮挡表征体系相比矩形方法,遮挡模式的数量提升了3个数量级,显著增强了对复杂遮挡场景的解析能力。



Fig. 14 Occlusion segmentation diagram

图14 遮挡分割示意图

3 损失函数

本文算法使用人脸识别损失 CosFace 与遮挡模式预测损失相结合的损失函数,如公式(5)所示。

$$L = L_{lmc} + L_{occ} \quad (5)$$

3.1 人脸分类损失

Softmax loss 函数是一种经典的人脸识别分类损失函数,如公式(6)、公式(7)所示。

$$L_{softmax} = \frac{1}{N} \sum_{i=1}^N -\log P_i = \frac{1}{N} \sum_{i=1}^N -\log \frac{e^{f_{y_i}}}{\sum_{j=1}^C e^{f_j}} \quad (6)$$

$$f_j = \mathbf{W}_j^T \mathbf{x} = \|\mathbf{W}_j\| \|\mathbf{x}\| \cos \theta_j \quad (7)$$

其中, $\mathbf{W}$ 为全连接层分类权重, $\mathbf{x}$ 为人脸的特征向量, $\mathbf{W}_j$ 为分类类别 j 对应的连接权重, θ_j 为第 j 个类别权重与特征 $\mathbf{x}$ 之间的角度。由于 Softmax 损失函数缺乏对类内和类间距离的度量,因此训练出的模型鲁棒性较差。CosFace 损失函数在原来 softmax 的基础上引入余弦间隔 margin 约束人脸特征,并且归一化了全连接层的权重与人脸特征,从而使特征空间被压缩到半径为 r 的超平面内,以提升遮挡人脸识别效果。所以本文采用 CosFace 作为人脸识别损失函数,如公式(8)所示。

$$L_{lmc} = \frac{1}{N} \sum_i -\log \frac{e^{s(\cos(\theta_{y_i}, j) - m)}}{e^{s(\cos(\theta_{y_i}, j) - m)} + \sum_{j \neq y_i} e^{s(\cos(\theta_{y_i}, j) - m)}} \quad (8)$$

$$\text{其中, } \mathbf{W} = \frac{\mathbf{W}^*}{\|\mathbf{W}^*\|}, \mathbf{x} = \frac{\mathbf{x}^*}{\|\mathbf{x}^*\|}, \cos(\theta_j, i) = \mathbf{W}_j^T \mathbf{x}_i.$$

3.2 遮挡模式损失

仅靠 CosFace 损失函数监督深度特征抑制模块是不够的,因为 CosFace 并没有处理遮挡问题的有效解决方案,所以本文采用 CosFace 损失函数与改进后的遮挡模式损失函数相结合的综合损失函数。遮挡模式损失本质上是一个交叉熵损失函数,可计算真实遮挡位置与预测遮挡位置的损失值。人脸图片被网格化之后,每个网格区域有两种可能(遮挡与被遮挡)。遮挡模式预测模块使用特征掩码作为输入,输出每个网格区域(10*10)对应的两个预测值。遮挡模式预测模块网络结构如图15所示。为了防止过拟合,本文将 Dropout 参数设置为 0.5。

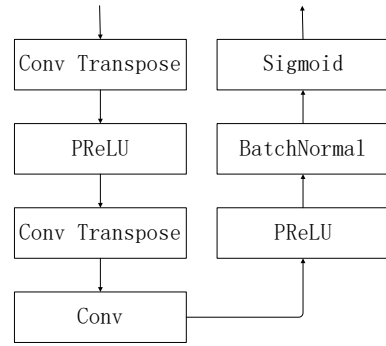


Fig. 15 Network structure of occlusion pattern prediction module

图15 遮挡模式预测模块网络结构

遮挡模式损失函数需要将人脸网格化为 10*10 的区域,对每个区域预测遮挡概率(二分类问题)。损失函数定义为交叉熵,如公式(9)所示。

$$L_{occ} = -\frac{1}{N \times 100} \sum_{i=1}^N \sum_{j=1}^{100} [y_{ij} \log(p_{ij}) + (1 - y_{ij}) \log(1 - p_{ij})] \quad (9)$$

其中, N 表示批量样本数, y_{ij} 表示第 i 个样本中第 j 个网格的真实遮挡标签(0 表示未遮挡, 1 表示遮挡), p_{ij} 表示模块正确预测第 i 个样本中第 j 个网格遮挡情况的概率值。

4 实验方法

4.1 训练配置

本文使用 Ubuntu 22.04.4 LTS 系统下的 Pytorch 进行程序编写,使用 GPU(显卡型号为 NVIDIA GeForce RTX 4060 Ti)训练整个网络。软件环境配置为 Python3.11、Cuda11.8、Torch2.0.0。

4.2 遮挡数据集构建

本文在公开的 WebFace (ASIA-WebFace) 数据集和 LFW 数据集基础上,使用遮挡算法构建遮挡人脸识别数据集 WebFace_Occ 用于训练神经网络,并使用相同的遮挡算法构建 LFW_Occ 用于验证对遮挡人脸识别的性能。为减少图片背景对识别效果的影响,对原始数据集进行人脸 48 关键点对齐并裁剪。裁剪后的 WebFace 数据集有 10 575

个人脸身份,共计 434 455 张图片,LFW 数据集有 5 749 个人脸身份,共计 13 233 张图片,之后使用裁剪好的数据集进行随机的遮挡操作。为了让实验过程更符合现实生活,本文选取 3 种遮挡图片模拟现实中的遮挡情况,分别为眼镜(100 张)、墨镜(100 张)和口罩(95 张)。部分遮挡样例图如图 16 所示。

模拟随机遮挡数据的过程如下:

输入:人脸图片,随机遮挡图片,所有遮挡模式对应的

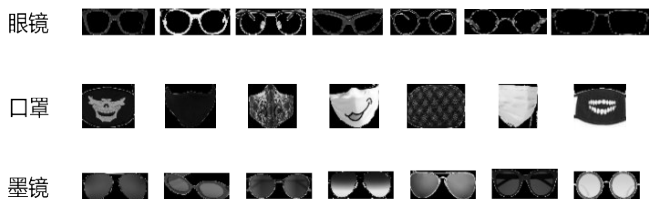


Fig. 16 Samples of partial occlusion

图 16 部分遮挡样例图

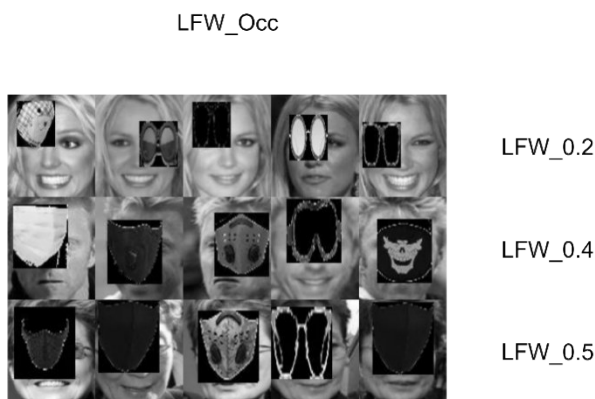


Fig. 17 Partial samples of occlusion dataset

图 17 遮挡数据集部分样本

4.3 评价指标

关于人脸识别的评价指标有很多,本文的实验场景属于 1:N 开集人脸识别,所以本文使用准确率(Accuracy)、正确接受率 TAR(True Acceptance Rate)与错误拒绝率 FAR(False Acceptance Rate)作为评价指标。

准确率是指所有样本中,预测正确的样本占总样本的比例。计算方式如公式(10)所示。

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (10)$$

其中,TP(True Positive)为被正确分类的正例数据,TN(True Negative)为被正确分类的负例数据,FP(False Positive)为被错误分类为正例的负例数据,FN(False Negative)为被错误分类为负例的正例数据。

在指定阈值下,TAR是指被正确预测为正例的样本占所有正样本的比例,FAR是指被错误预测为正例的样本占所有负样本的比例。计算公式如公式(11)所示,本文在 FAR 为 $1e-4$ 的条件下对实验结果进行评价。

$$\text{TAR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \text{FAR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (11)$$

框 box_label,遮挡比例 r(遮挡部分占整张图片的比例)

输出:遮挡后的人脸图片,label_occ

1. 获取的宽、高(w,h)
2. 随机取的宽
3. 重构为(, $r \cdot h \cdot w / \text{occ}_w$)
4. 随机取遮挡中心点(x,y)
5. 计算遮挡区域 box
6. 将粘贴到中对应 box 的区域,得到
7. 计算 box 与所有 box_label 的交并比,将交并比最小的 box_label 作为 label_occ

为了保证模型能同时实现遮挡人脸与无遮挡人脸识别,以 3:1 的比例混合遮挡数据与无遮挡数据构成训练集 Webface_occ。其中,遮挡数据的遮挡比例在 0.1:0.05:0.5 范围内随机选择,保证了训练集中遮挡范围、遮挡位置与遮挡数量的随机性。

所有模拟的遮挡数据如图 17 所示。

4.4 实验结果分析

4.4.1 消融实验

本文在不同程度随机遮挡的 LFW 数据集上进行模型的消融实验,评估了模型各个模块的有效性。其中,Baseline 表示使用 FROM 网络进行实验,Baseline_Loss 表示使用多尺度特征融合增强后的 FROM 进行训练。具体的实验设置如表 2 所示。

Table 2 Ablation experiment settings

表 2 消融实验设置

方法	FROM	Loss	MFPP
Baseline	√		
Baseline_Loss	√	√	
Ours	√	√	√

(1)Baseline。在经过随机遮挡处理后的 WebFace_Occ 数据集上直接训练 FROM 网络,损失函数为 CosFace($m=0.5, s=30$),训练 30 个 epoch,初始学习率设置为 0.1,学习率优化器使用 SGD (Weight decay=0.000 5, Momentum=128),Batchsize 设置为 128。训练 10 个 epoch 后将学习率

衰减为 0.01, 之后每训练 5 个 epoch, 学习率衰减为原来的十分之一, 最后学习率衰减为 0.000 01。

(2)Baseline_Loss。在 WebFace_Occ 数据集上训练, 在 Baseline 模型基础上引入多级特征融合模块的 FROM, 初始学习率设置为 0.1, 每训练 5 个 epoch, 学习率衰减为原来的十分之一, 其他实验设置与 Baseline 一致。

(3)Ours。在 WebFace_Occ 数据集上训练, 同时引入多级特征融合模块与改进后的损失函数, 其他实验设置与 Baseline 一致。

通过表 3 的消融实验结果可知, 本文算法在 LFW_Occ 数据集上的准确率与 TAR 均为最高, 证明了算法各个模块的有效性。

Table 3 Results of ablation experiment on occlusion datasets

表 3 遮挡数据集下消融实验结果				
数据集	指标	Baseline	Baseline_Loss	Ours
LFW_Occ	Acc	91.43	93.75	94.23
	Tar@1e-3	49.70	69.55	75.63
	Tar@1e-4	45.40	61.30	68.43

Baseline 模型在 LFW_Occ 下的准确率为 91.43%, 容忍度在千分之一下的 TAR 为 49.70%, 容忍度为万分之一的 TAR 为 45.4%, 有超过一半同一身份的图像对被预测为负例。一方面, FROM 没有充分利用好浅层特征, 只是简单地使用特征金字塔进行特征融合, 模型不能充分学习遮挡物体的细节特征; 另一方面, FROM 针对矩形遮挡设计的损失函数无法很好地处理非矩形形状的遮挡。

Baseline_Loss 相比 Baseline 准确率提升了 2.32%, 千分之一容忍度下的 TAR 提升了 19.85%, 万分之一容忍度下的 TAR 提升了 15.9%。这是因为 Baseline_Loss 引入改进的遮挡模式损失函数后, 模型学习到如何准确找到非矩形形状的不规则遮挡物体, 从而可以准确判断哪些特征被遮挡影响而导致损坏, 并且在抑制损坏特征的同时尽量保留人脸特征。因此, TAR 值得到大幅提升, 证明了改进后的遮挡模式损失函数的有效性。

Ours 相比 Baseline_Loss 准确率提升了 0.48%, 千分之一容忍度下的 TAR 提升了 6.08%, 万分之一容忍度下的 TAR 提升了 7.13%。这是因为 Ours 引入了 MFPF 模块, 使模型在充分利用深层特征时, 也不会忽略浅层特征的重要性, 同时提升了模型学习遮挡物体细节信息和人脸图片全局信息的能力, 证明了 MFPF 模块对多尺度融合人脸特征的有效性。

综上可得, EFROM 可以使用融合后的深层特征准确进行遮挡人脸识别, 使用融合后的浅层特征生成较为准确的特征掩码, 并利用特征掩码抑制融合后深层特征图中的损坏特征, 提升对遮挡人脸的识别性能。

4.4.2 LFW 对比实验

为进一步证明本文方法的有效性, 将本文方法与部分目前主流的人脸识别算法在公开的 LFW 数据集上进行测试, 结果如表 4 所示。

Table 4 Accuracy comparison of different methods on the LFW dataset

表 4 LFW 数据集上不同方法的准确率比较		
方法	Accuracy/%	训练集大小/M
SphereFace	99.42	0.49
CosFace	99.73	5.0
ArcFace	99.53	0.49
MobileFacenet	99.35	0.49
PDSN	99.20	3.80
FROM	99.38	0.49
MSML	99.40	0.49
FFR_net	99.43	0.98
AdaFace	99.82	5.10
GhostFaceNets	99.87	5.10
Ours	99.44	0.43

结果显示, 本文算法在 LFW 数据集上的准确率为 99.44%, 在使用较少训练数据的情况下超过了部分主流算法。这是因为本文在训练网络时使用的是混合遮挡数据集, 对于没有被遮挡的人脸图片, 网络会生成对原始特征图几乎无影响的特征掩码, 所以本文算法在无遮挡人脸数据集上的测试精度也能达到较高水准。然而, AdaFace、GhostFaceNets 和 CosFace 都取得了更高的测试准确率, 主要原因是使用的训练数据足够多, 增强了模型的泛化能力, 证明了训练数据的规模会对识别准确率产生较大影响^[22]。

4.4.3 LFW 遮挡对比实验

将本文方法和目前主流方法在 LFW_Occ 数据集上进行测试, 实验结果如图 18 所示。

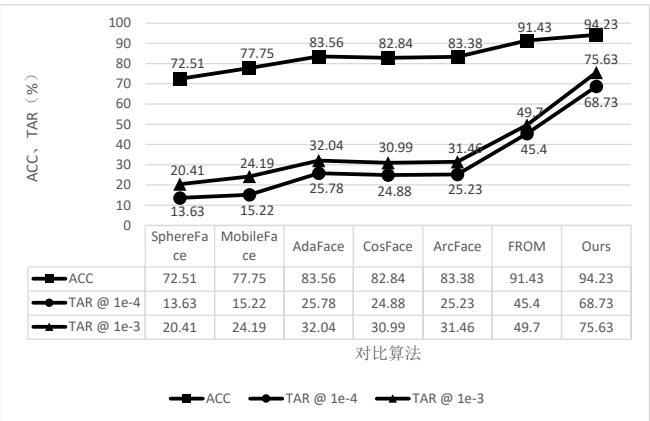


Fig. 18 Comparison of experimental results of different algorithms on the LFW_Occ dataset

图 18 LFW_Occ 数据集上不同算法的实验结果比较

结果显示, 本文算法在 LFW_Occ 数据集上的准确率与 TAR 都是最高的, 分别为 94.23% 和 68.73%, 因此本文算法能够实现遮挡情况下的人脸识别。在对比算法中, FROM 的准确率与 TAR 分别为 91.43% 和 45.4%, 低于本文算法, 这是因为 FROM 擅长处理矩形遮挡, 而面对非矩形遮挡时性能下降。CosFace、ArcFace、AdaFace 都取得了较高的 TAR, 得益于算法本身的优势, 能够识别出遮挡面积较少的人脸图片。SphereFace 与 MobileFace 在本文遮挡数

据集上的效果很差,主要因其本身缺乏处理遮挡的机制,而且更侧重于轻量化,仅靠数据增强的方式处理遮挡难以获得好的效果。

4.4.4 效率—精度平衡分析

EFROM 在无遮挡场景(LFW)和遮挡场景(LFW_Occ)中均展现出优异的性能(LFW ACC: 99.02%, LFW_Occ ACC: 94.23%),但其计算复杂度较高(FLOPs 6.76 GMac, 参数量 47.42 M),推理速度仅为 61.5 ms/帧,难以满足对实时性要求高的场景(如移动端人脸解锁、视频监控)。为平衡模型的效率与精度,本文依次通过剪枝、量化、知识蒸馏等方法优化模型,最终提出压缩方案。

结构化剪枝实验结果如表 5 所示。本文采用通道级结构化剪枝方式,并用通道权重的 L1 范数对通道的重要性进行评估,在剪枝后以 $1e-4$ 的学习率微调 10 个 epoch 以恢复模型性能。

Table 5 Results of structured pruning experiment

表 5 结构化剪枝实验结果

剪枝率	FLOPs	参数量	推理时间	LFW_Occ	TAR@1e-4
/%	/GMac	/M	/ms	ACC/%	/%
0	6.76	47.42	61.5	94.23	68.73
20	5.21	37.94	49.1	94.11	68.12
40	3.87	28.45	37.5	93.88	67.35
60	2.34	18.97	26.1	89.35	60.29

由实验结果可知,剪枝率与精度损失呈非线性关系。当剪枝率为 40% 时,模型体积减少了 40%,推理速度提升了 2.5 倍,精度损失仅为 0.35% (94.23%→93.88%)。但当剪枝率超过 40% 后精度骤降,因此需避免过度剪枝。

量化训练实验结果如表 6 所示。本文分别采用动态

量化(Post-Training Quantization)与量化感知训练(Quantization-Aware Training)两种方法,动态量化从原始 32 位浮点型数据量化为 8 位整型数据,量化感知训练以 $1e-5$ 的学习率训练 20 个 epoch。

Table 6 Quantified experimental results

表 6 量化实验结果

量化方案	模型体积/ MB	推理时间/ ms	LFW_Occ ACC/%	TAR@1e-4/%
FP32(原始)	181	61.5	94.23	68.73
INT8 动态量化	45	24.1	92.37	65.02
QAT 量化感知	47	25.1	93.91	67.29

由实验结果可知,INT8 动态量化使模型体积压缩至 $1/4$,推理速度提升了 1.7 倍,但精度损失较大 (-1.86% ACC)。QAT 通过训练补偿量化误差,使精度损失减少至 0.32%,更适合高精度需求场景。

知识蒸馏实验结果如表 7 所示。本文用完整模型作为教师模型指导轻量化的学生模型(40% 剪枝+QAT 量化感知),在尽量满足实时性的同时,使精度达到最优。

根据上述实验结果,根据场景需求的不同,给出本文算法的场景化推荐方案,如表 8 所示。

Table 7 Experimental results of knowledge distillation

表 7 知识蒸馏实验结果

学生模型	FLOPs	参数量	推理时间	LFW_Occ	TAR@1e-4
	/GMac	/M	/ms	ACC/%	/%
原始模型	6.76	47.42	61.5	94.23	68.73
40% 剪枝+INT8 量化	3.87	28.45	20.5	91.12	64.35
40% 剪枝+QAT	3.87	28.45	25.1	92.34	65.89
20% 剪枝+QAT	5.21	37.94	33.2	93.82	66.47

Table 8 Recommended model compression schemes

表 8 模型压缩推荐方案

场景需求	推荐方案	性能指标	优势分析
实时性优先	40% 剪枝 + INT8 量化	推理时间 20.5 ms, ACC 91.12%	速度最快(3 倍加速),满足移动端实时处理需求(49 FPS)
精度优先	20% 剪枝 + QAT 量化	推理时间 33.2 ms, ACC 93.82%	精度损失最小(仅 0.41%),适合安防等高精度场景
平衡型部署	40% 剪枝 + QAT 量化	推理时间 25.1 ms, ACC 92.34%	速度与精度均衡(2.4 倍加速,精度损失 1.89%)
极端轻量化	学生模型(蒸馏+40% 剪枝)	推理时间 20.5 ms, ACC 91.12%	体积最小(28.45 M),适配资源受限设备

5 结语

本文提出一种增强 FROM 的遮挡人脸识别网络,根据人脸图片被遮挡的程度自适应地生成正确的特征掩码,由特征掩码可抑制被遮挡导致损坏的人脸特征,从而提升遮挡人脸识别效果。为加强模型的人脸特征提取能力,引入通道注意力机制与多级特征融合模块 MFPPF;为提升掩码对损坏特征的抑制能力,改进了遮挡模式损失函数。实验结果表明,本文方法优于目前主流的人脸识别算法,具有较强的鲁棒性。然而,本文方法尚有一些不足之处。例如,图片网格化模拟的遮挡并不能完美表示遮挡位置,总

是会有误差;模型的体积较大,实时性较差。针对以上问题,未来的研究方向一是优化损失函数,以更好地引导网络生成正确的特征掩码;二是优化模拟数据的算法,以更好地模拟现实场景,或直接使用真实数据进行研究。

参考文献:

[1] GUO J B, MA X. Face lmgae recognition method using non-convex mixed norm error coding[J]. Computer Engineering, 2025, 51(3): 208-215.
郭俊博,马祥. 非凸混合范数误差编码人脸图像识别方法[J]. 计算机工程, 2025, 51(3): 208-215.

[2] HE Z H, WANG H, CAO W M. et al. Face recognition based on center bias estimation and adaptive margin [J]. Chinese Journal of Electronics, 2024, 52(8): 2866-2877.
何志浩,王浩,曹文明,等. 基于中心偏差估计和自适应间隔的人脸识

- 别算法[J]. 电子学报, 2024, 52(8): 2866–2877.
- [3] YANG Y L, LI Y, ZHANG Y T, et al. Review of multi-modal research methods for face recognition[J]. Journal of Computer Applications, 2025, 45(5): 1645–1657.
杨雅莉, 黎英, 章育涛, 等. 面向人脸识别的多模态研究方法综述[J]. 计算机应用, 2025, 45(5): 1645–1657.
- [4] LIU H S, NI Z. Face recognition simulation integrating local block reconstruction and linear regression[J], Computer Simulation, 2024, 41(11): 470–474.
刘海松, 倪震. 融合局部分块重构与线性回归的人脸识别仿真[J]. 计算机仿真, 2024, 41(11): 470–474.
- [5] DUAN W T, ZHI M. Ross-age face recognition method incorporating multi-level features [J]. Computer Engineering and Design, 2024, 45(12): 3732–3738.
段文涛, 智敏. 融合多层次特征的跨年龄人脸识别方法[J]. 计算机工程与设计, 2024, 45(12): 3732–3738.
- [6] QIU H B, GONG D H, LI Z F, et al. End2End occluded face recognition by masking corrupted features [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(10): 6939–6952.
- [7] LI X, ZHANG G, CUI H, et al. MCANet: a joint semantic segmentation framework of optical and SAR images for land use classification [J]. International Journal of Applied Earth Observation and Geoinformation, 2022, 106: 102638.
- [8] XU G P, LIAO W T, ZHANG X, et al. Haar wavelet downsampling: a simple but effective downsampling module for semantic segmentation [J]. Pattern Recognition, 2023, 143: 109819.
- [9] NI Z L, CHEN X H, ZHAI Y J, et al. Context-guided spatial feature reconstruction for efficient semantic segmentation [DB/OL]. <https://arxiv.org/pdf/2405.06228>.
- [10] RAHMAN M M, MUNIR M, MARCULESCU R. EMCAD: efficient multi-scale convolutional attention decoding for medical image segmentation [DB/OL]. <https://arxiv.org/pdf/2405.06880>.
- [11] SIFRE L, MALLAT S. Rigid-motion scattering for texture classification [DB/OL]. <https://arxiv.org/pdf/1403.1687>.
- [12] HU J, SHEN L, SUN G, et al. Squeeze-and-excitation networks [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 7132–7141.
- [13] LIU W Y, WEN Y D, YU Z, et al. Sphereface: deep hypersphere embedding for face recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 212–220.
- [14] WANG H, WANG Y T, ZHOU Z D, et al. CosFace: large margin cosine loss for deep face recognition [DB/OL]. <https://arxiv.org/pdf/1801.09414>.
- [15] SONG L X, GONG D H, LI Z F, et al. Occlusion robust face recognition based on mask learning with pairwisefifferential siamese network [C]// 2019 IEEE/CVF International Conference on Computer Vision, 2019: 773–782.
- [16] KIM M, JAIN A K, LIU X. AdaFace: quality adaptive margin for face recognition [DB/OL]. <https://arxiv.org/pdf/2204.00964>.
- [17] CHEN S, LIU Y, GAO X, et al. Mobilefacenet: efficient cnns for accurate real-time face verification on mobile devices [DB/OL]. <https://arxiv.org/pdf/1804.07573>.
- [18] DENG J K, GUO J, XUE N 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, 2019: 4690–4699.
- [19] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117–2125.
- [20] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759–8768.
- [21] ODENA A, DUMOULIN V, OLAH C. Deconvolution and checkerboard artifacts [J]. Distill, 2016(10): e3.
- [22] ALANSARI M, HAY O A, JAVED S, et al. GhostFaceNets: lightweight face recognition model from cheap operations [J]. IEEE Access, 2023, 11: 35429–35446.

(责任编辑:黄健)