

• 研究构想(Conceptual Framework) •

面向实际应用的微表情分析：从数据采集到智能部署*

李婧婷 赵琳 东子朝 王甦菁

(认知科学和心理健康全国重点实验室(中国科学院心理研究所), 北京 100101)
(中国科学院大学心理学系, 北京 100049)

摘要 微表情指个体试图抑制真实情绪时无意识流露的短暂面部动作, 其非侵入性特征在国家安全、公共安全等领域具有重要应用价值。针对实际场景中存在的生态效度不足、运动干扰及数据隐私等问题, 研究基于生理学与行为心理学机制, 构建高生态效度的微表情诱发范式, 开发面周肌电信号辅助编码系统, 并建立多场景动态微表情数据库。通过设计头部运动与口型变化干扰消除模块, 结合自监督学习框架解决小样本识别问题, 同时利用异步联邦学习实现跨场景隐私保护下的模型部署。该研究通过心理学与计算机科学的交叉融合, 提出兼顾理论机制与实际应用的微表情分析框架, 为多领域应用提供技术支持。

关键词 微表情智能分析, 情感计算, 高生态效度, 实际应用

分类号 B842

1 问题提出

情绪波动不仅揭示了人的内心世界, 更是人类复杂心理状态的直观体现。在众多情绪线索中, 面部表情被认为是最直观的情绪窗口。然而, 成语“强颜欢笑”和“笑里藏刀”提醒我们, 表情常被刻意控制, 遮盖真实的情感流露。微表情作为短暂且细微的面部动作, 往往在无意识或无法控制时流露, 成为揭示真实情绪的关键线索。因此, 微表情在公共安全、国家安全等领域展现出了巨大的应用前景。

微表情的研究和应用依赖扎实的理论基础。早在 1859 年, 查尔斯·达尔文(Charles Darwin) 在《人与动物的情感表达》中指出, 某些面部表情即使付出巨大努力也难以抑制(Darwin, 1872)。神经生理学研究发现, 自主表情和非自主表情分别受锥体运动系统和锥体外运动系统控制(Rinn, 1984)。1969 年, Paul Ekman 教授认为自主与非自主表情的冲突可能导致微表情的产生, 这种表情

可能源于对初步表情冲动的自我抑制的泄露, 或在正常表情过程中被意外截断(Ekman & Friesen, 1969)。因此, 微表情能够体现个体的真实情绪, 并且具有转瞬即逝、弱不可见、局部发生的特点。

人工检测和识别微表情极具挑战性, 即使是经过专业培训的专家, 其识别准确率也仅略高于随机猜测(Frank et al., 2009)。这一现象揭示了人工识别微表情的局限性, 尤其是对人力和时间的高需求, 凸显了提升智能微表情分析技术的紧迫性。当前微表情智能分析面临多重挑战: 首先, 不同情境下微表情的生成机制尚未完全明确, 制约了基础理论的发展; 其次, 具有高生态效度的微表情样本稀缺, 限制了深度学习模型的训练与应用效果; 同时, 人工标注的高难度导致样本规模难以满足算法需求; 此外, 头部姿态变化和口型动作等干扰因素显著影响分析的准确性; 最后, 数据隐私问题阻碍了跨场景数据的有效整合与共享。因此, 构建高效的数据采集体系并开发适配的智能分析技术, 是突破微表情研究瓶颈、推动实际应用的关键路径。

2 研究现状

2.1 微表情智能分析研究现状

如图 1 所示, 自 2009 年以来, 微表情智能分

收稿日期: 2025-03-06

* 国家自然科学基金项目(62476269, 62276252), 中国科学院青年创新促进会。

通信作者: 王甦菁, E-mail: wangsujing@psych.ac.cn

析得到了广泛关注,这一进展得益于一系列高质量微表情数据库的发布。然而,目前可用的微表情数据仅约10,000个,数据量较少且生态效度不足,限制了深度学习技术在微表情分析中的应用发展。此外,相较微表情识别(对已知微表情进行分类),微表情检测(在视频中发现并定位微表情的起始与终止时刻)的研究仍较为稀缺,且面临严峻挑战。接下来,本节将从微表情数据库、检测与微表情识别三个方面概述并分析微表情智能分析的研究现状。

2.1.1 微表情数据库

生态效度是决定微表情数据库是否适用于真实情景下微表情分析的关键因素,并与微表情诱发范式直接相关。本研究将已发表的微表情数据库按诱发方法分为三代,生态效度逐渐增加。

第一代微表情数据库:在微表情分析早期阶段,微表情样本采集是录制表演者在观察标准表情样本后试图摆出的转瞬即逝的面部表情。USF-HD (Shreve et al., 2011)和Polikovsky (Polikovsky et al., 2009)数据库是两个摆拍微表情数据库。然而,微表情被认为是自发的,难以伪装,模仿表情仅包括外部表现,缺乏真实性。因此,需要更自然的微表情诱发方式,以提升数据库的生态效度。

第二代微表情数据库:为解决上述问题,心理学研究人员试图用情绪刺激诱发微表情。例如,“面无表情”范式要求被试在观看情绪刺激材料时保持面无表情的状态,这可以诱发自发微表情,同时避免不想管面部动作。CASME系列((Li et al., 2023; Qu et al., 2018; Yan et al., 2013; Yan et al.,

2014)、SMIC (Li et al., 2013)、SAMM (Davison et al., 2016)、MMEW (Ben et al., 2022)、4DME (Li, Cheng, et al., 2022)、DFME (Zhao et al., 2024)等数据库就是通过这种方法收集的。尽管这些微表情样本自发性较强,且在控制良好的实验环境下收集,但它们仍与现实情境存在差距。因此,如果要将微表情的检测和识别应用于实际场景,还需要进一步提高样本的生态效度。

第三代微表情数据库:为解决上述限制,从更高生态效度中收集微表情检测样本是推动微表情分析研究的必然选择。Husak 等人发布了自然环境下的数据库 MEVIEW (Husák et al., 2017, February),总共有40个标注微表情,样本来自扑克游戏和电视采访视频。然而,这些样本也存在不可控因素,如镜头焦距、头部移动等。目前的微表情分析研究仍需在良好实验室环境中收集。为此,CAS(ME)³中发布了通过模拟犯罪场景收集的微表情样本,在提高生态效度的同时,消除了不可控因素(Li, Dong, et al., 2023)。然而,CAS (ME)³中高生态效度部分仅包含31名被试的166个样本,并且交互情境局限于模拟审讯问答,样本量与应用场景仍有限。因此,为了促进微表情分析的实际应用,未来需要构建更多符合真实场景的诱发范式,并采集更大规模的微表情数据。

2.1.2 微表情检测

微表情检测是在长视频中准确定位微小短暂的微表情片段,主要有两种思路:一是比较帧间特征差,二是通过机器学习对微表情帧和非微表情帧分类。

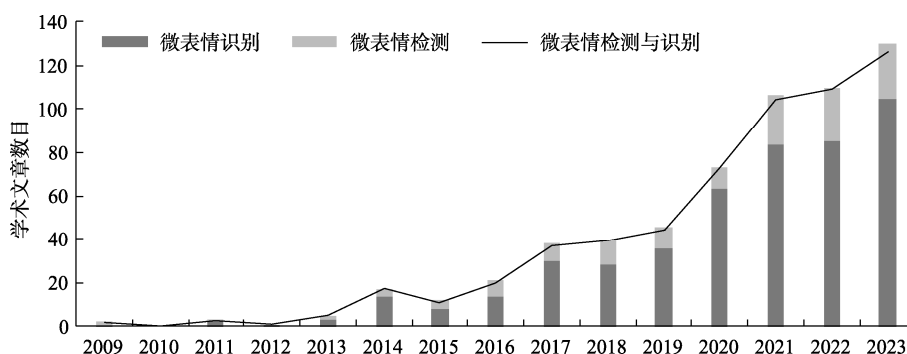


图1 基于计算机科学的微表情研究趋势

首先, 帧差法是计算时间窗口中的特征差异, 通过阈值筛选出动作幅度接近微表情的脸部运动。常用特征包括: 局部二值模式(LBP) (Moilanen et al., 2014, August)、光流特征 MDMO (Zhang et al., 2020)、3D-HOG (Davison et al., 2018, May)、局部时空模式(Li, Soladié, & Séguier, 2023)等。其次, 为了增强检测方法区分微表情与其他面部运动的能力, 基于深度学习的微表情检测方法逐渐成为焦点, 例如长视频中多尺度检测微表情片段的卷积神经网络(MESNet) (Wang et al., 2021), 局部双线卷积神经网络(Pan et al., 2022), LGSNet (Yu et al., 2024), 基于 3DCNN 利用时空定向参考帧的检测方法(Yap et al., 2022, October), 动作单元感知图卷积网络(Yin et al., 2023), 融合 VideoMAE 的检测模型等(Xu et al., 2023, October)。然而, 当前深度学习方法仍受到小样本的限制, 无法应用到实际场景中。

同时, 微表情国际挑战大赛(MEGC)设置的长视频中的微表情检测任务(Davison et al., 2023, October; Li et al., 2021, October; Li, Yap, et al., 2022, October; Li et al., 2020; See et al., 2019), 极大地促进了微表情检测任务的学术发展。比较历年前三名的方法, 发现传统帧差法通过有效的预处理、滤波和适当的阈值设置, 能去除轻微头部运动和外部干扰, 实现较好的检测性能。然而, 在真实场景中, 头部姿势、光照变化等干扰因素无法通过简单的对齐、滤波等消除, 亟需构建有效的干扰消除模块, 实现复杂长视频中的微表情检测。

2.1.3 微表情识别

微表情识别是指根据特定的情绪类型, 对微表情片段进行分类。主要有两类方法: 手工特征法和深度学习方法。常见的手工特征包括: 基于 LBP-TOP、HOG 和光流的方法等。例如李晓白等人采用了 HOG 的变体(Li et al., 2018), 随后黄晓华等人提出了基于积分投影的 LBP-TOP 以增强微表情分类间差异(Huang et al., 2019); 王甦菁等人提出在彩色空间中分别提取 LBP-TOP 特征(Wang et al., 2015), 刘永进等人提出了基于主方向平均光流的微表情识别方法(Liu et al., 2016)。但由于微表情具有持续时间短、强度低的特性, 这些特征对微表情的表征性有待提升。

近年, 结合深度学习的微表情识别成为了主

要趋势, 并且识别率在不断提升。贡晔等人提出了解耦面部动作信息和 ID 信息的微表情识别框架(Xie et al., 2022); 洪晓鹏等人提出了基于 Transformer 时空特征提取的微表情识别方法(Zhang et al., 2022); 宗源等人利用基于稀疏 Transformer 网络提取微表情相关特征(Zhu et al., 2022); 毛启荣等人通过区域启发式关系推理网络实现部分遮挡下的客观基于类别的微表情识别(Mao et al., 2022); 王上飞等人将在宏表情中训练得到的网络应用在微表情分类任务上, 从而提升识别性能(Xia et al., 2020, October); 陈恩红等人提出基于图神经网络的微表情识别方法(Zhang, Wang et al., 2023)。

基于深度学习的微表情分析方法面临小样本问题, 主要有三个原因: 一是小样本问题会造成模型的过拟合问题; 二是迁移学习相关方法对于微表情分析性能提升的效果有限; 三是微表情分析的样本数和网络参数都受到样本数的限制。因此, 构建一个大规模的微表情数据库对推动深度学习在微表情智能分析中的发展至关重要。

2.2 相关技术研究进展

2.2.1 基于肌电的面部表情识别研究现状

在过去几十年中, 精确的自动面部表情标注一直是情感计算领域的难题, 基于计算机视觉的标注是最常见的方法。近年来, 可穿戴肌电设备逐步成为替代方案。例如, Ang 等人使用 3 通道面部 EMG 信号识别生气、开心和悲伤表情(Ang et al., 2004, November)。Chen 等人设计了一款头戴设备, 识别眉毛运动相关的 5 种表情(Chen et al., 2015)。Sato 等人也设计可穿戴面部肌电采集设备来测量情绪的效价(Sato et al., 2021)。与传统的使用 8 组电极的面部肌电测量相比, Schultz 等人只用 4 组电极(前额皱眉肌、额肌、颧骨大肌), 而表情的识别率只减少了不到 5% (Schultz & Pruzinec, 2010)。Hamed 等人通过 3 组电极, 分别放在额肌和颞肌上, 10 种面部动作的区分准确率达 87% (Hamed et al., 2013)。Lu 等人通过基于面部动作单元的面部肌电测量首次刻画了微表情的运动强度(Lu et al., 2022, October)。

肌电信号在面部动作识别中已得到广泛应用且识别率较高。然而, 传统实验中将电极直接放置在面部肌肉上, 虽然确保了信号采集质量, 却带来了不自然的感觉, 且电极的重量可能影响面

部运动。为解决此问题, Gruebler 等人采用远端电极放置方式进行面部肌肉信号采集, 结果显示对微笑和皱眉的识别效果良好(Gruebler & Suzuki, 2014)。Monica 等人也通过远端电极和计算机视觉方法, 成功实现了微笑动作单元的检测(Perusquia-Hernández et al., 2021, December)。

由于微表情的微弱和短暂特性, 视觉层面的标注非常困难。因此, 通过设计面周肌电信号捕捉系统, 可以有效提升微表情编码的效率。

2.2.2 联邦学习的研究现状

在众多资源受限的边缘设备中, 存在大量的数据集。联邦学习技术能够利用这些分散的数据来训练全局模型, 从而提高数据表示精度和模型适用性(Konečný et al., 2017)。同时在保证数据传输合规性的前提下, 满足隐私保护需求(Bonawitz et al., 2017)。联邦学习系统由中央服务器和子节点(边缘设备)组成, 子节点在私有数据上本地训练模型, 服务器聚合各子节点的模型并更新全局模型。联邦学习已在多个领域中得到了应用。如, 谷歌在安卓用户的键盘应用中实施联邦学习, 推动推荐系统的发展(Mansour et al., 2020); Cohen 等人指出联邦学习可以在保证患者隐私的同时, 解决数据量小和标签不足的问题(Price & Cohen, 2019)。

联邦学习的同步方法常因滞后节点而导致

同步效率下降。为解决此问题, 异步训练广泛应用于传统的分布式随机梯度下降, 以应对滞后节点和异构延迟(Lian et al., 2018, July)。例如, FedMDS 框架(Zhang, Liu, et al., 2023), FedAAM (Lu et al., 2024)。

随着异步联邦技术的成熟, 微表情应用场景中数据隐私和传输困难问题得到了有效解决。通过异步联邦学习框架, 可以在考虑节点样本差异和传输延迟的情况下, 实现面向多应用场景微表情智能分析模型的部署。

2.3 本文贡献

微表情作为个体试图隐藏真实情绪时流露的面部线索, 广泛应用于医疗、公共安全等领域。然而, 微表情智能分析的应用面临诸多挑战, 如小样本问题、复杂场景下模型性能不足, 以及数据隐私与传输限制等。为推动微表情分析在特定场景的实践应用(如图 2 所示), 研究围绕核心挑战构建系统性解决方案: 针对高生态效度样本稀缺问题, 基于情绪-行为心理学原理设计多模态真实交互情境的微表情诱发范式; 针对人工标注效率瓶颈, 研发面周肌电信号驱动的自动化编码辅助系统以提升标注精度; 为克服头部运动与口型变化对检测的干扰, 开发基于运动补偿机制的干扰消除模块; 针对深度学习模型表征能力不足, 构建垂直领域适配的自监督学习架构以增强微

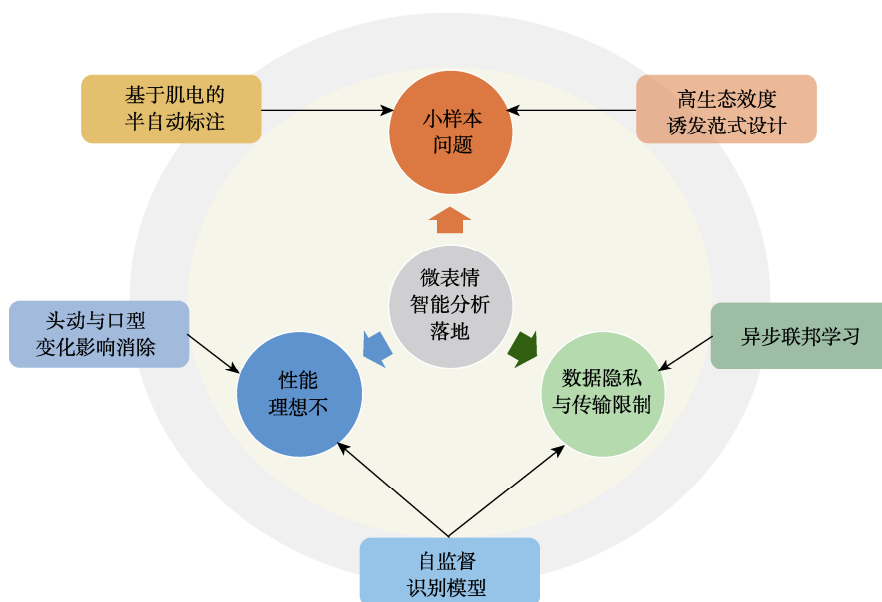


图2 研究难点及对应思路

表情特征提取的鲁棒性；针对数据隐私与传输限制，创新性融合异步联邦学习框架与离线介质传输协议，实现安全合规的模型迭代更新。基于心理学、计算机科学等多学科的交叉融合，本研究旨在建立具有扎实理论基础的微表情智能分析框架，推动微表情技术的性能提升，并促进非接触式心理状态监测的实际应用。

3 研究构想

如图 3 所示，面对微表情智能分析在特定应用场景中遇到的核心挑战，本研究首先基于生理与行为心理学研究，设计基于面周肌电的微表情辅助编码方法，构建不同场景下多模态、高生态效度数据库；在数据基础上，本研究着力解决在特定应用场景中微表情智能分析的技术难题，并基于联邦学习实现不同场景下的微表情智能分析部署。通过将心理学、计算机科学等多个学科的知识和技术相结合，本研究旨在建立一个既有理论基础又能满足实际应用需求的微表情智能分析框架。

3.1 基础理论和模型的研究构想

3.1.1 特定应用场景下微表情机制研究

本研究针对高生态效度微表情数据稀缺和微表情行为及生理机制不明的挑战，依托心理学研究基础，开展交互情境下的微表情机制研究。首先，设计有效诱发高生态效度微表情的实验范式，采集异构场景下的微表情数据。随后，设计基于面周肌电的辅助编码系统，克服人工编码的时间

与人力消耗问题，构建大规模高生态效度微表情数据库。并基于采集数据，研究微表情在不同交互情境中的行为与生理变化规律。

3.1.2 特定场景下微表情智能分析基础模型

头部姿态和口型变化会干扰对于真实、复杂场景中获取的视频数据的检测，导致微表情的时序和空间特征扭曲，进而影响检测的准确性和召回率。本研究将开发高效的可插拔处理算法，准确辨别并剔除由头部运动和说话引起的干扰，确保微表情特征的准确提取。通过大规模数据采集、干扰消除模块和成熟的微表情检测算法，为特定场景的应用提供高效且成本较低的解决方案。

3.1.3 基于大模型垂直领域的微表情自监督学习模型

为应对微表情分析中的小样本问题，本研究提出基于大模型垂直领域的微表情自监督学习模型。借助视觉大模型的发展，本研究基于垂直领域构建了两级下游任务。首先，基于大量未标注的表情数据，学习面部动作单元强度的变化模式；然后，使用少量已标注数据进行微表情识别训练，克服小样本限制。通过大模型的图像特征挖掘能力，逐级学习面部动作变化模式与微表情特征，提升微表情识别性能。

3.1.4 异步微表情联邦智能分析系统

鉴于微表情智能分析应用场景中的数据安全问题，本研究计划采用异步联邦学习的部署方法。在不直接共享原始数据的情况下，利用来自

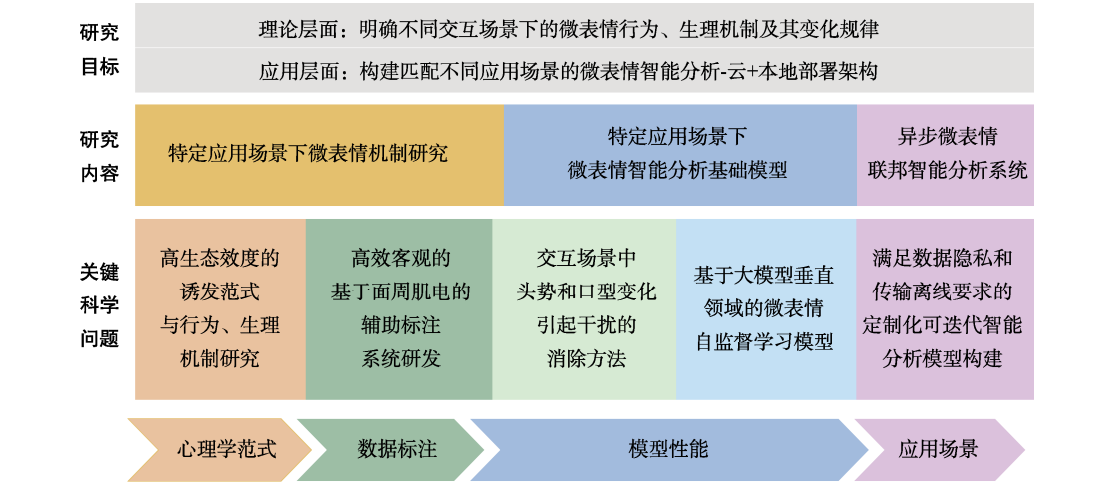


图 3 研究框架

不同来源的数据进行模型训练,提升模型的泛化能力和准确性。联邦学习保障了数据的隐私和安全,同时使模型能从多源数据中学习并提高性能。为了增强模型对小样本和特定场景的适应能力,子节点层面将结合自监督学习和强化学习。此外,该体系结构不仅实现了高效的模型训练和更新,还能够适应动态变化的应用环境,为微表情智能分析技术的发展和其在敏感领域的应用落地提供了安全保障。

3.2 关键技术的研究构想

技术路线总体思路见图4。

3.2.1 特定应用场景下微表情机制研究

数据是智能分析系统的核心。鉴于高生态效度的微表情样本稀缺,本研究基于心理学理论设计多个实际应用场景的微表情诱发范式,通过多种设备采集行为和生理数据,并采用肌电辅助编码系统,高效构建大规模微表情数据库。通过分析大量样本,研究面部微表情、生理信号与心理状态的关系,深入理解隐藏情绪下的行为模式,为微表情智能分析系统提供理论和数据支持。

1) 高生态效度心理学范式设计

正如第2节中的描述,相较于现有微表情数据库普遍采用的单向被动式诱发范式(如通过观看视频刺激对微表情进行诱发),本研究创新性地构建了多模态交互的微表情诱发框架:一方面通过经典测谎范式建立情景压力,另一方面引入通过角色扮演的人际交互范式构建自然社交场景。

这种双重范式的结合应用可以帮助我们获得以下进展:1)实验刺激除了被动接受以外,还增加了动态交互,这使得微表情的诱发产生机制更加符合人们人际交往过程中的真实心理;2)研究通过将受试置于具有现实映射关系的任务环境中,有效提升了情绪诱发的心理真实性。

综上所述,通过4种不同生态效度的实验范式,提升诱发微表情的生态效度,从多角度和情境探讨微表情的多样性及其复杂机制,进一步挖掘微表情在真实场景中的应用潜力,提升自动化分析的准确性。此外,基于多种实验范式的多维度框架,有助于深入理解微表情与情绪、认知过程之间的复杂关系,为其在特定场景中的应用提供理论支撑。

a) 视频诱发范式

此范式为传统诱发范式,用于微表情数据库的基准参考。具体而言,使用具有高情绪效价的视频片段作为诱发表情的刺激材料,包括7种情绪(高兴、生气、悲伤、恐惧、生气、惊讶和中性)。每种情绪2~3个视频,每个视频时长1~3分钟。这些视频均为CASME数据库系列中所使用的诱发材料。

b) 互动情景下的故意欺骗范式

实验采用面对面提问的方式,对问题进行“是/否”的回答,每2名被试为一组,分别扮演“说谎者”和“测谎者”,一轮提问后角色互换。实验开始前,被试需根据实际情况填写一份包含基线问题、事实问题和喜好问题的问卷。

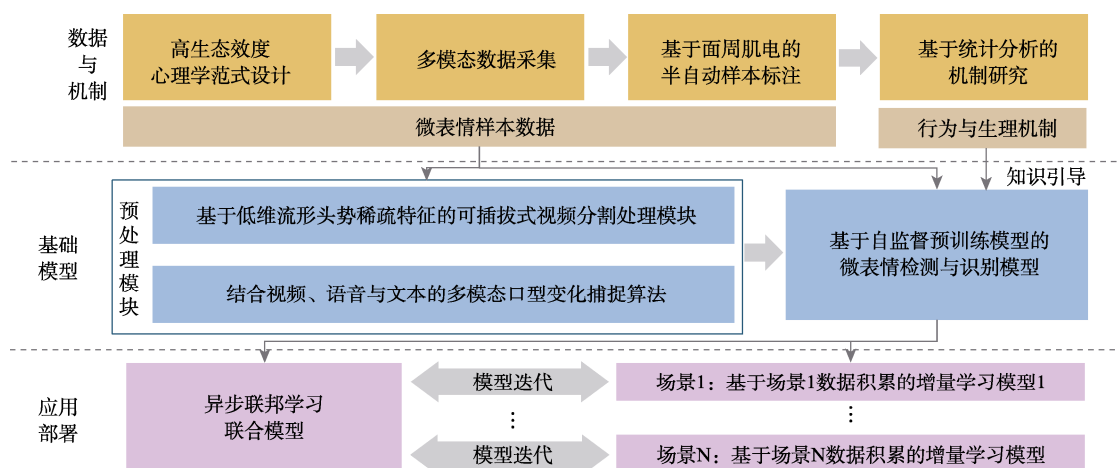


图4 技术路线总体思路

c) 主动说谎范式

在该范式中, 主试会设定一种演讲主题, 例如“我的童年”, 要求被试坐在椅子上根据主题进行自由演讲, 演讲的内容必须全部说谎, 同时告知被试, 实验结束后有他人观看此演讲视频, 如果能让对方相信演讲内容是真实的, 将会获得金钱奖励。

d) 模拟犯罪范式

本研究布置了模拟犯罪室和模拟审讯室。在模拟犯罪室中, 被试可以打开箱子找到目标物品。根据行为不同, 将被试分为犯罪者、知情者和无辜者。在模拟审讯室中, 被试被告知实验室发生盗窃案, 并成为重点嫌疑人。他需要与新研制的计算机测谎算法对抗, 成功证明自己清白可获得高额奖励, 否则将扣除部分被试费。

2) 多模态数据采集

在科学研究特别是情感计算和微表情分析领域中, 数据采集的精确性至关重要, 其质量直接影响研究的准确性和可靠性。为了尽可能地捕捉到细微的面部表情变化和相关的生理信号, 如图 5 所示, 本研究采用了高性能设备组合进行数据的采集和分析。

本研究使用 4K、120fps 高清相机进行视频录制。4K 分辨率确保面部细节清晰呈现, 为后续智能分析提供丰富面部空间信息。120fps 高帧率保证了面部动作的连续性, 支持后续的动态特征提取。同时, 使用深度相机获取场景的深度信息, 有效排除头部或身体移动造成的视觉干扰, 提升微表情检测和识别的准确性, 特别在复杂背景或光照条件下。

此外, 使用多导生理仪记录参与者的生理变化, 包括呼吸、脉搏、心率、皮电。微表情的出现往往伴随着生理反应, 如心跳加速、呼吸变化等, 这些数据为微表情研究提供了客观指标。通过综合分析面部微表情和生理数据, 可以全面理解情绪状态, 深入探索微表情的发生机制和生理基础。

本研究还将设置面部温度相机。面部温度的变化可以反映血液流动的变化, 间接指示情绪被诱发时身体的生理变化。面部温度和微表情的无接触采集方式能为测谎领域提供新的应用方法, 进一步拓展微表情分析在心理学和生理学中的研究边界。

3) 基于面周肌电的辅助编码系统

对于无论是以数据为驱动的深度学习方法还是基于统计学的行为机制分析, 面部微表情的数据库样本量明显不足。在数据库构建过程中, 微表情标注是一个需要精确完成的工作, 往往需要受过训练的标注人员对视频逐帧观察确定面部动作的起始帧、峰值帧和结束帧, 这种方法非常耗费时间和人力。因此, 本研究提出使用面部肌电信号辅助标注人员对表情数据库实现半自动化标注。

为了实现面周肌电信号的采集, 本研究设计一款多通道灵活配置的肌电信号采集系统。直接在面部肌肉活动区域粘贴电极和相关导线, 可能会对被试产生心理或物理上的负担, 从而影响其情绪的自然流露, 干扰微表情的真实性。并且, 电极及其导线会遮挡面部关键区域, 这对于后续需要利用计算机视觉技术进行面部表情(尤其是



图 5 数据采集场景

微表情)视频分析的研究而言,会严重影响模型对相关面部细节特征的提取与识别效果。因此,我们创新性地提出了面周肌电采集方案。如图6所示,通过将电极策略性地布置在面部周围区域,我们旨在完全避免遮挡面部表情,既干扰面部肌肉的自然运动,又能利用生物电信号的体积传导效应,客观地捕捉到由源头肌肉活动产生的肌电信号。即使电极未直接贴附于源肌肉,相关的肌肉活动信号仍能通过组织传导被周围电极检测到。我们前期通过与商用肌电系统的对比,证明了该自研肌电系统的可靠性(Zhang, Huang et al, 2023)。

关于信号采集,正如上一段中的描述,我们并非直接测量目标肌肉活动,而是捕捉源肌肉电信号传递到临近面周区域的“串扰”信号,即源肌肉电活动传播至邻近非目标肌肉区域被检测到的现象。为了有效处理这些混合信号,我们采用了分组单级配置方案(如图7所示)。具体而言,依据面部肌肉解剖结构及与面部动作单元的对应关系,我们将面部分为6个对称区域(左右各三)。由于现有研究结果显示,人脸的三庭和五眼的基本比例关系以及鼻子、嘴巴和眼睛的位置结构关系不会因为个体不同或表情不同而发生改变(Zhang, 2022)。

在此基础上,结合面部神经支配的对称性以及情绪的动态表达特征,我们对面部区域进行了划分。在划分时,我们首先依据垂直顺序对面部区域进行了定位,将其首先分为了上区域-额区(覆盖额肌与皱眉肌,主要是眉眼区域的运动),中区域-眼颧区(覆盖眼轮匝肌与颧大肌等肌肉群,主要是表达一些核心情绪,如快乐),下区域-口颌区(覆盖降嘴角肌与颈阔肌等肌肉,主要是用来表达个体的社交意图);之后,结合左右半脸在表达情绪时的非对称激活情况(如,



图6 面周肌电采集实例

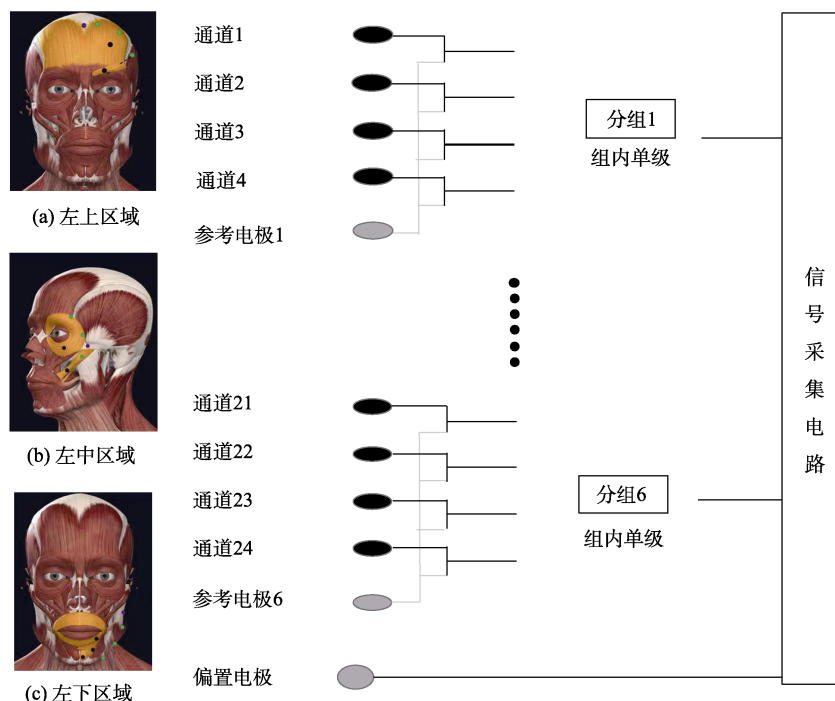


图7 面周肌电分组设计

在进行情绪表达时,左脸的情绪强度较右脸的强度来说更加强烈);最后,在产生情绪时面部区域的启动顺序也是不同的,从前到后依次为:上区域、中区域、下区域,这也符合“认知-情绪-行为”的神经传导路径。综上所述,我们将面部划分为6个区域:左上(包含左额肌、左皱眉肌)、左中(包含左眼轮匝肌、左颧大肌)、左下(包含左降口角肌、左颈阔肌)、右上(包含右额肌、右皱眉肌)、右中(包含右眼轮匝肌、右颧大肌)、右下(包含右降下唇肌、右颊肌)。

在每个区域内,如图7所示采用单级电极配置(4个记录电极加1个参考电极),旨在最大化地采集该区域内的局部“串扰”信号,这些信号主要源自该区域下的目标AU相关肌肉(如皱眉肌、颧大肌等)。通过这种分组设计和区间距离的考量,我们将混合信号源主要限制在目标区域内,从而显著减弱了来自其他面部区域肌肉活动的干扰。这样,每组获得的是所在区域的“串扰”,使得采集到的肌电信号更能反映目标区域的活动状态,提高了信号的相对准确性和区域特异性。然而,远端测量肌电图意味着在面部产生动作时,一个电极通道所采集到的信号可能是多个肌肉源产生的信号不同程度的结合。我们采用独立成分分析(ICA)的信号处理算法,从采集到的混合肌电信号中分离出与特定面部动作单元相关的独立信号源。随后,对肌电信号起始和终止时刻进行检测。具体而言,表面肌电信号(sEMG)是大量运动单元动作电位(MUAPs)的叠加,普遍认为实际的sEMG信号幅值分布是非高斯的,尤其是在肌肉活动期间(Sato & Kochiyama, 2023),这满足了ICA的一个关键假设。但是,面部肌肉密集且存在协同运动,单个电极记录的信号确实可能包含来自多个肌肉源的活动以及串扰,这使得源信号完全独立的假设在实践中难以完美满足。然而,这也恰恰是应用ICA处理面部肌电信号的一个重要动机。正如Sato等人指出,面部肌电信号容易受到串扰和非表情动作(如说话、咀嚼)的干扰,ICA能够分离这些混合信号,从而减少串扰影响,提取更纯净的目标肌肉活动信号(Sato & Kochiyama, 2023)。我们提出的结合ICA的多通道区域采集设计,正是为了在信号混合的情况下,尽可能地分离出主导的、相对独立的信号源,以提高区域活动检测的准确性。虽然不能保证分离

出绝对生理意义上的独立肌肉单元信号,但ICA在此作为一种有效的信号处理工具,用于提高目标信号的信噪比和区分度。此外需要强调的是,本文使用ICA的主要目标是精确地定位面部肌肉动作发生的区域以及动作起止的时间,而非精确量化每个独立肌肉单元(或ICA分离出的成分)的运动强度。

在本研究中,面周肌电系统的核心价值在于为编码人员提供客观、实时的面部肌肉活动信息,作为一种高效的辅助工具。鉴于微表情具有持续时间短、幅度微弱的特点,人眼往往难以准确捕捉和编码,而肌电信号能够精确地记录下这些细微的肌肉活动。因此,借助同步记录的肌电数据,编码人员能够获得关于特定AU(面部动作单元)是否激活、激活强度与持续时间等方面的客观生理指标,从而显著提升微表情标注的速度与准确度。需要特别澄清的是,尽管视频采集过程中被试佩戴了面周肌电装置,但由于我们采用的是面部周围的电极布局,并未遮挡与表情表达直接相关的核心面部区域。因此,这些带有精确时间戳的同步视频样本本身亦可直接用于面部微表情识别模型的训练。由此可见,面周肌电技术的应用具有双重优势:一方面,它极大地提升了训练数据集标签的质量和标注效率;另一方面,同步采集的视频样本有效扩充了微表情数据库,有助于缓解该领域长期存在的小样本问题。

综上所述,我们提出的面周肌电采集方案旨在平衡客观生理信号测量的需求与避免干扰被试自然表达及后续视觉分析的需求,通过特定的电极配置和信号处理策略来确保信号质量,并将其定位为提升微表情人工编码效率与准确性的有效辅助手段,进而为后续高质量数据集的构建和智能识别模型的研究奠定基础。

4) 基于统计分析的微表情机制研究

在对标注的微表情样本进行数据清洗与整理后,首先对数据进行描述性统计分析,包括微表情频率与持续时间,以及生理数据特征如心率、皮肤电活动等的平均值、标准差等。然后,进行范式间的比较分析,即被动说谎、主动表达(说谎)、模拟犯罪三种范式下微表情的差异,基于 t 检验对不同范式下的数据进行比较,计算 p 值以确定结果的显著性。基于统计测试的结果,对比较进行解释,讨论不同范式下的差异可能的原因

和意义。此外,基于皮尔逊相关系数探索微表情频率、强度与生理反应(如心率变化)之间的关联。例如,计算微表情频率(X)和心率(Y)之间的皮尔逊相关系数,并根据得到的 r 值来评估它们之间的相关强度。

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

其中 X_i 和 Y_i 是单个观测值, $\bar{X}$ 和 $\bar{Y}$ 是各自的平均值。

3.2.2 特定场景下微表情智能分析基础模型

1) 基于低维流形头势稀疏特征的视频分割技术

交互过程中的头动对微表情检测影响严重。在预处理过程中,以往的处理流程往往是基于面部刚性结构,即鼻子区域,去除头部动作。然而,这种方法只能处理视频中头部动作幅度非常小的情况。考虑到个体在交互过程中,特别是在微表情流露的过程中,很少会发生头部姿态剧烈变化的情况。如图 8 所示,本研究提出基于低维流形头势稀疏特征的视频分割方法,将复杂长视频分割成头势变化平缓的短视频片段进行微表情检测。

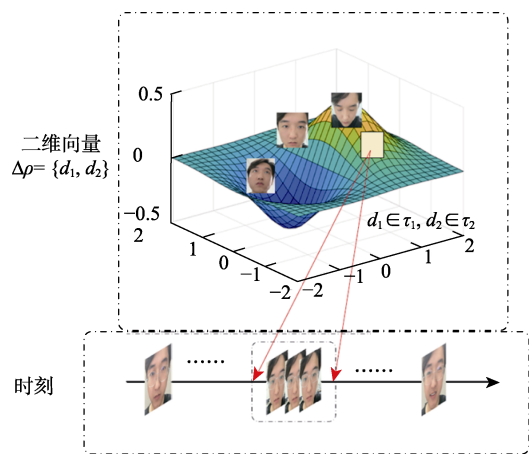


图 8 基于头势稀疏流形的时域分割

具体而言,首先通过采用计算机视觉工具提取人物头部姿态信息,基于低维流形提取头势稀疏特征。然后,通过基于局部线性嵌入的流形学习,构建数据点之间的邻接图,保持数据点的局部或全局几何结构,最终获得数据在低维空间中

的嵌入表示,揭示头部运动数据中的内在几何结构,同时保留头部运动的关键特性。在得到低维表示后,通过稀疏编码技术选取最有代表性的特征,这些特征能够以最少的元素准确地重构头部运动的平抑和旋转动态。

2) 结合视频、语音与文本的多模态口型变化捕捉算法

除了头动对微表情检测的影响,因说话导致的口型变化也是很重要的干扰因素。仅仅凭借视觉信号的处理,可能会导致一些非说话的口型动作被消除,进而损失对应的情绪表达。因此,如图 9 所示,本研究提出基于文本、语音与视觉信号的多通道算法,不仅能够捕捉口型变化,也能基于语言信息进行情感背景信息提取。

首先利用 Wav2Vec 网络从视频的音频轨道,将原始音频波形提取文本信息。随后,使用音素对齐技术将提取的文本与视频中的对话进行时间对齐。基于对齐结果,对视频中对话的片段进行分割。一方面,由于微表情的产生往往是在交互的过程中,所以通过对该阶段的视频和语音信号进行特征提取,学习视频中体现的情绪线索。另一方面,通过声纹特征注册与匹配,区分主试(提问者)和被试(微表情研究对象),选择视频说话时的视频片段。在确定说话人的说话的時刻后,通过面部预处理,提取下面部特征并基于滑动窗口进行特征差的比较,基于特定的策略对口型变化进行提取。在后续的微表情检测过程中,将对存在口型变化的片段进行特殊处理,即只关注其上面部。此外,基于文本特征、语音特征与视频特征,通过特征融合,实现情绪背景信息的提取。

3) 基于大模型垂直领域的微表情自监督学习模型

微表情识别的基础是模型能够很好地学习面部的动作特征。但是微表情本身微弱的动作特性和样本量少,导致具有鲁棒性的微表情识别算法难以实现。同时,视觉大模型具备了极强的图像特征学习能力。如图 10 所示,借助垂域大模型,本研究基于样本规模设置两级下游任务。首先,借助规模较大的表情数据来发展大模型通过自监督学习任务挖掘面部表情特征的能力;然后,在较小规模的微表情数据集上进行进一步的细化调整。通过从大型模型到具体细分领域的层层递进,提高微表情识别的效能。

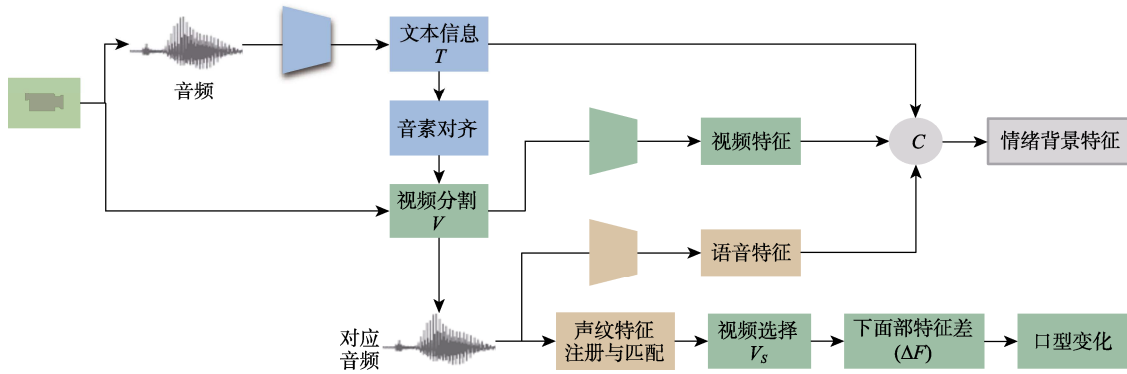


图 9 口型变化捕捉算法流程图

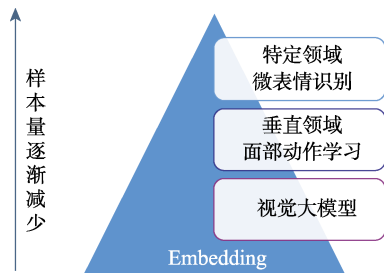


图 10 垂域视觉大模型的两级任务

首先, 针对垂直领域的面部动作学习, 本研究通过面部动作单元(AU)强度来构建样本对。具体而言, 大量表情视频数据库中的视频记录了丰富的面部表情。在这些视频中挑选合适的样本作为候选集, 得到更适合自监督网络学习脸部动作特征的样本。首先使用[OpenFace]工具包来检测 AU 的激活状态信息。随后, 为了在对比过程中排除个体生物特征信息的干扰, 构建起始帧-峰值帧(onset-apex)的样本对, 使得网络聚焦面部动作。因此, 该步骤包含两个阶段, 一个部分是动作幅度最大的峰值帧(apex)的筛选, 一个是对应 apex 的起始帧(onset)的推断。AU 强度大于设定阈值的帧, 被设定为当前样本对的峰值帧。根据约定俗成的微表情的持续时间定义, 即小于 500 ms。同时, 动作幅度最大的帧大部分分布在微表情片段的中间区域。因此, 在得到峰值帧后, 根据视频帧率, 此处以 30fps 为例, 往前推 8~9 帧作为起始帧候选集。同样的, 对于非 RGB 衍生的模态, 例如深度模态, 我们也在深度模态序列中根据时间保留相同的片段。

随后, 通过帧差异估计突出显示面部运动显著发生的区域, 提供微表情的时序变化的表征。具体而言, F_i^O 和 F_i^A 分别表示第 i 个候选片段中在

RGB 视图下的起始帧和峰值帧, 其中 k 代表有标签样本的数量。通过将峰值帧减去起始帧, 我们得到了帧差 $D_i = F_i^A - F_i^O$ 。此外, 我们将这种方法扩展到微表情的其他模态, 例如深度和灰度。对于每种模态中的任意起始帧和峰值帧, 分别表示为 E_i^A 和 E_i^O , 我们计算帧差异为 $D_i^E = E_i^A - E_i^O$ 。

本研究提出多视角对比学习(Multi-view Contrastive Learning, MvCL)对来自两个不同分布的样本进行区分。具体地, 针对每一段候选片段, 提供一段对应的多模态信息片段(附加模态), 如灰度图或深度图。两种不同模态间存在相互关联的“互信息”。因此, 同一种面部表情的 RGB 图像和附加模态分别经过特征提取器后得到的特征表示在高维特征空间中的距离应当彼此靠近, 同时推远与其它表情特征表示之间的距离。图 11 展示了使用 RGB 和附加模态的自监督对比学习过程。

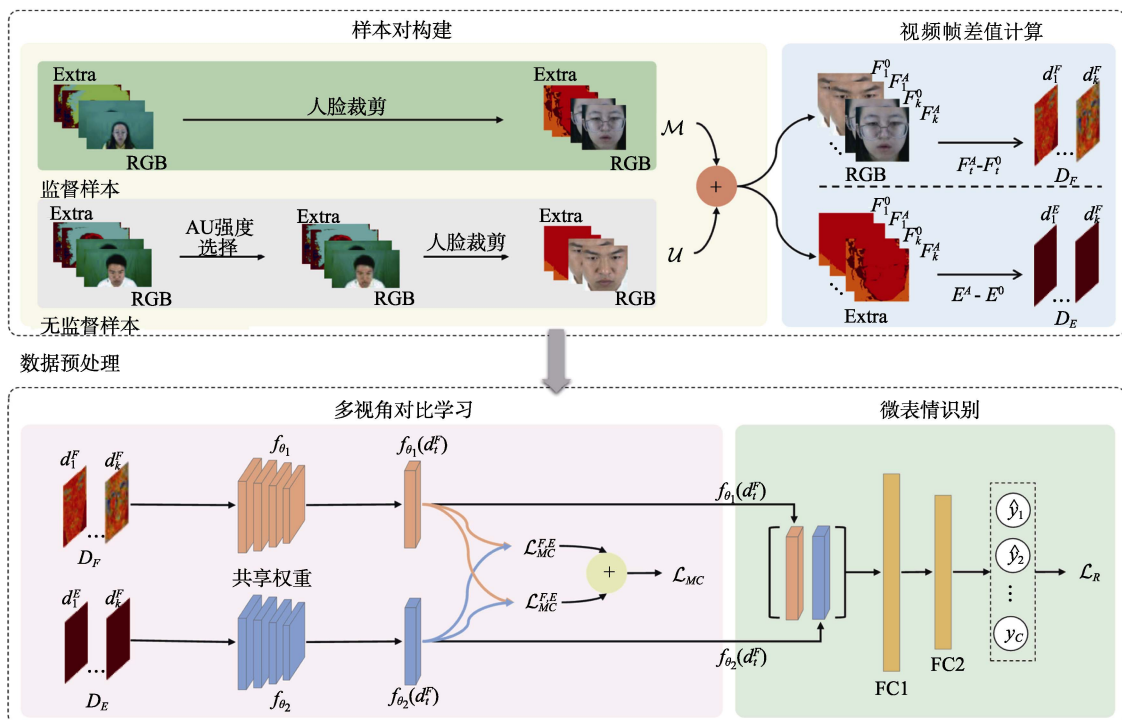
通过使用多视角对比进行自监督对比学习, 可以有效地学习 RGB 图像和额外模态两种不同视角下的微表情特征表示。这种多模态自监督学习方法可以利用两种模态之间的互信息, 提高模型的泛化能力和特征表达能力, 进而在微表情识别任务中取得更好的性能。

预训练结束后, 将自监督对比学习预训练得到的特征表示迁移到下游微表情识别任务中。具体来说, 首先冻结预训练网络中除线性分类层以外的所有层, 并在此基础上通过交叉熵损失训练一个分类器, 进行最终的微表情分类。

3.2.3 异步微表情联邦智能分析系统

1) 基于异步联邦学习的模型体系构建

由于微表情应用场景限制, 本研究设计基于异步联邦的微表情智能分析部署架构。如图 12 所



利用双视角自监督对比学习的微表情识别

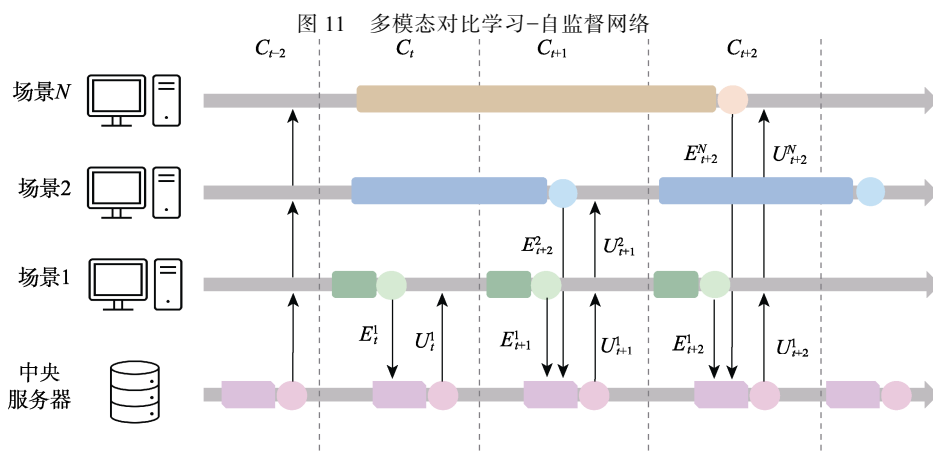


图 12 异步联邦学习框架

示首先基础模型构建中央服务器智能分析网络, 然后, 对于不同的场景, 在不同的通信轮次, 进行基于延迟轮次和场景样本分布的参数上传和迭代。

具体而言, 中央服务器模型, 即中央服务器托管基础智能分析网络, 负责协调学习过程并聚合来自各个节点的更新。同时, 不同的节点在不同的时间点向中央服务器发送更新, 服务器不需要等待所有节点同步完成。

设 θ 为节点 i 在时间 t 的模型参数, 中央服务器在接收到更新后进行参数聚合。在参数整合策略中, 深入考虑了三个主要影响因素, 以实现更加精准和高效的模型训练过程。首先, 针对不同节点场景下的样本数目差异, 本研究采用了样本参数权重的设置方法, 确保样本量较大的节点在参数整合过程中获得相对较大的权重, 这一策略有助于提升模型整体的学习效率。

和准确性。

其次, 考虑到各节点与中央服务器更新轮次的不一致性, 本研究引入了过时参数概念, 通过设定与通信轮次负相关的递减函数来调整过时参数的影响力, 以此确保中央服务器能够有效利用最新的全局信息进行参数更新。

最后, 为了应对实际应用中可能出现的通信延迟问题, 本研究还设置了本地节点迭代参数, 鼓励各节点在等待中央服务器更新的同时, 进行本地模型的自我迭代和优化, 从而不断增强模型的适应性和鲁棒性。

通过这三个策略的有机结合, 本研究旨在探索一种新型的参数整合机制, 既能够充分利用联邦学习的分布式特性, 又能确保模型训练的高效性和准确性。

2) 基于特定场景数据积累的增量学习模型

针对在子节点模型构建初期阶段微表情样本缺乏的问题, 本研究提出一个逐步积累和优化模型的策略。如图 13 所示, 首先, 利用现有的基础模型和大量未标注的数据, 构建自监督的面部动作学习模型。随后基于人工标注获取的一小批微表情样本, 对自监督模型在下游任务进行微调和优化。此外, 项目还将采用增量学习方法, 以逐步反馈的方式逐步提高模型的准确性和鲁棒性。这种方法不仅能够确保模型在早期阶段就具备一定的识别能力, 还能够随着时间的推移和数据的积累不断完善, 最终实现高效准确的微表情识别。

具体而言, 根据应用场景的需求, 设计一个不断适应新数据的学习过程。在这个过程中, 我们维护一个动态增长的数据集, 由已经标注的样本构成。

$$D_A = D_L \cup D_R$$

其中, D_L 为人工标注的样本集, D_R 为模型准确识别的样本集。累积的数据集为两者的并集 D_A 。随着时间的推移和模型的运行, D_L 和 D_R 会逐渐增加, 从而使得 D_A 不断扩大。在每个迭代步骤中, 模型会使用当前累积的数据集 D_A 进行训练, 以提高其对微表情的识别能力。新样本被模型预测后, 如果预测结果置信度较高, 则这些样本可以被自动加入到 D_R 中; 否则, 可能需要人工标注后才加入到 D_L 中。

4 理论构建与创新

微表情作为一种短暂微小的面部运动, 肉眼几乎不可见, 导致其智能分析的发展和应用面临诸多挑战。同时, 基于微表情的情绪识别在公共和国家安全等领域具有广泛应用潜力。本研究面向特定应用场景, 提出了一套实用的微表情智能分析方案, 特色和创新之处如下:

首先, 在心理学层面, 开展了高生态效度微表情诱发范式及其行为和生理机制研究。既往的微表情数据库多通过“面无表情”诱发范式采集, 生态效度与真实场景差异较大。本研究设计了基于行为和情绪心理学的交互情境微表情诱发范式, 采集接近真实情境的微表情数据。同时, 通过对生理和行为数据的心理学统计分析, 揭示不同交互场景下微表情的行为和生理机制, 为基于微表情的情绪监测奠定理论基础。

其次, 技术层面, 本研究构建了一个敏锐且鲁棒的微表情智能分析基础模型。研究基于面部肌肉动作拓扑结构设计分布式面周肌电采集系统, 提升

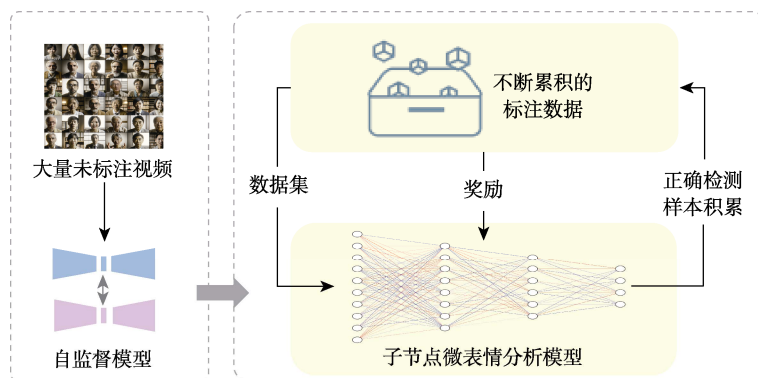


图 13 基于增量学习的微表情分析

了微表情编码效率。同时,采用头势分割和多模态口型变化识别技术,减少了复杂视频中的干扰因素,提升微表情检测准确性。通过自监督学习模型,在微表情样本量有限的情况下增强微表情识别性能。

最后,在应用推广方面,本研究构建匹配不同特定应用场景的微表情智能分析-云+本地部署架构。针对数据隐私和传输限制的问题,采用异步联邦学习框架,结合增量学习的子节点模型和中央服务器模型,支持在真实应用场景中的部署与验证。

参考文献

- Ang, L. B. P., Belen, E. F., Bernardo, R. A., Boongaling, E. R., Briones, G. H., & Coronel, J. B. (2004). Facial expression recognition through pattern analysis of facial muscle movements utilizing electromyogram sensors. *2004 IEEE Region 10 Conference TENCON 2004*, C, Vol. 3. (pp.600–603). Chiang Mai, Thailand. <https://doi.org/10.1109/TENCON.2004.1414843>
- Ben, X., Ren, Y., Zhang, J., Wang, S.-J., Kpalma, K., Meng, W., & Liu, Y.-J. (2022). Video-based facial micro-expression analysis: A survey of datasets, features and algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9), 5826–5846. <https://doi.org/10.1109/TPAMI.2021.3067464>
- Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (pp.1175–1191). Dallas, TX, USA. <https://doi.org/10.1145/3133956.3133982>
- Chen, Y., Yang, Z., & Wang, J. (2015). Eyebrow emotional expression recognition using surface EMG signals. *Neurocomputing*, 168, 871–879. <https://doi.org/10.1016/j.neucom.2015.05.037>
- Darwin, C. (1872). *The descent of man, and selection in relation to sex* (Vol. 2). D. Appleton.
- Davison, A., Merghani, W., Lansley, C., Ng, C.-C., & Yap, M. H. (2018). Objective micro-facial movement detection using FACS-based regions and baseline evaluation. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, (pp.642–649). <https://doi.org/10.1109/FG.2018.00101>
- Davison, A. K., Lansley, C., Costen, N., Tan, K., & Yap, M. H. (2016). Samm: A spontaneous micro-facial movement dataset. *IEEE Transactions on Affective Computing*, 9(1), 116–129.
- Davison, A. K., Li, J., Yap, M. H., See, J., Cheng, W.-H., Li, X., Hong, X., & Wang, S.-J. (2023). MEGC2023: ACM multimedia 2023 ME grand challenge. *Proceedings of the 31st ACM International Conference on Multimedia*, (pp.9625–9629). <https://doi.org/10.1145/3581783.3612833>
- Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, 32(1), 88–106. <https://doi.org/10.1080/00332747.1969.11023575>
- Frank, M., Herbasz, M., Sinuk, K., Keller, A., Kurylo, A., & Nolan, C. (2009). I see how you feel: Training laypeople and professionals to recognize fleeting emotions. *The Annual Meeting of the International Communication Association. Sheraton New York, New York City* (pp.1–35).
- Gruebler, A., & Suzuki, K. (2014). Design of a wearable device for reading positive expressions from facial EMG signals. *IEEE Transactions on Affective Computing*, 5(3), 227–237. <https://doi.org/10.1109/TAFFC.2014.2313557>
- Hamed, M., Salleh, S.-H., Astaraki, M., & Noor, A. M. (2013). EMG-based facial gesture recognition through versatile elliptic basis function neural network. *BioMedical Engineering OnLine*, 12(1), 73. <https://doi.org/10.1186/1475-925X-12-73>
- Huang, X., Wang, S.-J., Liu, X., Zhao, G., Feng, X., & 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. <https://doi.org/10.1109/TAFFC.2017.2713359>
- Husák, P., Cech, J., & Matas, J. (2017). Spotting facial micro-expressions “in the wild”. In *22nd Computer Vision Winter Workshop (Retz)* (pp.1–9). <http://cmp.felk.cvut.cz/~cechj/ME/>
- Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2017). Federated learning: Strategies for improving communication efficiency (No. arXiv:1610.05492). *arXiv*. <https://doi.org/10.48550/arXiv.1610.05492>
- Li, J., Dong, Z., Lu, S., Wang, S.-J., Yan, W.-J., Ma, Y., Liu, Y., Huang, C., & Fu, X. (2023). CAS(ME)³: A third generation facial spontaneous micro-expression database with depth information and high ecological validity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3), 2782–2800. <https://doi.org/10.1109/TPAMI.2022.3174895>
- Li, J., Soladié, C., & Séguier, R. (2023). Local temporal pattern and data augmentation for spotting micro-expressions. *IEEE Transactions on Affective Computing*, 14(1), 811–822. <https://doi.org/10.1109/TAFFC.2020.3023821>
- LI, J., Wang, S.-J., Yap, M. H., See, J., Hong, X., & Li, X. (2020). MEGC2020—The third facial micro-expression grand challenge. *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, 777–780. <https://doi.org/10.1109/FG47880.2020.00035>
- Li, J., Yap, M. H., Cheng, W.-H., See, J., Hong, X., Li, X., ... Dong, Z. (2022). MEGC2022: ACM multimedia 2022 micro-expression grand challenge. *Proceedings of the 30th ACM International Conference on Multimedia* (pp.7170–7174). <https://doi.org/10.1145/3503161.3551601>
- Li, J., Yap, M. H., Cheng, W.-H., See, J., Hong, X., Li, X., & Wang,

- S.-J. (2021). 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* (pp.5700–5701). <https://doi.org/10.1145/3474085.3478579>
- Li, X., Cheng, S., Li, Y., Behzad, M., Shen, J., Zafeiriou, S., Pantic, M., & Zhao, G. (2022). 4DME: A spontaneous 4D micro-expression dataset with multimodalities. *IEEE Transactions on Affective Computing*, 14(4), 3031–3047.
- Li, X., Hong, X., Moilanen, A., Huang, X., Pfister, T., Zhao, G., & 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. <https://doi.org/10.1109/TAFFC.2017.2667642>
- Li, X., Pfister, T., Huang, X., Zhao, G., & Pietikäinen, M. (2013). A spontaneous micro-expression database: Inducement, collection and baseline. *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (pp.1–6). <https://doi.org/10.1109/FG.2013.6553717>
- Lian, X., Zhang, W., Zhang, C., & Liu, J. (2018). Asynchronous decentralized parallel stochastic gradient descent. *Proceedings of the 35th International Conference on Machine Learning* (pp.3043–3052). <https://proceedings.mlr.press/v80/lian18a.html>
- Liu, Y.-J., Zhang, J.-K., Yan, W.-J., Wang, S.-J., Zhao, G., & Fu, X. (2016). A main directional mean optical flow feature for spontaneous micro-expression recognition. *IEEE Transactions on Affective Computing*, 7(4), 299–310. *IEEE Transactions on Affective Computing*. <https://doi.org/10.1109/TAFFC.2015.2485205>
- Lu, R., Zhang, W., Li, Q., He, H., Zhong, X., Yang, H., Alazab, M. (2024). Adaptive asynchronous federated learning. *Future Generation Computer Systems*, 152, 193–206. <https://doi.org/10.1016/j.future.2023.11.001>
- Lu, S., Li, J., Wang, Y., Dong, Z., Wang, S.-J., & Fu, X. (2022). A more objective quantification of micro-expression intensity through facial electromyography. *Proceedings of the 2nd Workshop on Facial Micro-Expression: Advanced Techniques for Multi-Modal Facial Expression Analysis* (pp.11–17). <https://doi.org/10.1145/3552465.3555038>
- Mansour, Y., Mohri, M., Ro, J., & Suresh, A. T. (2020). *Three approaches for personalization with applications to federated learning* (No. arXiv:2002.10619). arXiv. <https://doi.org/10.48550/arXiv.2002.10619>
- Mao, Q., Zhou, L., Zheng, W., Shao, X., & Huang, X. (2022). Objective class-based micro-expression recognition under partial occlusion via region-inspired relation reasoning network. *IEEE Transactions on Affective Computing*, 13(4), 1998–2016. <https://doi.org/10.1109/TAFFC.2022.3197785>
- Moilanen, A., Zhao, G., & Pietikäinen, M. (2014). Spotting rapid facial movements from videos using appearance-based feature difference analysis. *2014 22nd International Conference on Pattern Recognition* (pp.1722–1727). <https://doi.org/10.1109/ICPR.2014.303>
- Pan, H., Xie, L., & Wang, Z. (2022). Spatio-temporal convolutional emotional attention network for spotting macro- and micro-expression intervals in long video sequences. *Pattern Recognition Letters*, 162, 89–96. <https://doi.org/10.1016/j.patrec.2022.09.008>
- Perusquía-Hernández, M., Dollack, F., Tan, C. K., Namba, S., Ayabe-Kanamura, S., & Suzuki, K. (2021). Smile action unit detection from distal wearable electromyography and computer vision. *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)* (pp.1–8). <https://doi.org/10.1109/FG52635.2021.9667047>
- Polikovsky, S., Kameda, Y., & Ohta, Y. (2009). Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor. *3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009)*, P16/1–P16/6. <https://doi.org/10.1049/ic.2009.0244>
- Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine*, 25(1), 37–43.
- Qu, F., Wang, S.-J., Yan, W.-J., Li, H., Wu, S., & Fu, X. (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. <https://doi.org/10.1109/TAFFC.2017.2654440>
- Rinn, W. E. (1984). The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin*, 95(1), 52–77. <https://doi.org/10.1037/0033-2909.95.1.52>
- Sato, W., & Kochiyama, T. (2023). Crosstalk in facial EMG and its reduction using ICA. *Sensors*, 23(5), 2720. <https://doi.org/10.3390/s23052720>
- Sato, W., Murata, K., Uraoka, Y., Shibata, K., Yoshikawa, S., & Furuta, M. (2021). Emotional valence sensing using a wearable facial EMG device. *Scientific Reports*, 11(1), 5757.
- Schultz, I., & Pruzinec, M. (2010). *Facial expression recognition using surface electromyography* [Unpublished doctoral dissertation]. Karlsruhe Institute of Technology. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=7370f7109318a0ed91ae8a87371bb01d774e696e>
- See, J., Yap, M. H., Li, J., Hong, X., & Wang, S.-J. (2019). MEGC 2019 – the second facial micro-expressions grand challenge. *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)* (pp.1–5). <https://doi.org/10.1109/FG.2019.8756611>
- Shreve, M., Godavarthy, S., Goldgof, D., & Sarkar, S. (2011). Macro- and micro-expression spotting in long videos using spatio-temporal strain. *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)* (pp.51–56). <https://doi.org/10.1109/FG.2011.5771451>
- Wang, S.-J., He, Y., Li, J., & Fu, X. (2021). MESNet: A convolutional neural network for spotting multi-scale micro-expression intervals in long videos. *IEEE Transactions on Image Processing*, 30, 3956–3969. <https://doi.org/10.1109/TIP.2021.3064258>
- Wang, S.-J., Yan, W.-J., Li, X., Zhao, G., Zhou, C.-G., Fu, X., Yang, M., & Tao, J. (2015). Micro-expression recognition using color spaces. *IEEE Transactions on Image Processing*, 24(12), 6034–6047. <https://doi.org/10.1109/TIP.2015.2496314>

- Xia, B., Wang, W., Wang, S., & Chen, E. (2020). Learning from macro-expression: A micro-expression recognition framework. *Proceedings of the 28th ACM International Conference on Multimedia* (pp.2936–2944). <https://doi.org/10.1145/3394171.3413774>
- Xie, T., Sun, G., Sun, H., Lin, Q., & Ben, X. (2022). Decoupling facial motion features and identity features for micro-expression recognition. *PeerJ Computer Science*, 8, e1140.
- Xu, K., Chen, K., Sun, L., Lian, Z., Liu, B., Chen, G., Sun, H., Xu, M., & Tao, J. (2023). Integrating VideoMAE based model and optical flow for micro- and macro-expression spotting. *Proceedings of the 31st ACM International Conference on Multimedia* (pp.9576–9580). <https://doi.org/10.1145/3581783.3612868>
- Yan, W.-J., Li, X., Wang, S.-J., Zhao, G., Liu, Y.-J., Chen, Y.-H., & Fu, X. (2014). CASME II: An improved spontaneous micro-expression database and the baseline evaluation. *PLOS ONE*, 9(1), e86041. <https://doi.org/10.1371/journal.pone.0086041>
- Yan, W.-J., Wu, Q., Liu, Y.-J., Wang, S.-J., & Fu, X. (2013). CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces. *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (pp.1–7). <https://doi.org/10.1109/FG.2013.6553799>
- Yap, C. H., Yap, M. H., Davison, A., Kendrick, C., Li, J., Wang, S.-J., & Cunningham, R. (2022). 3D-CNN for facial micro- and macro-expression spotting on long video sequences using temporal oriented reference frame. *Proceedings of the 30th ACM International Conference on Multimedia* (pp.7016–7020). <https://doi.org/10.1145/3503161.3551570>
- Yin, S., Wu, S., Xu, T., Liu, S., Zhao, S., & Chen, E. (2023). AU-aware graph convolutional network for macroand micro-expression spotting. *2023 IEEE International Conference on Multimedia and Expo (ICME)* (pp.228–233). <https://doi.org/10.1109/ICME55011.2023.00047>
- Yu, W.-W., Jiang, J., Yang, K.-F., Yan, H.-M., & Li, Y.-J. (2024). LGSNet: A two-stream network for micro- and macro-expression spotting with background modeling. *IEEE Transactions on Affective Computing*, 15(1), 223–240. <https://doi.org/10.1109/TAFCC.2023.3266808>
- Zhang, J., Huang, S., Li, J., Wang, Y., Dong, Z., & Wang, S.-J. (2023). A perifacial EMG acquisition system for facial-muscle-movement recognition. *Sensors*, 23(21), Article 21. <https://doi.org/10.3390/s23218758>
- Zhang, L. (2022). Animation expression control based on facial region division. *Scientific Programming*, 2022(1), 5800099. <https://doi.org/10.1155/2022/5800099>
- Zhang, L., Hong, X., Arandjelović, O., & Zhao, G. (2022). Short and long range relation based spatio-temporal transformer for micro-expression recognition. *IEEE Transactions on Affective Computing*, 13(4), 1973–1985. <https://doi.org/10.1109/TAFCC.2022.3213509>
- Zhang, L.-W., Li, J., Wang, S.-J., Duan, X.-H., Yan, W.-J., Xie, H.-Y., & Huang, S.-C. (2020). Spatio-temporal fusion for macro- and micro-expression spotting in long video sequences. *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)* (pp.734–741). <https://doi.org/10.1109/FG47880.2020.00037>
- Zhang, Y., Liu, D., Duan, M., Li, L., Chen, X., Ren, A., Tan, Y., & Wang, C. (2023). FedMDS: An efficient model discrepancy-aware semi-asynchronous clustered federated learning framework. *IEEE Transactions on Parallel and Distributed Systems*, 34(3), 1007–1019. <https://doi.org/10.1109/TPDS.2023.3237752>
- Zhang, Y., Wang, H., Xu, Y., Mao, X., Xu, T., Zhao, S., & Chen, E. (2023). Adaptive graph attention network with temporal fusion for micro-expressions recognition. *2023 IEEE International Conference on Multimedia and Expo (ICME)* (pp.1391–1396). <https://doi.org/10.1109/ICME55011.2023.00241>
- Zhao, S., Tang, H., Mao, X., Liu, S., Zhang, Y., Wang, H., Xu, T., & Chen, E. (2024). DFME: A new benchmark for dynamic facial micro-expression recognition. *IEEE Transactions on Affective Computing*, 15(3), 1371–1386. <https://doi.org/10.1109/TAFCC.2023.3341918>
- Zhu, J., Zong, Y., Chang, H., Xiao, Y., & Zhao, L. (2022). A sparse-based transformer network with associated spatiotemporal feature for micro-expression recognition. *IEEE Signal Processing Letters*, 29, 2073–2077. <https://doi.org/10.1109/LSP.2022.3211200>

Micro-expression analysis for practical applications: From data acquisition to intelligent deployment

LI Jingting, ZHAO Lin, DONG Zizhao, WANG Su-Jing

(State Key Laboratory of Cognitive Science and Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China) (Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Micro-expressions, subtle facial movements that occur when individuals attempt to conceal their true emotions, serve as highly efficient non-invasive cues for monitoring psychological states. They hold immense potential for application in fields such as national and public security. However, implementing

micro-expression analysis in specific application scenarios poses numerous challenges. In response to these challenges, an approach is taken that initially designs micro-expression elicitation paradigms with high ecological validity based on physiological and behavioral psychology research. This is complemented by the development of an auxiliary coding system based on facial peripheral electromyography and the creation of a large-scale micro-expression database to deeply explore the physiological and behavioral mechanisms of micro-expressions in interactive environments. Furthermore, efforts are focused on addressing the difficulties of constructing micro-expression analysis models in specific application scenarios. This includes designing interference elimination modules to avoid the impact of head movements and lip movement changes on micro-expression spotting. Additionally, self-supervised recognition models are built based on large-scale vertical domain vision models to address the challenge of small sample sizes. Moreover, considering data privacy and transmission constraints, intelligent micro-expression analysis deployment in different scenarios is realized based on asynchronous federated learning. Through the interdisciplinary integration of psychology and computer science, this endeavor aims to construct a micro-expression analysis framework that balances theory and practical application. This not only enhances the accuracy and applicability of micro-expression analysis but also paves the way for its application in more fields.

Keywords: micro-expression intelligent analysis, affective computing, high ecological validity, practical application