智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

心理学视角下的自动表情识别

颜文靖, 蒋柯, 傅小兰

引用本文:
颜文靖,蒋柯,傅小兰. 心理学视角下的自动表情识别[J]. 智能系统学报, 2022, 17(5): 1039–1053.
YAN Wenjing,JIANG Ke,FU Xiaolan. Automatic facial expression recognition from a psychological perspective[J]. *CAAI Transactions on Intelligent Systems*, 2022, 17(5): 1039–1053.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202112056

DOI: 10.11992/tis.202112056
网络出版地址: https://kns.cnki.net/kcms/detail/23.1538.TP.20220617.1819.008.html

心理学视角下的自动表情识别

颜文靖[1], 蒋柯[1], 傅小兰[2,3]
（1. 温州医科大学 精神医学学院 浙江省阿尔茨海默病研究重点实验室, 浙江 温州 325015; 2. 中国科学院心理研究所 脑与认知科学国家重点实验室, 北京 100101; 3. 中国科学院大学 心理学系, 北京 100049）

摘　要: 自动表情识别是心理学与计算机科学等深度交叉的前沿领域。情绪心理学、模式识别、情感计算等领域的研究者发展表情识别相关的理论、数据库和算法, 极大地推动了自动表情识别技术的进步。文章基于心理学视角, 结合我们前期开展的相关工作, 首先梳理自动表情识别的心理学基础、情绪的面部表达方式、表情数据的演化、表情样本的标注等方面的理论观点与实践进展, 然后分析指出自动表情识别面临的主要问题, 最后基于预测加工理论的建构观点, 提出注重交互过程中的表情“理解”, 有望进一步提高自动表情识别的有效性, 并预期这可能是自动表情识别研究的未来发展方向。

关键词: 自动表情识别; 基本情绪理论; 情绪维度理论; 表情数据库; 建构论; 情绪标注; 微表情; 面部动作
中图分类号: TP202; F407　　**文献标志码:** A　　**文章编号:** 1673−4785(2022)05−1039−15

中文引用格式: 颜文靖, 蒋柯, 傅小兰. 心理学视角下的自动表情识别 [J]. 智能系统学报, 2022, 17(5): 1039–1053.
英文引用格式: YAN Wenjing, JIANG Ke, FU Xiaolan. Automatic facial expression recognition from a psychological perspective[J]. CAAI transactions on intelligent systems, 2022, 17(5): 1039–1053.

Automatic facial expression recognition from a psychological perspective

YAN Wenjing[1], JIANG Ke[1], FU Xiaolan[2,3]
(1. School of Mental Health, Key Laboratory of Alzheimer’s Disease of Zhejiang Province, Wenzhou Medical University, Wenzhou 325015, China; 2. State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China; 3. Department of Psychology, University of the Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Automatic facial expression recognition is an interdisciplinary and frontier field, spanning psychology, computer science, and other research areas. Researchers in the fields of emotional psychology, pattern recognition, and affective computing develop expression recognition-related theories, databases, and algorithms, greatly progressing the automatic facial expression technologies. Combining the previous related work, the article first discusses the theoretical perspectives and practical advances in the psychological basis of automatic facial expression recognition, facial expression approaches to emotions, facial expression database development, and emotion annotations. Then, it analyzes and highlights the primary issues in automatic expression recognition. Finally, based on the constructivism of the predictive processing theory, it proposes that attention must be paid to “understanding” the facial expressions in interpersonal interaction to further improve the effectiveness of automatic facial expression recognition and be the future research direction.

Keywords: automatic expression recognition; basic emotion theory; dimension theory in emotion; database of facial expressions; constructivism; emotion annotation; micro-expressions; facial actions

如果机器能够像人类一样, 通过识别表情来了解他人的情绪状态, 会是件多么美妙的事情。为实现这个美好的愿望, 几十年来心理学与计算机科学等领域的研究者付出了巨大的努力, 构建理论、采集数据和研发算法, 推动自动表情识别研究不断取得新进展。心理学在为自动表情识别提供思路和启发的同时, 其情绪心理学分支也得

收稿日期: 2021−12−30.　**网络出版日期:** 2022−06−20.
基金项目: 温州市科技计划项目(G20210027).
通信作者: 傅小兰. E-mail: fuxl@psych.ac.cn.

以蓬勃发展，并影响着自动表情识别的未来发展方向。我们前期围绕情绪的相关问题（尤其是微表情），在心理学和计算机科学等学科交叉领域开展工作，考察了情绪与表情的关系、微表情的行为特点，构建了 3 个微表情数据库和一个伪装表情数据库，研发微表情和伪装表情自动识别与检测算法等。

虽然自动表情识别已经取得了重大进展，但是依然存在着一些问题，导致实际应用中存在困难。我们在研究过程中也产生了有关情绪的面部表达及数据标注等方面的困惑，并进行了反思。本文基于心理学视角，首先系统地梳理自动表情识别的心理学基础、情绪的面部表达方式、表情数据的演化、表情样本的标注方法等方面的理论观点与实践进展，然后分析指出自动表情识别面临的主要问题，最后基于心理学的建构论，提出在人际交互过程中进行表情“理解”有望进一步提高自动表情识别的有效性，并预期这可能是自动表情识别研究的未来发展方向。本文是一篇从心理学视角下思考自动表情识别的理论性文章，而非综述性或实证性文章。主要梳理表情识别的心理学基础、情绪的面部表达方式、表情数据的演化、表情样本的标注等方面的理论观点与实践进展，对计算机识别出的“情绪”进行心理学视角的思考。

1　表情识别的心理学基础

情绪心理学中两大流派——基本情绪理论（basic emotion theory）和维度论（dimension approach）——几乎是所有自动表情识别的心理学基础。其中基本情绪理论处于主流地位，因为它有清晰的理论框架，结构化的系统，且与人们的常识体验相吻合。

1.1　基本情绪理论

早在 1872 年，达尔文在《人类与动物的表情》一书中对表情进行了分类[1]。20 世纪 60 至 70 年代，Ekman[2-3] 总结了基本情绪具有的 11 个特点，包括特定的普遍性信号（distinctive universal signal）、灵长类动物共有 (present in other primates)、特定的生理反应 (distinctive physiological response)、特定的普遍诱发事件 (distinctive universals in antecedent events)、一致的情绪性反应（coherence among emotional response）、特定的主观感受（distinctive subjective feeling）等。Ekman 等认为，人类拥有几类基本情绪，诸如高兴、悲伤、厌恶、愤怒、惊讶、恐惧等；这几类基本情绪是离散的、相互独立的；每类情绪都有其特定的主观体验、生理反应与行为表现[2]（见图 1）；基本情绪能够被全人类识别。以这些观点为核心的理论被称为基本情绪理论。自动表情识别领域中的工作大多数是根据基本情绪理论进行情绪分类的[4]。

图 1　基本情绪对应的原型表情示例 (模特为本文第一作者)

Fig. 1　An example of the prototypical facial expressions corresponding to the basic emotion theory (the model is the first author of this paper)

基本情绪理论认为每种情绪都是一个整体。例如，高兴意味着我们内心有愉悦的体验，身体上有心跳加速等生理活动，并可能还有对应的外显动作，如手舞足蹈、眉飞色舞等。这是一个封装好的系统，一旦触动某种情绪则会引发一系列完整、特定的反应[3,5]。基本情绪理论顺应了人类认识活动的一般趋势：对纷繁复杂的事物进行分析，形成清晰的、结构化的知识体系。使用这些简洁的类别标签，我们可以把复杂的情绪过程与性格特征归属为简单的类别，这不仅与多数人的生活体验相契合，也便于人们理解这些心理现象并进行沟通交流，同时也为机器自动表情识别提供了一个结构化的理论框架。

根据基本情绪理论，不同的情绪类型是离散的，相互独立的，有特定的诱发原因、主观体验、生理唤醒和行为反应，那么主观的情绪体验一定会反映在生理与行为上，即个体会表达出可观测的信号以区分内在的情绪体验，内在情绪体验与外在信号的关系是有效的（valid）、特异的（specific）和普遍的（generalized）。所以，通过提取面部动作[6]、肢体动作[7]、语言内容[8]、音频信号[9]、外周生理变化（如心率、血压、皮肤电）[10]，和中枢神经变化（如脑电波、血氧消耗）[11] 等特征，研究者就可以推测个体内在的主观情绪体验。

1.2　情绪的维度论

情绪的维度论由来已久。一个经典的情绪维

度论定义是: 可伴随特定生理活动的正性或负性体验[12]。维度取向曾经一度占据着情绪理论的主流。早在 19 世纪末, “心理学之父”冯特就认为情绪是可以通过愉快–不愉快、激动–平静、紧张–松弛 3 个维度来描述的。Osgood[13] 通过研究发现, 个体在对各种刺激进行判断时, 都会关注其在价值、活力和力量这 3 个因素上的语义差别, 而这些语义差别因素在本质上是情感性的, 是对刺激进行分类的基础。Mehrabian 等[14] 提出了情绪状态的“愉悦度–唤醒度–支配度”三维度模型(pleasantness-arousal-dominance, PAD)。在对 PAD 模型的深入研究中, Russell[15] 发现, 情绪的支配度更多地与其认知活动有关, 愉悦和唤醒两个维度就可以解释绝大部分情绪变异。2008 年, 国内引入了 PAD 情绪量表, 它可以从愉悦度、激活度和优势度上评定心境或情绪状态[16]。Watson 等[17] 采取自陈式情绪研究方法, 提出积极–消极情感模型(PANA), 他们认为积极情感(positive affect, PA)和消极情感(negative affect, NA)是两个相对独立的、基本的维度。

如果使用情绪维度来标注表情样本, 并不需要给出一个明确的情绪类别标签; 情感的维度模型似乎可以在连续的尺度上对每种情绪强度的微小变化进行编码。也有很多学者试图将维度论和基本情绪理论结合, 将基本表情放在两三个维度形成的坐标系中的合适位置, 如情绪的环形模型(circumplex model of affect)[15]。不过, 每一种情绪都是非常复杂的, 虽然我们可以用几个维度来表达某种情绪的主要特点, 但却无法充分地解释或理解这种情绪。

2　情绪的面部表达方式

显然, 6 种基本情绪似乎不足以涵盖我们复杂多样的情绪与对应的表情表达, 而且人类擅长伪装, 表情与情绪有时并不能很好地对应。此外, 表情还受到特定社会文化条件下的展示规则(display rule)的影响。因此, 除了研究基本表情类别, 许多研究者也开始关注微表情、复杂表情和结合其他线索的表情。

2.1　基本表情类型

基本情绪理论把情绪分成几个基本类别, 诸如: 高兴、悲伤、惊讶、恐惧、厌恶、愤怒等[2]。这 6 种基本情绪似乎是泾渭分明的, 且适用于所有人。但是, 科学研究和实践应用都表明, 依靠 6 种基本情绪的分类方式无法涵盖和解释复杂的情绪现象。

最近 Daniel Cordaro 和 Dacher Keltner(两人都曾是 Ekman 的学生)等[18-19] 进行了一系列跨文化研究, 扩展了基本情绪的清单。他们使用情绪编码范式, 系统地分析来自 5 种不同文化背景个体的 22 种情绪表现, 提出了情绪的国际核心模式(international core patterns, ICPs), 即, 在不同文化中存在着 22 种普遍的面部情绪表达规律, 而同时也会受到文化的一些影响。除了最初的 6 种情绪外, 这些研究还提供了在面部和声音表达中出现的情绪如娱乐、敬畏、满足、欲望、尴尬、痛苦、解脱和同情等情绪的证据。表情类型增加到 20 多个, 对表情数据库的建立以及自动表情识别的准确率都提出了新的挑战。

2.2　微表情和伪装表情

微表情是人们隐藏或抑制自己的真实情绪时出现非常迅速泄露的面部动作[20-22]。研究者以时长(根据微表情快速的特点)对微表情进行操作性定义。现在越来越多的研究者将小于 500 ms 的表情定义为微表情[22]。微表情已成为自动表情识别研究的新热点, 因为人们普遍认为微表情泄露了个体的真实表情, 能够反映其真实情绪。

早在《人类与动物的表情》一书中, 达尔文就开始关注难以抑制的情绪表达[1,23]。弗洛伊德也提出人们的情绪都会以某种形式表达出来[24]。Haggard 等[24] 在寻找治疗师和病人之间的非言语交流特征、观察心理治疗动态图片时, 发现了一种“微小瞬间表情(micro-momentary facial expressions)”, 并认为其与压抑和自我防御机制有关。神经心理学的研究发现, 自主表情和非自主表情分别受锥体束(pyramidal tract)和外锥体束(extrapyramidal tract)控制[25]。因此, Ekman 等[21,26] 假设微表情是自主表情和非自主表情之间对抗的产物。

我们前期在实验室里, 通过诱发被试(心理学实验参与者)的情绪(非自主的), 同时要求其伪装自己的表情(自主的), 探索微表情的诱发方法和出现条件[22]。我们基于收集的数据, 拟合不同条件下微表情表达的特点, 描述了自然诱发的微表情的时间和空间特征。结合前人的研究与假设, 我们总结了微表情的表达机制, 提出微表情既可能是个体在自主抑制其情绪表达时真实情绪的泄露, 也可能是个体在正常表达真实表情后因主动抑制而终止的真实情绪表达(见图 2)。至于微表情识别方面的研究, 不是本文的关注点, 感兴趣的读者可以查阅已经发表的综述性文章。

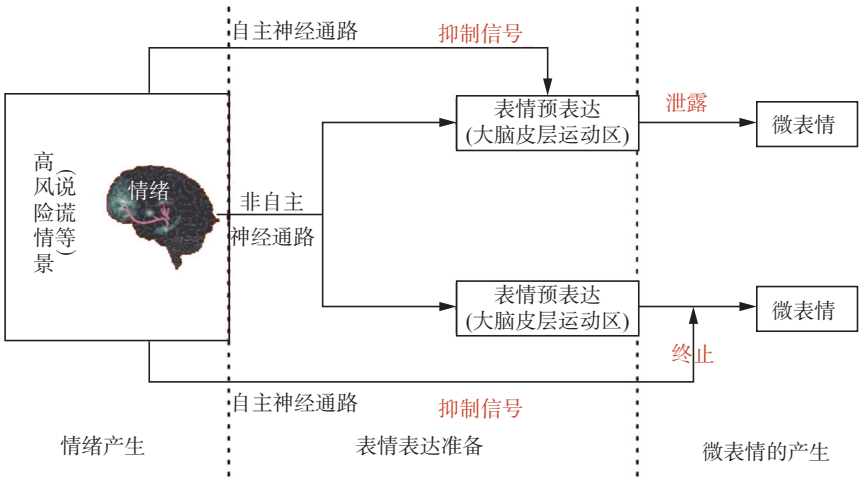

图 2　微表情的产生机制

Fig. 2　The production mechanism of micro-expressions

2.3　复杂表情

一些学者关注“复杂表情”，以期能更好地解释“不太标准”的表情。一篇发表在 PNAS 上的颇有影响力的文章对复杂表情的定义是：复杂表情是由基本表情组合而成的[27]。实验者要求参与者学习原型表情，并且努力摆出原型表情的组合，然后筛选出可以明确识别表情的图片。在此基础上，研究者对这些复杂表情的类别进行分类，得到了较高的准确率，认为该实验证明了 22 种情绪类别的表达和识别是一致的。

Li 等[28] 从社交网络中收集了表情图片，招募 315 名参与者对数以万计的图片进行标注，筛选出多标签的表情图片，建立了一个复杂表情数据库 RAF-ML。该数据库的标注采用的是 6 种基本情绪的标签。如果某个标签的选择人数超过 20%，则标定为存在该种情绪；如果有 2 个以上的标签有 20% 人选择，则定义为多标签（复杂）情绪。这个研究使用的是复杂表情的“操作性定义”。

值得一提的是，虽然关于复杂情绪与表情的研究工作大多是在基本情绪理论框架下开展的，但是该理论的领袖人物 Ekman 早期并不认同“复杂情绪”这个概念。Ekman[2] 认为在生理反应与行为表达上缺乏存在复杂表情的证据。在他看来，所谓复杂的情绪只是多个基本表情的序列呈现，是混合（mixed）而非融合（blend）。

2.4　表情的多模态信息

在过去的 20 年里，对情绪识别的研究已经超越了对 6 种情绪的静态描述，开启了一种多模态的、动态的行为模式，涉及面部动作、发声、身体运动、凝视、手势、头部运动、触摸，甚至气味[29] 的描述情绪表达的方式。例如，凝视模式和头部动作与尴尬[30]、自豪[31] 和敬畏[32] 的体验，以及相应的表达信号交织在一起。Keltner 等[33] 认为，当考虑到不同的模态时，我们就应该认可存在 24 种情绪状态的独特表达。

既然情绪表达是多方面的，那么在表情提供的信息不充分的情况下，就可以加入其他通道的信息，如肢体动作、皮肤温度、语言内容、语气语调、外周生理信号和中枢神经活动等。理论上而言，多模态信息互相补充可以得到更加完整的、更加确定的信号，因此应能获得更好的情绪识别结果，而很多实证研究也证明了这一点。如果多模态信息能够让我们更准确地识别情绪，那么，对机器来说，只要能获得足够的多模态数据，就能够通过深度学习，建立良好的情绪预测模型。

3　表情数据的演化

从最初的 6 种基本表情到更多类型的表情，从摆拍表情到自然表情，从实验室场景中的表情到自然场景（in-the-wild）中的表情，从静态表情图片到动态表情视频，从表情的单一面部动作模式到表情的多模态信息，从小样本到大样本，表情数据库的建设取得了巨大的进展，这是情绪心理学家和情感计算科学家共同努力的结果。

研究者提升机器识别人类情绪的准确性的工作主要集中在基于表情数据库训练出一个计算快速的、鲁棒性高的模型[34-35]，努力使机器能够基于表情准确分类表达者内心情绪的状态。显而易见的是，自动表情识别的准确性在很大程度上受制于数据库中样本标注的质量。

3.1　从摆拍表情到自发的自然表情

早期的表情数据库里大多是摆拍（posed）的原型表情，如 CK+[36]、JAFFE[37]、MUG[38]、RaFD[39]。近年来的表情数据库更加关注表情样本的自发性

(spontaneous)和自然性。有些研究者通过材料刺激或者做某些任务来实现情绪的诱发, 如 DISFA[40]、Belfast Database[41]、MMI[42]、Multi-PIE[43] 等。

构建微表情数据库也同样经历了从摆拍表情到自然诱发表情的过程[44]。我们过去所做的微表情数据库, 就是在实验室里, 采用情绪性视频作为诱发材料, 通过让参与者观看视频来激发参与者的情绪和表情。为了更好地记录被试情绪激发点, 又不干扰其情绪体验, 我们要求被试在有情绪反应时进行按键操作, 以便于在编码时过滤无情绪意义的面部动作。在观看情绪视频结束后, 收集被试情绪体验的主观报告(见图 3)。整理编码之后, 构建了 CASME 系列数据库[45-47]。使用类似的方法, 我们也构建了伪装表情数据库 MFED[48]。当然我们也明确地意识到, 这些在实验室里诱发出的表情样本依然缺乏生态效度。

图 3　微表情诱发范式流程

Fig. 3　The elicitation approach for micro-expressions

既然实验室样本的生态效度不够, 那么就有必要高度关注现实场景(in-the-wild)中的自然表情。与在实验室里诱发得到的表情相比, 现实场景中的自然表情在光照、脸部姿势、尺寸和面部遮挡等方面都有很大的变化, 因此对其分类更具挑战性, 但在实际应用中也更为重要。当前很多数据库从网上(如网页、社交媒体、视频等)抓取大量的表情图片, 并假设它们是相对自然的(不过这些图片中仍有不少是摆拍的), 如 EmotioNet[49]、AffectNet[34]、RAF-DB[50]。自 2013 年以来, FER2013 和 Emotion Recognition in the Wild(EmotiW)[51-52] 等情感识别竞赛基于真实世界场景中收集的相对充足的训练数据, 这也促进了自动表情识别从实验室场景到自然环境的过渡。

3.2　从静态表情图片到动态表情视频

在现实世界中, 人们的表情是一个动态的过程。一个完整的表情可区分为启动阶段(onset phase)、高峰阶段(apex phase)和恢复阶段(offset phase)。而静态表情图片仅仅展示了高峰阶段的一瞬间。

在基于静态表情图片的自动表情识别方法中, 特征表示只用当前单一图像的空间信息进行编码, 而在基于动态表情视频的识别方法中, 则会考虑输入表情序列中连续帧之间的时间关系。对序列(视频)数据进行识别已经成为一种趋势。Li 等[6] 总结了不同类型的方法在动态数据上的相对优势, 包括代表空间和时间信息的能力、对训练数据大小和帧数的要求(可变或固定)、计算效率和性能。心理学的研究也证明, 动态表情能够提供更多的有效信息, 包括区分真实与伪装的表情[53]。例如, 真实笑容的时长一般是在 500~4 000 ms, 而伪装笑容的时长则可能过长或过短[54]; 与非真实笑容相比, 真实笑容的启动时长和恢复时长都更长[55-56]。

3.3　从表情的单一面部动作模式到表情的多模态信息

人类在现实应用中的情绪表达涉及到不同的通道, 而面部表情只是其中一种。所以, 越来越多的多模态表情数据库被建构出来, 如 EU Emotion Stimulus[57]、BAUM-1[58]、AFEW[51]。其中, 最常见的是表情与声音结合的多模态数据库。例如, AFEW 数据库包含了从不同电影中收集的视频片段, 这些视频片段具有自发的表情、各种头部姿势、遮挡和照明, 有时间和多模态信息, 提供了不同环境条件下音频和视频方面的样本。多模态情感分析往往通过处理这些不同的模态来分析人类对某一事物的观点(通常区分为积极的或消极的)[59]。

3.4　从小样本到大样本

在实验室里诱发个体的情绪进而采集表情样本并进行标注, 是一种效率较低的构建表情数据库的方法, 但具有较高的效度, 可以较为明确地区分情绪类型。这些数据库中模特的数量往往在几十到几百人之间。

为了满足深度学习的大数据需求, 很多研究者从网上抓取图片与视频作为样本。这些样本往往无法确定当事人自身的主观体验, 而只能使用

观察者的他人主观标注。典型的数据库是 EmotioNet[49], 包含了百万图像。值得注意的是, 尽管这个表情数据集规模非常大, 但它并非完全由人工标注, 而是通过半自动的方式标注的, 所以可能存在很多噪声。另一个百万级别的表情数据库 AffectNet[34], 是用 6 种不同语言和 1 250 个与情绪相关的关键词在 3 个网络引擎上进行收集的, 并进行了情绪类别和维度(效价和唤醒度)的标注。

4　表情样本的标注方法

目前, 监督学习依然是情绪识别建模中最常用的方法。这需要为可观察到的外在行为与生理信号提供其情绪标注(即 ground-truth)。研究者基于不同的理论和不同的技术对表情样本进行标注, 有基于基本情绪理论或维度论的, 有基于主观或客观, 也有基于行为或生理的。不同的标注取向各有优缺点, 也决定了机器最后的输出结果。主流的标注方式来自基本情绪理论对基本情绪的划分, 诸如高兴、惊讶、厌恶、悲伤、愤怒、恐惧等。一些研究者会使用一些变式或者更多的情绪类型。而另一些研究者会(往往是同时)使用情感维度来标注, 如愉悦度、唤醒度和优势度。研究者们给行为或者生理信号标注情绪的方法既有主观的也有客观的。

4.1　体验者主观标注

体验者的自我报告是目前最具有分辨力的情绪测量方法[60], 因为情绪本质上是一种主观体验。其操作过程一般是先诱发出当事人的某种情绪体验, 然后要求体验者描述自己的情绪。例如, 研究者用一个刺激物来唤起当事人的情绪, 如情绪性的图片、视频, 或者对某一事件的描述, 如“你的表哥刚刚去世, 你感到非常悲伤”[19]。但对大多数人来说, 描述自己的主观情绪体验并非一件容易的事。这需要体验者具有较好的情绪感受能力, 愿意且能够表达出自己的情绪体验。另外, 个体在关注自己的情绪时往往会影响自己的情绪体验[61]。因此, 除了要求参与者描述他们的感受外, 更常用的方法是要求参与者从一组情绪形容词中选择自己当时体验到的情绪并对情绪进行评分[22,47]; 有时候还使用事后回溯的方式[62]。虽然词表可能有很多候选词, 但是研究者最终往往会将候选词简化为若干种“基本情绪”。参与者所体验到的情绪, 都可以被归类于基本情绪中的某一个“家族”, 例如, 高兴包含了兴奋、满足、愉快、舒适等一系列的积极情绪体验。

4.2　观察者主观标注

许多表情数据(如从网上抓取图片与视频)并没有当事人主观体验的任何信息, 所以研究者只能使用观察者的他人主观标注, 即要求观察者在观看相关表情材料后, 判断该材料对应的情绪类型。观察者主观标注的大部分材料是非实验室场景下拍摄的。由于这些表情往往不那么“标准”, 使得基于面部动作(AU)组合来判断表情的方法难以实现。因此研究者会通过“众包”的方法, 让一定数量的观察者为每一张图片进行情绪类型的标注从而达到一定程度的“标准化”。这种方法蕴涵的假设是: 情绪识别在人类中是普遍的, 具有跨文化的一致性; 人的判断是可靠的、特异的和具有普遍性的; 表情的表达者(编码者)与接收者(解码者)之间的信息沟通是通畅的。近期有一些表情数据库就是用这种方法进行标注的, 如 RAF-ML[28]、AffectNet[34]。

4.3　基于行为的客观标注

除了主观标注的方式外, 有研究者还采用一些客观标准来标注情绪。最常见的做法是事先定义一些情绪的动作单元(AU)组合。这种情绪–表情关系表一般参照 FACS(facial action coding system)研究手册[47] 或者由研究者自己设定。FACS 是一个基于解剖学的描述面部动作的工具, 用于描述所有视觉上可识别的面部运动。该系统由 Paul Ekman 和 Wallace V. Friesen 于 1978 年创立, 由 Ekman, Friesen 和 Joseph C. Hager 于 2002 年予以更新[63]。他们根据面部肌肉的解剖学特点及其外部表现特点, 将面部动作划分成几十个相对独立的动作单元(action unit, AU)。AU 表现为一个或多个面部肌肉的收缩或放松, 例如皱眉、抿嘴等。FACS 可以对面部各种动作的位置、形态、强度和时长进行相对客观地标记, 是目前最常用的描述面部动作的编码工具。

进行 FACS 编码十分耗时, 尤其是对视频进行逐帧编码的时候需要耗费大量时间成本。所以, 许多研究者努力研发基于计算机的自动编码系统[64-66]。2020 年 EmotioNet 挑战赛中, 有研究者通过 100 万张图像训练了非刚性的面部肌肉运动(主要是前 17 个 AU)和刚性的头部运动(最后 6 个 AU)的 FACS 编码算法。他们将 AU 识别问题作为一个多任务学习问题, 前 17 个 AU 准确率为 94.9%, 精确性和召回率的综合指标(称为 F1, 范围从 0 到 1)在验证集中达到 0.746, 在挑战赛的测试集中也达到了 0.730 6 的最终成绩[67]。

我们的研究结果也显示, 基于 AU 的标注方法结构化水平很高, 完全以表面形态(几何特征、纹理特征)为基础, 这种方法非常“适合”计算机

视觉和模式识别技术。所以, 许多数据库也选择基于 AU 组合来做情绪标注, 并获得了令人满意的效果, 如 Emotionet[49]。在情绪标注过程中, 有些数据库的开发人员基于 AU 组合的同时, 也尽可能地考虑主观报告与视频的内容[46-47]。但是, 标注准确性依然会受到情绪体验与表情之间的一致性水平的约束, 因为只有提供了一致的表面形态标准, 计算机才可以对表情特征做很好的分类。

4.4　情感维度标注

非摆拍条件下的表情照片中, 符合原型表情的动作组合较少, 所以基于原型表情模板进行情绪类型的标注比较困难。而基于 FACS 提供的“核心 AU”分析也很难确认某个表情的情绪类别。而根据情感维度模型, 则没有必要假设独立的离散的情绪类型。这种观点认为, 少量的两极维度可以作为情感体验和情感识别的基本构件[15]。这也是为什么许多非摆拍的样本也标注了维度, 如 AFEW-VA[68], AffectNet[34]。

从愉快到不愉快的效价(Valence)维度在定义情绪体验和表达方面至关重要。这一维度能够被人类自动地、快速地识别出来, 而且具有普遍性[69]。毕竟, 积极和消极的情感状态位于情感空间的相反位置, 它们以一种非常不同的方式被传达[69]。所以, 效价似乎是非常容易标注的, 而唤醒度(Arousal)的标注比较困难。例如, 哭泣是唤醒程度低的情绪吗? 生闷气的唤醒程度是否比哭泣高呢, 高多少呢? 而且, 在较低的效价和唤醒度状态下, 人们哪怕有情绪体验, 也往往面无表情。

5　自动表情识别面临的主要问题

在实践中, 从数据的标注到计算机的识别, 我们常常会遇到一些困难。在数据标注过程中, 我们很难确定这些表情是否确切地反映了某种情绪。虽然在数据采集过程中, 我们收集了主观评价、评估了视频的情绪特点并进行了面部动作编码, 但是却发现主观评估与面部动作有时并不匹配(基于基本情绪理论的观点应该是匹配的)。而且, 我们还发现巨大的个体差异, 例如, 有些人看到恶心的内容会表现出大笑, 但是这个大笑并不等于“高兴”, 然而当事人又说不清是什么情绪。于是, 虽然基于数据库的自动表情识别准确率非常高, 但是在现实生活情景中的识别准确率往往不是很高, 难以应用于实践。

5.1　问题一: 表情与真实情绪体验的一致性

我们前期在微表情数据库的构建以及微表情分析等领域做了一些颇有成效的工作, 但也发现情绪与表情的一致性并没有理论预期得那么高。同时, 大量研究也表明, 人的内在情绪体验和外在表情、生理信号之间的相关性较低。

Durán 等[70] 进行了一项荟萃分析(元分析), 其包含了 37 篇关于情绪体验与原型表情之间关系的研究。研究通过计算相关系数, 来确定一种情绪与所设定表达之间的一致性程度(见表 1)。荟萃分析的结果显示, 高兴与典型笑容的总体相关系数是 0.40(95% 的置信区间为 0.31~0.49)。如果我们把高兴(Happiness)和好玩(Amusement)看作是两种相互独立的情绪, 那么与微笑相关的总体估计值是: 快乐为 0.27[0.16, 0.39], 好玩为 0.52 [0.43, 0.62]。而参与者在高兴时出现典型笑容的概率是 0.41[0.08, 0.73]。如果把高兴和好玩分开考虑, 则高兴的概率为 0.12[0.06, 0.18], 好玩的概率为 0.47[0.09, 0.84]。

表 1　情绪与原型表情表达关系的元分析结果(Duran, 2017)

Table 1　The meta-analysis for the relationship between felt emotions and prototypical facial expressions

情绪	被试数	相关系数(95%置信区间)	被试数	反应概率(95%置信区间)
高兴+好玩	1 398	0.40[0.31, 0.49]	217	0.41 [0.08 0.73]
高兴	732	0.27[0.16, 0.39]	98	0.12[0.06, 0.18]
好玩	666	0.52[0.43, 0.62]	119	0.47[0.09, 0.84]
惊讶	168	0.24 [0.04, 0.44]	515	0.09 [0.05, 0.14]
厌恶	187	0.24 [0.10, 0.37]	279	0.32 [0.14, 0.50]
悲伤	247	0.41 [0.20 0.63]	119	0.21 [0.14, 0.29]
愤怒	281	0.22 [0.11, 0.33]	133	0.28 [0.20, 0.35]
恐惧	60	0.11 [–0. 14, 0.36]	170	0.34 [0.00, 0.74]

在所有测试的情绪类别中, 除了恐惧之外, 其他情绪与原型表情的相关系数均高于随机水平。然而, 高于随机水平并不能说明特定情绪可以对应到特定表情。实际上, 它们之间的相关性很

弱。进一步的荟萃分析[71]考察了来自 76 项研究的 131 个效应大小，共计 4 487 名参与者，也获得了类似的结果：原型表情与愤怒、厌恶、恐惧、快乐、悲伤或惊讶情绪的测量之间的总体相关系数为 0.31（弱相关），在情绪事件中观察到对应的标准面部动作的平均概率是 0.22。

以上这些研究结果表明，人们其实很难根据他人的面部动作有效地预测其内在情绪状态。从生活经验的角度看，这个结果并不意外。我们以“恐惧”情绪为例，面对潜在的危险，人和动物都可能产生所谓的 Freeze（呆若木鸡）、Fight（狗急跳墙）、Flight（逃之夭夭）等多种反应模式。在主观体验、生理唤醒和行为表现等方面，个体的表达方式千差万别，而在许多研究中都只用单一的恐惧反应来描述它们。然而，有研究表明这些恐惧情绪的行为表达所对应的神经环路也不同，不应该被归为同一类型[61]。

Barrett 等[72]指出了基本情绪理论相关研究中的 3 个关键缺陷：1）可靠性（reliability）有限，即同一情绪类别的实例既不能通过一套共同的面部动作可靠地表达，也不能从一套面部动作去推论个体的情绪；2）缺乏特异性（specificity），即不同的面部动作和对应的情绪类别之间没有独特的映射关系，即被标注为微笑的识别标签，并不一定是高兴的表情，皱眉也不一定是愤怒的表情；3）有限的普遍性（generalization），即没有充分的证据表明情绪表达的跨文化一致性。由于先前的跨文化证据往往存在方法上的缺陷，而这些缺陷导致了一种普遍的误解，即对情绪与面部动作之间关联性的误解，这一误解又进一步限制了这一证据在其他用途中的转化。Barrett 等[73]的总体结论是明确的：“从一个微笑中推断出快乐，从一个皱眉中推断出愤怒，或从一个皱眉中推断出悲伤，这样的推断是不可能具有足够信心的；而目前的许多技术正在运用这些错误的推断，并且这些错误的推断往往被认为是科学事实”。

5.2　问题二：人工标注的准确性

表情与真实情绪体验的一致性不高，会导致人工标注的有效性受到质疑。

如前所述，许多表情数据库的编码是基于行为的客观标注，即基于情绪–表情对应表。虽然 FACS 提供了一个情绪–表情对应表，但是后来的研究者在实际使用中并没有严格地参照。实际上，情绪与 AU 组合的映射关系哪怕在各个支持基本情绪理论的研究者眼里也没有达成一致[19,63]。而如今，越来越多的研究发现情绪与表情的相关性不高，这意味着基于 AU 确定表情的情绪类型可能是不准确的。而且，各数据库的标注标准差异也非常大。以悲伤为例，有的认为是 4+15[49]，有的则认为应该是 1+4+15 或 11 或 6+15[74]。

另一些表情数据库是根据观察者的判断进行标注的。之所以这样做，是基于下述（基本情绪理论的）假设：人的判断是可靠的、特异的和具有普遍性的；表情的表达者（编码者）与接收者（解码者）之间的信息沟通是通畅的。但是，该假设可能并不成立。例如，越来越多的研究表明，当人们推断面部结构中的情感含义时，背景是一个重要的、有时甚至是主导性的信息来源[75-76]。这个背景信息可以是观察者的状态、事件的前因后果、表达者所处的场景等[77]。也就是说人们是基于多方面的信息去理解对方的情绪，而不仅仅是根据个体的表情。这时候的情绪标签，很难保证反映了图片中个体的内在情绪体验。此外，观察者主观标注的方法还存在一个统计上的悖论。基本情绪理论通过高于随机水平的“表情识别能力”来证明基本情绪的存在，并以此标注“正确答案”。但是，人们的识别能力存在着个体差异且经常会存在“识别错误”，如混淆愤怒与厌恶、惊讶与恐惧等情况，因此单纯靠人的主观判断似乎是不可靠的。一群普通人进行情绪评估得到的“平均答案”作为“标准答案”来训练计算机，其结果也只是计算机的情绪识别水平会更接近“平均水平”。

主观报告似乎是情绪标注的一个可靠方式。一些数据库的开发人员基于 AU 组合的同时，也尽可能地考虑主观报告与视频的内容，如[46-47]。但是，基于体验者主观标注的方法存在两个问题，一是个体很难准确地描述自己的情绪体验。情绪的变异性过大导致难以被收敛到简单的标签；二是参与者被迫用几个预置设定的情绪词来表征自己的真实情绪，这种“迫选”式的设定可能会歪曲当事人的真实情绪体验[72,78]。而且，标注准确性依然会受到情绪体验与表情之间的一致性水平的约束——只有提供了一致的表面形态标准，机器才可以对表情特征做很好的分类。

如果采用的是维度标注方法，也需要关注下述两个问题：第一，效价与唤醒度的评分本身没有标准，主观性非常强。每个材料的标注可能都只是由一个人或者两个人来完成的[34,68]，重测信度较低[41,79]。另外，标注很大程度上基于情绪体验者的外部表现，而表情难以反映其内心的情绪，或内心的情绪常常不会反映在外部。例如，一般认为悲伤情绪可能会被认为处在低效价和低唤醒度象限里，但是当我们能够看到一个人明显的悲伤表情时，往往意味着此时他（她）的情绪体验激烈，

唤醒度可能很高。又如, 唤醒水平低且效价较高时, 人往往是处于舒适满足的状态, 这个时候大部分情绪体验者是面无表情的。这也许解释了为什么在 AFEW-VA 数据库中低效价象限中样本很少。第二, 效价和唤醒两个维度构成的环形模型[15]并不能解释大多数具体的情绪事件。Russell[78] 也认为情感维度模型并没有对典型的情绪事件提供足够丰富的解释。例如, 该模型未能充分解释恐惧、嫉妒、愤怒和羞愧有什么差异, 也无法解释观察者是如何区分它们的。近年来, 建构论的观点认为, 效价与唤醒两个核心要素仅仅是情绪的组成部分, 还需要对自身、环境等信息的整合, 才能形成特定的情绪。Russell[78] 的比喻是: 星座是最后赋予的意义解释, 而其中的星星只是各个成分。所以, 就算机器能够计算出某个人某时某刻的效价与唤醒度, 也不能输出一个人们能够理解的"情绪"结果。此外, 还存在一个更加具有挑战性的质疑: 评分者基于外部反应的主观标注(效价与唤醒度)本身也可能是不准确的。

5.3　问题三: 情绪与表情的变异性

当我们尝试用一个标签代表一类情绪或表情时, 会遇到一些困难。

例如, 在实验室诱发笑容(标注为 happiness)似乎是非常容易的——给参与者看一些喜剧片的搞笑片段就可以了, 但这种大笑并不意味着参与者的内心是愉悦幸福的。我们中了大奖、表白成功、获得学术奖项或者吃一顿美食时候的愉悦感与幸福感, 和观看视频产生的"好玩(amusing)"体验相去甚远。而且还有不少研究者发现, 人们在体验到幸福快乐的时候并不一定会笑, 而是在跟其他人进行交互的时候才会频繁地笑[5]。更有甚者, 有些被试看到恶心的内容会表现出大笑。

再如, 以观看恐怖片时诱发情绪过程为例。虽然我们知道电影中的场景非常可怕, 但也知道自己是安全的, 所以很多人乐于体验那种刺激的"愉悦感"。当出现某些恐怖场景时, 我们会选择一种回避的状态, 但是这种回避只是眯着眼睛或者转过头去。如果在森林里遇到危险物(如老虎之类的野兽), 我们可能会吓得僵直, 或者睁大眼睛寻找逃跑的路, 或者张大嘴巴发出惊叫以寻求帮助或吓退对象。这些反应都是根据当时情境做出的适应性反应[69]。对比看恐怖片和身处真实的危险场景这两种情况, 虽然我们把其中的情绪体验都叫做恐惧, 但实际上无论是主观体验还是行为反应都截然不同, 似乎不应该归为同一类。

这意味着, 情绪与表情的一致性可能没那么高, 个体的主观报告没有那么清晰准确, 而观察者也很难基于其表现确认其真实的情绪体验。例如, 我们见到他人打招呼时, 往往会伴随着微笑, 目的是让别人觉得"见到你很高兴", 而非真实的主观高兴的情绪体验; 而这时如果让机器进行识别, 机器会将这种表现识别为"高兴", 但不一定能反映人们内心的真实状态。又例如, 一些运动员在战胜对手时, 狂喜中却出现十分"痛苦"的表情[75], 机器可能会将其识别为"悲伤"或者"厌恶"; 许多抑郁症患者同样会面带微笑[80], 但是内心往往是不快乐的。于是, 在基本情绪理论基础上的自动表情识别系统会出现生态效度较低的问题, 即, 虽然基于数据库的表情识别准确率非常高, 但是在现实生活情景中的应用价值却很有限。

6　表情识别的未来进路

以上问题表明, 情绪与表情的关系很复杂, 表情样本数据的效度比较低, 自动表情识别仍然面临巨大的挑战。一方面, 现实中的大部分人的表情不是以原型表情的形式出现, 甚至与这些原型表情根本不相似。于是, 基于刻板的表情模板去识别现实情景中的表情几乎不可能。另一方面, 人类会根据现实情景和自己的经验来理解他人的情绪, 而不太依赖于面部肌肉、皮肤的形状与纹理来做判断, 即不太会受到"长什么样"的干扰。也就是说, 人对他人情绪的识别是"格式塔式的(gestalt)"而不是"刻板分类的"——人类的情绪识别方式与机器的识别方式相去甚远。未来工作中, 我们可能需要明确表情识别的目标, 以及尝试从基于预测加工理论的建构论观点来理解情绪。

6.1　表情识别的目标

自动表情识别的目标是准确识别他人的情绪类型, 还是努力理解人类的情绪并学习人类的情绪识别方式?

如果是前者, 则识别任务的设定必须是基于"表情、语言、生理信号能够准确反映人的情绪"这一理论假设。如果计算机识别成绩能超越人类的识别成绩, 则表明计算机工作的成绩优于常人。如果是后者, 工作重点则是理解人类的情绪, 并让计算机尽可能模仿人类的情绪识别方式。在这种模式下, 不再关注计算机的情绪识别是否比人类更准确, 而是计算机的情绪识别是否接近人类识别的成绩。例如, 张三现在内心很悲伤, 但是他笑得很开心的样子, 那么理想的识别模型应该将这个表情识别成悲伤还是高兴呢? 如果识别为"悲伤"则体现了"察言观色"的真正目的, 即"理解人"的心理活动; 如果识别为"高兴", 体现为模仿人的目标, 即"像大多数人"一样识别他人的表情。

计算机表情识别的目标选择与应用场景存在关联。在一些场景中，我们训练计算机是为让它了解人们内心的真实情绪，即所谓“读懂对方”，例如共情、测谎等任务。而有时候，我们仅仅希望机器能够像人一样，能看出对方希望展示的情绪状态（如打招呼时高兴的表情），或者能借助场景与经验推测对方的情绪。那么，训练计算机情绪识别时，首先应该考虑应用场景和明确的任务目标。

然而，当前很多研究者并没有考虑这两个目标的差异，在情绪识别模型建构时，往往默认情绪识别的目标是努力通过测量外部信号推测人的情绪类型。更具体而言，即是识别并区分几种有限的基本情绪类型，如高兴、悲伤等。这一目标往往事先假设了“外部信号与内部情绪是一致的”。唯有这样，情绪识别模型才能满足反向推断的要求，即，根据外在表现推断内心情绪[72]。但是，这一假设实际上可能并不成立（见 6.1 节）。对于第二个目标，即让机器尽量模仿人，似乎只需要找一群有代表性的普通人，根据情绪词表来进行情绪类别的标注，即“众包”（Crowdsourcing）[50]。只要众包的数据量足够，似乎机器就能够像人一样识别他人的情绪了。然而，这种识别并不是真正模仿了人类的表情识别方式（见 6.2 节）。

我们分析表情识别的目标，并反思情绪的本质，以及在表情识别领域人工智能的角色和定位。研究发现，无论是情绪的表达还是情绪的识别，都不仅仅是一个“分类”的过程，而是一个建构的过程[81]。按照这种建构取向，情绪本身并不存在“可分类”的信息，或者说这些情绪类型本质上并不存在——情绪类型只是人们在交互过程中的建构。如果情绪本身在概念意义上缺乏足够的结构性特征，那么，关于情绪的类别化也就没有充分的标准，进而也无法通过数据库所提供特征与标注并训练出一个计算化模型。因此，前述情绪识别的两个目标都无法实现。

6.2　交互中的建构与情绪理解

前期的实践结果显示，基于基本情绪理论训练计算机识别系统似乎无法精确地反映人类情绪的本质，也难以在实践中获得有价值的应用效果。因此，我们需要更深入地理解人的情绪识别特点。

我们可能很难根据某一瞬间（一张图片）正确断定一个人的情绪。多数情绪识别是在交互过程中慢慢确认的，需要不断地修正原来的判断[82]。这就是面部表达的行为生态学观点（behavioral ecology view of facial displays, BECV）。也就是说，一个人对另一个人的表情识别是在持续不断地交互过程中建构的。对一个人的愤怒表达，有许多解释的角度，如攻击的语言内容是指向自己的还是维护自己的（在骂别人）。个体从情绪情景中所感受到的情绪特征，绝不只是用愤怒或者不愤怒这个维度来评价的。接收者可能会考虑情绪表达者是否对自己有恶意、是否在呵护自己等角度来进行“识别”，进而形成不同的情绪体验，并做出不一样的行为反应。因此，整个过程的动态性和复杂性只能在持续地建构过程中才能实现。相应地，用简单的情绪分类来理解情绪并不真正符合日常生活中人们的情绪体验与行为反应。总之，个体对恐惧、愤怒、喜悦和悲伤等情绪的体验都是融合了情感表征、身体知觉、对象知觉、评价观念和行为冲动等内容而形成的整体性体验。从这个角度来看，情绪并非一个静态结构，而是一个建构过程。

建构论的观念最初源自 20 世纪初的社会学、人类学和社会心理学的社会互动理念，后经皮亚杰、维果斯基等的阐释与倡导，到 20 世纪末形成了一股强调社会互动和生成认知，强调动作导向的哲学、社会学和心理学思潮。建构论反对古老的理性主义，强调知识不是人出生时预留在头脑中的；它也反对经验主义，认为知识不是物理的或社会的环境给主体的认知碎片组合而成的。建构论认为，知识是主客体互动过程中生成的[83]。按照这样的观点，情绪识别不是基于人先天拥有的对“基本情绪”的表达和识别知识；也不是通过条件反射式的经验学习而获得的能力。情绪本身——包括表达与识别——是人际互动过程中逐渐生成的体验。

2013 年，Clark[84] 提出了一个基于贝叶斯计算和神经科学的预测加工理论（predictive processing）。根据预测加工理论，我们不再需要通过外在的知觉信号或行动去推测个体内在的“本质”状态，因为那种将个体的外部表现当作其内在状态表征的观念早已化作“老生常谈”（stale old debates），应该被抛弃了[85]。在预测加工理论的框架中，脑被看作是一个基于贝叶斯概率理论来评估环境信息的计算机。在个体与环境的互动过程中，大脑对互动进程中的先验概率（prior probability）、预测信号（prediction-signal）、后验概率（posterior probability）、似然性（likelihood）等进行实时地评估和计算，从而实现最小知觉偏差（minimise prediction error）。通过最小知觉偏差，个体与环境的互动得以维持在适度的平衡范围内，也就是大脑实现的“最佳猜度”（best guess）。关键是，这种最佳猜度是行动导向的（action-oriented），即，是

在个体与环境的互动过程中形成和调节的[74]。因此, 预测加工理论实际上是一种基于贝叶斯计算的建构论。

从预测加工理论的建构论视角来看, 我们不应该努力地做所谓的“情绪分类”, 即, 不再基于外部的行为与生理指标来推测当事人的内部有哪种情绪状态; 而应该基于个体与他人及情境的互动与建构去做“情绪理解”。唯有这样, 我们在前面陈述的情绪识别所遭遇的诸多困难有可能得以化解。

在日常生活中, 如果我们一开始就给他人的反应贴上具体的情绪标签, 那么, 很可能会因为情绪标签的片面性而误解了对方的情绪, 或者因为语言的抽象性而抽离了对方反应的生态意义, 使得具身(embody)的情绪体验变成了一个抽象的情绪识别命题。在现实的交互过程中, 情绪体验是非常具体而鲜活的。人们会在交互过程中不断地建构、修正对他人情绪的理解。例如, 当我看到一个人独自安静地坐在角落, 可能会先形成一个假设: 他现在不高兴。基于这个假设, 我会进一步预测: 如果我现在去和他开玩笑, 极有可能会激惹他。这个预测更进一步激发我的下一个假设: 在这个情景中我最好不要去打扰他。这个假设将继续触发了我对与这个人互动的下一步的预测……; 在这个过程中, 每一个环节上的假设都会激活下一步的预测, 这个预测又进一步成为下一个环节的假设……。对当事人而言, 在特定情景中, 时刻 T 的假设–预测链必然是以时刻 T_{-1} 的假设–预测链为前提而建构的, 而在时刻 T_{-1} 之前, 还有时刻 T_{-2}……[86]

同时, 从对方的角度来看, 那个静静坐在角落的人可能原本并没有特别的情绪, 只是安静地坐在那里。但他觉察到我靠近他, 又安静地离开, 他也会形成一系列的假设–预测链, 例如: 这个人平常会与我开玩笑, 今天却表现得很冷漠, 也许是他对我有不满意; 也许是因为我之前什么事情令他不高兴了……在这样的假设–预测链中, 当事人事实上体验到了某种不高兴的情绪。因此, 他的不高兴情绪并不是从一开始被我“识别”出来的, 而是在我和他的互动过程中建构出来的。

如果我在进入这个情景中的第一时间形成的假设是: 他看起来很安静, 我可以过去和他开开玩笑……则在我与他之间将形成另外不同的互动模式, 双方也将在另一种互动模式中建构另外的情绪体验。总之, 在这个过程中, 这个人的情绪体验和表达, 以及周围人对他的情绪的理解都不是根据一个时刻的静态的表现就确定了的, 而是在双方的互动过程中, 根据反馈信息逐渐地校准关于对方的情绪的评估, 并最终让当事人的情绪体验与之前的预测逐渐靠近, 实现了最小知觉偏差。预测加工理论为这个互动建构过程提供了一个可计算的模型。

7　结论

综上所述, 自动表情识别作为心理学与计算机科学等深度交叉的前沿领域, 受到了众多专家的关注。我们梳理自动表情识别的心理学基础、情绪的面部表达方式、表情数据的演化、表情样本的标注等方面的理论观点与实践进展, 然后分析指出自动表情识别面临的主要问题, 最后基于预测加工理论的建构观点, 提出注重交互过程中的表情“理解”。我们认为, 情绪理解是动态的过程, 需要根据事件的进展而不断建构并修正自己的解释。因此, 自动表情识别的研究重点应该着眼于对个体在与其他人或场景进行互动过程中的心理体验的理解。基于此, 我们有理由期待, 自动表情识别的有效性可以进一步提高, 并开启表情识别的 2.0 时代。

参考文献:

[1] DARWIN C. The expression of the emotions in man and animals[M]. New York: Appleton and Company, 1872.

[2] EKMAN P. An argument for basic emotions[J]. Cognition and emotion, 1992, 6(3/4): 169–200.

[3] EKMAN P, CORDARO D. What is meant by calling emotions basic[J]. Emotion review, 2011, 3(4): 364–370.

[4] SARIYANIDI E, GUNES H, CAVALLARO A. Automatic analysis of facial affect: a survey of registration, representation, and recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(6): 1113–1133.

[5] CRIVELLI C, FRIDLUND A J. Inside-out: from basic emotions theory to the behavioral ecology view[J]. Journal of nonverbal behavior, 2019, 43(2): 161–194.

[6] LI Shan, DENG Weihong. Deep facial expression recognition: a survey[J]. IEEE transactions on affective computing, 2020(99): 1.

[7] NOROOZI F, CORNEANU C A, KAMIŃSKA D, et al. Survey on emotional body gesture recognition[J]. IEEE transactions on affective computing, 2021, 12(2): 505–523.

[8] SAILUNAZ K, DHALIWAL M, ROKNE J, et al. Emo-

tion detection from text and speech: a survey[J]. Social network analysis and mining, 2018, 8(1): 1–26.

[9] ZHAO Huijuan, YE Ning, WANG Ruchuan. A survey on automatic emotion recognition using audio big data and deep learning architectures[C]//2018 IEEE 4th International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing, (HPSC) and IEEE International Conference on Intelligent Data and Security. Omaha, IEEE, 2018: 139−142.

[10] SHU Lin, XIE Jinyan, YANG Mingyue, et al. A review of emotion recognition using physiological signals[J]. Sensors (Basel, Switzerland), 2018, 18(7)–2074.

[11] ALARCÃO S M, FONSECA M J. Emotions recognition using EEG signals: a survey[J]. IEEE transactions on affective computing, 2019, 10(3): 374–393.

[12] SCHACHTER S, SINGER J E. Cognitive, social, and physiological determinants of emotional state[J]. Psychological review, 1962, 69: 379–399.

[13] OSGOOD C E. Dimensionality of the semantic space for communication via facial expressions[J]. Scandinavian journal of psychology, 1966, 7(1): 1–30.

[14] MEHRABIAN A, RUSSELL J A. An approach to environmental psychology[M]. [S. l.]: The MIT Press, 1974.

[15] RUSSELL J A. A circumplex model of affect[J]. Journal of personality and social psychology, 1980, 39(6): 1161–1178.

[16] 李晓明, 傅小兰, 邓国峰. 中文简化版 PAD 情绪量表在京大学生中的初步试用 [J]. 中国心理卫生杂志, 2008, 22(5): 327–329.
LI Xiaoming, FU Xiaolan, DENG Guofeng. Preliminary application of the abbreviated PAD emotion scale to Chinese undergraduates[J]. Chinese mental health journal, 2008, 22(5): 327–329.

[17] WATSON D, TELLEGEN A. Toward a consensual structure of mood[J]. Psychological bulletin, 1985, 98(2): 219–235.

[18] CORDARO D T, KELTNER D, TSHERING S, et al. The voice conveys emotion in ten globalized cultures and one remote village in Bhutan[J]. Emotion (Washington, D C), 2016, 16(1): 117–128.

[19] CORDARO D T, SUN Rui, KELTNER D, et al. Universals and cultural variations in 22 emotional expressions across five cultures[J]. Emotion (Washington, D C), 2018, 18(1): 75–93.

[20] 梁静, 颜文靖, 吴奇, 等. 微表情研究的进展与展望 [J]. 中国科学基金, 2013, 27(2): 75–78, 82.
LIANG Jing, YAN Wenjing, WU Qi, et al. Recent advances and future trends in microexpression research[J]. Bulletin of National Natural Science Foundation of China, 2013, 27(2): 75–78, 82.

[21] EKMAN P, FRIESEN W V. Nonverbal leakage and clues to deception[J]. Psychiatry, 1969, 32(1): 88–106.

[22] YAN Wenjing, WU Qi, LIANG Jing, et al. How fast are the leaked facial expressions: the duration of micro-expressions[J]. Journal of nonverbal behavior, 2013, 37(4): 217–230.

[23] EKMAN P. Darwin's contributions to our understanding of emotional expressions[J]. Philosophical transactions of the Royal Society of London Series B, Biological sciences, 2009, 364(1535): 3449–3451.

[24] HAGGARD E A, ISAACS K S. Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy[M]// Methods of Research in Psychotherapy. Boston, MA: Springer, 1966: 154−165.

[25] RINN W E. The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions[J]. Psychological bulletin, 1984, 95(1): 52–77.

[26] EKMAN P. Lie catching and microexpressions[M]//The Philosophy of Deception. [S. l.]: Oxford University Press, 2009: 118−136.

[27] DU Shichuan, TAO Yong, MARTINEZ A M. Compound facial expressions of emotion[J]. PNAS, 2014, 111(15): E1454–E1462.

[28] LI Shan, DENG Weihong. Blended emotion in-the-wild: multi-label facial expression recognition using crowdsourced annotations and deep locality feature learning[J]. International journal of computer vision, 2019, 127(6/7): 884–906.

[29] KELTNER D, TRACY J, SAUTER D A, et al. Expression of emotion[J]. Handbook of emotions, 2016: 467–482.

[30] KELTNER D. Signs of appeasement: evidence for the distinct displays of embarrassment, amusement, and shame[J]. Journal of personality and social psychology, 1995, 68(3): 441–454.

[31] TRACY J L, ROBINS R W. Putting the self into self-conscious emotions: a theoretical model[J]. Psychological Inquiry, 2004, 15(2): 103–125.

[32] CAMPOS B, SHIOTA M N, KELTNER D, et al. What is shared, what is different? Core relational themes and expressive displays of eight positive emotions[J]. Cognition & emotion, 2013, 27(1): 37–52.

[33] KELTNER D, CORDARO D T. Understanding multimodal emotional expressions[EB/OL]. The science of facial expression: Oxford University Press, 2017: 57−75. [2020−01−01] . http://emotionresearcher.com/wp-content/uploads/2015/08/Keltner-and-Cordaros-Original-Paper-With-Changed-Text-Highlighted.pdf

[34] MOLLAHOSSEINI A, HASANI B, MAHOOR M H. AffectNet: a database for facial expression, valence, and arousal computing in the wild[J]. IEEE transactions on affective computing, 2019, 10(1): 18–31.

[35] YAN Wenjing, LI Shan, QUE Chengtao, et al. RAF-AU Database: In-the-Wild Facial Expressions with Subjective Emotion Judgement and Objective AU Annotations [C]//Asian Conference on Computer Vision. Cham: Springer, 2021: 68−82.

[36] LUCEY P, COHN J F, KANADE T, et al. The Extended Cohn-Kanade Dataset (CK): a complete dataset for action unit and emotion-specified expression[C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. San Francisco, IEEE, 2010: 94−101.

[37] LYONS M, AKAMATSU S, KAMACHI M, et al. Coding facial expressions with Gabor wavelets[C]//Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition. Nara, Japan. IEEE, 1998: 200−205.

[38] AIFANTI N, PAPACHRISTOU C, DELOPOULOS A. The MUG facial expression database[C]//11th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 10. Desenzano del Garda, Italy. IEEE, 2010: 1−4.

[39] LANGNER O, DOTSCH R, BIJLSTRA G, et al. Presentation and validation of the radboud faces database[J]. Cognition and emotion, 2010, 24(8): 1377–1388.

[40] MAVADATI S M, MAHOOR M H, BARTLETT K, et al. DISFA: a spontaneous facial action intensity database[J]. IEEE transactions on affective computing, 2013, 4(2): 151–160.

[41] SNEDDON I, MCRORIE M, MCKEOWN G, et al. The Belfast induced natural emotion database[J]. IEEE transactions on affective computing, 2012, 3(1): 32–41.

[42] VALSTAR M, PANTIC M. Induced disgust, happiness and surprise: an addition to the MMI facial expression database[C]//Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect. 2010: 65.

[43] GROSS R, MATTHEWS I, COHN J, et al. Multi-PIE[J]. Image and vision computing, 2010, 28(5): 807–813.

[44] BEN X, REN Y, ZHANG J, et al. Video-based facial micro-expression analysis: a survey of datasets, features and algorithms[J]. IEEE transactions on pattern analysis and machine intelligence, 2021.

[45] QU F, WANG S-J, YAN W-J, et al. CAS (ME)^2: A Database for Spontaneous Macro-Expression and Micro-Expression Spotting and Recognition[J]. IEEE transactions on affective computing, 2017, 9(4): 424–436.

[46] YAN Wenjing, QI Wu, LIU Yongjin, et al. CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces[C]//2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Shanghai, IEEE, 2013: 1−7.

[47] YAN Wenjing, LI Xiaobai, WANG Sujing, et al. CASME II: an improved spontaneous micro-expression database and the baseline evaluation[J]. PLoS one, 2014, 9(1)–e86041.

[48] MO Fan, ZHANG Zhihao, CHEN Tong, et al. MFED: a database for masked facial expression[J]. IEEE access, 9: 96279–96287.

[49] BENITEZ-QUIROZ C F, SRINIVASAN R, MARTINEZ A M. EmotioNet: an accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, IEEE, 2016: 5562−5570.

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

[51] DHALL A, GOECKE R, GHOSH S, et al. From individual to group-level emotion recognition: EmotiW 5.0[C]//ICMI '17: Proceedings of the 19th ACM International Conference on Multimodal Interaction. New York: ACM, 2017: 524−528.

[52] DHALL A, MURTHY O V R, GOECKE R, et al. Video and image based emotion recognition challenges in the wild: EmotiW 2015[C]//ICMI '15: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. New York: ACM, 2015: 423−426.

[53] GUO Hui, ZHANG Xiaohui, LIANG Jun, et al. The dynamic features of lip corners in genuine and posed smiles[J]. Frontiers in psychology, 2018, 9: 202.

[54] EKMAN P, FRIESEN W V. Felt, false, and miserable smiles[J]. Journal of nonverbal behavior, 1982, 6(4):

238–252.

[55] HESS U, KLECK R E. Differentiating emotion elicited and deliberate emotional facial expressions[J]. European journal of social psychology, 1990, 20(5): 369–385.

[56] SCHMIDT K L, BHATTACHARYA S, DENLINGER R. Comparison of deliberate and spontaneous facial movement in smiles and eyebrow raises[J]. Journal of nonverbal behavior, 2009, 33(1): 35–45.

[57] O'REILLY H, PIGAT D, FRIDENSON S, et al. The EU-Emotion Stimulus Set: a validation study[J]. Behavior research methods, 2016, 48(2): 567–576.

[58] ZHALEHPOUR S, ONDER O, AKHTAR Z, et al. BAUM-1: a spontaneous audio-visual face database of affective and mental states[J]. IEEE transactions on affective computing, 2017, 8(3): 300–313.

[59] SOLEYMANI M, GARCIA D, JOU B, et al. A survey of multimodal sentiment analysis[J]. Image and vision computing, 2017, 65: 3–14.

[60] REISENZEIN R, STUDTMANN M, HORSTMANN G. Coherence between emotion and facial expression: evidence from laboratory experiments[J]. Emotion review, 2013, 5(1): 16–23.

[61] ROSENBERG E L, EKMAN P. Coherence between expressive and experiential systems in emotion[J]. Cognition and emotion, 1994, 8(3): 201–229.

[62] QU Fangbing, YAN Wenjing, CHEN Yunsin, et al. "You Should Have Seen the Look on Your Face…": Self-awareness of Facial Expressions[J]. Frontiers in psychology, 2017, 8: 832.

[63] EKMAN P, FRIESEN W, HAGER J. FACS Investigator's Guide (The Manual on CD Rom)[M]. Salt Lake: Network Information Research Corporation, 2002.

[64] BENITEZ-QUIROZ C F, SRINIVASAN R, MARTINEZ A M. Discriminant functional learning of color features for the recognition of facial action units and their intensities[J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 41(12): 2835–2845.

[65] CHU Wensheng, DE LA T, COHN J F. Learning spatial and temporal cues for multi-label facial action unit detection[C]// 017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). Washington, IEEE Computer Society, 2017: 25−32.

[66] WANG Shangfei, PENG Guozhu, CHEN Shiyu, et al. Weakly supervised facial action unit recognition with domain knowledge[J]. IEEE transactions on cybernetics, 2018, 48(11): 3265–3276.

[67] WANG Pengcheng, WANG Zihao, JI Zhilong, et al. TAL EmotioNet challenge 2020 rethinking the model chosen problem in multi-task learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, IEEE, 2020: 1653−1656.

[68] KOSSAIFI J, TZIMIROPOULOS G, TODOROVIC S, et al. AFEW-VA database for valence and arousal estimation in-the-wild[J]. Image and vision computing, 2017, 65: 23–36.

[69] CARROLL J M, RUSSELL J A. Do facial expressions signal specific emotions? Judging emotion from the face in context[J]. Journal of personality and social psychology, 1996, 70(2): 205–218.

[70] DURÁN J I, REISENZEIN R, FERNÁNDEZ-DOLS J M. Coherence Between Emotions and Facial Expressions[M]. [S. l.]: Oxford University Press, 2017.

[71] DURÁN J I, FERNÁNDEZ-DOLS J M. Do emotions result in their predicted facial expressions? A meta-analysis of studies on the co-occurrence of expression and emotion[J]. Emotion (Washington, D C), 2021, 21(7): 1550–1569.

[72] BARRETT L F, ADOLPHS R, MARSELLA S, et al. Emotional expressions reconsidered: challenges to inferring emotion from human facial movements[J]. Psychological science in the public interest: a journal of the American Psychological Society, 2019, 20(1): 1–68.

[73] BARRETT L F. Psychological construction: the Darwinian approach to the science of emotion[J]. Emotion review, 2013, 5(4): 379–389.

[74] ZHANG Xing, YIN Lijun, COHN J F, et al. BP4D-Spontaneous: a high-resolution spontaneous 3D dynamic facial expression database[J]. Image and vision computing, 2014, 32(10): 692–706.

[75] AVIEZER H, TROPE Y, TODOROV A. Body cues, not facial expressions, discriminate between intense positive and negative emotions[J]. Science, 2012, 338(6111): 1225–1229.

[76] BARRETT L F, MESQUITA B, GENDRON M. Context in emotion perception[J]. Current directions in psychological science, 2011, 20(5): 286–290.

[77] AVIEZER H, HASSIN R, BENTIN S, et al. Putting facial expressions back in context[EB/OL]. 2008. [2020–01–01]. http://cel.huji.ac.il/publications/pdfs/Aviezer_et_al_2008_Chapter_in_First_Impressions.pdf

[78] RUSSELL J A. Core affect and the psychological construction of emotion[J]. Psychological review, 2003, 110(1): 145–172.

[79] VALSTAR M, SCHULLER B, SMITH K, et al. AVEC 2014: 3D dimensional affect and depression recognition

challenge[C]//AVEC '14: Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge. New York: ACM, 2014: 3−10.

[80] FIQUER J T, MORENO R A, BRUNONI A R, et al. What is the nonverbal communication of depression? Assessing expressive differences between depressive patients and healthy volunteers during clinical interviews[J]. Journal of affective disorders, 2018, 238: 636–644.

[81] RUSSELL J A, FERNÁNDEZ DOLS J M. The science of facial expression[M]. New York: Oxford University Press, 2017.

[82] CRIVELLI C, FRIDLUND A J. Facial displays are tools for social influence[J]. Trends in cognitive sciences, 2018, 22(5): 388–399.

[83] CAMPBELL R L. Constructive processes: abstraction, generalization, and dialectics[M]//The Cambridge Companion to Piaget. Cambridge: Cambridge University Press, 2009: 150−170.

[84] CLARK A. Whatever next? Predictive brains, situated agents, and the future of cognitive science[J]. The Behavioral and brain sciences, 2013, 36(3): 181–204.

[85] CLARK A. Predicting peace: The end of the representation wars[M]. Open MIND. Frankfurt am Main: MIND Group, 2015.

[86] KWISTHOUT J, BEKKERING H, VAN ROOIJ I. To be precise, the details don't matter: on predictive processing, precision, and level of detail of predictions[J]. Brain and cognition, 2017, 112: 84–91.

作者简介:

颜文靖, 副教授, 主要研究方向为情绪与表情、测谎以及心理健康。主持国家自然科学基金项目和浙江省自然科学基金项目各一项, 获 2018 年度第八届吴文俊人工智能科学技术自然科学奖一等奖。第一作者论文在 Google Scholar 上被引 1000 余次。

蒋柯, 教授, 中国心理学学会理论心理学与心理学史分会委员,浙江省社会心理学会理事。主要研究方向为推理与决策的认知行为特征、人工智能的理论基础与认知逻辑、心灵哲学、进化心理学。主持国家社会科学基金项目、教育部项目, 以及温州市哲学社会科学项目等课题。发表学术论文 50 余篇; 出版专著、译著和教材等共 7 部。

傅小兰, 研究员, 中国心理学会常务理事、原理事长、原秘书长, 国务院学位委员会学科评议组心理学组成员, 《心理学报》主编, 主要研究方向为认知心理学、情绪心理学和说谎心理学。2013 年和 2017 年分别获北京市科学技术奖二等奖, 2018 年获吴文俊人工智能科学技术奖自然科学奖一等奖, 2021 年获得中国电子学会科学技术奖技术发明一等奖。承担和参与科技项目 30 余项, 发表学术论文 380 余篇, 主编著作 10 部和译著 13 部。