

DOI:10.19651/j.cnki.emt.2518612

基于双重注意力校准机制的口罩遮挡人脸识别

傅鹏飞¹ 徐巍² 刘怀广³ 刘源洞³ 刘劲威¹

(1. 武汉科技大学机械工程学院 武汉 430081; 2. 武汉科技大学冶金装备及其控制教育部重点实验室 武汉 430081;
3. 机械传动与制造工程湖北省重点实验室 武汉 430081)

摘要: 针对口罩遮挡下注意力机制存在的多维度动态协同能力不足与细粒度抑制欠缺等问题,本文提出一种基于双重注意力校准的鲁棒识别方法,在通道和空间两个维度进行遮挡区域的动态校准。其中通道维度基于全局统计量抑制污染通道的异常响应,空间维度则定位遮挡区域并削弱其梯度传播,实现了从粗粒度筛选到细粒度增强的动态校准。在此基础上,通过加权交叉熵损失和三元组损失进一步引导模型聚焦了局部未遮挡区域的特征表达,从而扩大类间特征距离间隔。实验结果表明,本文提出的双重注意力校准机制经过通道维度特征筛选与空间维度区域增强的协同作用,本文算法在 LFW 与 AgeDB-30 的掩膜场景下,相比于 ArcFace 算法,分别提高了 6% 和 7.2% 的准确率,在真实遮挡数据集 MAFA 数据集上则提高了 7.3%,验证了其在复杂遮挡场景下的识别鲁棒性。

关键词: 口罩遮挡人脸识别;注意力机制;双重注意力校准;加权交叉熵损失;三元组损失

中图分类号: TN432 **文献标识码:** A **国家标准学科分类代码:** 610416

Mask occlusion facial recognition based on dual attention calibration mechanism

Fu Pengfei¹ Xu Wei² Liu Huaiguang³ Liu Yuanjiong³ Liu Jinwei¹

(1. School of Mechanical Automation, Wuhan University of Science and Technology, Wuhan 430081, China;

2. Key Laboratory of Metallurgical Equipment and its Control, Ministry of Education,
Wuhan University of Science and Technology, Wuhan 430081, China;

3. Hubei Provincial Key Laboratory of Mechanical Transmission and Manufacturing Engineering, Wuhan 430081, China)

Abstract: This paper proposes a robust recognition method based on dual attention calibration to address the issues of insufficient multi-dimensional dynamic collaboration and fine-grained suppression in the attention mechanism under mask occlusion. The method dynamically calibrates the occlusion area in both channel and spatial dimensions. The channel dimension is based on global statistics to suppress abnormal responses of polluted channels, while the spatial dimension locates occluded areas and weakens their gradient propagation, achieving dynamic calibration from coarse-grained screening to fine-grained enhancement. On this basis, the weighted cross entropy loss and triplet loss are used to further guide the model to focus on the feature expression of locally unobstructed areas, thereby expanding the inter class feature distance interval. The experimental results show that the dual attention calibration mechanism proposed in this paper, through the synergistic effect of channel dimension feature screening and spatial dimension region enhancement, has improved accuracy by 6% and 7.2% respectively compared to the ArcFace algorithm in mask scenes of LFW and AgeDB-30, and by 7.3% on the real occlusion dataset MAFA dataset, verifying its recognition robustness in complex occlusion scenes.

Keywords: mask covering facial recognition; attention mechanism; dual attention calibration; weighted cross entropy loss; triplet loss

0 引言

在安防闸机、金融支付等高安全性应用场景中,口罩遮

挡引发的面部识别准确率下降已成为生物特征识别领域的关键挑战。该问题的本质在于口罩导致口罩区域生物特征缺失,致使下半脸关键特征提取困难。尽管基于注意力机

制的方法凭借高计算效率、强可解释性及深度学习框架兼容性等优势成为主流方案,但现有方法多局限于通道注意力与空间注意力的孤立使用或简单叠加。由于存在多维度动态协同不足、参数自适应适配能力缺失、遮挡抑制粒度粗糙等缺陷,此类方法在复杂遮挡场景中难以满足鲁棒性需求。

在多维度协同动态性中, Hu 等^[1]提出的 SENet (squeeze excitation networks) 过度依赖通道统计,弱化了空间细节感知能力,在口罩遮挡场景下,眼周轮廓等细节特征模糊,影响识别精度; Wang 等^[2]的方法虽能抑制异常通道,但却忽视了空间定位,在遮挡区域的识别中,误判率有所上升。张庆辉等^[3]指出,现有方法缺少级联协同架构,导致从粗到细的动态校准过程失效,难以应对遮挡尺度的变化。Liu 等^[4]的并行双注意力机制因分支独立,割裂了跨维关联,在复杂场景下特征一致性降低;张铮等^[5]的固定阈值掩膜难以适配动态遮挡的多样性,如随机物体形状差异,使得协同机制僵化。在参数适配策略方面,其静态性限制了动态优化。Woo 等^[6]的 CBAM (convolutional block attention module) 采用无参数加权的串行融合策略,限制了跨维度注意力交互的动态性; Qiu 等^[7]的模型以动态掩膜除遮挡特征,增强类间可分性,但遮挡差异致局部特征响应不足。王丹淋等^[8]的强化学习方法因奖励函数忽略梯度传播路径,未有效抑制遮挡区域的反向传播,导致自适应抑制失效;刘祥丰等^[9]验证了静态掩膜无法根据遮挡比例调整权重,遮挡干扰持续侵蚀分类置信度。在抑制粒度的粗放中, Zhang 等^[10]的局部区域注意力因缺乏动态空间掩膜调制,无法细化口罩边缘噪声的抑制粒度; Wang 等^[11]的全局注意力池化因固定阈值导致低置信度区域如口罩遮挡未被有效过滤,冗余特征干扰显著。

针对以上问题,本文提出基于双重注意力校准机制的口罩遮挡人脸识别,主要贡献如下:

1) 提出双重注意力校准机制(dual attention calibration mechanism, DACM),采用通道空间双维度级联架构;通道维度通过全局统计抑制污染通道异常响应,空间维度定位遮挡区域并削弱其梯度传播,引导网络聚焦未遮挡关键区域。这种双维度协同方式实现了从粗粒度筛选到细粒度增强的动态校准,显著提升遮挡敏感区域的鲁棒性。

2) 采用双重损失联合优化:其中加权交叉熵损失(weighted cross entropy loss, WCE)根据输入样本的遮挡区域占比动态调整类平衡系数 λ 分类损失权重,通过差异化监督信号增强未遮挡区域权重至 1、抑制遮挡区域至 0,降低污染区域对决策边界的干扰;边界增强三元组损失引入基于三元组关系的动态边距 α 约束,迫使同类遮挡样本在特征流形上保持紧凑性,同时扩大跨类最小间隔。

1 方 法

针对口罩遮挡导致的人脸特征缺失问题,本文提出一种基于双重注意力校准机制的鲁棒识别方法。如图 1 所示,该方法以 ResNet34^[12] 作为主干网络构建多尺度特征提取框架;输入图像中的口罩掩膜经过浅层卷积层捕获局部纹理与几何结构特征后,引入双重注意力校准模块执行特征优化,该模块在通道和空间两个维度动态校准遮挡区域,并生成单通道注意力权重 w_s ,并与另外输入网络的二值掩膜计算加权交叉熵损失强化局部特征表达;进一步地,主干网络提取到的人脸特征向量 x_i 结合三元组损失优化特征嵌入的判别性,通过反向传播,两种损失在抑制遮挡区域和增强有效特征之间实现动态平衡,最终完成身份识别。

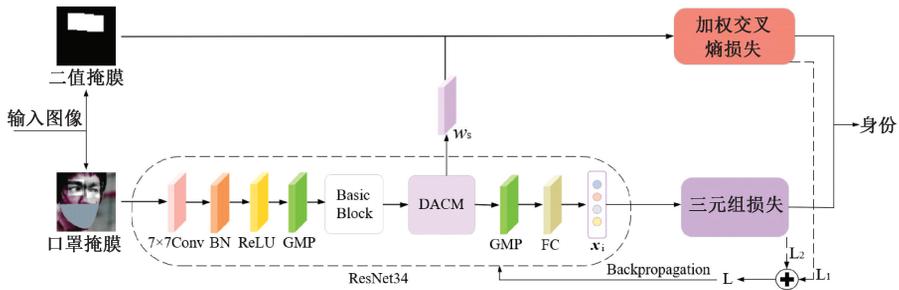


图 1 网络整体框架

Fig. 1 Overall framework of the network

1.1 残差网络特征提取

在口罩遮挡人脸识别任务中, ResNet34 作为特征提取模块的架构优势源于其残差学习机制的数学思想。如图 2 所示,该网络通过跨层跳跃连接构建残差映射;设输入特征为 x ,经两个 3×3 卷积层构成的残差函数 $H(x)$ 处理后,输出特征表述如式(1)所示。

$$y = x + H(x) \quad (1)$$

其核心价值在于通过梯度传播的数学约束解决深层网络训练难题。反向传播时,梯度分解式如式(2)所示。

$$\frac{\partial y}{\partial x} = 1 + \frac{\partial H(x)}{\partial x} \quad (2)$$

其中,恒等项 1 确保梯度在跨层传播中保持完整,有效规避了传统链式网络因梯度连乘引发的消失问题。这一特性不仅增强了网络对遮挡区域稀疏特征的代表能力,

还能通过深层语义信息弥补局部特征缺失。然而, ResNet34 对判别性特征的自主筛选机制存在局限,当面对口罩大面积遮挡时,模型易被无效区域的噪声特征误导。为此,本文提出双重注意力校准模块,通过对遮挡区域进行抑制与有效特征增强,进一步提升了模型的遮挡鲁棒性。

1.2 双重注意力校准机制

在口罩遮挡人脸识别任务中,局部特征缺失会导致模型依赖被污染的通道响应或错误的空间区域。传统单一路径的注意力机制往往仅依靠通道或空间维度关系,而本文提出的双重注意力校准机制经过通道维度特征筛选与

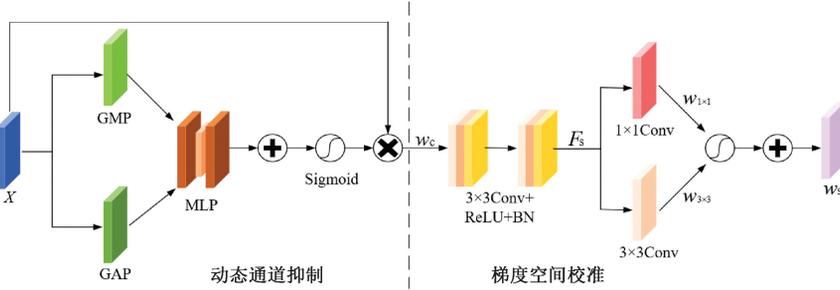


图3 DACM 结构示意图

Fig. 3 Schematic diagram of DACM structure

1) 动态通道抑制

输入特征图 $X \in R^{C \times H \times W}$ 首先通过全局平均池化(global average pooling, GAP)与全局最大池化(global max pooling, GMP)提取特征图的通道特征统计信息,并利用多层感知机(multilayer layer perceptron, MLP)学习通道间的非线性依赖关系,生成通道注意力权重 w_c , 该计算过程如式(3)所示。

$$w_c = X \odot \sigma(\text{MLP}(\text{GAP}(X)) + \text{MLP}(\text{GMP}(X))) \quad (3)$$

其中, σ 为 Sigmoid 函数, \odot 表示逐通道相乘, MLP 由 Linear-ReLU-Linear 构成。GAP 通过聚合通道特征的均值分布,能够有效捕捉口罩完整覆盖等全局遮挡信息;而 GMP 则通过提取通道特征的极值响应,可精准检测口罩边缘渐变等局部异常区域。二者的协同可全面感知遮挡特征,该设计受 CBAM 启发,但区别于其串行结构,本方法将双池化嵌入同一通道注意力分支以实现特征融合。

为量化通道级污染程度,定义隐式通道污染指数 C_k , 其通过统计各通道的 GAP 与 GMP 分布特性,结合动态阈值判定污染类型,该表达如式(4)所示。

$$C_k = \begin{cases} 1, & \mu_k \leq \alpha \text{ 或 } |\mu_k - \gamma_k| > \beta \\ 0, & \text{其他} \end{cases} \quad (4)$$

其中, μ_k 与 γ_k 分别为第 k 个通道的 GAP 与 GMP 值。 α 为低激活阈值, β 为最大允许偏差,它们由动态通道选择方法^[13]进行隐式优化,即通过特征响应统计量自适应调整阈值,降低污染通道的权重并增强有效通道贡献,克服传

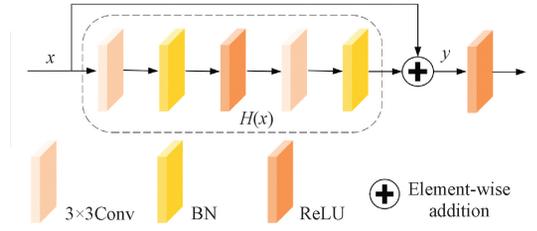


图2 残差块结构示意图

Fig. 2 Schematic diagram of residual block structure

空间维度区域增强的协同作用,显著缓解了由于遮挡带来的干扰问题,该模块的处理过程如图3所示。

统人工参数设定而导致的模型性能局限。当 $C_k = 1$ 时,判定该通道受污染: $\mu_k \leq \alpha$ 时,属于全局均匀污染情况,对应整体被口罩遮挡, $|\mu_k - \gamma_k| > \beta$ 时属于局部异常污染,对应存在局部遮挡干扰。因此,通道注意力权重 w_c 的生成本质是 MLP 对于 μ_k 以及 γ_k 的非线性组合,则 w_c 的生成可等价如式(5)所示。

$$w_c = X \odot \sigma(\text{MLP}(C_k(\mu_k, \gamma_k))) \quad (5)$$

式中: MLP 通过参数学习适应 C_k 中的阈值逻辑,将污染程度转化为权重抑制信号, Sigmoid 函数确保权重归一化以削弱污染通道的影响。

2) 梯度空间校准

通道注意力虽然可以抑制污染通道的全局响应,但对于同通道内遮挡区域的空间差异感知有限。为此,级联空间注意力模块进行局部细化:首先通过两层 3×3 卷积以及 BN 与 ReLU 操作来提取局部特征如眼部轮廓,强化遮挡过渡区域的特征表达,该过程如式(6)所示。

$$F_s = \text{ReLU}(\text{BN}(\text{Conv}_2(\text{ReLU}(\text{BN}(\text{Conv}_1(w_c)))) \quad (6)$$

随后, 1×1 卷积将通过式(7)将特征映射至空间域,其中 $W_{1 \times 1}^{(c)}$ 为卷积核参数,当通道 c 被判定为受污染时,通过梯度反传来对 $W_{1 \times 1}^{(c)}$ 进行负向优化,实现对污染通道的全局特征响应抑制。针对通道内局部污染区域,则采用 3×3 卷积构建局部感受野,如式(8)所示, $W_{3 \times 3}^{(m, n, c)}$ 为卷积核参数,通过邻域偏移($m, n \in \{-1, 0, 1\}$)学习空间模式,对口罩边缘等遮挡过渡区域赋予负权重以抑制响应,而对眼部周边等未遮挡区域则增强特征表达。

$$w_{1 \times 1} = \sum_{c=1}^C W_{1 \times 1}^{(c)} \cdot F_s^{(i,j,c)} \quad (7)$$

$$w_{3 \times 3} = \sum_{m=-1}^1 \sum_{n=-1}^1 \sum_{c=1}^C W_{3 \times 3}^{(m,n,c)} \cdot F_s^{(i+m,j+n,c)} \quad (8)$$

最终,双卷积输出经线性叠加与截断约束生成单通道注意力权重 w_s , 该过程如式(9)所示。

$$w_s = \text{clip}(\sigma(w_{1 \times 1}) + \sigma(w_{3 \times 3}), 0, 1) \quad (9)$$

其中,使用 clip() 函数旨在对双卷积融合后的叠加权重进行最终截断,与单一路径中通过 Sigmoid 实现的范围约束协同作用,共同确保权值严格位于(0,1)范围内。

在反向传播的过程中,权重 w_s 通过梯度衰减机制约束网络对遮挡区域的学习:若位置 (i, j) 被判定为遮挡区域即 $w_s^{(i,j)} \approx 0$, 则特征图 X 的梯度 $\frac{\partial L}{\partial X^{(i,j)}}$ 更新如式(10)所示。

$$\frac{\partial L}{\partial X^{(i,j)}} = \frac{\partial L}{\partial X_c^{(i,j)}} \cdot w_s^{(i,j)} \quad (10)$$

其中, L 为损失函数表达式, $X_c^{(i,j)}$ 为生成通道注意力 w_c 校准后的特征图。由于遮挡区域权重 $w_s^{(i,j)} \approx 0$, 其梯度更新量趋近于 0, 从而阻止网络依赖无效特征更新参数。这种从通道和空间两个维度的校准机制可以显著降低模型对于遮挡区域的敏感性,同时增强对于未遮挡区域的特征聚焦。

1.3 损失函数

1) 加权交叉熵损失

特征图经过双重注意力校准模块后,生成单通道空间注意力权重 $w_s \in [0, 1]^{H \times W}$, 其每个位置 (j, k) 的值 $W_s^{(j,k)}$ 表示该区域属于面部关键特征的概率。同时,输入的二值掩膜 $Y \in [0, 1]^{H \times W}$ 直接作为监督信号,其中 $Y^{(j,k)} = 1$ 对应白色区域位置即代表未遮挡的关键区域如眼部周边区域,而 $Y^{(j,k)} = 0$ 对应黑色区域即对应遮挡或背景区域。通过对比权重 w_s 与二值掩膜 Y , 采用加权二元交叉熵损失函数量化预测与真值的差异,表达式如式(11)所示。

$$L_1 = - \sum_{j,k} [\lambda Y^{(j,k)} \log W_s^{(j,k)} + (1 - Y^{(j,k)}) \log(1 - W_s^{(j,k)})] \quad (11)$$

其中, λ 为类平衡系数,用于缓解遮挡区域占比较大时所造成的训练偏差,该损失函数通过梯度回传来进行不断优化,即对于 $Y^{(j,k)} = 1$ 的位置,会引导 $W_s^{(j,k)}$ 趋近于 1, 保留并强化输入特征图特征 $X^{(i,j)}$, 从而达到增强关键区域的目的;对于 $Y^{(j,k)} = 0$ 的位置,会引导 $W_s^{(j,k)}$ 趋近于 0, 从而削弱噪声特征的影响。该过程如图 4 所示,随着训练迭代,注意力权重逐渐聚焦于有效特征区域如眼部周边区域,从而补偿由于口罩遮挡所导致的局部信息丢失。

2) 三元组损失

在加权交叉熵损失优化注意力权重的基础上,为进一步增强特征区分度,模型联合优化三元组损失。具体的,输入人脸图像在经过残差网络与双重注意力校准模块处

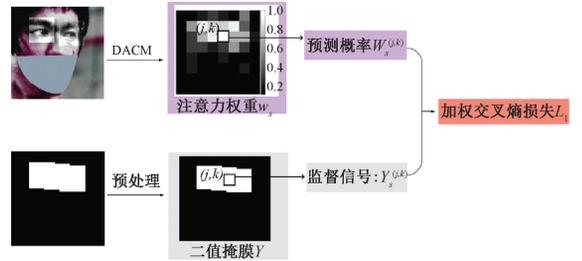


图 4 加权交叉熵损失计算流程

Fig. 4 Weighted cross entropy loss calculation process

理后,会提取生成深层特征向量 x_i , 三元组损失通过约束同类样本与异类样本在特征空间中的相对距离,引导模型学习更具判别性的嵌入表示。给定锚点样本 A(Anchor)、正样本 P(同一身份)与负样本 N(不同身份),该损失函数定义如式(12)所示。

$$L_2 = \frac{1}{M_{(A,P,N) \in M}} \max(0, D(A,P) - D(A,N) + \alpha) \quad (12)$$

其中, $D()$ 表示特征向量间的欧氏距离^[14], $\alpha > 0$ 为预设边距, M 为所有三元组集合,该优化处理条件如式(13)所示,当满足该条件时,损失函数产生非零梯度,通过反向传播不断拉近 A 与 P 的距离,同时加大 A 与 N 的距离,该过程如图 5 所示,模型在训练过程中不断调整参数扩大不同身份的特征差异,从而提高整体识别性能。

$$D(A,P) + \alpha \geq D(A,N) \quad (13)$$

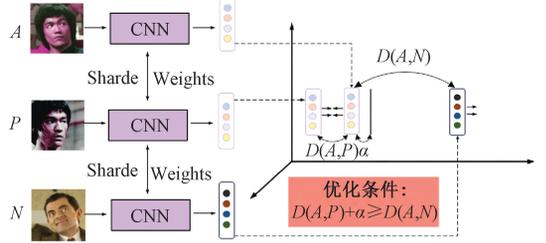


图 5 三元组损失计算流程

Fig. 5 Triple loss calculation process

3) 综合损失

结合两种损失函数,最终形成综合损失如式(14)所示,通过不断优化该损失,结合局部聚焦与全局判别的协同机制,逐步增强特征表征的捕获能力及类别间的区分度。

$$L = L_1 + L_2 \quad (14)$$

其中, L 为最终综合损失,而 L_1 和 L_2 分别代表加权交叉熵损失和三元组损失。

2 实 验

2.1 实验环境

实验在 Featurize 云服务器平台上进行,使用了 PyTorch 深度学习框架以及 Python 语言构建神经网络。

开发环境基于 Windows10 操作系统,并配置了 NVIDIA RTX 3080 GPU 以加速模型训练过程。模型超参数设置如下:每批次处理的图像数量为 40;优化器选用 adam^[15],学习率设为 0.001;在三元组损失中,困难样本的参数值设为 100 000。

2.2 实验数据集

训练集采用了数据增强处理后的 VggFace2-Mask^[16]数据集及其对应的二值掩膜,包含超过 9 100 个个体的约 3.31×10^6 张面部图像。通过合成遮挡实现了对关键遮挡参数如位置、面积、形状的精准确控,满足遮挡鲁棒性研究中单一变量分析的需求。该数据集的操作预处理如图 6 所示。

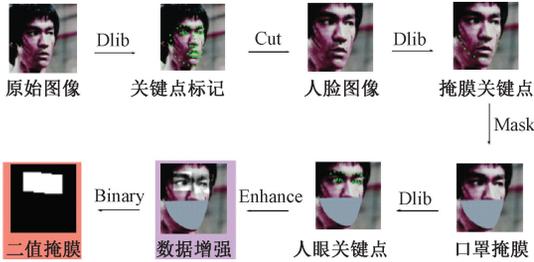


图 6 数据集处理流程

Fig. 6 Dataset processing flow

首先,基于 FERET (face recognition technology) 数据集标准^[17],采用自适应裁剪对齐人脸区域,使其占比图像面积的 $72\% \pm 3\%$,以消除背景干扰。接着,根据面部关键点构建口部遮挡掩膜,覆盖面部 $36\% \pm 4\%$ 区域,以此来模拟真实口罩覆盖,并填充中性灰度 (RGB: 160, 160, 160) 以避免颜色特征依赖。为进一步强化模型对判别性区域的关注,利用 Dlib^[18] 定位眼周关键点,构建自适应尺寸的注意力增强区域,占面部宽度 40%,并采用自适应直方图均衡化来优化局部特征表达。最终生成数据增强图像以及二值掩膜,用于联合损失函数的区域权重分配,实现与 DACM 模块的端到端协同优化。经上述处理的部分数据集如图 7 所示。

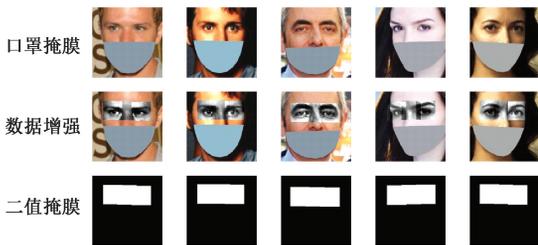


图 7 数据集处理示意

Fig. 7 Dataset processing diagram

本文选用 LFW^[19] (labeled faces in the wild)、AgeDB-30^[20] (aging database with 30 images per subject)、MAFA^[21] (masked faces) 作为测试集,该三种数据集部分图像如图 8 所示,其中 LFW 与 AgeDB-30 通过预处理合成

口罩掩膜模拟遮挡场景,而 MAFA 则使用真实佩戴口罩的原始图像,未引入任何遮挡合成操作。此设计旨在分别验证算法在合成遮挡与真实遮挡场景下的泛化能力。



图 8 三种数据集部分示意

Fig. 8 Partial schematic of three datasets

LFW 包含超过 13 000 张来自 5 749 个个体的网络面部图片,涵盖多样光照、姿态及遮挡场景,是人脸验证任务的基准数据集;AgeDB-30 包含约 160 00 张图像,覆盖 16~62 岁年龄跨度,用于评估模型对皱纹增生、面部轮廓形变等年龄特征的鲁棒性;MAFA 包含 2 500 张真实遮挡图像,涵盖医用口罩、N95 及布质口罩等多种类型,遮挡面积占比 $15\% \sim 40\%$ 。通过上述三种数据集的测试,充分验证了本文算法的可行性。

2.3 模型实验结果与对比分析

1) LFW 对比试验

为确保算法的有效性,将本文所提出的算法与 AroFace^[22]、TransFace^[23]、MinusFace^[24]、DSL-FIQA^[25]、DINO^[26]、ArcFace^[27]、ResNet50 以及添加了 CBAM 模块的 ResNet34 在有口罩掩膜的 LFW 数据集上进行了测试,结果如图 9 所示。

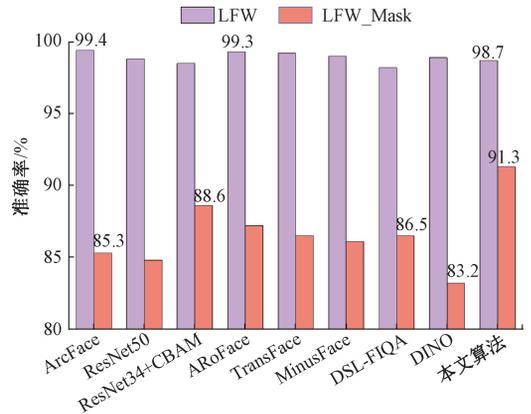


图 9 LFW 有无掩膜下不同算法对比

Fig. 9 Comparison of different algorithms with and without mask in LFW

通过分析可知,本文算法在有口罩掩膜情况下进行人脸识别的任务中,相比于其他几种算法,表现出了显著的优势。具体来说,在无掩膜情况下,本文算法的准确率达到 98.7%,略低于 ArcFace 的 99.4%,但差距在 1% 以内是因为本文模型针对遮挡场景优化,略微牺牲了无遮挡场

景的特征泛化性。在掩膜情况下,本文算法以 91.3% 的准确率显著优于其他算法,证明了本文算法缓解了由于遮挡所带来的影响,因而具有更高的识别精度。

2) AgeDB-30 对比试验

本文提出的算法在有无掩膜的 AgeDB-30 数据集上进行测试的结果如图 10 所示。

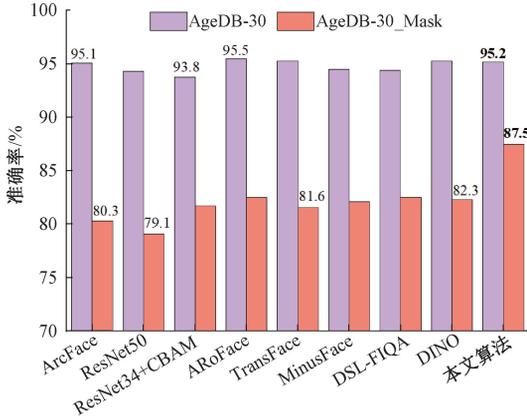


图 10 AgeDB-30 有无掩膜下不同算法对比

Fig. 10 Comparison of different algorithms under AgeDB-30 with and without mask

通过分析可得,本文算法在 AgeDB-30 数据集上的准确率略低于 LFW,但较之其他算法,仍旧得到最佳的表现,在无掩膜情况下,本文算法的准确率为 95.2%,与最低的算法 ResNet34+CBAM 相较,得到了 1.4% 的提升,说明在正常场景下本文算法具备更强的特征提取能力;在掩膜情况下,本文算法以 7.2% 的优势高于 ArcFace 算法,进一步说明了本文算法对于口罩遮挡干扰具有更强的鲁棒性。

3) 模型参数时间对比

将本文方法与其他算法的参数数量和推理效率进行对比如图 11 所示,结果表明:本文算法参数数量仅为 23.1 M,推理时间为 68 ms,显著优于 ARoFace 的 98 M、45.1 ms,证明了在处理掩膜遮挡等复杂场景时仍保持高精度与鲁棒性。与 ResNet34+CBAM 的 21.4 M 相比,其参数数量虽然略低,但识别性能显著弱于本文算法,说明了本文设计在高精度需求下实现了性能与效率的最优平衡。

4) DACM 模块实验分析

为验证 DACM 模块的最优部署策略,本研究在 ResNet34 架构中设计了多组对比实验,各方案设计及其实验结果分别如图 12 与 13 所示。

实验结果表明,在方案三当中,当 DACM 置于输入层与首层残差块之间时,虽能强化初级特征提取,但受限于网络浅层的低语义信息,导致注意力机制难以精准定位关键特征,最终识别精度提升有限;方案四中,将模块嵌入残差块内部虽能利用中层特征兼具语义与细节的优势,但其性能仍存在优化空间。而本研究提出的在残差块后嵌入 DACM 模块的方案二展现出显著优势,与方案一相比,在

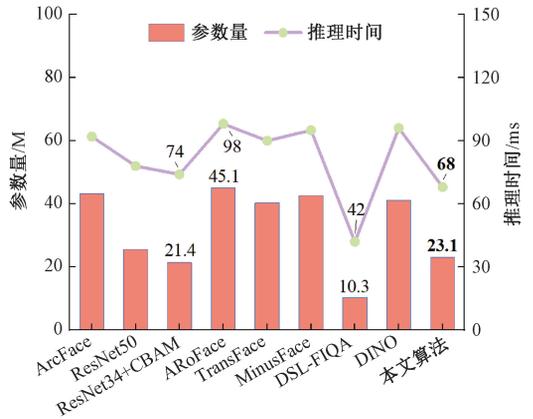


图 11 不同算法参数数量对比

Fig. 11 Comparison of parameter quantities for different algorithms

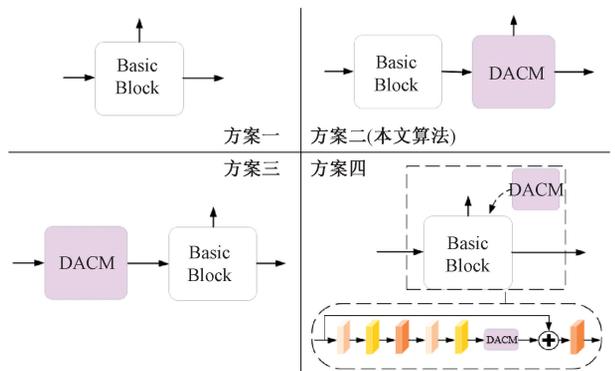


图 12 DACM 部署方案

Fig. 12 DACM deployment plan

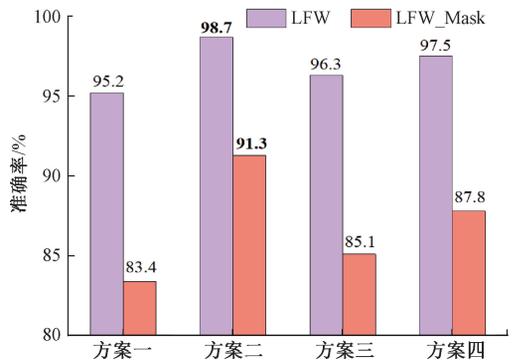


图 13 DACM 部署结果

Fig. 13 DACM deployment results

有无掩膜条件下分别实现 3.5% 和 7.9% 的精度提升。该设计通过深层特征与注意力机制的有效耦合,既避免了浅层特征的语义匮乏问题,又规避了中层特征的空间信息损失,充分验证了模块部署策略的合理性与有效性。

5) 消融实验

本文使用 ResNet34 网络作为 Baseline,选择性使用数据增强方法、DACM 模块以及加权交叉熵损失 WCE,使用 LFW 数据集分别在有无掩膜情况下进行测试,并以准确

率为代表验证各个模块所存在的有效性,实验的设计结果如表1所示。

表1 消融实验的设计与结果

Table 1 Design and results of ablation experiment

方法	准确率/%	
	LFW	LFW_Mask
Baseline	94.1	81.6
+数据增强	94.4	82.3
	(+0.3)	(+0.7)
+DACM	96.8	87.2
	(+2.7)	(+5.6)
+WCE	95.2	85.1
	(+1.1)	(+3.5)
本文算法	98.7	91.3
	(+4.6)	(+9.7)

由表1分析可得,未添加任何模块时,模型在 LFW 数据集上的识别准确率为 94.1%, LFW_Mask 上则为 81.6%,表明存在的口罩掩膜对面部识别构成挑战。使用数据增强方法之后,相较于 Baseline,在有无掩膜情况下,分别提高了 0.3% 以及 0.7%,表明口罩掩膜对于遮挡情况有所改善;在嵌入 DACM 模块时,模型在 LFW 上达到了 96.8%,增幅近 3%,证明 DACM 模块可以充分降低遮挡区域的影响。而在 LFW_Mask 上提升近 6%,表明了模型在掩膜情况下表现出的适应性显著增强;当嵌入 WCE 模块时,模型分别提高了 1.1% 以及 3.5%,说明引入加权交叉熵损失强化未遮挡部分的重要性,补偿遮挡造成的信息缺失,该方案是确实可行的,最后在本文算法中,结合所有措施,算法得到了 4.6% 以及 9.7% 的涨幅,这表明结合细致的注意力动态校准和精确的损失优化策略,可以极大提升模型在带有掩膜条件下的识别精度。

6) 真实遮挡场景验证

为验证算法在真实遮挡场景下的鲁棒性,本文在真实遮挡数据集 MAFA 上进行测试。表2展示了本文算法与对比方法的性能差异。

表2 真实遮挡场景性能对比

Table 2 Performance comparison on real-world masked faces

方法	MAFA 准确率/%
ArcFace	78.3
ResNet50	82.1
ResNet34+CBAM	83.5
本文算法	85.6

实验结果表明,本文算法在真实遮挡数据集 MAFA 上达到了 85.6% 的准确率,显著优于 ResNet34+CBAM

的 83.5% 与 ResNet50 的 82.1%。相较于其他两种合成掩膜数据,准确率有所下降,主要源于真实场景中口罩材质多样性^[28]如反光 N95、布质纹理、佩戴位置偏移与合成数据的预设差异。尽管如此,本文算法依旧显著优于 ArcFace 的 78.3% 以及其他对比方法,实验验证了本文算法可以通过动态聚焦眼部、眉部等未遮挡区域,能够有效抑制复杂遮挡带来的干扰问题。同时,凭借其无需额外调整学习率,优化器等其他参数即可实现跨域泛化的优势,证明了实际部署的价值。

7) 特征图可视化分析

为评估本文算法对特征聚焦的有效性,本研究采用梯度加权类激活映射^[29](gradient-weighted class activation mapping, Grad-CAM),对 ResNet50、ResNet34+CBAM 及本文算法的最后一层卷积层特征进行分析,生成特征响应热力图,结果如图14所示。图14(a)展示了输入网络的数据增强图像及相应的二值掩膜图。图14(b)、(c)和(d)的第一张分别呈现了在 ResNet50、ResNet34+CBAM 和本文算法3种不同算法下,输入图像各区域对神经网络输出类别的“贡献”程度,其中红色表示“贡献”较大,颜色由黄色到蓝色则表示“贡献”程度逐渐降低。从图14(b)可以看出,在 ResNet50 算法中,口罩区域仍占据较大比重。图14(c)中,尽管在 ResNet34+CBAM 算法下人眼周边区域的“贡献”较为显著,但口罩区域的权重分布仍高于本文算法(图14(d))中的情况。进一步对比图14(b)(c)、(d)的

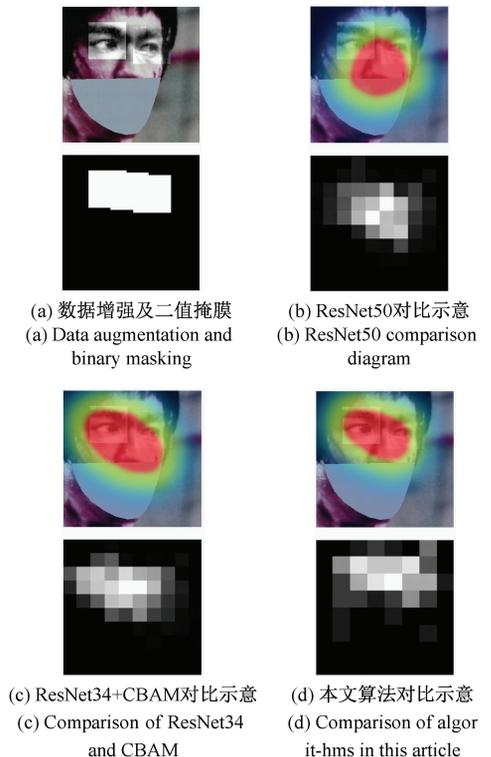
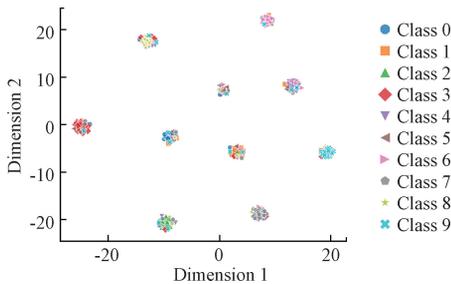


图14 数据可视化示意

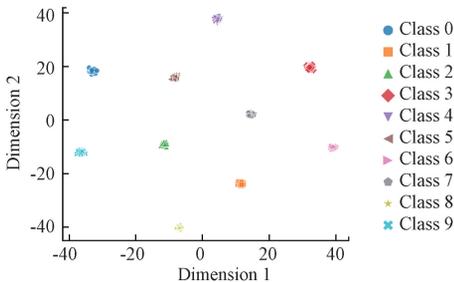
Fig. 14 Data visualization diagram

第二张的二值掩膜可以发现,本文算法(图 14(d))对人眼周边区域的权重分配更为集中,相较于 ResNet50 和 ResNet34+CBAM 两种算法,本文算法在关键特征提取方面具有显著优势,精准性和针对性更强。

为了验证本文方法在身份分类方面的性能,在 LFW 数据集中选取了 10 个不同的身份类别,每个类别均包含无遮挡及口罩掩膜的图像。对这些图像进行特征提取操作得到 512 维特征,并运用 t-SNE(t-distributed stochastic neighbor embedding)方法将其降至二维特征,并将特征进行可视化,所得结果如图 15 所示。经过 DACM 模块处理后,同类样本聚集成紧凑簇群,其簇内紧密度显著提升;同时异类样本间形成明确分界,簇间分离度同步增强。该现象表明模型能有效过滤遮挡区域的冗余信息,所提取特征具有更强的身份鉴别能力。不同身份特征在高维空间中呈现显著可分性分布,充分验证了本文方法在提升特征判别性方面的有效性。



(a) 基准模型(无DACM)
(a) Benchmark model (without DACM)



(b) 本文算法(存在DACM)
(b) Algorithm in this article (with DACM present)

图 15 使用 t-SNE 进行特征分布的可视化

Fig. 15 Visualization of feature distribution using t-SNE

3 结 论

本文提出双重注意力校准算法,通过多维协同与细粒度抑制优化,显著提升口罩遮挡下的人脸识别鲁棒性。通过通道注意力机制抑制污染通道的全局干扰,结合空间注意力模块定位局部遮挡区域并削弱其梯度传播,构建的双维度协同校准机制在掩膜 LFW 数据集上实现 91.3% 的识别准确率,较 ArcFace 和 ResNet50 模型分别提升 6.0% 和 3.8%。消融实验表明,移除双重注意力模块将导致 LFW_Mask 准确率下降 5.6%。同时,加权交

叉熵损失与三元组损失的联合优化策略贡献 3.5% 的性能增益,在真实遮挡数据集 MAFA 上达到 85.6% 的准确率,显著优于 ArcFace 算法的 78.3%。然而,现有方法在遮挡条件下仅考虑了口罩带来的影响,这可能制约实际应用场景中的识别效果。后续研究将进一步探索帽子、墨镜等复合遮挡条件下的算法适应性,致力于构建更具通用性的识别框架。

参考文献

- [1] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.
- [2] WANG F, JIANG M Q, QIAN C, et al. Residual attention network for image classification [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6450-6458.
- [3] 张庆辉, 张媛, 张梦雅. 有遮挡人脸识别进展综述 [J]. 计算机应用研究, 2023, 40(8): 2250-2273.
ZHANG Q H, ZHANG Y, ZHANG M Y. Advances in occluded face recognition: A survey [J]. Application Research of Computers, 2023, 40(8): 2250-2273.
- [4] LIU C, HIROTA K, DAI Y. Patch attention convolutional vision transformer for facial expression recognition with occlusion [C]. IEEE International Conference on Information Sciences. Piscataway: IEEE, 2023: 781-794.
- [5] 张铮, 芦天亮, 曹金璇. 基于分割和多级掩膜学习的遮挡人脸识别方法 [J]. 计算机科学与探索, 2024, 18(7): 1814-1825.
ZHANG ZH, LU T L, CAO J X. Occluded face recognition method based on segmentation and multi-level mask learning [J]. Computer Science and Exploration, 2024, 18(7): 1814-1825.
- [6] WOO S, PARK J, LEE J, et al. CBAM: Convolutional block attention module [C]. European Conference on Computer Vision, 2018: 3-19.
- [7] QIU H, GONG D, LI Z, et al. End2End occluded face recognition by masking corrupted features [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(10): 6939-6952.
- [8] 王丹淋, 张辅超, 刘昕, 等. 基于强化学习的模块化有遮挡人脸识别算法研究 [J]. 电子设计工程, 2025, 33(5): 157-161.
WANG D L, ZHANG F CH, LIU X, et al. Research on modular occluded face recognition algorithm based

- on reinforcement learning [J]. *Electronic Design Engineering*, 2025, 33(5): 157-161.
- [9] 刘祥丰. 基于注意力深度神经网络的遮挡人脸识别研究与应用[D]. 徐州:中国矿业大学, 2021.
- LIU X F. Research and application of occluded face recognition based on attention deep neural network[D]. Xuzhou: China University of Mining and Technology, 2021.
- [10] ZHANG J, ZHU W, HUANG J, et al. LAMM-ViT: AI face detection via layer-aware modulation of region-guided attention [J]. *ArXiv preprint arXiv: 2505.07734*, 2025.
- [11] WANG K, PENG X, YANG J, et al. Suppressing uncertainties for large-scale facial expression recognition[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 6897-6906.
- [12] 刘帅, 嵇小辅. 基于改进 ResNet34 卷积神经网络的硅棒缺陷检测[J]. *计算机测量与控制*, 2025, 33(2): 37-43.
- LIU SH, JI X F. Silicon rod defect detection based on improved ResNet34 convolutional neural network [J]. *Computer Measurement & Control*, 2025, 33(2): 37-43.
- [13] TIWARI A, CHATURVEDI A. A novel channel selection method for BCI classification using dynamic channel relevance [J]. *IEEE Access*. Washington: IEEE, 2021, 9: 126698-126716.
- [14] 张东, 高丙朋. 基于改进 MobileNet 网络的人脸识别方法[J]. *山西大学学报(自然科学版)*, 2023, 46(1): 147-153.
- ZHANG D, GAO B P. Face recognition method based on improved MobileNet network[J]. *Journal of Shanxi University(Natural Science Edition)*, 2023, 46(1): 147-153.
- [15] 赵志杰, 张艳艳, 毛翔宇. 基于改进 Adam 优化算法的中文短文本分类方法[J]. *电子测量技术*, 2022, 45(23): 132-138.
- ZHAO ZH J, ZHANG Y Y, MAO X Y. Chinese short text classification method based on improved Adam optimization algorithm [J]. *Electronic Measurement Technology*, 2022, 45(23): 132-138.
- [16] 王景. 基于深度学习算法的视频监控人脸识别系统研究[J]. *长江信息通信*, 2023, 36(8): 128-131.
- WANG J. Research on video surveillance face recognition system based on deep learning algorithm[J]. *Changjiang Information and Communication*, 2023, 36(8): 128-131.
- [17] ABDULWAHID A A. Classification of ethnicity using efficient CNN models on MORPH and FERET datasets based on face biometrics [J]. *Applied Sciences*, 2023, 13(12): 7288.
- [18] 廖冬杰. 基于 Dlib 和变种 Transformer 的哈欠检测方法[J]. *汽车技术*, 2023(3): 42-48.
- LIAO D J. Yawning detection method based on Dlib and variant Transformer[J]. *Automotive Technology*, 2023(3): 42-48.
- [19] 季瑞瑞, 谢宇辉, 骆丰凯, 等. 改进视觉 Transformer 的人脸识别方法[J]. *计算机工程与应用*, 2023, 59(8): 117-126.
- JI R R, XIE Y H, LUO F K, et al. Improving the face recognition method of visual Transformer [J]. *Computer Engineering and Applications*, 2023, 59(8): 117-126.
- [20] 张飞翔. 面向卷积神经网络的人脸识别技术研究[D]. 上海:上海交通大学, 2020.
- ZHANG F X. Research on facial recognition technology for convolutional neural networks [D]. Shanghai: Shanghai Jiao Tong University, 2020.
- [21] GE S, LI J, YE Q, et al. Detecting masked faces in the wild with LLE-CNNs[C]. *IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017: 2682-2690.
- [22] SAAEDABADI M S E, MALAKSHAN S R, DABOUEI A, et al. ARoFace: Alignment robustness to improve low-quality face recognition[C]. *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2024: 308-327.
- [23] DAN J, LIU Y, XIE H, et al. TransFace: Calibrating transformer training for face recognition from a data-centric perspective [C]. *IEEE/CVF International Conference on Computer Vision*. IEEE, 2023: 20642-20653.
- [24] MI Y, ZHONG Z, HUANG Y, et al. Privacy-preserving face recognition using trainable feature subtraction[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024: 297-307.
- [25] CHEN W T, KRISHNAN G, GAO Q, et al. DSL-FIQA: Assessing facial image quality via dual-set degradation learning and landmark-guided transformer[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

- IEEE, 2024: 2931-2941.
- [26] ZHAO Z A, WANG S, CHEN M X, et al. Enhancing human detection in occlusion-heavy disaster scenarios: A visibility-enhanced DINO (VE-DINO) model with reassembled occlusion dataset[J]. Smart Cities. Lausanne: EPFL Press, 2025, 8(1): 12.
- [27] DENG J, GUO J, XUE N, et al. ArcFace: Additive angular margin loss for deep face recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019: 4690-4699.
- [28] CRISAFI F, SMEDILE F, YAKIMOV M M, et al. Bacterial biofilms on medical masks disposed in the marine environment: A hotspot of biological and functional diversity [J]. Science of the Total Environment. Amsterdam: Elsevier, 2022, 837: 155731.
- [29] SANTOS R, LIU Y, WANG Q, et al. Grad-CAM: The impact of large receptive fields and other caveats [J]. Computer Vision and Image Understanding, 2025, 234: 103-115.

作者简介

傅鹏飞, 硕士研究生, 主要研究方向为人脸识别与检测。

E-mail: 1710698374@qq.com

徐巍(通信作者), 博士, 主要研究方向为机器学习和工业自动化检测。

E-mail: xuwei@wust.edu.cn

刘怀广, 博士, 副教授, 主要研究方向为机器视觉和模式识别。

刘源洞, 博士, 副教授, 主要研究方向为机器视觉和智能控制。

刘劲威, 硕士研究生, 主要研究方向为图像处理与应用。