

中图法分类号: TP391.1 文献标识码: A 文章编号: 1006-8961(2022)06-2008-28

论文引用格式: Yao H X, Deng W H, Liu H H, Hong X P, Wang S J, Yang J F and Zhao S C. 2022. An overview of research development of affective computing and understanding. Journal of Image and Graphics, 27(06): 2008-2035 (姚鸿勋, 邓伟洪, 刘洪海, 洪晓鹏, 王甦菁, 杨巨峰, 赵思成. 2022. 情感计算与理解研究发展概述. 中国图象图形学报, 27(06): 2008-2035) [DOI: 10.11834/jig.220085]

情感计算与理解研究发展概述

姚鸿勋^{1*}, 邓伟洪², 刘洪海¹, 洪晓鹏¹, 王甦菁³, 杨巨峰⁴, 赵思成⁵

1. 哈尔滨工业大学, 哈尔滨 150006; 2. 北京邮电大学, 北京 100876; 3. 中国科学院心理研究所, 北京 100083;
4. 南开大学, 天津 300071; 5. 美国哥伦比亚大学, 纽约 10032, 美国

摘要: 情感在感知、决策、逻辑推理和社交等一系列智能活动中起到核心作用, 是实现人机交互和机器智能的重要元素。近年来, 随着多媒体数据爆发式增长及人工智能的快速发展, 情感计算与理解引发了广泛关注。情感计算与理解旨在赋予计算机系统识别、理解、表达和适应人的情感的能力来建立和谐人机环境, 并使计算机具有更高、更全面的智能。根据输入信号的不同, 情感计算与理解包含不同的研究方向。本文全面回顾了多模态情感识别、孤独症情感识别、情感图像内容分析以及面部表情识别等不同情感计算与理解方向在过去几十年的研究进展并对未来的发展趋势进行展望。对于每个研究方向, 首先介绍了研究背景、问题定义和研究意义; 其次从不同角度分别介绍了国际和国内研究现状, 包括情感数据标注、特征提取、学习算法、部分代表性方法的性能比较和分析以及代表性研究团队等; 然后对国内外研究进行了系统比较, 分析了国内研究的优势和不足; 最后讨论了目前研究存在的问题及未来的发展趋势与展望, 例如考虑个体情感表达差异问题和用户隐私问题等。

关键词: 情感计算; 情感识别; 孤独症; 图像识别; 表情识别

An overview of research development of affective computing and understanding

Yao Hongxun^{1*}, Deng Weihong², Liu Honghai¹, Hong Xiaopeng¹, Wang Sujing³,
Yang Jufeng⁴, Zhao Sicheng⁵

1. Harbin Institute of Technology, Harbin 150006, China; 2. Beijing University of Posts and Telecommunications, Beijing 100876, China; 3. Institute of Psychology, Chinese Academy of Sciences, Beijing 100083, China;
4. Nankai University, Tianjin 300071, China; 5. Columbia University, New York 10032, USA

Abstract: Humans are emotional creatures. Emotion plays a key role in various intelligent actions, including perception, decision-making, logical reasoning, and social interaction. Emotion is an important and dispensable component in the realization of human-computer interaction and machine intelligence. Recently, with the explosive growth of multimedia data and the rapid development of artificial intelligence, affective computing and understanding has attracted much research attention. It aims to establish a harmonious human-computer environment by giving the computing machines the ability to recognize, understand, express, and adapt to human emotions, and to make computers have higher and more comprehensive intelligence. Based on the input signals, such as speech, text, image, action and gait, and physiological signals, affective

收稿日期: 2021-01-21; 修回日期: 2022-04-06; 预印本日期: 2022-04-13

* 通信作者: 姚鸿勋 h.yao@hit.edu.cn

基金项目: 国家自然科学基金项目(61831022)

Supported by: National Natural Science Foundation of China(61831022)

computing and understanding can be divided into multiple research topics. In this paper, we will comprehensively review the development of four important topics in affective computing and understanding, including multi-modal emotion recognition, autism emotion recognition, affective image content analysis, and facial expression recognition. For each topic, we first introduce the research background, problem definition, and research significance. Specifically, we introduce how such topics were proposed, what the corresponding tasks do, and why it is important in different applications. Second, we introduce the international and domestic research on emotion data annotation, feature extraction, learning algorithms, performance comparison and analysis of some representative methods, and famous research teams. Emotion data annotation is conducted to evaluate the performances of affective computing and understanding algorithms. We briefly summarize how categorical and dimensional emotion representation models in psychology are used to construct datasets and the comparisons between these datasets. Feature extraction aims to extract discriminative features to represent emotions. We summarize both hand-crafted features in the early years and deep features in the deep learning era. Learning algorithms aim to learn a mapping between extracted features and emotions. We also summarize and compare both traditional and deep models. For a better understanding of how existing methods work, we report the emotion recognition results of some representative and influential methods on multiple datasets and give some detailed analysis. To better track the latest research for beginners, we briefly introduce some famous research teams with their research focus and main contributions. After that, we systematically compare the international and domestic research, and analyze the advantages and disadvantages of domestic research, which would motivate and boost the future research for domestic researchers and engineers. Finally, we discuss some challenges and promising research directions in the future for each topic, such as 1) image content and context understanding, viewer contextual and prior knowledge modeling, group emotion clustering, viewer and image interaction, and efficient learning for affective image content analysis; 2) data collection and annotation, real-time facial expression analysis, hybrid expression recognition, personalized emotion expression, and user privacy. Since emotion is an abstract, subjective, and complex high-level semantic concept, there are still some limitations of existing methods, and many challenges still remain unsolved. Such promising future research directions would help to reach the emotional intelligence for a better human-computer interaction.

Key words: affective computing; emotion recognition; autism; image recognition; expression recognition

0 引言

情感(emotion)一词源于希腊文“pathos”,最早用来表达人们对悲剧的感伤之情。情感在感知、决策、逻辑推理和社交等一系列智能活动中起到核心作用,甚至有研究显示“人类交流中80%的信息都是情感性的信息”。由于情感在人类信息沟通中的重大意义,情感计算是实现人机交互过程必不可少的部分,也是让机器具有智能的重要突破口。情感计算的概念由美国麻省理工学院媒体实验室 Picard 教授提出,并于1997年正式出版书籍《Affective Computing(情感计算)》。在书中, Picard 教授指出“情感计算是与情感相关,来源于情感或能够对情感施加影响的计算”,开辟了情感计算研究的先河。情感计算与理解旨在赋予计算机系统识别、理解、表达和适应人的情感的能力来建立和谐人机环境,并使计算机具有更高、更全面的智能。根据输入信号

的不同,情感计算与理解包含不同的研究方向。

1.1 多模态情感识别

多模态情感识别通过提取图像、视频、音频、文本和生理信号等多种模态数据中的情感信号,完成情感的分类、回归、检测和检索任务。多模态情感识别发展的早期,主要通过进行引导参与者产生目标情感的实验,记录参与者的生理信号、声音和面部表情的方式收集数据,支持多模态情感识别的研究(Koelstra 等,2012;Soleymani 等,2012)。近几年,多媒体技术快速发展,越来越多的用户在社交平台上发表关于电影、餐厅和电子商品的评价内容。研究者们开始搜集这些有情感倾向的图像、视频、音频和文本评价内容来构造多模态情感研究的数据集(Zadeh 等,2018c;Yu 等,2020)。随着数据量的增长和计算能力的突破,研究者们开始使用深度神经网络来分析多模态情感。很多研究使用基于 Transformer 的网络进行识别任务(Rahman 等,2020)。与之前的深度模型相比,基于 Transformer 的网络具有

更好的并行计算效率和更好的建模远距离特征优势。多模态情感识别的研究可以进一步提升人们的生活质量。在舆情分析方面,可以分析用户对新冠疫情等事件的情感倾向;在商业智能方面,可以分析用户对于商品的满意度,设计引起用户积极情感的广告;在健康方面,可以分析驾车、上课等状态下的情感程度,针对性地给出警示,提高工作效率。

1.2 孤独症情感识别

孤独症谱系障碍 (autism spectrum disorder, ASD), 又称自闭症, 是儿童时期最常见的神经发育障碍疾病之一, 其临床表现主要为社交沟通障碍、刻板行为和兴趣狭隘 (Centers for Disease Control and Prevention, 2016)。根据美国疾病控制与预防中心 (Centers for Disease Control and Prevention, CDC) 2021 年的调查报告, 孤独症儿童患病率从 2005 年的 1/166 增长到了目前的 1/44, 且呈现逐年增加的趋势。《中国自闭症教育康复行业发展状况报告 III》指出, 中国孤独症发病率约有 1%, 目前已约有超 1 000 万孤独症谱系障碍人群, 其中 12 岁以下的儿童达 200 多万, 而且孤独症儿童的数量每年以接近 20 万的数字递增。孤独症会严重影响患儿的日常生活, 并给家庭和社会带来巨大的花销, 据美国报道, 孤独症儿童终身花销大约为 200 万美元到 240 万美元。目前, 孤独症尚无特效治疗方法, 主要以早期诊断、干预为主。然而, 目前孤独症早期行为干预大多依赖于临床医师和专业人员通过行为学观察并制定相应的干预方案, 这种诊疗模式以临床医师和临床专业人员为主导, 其耗时较长, 非常不利于该病的治疗。美国早在 1943 年就对孤独症有了第 1 次确诊, 但在中国, 直到 1982 年陶国泰教授才确诊了我国第 1 例孤独症, 经过几十年的发展, 我国孤独症临床和基础研究取得了一定的进步。但迄今为止, 我国孤独症诊疗仍面临诸多困难: 专业诊疗队伍缺乏、无基于国情行业标准和诊疗指南等。社交沟通障碍是自闭症的核心症状之一, 主要体现在社交情境中患儿情感功能失调。根据《国际功能、残疾和健康分类 (儿童和青少年版)》, 孤独症患儿的社交情感功能可以界定为两大方面: 情感识别与理解以及情感调节与表达。孤独症社交情感分析可为孤独症临床康复提供新的技术手段和科学方法, 也为揭示孤独症病理机制奠定了基础。

1.3 情感图像内容分析

Minsky (1970 年图灵奖获得者) 曾指出, “问题不在于智能机器是否会有情感, 而是没有情感的机器能否智能” (Minsky, 1986)。情感在机器和人工智能领域起到至关重要的作用, 能够影响人们的判断与决策。随着移动相机和社交平台的广泛普及, 人们习惯于用图像、视频和文字等多媒体数据在线分享和表达自己的观点 (Zhao 等, 2020b)。识别这些数据中的情感内容, 可以帮助理解用户的行为和情感。“一图胜千言”, 图像可以传递丰富的语义信息。情感图像内容分析 (affection image content analysis, AICA) 的目标是理解认知层次的语义信息, 识别图像对特定观看者或大多数人会诱发的情感 (Zhao 等, 2021)。使用 AICA 自动地推断人们的情感状态, 可以帮助检测他们的心理健康、发现情感异常, 阻止他们对自己、甚至对整个社会进行的极端行为 (Zhao 等, 2021)。

情感图像内容分析 (AICA) 在心理学和行为学的研究基础上出现, 例如基于 IAPS (International Affective Picture System) 数据集研究视觉刺激与情感之间的关联 (Lang 等, 1997; Mikels 等, 2005)。早期情感图像内容分析的方法多是基于手工设计的特征, 例如低层次的全局 Wiccest 和 Gabor 特征 (Yanulevskaya 等, 2008)、艺术元素 (Machajdik 和 Hanbury, 2010)、中层次的艺术原理 (Zhao 等, 2014) 和高层次的形容词名词对 (Borth 等, 2013)。2014 年, 研究者将大规模数据集上预训练的卷积网络参数迁移到 AICA 领域 (Xu 等, 2014)。为了解决情感感知的主观性问题, 研究者们提出了个性化情感预测 (Xu 等, 2014; Yang 等, 2013; Zhao 等, 2016; Rui 等, 2017) 和情感分布学习 (Zhao 等, 2015, 2017c, 2020a; Peng 等, 2015) 策略。近几年, 领域自适应 (Zhao 等, 2018b, 2019b; Lin 等, 2020) 和零样本学习 (Zhan 等, 2019) 也应用到 AICA 领域, 来解决情感标签缺失问题。

1.4 面部表情识别

面部表情 (facial expression) 作为人类视觉最杰出的能力之一, 是非语言交流的一种重要形式 (Ekman, 1965)。人脸表情分析 (facial expression analysis, FEA) (Tian 等, 2011) 则属于计算机识别和解释人类情感状态的多学科研究领域——即情感计算的一部分, 并建立在计算机视觉技术的基础之上, 通

过分析不同来源数据(如静态图像和视频)中的人脸表情,来直观地揭示人物情绪状态。在医疗健康、刑侦检测、广告娱乐和在线教育等系列场景中均有着广泛的应用。

早在19世纪,关于人脸表情的研究就已经展开,早期主要集中在心理学和生物学方面。1862年,神经学家杜兴通过其著名的面部电流刺激实验,论证了面部肌肉是如何产生表情的(Emery等,2015)。紧接着,1872年,达尔文从进化论的角度出发,在其著作《人类和动物的情感表达》中指出,人和动物拥有与生俱来的情绪和共同的情感生物起源(Darwin,2015)。该研究对心理学家埃克曼产生了巨大的影响。他在20世纪60年代开展的首批跨文化实地研究中支持了达尔文的假说,即表情具有普适性(Ekman和Friesen,1971)。基于该研究,艾克曼定义了6种人类普遍表达的基本离散表情,即悲伤、恐惧、愤怒、开心、惊讶和厌恶(Ekman等,1987)。此外,还有一种基于局部外观的客观描述形式,即面部动作编码系统(facial action coding system, FACS)(Ekman和Friesen,1978)。该系统根据人脸解剖学的特点,将面部肌肉划分成若干动作单元(action unit, AU),来描述人脸表情的组成和变化。目前观察到的动作单元AU组合已经达到7000余种。

1.5 面部微表情识别

与普通面部表情,即宏表情(macro-expression)不同,微表情(micro-expression)是一种转瞬即逝的自发人类脸部表情(Haggard和Isaacs,1966)。当人们在高风险高压力的环境下希望隐藏其真实情感时,往往容易在脸部区域产生微表情(Ekman和Friesen,1969)。这就意味着,一方面在受到有效刺激的时候,人类最初的反应绝大多数都是不受思维控制的,因此能够映射其心理真实的变化和状态;另一方面,这些反应会很快被控制和修正,因此持续时间非常简短(通常在1/25~1/2 s内)且动作幅度很小。心理学研究表明,没有受过专门训练的人进行微表情识别,仅仅略好于随机猜测(Ekman,2003)。由于微表情这种特殊性,研究这类自发反应就显得特别重要且难以替代。

虽然在心理学上已有长达半个多世纪的研究,但是在计算机视觉、人工智能领域却只有短短10余年的发展(Zhao和Li,2019)。2011年,芬兰Oulu大

学团队提出了一种基于帧插值和多核学习(multiple kernel learning, MKL)的微表情识别方法,并建立了首个自发微表情识别数据集(Pfister等,2011)。自此以后,越来越多的学者尝试用计算机视觉的方法研究自动微表情分析。

2 国际研究现状

2.1 多模态情感识别

在多模态情感识别中,可以根据多种模态的情感信号来分析人的情感。这些不同的情感模态可以分为显性和隐性两类,显性的模态从人的身体变化来表现情感,如人脸表情、眼球移动、声音、动作、步态和脑电图等,这些模态信号可以被观察并记录;隐性模态指从多媒体平台上获取的用户信息,如用户发表的图像、视频、音频和文本,这些模态信息在数字设备之间传输并存储。

多模态情感识别的数据集构建包括获取数据和情感标注两个步骤。获取数据的方法可以分为两类,第1类在特定的场景下记录实验信息作为多模态情感数据。例如英国伦敦大学的Sander团队构建的DEAP(database for emotion analysis using physiological signals)(Koelstra等,2012),记录参与者观看音乐剧时的人脸视频、脑电图等信号。第2类直接从多媒体平台获取用户上传的图像、视频、音频和文本模态的数据。例如美国卡内基梅隆大学的Louis-Philippe Morency教授团队从YouTube收集用户上传的独白视频,构建了具有代表性的CMU-MOSEI(Carnegie Mellon University—Multimodal Opinion Sentiment and Emotion Intensity)数据集(Zadeh等,2018c)。情感标注通常由多位参与者在情感维度打分或报告感受的情感类别。另外,一些数据集并不需要进行情感的标注。例如,EMODB数据集(Burkhardt等,2005)中,演员表演的每个句子都对应一个目标情感、Multi-ZOL的每条用户评价都包含一个用户给出的打分(Xu等,2019)。

多模态情感识别面临来自多模态融合和情感识别两方面的挑战。多模态融合方面,第1个挑战是数据缺失,在获取数据的阶段,由于传感器的故障等难以避免的情况,经常出现特定模态的数据不完整的问题;第2个挑战是跨模态不一致,在社交平台上,例如用户发表的图像与文本信息有时并无语义

上的关联;第3个挑战是跨模态信息不均衡,例如,新闻通常包含很大篇幅的文本内容和少数的几幅图像。情感识别方面,第1个挑战是情感的鸿沟,即从特定模态上提取的特征,缺乏情感上的区分性;第2个挑战是感知的主观性,由于文化、性格等因素的差异,不同的人感受到相同的刺激时,产生的情感可能差别很大;第3个挑战是标签的噪声和缺失,由于主观性,情感标注通常需要统计多位参与者的投票,因此通常自动地获取网络上的标记或关键词作为标注。这种情况下,标签通常包含很多噪声。

多模态情感识别的计算方法包括情感模态的表示方法和情感模态的融合方法、多模态情感识别的分类方法、多模态情感识别的领域自适应方法。情感模态的表示,是存储和利用模态信息的基础。关于文本信息,词语表示是分析文本内容的基础。目前主要通过词到向量(word to vector, word2vec)、全局向量(global vectors, GLOVE)和基于变换器的双向编码器表示技术(bidirectional encoder representations from transformers, BERT)等训练方法获取词语的表示向量(Xu等,2019;Poria等,2017;Zadeh等,2018b)。关于音频信息,先转换成频谱图等图形化的表示,再用卷积神经网络(convolutional neural network, CNN)提取特征的方法在很多大规模任务中有很好的表现;关于图像信息,关注图像情感区域的表示方法有很好的竞争力(Yang等,2018a)。另外,如果图像中包含人脸信息,那么人脸表情则是一个很有用的线索;关于视频信息,由于包含一段有序的图像,最近很多工作使用加入时序的3维卷积提取特征表示;关于生理信号,脑电图是其中最具有代表性的一种信号。脑电图由多个通道组成,因此加入通道注意力的卷积神经网络可以更有效地提取表示信号的特征。

情感模态的融合是多模态情感识别的一个关键步骤,旨在结合不同情感模态的表示信息,完成情感识别任务。一般来说,情感模态融合包含模型无关融合、基于模型融合两类方法。模型无关的融合方法可以分为先融合、晚融合、混合融合3类。先融合也称为特征融合,将不同情感模态的表示在时间上同步后,连接为单个特征表示;后融合也称为决策融合,将每个情感模态的识别结果进行集成。与先融合相比,后融合具有更好的灵活性和鲁棒性。混合

融合使用一个统一的框架将先融合与后融合结合起来并利用二者的优势,这种情况下进行计算的成本较高。目前,由于模型无关的融合方法难以表示多模态数据的复杂情况,基于模型的融合方法获得了更多的关注。针对浅层模型,基于支持向量机(support vector machine, SVM)等核函数和基于图的融合方法最具有代表性;针对深层模型,通常使用基于张量计算、注意力机制和神经网络的方法进行融合。

针对多模态情感识别的计算方法,本文在5个通用数据集上进行了非深度方法与深度方法的定量比较,在表1中展示结果。比较的非深度方法包括支持向量机(SVM)、随机森林(random forest, RF)和三模态隐马尔可夫(tri-modal hidden Markov model, THMM),深度方法包括多视图长短期记忆(multi-view long short-term memory, MV-LSTM)、双向上下文长短期记忆(bi-directional contextual LSTM, BC-LSTM)、张量融合网络(tensor fusion network, TFN)、多注意力循环网络(multi-attention recurrent network, MARN)和记忆融合网络(memory fusion network, MFN)。观察表1中的结果可以发现,与非深度的分类方法相比, MARN(Zadeh等,2018b)、MFN(Zadeh等,2018a)等5个深度的方法在多数情况下具有更好的表现。另外,数据集不同,性能最好的模型也不同。在YouTube(Morency等,2011)、ICT-MMMO(Wöllmer等,2013)数据集上, MFN有明显的优势;在IEMOCAP(Busso等,2008)数据集上, MARN更有竞争力。另外,在CMU-MOSI(Busso等,2008)数据集上使用基于Transformer进行词嵌入的多模态情感分类方法进行定量实验,在表2中展示结果,其中FT(fine-tuning)和MAG(multimodal adaptation gate)分别代表微调和多模态适应门。表2采用了BERT、超长网络(extra long network, XLNet)两种词嵌入方法。与表1结果相比,基于Transformer的深度网络明显具有更好的性能,这是目前最有代表性的方法。通过进一步比较可以发现,基于XLNet的方法具有更稳定的优势。

多模态情感识别的领域自适应方法,旨在从经过标注的源域上学习模型,将模型迁移到无标注的目标域上同样可以有很好的表现。领域自适应的方法是用来解决标签噪声挑战的一个重要手段(Yu等,2021)。目前的工作致力于深度无监督领域自

表 1 非深度方法与深度方法在 5 个常用多模态情感识别数据集的定量比较

Table 1 Quantitative comparisons of some representative non-deep and deep methods on five widely-used multi-modal emotion recognition datasets

数据集	CMU-MOSI					YouTube		ICT-MMMO		MOUD		IEMOCAP					
train : val : test	1 284 : 229 : 686					30 : 5 : 11		11 : 2 : 4		49 : 10 : 20		3 : 1 : 1					
	$A_2 \uparrow$	$F1 \uparrow$	$A_7 \uparrow$	$M \downarrow$	$C \uparrow$	$A_3 \uparrow$	$F1 \uparrow$	$A_2 \uparrow$	$F1 \uparrow$	$A_2 \uparrow$	$F1 \uparrow$	$A_9 \uparrow$	$F1 \uparrow$	$M_v \downarrow$	$C_v \uparrow$	$M_A \downarrow$	$C_A \uparrow$
SVM	71.6	72.3	26.5	1.10	0.56	42.4	37.9	68.8	68.7	60.4	45.5	24.1	18.0	0.25	0.06	0.55	0.54
RF	56.4	56.3	21.3	-	-	49.3	49.2	70.0	69.8	64.2	63.3	27.3	25.3	-	-	-	-
THMM	50.7	45.4	17.8	-	-	42.4	27.9	53.8	53.0	58.5	52.7	23.5	10.8	-	-	-	-
MV-LSTM	73.9	74.0	33.2	1.02	0.60	45.8	43.3	72.5	72.3	57.6	48.2	31.3	26.7	0.26	0.02	0.51	0.62
BC-LSTM	73.9	73.9	28.7	1.08	0.58	47.5	47.3	70.0	71.1	72.6	72.9	35.9	24.1	0.25	0.07	0.59	0.40
TFN	74.6	74.5	28.7	1.04	0.59	47.5	41.0	72.5	72.6	63.2	61.7	36.0	34.5	0.25	0.04	0.52	0.55
MARN	77.1	77.0	34.7	0.97	0.63	54.2	52.9	86.3	85.9	81.1	81.2	37.0	35.9	0.24	0.10	0.50	0.65
MFN [*]	77.4	77.3	34.1	0.97	0.63	61.0	60.7	87.5	87.1	81.1	80.4	36.5	34.9	0.24	0.11	0.48	0.65

注: A_N 指情感识别的准确率, N 代表情感类别的数量。 C 指皮尔逊系数, M 代表结果与真实标注的平均绝对值误差, 其中下标 V 和 A 分别代表与标注中的积极性 (valence)、激烈程度 (arousal) 相关。 \uparrow 表示对应的值越大, 模型性能越好; \downarrow 表示对应的值越小, 模型的性能越好。 $-$ 表示无数据, train、val、test 为训练集、校验集、测试集。

表 2 在 CMU-MOSI 数据集上分别使用 BERT 和 XLNet 作为词嵌入对多模态情感识别分类方法定量比较结果

Table 2 Quantitative comparisons of some representative methods for multi-modal emotion recognition on the CMU-MOSI dataset using BERT and XLNet as word embeddings

基于 BERT 计算结果/基于 XLNet 计算结果				
	$A_2 \uparrow$	$F1 \uparrow$	$M \downarrow$	$C \uparrow$
TFN	74.8/78.2	74.1/78.2	0.96/0.91	0.65/0.71
MARN	77.1/78.3	77.9/78.8	0.94/0.92	0.69/0.71
MFN	78.2/78.3	78.1/78.4	0.91/0.90	0.70/0.71
FT	83.5/84.7	83.4/84.6	0.74/0.68	0.78/0.81
MAG	84.2/85.7	84.1/85.6	0.71/0.68	0.80/0.82

适应的方法,采用两条分支的结构。一条分支在标记的源域上训练多模态情感模型,另一条分支对齐源域与目标域的数据。

国际上比较有代表性的研究团队包括美国卡内基梅隆大学的 Louis-Philippe Morency 教授团队、英国伦敦帝国理工学院的 Björn W. Schuller 教授团队、美国罗彻斯特大学的 Ehsan Hoque 教授团队、新加坡南洋理工大学的 Erik Cambria 教授团队、新加坡科技设计大学的 Soujanya Poria 教授团队等。上述团队的研究方向各有侧重,例如 Louis-Philippe

Morency 教授团队关注多模态情感识别的计算方法,Björn W. Schuller 教授团队关注在开放环境的情感分析,Ehsan Hoque 教授团队、Erik Cambria 教授团队关注对话中的情感识别,Soujanya Poria 教授团队关注情感模态的融合方法。

2.2 孤独症情感识别

目前,孤独症情感识别主要集中在面部情绪识别。面孔是人类表达、认知情感的重要工具和途径,正确的表达和识别面部表情是人类生存的一项重要技能,面部表情的识别和理解能力对于儿童有着重要的社会意义 (Theeuwes 和 Van der Stigchel, 2006; Tomasello 等, 2005)。与正常儿童相比,孤独症患儿在面部信息识别方面存在着障碍,例如孤独症患儿不能识别和理解表情,进而无法进行正常的社交或情感交流。随着计算机视觉技术的发展,目前已经存在多种情绪/情感分类算法,它们在孤独症筛查和干预方面发挥着重要作用。然而,这些算法主要是在成人数据集上进行训练的,因此无法推广到孤独症患者进行应用。为了解决上述问题,美国斯坦福大学 Kalantarian 等人 (2019) 通过一种移动猜谜游戏来收集和标记孤独症患儿的情感数据。实验结果表明,分类器在孤独症数据集上的识别能力能够得到较大提升。随后,通过与主流云提供商相互合作,Kalantarian 等人 (2020) 将他们的面部情感

分析模型提供给消费者以便进一步对模型进行迭代和优化。同样来自美国斯坦福大学, Washington 等人(2019)认为目前的大多数工作都侧重于通过面部感知和学习来训练孤独症患儿的情感识别能力, 较少有方法通过协作/交互的方式让孤独症患儿能够意识到情感的存在。因此, 他们通过一个协作游戏告知孤独症患儿彼此间的情绪, 以减轻患儿的压力, 使其快乐参与游戏。与利用视觉进行情绪识别不同, 美国斯坦福大学的其他一些学者(Nag 等, 2020)通过可穿戴智能眼镜分析孤独症患儿的凝视模式以识别他们的情绪, 作者认为可穿戴技术或许能够将情绪识别带入自然的社交互动和环境中, 因此能够更好地探索情绪对社交的影响。美国麻省理工学院的学者在孤独症筛查和干预方面也有很深入的研究, 例如 Rudovic 等人(2018)利用深度学习技术定制了个性化的机器学习框架, 用于在机器人辅助自闭症治疗期间自动感知儿童的情感状态和参与度。此外, Narain 等人(2020a, b)围绕非语言发声(如叹息、咕噜声和单音节声音)进行研究, 他们认为对于患有轻微孤独症的人, 这些发声包含重要的个人情感和交换信息。英国伦敦学院 Palser 等人(2021)认为孤独症患者在理解情绪方面是有差异的。通过实验, 他们发现孤独症儿童身体情感地图的潜在差异与内在的感受信号处理有关(如心跳), 因此可以通过内在的生理信号分析这种差异化。美国匹兹堡大学 Conner 等人(2020)研究发现许多孤独症患儿在管理情绪(情绪调节)和焦虑方面存在问题。通过1 000份调查问卷, 他们认为虽然情绪调节和焦虑是相互关联的, 但情绪调节问题可能是治疗自闭症焦虑症的重要目标。英国纽卡斯尔大学 Ola 和 Gullon-Scott(2020)研究发现感知和体验情绪共享相同的神经系统, 因此能够识别自己的情绪可能有助于识别他人的能力。上述国际研究大都在探究影响孤独症患儿的外在和内在因素, 进而分析导致孤独症典型症状的原因, 为临床的筛查和干预提供依据。

2.3 情感图像内容分析

情感图像内容分析一般包括数据集构建、情感特征提取和分类模型学习等步骤(Zhao 等, 2021)。在 AICA 研究初期, 数据集通常来自于心理学或艺术学领域, 数据集规模通常较小, 例如 IAPS(Lang 等, 1997)、Abstract(Machajdik 和 Hanbury, 2010)、

GAPED(Geneva affective picture database)(Dang-Glauser 和 Scherer, 2011)和 MART(Museum of Modern and Contemporary Art of Trento and Rovereto)(Alameda-Pineda 等, 2016)等。其中 IAPS(Lang 等, 1997)是最常用的视觉情感刺激数据集, 它来自于情感实验分析和心理学注意力研究中, 包含 1 182 幅现实风格自然图像, 由约 100 位本科生标注为 9 种情感等级。IAPS 的一个子集 IAPSa(Mikels 等, 2005)由 20 位本科生标注为 8 种离散情感类别。随着社交网络的快速发展, 多个大规模数据集通过爬取网络数据的方式产生, 例如 FI(You 等, 2016)、VSO(Borth 等, 2013)、Emotion6(Peng 等, 2015)、T4SA(Twitter for sentiment analysis)(Vadicamo 等, 2017)和 LUCFER(labeled University of Central Florida emotion recognition)(Balouchian 等, 2019)等。VSO 数据集(Borth 等, 2013)用 1 000 多个形容词—名词对作为检索词从 Flickr 搜索并下载约 50 万幅图像。标注图像的 ANP 包含在图像对应的标题、标记或者描述等元数据中。采用 Plutchik 轮盘的 8 种基本情感和 3 种强烈程度作为情感模型。Emotion6(Peng 等, 2015)同样使用来自于 Flickr 的图像, 基于优势度—激活度(valence-arousal, VA)分数和离散的情感分布两种情感模型进行标注。FI(You 等, 2016)是一个基于 Mikels 的情感模型构建的常用的大规模情感图像数据集, 其包含的图像通过使用 8 种情绪作为关键词在 Flickr 和 Instagram 搜索得到, 并雇佣 200 余位 AMT 员工标注图像。

情感特征提取在 AICA 中起到至关重要的作用。研究者们提出了多种有效的视觉特征(从传统的手工特征到近期的深度特征)来表示情感。研究初期, 各种低层次特征用来表示情感内容。例如, Machajdik 和 Hanbury(2010)融合了不同类型的特征, 包括各种颜色和纹理表示, 是低层次特征发展过程的一个里程碑; Lu 等人(2012)系统地研究视觉形状对于图像情感的影响; Sartori 等人(2015)基于 Itten 的颜色轮盘, 研究了抽象画中不同的颜色组合与图像情感之间的关联。低层次特征容易从已有的计算机视觉特征中迁移而来, 但却缺乏合理的解释, 并且与情感之间的关联较弱。相较于低层次特征, 中层次情感特征更易于理解, 并且与情感之间的关联更加强烈。例如, Yuan 等人(2013)提出了一个称为 Stribute 的中层属性特征, 并且考虑了检测到

的人脸表情; Rao 等人(2016)考虑了基于多尺度结构的情感特征挖局,使用不同的分割方法提取每幅图像的多尺度模块,然后提取图像的尺度不变特征变换(scale-invariant feature transform, SIFT)特征,使用视觉词包(bag of visual words, BoVW)的方法编码每个模块,结果证明 BoVW 可以较好地描述不同区域的情感信息。高层次特征指容易理解且可以直接使观看者产生情感的图义信息。最具代表性的高层特征是 SentiBank, 一个大规模视觉情感本体论。它包含 1 200 个概念,每一个概念表示一个形容词—名词对,例如可爱的小孩,这些概念包含很明显的语义信息。Jou 等人(2015)对 SentiBank 进行了扩展,提出了一个大规模多语言视觉情感本体论。

近年来,随着卷积神经网络的快速发展,基于学习的特征在 AICI 研究中越来越受重视。此类方法大致可分为两类:全局特征和局部特征。全局特征平等地处理和对待图像中每个区域;基于心理学中关于情感区域的研究,局部特征侧重于提取包含丰富信息的局部特征。Chen 等人(2014)使用卷积神经网络对 1 200 个形容词—名词对进行分类,所提出的基于深度网络的 DeepSentiBank 模型取得的性能明显好于不用深度网络的 SentiBank。Xu 等人(2014)将大规模数据集(ImageNet)训练的卷积神经网络参数迁移到情感预测任务,发现 FC7 层激活后得到的特征优于 FC8 层。You 等人(2015)所提出的基于弱标记图像训练的渐进卷积神经网络极具代表性,将预测结果中情感极性有较大差异的训练样本保留到下一轮训练,使得噪声数据逐渐移除。Rao 等人(2020)提出了一个包含 AlexNet、美学网络、纹理网络 3 条并行分支的端到端网络结构,通过融合卷积神经网络生成多层深度特征,进而捕捉图像不同类型的信息。局部特征因其捕捉包含丰富信息的情感区域的能力,受到了越来越多的关注。Chen 等人(2015)考虑细粒度的信息,利用不同尺度提取局部块的特征,然后通过 Fisher Vector 集成为一个统一的表示。Liu 等人(2016a)在计算视觉显著性区域时检测人脸表情和情感的物体,一起构成情感因子。You 等人(2017)在可描述的图像属性基础上,使用注意力机制发现引起观看者情感的局部区域,从这些区域中提取的特征提高了 AICA 的性能。Zhao 等人(2019a)将基于空间注意力和通道注意力的特征连接后作为优势度—激活度—控制度

(valence-arousal-dominance, VAD)视觉情感回归任务的表示。为了有效地使用不同层得到的信息, Rao 等人(2019)提出了一个多层次基于区域的卷积神经网络框架,找出局部区域的情感表示。研究表明,全局特征和局部特征都可以在一定程度上决定图像的情感,结合局部特征应该比全局特征会生成更具判别性的情感图像表示(Zhao 等,2021)。

情感特征提取之后学习分类模型。在 AICA 领域,已有分类模型多数集中在大众化情感识别、个性化情感预测和情感分布学习,以及从噪声数据或少量标签中学习等任务。早期方法大多使用传统分类器对大众化情感进行识别,常用的分类器包括支持向量机 SVM (Zhao 等, 2014)、朴素贝叶斯 (Machajdik 和 Hanbury, 2010) 和非线性矩阵补全 (Alameda-Pineda 等, 2016) 等。例如, Ahsan 等人(2017)通过训练的卷积神经网络模型检测活动概念,然后使用 SVM 分类器将视觉属性映射到具体的情感。近期研究大多通过设计各种各样基于学习的方法来识别图像情感。考虑到情感的特性,研究者通过改进传统的交叉熵损失函数和均方误差损失函数来改进 AICA 的性能。例如, Zhao 等人(2019a)提出了极性一致的均方误差损失函数用于图像情感的回归。Fan 等人(2017)提出一个称为焦点通道的第 4 条通道,用于图像中焦点物体的掩码,或者显著性图表示,结果证明消极情感主要是由焦点区域引出的,图像的其他区域很难造成影响,而积极情感则由焦点区域和其他区域一起决定。为了解决情感感知的主观性问题, Peng 等人(2015)使用概率分布的标注方法构造了 Emotion6 数据集,并且使用卷积神经网络回归作为情感回归模型。Wang 等人(2015b)根据视觉内容和相关文本信息的关系,在网络图像上进行无监督情感分析。领域自适应研究如何将带有标注的源域数据上训练的模型,迁移到另一个稀疏标注或无标注的目标域。Zhao 等人(2018b)首先研究了情感分布学习的域适应问题,所提出的 EmotionGAN 对抗模型通过生成与目标域相近的中间域来实现源域和目标域的像素级对齐。语义一致性约束保证了中间域的图像能够保留源域的情感信息。此外, Zhao 等人(2019b)利用 CycleGAN 代替 GAN,来解决传统 GAN 的不稳定性和容易训练失败的问题,在不需要匹配图像对的情况下,学习大众化情感的跨域迁移。Lin 等人(2020)研

究了图像情感二分类任务的多源域迁移问题,设计了一个多源域情感对抗生成网络,挖掘多源域、目标域图像具有相似分布的统一情感隐藏空间。Panda 等人(2018)研究了情感图像内容分析的域泛化问题,以此解决不同数据集图像存在明显差异的问题。

国际上比较有代表性的研究团队包括美国哥伦比亚大学 Shih-Fu Chang 教授团队、美国加州大学伯克利分校 Kurt Keutzer 教授团队、美国宾夕法尼亚州立大学 James Z. Wang 教授团队、美国罗彻斯特大学 Jiebo Luo 教授团队、美国中佛罗里达大学 Hassan Foroosh 教授团队、美国康奈尔大学 Tsuhan Chen 教授团队、意大利特伦托大学 Nicu Sebe 教授团队、奥地利信息工程研究所 Allan Hanbury 教授团队、澳大利亚悉尼科技大学 Min Xu 教授团队、新加坡国立大学 Tat-Seng Chua 教授团队等。上述团队的研究侧重点各有不同,例如,Shih-Fu Chang 教授团队和 Jiebo Luo 教授团队主要关注社交网络中图像情感的分析,James Z. Wang 教授团队、Allan Hanbury 教授团队和 Nicu Sebe 教授主要研究如何设计有效的手工特征,Tsuan Chen 教授团队主要关注情感分布的研究,Tat-Seng Chua 教授团队主要研究情感在推荐系统中的应用,而 Kurt Keutzer 教授团队主要关注 AICA 中的领域自适应问题。

2.4 面部表情识别

随着计算机视觉、模式识别和人工智能等学科的发展,目前领域在基本表情上真实世界中的识别结果已经可以达到很高的水平。例如在 RAF-DB 数据集上(Li 等,2017;Li 和 Deng,2019b),最新技术已将 7 类基本表情分类结果提高到 90% 以上(Xue 等,2021)。然而,越来越多的研究表明,基本表情并不能完整包含人类所有情绪。心理学研究中还有一些更为复杂且全面的情感描述形式。普拉切克的情绪轮盘模型指出,单个基本情绪具有不同的强度,并且可以互相混合,从而派生出更多复杂的情绪(Plutchik,2001)。另一种维度描述模型则是将大量不同类别的情感投射到一个连续的维度空间里(Mehrabian,1996)。这些研究都表明,传统的 7 类基本表情过于局限,无法涵盖自然交流场景中更多复杂的情绪。为了进一步拓展表情模型的描述范围,相关研究通过将不同基本表情组合起来,提出了混合表情的概念(Nummenmaa,1988;Martin 等,2006)。其中,复合表情作为一项特例,指由两种不

同基本表情组合而来的表情类别(Du 和 Martinez,2015;Du 等,2014)。

国际上有很多广泛用于算法评估的真实世界人脸表情数据集和相应评估准则。FERPlus 数据集(Barsoum 等,2016)在 FER2013(Goodfellow 等,2013)的基础之上通过众包对其进行了更加精准的 8 类基本表情标注(7 类基本表情和蔑视表情)。该数据集包含 28 558 幅训练图像,3 579 幅验证图像和 3 573 幅测试图像。EmotiNet(Benitez-Quiroz,2016)是一个包含从互联网收集的一百万幅面部表情图像的大型数据库。其中有 950 000 幅图像由机器自动标注上动作单元(AU)标签,而剩余 250 000 幅图像则是由手动标注了 11 类 AU。该数据库所提供的 23 类基本表情和复合表情标签并非由标注者直接标注,而是由 AU 组合推断而来,因此存在较大误差。AffectNet 数据集(Mollahosseini 等,2019)包含从互联网收集而来的超过一百万幅图像,其中 450 000 幅图像具有手动标注的 8 类基本表情标签(7 类基本表情和蔑视表情)。此外,该数据库也对图像进行了维度空间的标注。AFEW 7.0(Dhall,2019)包含由电影片段剪辑而来的 1 809 个视频数据,其中分别用于训练、验证和测试的视频数分别为 773,383 和 653,每一个视频也进行了 7 类基本表情的标签标注。ExpW(Zhang 等,2018)数据集则包含了 91 793 幅从互联网收集而来的图像数据,每幅图像被标注上了 7 类基本表情标签。Aff-wild2(Kollias 和 Zafeiriou,2019)是第 1 个同时针对效价—唤醒 2 维连续情感估计、7 类基本表情识别和面部动作单元检测这 3 种任务都进行标注的真实世界数据集。其中有 558 个视频包含了效价—唤醒度标注,63 个视频包含了 8 类 AU 标签,84 个视频包含了 7 类基本表情标签。

2.5 面部微表情分析与识别

本节从数据库构建、微表情预处理方法、微表情检测和微表情识别几个方面阐述相关的进展。

2.5.1 微表情数据库构建

考虑到微表情产生的机理以及相关研究目前仍然处于初期发展阶段,如何构建合理的数据集一直是研究者们关注的焦点问题。芬兰 Oulu 大学的团队在 2013 年扩展了此前的 Pfister 等所用的数据集(Pfister 等,2011),并正式发布了 SMIC(spontaneous micro-expression database)(Li 等,2013)。Husák 等

人(2017)对视频网站上的扑克牌游戏视频进行标注,构建了 MEVIEW (microexpression videos in the wild)数据集。Davison 等人(2018)构建了 SAMM (the spontaneous actions and micro-movements dataset)数据集。

2.5.2 微表情预处理方法

微表情存在动作幅度小、持续时间短的特点,往往需要专门的方法进行相应预处理。因此微表情预处理也成为微表情分析任务中一个显著技术难点。其中,拉格朗日视频运动放大法采用拉格朗日视角(Lagrangian view)对运动进行描述和操作并被引入微表情分析领域(Ngo 等,2018)。

2.5.3 微表情检测

在自动微表情分析系统中,在进行微表情识别任务之前需要先进行微表情检测(Li 等,2018)。微表情检测,可以分为对于脸部整体运动的检测和基于脸部运动单元(AU)(Ekman 等,2002)的检测。如局部描述子的差分(Moilanen 等,2014)、图像区域的亮度变化(Husák 等,2017)以及光流的运动向量(Patel 等,2015)等都已经用于检测微表情。

2.5.4 微表情识别

国际上早期微表情识别研究普遍使用手工设计特征,如局部二值模式(LBP)(Li 等,2013)及其时域扩展 LBP-TOP(Pfister 等,2011)进行识别。除此之外,几何特征(Pfister 等,2011)、基于光流特征的运动信息(Happy 和 Routray,2019)也受到了越来越多的关注。随着机器学习乃至深度学习在微表情分析领域的应用,Patel 等人(2016)使用特征迁移将深度学习技术引入微表情识别,并基于特征选择技术选择在最坏情况下性能最好的特征组合以缓解深度学习模型在小样本微表情数据集上的过拟合问题。该工作一般被认为是深度学习在微表情识别领域的首次成功应用。Kim 等人(2016)将卷积神经网络(CNN)和长短期记忆网络结合,分别利用 CNN 和 LSTM 提取空间和时间特征。早期的深度学习方法在性能上尚不能与传统方法性能相比。为此,研究者们持续改进基于深度学习的微表情识别方法与技术,使其发展迅速。

3 国内研究进展

3.1 多模态情感识别

多模态情感识别在国内也引起了广泛的关注。

多模态情感识别的数据集方面,中国科学院自动化研究所的毛文吉研究员团队从中关村在线网站收集了 28 469 条评论,构建基于汉语的 Multi-ZOL 图文情感数据集(Xu 等,2019);清华大学的徐华教授团队从影视剧和综艺节目中搜集 2 281 个视频片段,构建基于汉语的 CH-SIMS 视频情感数据集(Yu 等,2020)。这些数据集为基于汉语文本的多模态情感识别发展奠定了基础。

多模态情感识别的计算方法方面,毛文吉研究员团队结合情感倾向,提出基于深度网络的特征融合方法(Xu 等,2019)。分别提取各模态的特征,送入多层交互记忆网络。在网络的每一层中都对不同模态的特征进行交互,实现跨模态的融合。为了解决长时间序列多模态融合的遗忘问题,中山大学的胡海峰教授团队将多模态情感数据按照时间划分为多个部分,并对每一个时间块的多模态数据进行显式融合(Mai 等,2022)。哈尔滨工业大学的秦兵教授团队在 Wu 等人(2021)提出的以文本为主的多模态信息中融合新思路。与只使用文本信息进行情感识别的方法相比,非文本信息从共享语义、独享语义两个方面提高识别能力。一方面,图像、音频等模态的语义与文本提供的语义相同,这些重复的信息可以对原有语义进行增强,这是多模态信息的共享语义;另一方面,非文本模态可以提供与文本语义不同的信息,这些语义信息可以帮助模型更准确地进行情感识别,这是多模态信息的独享语义。基于这两种多模态的语义信息,提出一个以文本为中心的多模态融合框架。厦门大学的纪荣嵘教授团队在 Ji 等人(2019)提出双层的多模态超图情感识别方法中,显式地对不同模态之间的相关性建模。在模型的第 1 层学习数据的特征和相关性进行情感识别,在模型的第 2 层对各模态特征的相关性进行学习。清华大学的徐华教授团队在 Yu 等人(2020)研究中证明单模态的标签可以为多模态情感识别提供帮助。因此在 Yu 等人(2021)的研究中,徐华教授团队提出自监督训练单模态情感预测的方法。在训练的过程中,为多模态和每个单一模态分别维护标签中心,基于每个单一模态标签与中心的距离,应与多模态标签与中心距离一致的假设,为每个单一模态提供训练过程中的监督信息。在预测阶段单模态预测结果完成识别任务。

多模态情感识别的应用方面,毛文吉研究员团

队提出了针对用户评价情感倾向分析的数据集和模型方法(Xu等,2019)。之前的工作全部使用单一文本模态进行情感倾向分析,这是首次使用多模态信息进行情感倾向分析的工作。厦门大学的纪荣嵘教授团队针对微博数据进行多模态的情感分析(Chen等,2018a)。使用数据中包含的表情符号作为噪声标签,使用概率图形模型提取具有情感区分性的多模态特征并过滤标签中的噪声。南开大学的杨巨峰教授团队提出了针对用户旅途情感倾向分析的数据集和模型方法(Wang等,2021a)。在这篇工作中,将旅途分为值机、等待、途中和延迟4种事件。使用深度模型分别提取文本特征、图像特征,经过先融合后,预测情感倾向和事件类型。

国内比较有代表性的研究团队包括清华大学的徐华教授团队、中国科学院自动化研究所的毛文吉研究员团队、哈尔滨工业大学的秦兵教授团队、厦门大学的纪荣嵘教授团队、中山大学的胡海峰教授团队和南开大学的杨巨峰教授团队等,其中各个团队的侧重点有所不同。如徐华教授团队关注多模态情感识别的分类方法、毛文吉教授团队关注多模态情感倾向分析、秦兵教授团队关注文本为主的多模态情感分析方法、纪荣嵘教授团队关注社交平台的多模态情感识别、胡海峰教授关注多种情感模态的融合方法、杨巨峰教授团队关注包含图文信息的多模态情感识别方法。

3.2 孤独症情感识别

在国内,也有众多的学者对孤独症患儿的情感识别和理解进行研究,并提出了一些用于孤独症情感干预的方法和技术手段,他们大都认为孤独症患儿的情感表达和其社交障碍之间存在着关联。香港协康会提出“情感表达和互动”是孤独症干预比较困难的部分,因此将“情感表达”分解为多个次范畴,如面部表情、身体语言和恰当的情绪等,然后通过设计针对性的干预范式对不同的次范畴进行训练。台湾成功大学Tsai等人(2021)通过使用3D虚拟场景提高孤独症儿童正确识别情绪的能力。同时,他们使用第三人称视角角色扮演游戏来教授孤独症患儿社交技能并帮助他们加深理解6种(愤怒、恐惧、惊讶、厌恶、快乐和悲伤)基本情绪。中山大学附属第三医院邹小兵医生(邹小兵,2019)也曾指出矫正不良情绪和行为是突破自闭症社交障碍的重要一环,他呼吁家长对孤独症患儿进行情绪调控。

南开大学王崇颖教授(Sullivan和Wang,2020)同样认为社交情绪干预是孤独症干预的核心问题之一。北京大学易莉研究员(Wang等,2018)探讨了孤独症患儿的眼睛回避是否会受到情绪表达的影响。实验结果表明当面对愤怒的面孔时,孤独症患儿比正常儿童对眼睛的注意力更少,而且持续时间长,这项研究不仅扩展了凝视厌恶假设,而且对孤独症的治疗和筛查也有影响。昆山杜克大学李明教授等人(Pan等,2021)设计了“社交性微笑”的筛查范式,他们通过集成语音识别和计算机视觉技术,提出了一个机器学习框架以实现在特定范式下对受试儿童的行为进行分析。华东师范大学陈靛影教授等人(廖梦怡等,2021)探索了融合多模态数据的孤独症患儿智能化识别方法,该方法可以根据数据来源和时间同步性将数据进行分层融合,因此可以提高孤独症患儿的识别准确率。同样适用多源数据融合,哈尔滨工业大学(深圳)刘洪海教授等人(Wang等,2021b;Liu等,2020)通过“五不(不看、不应、不指、不说、不当)”行为、社交情感等设计出了国内首台孤独症早期辅助筛查平台和机器辅助干预系统,实现孤独症早期筛查自动化评估和干预。

对于通过行为对孤独症患儿进行诊疗的方法,社交情绪/情感是其中非常重要的一环,它对提升孤独症患儿的社交能力非常重要。目前,如何捕捉和量化孤独症患儿社交情绪非常困难,国内外尚未出现针对孤独症社交情绪诊疗的机器智能辅助系统。

3.3 情感图像内容分析

国内研究者针对情感图像内容分析的研究也涵盖了数据集构建、情感特征提取和分类模型学习等步骤。国内构造的数据集具有代表性的是哈尔滨工业大学姚鸿勋教授团队的 IESN (image-emotion-social-net)数据集(Zhao等,2014)和南开大学杨巨峰教授团队的 Comics数据集(She等,2019)、Flickr_LDL数据集和 Twitter_LDL数据集(Yang等,2017b)。IESN数据集构建的目的是实现个性化的情感预测,包括从一万余位用户上传至 Flickr 的一百万余幅图像,并且收集了与图像相关的标记、描述、评论和上传者的社会背景等各种元数据。数据集的每幅图像都提供了上传者想表达的情感,以及观看者实际感受到的情感。基于关键字搜索,该模型使用 VAD 模型和 Mikels 模型进行标注。Comics(She等,2019)包括从海贼王、蜘蛛侠等70幅漫画中选出的11821

幅图像。10位参与者(平均年龄20.3岁)使用 Mikels 的8种情绪类别标注。将该数据集进一步分为两个子集:动漫子集包含欧洲、美国的现实风格动漫,漫画子集包含亚洲漫画风格的抽象图像。Flickr_LDL 数据集(Yang 等,2017b)和 Twitter_LDL 数据集(Yang 等,2017b)是针对情感的模糊性和多义性提出的数据集。Flickr_LDL 从 VSO 数据集抽取 10 700 幅图像,11 位参与者观看每一幅图像,然后根据 Mikels 情感进行标注。Twitter_LDL 通过多个情感关键词在 Twitter 数据集上获取数据,8 位观看者对 10 045 幅图像使用 Mikels 情感进行标注。

国内研究者很早就开始了低层次手工情感特征的设计,华南理工大学王伟凝教授团队研究了线条方向与图像情感之间的关联(Wang 等,2004);基于心理学的颜色理论,他们在一个正交的3维情感空间中构造了亮度—冷系—暖系、饱和度—冷系—暖系—对比度以及对比度—锐度表示(Wang 等,2006)。在中层特征提取方面,国内研究者也有涉及。例如,清华大学贾珈教授团队设计了一组可解释的、容易理解的特征(Wang 等,2013),用于表示图像前景区域与背景区域的对比;哈尔滨工业大学姚鸿勋教授团队提出了基于艺术原理的特征(Zhao 等,2014),通过量化平衡、强调、和谐、多样和渐变等艺术原理,改进了图像情感识别的性能。近年来,国内在深度情感特征提取方面也展开了系列工作。例如,中国科学院大学黄庆明教授团队使用卷积神经网络的不同层来提取多层次特征(Zhu 等,2017),并且使用双向门控循环单元结构来捕捉不同层之间的依赖关系。南开大学杨巨峰教授团队在深度全局特征和局部特征提取上展开了多个原创性工作:提出了一个由不同层的 Gram 矩阵元素组成的情感表示方法(Yang 等,2018b);提出了使用离线物体检测工具生成候选边界框,结合在去重后区域中提取的特征和全局图像的特征进行情感分类(Yang 等,2018c);提出了一个包含分类分支和检测分支的统一结构,在检测分支通过结合所有基于类别的特征表示,生成弱监督的情感图,将全局特征和情感图耦合,得到完整的局部信息(Yang 等,2018a);在低层次和高层次分别添加极性注意力和情感注意力,通过跨层次的双线性池融合不同层次的特征,生成最终的情感表示(Yao 等,2019)。近年来,西安电子科技大学高新波教授团队在物体与情感关系的挖掘上

作出了重要贡献:基于心理学“刺激—机体—反应”框架,提出了选择图像中可能诱发情感的不同刺激,并为这些刺激提取不同的深度特征(Yang 等,2021b);提出了基于图卷积网络的场景—物体相关情感推理网络,来挖掘图像里物体与物体以及物体与场景之间的交互(Yang 等,2021a)。

在大众化情感分类模型上,西安电子科技大学高新波教授团队提出了层次化的交叉熵损失函数来加大对错误分类样本的惩罚(Yang 等,2021a)。国内研究者关于个性化情感预测和情感分布学习的研究走在了国际前列。具备代表性的关于个性化情感预测的两个工作出自清华大学朱文武教授团队和哈尔滨工业大学姚鸿勋教授团队。前者提出了基于用户兴趣和社会影响的个体情感预测模型(Yang 等,2013)。用户的兴趣通过个性化词典构造和基本颜色特征聚类得到,社会影响通过计算不同用户对相同微博的情感相似度得到。随后,该团队使用概率图模型扩展了赋予权重的过程(Rui 等,2017)。后者考虑了可能影响情感的多种因素(Zhao 等,2016,2018a),如:图像的视觉内容、用户的社会背景、情感随时间的变化以及图像的位置信息等。结合这些因素,文章提出了迭代多任务超图学习方法,对用户、目标图像和历史图像集构建超边,通过半监督学习方法同时为多个用户进行个性化情感预测。图像情感分布学习任务上比较具有代表性的国内团队包括:南开大学杨巨峰教授团队、清华大学丁贵广教授团队、哈尔滨工业大学姚鸿勋教授团队和天津大学刘安安教授团队。例如,姚鸿勋教授团队将情感分布学习当做共享稀疏学习问题建模(Zhao 等,2015);丁贵广教授团队提出了带权重的多模态共享稀疏学习(Zhao 等,2017a)和带权重的多模态条件概率神经网络(Zhao 等,2017b),自动学习不同特征的权重;杨巨峰教授团队提出了改进的条件概率神经网络 BCPNN(binary conditional probability neural network)和 ACPNN(augmented conditional probability neural network)(Yang 等,2017a),分别使用二进制编码替换单个整数表示标签和向真实标签添加噪声,使得训练的模型更加鲁棒;杨巨峰教授团队还提出了一个同时计算 Kullback-Leibler (KL) 损失和 softmax 损失,并且将单个情感标签转变成概率分布的统一框架(Yang 等,2017a);天津大学刘安安教授团队提出将低秩和协方差正则化加入一个框架中进

行情感分布学习,确保了回归系数的结构稀疏性(Liu等,2018)。丁贵广教授团队和姚鸿勋教授团队在连续的情感空间中使用混合高斯模型建模连续分布(Zhao等,2017c),用期望最大化算法预测参数值,使用共享稀疏回归(shared sparse regression, SSR)作为学习的模型,并且扩展至多任务SSR来挖掘不同任务之间的相关性,通过使用合适的跨任务共享参数来预测不同测试图像的参数。

国内比较有代表性的研究团队包括哈尔滨工业大学姚鸿勋教授团队、清华大学朱文武教授团队、丁贵广教授团队和贾珈教授团队、南开大学杨巨峰教授团队、中国科学院大学黄庆明教授团队、西安电子科技大学高新波教授团队和李雷达教授团队等。上述团队的研究各有侧重,例如姚鸿勋教授团队侧重在中层情感特征设计和个性化情感预测,丁贵广教授团队主要关注图像情感的离散和连续概率分布学习,杨巨峰教授团队在图像情感的离散分布学习和深度局部特征挖掘等多个方向都有涉及,高新波教授团队近年来主要挖掘图像内不同物体之间的关系及其与情感之间的映射。

3.4 面部表情识别

近年来,国内对人脸表情识别领域的发展有了很大的进展,包含一些广泛用于算法评估的真实世界人脸表情数据集和相应评估准则。如RAF-DB数据集(Li等,2017;Li和Deng,2019b)包含了从互联网下载的近3万幅高度多样化面部图像。通过手动的众包标注和标签可靠性估计,该数据库为样本提供了精确的7类基本表情标签和12类复合表情标签。RAF-ML(Li和Deng,2019a)是第1个真实世界混合表情数据集,其包含了类别更为丰富的4908幅多标签表情图像样本。RAF-AU(real-world affective faces action unit)(Yan等,2020)则是在RAF-ML数据集的基础上对其进行了26类AU标签的手动标注。

算法方面,也有一些具有代表性的算法。下面分别从不确定性学习、关系学习和解耦学习这几个方面对该领域内近年来具有代表性的表情分析算法进行介绍。

3.4.1 不确定性学习

表情标注具有主观性和差异性,因此数据集中的标签噪声难以避免。针对该问题,相关研究提出了样本的不确定性学习算法,通过自动纠正不确定

性大的样本标签,提高模型的泛化能力。Wang等人(2020a)通过全连接层学习来计算出每个样本对应的权重,并将该权重作用到softmax指数部分,使得不确定性高的样本对应的重要性权重更低,从而降低噪声样本带来的负面影响。最后通过对不确定性高的样本进行重标注实现噪声清洗。She等人(2021)进一步考虑了标签分布来解决标注模糊的问题。通过依次dropout某一类别的样本,分别学习独立的网络分支,来获取每个样本的潜在标签分布,从而巧妙地分散噪声标签的影响。此外,不确定性的绝对数值很难定义,但相对大小比较好确定。Zhang等人(2021e)使用不确定度作为权重对图像的特征进行加权混合,通过对比进行不确定度的学习,设计了一种特征mixup的机制,利用损失函数“公平”地从混合特征中同时识别出两种表情。在降低损失函数的过程中,通过整个训练中大量pair的对比,自动学习到图像的不确定度值。

3.4.2 关系学习

各种情感类别或者所学情感特征之间存在着关联,利用这种内在关联知识,可以在小样本条件下获得更加稳定的情感识别性能。标签分布学习是一种典型的类别间的关系学习方法。不同情感类别之间存在着潜在的相关关系,对此Wang和Geng(2021)提出了标签分布流形学习算法,通过挖掘标签分布中隐藏的流形结构来同时编码标签之间的全局和局部关系。除了利用类别间的关系,还可以结合图神经网络来建模潜在动作特征之间的关系。例如,Ruan等人(2021)利用图神经网络来计算每幅图像包含的潜在表情成分特征之间的关系。Song等人(2021)则是结合了多种关系学习方法。首先引入了贝叶斯网络来建模面部动作之间的关系,然后将该习得的分布作为图卷积的输入邻接矩阵来进一步建模特征之间关系。此外,随着Transformer技术(Vaswani等,2017)在计算机视觉中取得的进展和优秀成果(Dosovitskiy等,2020;Liu等,2021b),部分工作也将其运用到了表情分析中。Jacob和Stenger(2021)直接提取各个AU的表示特征作为Transformer的输入,利用Transformer的自注意力机制建模了AU之间的关系。Xue等人(2021)则是在视觉变换器(vision transformer, ViT)网络(Dosovitskiy等,2020)中结合了dropout机制,取得了目前最佳的表情识别结果。

3.4.3 解耦学习

在情感识别中,存在着人物(身份、年龄、性别和种族等)、采集噪声(遮挡、低分辨率等)和姿态变化等与情感无关的干扰因素,如何在情感特征中解耦出这些干扰噪声,从而获得能够有效建模情绪的判别性特征,是情感识别的根本问题。Xu 等人(2020)利用对抗学习分别获得了表情和各个脸型属性(种族、年龄和性别)的独有特征,通过人为去除人脸属性特征,获得了对各个属性公平的表情识别性能。Niu 等人(2020)提出了生理信号与噪声解耦的交叉检验方法,在有效地从信号中分离出噪声信号的同时,可以生产伪信号,起到样本增广的作用。此外,姿态的整体变化会淹没面部的细微动作,而细微动作正是反映情感的有效特征。对此,Li 等人(2022)提出了头部姿态变化与面部动作的分解方法,利用分解信号的合并验算以及姿态变化和面部动作分别构造正负变换,构造了一个基于视频的自监督 AU 学习框架。对于面部遮挡干扰,Wang 等人(2020b)提出采用注意力机制让网络降低对遮挡度高面部区域的学习权重。人物身份也是另一项待解耦因素,不同个体表达情感的方式以及其面部身份信息对表情识别有着较大干扰。Zhang 等人(2021d)采用偏差计算模块,直接从人脸特征中减去身份属性特征来获取身份不变特征。

3.5 面部微表情分析与识别

3.5.1 微表情数据库构建

国内方面,中国科学院心理研究所团队利用先后构建的 CASME(the Chinese Academy of Sciences Micro-expression)(Yan 等,2013)、CASME II(Yan 等,2014a)和 CAS(ME)²(Qu 等,2018)数据集,并与山东大学和复旦大学等的学者联合发布了 MMEW(micro-and-macro expression warehouse)数据库(Ben 等,2021)。从 SMIC(spontaneous micro-expression database)到 CASME(the Chinese Academy of Sciences micro-expression dataset)系列再到 MMEW 数据集,研究者们基本采用了一致的微表情诱发策略,即通过让被试在尽量保持中性表情的前提下观看带有较强情绪性的视频片段并伺机采集诱发的微表情。之前的数据库主要侧重微表情识别方面的验证,近期有一些工作也开始关注微表情检测(spotting)的数据和实验条件。由于 CAS(ME)²和 SAMM 同时提供了包含宏表情和微表情的长序列,

因此这两个数据集可以用来开展微表情检测的研究。近期 Tran 等人(2021)在 SMIC 数据集中长序列版本 SMIC-VIS-E 上进行了扩展和标注,提供了 SMIC-E-Long 数据集,并对测试协议和性能评价方法进行了探讨。

微表情分析的研究者一直朝着更多人、更高分辨率和更多微表情样本的方向努力构建数据集,近期的数据集还都增加了关于人脸运动单元。但目前主要的微表情研究数据集仍然只能提供涵盖 30~40 个人的大概 190~300 个微表情样本,这为深度学习模型的训练带来了极大的困难。为此,CAS(ME)²和 MMEW 数据集通过提供同时采集的宏表情和微表情样本来为深度模型的预训练提供一些额外的数据选择。此外,研究者们尝试设计新的测试协议,将多个数据库进行合并或者组合从而开展实验。这方面的尝试包括将 SMIC-HS、CASME II 和 SAMM 数据集进行合并的混合测试协议(See 等,2019)以及跨数据库测试协议(Yap 等,2018;Zong 等,2018a,2019;Xia 等,2019b;Zhang 等,2020)等。

3.5.2 微表情预处理方法

国内,欧拉视频运动放大方法(Eulerian video magnification, EVM)(Wu 等 2012)在 2015 年被引入微表情识别任务(Li 等,2015,2018)。Lei 等人(2020a)进一步尝试了基于学习的视频运动放大方法(Oh 等,2018),并取得较好的效果。在时域插值方面,基于图模型的插值方法(Zhou 等,2011,2014)得到广泛应用。在此基础上,研究者将运动放大和帧插值两个模块进行统一建模,形成了单一预处理模块。该方法不仅提高了运算速度,还进一步提升了识别性能(Peng 等,2019;Hong 等,2019)。Xia 等人(2019a)使用了多种运动放大倍数的预处理方式。此外,基于数据生成的预处理方式也开始受到关注。Xie 等人(2020)使用 AU 强度可控的生成对抗网络(generative adversarial network, GAN)生成。Liong 等人(2020)则使用条件生成对抗网络对提取的光流图进行增强。

3.5.3 微表情检测

国内关于脸部信息变化的有效刻画,如局部描述子的差分(Li 等,2018)、光流的运动向量(He, 2021)、光流主方向的模最大差分(Wang 等,2017)以及光流的夹角和模信息(Guo 等,2021b)等都被用来检测微表情。Yan 等人(2014b)、Liong 等

人(2015)、Han 等人(2018)以及 Li 等人(2021a)则尝试了针对峰值帧(apex frame)的检测方法。Li 等人(2020a)分析了微表情发生时的局部时域模式,并提出了一种基于 S-模式的微表情检测方法。基于动作单元 AU 的检测是近年来的热点,Zhang 等人(2021d)根据不同 AU 活动的面部范围设计了新的面部关键子区域分割方法,再根据分割后的关键区域,将全脸 AU 检测这个大型的多标签分类问题转化为多个在各个关键子区域内的小型多标签分类问题。Li 等人(2021b)引入 attention 机制进行 AU 的检测。

3.5.4 微表情识别

早期的微表情识别研究者普遍使用手工设计特征。广泛使用如局部二值模式(local binary pattern, LBP)(Li 等,2013)、STLBP-IP(spatiotemporal local binary pattern with integral projection)(Huang 等,2015)、STCLQP(spatiotemporal completed local quantized patterns)(Huang 等,2016)、HSTLBP-IP(hierarchical spatiotemporal local binary pattern with integral projection)(Zong 等,2018b)、DiSTLBP-RIP(discriminative spatiotemporal local binary pattern with revisited integral projection)(Huang 等,2019)等。Hong 等人(2016a)提出了 TOP 类特征的改进实现,加速了描述子的计算效率。Li 等人(2015,2018)对 LBP、HIGO(histograms of image gradient orientation)和 HOG(histograms of oriented gradients)的时空变种进行对比,发现只体现梯度方向信息而不反映梯度能量信息的 HIGO-TOP 描述子在微表情分析上相比 LBP-TOP 和 HOG-TOP 更为高效。在此基础上构建了第 1 个完整的微表情检测与识别的系统,并进行了人机对战实验。该工作受到了 MIT Technology Review 等的专文报道(<https://www.technology-review.com/2015/11/13/10130/machine-vision-algorithm-learns-to-recognize-hidden-facial-expressions/>)。除了基于梯度的直方图特征之外,颜色空间特征(Wang 等,2015a)、协方差矩阵(Hong 等,2016b)和基于光流特征的运动信息(Liu 等,2016b,2021a;Xu 等,2017)也受到了越来越多的关注。

与其他计算机视觉任务相比,机器学习乃至深度学习在微表情分析领域的应用起步要相对晚一些,但近年来得到较快发展。Gan 等人(2019)采用 CNN 网络对峰值帧对应的光流图提取深度特征。Xia 等人(2019a)利用帧间差分 and 光流作为递归神

经网络(recurrent convolutional neural network, RCNN)的输入并提取特征,取得了较为显著的性能提升,在 SMIC 和 CASMEII 数据集上取得了超越传统手工特征的性能。Xia 等人(2020a)针对 RCNN 在微表情识别领域的应用存在的若干关键因素进行了深入的探讨。Lei 等人(2020a)则利用图关系网络对图像区域之间的关系进行学习并进一步与 AU 的建模相结合(Lei 等,2021)。Xie 等人(2020)利用图注意力机制对 AU 之间的关系进行学习。Kumar 和 Bhanu(2021)也采用图注意力机制对特征点和光流区域的关系进行刻画。KFC(key facial components guided micro-expression recognition)算法利用关键脸部成分的分割概率图和成分可感知的注意力机制对一阶、二阶运动信息进行调和(Su 等,2021)。Li 等人(2021a)提出了可进行局部和全局联合学习的深度微表情识别方法。GEME(gender-based micro-expression)方法借鉴多任务学习思想,引入性别分类作为相关任务,并设计了一个双流浅层网络以同时学习微表情和性别分类(Nie 等,2021)。Xia 等人(2020b)利用宏表情和微表情数据集分别训练两个表情—身份解耦网络,并利用宏表情数据得到的网络对微表情的网络进行微调。进一步地,Xia 和 Wang(2021)提出 MiMaNet 将宏表情识别任务的空间和时域模式相关知识迁移到微表情识别任务中并提高了性能。关于微表情识别的详细进展,可参考近期的几个综述工作(Ben 等,2021;Zhou 等,2021;Zhang 和 Arandjelović,2021;Li 等,2021c)。

4 国内外研究进展比较

4.1 多模态情感识别

综合国内外关于多模态情感识别的研究现状,1)从多模态情感识别的数据来看,目前国内缺乏大规模的数据集。多模态视频数据集中,基于汉语的 CH-SIMS 包含 2 281 个视频片段(Yu 等,2020)。国际上常用的 CMU-MOSEI 数据集包含 23 453 个视频片段(Zadeh 等,2018c)。随着深度模型的发展,对数据量有了更高的要求。2)国内缺乏包含显性模态信息的数据集。直接表示身体变化的物理信号是进行情感识别的一个重要模态。国际上的 DEAP(database for emotion analysis using physiological signals)、MAHNOB-HCI(multimodal analysis of human

nonverbal behaviour in real-world settings-human machine interaction)等数据集包含了丰富的身体信号(Koelstra等,2012;Soleymani等,2012)。3)在计算方法的创新上国内已经具有优势,如中国科学院自动化研究所毛文吉研究员团队提出多层交互记忆网络计算多模态情感倾向(Xu等,2019)、中山大学胡海峰教授团队在时间上分布进行多种情感模态的融合(Mai等,2022)、哈尔滨工业大学秦兵教授团队从共享语义、独享语义两方面设计情感模态的融合方法(Wu等,2021a)、厦门大学纪荣嵘教授设计双层超图模型计算情感模态之间的相关性(Ji等,2019)、清华大学徐华教授团队提出自监督单模态情感预测辅助多模态情感识别的方法(Yu等,2021)、南开大学杨巨峰教授团队提出用多模态注意力机制学习新闻内容的情感表征(Guo等,2021)。4)在多模态情感识别的应用上,国内也做出了更多的尝试,如分析手机评价的情感(Xu等,2019)、微博数据的情感(Chen等,2018a)、用户旅途的情感(Wang等,2021a)以及文本和图像构成的新闻情感(Guo等,2021a)。

4.2 孤独症情感识别

国际上孤独症相关研究起步较早,因此孤独症患儿的诊断、干预以及生活都有专业的医疗团队提供帮助。然而,国内的孤独症专业诊疗队伍缺乏,相关医疗资源短缺,孤独症患儿的诊疗和教育状况不容乐观。此外,国外具有比较完善的评估量表、诊疗指南和干预手册,因此很方便进行家庭干预或社区干预。国内尚无专业和权威的孤独症诊疗手册,因此无法进行家庭和社区的全面推广。此外,国内缺乏具有资质的孤独症临床医师和专业诊疗机构。对于孤独症情感或情绪方面的研究,国外学者比较注重分析情感与社交能力、情感与内在生理信息以及情感与病理或干预机制之间的关系。国内学者大都关注于如何对孤独症患儿的情感识别和理解能力进行训练,常常忽略干预效果的评估,也缺乏对干预机制和机理的探究。因此,国内相关研究学者应该联合起来,建立符合中国国情的孤独症诊疗手册,并加大对相关领域的投入,培养专业医师队伍,鼓励相关产业的发展。在国内外,目前均缺乏有效的孤独症情感干预系统(设备),因此开发相关诊疗设备意义重大,不仅可以促进国内孤独症相关研究的发展,也可以为家庭和社会创造巨大的经济效益。

4.3 情感图像内容分析

综合上述国内和国际关于情感图像内容分析的研究进展,可以发现国内研究者在AICA方向的研究开展相对较晚,但近年来国内在该方向上持续发力,与国际上的差距逐步缩小,甚至在某些具体任务上已经走在国际前沿。1)现有的大规模数据集主要是由国外研究者构造,例如FI、T4SA和LUCFER等。国内在百万级甚至更大规模数据集的构建上需继续发力。2)国内在个性化情感预测和情感分布学习任务上已经走在了国际前列。例如清华大学朱文武教授团队和哈尔滨工业大学姚鸿勋教授团队关于个性化情感的研究极具代表性;南开大学杨巨峰教授团队构建的关于情感分布学习的数据集FlickrLDL和TwitterLDL已经成为规模较大的标准数据集。3)国内在从噪声数据或少量标签数据中进行图像情感识别的工作有些滞后。现实场景中的数据很可能是有噪声的,并且由于标注的昂贵性和情感的复杂主观性,有标注的数据量可能是少量的,因此国内在这种特定情况下的AICA研究还需加强。4)国内基于AICA的具体应用较少。图像情感在很多实际应用中起到重要作用,如广告推荐、自闭症治愈和舆情监测等。Tat-Seng Chua教授团队和Jiebo Luo教授团队在基于情感的广告推荐和基于情感的COVID-19(corona virus disease 2019)舆情监测任务上取得了显著进展,而国内在开发基于AICA的应用上有待改进。

4.4 微表情分析与识别

微表情检测与识别的研究是由芬兰Oulu大学Matti Pietikäinen和赵国英团队首倡。之后国内的中国科学院心理研究所、东南大学、复旦大学和山东大学等单位迅速跟进并始终与前者保持密切合作。而在该领域较为活跃的英国曼彻斯特城市大学和马来西亚研究组也均与上述机构保持合作。因此,在微表情分析领域国内外研究者一直紧密合作、相互促进,国内外研究进展基本上是同步进行的。这一点,通过国内外学者共同组织的MEGC(Facial Micro-Expressions Grand Challenge)和FME(Workshop on Facial Micro-Expression)系列挑战赛的发展可以更清楚地看到。该系列挑战赛是来自中国、英国、马来西亚、芬兰和中国台湾的学者联合组织的一项微表情识别与检测挑战赛,从2018年开始,已经举办了4届。在该系列挑战赛和相关的报告(Yap等,

2018; See 等, 2019; Li 等, 2020b, 2021d) 上可以了解微表情研究领域中比较前沿的测试方式与方法。

5 发展趋势与展望

5.1 多模态情感识别

1) 从研究方法的角度考虑。多模态情感识别的发展可以加入一些创新的思维。例如, 人感受的情感可以由对话场景和年龄、文化等先验信息影响, 因此将先验信息建模到模型的学习过程是一个可以尝试的途径。结合显性和隐性情感模态的方法也是一个有意义的思路。显性模态容易被抑制且难以获取, 隐性模态难以实时反映情感信息, 利用两类模态信息对情感识别的优势进行互补是一个有趣的研究方向。

2) 从应用的角度考虑。多模态情感识别的发展可以做一些更贴近现实场景的探索。例如, 由于手机、摄像头等边缘设备计算能力的不足, 将多模态情感识别的大模型进行部署时会出现计算资源的限制。因此考虑对多模态情感识别模型的量化是一个很有意义的发展趋势。

5.2 孤独症情感识别

最近的很多国内外工作开始关注孤独症情感功能与社交能力之间的关系。例如, 电子科技大学成都脑科学研究院临床医院 (Zhang 等, 2021b) 设计了一系列电脑训练游戏来帮助提高孤独症患儿的面部表情识别能力以及社交能力。作者发现通过相关程序的干预, 孤独症患儿的表情识别能力可以得到显著的提升。与控制组的患儿相比, 孤独症患儿的社交能力也能够得到相应的提升, 但是对于提升的程度以及所持续的时间作者并未分析。此外, 美国哈佛大学医学院 (Zagury-Orly 等, 2022) 也指出孤独症患者人脸处理能力的缺陷和社交能力紧密相关, 但是这种关系会受到年龄、性别和智商的影响。在孤独症研究领域, 大多数工作还是主要关注患儿的认知能力或社交行为, 如怎么识别表情、如何回应他人的指令和如何表达需求等。虽然这些工作研究如何教会孤独症患儿识别和理解他人的情感, 但是较少有工作关注孤独症患儿如何体验和表达自己的情感。可能正是由于孤独症患儿对自己的情感体验和表达不足或异常导致了他们的社交沟通障碍, 因此研究孤独症患儿的情感体验和表达机制非常重要。

未来的研究趋势应该是探究孤独症患儿的情感体验和表达机制, 并分析孤独症情感与社交能力之间的联系。同时, 目前缺乏有效的孤独症情感干预系统, 这也是未来国内研究者竞相争逐的一个热点。

5.3 情感图像内容分析

尽管情感图像内容分析已经取得了显著的进步, 但仍然有一些开放的问题和方向, 值得心理学、认知科学、多媒体和机器学习等各行各业的人一起努力考虑和研究。1) 图像内容和上下文理解。准确分析图像内容可以改进 AICA 的性能, 使用手工特征指导生成可解释的深度特征值得研究。2) 观看者上下文和先验知识建模。观看者在看图像时的上下文信息可以影响情感, 结合这些背景因素可以改进 AICA 的性能。3) 群体情感聚类。一些兴趣爱好相似、背景相似的用户群体, 可能对同一幅图像产生相似的情感反应, 群体情感识别在推荐中起到很关键的作用。分析用户自己提供的性别、背景、兴趣和爱好等个人资料, 将用户分为不同类型的群体, 然后使用概率图模型或图卷积网络来表示不同因素的复杂关系值得尝试。4) 观看者与图像交互。除直接分析情感内容外, 还可以记录并分析观看者在看图像时的视听和生理反应 (例如面部表情、脑电信号), 并结合图像内容和观看者的反应进行综合建模可以更好地弥合情感鸿沟。5) 高效的 AICA 学习。高效性问题在 AICA 领域尚处于开放阶段。基于计算机视觉中已有方法 (例如剪枝、神经网络搜索等), 结合 AICA 的特性 (例如情感层次), 或许有效。如果在边缘设备上训练的模型可以在线增量式地学习, 将会更有意义。

5.4 面部表情识别

面部表情数据集的识别精度在稳步提升, 除了提高数据集上的准确率外, 真实世界的情感感知还有很多准确率之外的实际挑战。

1) 数据采集和标注问题。例如, 在诸如自闭症孤独症之类的心理疾病筛查应用场景中, 数据采集条件与真实应用场景仍存在较大差距。在进行采集环境的设置时, 必须结合心理学知识, 才能获得有效的数据和相应标注 (Zhang 等, 2021c; Tang 等, 2020)。目前该方面的数据采集量仍无法满足大规模神经网络训练的要求。

2) 实时表情分析问题。目前的训练数据大都是切分好的片段序列或者单幅图像, 而情绪往往可

能隐藏在某个瞬间(Ben等,2021)。如何采用高效检测手段来进行实时化的情感识别也是目前待解决的一项问题。从面部行为推断该个人明显情绪的线索通常是稀疏且不规则地分布在感兴趣的时间窗内,收集这种分布的上下文线索对进行实时化情绪判断是至关重要的(Sanchez等,2021)。

3)混合表情识别问题。由于细粒度混合表情的复杂性和模糊性,其表情类别数也远远超过6类基本表情,因此传统方法无法很好地区分混合表情。目前领域仍缺乏精心设计的算法对混合表情进行精准识别。

4)个体情感表达差异问题。由于不同被试个体之间在生理(性别、年龄等)和心理(种族文化等)方面上的差异,在相同诱发条件下,不同个体表达同一情感的面部方式也会存在较大的鸿沟(Xu等,2020;Chen和Joo,2021)。可以考虑采用迁移学习或增量学习的方法来减小这些差异,从而构建出具有更佳泛化能力的表情分析模型(Li和Deng,2020;Kara等,2021)。

5)用户隐私问题。随着各项表情识别应用的普及,用户的视觉隐私也成为了不可避免的问题。人们对于个人信息隐私保护关注的日益增长,视觉摄像头在生活场景中的实施也会受到阻碍(Rahulmathavan和Rajarajan,2017)。因此,目前领域还需要可靠准确的面部表情识别隐私保护方法。这方面的研究仍不足(Nakashima等,2015;Chen等,2018b;Ullah等,2021)。

5.5 微表情分析与识别

智能微表情分析在刑侦测谎、商业谈判以及治疗有效性评估等领域具有十分广阔的应用的前景。在美剧《别对我撒谎》(Lie to me)中,主角Cal Lightman的故事即以Paul Ekman博士的诸多案例和事迹改编而成。尽管经过10年左右的发展,在微表情的检测和识别方面都取得了一些进展,但也应该看到,目前绝大部分微表情分析的工作都是针对在实验室或者受控环境下采集的微表情样本来进行的。如何设计出可以检测、识别与分析自然产生的微表情的有效方法,仍然是一个亟待解决的研究方向。其次,在微表情数据集构建方面,仅仅使用诱发的范式进行采集很难创造大规模的数据集,因此一种潜在的趋势是提出更为科学的方法,充分利用来自社交媒体如bilibili等的大量视频并对其存在的微表

情进行标注。再者,动作单元AU与微表情之间是否有AU与宏表情类似的对应关系?对这个问题的探索将有助于进一步理解微表情和人类情感表达之间的关系,并为自动微表情分析提供进一步的依据。此外,如何充分利用相对来讲更为容易获得的宏表情数据提升微表情识别系统的性能,也是一个很有意义的研究方向。最后,微表情分析在应用落地方面,仍然存在诸多障碍,面临不少难题,这些都有待学术界和工业界的同行共同去努力应对。

致谢 本文由中国图象图形学学会情感计算与理解专业委员会组织撰写,专委会更多详情请见链接:<http://www.csig.org.cn/detail/3186>。

参考文献(References)

- Ahsan U, de Choudhury M and Essa I. 2017. Towards using visual attributes to infer image sentiment of social events//Proceedings of 2017 International Joint Conference on Neural Networks. Anchorage, USA; IEEE: 1372-1379 [DOI: 10.1109/IJCNN.2017.7966013]
- Alameda-Pineda X, Ricci E, Yan Y and Sebe N. 2016. Recognizing emotions from abstract paintings using non-linear matrix completion//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE: 5240-5248 [DOI: 10.1109/CVPR.2016.566]
- Balouchian P, Safaei M and Foroosh H. 2019. LUCFER: a large-scale context-sensitive image dataset for deep learning of visual emotions//Proceedings of 2019 IEEE Winter Conference on Applications of Computer Vision. Waikoloa, USA; IEEE: 1645-1654 [DOI: 10.1109/WACV.2019.00180]
- Barsoum E, Zhang C, Ferrer C C and Zhang Z Y. 2016. Training deep networks for facial expression recognition with crowd-sourced label distribution//Proceedings of the 18th ACM International Conference on Multimodal Interaction. Tokyo, Japan; ACM: 279-283 [DOI: 10.1145/2993148.2993165]
- Ben X Y, Ren Y, Zhang J P, Wang S J, Kpalma K, Meng W X and Liu Y J. 2021. Video-based facial micro-expression analysis: a survey of datasets, features and algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence; #3067464 [DOI: 10.1109/TPAMI.2021.3067464]
- Benitez-Quiroz C F, Srinivasan R and Martinez A M. 2016. EmotionNet: an accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE: 5562-5570 [DOI: 10.1109/CVPR.2016.600]
- Borth D, Ji R R, Chen T, Breuel T and Chang S F. 2013. Large-scale

- visual sentiment ontology and detectors using adjective noun pairs//Proceedings of the 21st ACM International Conference on Multimedia. Barcelona, Spain; ACM: 223-232 [DOI: 10.1145/2502081.2502282]
- Burkhardt F, Paeschke A, Rolfes M, Sendmeier W F and Weiss B. 2005. A database of German emotional speech//Proceedings of the 9th European Conference on Speech Communication and Technology. Lisbon, Portugal; ISCA: 1-4 [DOI: 10.21437/Interspeech.2005-446]
- Busso C, Bulut M, Lee C C, Kazemzadeh A, Mower E, Kim S, Chang J N, Lee S and Narayanan S S. 2008. IEMOCAP: interactive emotional dyadic motion capture database. Language Resources and Evaluation, 42(4): 335-359 [DOI: 10.1007/s10579-008-9076-6]
- Centers for Disease Control and Prevention. 2016. Key Findings from the ADDM Network: A Snapshot of Autism Spectrum Disorder. Community Report on Autism. Centers for Disease Control and Prevention
- Chen F H, Ji R R, Su J S, Cao D L and Gao Y. 2018a. Predicting microblog sentiments via weakly supervised multimodal deep learning. IEEE Transactions on Multimedia, 20(4): 997-1007 [DOI: 10.1109/TMM.2017.2757769]
- Chen J W, Konrad J and Ishwar P. 2018b. VGAN-based image representation learning for privacy-preserving facial expression recognition//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City, USA; IEEE: 1683-1692 [DOI: 10.1109/CVPRW.2018.00207]
- Chen M, Zhang L and Allebach J P. 2015. Learning deep features for image emotion classification//Proceedings of 2015 IEEE International Conference on Image Processing. Quebec City, Canada; IEEE: 4491-4495 [DOI: 10.1109/ICIP.2015.7351656]
- Chen T, Borth D, Darrell T and Chang S F. 2014. DeepSentiBank: visual sentiment concept classification with deep convolutional neural networks [EB/OL]. [2022-02-06]. <https://arxiv.org/pdf/1410.8586.pdf>
- Chen Y L and Joo J. 2021. Understanding and mitigating annotation bias in facial expression recognition//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada; IEEE: 14960-14971 [DOI: 10.1109/ICCV48922.2021.01471]
- Conner C M, White S W, Scahill L and Mazefsky C A. 2020. The role of emotion regulation and core autism symptoms in the experience of anxiety in autism. Autism, 24(4): 931-940 [DOI: 10.1177/1362361320904217]
- Dan-Glauser E S and Scherer K R. 2011. The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. Behavior Research Methods, 43(2): 468-477 [DOI: 10.3758/s13428-011-0064-1]
- Darwin C. 2015. The Expression of the Emotions in Man and Animals. Chicago: University of Chicago Press [DOI: 10.7208/9780226220802]
- Davison A K, Lansley C, Costen N, Tan K and Yap M H. 2018. SAMM: a spontaneous micro-facial movement dataset. IEEE Transactions on Affective Computing, 9(1): 116-129 [DOI: 10.1109/TAFFC.2016.2573832]
- Dhall A. 2019. EmotiW 2019: automatic emotion, engagement and cohesion prediction tasks//Proceedings of 2019 International Conference on Multimodal Interaction. Suzhou, China; ACM: 546-550 [DOI: 10.1145/3340555.3355710]
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Houselby N. 2020. An image is worth 16 x 16 words: transformers for image recognition at scale//Proceedings of the 9th International Conference on Learning Representations. [s. l.]: OpenReview.net
- Du S C and Martinez A M. 2015. Compound facial expressions of emotion: from basic research to clinical applications. Dialogues in Clinical Neuroscience, 17(4): 443-455 [DOI: 10.31887/DCNS.2015.17.4/sdu]
- Du S C, Tao Y and Martinez A M. 2014. Compound facial expressions of emotion. Proceedings of the National Academy of Sciences of the United States of America, 111(15): E1454-E1462 [DOI: 10.1073/pnas.1322355111]
- Ekman P. 1965. Communication through nonverbal behavior: a source of information about an interpersonal relationship//Tomkins S S and Izard C E, eds. Affect, Cognition and Personality. New York: Springer: 390-442
- Ekman P. 2003. Darwin, deception, and facial expression. Annals of the New York Academy of Sciences, 1000(1): 205-221 [DOI: 10.1196/annals.1280.010]
- Ekman P and Friesen W V. 1969. Nonverbal leakage and clues to deception. Psychiatry, 32(1): 88-106 [DOI: 10.1080/00332747.1969.11023575]
- Ekman P and Friesen W V. 1971. Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 17(2): 124-129 [DOI: 10.1037/h0030377]
- Ekman P and Friesen W V. 1978. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto: Consulting Psychologists Press
- Ekman P, Friesen W V and Hager J C. 2002. Facial Action Coding System—the Manual. Salt Lake City: Research Nexus Division of Network Information Research Corporation: 15-464
- Ekman P, Friesen W V, O'Sullivan M, Chan A, Diacoyanni-Tarlatzis I, Heider K, Krause R, Lecompte W A, Pitcairn T, Ricci-Bitti P E, Scherer K, Tomita M and Tzavaras A. 1987. Universals and cultural differences in the judgments of facial expressions of emotion. Journal of Personality and Social Psychology, 53(4): 712-717 [DOI: 10.1037/0022-3514.53.4.712]
- Emery A E H, Muntoni F and Quinlivan R. 2015. Duchenne Muscular Dystrophy. 4th ed. Oxford: Oxford University Press
- Fan S J, Jiang M, Shen Z Q, Koenig B L, Kankanhalli M S and Zhao Q. 2017. The role of visual attention in sentiment prediction//Pro-

- ceedings of the 25th ACM International Conference on Multimedia. Mountain View, USA: ACM; 217-225 [DOI: 10.1145/3123266.3123445]
- Gan Y S, Liong S T, Yau W C, Huang Y C and Tan L K. 2019. Off-ApexNet on micro-expression recognition system. *Signal Processing: Image Communication*, 74: 129-139 [DOI: 10.1016/j.image.2019.02.005]
- Goodfellow I J, Erhan D, Carrier P L, Courville A, Mirza M, Hamner B, Cukierski W, Tang Y C, Thaler D, Lee D H, Zhou Y B, Ramiah C, Feng F X, Li R F, Wang X J, Athanasis D, Shawe-Taylor J, Milakov M, Park J, Ionescu R, Popescu M, Grozea C, Bergstra J, Xie J J, Romaszko L, Xu B, Chuang Z and Bengio Y. 2013. Challenges in representation learning: a report on three machine learning contests//*Proceedings of the 20th International Conference on Neural Information Processing*. Daegu, Korea(South): Springer; 117-124 [DOI: 10.1007/978-3-642-42051-1_16]
- Guo W Y, Zhang Y, Cai X R, Meng L, Yang J F and Yuan X J. 2021a. LD-MAN: layout-driven multimodal attention network for online news sentiment recognition. *IEEE Transactions on Multimedia*, 23: 1785-1798 [DOI: 10.1109/TMM.2020.3003648]
- Guo Y F, Li B, Ben X Y, Ren Y, Zhang J P, Yan R and Li Y J. 2021b. A magnitude and angle combined optical flow feature for microexpression spotting. *IEEE MultiMedia*, 28(2): 29-39 [DOI: 10.1109/MMUL.2021.3058017]
- Haggard E A and Isaacs K S. 1966. Micromomentary Facial Expressions as Indicators of Ego Mechanisms in Psychotherapy//Gottschalk L A and Auerbach A H, eds. *Methods of Research in Psychotherapy*. Boston: Springer; 154-165 [DOI: 10.1007/978-1-4684-6045-2_14]
- Han Y H, Li B J, Lai Y K and Liu Y J. 2018. CFD: a collaborative feature difference method for spontaneous micro-expression spotting//*Proceedings of the 25th IEEE International Conference on Image Processing*. Athens, Greece: IEEE; 1942-1946 [DOI: 10.1109/ICIP.2018.8451065]
- Happy S and Routray A. 2019. Fuzzy histogram of optical flow orientations for micro-expression recognition. *IEEE Transactions on Affective Computing*, 10(3): 394-406 [DOI: 10.1109/TAFFC.2017.2723386]
- He Y H. 2021. Research on micro-expression spotting method based on optical flow features//*Proceedings of the 29th ACM International Conference on Multimedia*. [s.l.]: ACM; 4803-4807 [DOI: 10.1145/3474085.3479225]
- Hong X P, Peng W, Harandi M, Zhou Z H, Pietikäinen M and Zhao G Y. 2019. Characterizing subtle facial movements via riemannian manifold. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 15(S3): #94 [DOI: 10.1145/3342227]
- Hong X P, Xu Y Y and Zhao G Y. 2016a. LBP-TOP: a tensor unfolding revisit//*Proceedings of 2016 Asian Conference on Computer Vision*. Taipei, China: Springer; 513-527 [DOI: 10.1007/978-3-319-54407-6_34]
- Hong X P, Zhao G Y, Zafeiriou S, Pantic M and Pietikäinen M. 2016b. Capturing correlations of local features for image representation. *Neurocomputing*, 184: 99-106 [DOI: 10.1016/j.neucom.2015.07.134]
- Huang X H, Wang S J, Liu X, Zhao G Y, Feng X Y and Pietikäinen M. 2019. Discriminative spatiotemporal local binary pattern with revisited integral projection for spontaneous facial micro-expression recognition. *IEEE Transactions on Affective Computing*, 10(1): 32-47 [DOI: 10.1109/TAFFC.2017.2713359]
- Huang X H, Wang S J, Zhao G Y and Piteikäinen M. 2015. Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection//*Proceedings of 2015 IEEE International Conference on Computer Vision Workshop*. Santiago, Chile: IEEE; 1-9 [DOI: 10.1109/ICCVW.2015.10]
- Huang X H, Zhao G Y, Hong X P, Zheng W M and Pietikäinen M. 2016. Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns. *Neurocomputing*, 175: 564-578 [DOI: 10.1016/j.neucom.2015.10.096]
- Husák P, Ůech J and Matas J. 2017. Spotting facial micro-expressions "In the Wild"//*Proceedings of the 22nd Computer Vision Winter Workshop*. Retz, Austria: [s.n.]
- Jacob G M and Stenger B. 2021. Facial action unit detection with transformers//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE; 7676-7685 [DOI: 10.1109/CVPR46437.2021.00759]
- Ji R R, Chen F H, Cao L J and Gao Y. 2019. Cross-modality microblog sentiment prediction via bi-layer multimodal hypergraph learning. *IEEE Transactions on Multimedia*, 21(4): 1062-1075 [DOI: 10.1109/TMM.2018.2867718]
- Jou B, Chen T, Pappas N, Redi M, Topkara M and Chang S F. 2015. Visual affect around the world: a large-scale multilingual visual sentiment ontology//*Proceedings of the 23rd ACM International Conference on Multimedia*. Brisbane, Australia: ACM; 159-168 [DOI: 10.1145/2733373.2806246]
- Kalantarian H, Jedoui K, Dunlap K, Schwartz J, Washington P, Husic A, Tariq Q, Ning M, Kline A and Wall D P. 2020. The performance of emotion classifiers for children with parent-reported autism: quantitative feasibility study. *JMIR Mental Health*, 7(4): #e13174 [DOI: 10.2196/13174]
- Kalantarian H, Jedoui K, Washington P, Tariq Q, Dunlap K, Schwartz J and Wall D P. 2019. Labeling images with facial emotion and the potential for pediatric healthcare. *Artificial Intelligence in Medicine*, 98: 77-86 [DOI: 10.1016/j.artmed.2019.06.004]
- Kara O, Churamani N and Gunes H. 2021. Towards fair affective robotics: continual learning for mitigating bias in facial expression and action unit recognition//*Proceedings of the Workshop on Lifelong Learning and Personalization in Long-Term Human-Robot Interaction (LEAP-HRI)*, 16th ACM/IEEE International Conference on Human-Robot Interaction (HRI) [EB/OL]. [2022-02-06]. <https://>

- arxiv.org/pdf/2103.09233.pdf
- Kim D H, Baddar W J and Ro Y M. 2016. Micro-expression recognition with expression-state constrained spatio-temporal feature representations//Proceedings of the 24th ACM International Conference on Multimedia. Amsterdam, the Netherlands; ACM; 382-386 [DOI: 10.1145/2964284.2967247]
- Koelstra S, Muhl C, Soleymani M, Lee J S, Yazdani A, Ebrahimi T, Pun T, Nijholt A and Patras I. 2012. DEAP: a database for emotion analysis; using physiological signals. IEEE Transactions on Affective Computing, 3 (1): 18-31 [DOI: 10.1109/T-AFFC.2011.15]
- Kollias D and Zafeiriou S. 2019. Expression, affect, action unit recognition: aff-wild2, multi-task learning and arcface//Proceedings of the 30th British Machine Vision Conference. Cardiff, UK; BMVA Press
- Kumar A J R and Bhanu B. 2021. Micro-expression classification based on landmark relations with graph attention convolutional network//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Nashville, USA; IEEE; 1511-1520 [DOI: 10.1109/CVPRW53098.2021.00167]
- Lang P J, Bradley M M and Cuthbert B N. 1997. International Affective Picture System (IAPS): Technical Manual and Affective Ratings. NIMH Center for the Study of Emotion and Attention; 39-58
- Lei L, Chen T, Li S G and Li J F. 2021. Micro-expression recognition based on facial graph representation learning and facial action unit fusion//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Nashville, USA; IEEE; 1571-1580 [DOI: 10.1109/CVPRW53098.2021.00173]
- Lei L, Li J F, Chen T and Li S G. 2020a. A novel graph-TCN with a graph structured representation for micro-expression recognition//Proceedings of the 28th ACM International Conference on Multimedia. Seattle, USA; ACM; 2237-2245 [DOI: 10.1145/3394171.3413714]
- Li J T, Soladie C and Segulier R. 2020a. Local temporal pattern and data augmentation for micro-expression spotting [JB/OL]. IEEE Transactions on Affective Computing, https://ieeexplore.ieee.org/document/9195783 [DOI: 10.1109/T-AFFC.2020.3023821]
- Li J T, Wang S J, Yap M H, See J, Hong X P and Li X B. 2020b. MEGC2020 - the third facial micro-expression grand challenge//Proceedings of the 15th IEEE International Conference on Automatic Face and Gesture Recognition. Buenos Aires, Argentina; IEEE; 777-780 [DOI: 10.1109/FG47880.2020.00035]
- Li J T, Yap M H, Cheng W H, See J, Hong X P, Li X B and Wang S J. 2021d. FME '21: 1st workshop on facial micro-expression; advanced techniques for facial expressions generation and spotting//Proceedings of the 29th ACM International Conference on Multimedia. [s.l.]; ACM; 5700-5701 [DOI: 10.1145/3474085.3478579]
- Li S and Deng W H. 2019a. Blended emotion in-the-wild: multi-label facial expression recognition using crowdsourced annotations and deep locality feature learning. International Journal of Computer Vision, 127(6): 884-906 [DOI: 10.1007/s11263-018-1131-1]
- Li S and Deng W H. 2019b. Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition. IEEE Transactions on Image Processing, 28 (1): 356-370 [DOI: 10.1109/TIP.2018.2868382]
- Li S and Deng W H. 2020. A deeper look at facial expression dataset bias. IEEE Transactions on Affective Computing [DOI: 10.1109/T-AFFC.2020.2973158]
- Li S, Deng W H and Du J P. 2017. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE; 2584-2593 [DOI: 10.1109/CVPR.2017.277]
- Li X B, Hong X P, Moilanen A, Huang X H, Pfister T, Zhao G Y and Pietikäinen M. 2015. Reading hidden emotions: spontaneous micro-expression spotting and recognition [EB/OL]. [2022-02-06]. https://arxiv.org/pdf/1511.00423.pdf
- Li X B, Hong X P, Moilanen A, Huang X H, Pfister T, Zhao G Y and Pietikäinen M. 2018. Towards reading hidden emotions: a comparative study of spontaneous micro-expression spotting and recognition methods. IEEE Transactions on Affective Computing, 9(4): 563-577 [DOI: 10.1109/T-AFFC.2017.2667642]
- Li X B, Pfister T, Huang X H, Zhao G Y and Pietikäinen M. 2013. A spontaneous micro-expression database: inducement, collection and baseline//Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Shanghai, China; IEEE; 1-6 [DOI: 10.1109/FG.2013.6553717]
- Li Y, Zeng J B and Shan S G. 2022. Learning representations for facial actions from unlabeled videos. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44 (1): 302-317 [DOI: 10.1109/TPAMI.2020.3011063]
- Li Y T, Huang X H and Zhao G Y. 2021a. Joint local and global information learning with single apex frame detection for micro-expression recognition. IEEE Transactions on Image Processing, 30: 249-263 [DOI: 10.1109/TIP.2020.3035042]
- Li Y T, Huang X H and Zhao G Y. 2021b. Micro-expression action unit detection with spatial and channel attention. Neurocomputing, 436: 221-231 [DOI: 10.1016/j.neucom.2021.01.032]
- Li Y T, Wei J S, Liu Y, Kauttonen J and Zhao G Y. 2021c. Deep learning for micro-expression recognition: a survey [EB/OL]. [2020-12-21]. https://arxiv.org/pdf/2107.02823.pdf
- Liao M Y, Chen J Y, Wang G S and Peng S X. 2021. Intelligent identification of children with autism spectrum disorder integrating multimodal data and its effectiveness. Chinese Science Bulletin, 66(20): 2618-2628 (廖梦怡, 陈靓影, 王广帅, 彭世新. 2021. 融合多模态数据的自闭症谱系障碍儿童智能化识别及其有效性. 科学通报, 66(20): 2618-2628) [DOI: 10.1360/TB-2020-1635]
- Lin C, Zhao S C, Meng L and Chua T S. 2020. Multi-source domain adaptation for visual sentiment classification//Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto, USA;

- AAAI: 2661-2668 [DOI: 10.1609/aaai.v34i03.5651]
- Liong S T, Gan Y S, Zheng D N, Li S M, Xu H X, Zhang H Z, Lyu R K and Liu K H. 2020. Evaluation of the spatio-temporal features and GAN for micro-expression recognition system. *Journal of Signal Processing Systems*, 92 (7): 705-725 [DOI: 10.1007/s11265-020-01523-4]
- Liong S T, See J, Wong K, Le Ngo A C, Oh Y H and Phan R. 2015. Automatic apex frame spotting in micro-expression database//Proceedings of the 3rd IAPR Asian Conference on Pattern Recognition. Kuala Lumpur, Malaysia: IEEE: 665-669 [DOI: 10.1109/ACPR.2015.7486586]
- Liu A A, Shi Y D, Jing P J, Liu J and Su Y T. 2018. Structured low-rank inverse-covariance estimation for visual sentiment distribution prediction. *Signal Processing*, 152: 206-216 [DOI: 10.1016/j.sigpro.2018.06.001]
- Liu H Y, Xu M, Wang J Q, Rao T R and Burnett I. 2016a. Improving visual saliency computing with emotion intensity. *IEEE Transactions on Neural Networks and Learning Systems*, 27 (6): 1201-1213 [DOI: 10.1109/tnnls.2016.2553579]
- Liu J J, Wang Z Y, Xu K, Ji B, Zhang G Y, Wang Y, Deng J X, Xu Q, Xu X and Liu H H. 2020. Early screening of autism in toddlers via response-to-instructions protocol. *IEEE Transactions on Cybernetics*, 1-11 [DOI: 10.1109/TCYB.2020.3017866]
- Liu Y J, Li B J and Lai Y K. 2021a. Sparse MDMO: learning a discriminative feature for micro-expression recognition. *IEEE Transactions on Affective Computing*, 12 (1): 254-261 [DOI: 10.1109/TAF-FC.2018.2854166]
- Liu Y J, Zhang J K, Yan W J, Wang S J, Zhao G Y and Fu X L. 2016b. A main directional mean optical flow feature for spontaneous micro-expression recognition. *IEEE Transactions on Affective Computing*, 7 (4): 299-310 [DOI: 10.1109/taffc.2015.2485205]
- Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, Lin S and Guo B N. 2021b. Swin transformer: hierarchical vision transformer using shifted windows//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 9992-10002 [DOI: 10.1109/ICCV48922.2021.00986]
- Lu X, Suryanarayan P, Adams R B, Li J, Newman M G and Wang J Z. 2012. On shape and the computability of emotions//Proceedings of the 20th ACM International Conference on Multimedia. Nara, Japan: ACM: 229-238 [DOI: 10.1145/2393347.2393384]
- Machajdik J and Hanbury A. 2010. Affective image classification using features inspired by psychology and art theory//Proceedings of the 18th ACM International Conference on Multimedia. Firenze, Italy: ACM: 83-92 [DOI: 10.1145/1873951.1873965]
- Mai S J, Hu H F, Xu J and Xing S L. 2022. Multi-fusion residual memory network for multimodal human sentiment comprehension. *IEEE Transactions on Affective Computing*, 13 (1): 320-334 [DOI: 10.1109/TAF-FC.2020.3000510]
- Martin J C, Niewiadomski R, Devillers L, Buisine S and Pelachaud C. 2006. Multimodal complex emotions: gesture expressivity and blended facial expressions. *International Journal of Humanoid Robotics*, 3 (3): 269-291 [DOI: 10.1142/S0219843606000825]
- Mehrabian A. 1996. Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14 (4): 261-292 [DOI: 10.1007/BF02686918]
- Mikels J A, Fredrickson B L, Larkin G R, Lindberg C M, Maglio S M and Reuter-Lorenz P A. 2005. Emotional category data on images from the international affective picture system. *Behavior Research Methods*, 37 (4): 626-630 [DOI: 10.3758/BF03192732]
- Minsky M L. 1986. *The Society of Mind*. New York: Simon and Schuster
- Moilanen A, Zhao G Y and Pietikäinen M. 2014. Spotting rapid facial movements from videos using appearance-based feature difference analysis//Proceedings of the 22nd International Conference on Pattern Recognition. Stockholm, Sweden: IEEE: 1722-1727 [DOI: 10.1109/ICPR.2014.303]
- Mollahosseini A, Hasani B and Mahoor M H. 2019. AffectNet: a database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10 (1): 18-31 [DOI: 10.1109/TAF-FC.2017.2740923]
- Morency L P, Mihalcea R and Doshi P. 2011. Towards multimodal sentiment analysis: harvesting opinions from the web//Proceedings of the 13th International Conference on Multimodal Interfaces. Alicante, Spain: ACM: 169-176 [DOI: 10.1145/2070481.2070509]
- Nag A, Haber N, Voss C, Tamura S, Daniels J, Ma J, Chiang B, Ramachandran S, Schwartz J, Winograd T, Feinstein C and Wall D P. 2020. Toward continuous social phenotyping: analyzing gaze patterns in an emotion recognition task for children with autism through wearable smart glasses. *Journal of Medical Internet Research*, 22 (4): #e13810 [DOI: 10.2196/13810]
- Nakashima Y, Koyama T, Yokoya N and Babaguchi N. 2015. Facial expression preserving privacy protection using image melding//Proceedings of 2015 IEEE International Conference on Multimedia and Expo (ICME). Turin, Italy: IEEE: 1-6 [DOI: 10.1109/ICME.2015.7177394]
- Narain J, Johnson K T, Ferguson C, O'Brien A, Talkar T, Weninger Y Z, Wofford P, Quatieri T, Picard R and Maes P. 2020a. Personalized modeling of real-world vocalizations from nonverbal individuals//Proceedings of 2020 International Conference on Multimodal Interaction. [s. l.]: ACM: 665-669 [DOI: 10.1145/3382507.3418854]
- Narain J, Johnson K T, O'Brien A, Wofford P, Maes P and Picard R. 2020b. Nonverbal vocalizations as speech: characterizing natural-environment audio from nonverbal individuals with autism//Proceedings of the Laughter and Other Non-Verbal Vocalisations Workshop 2020. Bielefeld, Germany: [s. n.]
- Ngo A C L, Johnston A, Phan R C W and See J. 2018. Micro-expression motion magnification: Global lagrangian vs. local eulerian approaches//Proceedings of the 13th IEEE International Conference

- on Automatic Face and Gesture Recognition (FG 2018). Xi'an, China; IEEE; #102 [DOI: 10.1109/FG.2018.00102]
- Nie X, Takalkar M A, Duan M Y, Zhang H M and Xu M. 2021. GEME: dual-stream multi-task GEnde-based micro-expression recognition. *Neurocomputing*, 427: 13-28 [DOI: 10.1016/j.neucom.2020.10.082]
- Niu X S, Yu Z T, Han H, Li X B, Shan S G and Zhao G Y. 2020. Video-based remote physiological measurement via cross-verified feature disentangling//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK; Springer: 295-310 [DOI: 10.1007/978-3-030-58536-5_18]
- Nummenmaa T. 1988. The recognition of pure and blended facial expressions of emotion from still photographs. *Scandinavian Journal of Psychology*, 29 (1): 33-47 [DOI: 10.1111/j.1467-9450.1988.tb00773.x]
- Oh T H, Jaroensri R, Kim C, Elgharib M, Durand F E, Freeman W T and Matusik W. 2018. Learning-based video motion magnification//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany; Springer: 663-679 [DOI: 10.1007/978-3-030-01225-0_39]
- Ola L and Gullon-Scott F. 2020. Facial emotion recognition in autistic adult females correlates with alexithymia, not autism. *Autism*, 24(8): 2021-2034 [DOI: 10.1177/1362361320932727]
- Palser E R, Galvez-Pol A, Palmer C E, Hannah R, Fotopoulou A, Pelliscano E and Kilner J M. 2021. Reduced differentiation of emotion-associated bodily sensations in autism. *Autism*, 25(5): 1321-1334 [DOI: 10.1177/1362361320987950]
- Pan Y R, Cai K J, Cheng M, Zou X B and Li M. 2021. Responsive social smile: a machine learning based multimodal behavior assessment framework towards early stage autism screening//Proceedings of the 25th International Conference on Pattern Recognition (ICPR). Milan, Italy; IEEE: 2240-2247 [DOI: 10.1109/ICPR48806.2021.9412766]
- Panda R, Zhang J M, Li H X, Lee J Y, Lu X and Roy-Chowdhury A K. 2018. Contemplating visual emotions: understanding and overcoming dataset bias//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany; Springer: 594-612 [DOI: 10.1007/978-3-030-01216-8_36]
- Patel D, Hong X P and Zhao G Y. 2016. Selective deep features for micro-expression recognition//Proceedings of the 23rd International Conference on Pattern Recognition. Cancun, Mexico; IEEE: 2258-2263 [DOI: 10.1109/ICPR.2016.7899972]
- Patel D, Zhao G Y and Pietikäinen M. 2015. Spatiotemporal integration of optical flow vectors for micro-expression detection//Proceedings of the 16th International Conference on Advanced Concepts for Intelligent Vision Systems. Catania, Italy; Springer: 369-380 [DOI: 10.1007/978-3-319-25903-1_32]
- Peng K C, Chen T, Sadovnik A and Gallagher A. 2015. A mixed bag of emotions: model, predict, and transfer emotion distributions//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA; IEEE: 860-868 [DOI: 10.1109/CVPR.2015.7298687]
- Peng W, Hong X P, Xu Y Y and Zhao G Y. 2019. A boost in revealing subtle facial expressions: a consolidated eulerian framework//Proceedings of the 14th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2019). Lille, France; IEEE: 1-5 [DOI: 10.1109/FG.2019.8756541]
- Pfister T, Li X B, Zhao G Y and Pietikäinen M. 2011. Recognising spontaneous facial micro-expressions//Proceedings of 2011 International Conference on Computer Vision. Barcelona, Spain; IEEE [DOI: 10.1109/ICCV.2011.6126401]
- Plutchik R. 2001. The nature of emotions; human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89(4): 344-350
- Poria S, Cambria E, Hazarika D, Majumder N, Zadeh A and Morency L P. 2017. Context-dependent sentiment analysis in user-generated videos//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada; Association for Computational Linguistics: 873-883 [DOI: 10.18653/v1/P17-1081]
- Qu F B, Wang S J, Yan W J, Li H, Wu S H and Fu X L. 2018. CAS (ME)2: a database for spontaneous macro-expression and micro-expression spotting and recognition. *IEEE Transactions on Affective Computing*, 9 (4): 424-436 [DOI: 10.1109/TAFFC.2017.2654440]
- Rahman W, Hasan K, Lee S, Zadeh A A B, Mao C F, Morency L P and Hoque E. 2020. Integrating multimodal information in large pre-trained transformers//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. [s.l.]: Association for Computational Linguistics: 2359-2369 [DOI: 10.18653/v1/2020.acl-main.214]
- Rahulamathavan Y and Rajarajan M. 2017. Efficient privacy-preserving facial expression classification. *IEEE Transactions on Dependable and Secure Computing*, 14(3): 326-338 [DOI: 10.1109/TDSC.2015.2453963]
- Rao T R, Li X X and Xu M. 2020. Learning multi-level deep representations for image emotion classification. *Neural Processing Letters*, 51(3): 2043-2061 [DOI: 10.1007/s11063-019-10033-9]
- Rao T R, Li X X, Zhang H M and Xu M. 2019. Multi-level region-based Convolutional Neural Network for image emotion classification. *Neurocomputing*, 333: 429-439 [DOI: 10.1016/j.neucom.2018.12.053]
- Rao T R, Xu M, Liu H Y, Wang J Q and Burnett I. 2016. Multi-scale blocks based image emotion classification using multiple instance learning//Proceedings of 2016 IEEE International Conference on Image Processing. Phoenix, USA; IEEE: 634-638 [DOI: 10.1109/ICIP.2016.7532434]
- Ruan D L, Yan Y, Lai S Q, Chai Z H, Shen C H and Wang H Z. 2021. Feature decomposition and reconstruction learning for effective facial expression recognition//Proceedings of 2021 IEEE/CVF

- Conference on Computer Vision and Pattern Recognition. Nashville, USA; IEEE: 7656-7665 [DOI: 10.1109/CVPR46437.2021.00757]
- Rudovic O, Lee J, Dai M, Schuller B and Picard R W. 2018. Personalized machine learning for robot perception of affect and engagement in autism therapy. *Science Robotics*, 3(19): #6760 [DOI: 10.1126/scirobotics.aao6760]
- Rui T, Cui P and Zhu W W. 2017. Joint user-interest and social-influence emotion prediction for individuals. *Neurocomputing*, 230: 66-76 [DOI: 10.1016/j.neucom.2016.11.054]
- Sanchez E, Tellamekala M K, Valstar M and Tzimiropoulos G. 2021. Affective processes; stochastic modelling of temporal context for emotion and facial expression recognition//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA; IEEE: 9070-9080 [DOI: 10.1109/CVPR46437.2021.00896]
- Sartori A, Culibrk D, Yan Y and Sebe N. 2015. Who's Afraid of Itten; using the art theory of color combination to analyze emotions in abstract paintings//Proceedings of the 23rd ACM International Conference on Multimedia. Brisbane, Australia; ACM: 311-320 [DOI: 10.1145/2733373.2806250]
- See J, Yap M H, Li J T, Hong X P and Wang S J. 2019. MEGC 2019-the second facial micro-expressions grand challenge//Proceedings of the 14th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2019). Lille, France; IEEE: 1-5 [DOI: 10.1109/FG.2019.8756611]
- She D Y, Sun M and Yang J F. 2019. Learning discriminative sentiment representation from strongly- and weakly supervised CNNs. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 15(S3): #96 [DOI: 10.1145/3326335]
- She J H, Hu Y B, Shi H L, Wang J, Shen Q and Mei T. 2021. Dive into ambiguity: latent distribution mining and pairwise uncertainty estimation for facial expression recognition//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA; IEEE: 6244-6253 [DOI: 10.1109/CVPR46437.2021.00618]
- Soleymani M, Lichtenauer J, Pun T and Pantic M. 2012. A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3(1): 42-55 [DOI: 10.1109/T-AF-FC.2011.25]
- Song T F, Cui Z J, Wang Y R, Zheng W M and Ji Q. 2021. Dynamic probabilistic graph convolution for facial action unit intensity estimation//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA; IEEE: 4843-4852 [DOI: 10.1109/CVPR46437.2021.00481]
- Su Y T, Zhang J Q, Liu J and Zhai G T. 2021. Key facial components guided micro-expression recognition based on first and second-order motion//Proceedings of 2021 IEEE International Conference on Multimedia and Expo. Shenzhen, China; IEEE: 1-6 [DOI: 10.1109/ICME51207.2021.9428407]
- Sullivan O A and Wang C Y. 2020. Autism spectrum disorder interventions in mainland China; a systematic review. *Review Journal of Autism and Developmental Disorders*, 7(3): 263-277 [DOI: 10.1007/s40489-019-00191-w]
- Tang C G, Zheng W M, Zong Y, Qiu N, Lu C, Zhang X L, Ke X Y and Guan C T. 2020. Automatic identification of high-risk autism spectrum disorder; a feasibility study using video and audio data under the still-face paradigm. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(11): 2401-2410 [DOI: 10.1109/TNSRE.2020.3027756]
- Theeuwes J and Van der Stigchel S. 2006. Faces capture attention; evidence from inhibition of return. *Visual Cognition*, 13(6): 657-665 [DOI: 10.1080/135062805000410949]
- Tian Y L, Kanade T and Cohn J F. 2011. Facial expression recognition//Li S Z and Jain A K, eds. *Handbook of Face Recognition*. London, UK: Springer: 487-519 [DOI: 10.1007/978-0-85729-932-1_19]
- Tomasello M, Carpenter M, Call J, Behne T and Moll H. 2005. Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences*, 28(5): 675-691 [DOI: 10.1017/S0140525X05000129]
- Tran T K, Vo Q N, Hong X P, Li X B and Zhao G Y. 2021. Micro-expression spotting; a new benchmark. *Neurocomputing*, 443: 356-368 [DOI: 10.1016/j.neucom.2021.02.022]
- Tsai W T, Lee I J and Chen C H. 2021. Inclusion of third-person perspective in CAVE-like immersive 3D virtual reality role-playing games for social reciprocity training of children with an autism spectrum disorder. *Universal Access in the Information Society*, 20(2): 375-389 [DOI: 10.1007/s10209-020-00724-9]
- Ullah A, Wang J, Anwar M S, Ahmad A, Nazir S, Khan H U and Fei Z S. 2021. Fusion of machine learning and privacy preserving for secure facial expression recognition. *Security and Communication Networks*, 2021: #6673992 [DOI: 10.1155/2021/6673992]
- Vadicamo L, Carrara F, Cimino A, Cresci S, Dell'Orletta F, Falchi F and Tesconi M. 2017. Cross-media learning for image sentiment analysis in the wild//Proceedings of 2017 IEEE International Conference on Computer Vision Workshops. Venice, Italy; IEEE: 308-317 [DOI: 10.1109/ICCVW.2017.45]
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin H. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 6000-6010
- Wang J and Geng X. 2021. Label distribution learning by exploiting label distribution manifold. *IEEE Transactions on Neural Networks and Learning Systems*: 1-14 [DOI: 10.1109/TNNLS.2021.3103178]
- Wang K, Peng X J, Yang J F, Lu S J and Qiao Y. 2020a. Suppressing uncertainties for large-scale facial expression recognition//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA; IEEE: 6896-6905 [DOI: 10.1109/

- CVPR42600. 2020. 00693]
- Wang K, Peng X J, Yang J F, Meng D B and Qiao Y. 2020b. Region attention networks for pose and occlusion robust facial expression recognition. *IEEE Transactions on Image Processing*, 29: 4057-4069 [DOI: 10.1109/TIP.2019.2956143]
- Wang L J, Guo W Y, Yao X X, Zhang Y X and Yang J F. 2021a. Multimodal event-aware network for sentiment analysis in tourism. *IEEE MultiMedia*, 28 (2): 49-58 [DOI: 10.1109/MMUL.2021.3079195]
- Wang Q D, Lu L, Zhang Q, Fang F, Zou X B and Yi L. 2018. Eye avoidance in young children with autism spectrum disorder is modulated by emotional facial expressions. *Journal of Abnormal Psychology*, 127(7): 722-732 [DOI: 10.1037/abn0000372]
- Wang S J, Wu S H, Qian X S, Li J X and Fu X L. 2017. A main directional maximal difference analysis for spotting facial movements from long-term videos. *Neurocomputing*, 230: 382-389 [DOI: 10.1016/j.neucom.2016.12.034]
- Wang S J, Yan W J, Li X B, Zhao G Y, Zhou C G, Fu X L, Yang M H and Tao J H. 2015a. Micro-expression recognition using color spaces. *IEEE Transactions on Image Processing*, 24(12): 6034-6047 [DOI: 10.1109/TIP.2015.2496314]
- Wang W N, Yu Y L and Jiang S M. 2006. Image retrieval by emotional semantics: a study of emotional space and feature extraction//*Proceedings of 2006 IEEE International Conference on Systems, Man and Cybernetics*. Taipei, China; IEEE: 3534-3539 [DOI: 10.1109/ICSMC.2006.384667]
- Wang W N, Yu Y L and Zhang J C. 2004. Image emotional classification: static vs. dynamic//*Proceedings of 2004 IEEE International Conference on Systems, Man and Cybernetics*. Hague, the Netherlands; IEEE: 6407-6411 [DOI: 10.1109/ICSMC.2004.1401407]
- Wang X H, Jia J, Yin J M and Cai L H. 2013. Interpretable aesthetic features for affective image classification//*Proceedings of 2013 IEEE International Conference on Image Processing*. Melbourne, Australia; IEEE: 3230-3234 [DOI: 10.1109/ICIP.2013.6738665]
- Wang Y L, Wang S H, Tang J L, Liu H and Li B X. 2015b. Unsupervised sentiment analysis for social media images//*Proceedings of the 24th International Conference on Artificial Intelligence*. Buenos Aires, Argentina; AAAI Press: 2378-2379
- Wang Z Y, Liu J J, He K S, Xu Q, Xu X and Liu H H. 2021b. Screening early children with autism spectrum disorder via response-to-name protocol. *IEEE Transactions on Industrial Informatics*, 17(1): 587-595 [DOI: 10.1109/TII.2019.2958106]
- Washington P, Paskov K M, Kalantarian H, Stockham N, Voss C, Kline A, Patnaik R, Chrisman B, Varma M, Tariq Q, Dunlap K, Schwartz J, Haber N and Wall D P. 2019. Feature selection and dimension reduction of social autism data//*Altman R B, Dunker A K, Hunter L, Ritchie M D, Murray T and Klein T E, eds. Pacific Symposium on Biocomputing 2020*. Kohala Coast, USA; [s. n.]: 707-718 [DOI: 10.1142/9789811215636_0062]
- Wöllmer M, Weninger F, Knaup T, Schuller B, Sun C K, Sagae K and Morency L P. 2013. Youtube movie reviews: sentiment analysis in an audio-visual context. *IEEE Intelligent Systems*, 28(3): 46-53 [DOI: 10.1109/MIS.2013.34]
- World Health Organization. 2007. *International Classification of Functioning, Disability and Health: Children and Youth Version*; ICF-CY. World Health Organization
- Wu H Y, Rubinstein M, Shih E, Gutttag J, Durand F and Freeman W. 2012. Eulerian video magnification for revealing subtle changes in the world. *ACM Transaction on Graphics*, 31(4): #65
- Wu Y, Lin Z J, Zhao Y Y, Qin B and Zhu L N. 2021. A text-centered shared-private framework via cross-modal prediction for multimodal sentiment analysis//*Proceedings of the Findings of the Association for Computational Linguistics*. [s. l.]: Association for Computational Linguistics [DOI: 10.18653/v1/2021.findings-acl.417]
- Xia B and Wang S F. 2021. Micro-expression recognition enhanced by macro-expression from spatial-temporal domain//*Proceedings of the 13th International Joint Conference on Artificial Intelligence*. Montreal, Canada; ijcai.org: 1186-1193 [DOI: 10.24963/ijcai.2021/164]
- Xia B, Wang W K, Wang S F and Chen E H. 2020b. Learning from macro-expression: a micro-expression recognition framework//*Proceedings of the 28th ACM International Conference on Multimedia*. Seattle, USA; ACM: 2936-2944 [DOI: 10.1145/3394171.3413774]
- Xia Z Q, Hong X P, Gao X Y, Feng X Y, Zhao G Y. 2019a. Spatio-temporal recurrent convolutional networks for recognizing spontaneous micro-expressions. *IEEE Transactions on Multimedia*, 22(3): 626-640 [DOI: 10.1109/TMM.2019.2931351]
- Xia Z Q, Liang H, Hong X P and Feng X Y. 2019b. cross-database micro-expression recognition with deep convolutional networks//*Proceedings of the 3rd International Conference on Biometric Engineering and Applications*. Stockholm, Sweden; ACM: 56-60 [DOI: 10.1145/3345336.3345343]
- Xia Z Q, Peng W, Khor H Q, Feng X Y and Zhao G Y. 2020a. Revealing the invisible with model and data shrinking for composite-database micro-expression recognition. *IEEE Transactions on Image Processing*, 29: 8590-8605 [DOI: 10.1109/TIP.2020.3018222]
- Xie H X, Lo L, Shuai H H and Cheng W H. 2020. Au-assisted graph attention convolutional network for micro-expression recognition//*Proceedings of the 28th ACM International Conference on Multimedia*. Seattle, USA; ACM: 2871-2880 [DOI: 10.1145/3394171.3414012]
- Xu C, Cetintas S, Lee K C and Li L J. 2014. Visual sentiment prediction with deep convolutional neural networks [EB/OL]. [2022-02-06]. <https://arxiv.org/pdf/1411.5731.pdf>
- Xu F, Zhang J P and Wang J Z. 2017. Microexpression identification and categorization using a facial dynamics map. *IEEE Transactions on Affective Computing*, 8(2): 254-267 [DOI: 10.1109/TAFFC.2016.2518162]
- Xu N, Mao W J and Chen G D. 2019. Multi-interactive memory network

- for aspect based multimodal sentiment analysis//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Palo Alto, USA; AAAI: 371-378 [DOI: 10.1609/aaai.v33i01.3301371]
- Xu T, White J, Kalkan S and Gunes H. 2020. Investigating bias and fairness in facial expression recognition//Proceedings of 2020 European Conference on Computer Vision. Glasgow, UK: Springer: 506-523 [DOI: 10.1007/978-3-030-65414-6_35]
- Xue F L, Wang Q C and Guo G D. 2021. TransFER: learning relation-aware facial expression representations with transformers//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada; IEEE: 3581-3590 [DOI: 10.1109/ICCV48922.2021.00358]
- Yan W J, Li S, Que C T, Pei J Q and Deng W H. 2020. RAF-AU database: in-the-wild facial expressions with subjective emotion judgement and objective AU annotations//Proceedings of the 15th Asian Conference on Computer Vision. Kyoto, Japan: Springer [DOI: 10.1007/978-3-030-69544-6_5]
- Yan W J, Li X B, Wang S J, Zhao G Y, Liu Y J, Chen Y H and Fu X L. 2014a. CASME II: an improved spontaneous micro-expression database and the baseline evaluation. PLoS One, 9(1): #e86041 [DOI: 10.1371/journal.pone.0086041]
- Yan W J, Wang S J, Chen Y H, Zhao G Y and Fu X L. 2014b. Quantifying micro-expressions with constraint local model and local binary pattern//Agapito L, Bronstein M M and Rother C, eds. Computer Vision - ECCV 2014 Workshops. Cham: Springer: 296-305 [DOI: 10.1007/978-3-319-16178-5_20]
- Yan W J, Wu Q, Liu Y J, Wang S J and Fu X L. 2013. CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces//Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Shanghai, China; IEEE: 1-7 [DOI: 10.1109/FG.2013.6553799]
- Yang J F, She D Y and Sun M. 2017a. Joint image emotion classification and distribution learning via deep convolutional neural network//Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia; AAAI: 3266-3272
- Yang J F, She D Y, Lai Y K, Rosin P L and Yang M H. 2018a. Weakly supervised coupled networks for visual sentiment analysis//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 7584-7592 [DOI: 10.1109/CVPR.2018.00791]
- Yang J F, She D Y, Lai Y K and Yang M H. 2018b. Retrieving and classifying affective images via deep metric learning//Proceedings of the 32nd AAAI Conference on Artificial Intelligence and the 30th Innovative Applications of Artificial Intelligence Conference and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence. New Orleans, USA; AAAI: #61
- Yang J F, She D Y, Sun M, Cheng M M, Rosin P L and Wang L. 2018c. Visual sentiment prediction based on automatic discovery of affective regions. IEEE Transactions on Multimedia, 20(9): 2513-2525 [DOI: 10.1109/TMM.2018.2803520]
- Yang J F, Sun M and Sun X X. 2017b. Learning visual sentiment distributions via augmented conditional probability neural network//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA; AAAI: 224-230
- Yang J Y, Gao X B, Li L D, Wang X M and Ding J S. 2021a. SOLVER: scene-object interrelated visual emotion reasoning network. IEEE Transactions on Image Processing, 30: 8686-8701 [DOI: 10.1109/TIP.2021.3118983]
- Yang J Y, Li J, Wang X M, Ding Y X and Gao X B. 2021b. Stimuli-aware visual emotion analysis. IEEE Transactions on Image Processing, 30: 7432-7445 [DOI: 10.1109/TIP.2021.3106813]
- Yang Y, Cui P, Zhu W W and Yang S Q. 2013. User interest and social influence based emotion prediction for individuals//Proceedings of the 21st ACM International Conference on Multimedia. Barcelona, Spain; ACM: 785-788 [DOI: 10.1145/2502081.2502204]
- Yanulevskaya V, van Gemert J C, Roth K, Herbold A K, Sebe N and Geusebroek J M. 2008. Emotional valence categorization using holistic image features//Proceedings of the 15th IEEE International Conference on Image Processing. San Diego, USA; IEEE: 101-104 [DOI: 10.1109/ICIP.2008.4711701]
- Yao X X, She D Y, Zhao S C, Liang J, Lai Y K and Yang J F. 2019. Attention-aware polarity sensitive embedding for affective image retrieval//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 1140-1150 [DOI: 10.1109/ICCV.2019.00123]
- Yap M H, See J, Hong X P and Wang S J. 2018. Facial micro-expressions grand challenge 2018 summary//Proceedings of the 13th IEEE International Conference on Automatic Face and Gesture Recognition. Xi'an, China; IEEE: 675-678 [DOI: 10.1109/FG.2018.00106]
- You Q Z, Jin H L and Luo J B. 2017. Visual sentiment analysis by attending on local image regions//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA; AAAI: 231-237
- You Q Z, Luo J B, Jin H L and Yang J C. 2015. Robust image sentiment analysis using progressively trained and domain transferred deep networks//Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin, USA; AAAI: 381-388
- You Q Z, Luo J B, Jin H L and Yang J C. 2016. Building a large scale dataset for image emotion recognition: the fine print and the benchmark//Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix, USA; AAAI: 308-314
- Yu W M, Xu H, Meng F Y, Zhu Y L, Ma Y X, Wu J L, Zou J Y and Yang K C. 2020. CH-SIMS: a Chinese multimodal sentiment analysis dataset with fine-grained annotation of modality//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. [s.l.]: Association for Computational Linguistics: 3718-3727 [DOI: 10.18653/v1/2020.acl-main.343]
- Yu W M, Xu H, Yuan Z Q and Wu J L. 2021. Learning modality-specific representations with self-supervised multi-task learning for mul-

- timodal sentiment analysis//Proceedings of the 35th AAAI Conference on Artificial Intelligence. [s.l.]: AAAI
- Yuan J B, McDonough S, You Q Z and Luo J B. 2013. Sentribute: image sentiment analysis from a mid-level perspective//Proceedings of the 2nd International Workshop on Issues of Sentiment Discovery and Opinion Mining. Chicago, USA: ACM; 10 [DOI: 10.1145/2502069.2502079]
- Zadeh A, Liang P P, Mazumder N, Poria S, Cambria E and Morency L P. 2018a. Memory fusion network for multi-view sequential learning//Proceedings of the 32nd AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th Innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18). New Orleans, USA: AAAI; 5634-5641
- Zadeh A, Liang P P, Poria S, Vij P, Cambria E and Morency L P. 2018b. Multi-attention recurrent network for human communication comprehension//Proceedings of the 32nd AAAI Conference on Artificial Intelligence and the 30th Innovative Applications of Artificial Intelligence Conference and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence. New Orleans, USA: AAAI; # 692
- Zadeh A A B, Liang P P, Poria S, Cambria K and Morency L P. 2018c. Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia: Association for Computational Linguistics; 2236-2246 [DOI: 10.18653/v1/P18-1208]
- Zagury-Orly I, Kroeck M R, Soussand L and Cohen A L. 2022. Face-processing performance is an independent predictor of social affect as measured by the autism diagnostic observation schedule across large-scale datasets. *Journal of Autism and Developmental Disorders*, 52(2): 674-688 [DOI: 10.1007/s10803-021-04971-4]
- Zhan C, She D Y, Zhao S C, Cheng M M and Yang J F. 2019. Zero-shot emotion recognition via affective structural embedding//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE; 1151-1160 [DOI: 10.1109/ICCV.2019.00124]
- Zhang L F and Arandjelović O. 2021. Review of automatic microexpression recognition in the past decade. *Machine Learning and Knowledge Extraction*, 3(2): 414-434 [DOI: 10.3390/make3020021]
- Zhang L F, Arandjelovic O and Hong X P. 2021d. Facial action unit detection with local key facial sub-region based multi-label classification for micro-expression analysis//Proceedings of the 1st Workshop on Facial Micro-Expression: Advanced Techniques for Facial Expressions Generation and Spotting. [s.l.]: ACM; 11-18 [DOI: 10.1145/3476100.3484462]
- Zhang Q Q, Wu R J, Zhu S Y, Le J, Chen Y S, Lan C M, Yao S X, Zhao W H and Kendrick K M. 2021b. Facial emotion training as an intervention in autism spectrum disorder: a meta-analysis of randomized controlled trials. *Autism Research*, 14(10): 2169-2182 [DOI: 10.1002/aur.2565]
- Zhang R J, Chen J Y, Wang G S, Xu R Y, Zhang K, Wang J D and Zheng W M. 2021c. Towards a computer-assisted comprehensive evaluation of visual motor integration for children with autism spectrum disorder: a pilot study. *Interactive Learning Environments*; 1-16 [DOI: 10.1080/10494820.2021.1952273]
- Zhang T, Zong Y, Zheng W M, Chen C L P, Hong X P, Tang C G, Cui Z and Zhao G Y. 2020. Cross-database micro-expression recognition: a benchmark. *IEEE Transactions on Knowledge and Data Engineering*, 34(2): 544-559 [DOI: 10.1109/TKDE.2020.2985365]
- Zhang W, Ji X P, Chen K Y, Ding Y and Fan C J. 2021d. Learning a facial expression embedding disentangled from identity//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE; 6755-6764 [DOI: 10.1109/CVPR46437.2021.00669]
- Zhang Y H, Wang C R and Deng W H. 2021e. Relative uncertainty learning for facial expression recognition [EB/OL]. [2020-12-20]. <https://openreview.net/forum?id=h1-ilmYbdea>
- Zhang Z P, Luo P, Loy C C and Tang X O. 2018. From facial expression recognition to interpersonal relation prediction. *International Journal of Computer Vision*, 126(5): 550-569 [DOI: 10.1007/s11263-017-1055-1]
- Zhao G Y and Li X B. 2019. Automatic micro-expression analysis: open challenges. *Frontiers in Psychology*, 10: #1833 [DOI: 10.3389/fpsyg.2019.01833]
- Zhao S C, Ding G G, Gao Y and Han J G. 2017a. Approximating discrete probability distribution of image emotions by multi-modal features fusion//Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia: IJCAI.org; 4669-4675 [DOI: 10.24963/ijcai.2017/651]
- Zhao S C, Ding G G, Gao Y and Han J G. 2017b. Learning visual emotion distributions via multi-modal features fusion//Proceedings of the 25th ACM International Conference on Multimedia. Mountain View, USA: ACM; 369-377 [DOI: 10.1145/3123266.3130858]
- Zhao S C, Ding G G, Gao Y, Zhao X, Tang Y B, Han J G, Yao H X and Huang Q M. 2020a. Discrete probability distribution prediction of image emotions with shared sparse learning. *IEEE Transactions on Affective Computing*, 11(4): 574-587 [DOI: 10.1109/TAFFC.2018.2818685]
- Zhao S C, Gao Y, Jiang X L, Yao H X, Chua T S and Sun X S. 2014. Exploring principles-of-art features for image emotion recognition//Proceedings of the 22nd ACM International Conference on Multimedia. Orlando, USA: ACM; 47-56 [DOI: 10.1145/2647868.2654930]
- Zhao S C, Jia Z Z, Chen H, Li L D, Ding G G and Keutzer K. 2019a. PDANet: polarity-consistent deep attention network for fine-grained visual emotion regression//Proceedings of the 27th ACM International Conference on Multimedia. Nice, France: ACM; 192-201 [DOI: 10.1145/3343031.3351062]

- Zhao S C, Lin C, Xu P F, Zhao S D, Guo Y C, Krishna R, Ding G G and Keutzer K. 2019b. CycleEmotionGAN: emotional semantic consistency preserved CycleGAN for adapting image emotions//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI: 2620-2627 [DOI: 10.1609/aaai.v33i01.33012620]
- Zhao S C, Ma Y S, Gu Y, Yang J F, Xing T F, Xu P F, Hu R B, Chai H and Keutzer K. 2020b. An end-to-end visual-audio attention network for emotion recognition in user-generated videos//Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI: 303-311 [DOI: 10.1609/aaai.v34i01.5364]
- Zhao S C, Yao H X, Gao Y, Ding G G and Chua T S. 2018a. Predicting personalized image emotion perceptions in social networks. IEEE Transactions on Affective Computing, 9(4): 526-540 [DOI: 10.1109/TAFFC.2016.2628787]
- Zhao S C, Yao H X, Gao Y, Ji R R and Ding G G. 2017c. Continuous probability distribution prediction of image emotions via multitask shared sparse regression. IEEE Transactions on Multimedia, 19(3): 632-645 [DOI: 10.1109/TMM.2016.2617741]
- Zhao S C, Yao H X, Gao Y, Ji R R, Xie W L, Jiang X L and Chua T S. 2016. Predicting personalized emotion perceptions of social images//Proceedings of the 24th ACM International Conference on Multimedia. Amsterdam, the Netherlands: ACM: 1385-1394 [DOI: 10.1145/2964284.2964289]
- Zhao S C, Yao H X, Jiang X L and Sun X S. 2015. Predicting discrete probability distribution of image emotions//Proceedings of 2015 IEEE International Conference on Image Processing. Quebec City, Canada: IEEE: 2459-2463 [DOI: 10.1109/ICIP.2015.7351244]
- Zhao S C, Yao X X, Yang J F, Jia G J, Ding G G, Chua T S, Schuller B W and Keutzer K. 2021. Affective image content analysis: two decades review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence [DOI: 10.1109/TPAMI.2021.3094362]
- Zhao S C, Zhao X, Ding G G and Keutzer K. 2018b. EmotionGAN: unsupervised domain adaptation for learning discrete probability distributions of image emotions//Proceedings of the 26th ACM International Conference on Multimedia. Seoul, Korea (South): ACM: 1319-1327 [DOI: 10.1145/3240508.3240591]
- Zhou L, Shao X Y and Mao Q R. 2021. A survey of micro-expression recognition. Image and Vision Computing, 105: #104043 [DOI: 10.1016/j.imavis.2020.104043]
- Zhou Z H, Hong X P, Zhao G Y and Pietikäinen M. 2014. A compact representation of visual speech data using latent variables. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(1): 181-187 [DOI: 10.1109/TPAMI.2013.173]
- Zhou Z, Zhao G and Pietikainen M. 2011. Towards a practical lipreading system//Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEEE: 137-144
- Zhu X E, Li L, Zhang W G, Rao T R, Xu M, Huang Q M and Xu D. 2017. Dependency exploitation: a unified CNN-RNN approach for visual emotion recognition//Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia: IJCAI.org: 3595-3601 [DOI: 10.24963/ijcai.2017/503]
- Zong Y, Huang X H, Zheng W M, Cui Z and Zhao G Y. 2018b. Learning from hierarchical spatiotemporal descriptors for micro-expression recognition. IEEE Transactions on Multimedia, 20(11): 3160-3172 [DOI: 10.1109/TMM.2018.2820321]
- Zong Y, Zheng W M, Hong X P, Tang C G, Cui Z and Zhao G Y. 2019. Cross-database micro-expression recognition: a benchmark//Proceedings of 2019 on International Conference on Multimedia Retrieval. Ottawa, Canada: ACM: 354-363 [DOI: 10.1145/3323873.3326590]
- Zong Y, Zheng W M, Huang X H, Shi J G, Cui Z and Zhao G Y. 2018a. Domain regeneration for cross-database micro-expression recognition. IEEE Transactions on Image Processing, 27(5): 2484-2498 [DOI: 10.1109/TIP.2018.2797479]
- Zou X B. 2019. Intervention principles for children with autism and BSR model. Chinese Journal of Child Health Care, 27(1): 1-6 (邹小兵. 2019. 孤独症谱系障碍干预原则与 BSR 模式. 中国儿童保健杂志, 27(1): 1-6) [DOI: 10.11852/zgetbjzz2018-1611]

作者简介



姚鸿勋, 1965年生, 女, 教授, 主要研究方向为计算机视觉智能。

E-mail: h.yao@hit.edu.cn

邓伟洪, 男, 教授, 主要研究方向为 лица图像识别。

E-mail: whdeng@bupt.edu.cn

刘洪海, 男, 教授, 主要研究方向为多模态人机交互感知与理解。E-mail: honghai.liu@hit.edu.cn

洪晓鹏, 男, 教授, 主要研究方向为情感计算。

E-mail: hongxiaopeng@ieee.org

王甦菁, 男, 副研究员, 主要研究方向为微表情的神经机制。

E-mail: wangsujing@psych.ac.cn

杨巨峰, 男, 教授, 主要研究方向为计算机视觉。

E-mail: yangjufeng@nankai.edu.cn

赵思成, 男, 博士后研究员, 主要研究方向为情感计算。

E-mail: schzhao@gmail.com