

# 年龄算子深度稀疏融合扩展表情识别

陈 斌, 樊飞燕, 张 睿

(南京师范大学信息化建设管理处, 江苏 南京 210023)

[摘要] 为解决不同年龄段的人群在表情特征上的特质差异,提出了年龄算子深度稀疏融合扩展表情识别模型.该模型利用高层级形式的线性字典序列作为稀疏表示的输入信号,将输入对象以线性组合的方式构造训练集,并在所有求解中以稀疏级别为指标选择最优解,进一步通过卷积神经网络的卷积、池化和全连接处理,通过融合扩展策略解决数据集相似度高、数量不足及分布不均的问题.在此基础上结合了年龄算子作为特征元素与表情特征一并提取并作为分类决策依据.通过深度稀疏融合扩展,并利用年龄算子的人体测量模型、内部角度构算和皮肤皱纹检测作为表情特征提取附加因子,经过多数据集实验结果比对,证明了本算法跨数据集的有效性和稳定性,并通过代表性跨年龄表情分类算法横向比对,证明了本算法的优势.该方法对跨年龄表情识别准确度和鲁棒性有较好效果,有一定研究价值和参考意义.

[关键词] 表情识别,深度稀疏,融合扩展,年龄算子,机器学习,卷积神经网络

[中图分类号] TP39 [文献标志码] A [文章编号] 1672-1292(2023)03-0043-10

## Age Operator Deep Sparse Fusion Extension of the Expression Recognition

Chen Bin, Fan Feiyan, Zhang Rui

(Informatization Office, Nanjing Normal University, Nanjing 210023, China)

**Abstract:** In order to solve the trait differences in expression characteristics of people at different ages, the age operator deep sparse fusion extended expression recognition model is proposed. The model uses the form of linear dictionary sequence as sparse input signal, the input object with linear combination of training set, and select the optimal solution, further by convolutional neural network convolution, pooling and full connection processing, through fusion extension strategy to solve the problem of data set similarity, insufficient quantity and uneven distribution, and on the basis of combining age operator as characteristic elements and expression features extraction and as the basis of classification decision. Through deep sparse fusion extension, the age operator anthropometric model, internal angle structure and skin wrinkles detection as expression feature extraction additional factors. The multiple data set experiment results prove the effectiveness and stability of the algorithm across data sets, and the representative across age expression classification algorithm transverse comparison proves the advantages of the algorithm. This method has good effects on the accuracy and robustness of cross-age expression recognition, and has certain research value and reference significance.

**Key words:** expression recognition, deep sparse, fusion extension, age operator, machine learning, convolutional neural network

人脸面部表情作为情感认知的重要载体,对其蕴含的重要信息的把握与理解,对准确有效地沟通而言极为重要.传统学科中,借助医学及心理学理论对表情分类定义、结构组成、成因分析及内在规律等研究领域的成果已较为成熟.人工智能背景下,基于模式识别及机器学习等新兴学科中,对快速准确的表情分类分析的研究也已然成为热点.表情识别具有广泛应用场景,可应用于安全驾驶、智能教室、视频会议、虚拟现实和认知科学等<sup>[1-3]</sup>.面部表情随着年龄增长会有相应变化<sup>[4]</sup>,这与脸部皮肤肌肉及心理状态的变化有较大的关联<sup>[5]</sup>.当前智能表情识别更多聚焦于广义分类,基于年龄或性别等差异化分类因素考虑较少,相应的标准数据集和研究基础也相对薄弱<sup>[6]</sup>.另一方面,自2020年以来在全球新型冠状病毒大流行的态势下,口罩的使用成为了生活常态,面部大部分区域被遮挡,在非受控场景下以及很多受控场景下的表情采集也都转而聚焦于以眼部区域为基础的表情研究<sup>[7]</sup>,这无疑为表情识别技术的发展提高了研究门槛.同时,基于年龄的表情分

收稿日期:2023-06-25.

基金项目:江苏省现代教育技术研究智慧校园专项课题项目(2021-R-96609).

通讯作者:陈斌,博士,高级工程师,研究方向:模式识别、机器学习、大数据分析. E-mail: njnuchenbin@njnu.edu.cn

析需求也提上日程. 例如,对针对儿童学习注意力检测、儿童心理危机干预、老年人疾病异常监测以及为老年人提供创新性服务以提高其生活质量等<sup>[8]</sup>. 到目前为止,大多数智能表情识别研究使用的数据集都是聚焦于一个较小的年龄跨度上,常见典型的表情数据集的设计基本如此,例如 Cohn Kanade(CK)<sup>[9]</sup>,Extended Cohn Kanade(CK+)<sup>[10]</sup>,Japanese Female Facial Expression(Jaffe)<sup>[11]</sup>,以及 MMI<sup>[12]</sup>等. 典型人脸表情分析通常分为 3 个主要阶段:人脸表情数据采集、人脸表情特征提取及人脸表情分类. Ekman 等<sup>[13]</sup>确定了代表性表情分类的 6 个类别:愤怒、厌恶、恐惧、快乐、悲伤和惊讶. 将面部表情分类解决方案分为两大类,针对一组图像序列进行处理以及针对单帧图像进行处理. 但不管是哪一种类型的表情图像处理,在影响表情分类的因素中,年龄因子是最为显著的<sup>[14-15]</sup>. 由于年龄相关的面部结构变化,老年人的表情结构解析会更加困难和复杂,年长者脸部皱纹和褶皱对正常表情解析带来干扰的同时,在皱纹和褶皱中也蕴含着情绪信息可供解析. 另外,面部肌肉组织的凹陷也将对表情识别造成干扰. Malatesta 等<sup>[16]</sup>的研究表明,老年人的表情含有更多的暗示信息,与年轻人相比在表达情绪时涉及到更少的脸部区域结构,他们在情绪表达上掺杂着更多的混合的表情. 另有研究表明,当情绪与记忆相关联时,老年人的表情变化及频率比年轻人更为突显. 针对以上问题,本文提出了一种基于年龄算子加权深度融合扩展表情识别模型,通过深度学习融合网络对表情识别的优势,结合增强网络对年龄算子的抗干扰性,以提高表情识别算法的跨年龄识别鲁棒性.

1 基本原理

1.1 表情识别原理及典型方法

人脸表情可以由单个或多个脸部区域五官变化配合完成,它是整体识别和特征识别共同作用的结果. 利用人工智能进行表情分析的原理,本质上就是对含有表情的人脸单帧或连续帧进行采集,并通过算法提取表情特征,最终将表情特征进行分类的过程. 所以表情识别过程可分为 3 个部分:人脸图像的获取与预处理、表情特征提取和表情分类<sup>[17]</sup>. 常用的特征主要有 3 种:灰度特征、运动特征及频率特征. 表情识别方法上,主要有:整体识别法和局部识别法、形变提取法和运动提取法、几何特征法和容貌特征法<sup>[18]</sup>.

1.2 典型表情数据集

常用典型表情数据集有数十种,每个数据集各有其特点. 以下对几种具有代表性的表情数据集进行简单介绍(典型表情数据集重要属性对比如表 1 所示).

1.2.1 Cohn-Kanade(CK)及 CK+数据集

CK 数据集于 2000 年被发布,曾经在相当长一段时间内作为最广泛使用的表情测试数据集<sup>[19]</sup>. 由于 CK 只包含静态图片,对后期连续帧检测分类支持有限,所以通过推出 CK+的扩展集弥补了其不足.

1.2.2 JAFFE 数据集

由日本九州大学发布的以 10 位日本青年女性每人 7 种面部表情为集合的数据集 The Japanese Female Facial Expression Database (JAFFE),共有 213 张照片<sup>[20]</sup>. 其中包含了愤怒(Angry),厌恶(Disgust),恐惧(Fear),高兴(Happy),悲伤(Sad),惊讶(Surprise),中性(Neutral). 每幅图像的表情通过 60 名参与者对其评判给出平均语义评分.

1.2.3 HUMAINE 数据集

由 Douglas-Cowie 等人创建,它由 3 个自然数据库和 6 个诱导反应数据库组成. 其中包含了 50 个表情片段,并且使用了一系列情感标签对这些片段进行描述,所有标签既可以全局使用,也可以单帧使用,随时间线变化而变化<sup>[21]</sup>. 该数据集由 125 位受试者数据组成,在不同地点独立开发完成.

1.2.4 MMI 数据集

MMI 数据集在 2002 年构建并发布,包括了从正面及侧面视角捕获的图像和视频,它包括了 75 个表现不同基本情绪的成年人及其 2 900 多个与表情反应有关图像或视频的数据<sup>[22]</sup>. 该数据集的提出解决了通

表 1 典型表情数据集重要属性对比  
Table 1 Comparison of important attributes of typical expression data sets

属性	CK+	JAFFE	HUMAINE	MMI	NVIE
受试数	123	10	8~125	79	103
样本数	593	213	50	2 900	3 600
特征类	愤怒	愤怒		愤怒	
	厌恶	厌恶		厌恶	愤怒
	恐惧	恐惧	愤怒	恐惧	厌恶
	高兴	高兴	沮丧	高兴	恐惧
	悲伤	悲伤	高兴	悲伤	高兴
	惊讶	惊讶	中性	惊讶	悲伤
静态/序列	是/是	是/否	否/是	是/是	是/是
分辨率	640×490 640×480	256×256	—	720×576	—
色阶	8 位灰度 24 位彩色	8 位灰度	24 位彩色	24 位彩色	24 位彩色

常表情数据集的一些关键遗漏,特别在于其包含了一个面部表情的全时间线变化模式的全过程,即从中性开始到一系列表情的开始、发展、顶端和偏移的各个阶段,再恢复到中性面孔的全流程。

### 1.2.5 NVIE 数据集

由中国科技大学安徽省计算与通信软件重点实验室建成并发布,是目前为止较为全面的人脸表情数据库<sup>[23]</sup>,由 863 项目支持。其中包含大约 100 名被试者 3 种光照条件下(正面、左侧、右侧光照)6 种表情的可见图像以及长波红外图像,该数据集表情又分为自发表情与人为表情,自发表情通过特定视频诱发采集,人为表情又分为戴眼镜与不戴眼镜两种情况。为进行自发表情或人为表情识别与情绪分析推理实验提供了充足的实验样本与数据。

### 1.3 跨年龄人脸数据库介绍

针对跨年龄人脸表情识别问题,近些年已有部分研究。首先在跨年龄人脸数据库方面,已经有一些代表性案例。FACES 发布于 2010 年,包含了 171 名受试者的 2 052 张照片。VGG Face2 表情库发布于 2017 年,包含了 9 131 个受试者的三百多万张图片。UMDFaces 表情库发布于 2016 年,包含了 8 277 名受试者三百六十多万静态图片及 3 107 名受试者的两万多个动态视频。FaceScrub 表情库发布于 2016 年,包含了 530 名受试者十万多张照片。以上数据集都包含了预估的年龄信息。CAFR 表情库发布于 2019 年,包含了 25 000 人的一千四百多万张图片,是目前最大的跨年龄数据库。

### 1.4 表情识别领域的共性问题

心理学家 Mehrabian 通过研究发现:情绪表达=7%的言词+38%的语音表情+55%的面部表情。面部表情是表达情绪的最主要、最自然和最直接的通道<sup>[24]</sup>。一个完整的情绪体验过程,由认知层面上的主观体验、生理层面上的生理唤醒、表达层面上的外显行为这三种层面的活动共同构成<sup>[25]</sup>。但在实践中,对表情特征的标注是最大的共性难题,因为在数据标注过程中往往发现表情与标签情绪很难完全对应。基本情绪标签的对应通常符合逻辑,但并不一定符合场景和特定情况。例如,同样是惊吓的表情,在观看恐怖片和遭遇真实惊吓场景时在基础特征标签分类上无法区分,但实际蕴含的表情含义是有区别的。另一方面,特征标注与受试者年龄和心理状况有较大关系,例如同样是高兴的表情,如果仅从眉眼口鼻等关键表情特征判断,年轻受试者较年长受试者有更高的识别率,但年长者表情内里高兴的成分可能反而高于前者。这些都意味着,情绪与表情的对应性存在一定偏差,个体的主观报告结果也未必非常准确,受试者的真实情绪并不一定能准确地被解析识别。所以在基本情绪理论基础上的自动表情识别往往会出现基于数据库的表情识别率虽然很高,但与实际场景和个体真实情况对应的识别准度却不能匹配,从而导致实际应用价值受限。

## 2 理论框架及建模

### 2.1 深度稀疏卷积神经网络

在模式识别领域,稀疏编码算法在图像去噪、聚类、去模糊和分类等诸多方面的突出表现已经引起很多研究者的强烈兴趣<sup>[26]</sup>,本课题组以及有关研究团队在以稀疏编码作为分类方式基础上在表情识别方面取得了显著的成绩<sup>[27]</sup>。稀疏表示使用高层级形式的线性字典序列作为输入信号,其字典学习能力的设计用于根据输入信息推断数据内部本质含义。稀疏表示本质上是基于压缩感知理论,稀疏表示分类的原理是将输入对象以线性组合的方式构造训练集,在所有解中以稀疏级别为指标选择最优解,每一个训练样本都被向量化并以列矩阵排列形成字典,属于同一个类的训练样本以近邻方式相互组合,而字典在逻辑上与类本身相分离,在基础的稀疏表示分类算法中,分类求解的表达式为

$$\text{ARGMIN}_C \|S - \text{DIC}_C\| + \beta \|C\|. \quad (1)$$

式中,ARGMIN 表示使目标函数取最小值时的变量值。其中  $S \in M^P$  表示矢量测试样本集, $\text{DIC} \in M^{P \times Q}$  表示字典序列, $C \in M^Q$  表示稀疏向量系数, $\beta$  表示正则化参数, $M$  表示输入对象的训练集矩阵, $P$  表示训练集矩阵的行向量, $Q$  表示训练集矩阵的列向量,稀疏表示  $S$  的计算使用的是最小化  $L_1$  范式。由于使用了稀疏向量系数  $C$ ,稀疏表示分类算法的残差量可以使用以下残差守恒方程计算,RESI 表示残差量,CLS 表示字典表分类, $C_k$  表示  $\text{CLS}_k$  类别的稀疏向量系数限制, $k$  表示向量的秩,这里除了字典表分类系数外其他项都将是零值,由此得到的就是最优稀疏非零解,进而稀疏向量系数  $C$  将从稀疏矩阵中选择相应非零解的列,并对类内残差进行容错,最终测试样本将会对残差最小的类别进行分配,



$$RESI_{CLS_k} \parallel S-DIC_{C_k} \parallel. \quad (2)$$

卷积神经网络(convolutional neural networks, CNN)是一种深度学习模型,它的优势在于处理带有网格性质的数据,它的原理类似图像受到动物视觉器官和神经的反应,自动适应性地由低层级到高层级的特征空间层次学习. CNN 作为一种数学建模,通常由卷积层、池化层和全连接层 3 种典型层级组成. 本课题组已在卷积神经网络伴残差堆叠方向上在表情识别领域有了一定的研究基础<sup>[28]</sup>. 深度网络的性能随着网络的深度和宽度的增加而提高,但其缺点是需要适配大量的参数,然而随着参数数量的增加,过拟合情况也会随之发生,并且计算量也会成几何倍数的增加. 有研究表明,通过将连接以及卷积转换为稀疏连接集可以有效解决这个问题<sup>[29]</sup>. 对于非均匀的稀疏矩阵数据,通常的算法都存在着低效的问题,所以如何能既保持网络的稀疏性质,又能保证密集矩阵的计算性能,是一个很重要的问题. 实践证明,将稀疏矩阵以聚类的方式转换为密集矩阵可以有助于其性能的提升,本文的框架模型便是在此思路基础上进行设计的.

## 2.2 融合扩展策略

虽然目前典型的人脸表情数据集有很多,但很多样本存在着相似度高的问题,并且尚无统一标准,这导致表情的分类标签类型基本是自定义的,且分类的分布很难均衡,由此导致的测试集学习效果不理想,各种分类学习效果不一致问题非常突出. 特别在特殊属性下的表情识别,例如本文重点关注的跨年龄人脸表情识别领域,本身代表性数据集就非常缺乏,再加上严重的分布差异,导致算法的学习结果难以达到预期,分类效果不能满足实际需要.

目前较有效解决该问题的方式是通过融合扩展策略对数据集进行有效的补充,既达到扩充数据集全局数量的目的,也达到补齐数据集的各个分类分布不均衡短板的目标. 通常在表情识别领域使用的图像生成方式为生成对抗网络方法,生成对抗网络模型分别由生成模型和判别模型构成,前者用于对训练集数据的分布进行探测,后者用于分辨待测数据的原生性及衍生性. 在此基础上添加条件约束作为生成对抗网络的条件扩展,以此形成带循环约束的生成对抗网络. 利用该方法的无监督概率学习手段,对待用测试数据集的样本分布进行学习,并自发产生与其相关联的扩展数据集,新产生的数据集对原数据集是全兼容的,但在全局数据量和类别分布上对原数据集进行了有益的补充. 关于该方法的实践本课题组在前期研究中已经有了多次的尝试<sup>[30]</sup>,在基于残差整流增强卷积神经网络的表情识别实验中以及双流增强融合网络微表情识别实验中均收获了较好的融合扩展结果,对最终训练、特征提取及分类效果都有明显的提升.

通过带循环约束的生成对抗网络的生成模型及判别模型对源数据集及目标数据集进行一致性比对和条件关联. 虽然在生成过程中会存在对抗损失,并且在循环条件判别过程中可能存在循环一致性损失,但这两种损失造成的目标数据集与源数据集的分布偏差均非常有限,不足以影响对最终训练集使用效果的影响,并且目标数据集的生成补充结果对该偏差进行了弥补,所以总体上是处于正偏离状态.

## 2.3 年龄算子深度稀疏融合扩展表情识别

深度稀疏卷积神经网络在表情图像分类方面具有优势,融合扩展方式弥补了其训练集不充分问题,提高了其分类算法的训练精度,使得即便是跨年龄段数据集也得以平均分布,但在其特征提取的卷积过程中,并未将年龄因子纳入考量并融合分析,这导致年龄因素中存在的表情信息没有被充分利用. 所以本课题组将年龄算子作为特征元素与表情特征一并提取并作为分类决策的依据.

对年龄算子的提取主要包括了人体测量模型、内部角度构算和皮肤皱纹检测. 人体测量模型通过尺寸大小来衡量人脸面部特征. 面部的标记点按照区域部位及角度有明确的定义,例如唇角为口部区域左右边缘,眉毛的内角起点在鼻梁根部. 通过这种定量检测和计量模式,采用欧式距离对特征点之间点位按照以下方程式进行计算. 其中  $m_1, m_2, n_1, n_2$  分别为沿着  $X$  坐标和  $Y$  坐标各自的像素位置,

$$Euclidean_{Metric} = \sqrt{(m_1 - n_1)^2 - (m_2 - n_2)^2}. \quad (3)$$

对于内部角度构算通常采用面部投影方式,一个人从幼儿到成年其面部投影中的特征标记点位置和距离是变化的,可以通过以下余弦公式对内部角度进行计算.

$$\alpha_1 = \cos^{-1} \left( \frac{K^2 + T^2 - L^2}{2KT} \right), \quad \alpha_2 = \cos^{-1} \left( \frac{K^2 + L^2 - T^2}{2KL} \right), \quad \alpha_3 = \cos^{-1} \left( \frac{T^2 + L^2 - K^2}{2TL} \right). \quad (4)$$

式中,  $K, T$  和  $L$  是由面部投影中特征标记点形成的三角形的边,  $\alpha_1, \alpha_2$  和  $\alpha_3$  为三角形对应的三个内部角.

随着年龄增长,人的皮肤纹理也会随着时间的推移产生日积月累的变化. 在这些变化中,最为明显的

会出现皱纹、眼袋或松弛等情况,通过边缘检测法可以对皮肤的褶皱进行探测,对探测到的皱纹总量进行统计,从而对年龄进行推算. 可以通过下列算式对皱纹总量进行计算,其中  $P$  表示特征区域总像素数,  $WP$  表示该特征区域白噪声像素数(即皱纹像素数),  $i$  自 1~5 分别表示脸部的前额区域、左眼睑区域、右眼睑区域、眼下脸颊区域和嘴唇周边区域. 年龄算子深度稀疏融合扩展表情识别框架如图 1 所示.

$$SUM_{wrinkles} = \sum_{i=1}^5 \frac{WP_i}{P_i} \quad (5)$$

式中,  $SUM$  表示对目标函数求和.

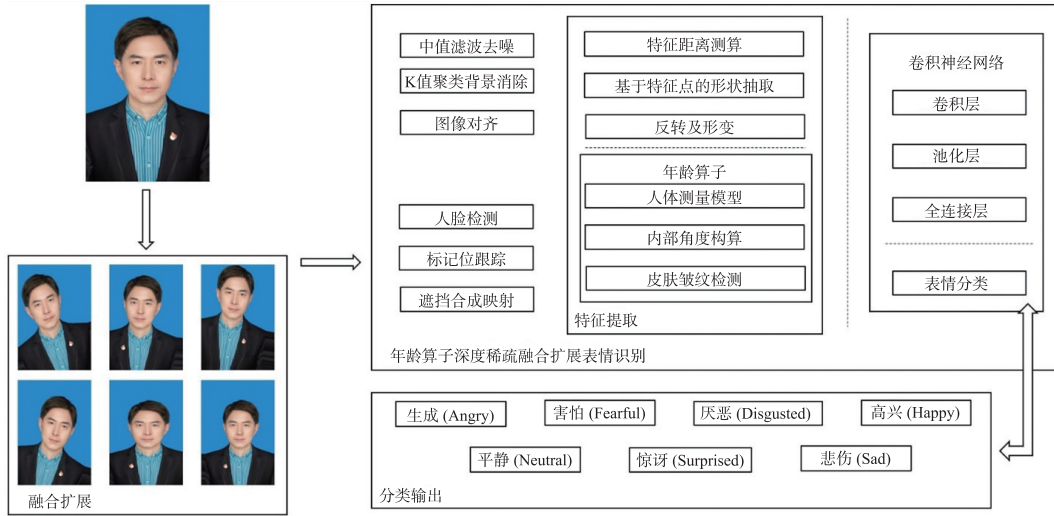


图 1 年龄算子深度稀疏融合扩展表情识别框架

Fig. 1 Age factor deep sparse fusion extension of the expression recognition framework

### 3 实验及结果

#### 3.1 实验环境及部署配置

本实验采用 64 位 Windows Server2019 操作系统,运行的硬件环境为浪潮 NF5280M6 机架式服务器, 2 \* Intel Xeon Gold 5318Y CPU, 128G DDR4 SDRAM 内存, 2 \* NVIDIA GeForce A10 GPU 加速卡. 数据运行平台采用 Anaconda3 2022.10(64-bit)集成包,是一个用于科学计算的 Python 发行版本,支持 Linux、Mac 和 Windows,包含了众多流行的科学计算、数据分析的 Python 包. 代码执行平台采用了由 JetBrains 发布的 PyCharm2022.3,是一种 Python 集成开发环境,带有一整套高效的 Python 语言开发工具. 机器学习算法运行平台采用的是 TensorFlow2.10. 由于人脸采集过程受光线的影响较敏感,故实验在自然光线充足的情况下,又增加了聚光灯补光,采集者统一在白色背景下进行受试以排除背景不一致带来的干扰. 图 2 为运行软硬件环境及现场测试环境展示.

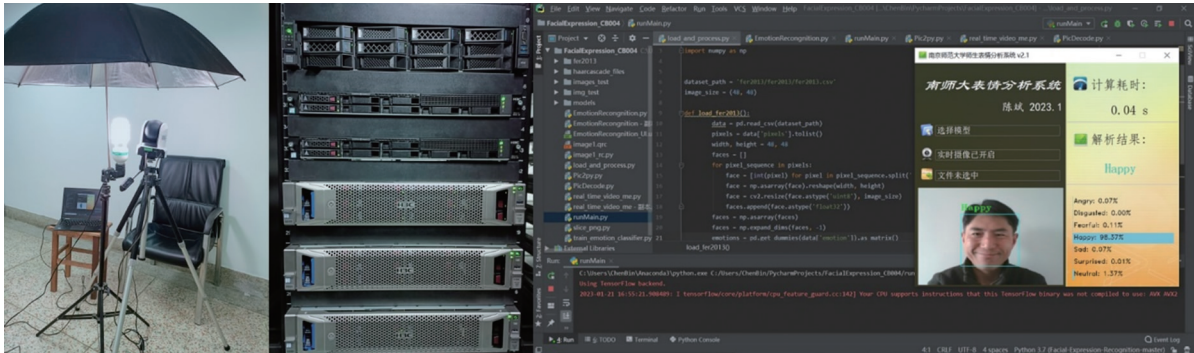


图 2 实验运行软硬件环境及现场测试环境展示

Fig. 2 Display of the experimental operation hardware and software environment and the testing environment



### 3.2 实验训练集融合扩展

因为需要对本文所提的年龄算子深度稀疏融合表情识别算法进行有效性验证,使用了跨年龄表情数据集中具有标志性的 FACES 数据集,并对该数据集进行融合扩展. FACES 数据集在 2010 年由 Ebner 团队提出,该团队推出此数据集的目的是解决表情数据集跨年龄方面研究不足的问题. 该数据集由 171 位受试者信息组成,分为 19~31 岁,39~55 岁及 69~80 岁 3 个区间年龄组,每个组分别由 58 名、56 名及 57 名受试者组成,在受试者组成数量上满足了平均分布的条件. 数据集涵盖了愤怒(Angry)、恐惧(Fearful)、厌恶(Disgust)、快乐(Happy)、中性(Neutral)、悲伤(Sad)6 种表情分类数据. 每一位受试者在这 6 种表情分类中均有两张代表性照片,这两张照片表情的采集受到专业指导后诱发拍摄. 采集过程中要求受试者素颜出境,正面对镜头,不佩戴任何饰物或眼镜,以将遮挡或干扰尽可能排除. 在对每位受试者采集若干张照片之后,由多名专业人员分别对每一张的情绪值和强度值进行综合评分,最终评分最高的两张挑选为该受试者在该表情的标签数据. 该数据集总量为:171×6×2=2 052 张. FACES 数据集表情图像示例如图 3 所示. 由于典型表情分类均为 7 类表情,所以在实验中对数据集缺失的惊讶(Surprised)分类进行了相似性添补. 经过对每一张照片进行 6 倍的融合扩展,包含相似性添补及融合扩展后的数据,最终本实验的训练数据集总量达到了 171×7×2×6=14 364 张. 对数据集进行有效的融合扩展对实验的训练效果及最终的



图 3 FACES 数据集表情图像示例

Fig. 3 Example of expression images for the FACES dataset

分类能力有明显的帮助,FACES 数据集融合扩展效果示例如图 4 所示.



图 4 FACES 数据集融合扩展效果示例

Fig. 4 Example of the fusion extension effect of the FACES dataset

### 3.3 实验结果分析

本实验是在近距受控补光采集条件下进行的,排除了遮挡和采光不足因素. 通过固定采集机位和限定受试者位置及面部角度,排除了不同形变造成的分类差异. 通过高性能算法和专业级 GPU 加速,排除了低效延迟带来的分析结果抖动. 实验按照年龄和性别不同进行分批次采集,每批次实验的训练集分为仅深度稀疏情况及深度稀疏融合增强相结合情况分别进行,4 名实验受试者(其中两名为本文作者)分别进行了每种情况下各十轮的实验,在每组实验结果中选择了总体识别率最高的为该组代表性实验结果数据,用于进行实验数据横向比对. 在分情况下的多轮实验中,从各组结果数据来看,深度稀疏与融合增强相结合的情况比仅深度稀疏后的训练集获得的效果更为理想. 分组分情况实验结果在表 2 中列出,表中情况 1 为仅深度稀疏情况,情况 2 为深度稀疏融合增强相结合情况.

表 2 分组分情况实验结果

Table 2 The experimental results of the different groups and case

表情类别	老年男性		老年女性		中青年男性		中青年女性	
	情况 1	情况 2	情况 1	情况 2	情况 1	情况 2	情况 1	情况 2
愤怒( Angry)	49.65%	59.34%	35.14%	41.57%	44.08%	51.04%	50.02%	62.20%
恐惧( Fearful)	31.04%	33.75%	40.12%	49.26%	39.69%	51.22%	54.61%	60.94%
厌恶( Disgust)	22.23%	30.66%	33.92%	37.32%	50.10%	57.50%	20.07%	27.25%
快乐( Happy)	89.67%	99.36%	87.47%	88.01%	91.33%	98.37%	77.29%	83.66%
中性( Neutral)	28.03%	34.34%	27.46%	38.86%	58.21%	72.63%	61.84%	77.72%
惊讶( Surprised)	31.38%	36.55%	25.82%	35.17%	53.80%	62.61%	26.57%	35.50%
悲伤( Sad)	45.96%	62.90%	26.61%	39.45%	47.17%	53.95%	44.03%	48.58%

注:情况 1 为仅深度稀疏情况,情况 2 为深度稀疏融合增强相结合情况.

从不同组的实验数据来看,快乐表情的分类准确率最高,最高值出现在男性年长者组别,高峰值为 99.36%. 厌恶表情的分类准确率最低,最低值出现在女性年轻者组别,低峰值为 27.25%. 造成该现象的主要原因有可能是老年男性的表情更趋近于内心真实状态的表达,而快乐的表情更加容易流露,另外快乐表情的五官特征更为明显,年龄因子(皮肤皱纹及内部角度等)更加清晰,而年轻女性的表情更趋向于内敛,特征信息相对不明显,且年龄因子较为弱化. 针对不同年龄组总体对比分析,总体而言年轻受试者组别较年老受试者组别有着更高的表情识别率,而相似年龄组中男性较女性有着更高的表情识别率. 图 5 为基于本文算法模型的实验系统例图截取,例图选择每组平均识别率最高的为代表图例.





图 5 基于本文算法模型的实验系统例图截取

Fig. 5 Example screenshot of the experimental system based on the algorithm model presented in this paper

3.4 基于不同数据集的实验结果对比分析

因为 FACES 数据集是公认的较具代表性的跨年龄数据集,所以基础实验在该数据集上进行. 为了验证本文提出模型的跨数据集的有效性和稳定性,选择了在 UMDFaces、FaceScrub 及 CAFR 共 3 个跨年龄数据集上进行了实验验证. UMDFaces 数据集虽然数据量较大,但图片均来自网络共享采集的非专业数据,并且图像均为非受控场景下自然获取,从最终实验结果看整体识别率最低且无论增强融合与否均无明显改善. FaceScrub 数据集具有大量正面采集示例,并且每个受试者有多达 200 张例图,但是一方面其跨年龄特性不足,大量图片集中在中青年年龄段,另一方面虽然图片量较大但图片多来自网络上公众人物或影视剧照的片段截取,并且缺乏专业分类基础,所以虽然本模型在年轻组别的某些分类上有较高的识别率,但在年长组别的识别率不够理想. CAFR 数据集既有按不同年龄段的标准表情分类数据,同时每个分类都有大量的标准采集图片,所以在各个年龄段上都获得了较高的识别率. 所以可以看出数据集的分布对分类算法还是有较大的影响,而分类中数量不同的影响可以通过增强融合改善. 基于不同数据集的实验结果横向对比分析如表 3 所示. 表中基于仅深度稀疏情况 1,及深度稀疏融合增强相结合情况 2 分别取各数据集实验下的跨年龄均值作为对比分析数据.

表 3 基于不同数据集的横向实验结果对比分析

表情类别	UMDFaces		FaceScrub		CAFR		FACES	
	情况 1	情况 2	情况 1	情况 2	情况 1	情况 2	情况 1	情况 2
愤怒 (Angry)	29.01%	30.83%	41.10%	46.73%	52.18%	57.13%	44.72%	53.54%
恐惧 (Fearful)	29.62%	29.07%	32.71%	40.21%	39.38%	43.06%	41.37%	48.79%
厌恶 (Disgust)	25.64%	26.90%	30.04%	36.44%	31.27%	36.95%	31.58%	38.18%
快乐 (Happy)	79.35%	82.06%	78.55%	85.48%	87.31%	90.22%	86.44%	92.35%
中性 (Neutral)	32.79%	34.18%	37.84%	41.52%	49.75%	54.08%	43.89%	55.89%
惊讶 (Surprised)	26.20%	26.33%	36.09%	37.81%	30.90%	40.66%	34.39%	42.46%
悲伤 (Sad)	26.58%	28.11%	38.70%	43.06%	45.83%	47.56%	40.94%	51.22%

3.5 典型跨年龄表情分类研究成果的比较

本节将有关跨年龄表情分类研究成果与本文提出算法的研究成果进行横向对比,以对实验结果的整



体情况作进一步说明. 使用深度学习的自动面部表情识别和年龄估计算法<sup>[31]</sup>,通过基于目标跟踪的脸部年龄评估及年龄特征提取分类,对各年龄段表情分类准确度最高平均值为 91.4%,最低平均值为 85.5%,高低平均差值为 5.9%. 低分辨率年龄相关斑点退化放大人脸表情识别算法<sup>[32]</sup>,通过局部斑点特征放大的方式促进了表情识别的准确率提升,对各年龄段表情分类准确度最高平均值为 72.3%,最低平均值为 58.2%,高低平均差值为 14.4%. 一种基于人脸性别和年龄双属性的深度条件随机森林人脸表情识别方法<sup>[33]</sup>,根据人脸性别和年龄双属性因子采用多条件随机森林方法进行人脸表情识别,对各年龄段表情分类准确度最高平均值为 89.3%,最低平均值为 70.5%,高低平均差值为 18.7%. 基于 CNN 的深度学习人脸表情识别算法<sup>[34]</sup>,通过构造深度卷积神经网络进行特征学习、选择和分类,使得输出和真实标签之间的差异最小化,对各年龄段表情分类准确度最高平均值为 88.8%,最低平均值为 47.9%,高低平均差值为 40.9%. 本文所提出的年龄算子深度稀疏融合扩展表情识别算法经过多轮实验验证,实验数据如前节所示,对各年龄段表情分类准确度最高平均值为 92.4%,最低平均值为 88.2%,高低平均差值为 4.2%. 典型跨年龄表情分类研究成果的比较分析情况如表 4 所示.

表 4 典型跨年龄表情分类研究成果的比较分析

Table 4 Comparative analysis of the research results of typical cross-age facial expression classification

研究方法	分类准确度		
	最高平均值	最低平均值	高低平均值差
基于目标跟踪的脸部年龄评估及年龄特征提取分类	91.4%	85.5%	5.9%
低分辨率年龄相关斑点退化放大人脸表情识别算法	72.3%	58.2%	14.4%
基于人脸性别和年龄双属性的深度条件随机森林人脸表情识别方法	89.3%	70.5%	18.7%
基于 CNN 的深度学习人脸表情识别算法	88.8%	47.9%	40.9%
年龄算子深度稀疏融合扩展表情识别算法(本文)	92.4%	88.2%	4.2%

通过比较以上代表性跨年龄表情算法的平均综合算力可以看出,多数算法在分类准确度的最高平均值上都较为理想(接近或高于 90%),本文提出的算法最高平均值略高,具有一定优势和研究价值. 最高平均值与最低平均值的差距可以在一定程度上反映算法的鲁棒性及分类分布均衡性,本文提出的算法高低平均值差较低,具有一定代表性和研究价值.

## 4 结论

针对不同年龄段的人群在表情特征上的特质差异,跨年龄表情分析的研究不足,以及跨年龄段数据集的缺乏问题,本文提出了一种年龄算子深度稀疏融合扩展表情识别方法,以 FACES 数据集为基础实验训练数据集,通过融合扩展方式对数据集进行了增补,并通过结合年龄算子对算法跨年龄准确度进行了提升,同时实验还在多个数据集上进行了横向比较. 从实验结论可以得出,本文提出的模型对跨年龄段表情分析识别率较高,鲁棒性较强,分类分布均衡性较稳定,整体效果较理想.

## [参考文献](References)

- [1] ZHANG T,ZHENG W M,CUI Z,et al. A deep neural network-driven feature learning method for multi-view facial expression recognition[J]. IEEE Transactions on Multimedia,2016,18(12):2528-2536.
- [2] YUAN X H,XIE L J,ABOUELENIEN M. A regularized ensemble framework of deep learning for cancer detection from multi-class,imbalanced training data[J]. Pattern Recognition,2018,77:160-172.
- [3] FANG B F,ZHANG Q,WANG H,et al. Personality driven task allocation for emotional robot team[J]. International Journal of Machine Learning and Cybernetics,2018,9(12):1955-1962.
- [4] MALATESTA C Z,IZARD C E,CULVER C,et al. Emotion communication skills in young,middle-aged,and older women[J]. Psychol Aging,1987,2:193-203.
- [5] SULLIVAN S,RUFFMAN T. Emotion recognition deficits in the elderly[J]. International Journal of Neuroscience,2004,114,403-432.
- [6] GONCALVES A R,FERNANDES C,PASION R,et al. Effects of age on the identification of emotions in facial expressions: A meta-analysis[J]. Peer J,2018,6:e5278.

- [7] DANTCHEVA A, ELIA P, ROSS A. What else does your biometric data reveal? A survey on soft biometrics[J]. IEEE Transactions on Information Forensic Security, 2016, 11(3): 441–467.
- [8] MACEDO J, COSTA F, JEFERSSON A, et al. A benchmark methodology for child pornography detection[C]//2018 31st Conference on Graphics, Patterns and Images. Parana, Brazil, 2018.
- [9] KANADA T, COHN J F, TIAN Y L, et al. Comprehensive database for facial expression analysis[C]//Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition. Grenoble, France; IEEE, 2000: 46–53.
- [10] LUCEY P, COHN J F, KANADE T, et al. The Extended Cohn-Kanade Dataset(CK+): A complete dataset for action unit and emotion-specified expression[C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. San Francisco, CA, USA; IEEE, 2010: 94–101.
- [11] LYONS M, AKAMATSU S, KAMACHI M, et al. Coding facial expressions with Gabor wavelets[C]//Proceeding Third IEEE International Conference on Automatic Face and Gesture Recognition. Nara, Japan; IEEE, 2018: 200–205.
- [12] PANTIC M, VALSTAR M, RADEMAKER R, et al. Web-based database for facial expression analysis[C]//2015 IEEE International Conference on Multimedia and Expo. Amsterdam, Netherlands; IEEE, 2005.
- [13] EKMAN P, ROLLS E T, PERRETT D I, et al. Facial expressions of emotion: An old controversy and new findings[J]. Philosophical Transactions; Biological Sciences, 335(1273): 69.
- [14] GUO G D, GUO R, LI X. Facial expression recognition influenced by human aging[J]. IEEE Transactions on Affective Computing, 2013, 4(3): 291–298.
- [15] WANG S F, WU S, GAO Z, et al. Facial expression recognition through modeling age-related spatial patterns[J]. Multimedia Tools and Applications, 2016, 75(7): 3937–3954.
- [16] MALATESTA-MAGAI C, JONAS R, SHEPARD B, et al. Type A behavior pattern and emotion expression in younger and older adults[J]. Psychology and Aging, 1992, 7(4): 551–561.
- [17] YANG D, ZHOU Q, WEN J. Interactive two-stream graph neural network for skeleton-based action recognition[J]. Journal of Electronic Imaging, 2021, 30(3): 033025.
- [18] ZHU D H, DU B, ZHANG L P. Two-stream convolutional net-works for hyperspectral target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(8): 6907–6921.
- [19] TEJA R S P, TEJA K S, DUBEY S R, et al. Spontaneous facial micro-expression recognition using 3D spatiotemporal convolutional neural networks[C]//2019 International Joint Conference on Neural Networks. Budapest, Hungary, 2019: 1–8.
- [20] ZACH C, POCK T, BISCHOF H. A duality based approach for realtime TV-L1 optical flow[C]//Joint Pattern Recognition Symposium. Berlin Heidelberg; Springer, 2007: 214–223.
- [21] 李勇, 李小竹, 蒋梦莹. 基于跨连接 LeNet-5 网络的面部表情识别[J]. 自动化学报, 2018, 44(1): 176–182.
- [22] GOYAL S J, UPADHYAY A K, JADON R S, et al. Reallife facical expression recognition systems: A review[J]. Smart Computing and Informatics, 2018, 77(1): 311–331.
- [23] WANG S F, LIU Z L, LÜ S L, et al. A natural visible and infrared facial expression database for expression recognition and emotion inference[J]. IEEE Transactions on Multimedia, 2010, 12(7): 682–691.
- [24] KELTNER D, EKMAN P. Facial Expression of Emotion[J]. Encyclopedia of Human Behavior, 2012: 173–183.
- [25] 孟昭兰. 情绪心理学[M]. 北京: 北京大学出版社, 2005.
- [26] FU Y, LAM A, SATO I, et al. Adaptive spatial-spectral dictionary learning for hyperspectral image restoration[J]. International Journal of Computer Vision, 2017, 122(2): 228–245.
- [27] 陈斌, 东一舟, 朱晋宁. 非控场景下主成分稀疏表示与低秩分解的人脸识别[J]. 液晶与显示, 2019, 34(8): 816–824.
- [28] 陈斌, 朱晋宁, 东一舟. 基于残差整流增强卷积神经网络的表情识别[J]. 液晶与显示, 2020, 35(12): 1299–1308.
- [29] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, Massachusetts, 2015: 1–9.
- [30] 陈斌, 朱晋宁. 双流增强融合网络微表情识别[J]. 智能系统学报, 2013, 18(2): 360–371.
- [31] SYEDA A R, YAZEED Y G, AHMAD J, et al. Automated facial expression recognition and age estimation using deep learning[J]. Computers Materials & Continua, 2022, 71(3): 5235–5252.
- [32] JO L, RACHEL A R, EMILIE M F R, et al. Caricaturing can improve facial expression recognition in low-resolution images and age-related macular degeneration[J]. Journal of Vision, 2019, 19(6): 1–22.
- [33] 廖海斌, 徐斌. 基于性别和年龄因子分析的鲁棒性人脸表情识别[J]. 计算机研究与发展, 2021, 58(3): 528–538.
- [34] FERNANDO A F, KEVIN H D, SILVIA R R, et al. Facial masks and soft-biometrics: Leveraging face recognition CNNs for age and gender prediction on mobile ocular images[J]. The Institution of Engineering and Technology, 2021, 10(5): 562–580.

[责任编辑: 陈 庆]