软件学报 ISSN 1000-9825, CODEN RUXUEW
[doi: 10.13328/j.cnki.jos.007630] [CSTR: 32375.14.jos.007630]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

# 基于深度学习的微表情分析综述*

胡晨星[1], 苗启广[1], 刘璨宇[1], 刘旭杰[1], 刘如意[1], 王 泉[1], 杨宗凯[1,2]

[1](西安电子科技大学 计算机科学与技术学院, 陕西 西安 710126)

[2](华中师范大学, 湖北 武汉 430079)

通信作者: 苗启广, E-mail: qgmiao@xidian.edu.cn; 杨宗凯, E-mail: zkyang@mail.ccnu.edu.cn

**摘 要**: 微表情能够揭示人类试图隐藏的真实情感, 在谎言检测、医疗护理、深层次情感识别等领域中具有重大应用价值. 不同于宏表情, 微表情是短暂且细微的面部运动, 微表情数据集构建有着较高的门槛且耗时耗力, 因此微表情面临着数据量小、模型难以捕获细微短暂变化等困难. 微表情分析有着两个重要研究分支: 微表情识别和微表情检测, 目前基于深度学习的方法已经成为研究主流. 针对上述情况, 全面综述近年来基于深度学习的微表情识别和检测代表性工作和研究现状. 首先, 对常用的开源微表情数据集全面进行总结, 然后分别介绍微表情识别和检测的代表性研究工作, 对其进行分类梳理并讨论优点和局限性. 接着, 介绍微表情分析常用的评估协议和指标, 整合讨论各个方法的性能表现. 最后对微表情分析的整体发展状况进行总结并对未来研究方向进行展望.

**关键词**: 微表情分析; 微表情识别; 微表情检测; 深度学习; 计算机视觉

**中图法分类号**: TP18

中文引用格式: 胡晨星, 苗启广, 刘璨宇, 刘旭杰, 刘如意, 王泉, 杨宗凯. 基于深度学习的微表情分析综述. 软件学报. http://www.jos.org.cn/1000-9825/7630.htm

英文引用格式: Hu CX, Miao QG, Liu CY, Liu XJ, Liu RY, Wang Q, Yang ZK. Survey on Micro-expression Analysis Based on Deep Learning. Ruan Jian Xue Bao/Journal of Software (in Chinese). http://www.jos.org.cn/1000-9825/7630.htm

## Survey on Micro-expression Analysis Based on Deep Learning

HU Chen-Xing[1], MIAO Qi-Guang[1], LIU Can-Yu[1], LIU Xu-Jie[1], LIU Ru-Yi[1], WANG Quan[1], YANG Zong-Kai[1,2]

[1](School of Computer Science and Technology, Xidian University, Xi'an 710126, China)

[2](Central China Normal University, Wuhan 430079, China)

**Abstract**: Micro-expressions can reveal genuine emotions that individuals attempt to conceal and therefore have significant application value in areas such as lie detection, healthcare, and fine-grained emotion recognition. Unlike macro-expressions, micro-expressions are brief and subtle facial movements. In addition, the construction of micro-expression datasets involves high costs and considerable effort, which results in limited data availability and makes it difficult for models to capture subtle and transient facial changes. Micro-expression analysis mainly consists of two important research branches: micro-expression recognition and micro-expression detection. Currently, deep learning-based methods dominate research in this field. Accordingly, this study comprehensively reviews representative works and recent advances in deep learning-based micro-expression recognition and micro-expression detection. Firstly, the study provides a systematic summary of commonly used open-source micro-expression datasets. Secondly, it separately introduces representative research on micro-expression recognition (MER) and micro-expression detection (MED), classifies and organizes these methods, and discusses their respective advantages and limitations. Furthermore, it presents commonly used evaluation protocols and metrics for micro-expression analysis and integrates discussions on the performance of various methods. Finally, it summarizes the overall development of micro-expression analysis and offers insights into future research directions.

**Key words**: micro-expression analysis; micro-expression recognition; micro-expression detection; deep learning; computer vision

---

* 基金项目: 国家自然科学基金 (62272364); 陕西省重点研发计划 (2024GH-ZDXM-47); 陕西省高等教育教学改革研究项目 (23JG003); 西安电子科技大学人工智能赋能研究生教育项目 (AIZS2501); 西安电子科技大学 2024 年研究生教育综合改革专项计划
收稿时间: 2025-07-24; 修改时间: 2025-11-23; 采用时间: 2025-12-30; jos 在线出版时间: 2026-07-15

面部表情蕴含着丰富的情感信息, 在情感分析中发挥着重要作用. 通常来讲, 面部表情可以分为宏表情和微表情, 这两类表情的主要视觉区别在于它们的持续时间和运动强度, 宏表情的持续时间通常在 0.5 s 以上, 并且运动覆盖较大面部区域[1,2], 相比之下微表情的持续时间通常在 0.065–0.5 s 之间[3−5], 运动细微. 此外, 宏表情可以被人为表演, 从而掩盖真实的情绪, 因此在某些特殊的场景下, 无法通过宏表情准确地分析出一个人的内在真实情感. 微表情则是一种无法控制的面部肌肉运动, 即便是刻意的抑制面部表情, 这些细微的运动仍然经常发生, 因此微表情分析成为揭示人类深层情感的重要手段[6,7], 在医疗护理、刑事讯问和情感分析等领域被广泛应用[8−11]. 依靠肉眼进行微表情识别非常困难, 这不仅需要观察者集中注意力来捕获微小的面部运动, 还需要拥有丰富的经验才能正确分析出微表情蕴含的情绪信息, 即使经过严格的训练, 人工平均也只能识别不足 50% 的微表情[12], 且人工识别微表情耗时费力、成本高昂, 因此开发基于计算机视觉的自动化微表情分析算法是必要的[13].

微表情分析核心研究任务有微表情识别和检测两个方面. 识别是将给定的微表情片段分类成不同的情感类型, 检测则是在给定的原始面部视频序列中定位到微表情发生的时间段或关键帧. 目前部分微表情分析领域综述仅聚焦于识别或检测单一任务, 未对两个研究方向进行系统性覆盖[14,15]. 除此之外, Ben 等人[16]整合了微表情检测与识别两大核心方向的研究进展及关键挑战, 并构建了新数据集“微表情与宏表情仓库 (MMEW)”. Zhao 等人[17]进一步拓展综述范围, 还纳入了微表情动作单元检测和微表情生成相关研究, 但这两个综述均未将深度学习在微表情分析中的应用进展作为核心议题展开深入探讨. 深度学习技术在许多传统计算机视觉任务中 (例如目标检测[18−20]、语义分割[21−23]等) 取得了卓越的进展, 基于深度学习的技术目前已经成为微表情分析研究的主流, 而且微表情分析整体正处于快速发展期, 对综述文章的时效性要求较高. 综上所述, 本文专注于深度学习且同时对识别和检测这两个微表情分析领域重要分支进行全面总结, 能够帮助研究人员快速掌握微表情分析的研究现状, 并获得对未来研究方向的参考.

本文整体组织结构如图 1 所示. 第 1 节介绍现有研究常用的微表情数据集, 从微表情诱发、收集和标注这 3 个方面总结了制作微表情数据集的要点, 介绍了常用的数据处理方式. 第 2、3 节分别介绍基于深度学习的微表情识别和检测方法的研究现状. 第 4 节介绍微表情分析评估指标. 第 5 节从微表情识别和微表情检测两个方面对近几年代表性研究工作的效果进行总结和讨论, 第 6 节对基于深度学习的微表情分析领域进行展望, 对未来研究方向提供参考.

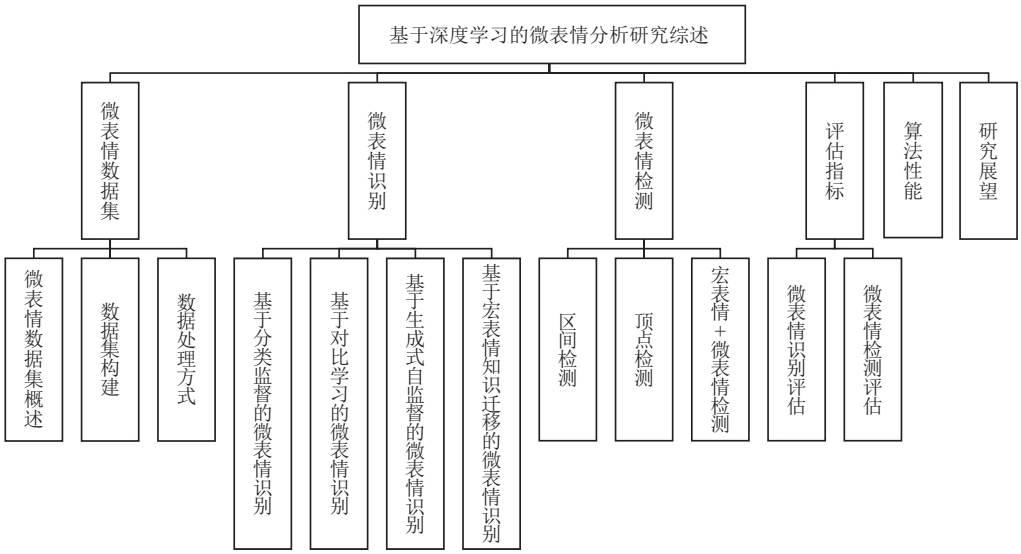



图 1　本文的组织结构

## 1 微表情数据集及常用数据处理方式

### 1.1 微表情数据集概述

深度学习中, 高质量大规模数据集的诞生能够极大推动相关领域的发展, 但由于微表情的特性 (自发性、短暂、细微等), 微表情数据集的构建面临诱发难、采集难、标注难的困境. 在 2010 年左右, 出现了两个专用于微表情研究的数据集 Polikovsky's[24]和 USF-HD[25], 但这两个数据集在构建时要求被试者故意表演对应的微表情, 与微表情的自发性相矛盾, 因此目前主流研究中基本不再使用这两个数据集. 实验环境下采集的微表情数据集需要采用合理的诱发手段来保证微表情的自发性, 尽管目前已有很多自发微表情数据集可供研究, 但相较于一些数据易得且标注简单的领域, 微表情数据集的规模和质量仍还有很大的进步空间. 我们将常用微表情数据集的关键参数汇总在表 1 中.

表 1 微表情数据集

| 数据集 | 年份 | 分辨率 | 帧率 (f/s) | 微表情样本 | 平均年龄 | AU | 类别标签 (数量) |
|---|---|---|---|---|---|---|---|
| CASME[26] | 2013 | 640×480<br>1280×720 | 60 | 195 | 22.03 | √ | 高兴 (5), 厌恶 (88), 恐惧 (2), 蔑视 (3), 悲伤 (6), 紧张 (28), 惊讶 (20), 抑制 (40) |
| CASME II[27] | 2014 | 640×480 | 200 | 247 | 22.03 | √ | 高兴 (33), 厌恶 (60), 惊讶 (25)<br>抑制 (27), 其他 (102) |
| CAS(ME)$^2$[28] | 2017 | 640×480 | 30 | 57 | 22.59 | √ | 积极 (51), 消极 (70), 惊讶 (43), 其他 (19) |
| CAS(ME)$^3$[29] | 2022 | 1280×720 | 30 | 1 109 | 22.74 | √ | 高兴 (64), 厌恶 (281), 恐惧 (93), 愤怒 (70), 惊讶 (201), 悲伤 (64), 其他 (170) |
| CASMEMG[30] | 2024 | 3840×2160 | 30 | 233 | 24.7 | √ | 高兴 (49), 愤怒 (80), 厌恶 (91), 恐惧 (52), 悲伤 (65), 惊讶 (40), 其他 (3) |
| MMEW[16] | 2021 | 1920×1080 | 90 | 300 | 22.35 | √ | 高兴 (36), 愤怒 (8), 惊讶 (89), 厌恶 (72), 恐惧 (16), 悲伤 (13), 其他 (66) |
| SAMM[31] | 2016 | 2040×1080 | 200 | 159 | 33.24 | √ | 高兴 (24), 愤怒 (20), 惊讶 (13), 厌恶 (8), 恐惧 (7), 悲伤 (3), 其他 (84) |
| SAMM LONG[32] | 2016 | 2040×1080 | 200 | 159 | 33.24 | √ | N/A |
| SMIC[33] HS/NIR/VIS | 2013 | 640×480 | 100/25/25 | 164/71/71 | 28.1 | × | HS: 积极 (51), 消极 (70), 惊讶 (43)<br>NIR: 积极 (28), 消极 (23), 惊讶 (20)<br>VIS: 积极 (28), 消极 (24), 惊讶 (19) |
| MEVEIW[34] | 2017 | 1280×720 | 25 | 29 | N/A | √ | 高兴 (5), 愤怒 (1), 惊讶 (13), 厌恶 (1), 恐惧 (3), 蔑视 (4), 未知 (7) |
| 4DME[35]<br>4D/GS/RGB/DEPTH | 2022 | 1200×1600<br>640×480<br>640×480<br>640×480 | 60<br>30<br>30<br>30 | 1 068 | 27.8 | √ | 积极 (34), 消极 (127), 惊讶 (30), 抑制 (6), 积极+惊讶 (13), 消极+惊讶 (8), 抑制+惊讶 (3), 积极+抑制 (8), 消极+抑制 (7), 其他 (31) |
| DFME[36] | 2023 | 1024×768 | 500/300/<br>200 | 7 526 | 22.43 | √ | 高兴 (992), 厌恶 (2 528), 恐惧 (892), 愤怒 (619), 蔑视 (401), 惊讶 (1 208), 悲伤 (635), 其他 (251) |

CASME 系列数据集由中国科学院心理研究所发布, 为微表情分析领域做出重大贡献, 目前超过 80% 的微表情分析文章都采用了系列中的至少一个数据集进行方法验证[29]. 其中 CASME[26]是系列首个自发微表情数据集, 包含 19 名受试者的 195 个自发微表情样本, 所有样本均以 60 f/s 的帧率记录, 数据依据不同的采集设备及环境被分为 A 类和 B 类两部分, 其中 A 类分辨率设置为 1280×720, 在自然光下拍摄, B 类分辨率为 640×480, 在有两盏 LED 灯打光的室内拍摄, A 类和 B 类中所有样本人脸区域的分辨率均为 150×190. CASME II[27]相较于 CASME 数据集拥有更大的数据规模 (来自 26 名参与者的 247 个样本)、更高的帧率 (200 f/s) 以及更高的面部区域分辨率 (280×340), CASME II 是目前微表情识别研究最常用的数据集之一. 为了给长视频中的微表情检测提供数据支撑,

开发首个开源长视频微表情数据集 CAS(ME)$^2$[28], 其中有 87 个同时包含宏表情和微表情的长视频以及裁剪好的 300 个宏表情样本和 57 个微表情样本, 样本的情感类别均为积极、消极、惊讶和其他这 4 类.

SMIC[33]数据集是最早发布的自发微表情数据集, 包含了 16 名被试者的 164 个微表情片段, 所有片段均记录在 SMIC-HS (高速) 数据集中, 来自 8 名被试者的 71 个片段也被记录在 SMIC-VIS (可见光) 和 SIMC-NIR (近红外) 数据集中. SAMM[31]数据集包含了来自 32 名被试者的 159 个微表情片段, 被试者包含了 13 个不同种族并且性别分布完全均衡, 数据集拥有极高的帧率 (200 f/s), 以及约 400×400 的面部区域分辨率. SAMM LONG 长视频数据集[32]在 SAMM 的基础上进行扩展, 147 段长视频中同时包含 343 个宏表情和 159 个微表情, 旨在为长视频中的宏表情和微表情检测提供支撑, 但该数据集未提供具体的情感标签. 此外, 早期还有一个自发微表情数据集 MEVIEW[34]被提出, 但与上述数据集不同的是, MEVIEW 由从互联网下载的两个真实场景的视频片段组成, 场景包括扑克游戏和电视采访, 这种真实场景下的微表情数据集拥有最自然的微表情流露, 但不可控场景导致数据有着光照不一、面部不全等缺点, 因此在目前的微表情研究中该数据集较少被使用.

上述数据集均为较早期出现, 为微表情分析领域奠定了坚实的基础, 但仍还存在一定的局限性, 首先是可供研究的有效微表情样本依旧太少, 其次在更多模态数据的探索上有所欠缺.

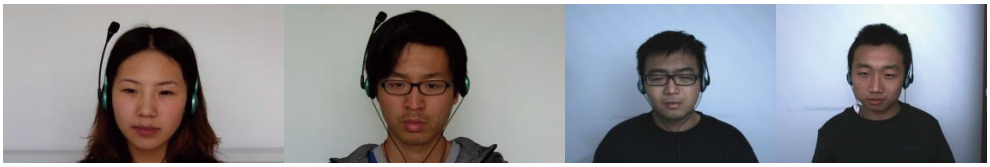
Ben 等人[16]于 2021 年的综述论文中发布微表情数据集 MMEW, 包含来自 36 名参与者的 300 个微表情样本和 900 个宏表情样本, 这些样本以 90 f/s 和 1920×1080 分辨率采集. Zhao 等人[36]在 2023 年推出 DFME 数据集, 是目前规模最大的微表情数据集, 包含 7 526 个高质量标注的微表情视频, 涵盖了多种高帧率 (200 f/s、300 f/s、500 f/s). 这些视频通过情感刺激实验从 671 名参与者中收集, 标注了 7 个情感类别: 快乐、悲伤、愤怒、蔑视、恐惧、厌恶、惊讶. 这些数据集的提出为微表情分析提供了更大规模的样本.

中国科学院心理研究所新推出的两个 CASME 系列微表情数据集 CAS(ME)$^3$[29]和 CASMEMG[30]以及 Li 等人[35]推出的 4DME 数据集在多模态角度做出大量探索. 其中 CAS(ME)$^3$[29]数据集中首次将深度信息引入, 构建了一个包含 RGB 图像、生理、语音以及深度信息的多模态自发微表情数据集, 并且 CAS(ME)$^3$ 尝试了一种新的微表情诱发范式, 通过模拟犯罪场景来收集更高生态效度的微表情样本, CAS(ME)$^3$ 由 3 部分组成, A 部分包含 100 名被试者的 1 300 段视频, 标注员标注了 943 个微表情以及 3 143 个宏表情, B 部分包含 116 名被试者的 1 508 个未标注视频片段, 可以为自监督方法提供大规模数据支撑, C 部分则是采用模拟犯罪范式采集的数据, 包含 31 名被试者的 166 个微表情和 347 个宏表情, 3 个部分总计包含了约 800 万帧视频, 1 109 个微表情及 3 490 个宏表情. CASMEMG[30]则探索了肌电信号 (electromyography, EMG) 与微表情外部特征的关联性, 旨在促进基于肌电信号的微表情机制和运动变化动态研究, 构建了一个包含其中包括 233 个微表情样本和 147 个宏表情样本的数据集, 每个样本均配有相应的 EMG 和视频数据. 4DME 数据集[35]则包含了 DI4D 视频、正面灰度视频、Kinect 彩色视频和 Kinect 深度视频这 4 种模态的表情数据, 在最终的标注手册中共有来自 41 名受试者的 267 个微表情和 123 个宏表情, 这 4 种模态的微表情样本总数为 1 068 个. 后文图 2 展示了部分数据集的微表情样本.

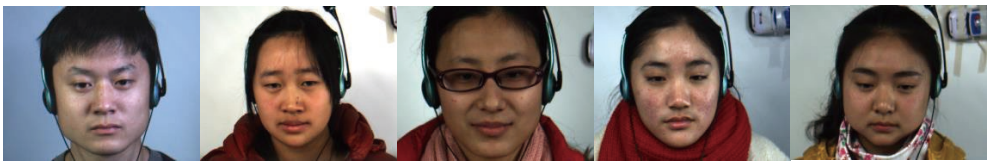
需要注意的是, 当前微表情情感类别标注尚未形成统一的客观标准与规范: 一方面, 微表情本身具有强度微弱、持续时间短的特性, 导致标注者难以完全正确地观察每个样本的运动细节, 即便针对同一批样本, 不同标注者也可能因主观认知、专业背景的不同给出不一致的标签; 另一方面, 不同团队数据集的标注体系或多或少存在差异, 这也导致跨数据集研究难度增加. 尽管多数数据集会通过多标注者一致性检验来优化标注质量, 但受上述客观局限与主观差异影响, 仍难以实现微表情情感标注的完全统一.

### 1.2 数据集构建

面部动作编码系统 (facial action coding system, FACS)[37]对面部表情分析影响深远, FACS 根据人类面部解剖结构对每个视觉上可区分的面部肌肉动作进行编码和分类, 这些动作被称为动作单元 (action unit, AU), AU 对应的面部运动描述及关联面部肌肉如表 2 所示. FACS 通过心理学研究观察证据为宏表情提供了 AU-情感对照表, 将 AU 的组合映射到对应的情感类别, 但微表情 AU 和情感类别的映射关系尚不明确, 即便如此, 大多数自发微表情数据集也会对样本进行 AU 标注并将 AU 作为微表情情感类别标注的参考依据之一.

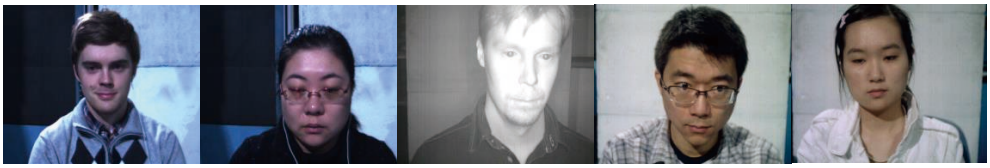


(a) CASME 数据集

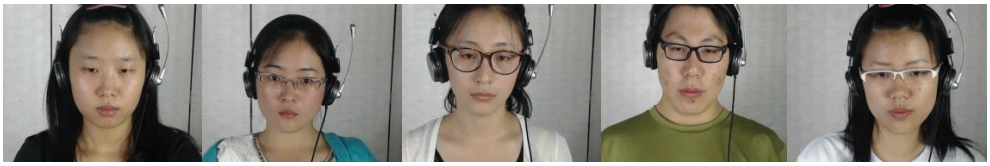


(b) CASME II 数据集

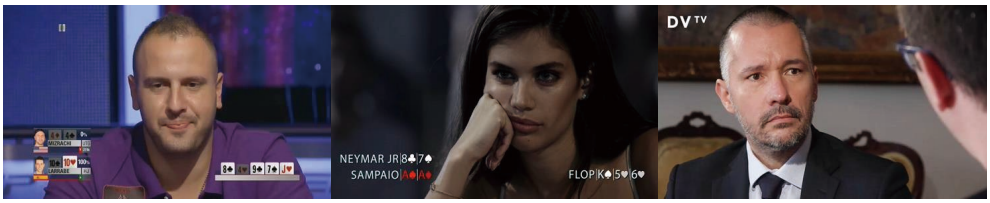
(c) SAMM 数据集



(d) SMIC 数据集



(e) CAS(ME)$^2$ 数据集




(f) MEVIEW 数据集

图 2　微表情数据集示例

建立微表情数据集通常要经过 3 个主要步骤: 微表情诱发、微表情采集、微表情标注, 其中微表情标注需要标定微表情发生片段、情感类别以及 AU. 负责 AU 的标注员需要拥有 FACS 资质, 负责情感类别的标注员则需要学习大量的微表情知识, 因此微表情数据集构建有很高的门槛, 此外, 由于微表情本身迅速且细微的特点, 标注者需要在海量原始视频中逐帧比对, 因此构建微表情数据集极为耗时耗力, 提出一个高质量开源的微表情数据集可以为领域做出巨大贡献. 接下来我们将分别介绍微表情诱发、采集和标注.

(1) 微表情诱发

合理的诱发范式可以营造微表情密集的场景以便于高效的收集微表情数据, 这通常需要满足两点: 高压、情感

抑制. 为了让参与者产生自发的微表情, 目前常用的做法是要求参与者全程保持面无表情, 通过诱发材料 (常为精心挑选的视频片段) 来诱发参与者产生无法抑制的、自发的微表情, 其中也会采用流露表情扣除被试费、填写复杂问卷等手段来让被试者产生克制表情流露的高压环境, 这种高压情感抑制的范式被研究人员广泛认可. 在 CAS(ME)$^3$[29] 数据集中首次提出了一种新的诱发范式, 设计了一个模拟犯罪场景并对嫌疑人进行审讯, 在审讯过程中进行微表情数据采集, 但整体仍处于严格控制的实验环境中进行. 需要注意的是, 微表情诱发场景的设计要符合心理学理论并严格遵循伦理道德规范, 尤其在进行特殊人群 (儿童、患者等) 的微表情采集时, 要保证诱发刺激的适度.

表 2 面部动作编码系统

| AU | 描述 | 面部肌肉 | AU | 描述 | 面部肌肉 |
|---|---|---|---|---|---|
| 1 | 眉毛内侧抬升 | 额肌, 内侧肌群 | 18 | 撅嘴 | 上唇切牙肌, 下唇切牙肌 |
| 2 | 眉毛外侧抬升 | 额肌, 外侧肌群 | 20 | 嘴唇拉伸 | 笑肌和颈阔肌 |
| 4 | 降低眉毛 | 皱眉肌, 降眉肌 | 22 | 漏斗嘴 | 口轮匝肌 |
| 5 | 上眼睑提升 | 提上睑肌 | 23 | 收紧嘴 | 口轮匝肌 |
| 6 | 脸颊上升 | 眼轮匝肌外圈 | 24 | 抿嘴 | 口轮匝肌 |
| 7 | 眼睑收紧 | 眼轮匝肌内圈 | 25 | 嘴唇分开 | 降下唇肌/放松颏肌/口轮匝肌 |
| 9 | 皱鼻 | 鼻肌和上唇鼻翼肌 | 26 | 放松下颌 | 咬肌, 放松颞肌和翼内肌 |
| 10 | 上唇抬起 | 提上唇肌眶下头 | 27 | 张大嘴巴 | 翼状肌和二腹肌 |
| 11 | 鼻唇沟加深 | 颧小肌 | 28 | 吸嘴唇 | 口轮匝肌 |
| 12 | 嘴角拉伸 | 颧大肌 | 41 | 眼睑下垂 | 提上睑肌放松 |
| 13 | 脸颊鼓起 | 提口角肌 | 42 | 眯眼 (slit) | 眼轮匝肌 |
| 14 | 酒窝 | 颊肌 | 43 | 闭眼 | 提上睑肌放松, 眼轮匝肌内圈 |
| 15 | 嘴角下压 | 降口角肌 | 44 | 眯眼 (squint) | 眼轮匝肌内圈 |
| 16 | 下嘴唇下压 | 降下唇肌 | 45 | 眨眼 (blink) | 提上睑肌放松, 眼轮匝肌内圈 |
| 17 | 下巴提起 | 颏肌 | 46 | 眨眼 (wink) | 提上睑肌放松, 眼轮匝肌内圈 |

(2) 微表情采集

微表情对噪声非常敏感, 因此要严格控制采集环境, 被试者需要全程保持头部稳定, 并且要设置合理的灯光以免出现阴影、光照不足、光照过量等情况. 由于微表情的短暂性, 高速摄像机可以捕获到更细致的时序运动变化, 保留更多的微表情信息, SAMM[31]、DFME[36]等数据集均采用了 200 f/s 以上的高速摄相机, 但需要注意的是, 更高速的摄像机意味着更大的微表情标注工作量.

(3) 微表情标注

微表情数据集标注通常有两个重要阶段: 样本选择阶段和样本标注阶段. 在样本选择阶段标注者需要对采集到的原始视频数据进行逐帧观察, 从大量原始数据中找出与情绪明确相关的表情片段, 在这些数据可能蕴含如眨眼、头部整体偏移等表情无关运动, 有的被试者也可能存在一些习惯性的面部肌肉运动, 这一步往往需要耗费大量的时间精力. 在样本标注阶段, 标注者需要根据 FACS 进行 AU 标注、对表情发生片段进行关键帧标注 (起始帧、顶点帧和结束帧), 由于微表情的情感类别和 AU 组合没有明确的映射关系, 情感类别标注需要标注者拥有丰富的微表情阅读经验. 为了保证数据标注的客观性和可靠性, 标注通常需要多名标注者分别独立的对所有数据进行标注, 并根据标注结果进行一致性检验.

### 1.3 数据处理方式

原始的微表情数据集中常包含大量干扰因素, 例如多余的背景信息、头部整体偏移等. 微表情的特有性质导致轻微的扰动都有可能使其对结果产生致命的影响, 因此合理的数据预处理手段在微表情分析中非常重要. 除了深度学习普适常用的图像的数据增强手段 (如随机旋转、裁剪等)[38]之外, 微表情分析中有一些特殊的处理方式. 从原始的微表情视频数据出发, 研究者们会采用人脸检测和人脸对齐来尽可能消除背景信息干扰, 缓解头部运动影响, 这两步往往在研究工作中是必要的. 微表情是极为细微的面部运动, 为了获得更显著的时空运动信息, 有大量工作会采用帧间光流或运动放大后的图像作为网络输入. 图 3 展示了常用数据处理方式效果示例, 展示样例均

取自 CASME Ⅱ 数据集[27], 接下来将逐一介绍这些处理方法.

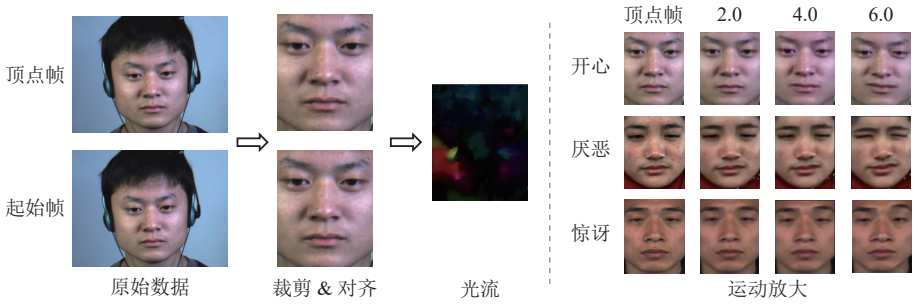



图 3 常用数据预处理方式

(1) 人脸裁剪对齐

原始的数据常包含大量面部区域以外的无关背景信息, 人脸检测目的就是裁剪出人脸区域从而去除背景干扰. 目前基于深度学习的人脸检测方法被广泛使用, 这些方法相较于传统人脸检测方法有着更强的泛用性和稳健性, 研究者通过 OpenCV、Dlib 和 OpenFcae 等开源库可以很便捷的使用人脸检测工具. 微表情是非常细微的面部肌肉运动, 很容易受到外界影响, 人脸对齐旨在依据面部关键特征点将两帧人脸图像对齐, 可以有效地缓解头部整体运动造成的影响, 也可以消除人脸检测裁剪导致的细微位置偏移. 人脸检测和对齐通常在微表情分析中是必要的.

(2) 运动放大

由于微表情的细微性, 可以显著增强细微运动强度的运动放大技术自然而然地被应用到微表情分析任务中. 早期的微表情分析工作常用欧拉视频放大 (Eulerian video magnification, EVM)[39]方法, 该方法从全局视角分析视频内容的时空变化, 通过滤波提取并放大特定频段的微小运动或颜色变化信号, 最终重构视频以实现微小变化的可视化增强. 此外, Lei 等人[40]首次将基于学习的运动放大技术 MagNet[41]引入到微表情分析中, MagNet 以两帧图像作为输入, 输出将两帧图像差异放大后的图像, 该图像拥有更显著的运动特征, 也可以直接使用 MagNet 中间层的多通道特征作为后续网络输入以减少运动放大带来的噪声. 需要注意的是, 运动放大的放大因子选择至关重要, 过小的放大因子可能无法起到应有的效果, 而过大的放大因子可能会导致引入过多的噪声, 在训练时可以对训练数据随机采样放大因子以进行数据增强.

(3) 光流

光流计算图像之间的像素运动方向和幅度, 代表了两帧图像之间的运动信息, 能有效地滤除身份特征影响. 在微表情研究中光流被广泛认为是高效的运动表示方法, 在基于深度学习的微表情分析研究中, 大量的工作以光流数据作为网络的输入, 因此对原始数据进行光流计算是微表情分析中必须要掌握的技术. 目前已有很多光流计算方法被提出, 如 Lucas-Kanade[42]、Farnebäck's[43]、TV-L1[44]和 FlowNet[45], 其中 TV-L1 光流在微表情分析领域最为常用, 它通过最小化总变差项来优化光流场, 具有对光照变化和噪声干扰的鲁棒性, 能处理大运动和快速运动并提供高精度稠密运动信息.

### 1.4 总 结

本节系统梳理了微表情数据集的核心资源、构建逻辑与关键预处理技术, 为后续介绍深度学习驱动的微表情分析奠定数据基础. 现有微表情数据集通过高帧率采集与 FACS 标准化编码保障数据质量, 近年来呈现大规模化、多模态化以及高生态效度化的发展趋势, 但仍普遍存在样本量稀缺、场景单一等问题. 数据集构建整体遵循诱发-采集-标注的三阶段流程, 高专业门槛且耗时耗力, 导致当前微表情数据集仍存在规模有限、质量不均等问题, 因此数据集的发展完善仍是未来微表情分析领域需持续突破的关键方向.

## 2 基于深度学习的微表情识别

早期的微表情识别研究主要基于传统的手工特征提取方法[46–48], 然而这些方法难以识别面部肌肉运动的细微

变化, 后续改进也存在较大限制. 随着深度学习的发展, 能够更好地提取图像高级特征信息的卷积神经网络逐渐成为主流, 自 2016 年左右, 逐渐出现一些工作将深度学习的方法引入微表情识别[49−51], 取得了一些令人鼓舞的进展, 但早期微表情识别任务缺乏统一的评估标准和验证策略, 无法客观的比较不同方法之间的性能优劣, 领域发展受到较大限制.

第 2 届人脸微表情挑战赛 MEGC 2019[52]上, 主办方按照消极 (抑制、愤怒、蔑视、厌恶、恐惧、悲伤)、积极 (高兴) 和惊讶 (惊讶) 方式统合了 SAMM[31]、SMIC[33]和 CASME II[27]这 3 个数据集的数据类别构建混合数据集, 并为微表情识别奠定了一个较为合理且统一的评估方式, 自此之后的微表情识别工作大都沿用本次挑战赛的评估方式进行验证. 基于上述原因, 本文主要关注 MEGC 2019 之后出现的工作.

### 2.1 基于分类监督的微表情识别

微表情识别本质上是一个分类任务, 即预测给定微表情的情感类别, 因此依赖预测值和真实标签计算交叉熵损失进行监督学习是最传统且普遍的做法, 这类工作通常更侧重于模型结构的改进. Verma 等人[53]提出 LEARNet 网络, 该网络通过横向连接获得多样性特征并结合累积层形成混合特征, 通过融入卷积层之间的交叉解耦关系保留微小但具有影响力的面部变化信息. Li 等人[54]提出了一种肌肉运动引导的双分支网络 MMNet, 该网络的主分支通过起始帧和顶点帧的帧差来提取运动特征, 轻量化的子分支以起始帧作为输入生成面部位置嵌入配合主分支将运动特征映射到特定面部区域. Wang 等人[55]提出 HTNet, 他们将面部划分为 4 个不同区域 (左唇区、左眼区、右眼区和右唇区), 并根据这 4 个区域分别分块, 利用 Transformer 提取 4 个区域特征并逐层聚合. Cai 等人[56]发现, 光流信息本身就是一种运动显著性的代表, 因此设计出光流驱动注意力网络 MFDAN, 该网络通过光流驱动注意力, 从而自适应的增强微表情图像上感兴趣区域的特征表示. Zhang 等人[57]则将通过顶点帧和起始帧计算的光流图均匀分块, 将每块视为一个独立节点, 利用图卷积网络提取图像特征. 还有工作借助 CAS(ME)$^3$ 提供的深度信息[58], 构建三维人脸结构, 提出一种轻量级点云图卷积神经网络 Lite-Point-GCN, 整合面部结构和运动特征, 在 CAS(ME)$^3$ 数据集上取得了 SOTA 的性能.

AU 代表了面部局部区域运动且和情感类别有一定关联, 因此成为面部表情分析中至关重要的信息, 有许多工作专门致力于 AU 检测或强度估计[59−61]. 在微表情识别中, 一些工作将 AU 信息引入来提高识别准确率. Lo 等人[62]提出 MER-GCN, 使用卷积神经网络提取 AU 的局部特征, 并用图卷积神经网络提取 AU 的图结构特征. Xie 等人[63]将 AU 识别与微表情识别相结合, 基于人脸肌肉运动关系设计 AU-GACN 网络提取显著的 AU 特征, 此外他们还提出一种基于生成对抗网络的微表情样本生成方法, 该方法能够利用不同动作单元的不同强度合成微表情训练样本. 然而这两项工作要么仅能学习到静态的 AU 模式[62], 要么无法提取时空特征[63], 因此 Wang 等人[64]使用多帧光流和可变形卷积提取时空特征, 将得到的 AU 识别结果利用 Transformer 获取关联信息, 并通过图卷积神经网络提取最终的 AU 特征, 最终将时空特征和 AU 特征融合以进行微表情识别. Lei 等人[40]将起始帧和顶点帧通过 MagNet 运动放大得到的中间层特征作为网络输入, 从而减少运动放大带来的噪声影响, 利用面部关键点构建图结构, 提出 Graph-TCN 网络分别提取边特征和节点特征从而进行分类. 次年, 他们提出 AU-GCN 网络[65], 分别使用可变形卷积 DConv 和 Transformer 提取节点特征和边特征, 又利用图卷积神经网络将 AU 信息融入, 进一步提高识别准确率.

结合了 AU 信息的方法虽然已取得了令人鼓舞的效果, 但需要注意的是, 微表情中 AU 的组合与情感类别之间没有明确的对应关系, 如后文图 4 所示, 尽管同样的 AU 在视觉上拥有相似的面部运动, 但可能对应完全不同的情感类别, 盲目地引入 AU 信息是否会对模型学习产生错误的引导从而影响泛化能力值得深思. 我们认为, 未来需要更加灵活的 AU 信息处理方式以及更大规模且高质量的数据才能真正捕获到 AU 组合和微表情类别的关联, 以辅助进行更好的微表情分析研究.

### 2.2 基于对比学习的微表情识别

对比学习从特征层面入手, 通过对比损失使同类样本的特征更相似, 不同类样本特征更相异. 在没有标签的情况下, 正样本通常由目标样本的数据增强版本构成, 并将其余所有样本视为负样本[66]. 2020 年, Khosla 等人[67]提

出监督对比学习, 利用数据的标签将同一类中的不同样本也视为正样本, 使得同类样本具有更紧密的特征分布. 由于微表情数据集样本数量有限且类别较少, 采用监督对比学习方法可以有效防止样本被误判, 因此 Zhi 等人[68]将监督对比学习引入微表情识别, 并使用双端光流采集来扩充数据集, 将起始帧-顶点帧光流和结束帧-顶点帧光流都作为有效样本. Fang 等人[69]在此基础上进一步研究, 使用 COC 网络更好的根据光流信息获得灵活的感受野, 使得来自同一部位的运动信息更容易被分组到同一集合中进行计算和更新, 并引入原型损失使同类样本更合理的收敛到相同特征域从而减轻样本类别不均衡为训练带来的不利影响.


图 4 相同 AU 标注对应不同情感类别

为了缓解微表情样本数量少的问题, Song 等人[70]提出了一种基于属性信息和跨模态对比学习的微表情识别方法, 模型框架如图 5 所示, 除了图像信息特征之外, 他们还利用 BERT 提取文本特征, 通过跨模态对比损失将 FACS 中 AU 标签代表的文本属性信息作为辅助模态特征嵌入到视觉网络, 从而提高微表情识别表征能力. 而 Wei 等人[71]提出了一种新颖的对比放大网络, 将同一视频序列中的不同帧视为负样本, 试图基于帧之间的强度差异学习出微表情视频序列中的强度线索, 并根据其与面部纹理特征序列逐元素相乘得到强度增强后的特征, 最后通过 LSTM 聚合整个序列的特征用于分类. 为了更好地利用先验知识, Ma 等人[72]模拟人类认知发展, 模拟从简单到复杂的学习过程, 利用进行过样本平衡的运动放大微表情数据集训练一个高级特征编码器, 使其能够提取高级微表情特征, 为模型提供稳健的初始化, 然后在元学习框架[73]内微调整合编码器特征, 对微表情进行分类验证.

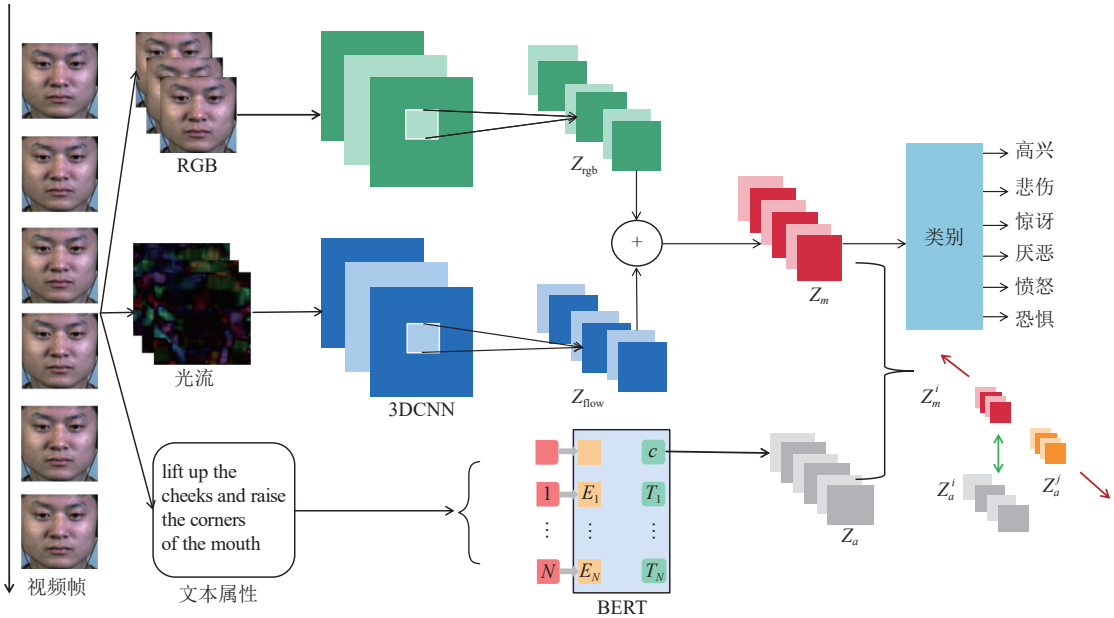



图 5 属性信息嵌入与跨模态对比学习模型框架[70]

### 2.3 基于生成式自监督的微表情识别

Transformer、BERT 等生成式预训练模型在自然语言处理取得重大成就[74−77], 这也对视觉领域产生了巨大影

响, 大量的工作使用生成式自监督学习来提升视觉任务性能[78−81]. 自监督预训练是一种解决特定领域数据匮乏的可行方法, 微表情识别领域也自然出现了一些利用了生成式自监督学习思想的工作, 这些工作往往设计生成式任务驱动自监督预训练, 从而得到更加泛用且鲁棒的微表情识别解决方案.

Nguyen 等人[82]提出了 Micron-BERT 模型, 该模型通过块注意力以及对角微注意力模块自动关注到感兴趣的面部细微变化并减少无关背景信息对性能的影响, 在预训练阶段利用 CAS(ME)$^3$ 数据集[29]构建包含 800 万帧的无标签训练数据集, 随机交换临近两帧的某些块, 并让网络根据交换后的图像生成原始图像, 临近的两帧往往差距细微, 这就使得模型的编码器拥有捕获细微面部变化的能力, 使用模型预训练后的编码器进行微表情识别任务取得了先进的性能.

利用生成式自监督方法, Nguyen 等人[82]将有效数据扩展到了上百万帧的无标签数据, 利用这些数据模拟微表情的细微变化来辅助编码器训练. 而一些研究人员从另一个角度入手, 通过生成式自监督让模型获得自动捕获微表情动态特征的能力.

大部分微表情识别工作默认采用光流代表动态特征作为模型输入, 为了消除对光流计算的依赖并获得更加适合微表情的高质量动态特征, Fan 等人[83]提出名为 SelfME 的自监督学习框架, 通过运动模式学习使得模型可以生成两帧之间的运动场, 根据得到的运动场进行微表情识别分类, 具体而言在运动场生成预训练阶段, 他们通过给定的源帧重建目标帧来形成一个生成式自监督任务, 从而隐式的学习运动特征, 此外他们根据面部动作的对称性开发对称对比视觉 Transformer, 以约束模型更好的学习面部运动信息. Zhai 等人[84]也提出 FRL-DGT, 通过生成式任务驱动位移生成模块网络预训练, 根据起始帧和峰值帧生成自适应的动态位移特征来摆脱对光流计算的依赖, 并在分类阶段结合全脸特征提取和基于 AU 的局部特征提取以进行微表情分类, 网络结构如图 6 所示.

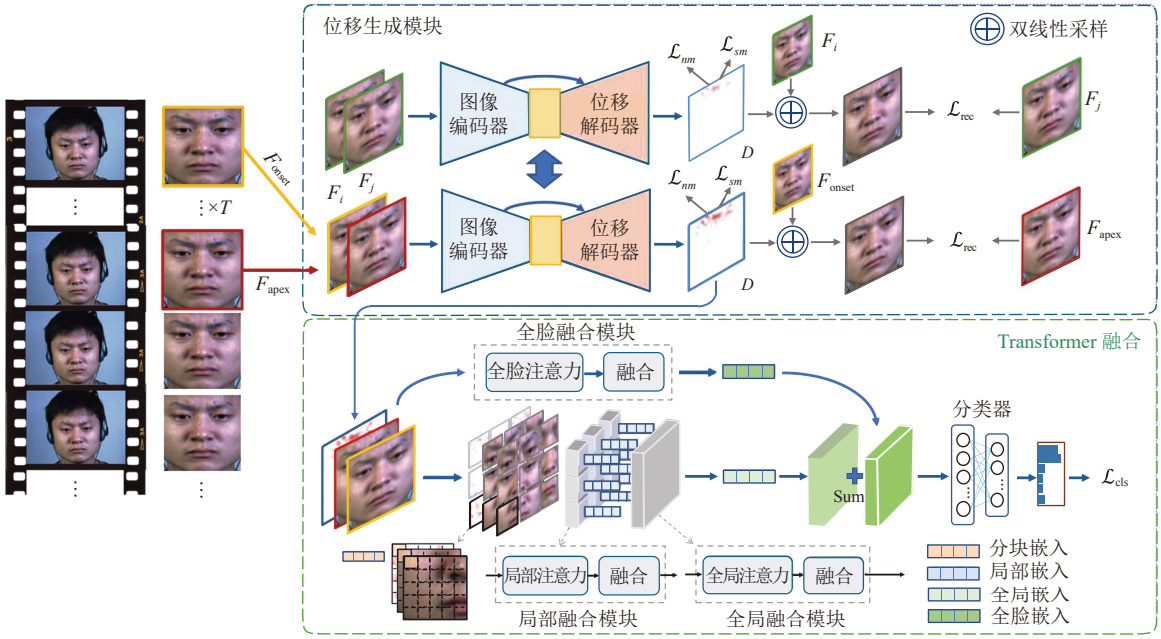



图 6 FRL-DGT 网络结构[84]

然而, 上述两种方法仍需要依赖微表情发生序列中的关键帧才能进行动态特征提取, Zhang 等人[85]提出了一种无需顶点帧标注、直接从发生序列提取动态特征的端到端微表情识别方法 SODA4MER, 通过自监督学习构建了面向定向局部形变的表征能力, 有效挖掘面部肌肉群之间的动态结构关系, 随后在预训练模型的基础上, 结合动力定型理论[86], 引导模型从人类情感加工机制角度建模情绪表达过程, 实现更贴近人类认知规律的微表情识别.

### 2.4 基于宏表情知识迁移的微表情识别

相较于微表情, 宏表情的数据采集标注更加容易, 有大量的开源数据集支持研究工作[87−92], 并且两者在面部行为上存在一定的相似之处, 因此有大量的工作致力于跨越语义差异, 将宏表情相关知识迁移到微表情识别.

探索深度学习在微表情识别上的应用初期时, Patel 等人[49]就探索迁移学习在微表情识别任务上的应用, 他们将在 ImageNet[93]和宏表情数据集 CK+[88]预训练的 VGG 网络迁移到小数据量的微表情识别任务上, 但整体模式较为简单. Liu 等人[94]则认为在 ImageNet 这种充斥着大量面部无关数据的数据集上进行预训练并不合理, 因此选择了 4 个宏表情数据集[88−91]样本作预训练, 在模型上采用了更加轻量级的视觉注意力网络 MobileVit[95]作为面部特征提取器.

宏表情与微表情在视觉表现上的主要区别在于运动强度差异, 这使得直接从宏表情中获取的知识难以有效迁移到微表情识别[14]. Liu 等人[96]认为微表情的顶点帧是宏表情不可避免的一个中间过程, 基于此假设提出表情放大缩小 (expression magnification and reduction, EMR), 将宏表情起始帧和顶点帧中的中间帧视为宏表情的缩小, 对微表情进行运动放大作为微表情的放大从而使得宏表情微表情有着更相似的运动强度, 结合文献 [97] 中提出的领域对抗自适应技术, 更好的弥合了宏表情和微表情之间的差异, 然而采用运动放大会为微表情引入大量噪声, 直接取宏表情的中间帧也无法做到均匀合理, 因此无法保证宏表情和微表情的相似性. Wang 等人[98]提出一种从宏表情到微表情的渐进式学习方法 (progressive learning from macro-expressions to micro-expressions, PLMaM), 让模型逐步学习不同强度级别的数据, 并使用自知识蒸馏模型使得每一步学习的知识可以转移到下一步. 此外, Xia 等人[99]利用两个不同的网络分别解耦微表情和宏表情中与表情相关的特征, 并引入对比损失和三元组损失引导网络提取出表情样本的共享特征辅助微表情识别.

然而上述采取宏表情辅助的方法往往欠缺对面部运动的时间特征捕捉, 为此 Xia 等人[100]在之前工作的基础上, 对宏表情和微表情的时空域的共享特征建模, 不仅可以捕获面部外观的静态纹理特征, 也可以学习出面部肌肉的动态时间特征. 具体而言, 他们分别针对宏表情和微表情预训练了一个双流模型, 并引入领域判别器和关系分类器, 领域判别器用于对齐宏表情和微表情特征, 关系分类器用于预测不同采样间隔的时间特征关系, 从而更好地辅助时空特征提取.

### 2.5 总 结

基于深度学习的微表情识别近年来已形成分类监督、对比学习、生成式自监督和宏表情知识迁移这 4 大主流技术路径. 这些方法围绕微表情分析中“数据稀缺”和“细微时空特征难捕捉”两大核心难点持续演进, 构建了从基础框架到进阶优化的完整研究体系, 我们将代表性方法进行整合, 绘制如图 7 所示的时间轴以便读者快速了解微表情识别研究发展的时间脉络.

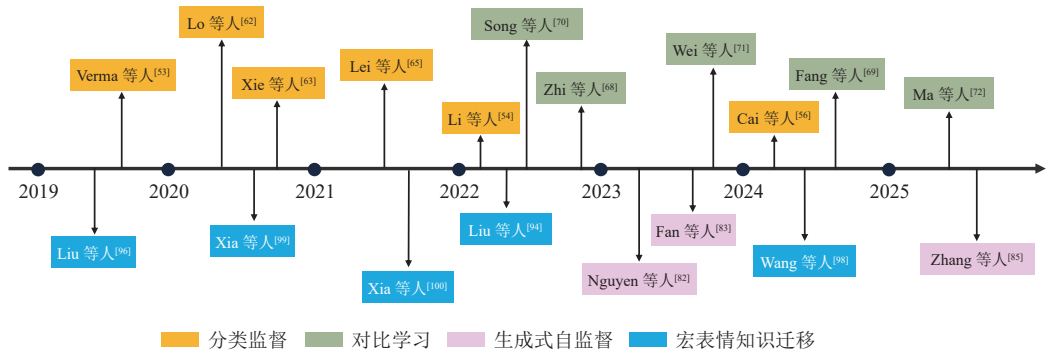



图 7 基于深度学习的微表情识别研究发展脉络

分类监督方法作为技术基石, 以情感类别预测为核心目标, 聚焦模型结构创新与动作单元 (AU) 信息深度融合, 强化对局部肌肉运动特征的捕捉, 不过其性能高度依赖标注数据规模, 泛化能力受限于数据质量. 对比学习方

法尝试从更多模态以及特征表征层面突破数据瓶颈, 通过监督对比、跨模态对比等设计, 让同类样本特征更聚合、异类样本特征更离散, 同时结合光流扩充、原型损失等策略, 有效缓解样本量少、类别不均衡带来的训练难题, 进一步提升模型对细微差异的区分度. 生成式自监督方法则开辟了数据高效利用的新路径, 借助预训练模型的强大泛化能力, 通过无标签数据的自监督任务 (如帧重建、动态位移生成) 挖掘深层特征, 提升模型的泛用性与鲁棒性. 宏表情知识迁移方法则充分利用宏表情海量标注数据的优势, 通过领域自适应、渐进式学习等技术, 弥合宏表情与微表情在运动强度上的差异, 提取两类表情的共性特征, 为微表情识别补充丰富的先验知识, 弥补微表情样本缺口.

需要注意的是, 这 4 大主流研究路径各有侧重, 但并非孤立发展, 近年来的一些研究已经呈现互补融合的趋势[72,85,98], 并将研究重点逐步转移到了解决真正应用瓶颈而非单纯的算法创新, 这种技术协同推动微表情识别持续向数据利用高效化和特征提取精准化的方向迈进, 为其在实际场景中的落地奠定了坚实基础.

## 3 基于深度学习的微表情检测

与微表情识别相比, 当前微表情分析中对微表情检测的重视程度严重不足, 但在实际应用中检测为后续微表情分析提供可靠信息, 是极为重要的一步. 微表情采集时参与者不可避免地会伴随一些表情无关运动 (如眨眼、吞咽、头部转动等), 这些动作在长视频中会累计更大的误差, 因此微表情检测相较于识别面临更大的挑战. 在早期研究中, 微表情检测通常采用传统的手工特征进行差异分析, 常见的包括局部二值模式[101,102]、光流[103−105]、方向梯度直方图[106]等, 尽管基于传统手工特征的方法取得了一定成效, 但整体泛化能力差、可拓展性低制约了其发展, 目前的研究重点逐步转向深度学习. 需要注意的是, 微表情检测通常包含两个概念容易被混淆的工作: 区间检测和顶点检测, 其中区间检测旨在从原始长视频片段中检测出微表情发生的区间, 而顶点检测旨在检测出微表情的顶点帧. 此外, MEGC2020[107]提出了一个新的挑战: 同时检测出视频序列中的宏表情和微表情, 这与宏表情和微表情交织的现实场景更加相符. 综上所述, 本节将从区间检测、顶点检测以及宏表情+微表情检测这 3 个方面对近几年采用深度学习的代表性工作进行介绍.

### 3.1 区间检测

Verburg 等人[108]使用光流向量计算方向光流直方图并通过一个由长短期记忆网络单元组成的循环神经网络来学习区分相关和不相关的微表情, 网络接受由滑动窗口生成的包含 25 个时间步的视频序列并判断该序列是否存在微表情, 最后通过非极大值抑制对候选区间进行后处理. Takalkar 等人[109]提出一种名为 LGAttNet 的注意力驱动检测网络, 网络整体架构如图 8 所示, 使用全局注意力流和局部注意力流的双流注意力网络来构建微表情检测模型, 全局注意力专注于整脸特征, 而局部注意力则聚焦于局部区域内的微小肌肉运动, 数据在输入网络前被预处理分成眼部区域和嘴部区域两部分, 这两部分被分别送入不同的局部注意力模块, 整脸图像则被送入全局注意力模块, 检测模块由 3 个全连接层组成, 根据局部和全局的特征组合将单个帧分类成是否为微表情帧, 该方法兼顾了全局和关键局部特征, 但仅独立的对单帧图像进行特征提取和分类, 忽略了微表情序列中的时序运动信息, 并且这项研究没有对预测结果进行区间整合.

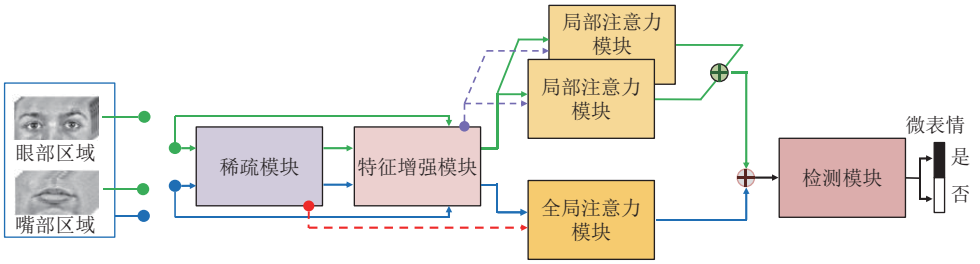



图 8 LGAttNet 模型框架[109]

上述工作通过滑动窗口的方法逐个分类短片段, 或者仅对每个帧进行分类, 并没有真正的从原始长视频中端到端的生成微表情区间, 受到一些目标检测[110,111]和时序动作定位[112,113]工作的启发, Wang 等人[114]提出 MESNet,

可以直接从原始的长视频中预测微表情所在区间, MESNet 由 3 个主要模块组成, 第 1 个模块通过 2D CNN 和 1D CNN 的组合从视频序列中提取空间特征和时间特征, 这样设计可以比 LSTM 有着更少的参数量, 对小样本任务更加友好; 第 2 个模块是时段提案网络 (clip proposal network, CPN), CPN 负责提供一些微表情候选片段, 该模块接收第 1 个模块的输出特征, 并通过 5 个不同感受野的卷积分支来输出 5 个不同尺度的时段提案, 与提案对应的时空特征将被输入到第 3 个模块; 第 3 个模块是一个分类回归网络 (classification regression network, CRN), CRN 将与提案对应的特征分类为微表情或非微表情, 并进一步回归属于微表情提案的时间边界. Guo 等人[115]使用 I3D 作为主干网络提取光流和原始图像视频特征, 然后使用多尺度 Transformer 捕获帧之间的时序相关性, 获得对面部运动更具代表性的特征以生成多尺度特征, 最后通过多头估计器对特征进行预测和回归.

运动放大在微表情识别中应用较广, 但长视频中应用运动放大技术会同时放大真实峰值和虚假峰值, 带来更大的干扰, 因此在微表情检测中的应用并不广泛. 为了应对上述问题, Zhou 等人[116]提出了小波卷积放大网络 WCMN, 以提高长视频中微表情检测的性能, 该模型结合了小波分解和一个可学习的运动放大模块, 以从特征层面增强微表情的光流, 具体来说, WCMN 利用离散小波变换 (discrete wavelet transform, DWT)[117]将视频的光流序列分解为 1 个低频分量和 3 个高频分量, 计算注意力得分, 并通过一个阈值来筛选放大注意力得分, 最终通过逐元素相乘实现特征放大.

### 3.2 顶点检测

Zhang 等人[118]首次将深度学习应用到微表情顶点检测, 提出 SMEConvNet, 为了增加峰值帧的训练样本量, 作者在训练时将微表情发生片段的每一帧都视为峰值帧, SMEConvNet 通过 CNN 组合全连接层的架构提取视频每一帧的特征得到维度 $X\times Y$ 的特征矩阵 $F$, 其中 $X$ 为视频帧数, $Y$ 为每帧的特征向量维度, 在后处理阶段, 将 $F$ 的每一行的元素与第 1 行中对应的元素相减并求平方和, 以得到 $X\times 1$ 大小的矩阵 $A$, 然后利用滑动窗口在 $A$ 以获得每个滑动窗口中所有值的和, 当滑动窗口与微表情片段重合时, 应得到最大的和值, 此时滑动窗口内的最大值即认为是顶点帧. Tran 等人[119]在视频的所有位置上滑动一个扫描窗口, 利用 LSTM 接收这些滑动窗口的特征, 并对每个滑动窗口进行预测得到是否为微表情的预测得分, 顶点帧由滑动窗口的中间位置确定. Yee 等人[120]将注意力机制引入, 以自适应地找到对微表情信息更为敏感的区域, 并为这些分配更多权重.

Liu 等人[121]提出了一种基于单分类理论 (one-class classification, OCC) 的长视频微表情检测方法, OCC 常应用于两类样本极端不均衡的情况, 仅使用易获得的正类样本训练模型, 通过特征工程以检测或识别离群的负类样本. 在微表情顶点检测中作者将长视频中占绝大多数的中性表情帧视为正类, 将微表情帧视为负类, 并仅使用正类样本进行训练, 模型输出异常分数, 最后将滑动窗口应用于异常分数序列, 通过极值分析定位微表情顶点帧.

### 3.3 宏表情+微表情检测

在现实场景中, 微表情常和宏表情交织在一起出现, 并且对于面部表情分析, 宏表情和微表情都极为重要. 因此为了应对更加复杂的真实场景, MEGC2020[107]提出新的挑战, 在长视频中同时检测宏表情和微表情发生区间, 该项挑战在 MEGC2021–2024[122−125]得以延续, 诞生了大量优秀的研究成果, 并且受此影响, 后续出现的大部分检测工作会同时对宏表情和微表情进行检测.

Pan 等人[126]提出了一种基于局部双线性卷积神经网络 (local bilinear convolutional neural network, LBCNN) 的方法, 将长视频序列中的微表情检测转换成细粒度图像识别问题, 利用卷积网络将每一帧人脸的 4 个局部区域 (左眉、右眉、鼻子、嘴巴) 和整脸进行特征提取, 并分类预测成微表情、宏表情或中性表情. 但该方法单独对每一帧进行特征提取和分类, 仅考虑了空间信息而忽略了时序信息. Yu 等人[127]提出一种基于位置抑制的检测网络 LSSNet, 该网络使用 I3D[128]模型作为主干网络提取视频和光流序列特征, 并使用特征金字塔生成候选框, 最后通过位置抑制模块减少间隔过长和过短的候选框. Yang 等人[129]提出了一种关注帧间 AU 相关性的端到端检测框架, 使用 OpenFace 2.0[130]逐帧进行 AU 检测, 将检测结果作为网络输入的一部分, 该框架不局限于特定深度学习模型, 可以适配各种主干网络以探索最优性能, 需要注意的是 OpenFace 等开源库中提供的相关工具均通过宏表情数据进行训练, 在使用这些工具辅助进行微表情特化任务时可能无法取得理想的效果. Leng 等人[131]提出顶点与边界

感知网络 ABPN 来同时检测顶点帧和边界帧, ABPN 由 3 个模块组成, 视频编码模块通过计算感兴趣区域的光流以表示视频特征; 概率评估模块负责评估每帧是否在表情区间内、是否为顶点帧或边界帧等, 该模块以一维卷积为主; 候选区间生成模块整合概率评估模块的预测结果生成候选区间, 该方法在主要特征提取上仍采用手工方法计算.

为了缓解数据严重不足带来的负面影响, 如图 9 所示, Xu 等人[132]提出了一种将预训练模型和光流法相结合的方法, 使用一个名为 VoxCeleb2[133]的大型视听说话人识别数据集对 VideoMAE[134]进行自监督预训练, 并使用宏表情和微表情数据做微调, 在后处理阶段使用光流法[135]输出的表情区间对结果进行联合矫正, 并且使用 Dlib 眨眼检测来排除眨眼造成的干扰, 该方法通过自监督预训练一定程度上缓解了数据不足带来的影响, 并且结合手工特征方法的优势来优化效果取得了出色的结果, 获得了 MEGC2023 的冠军. Zhang 等人[136]同样使用经过大规模预训练的 VideoMAE 做视频特征提取器, 将提取的特征通过特征金字塔生成多尺度候选片段特征以预测每一帧微表情和宏表情的概率以及该帧到区间前后边界的距离, 使用 Soft-NMS[137]对候选片段进行初步滤波, 并且提出了一种多起点光流滤波的方法对候选片段进一步细化处理以得到最终预测结果.

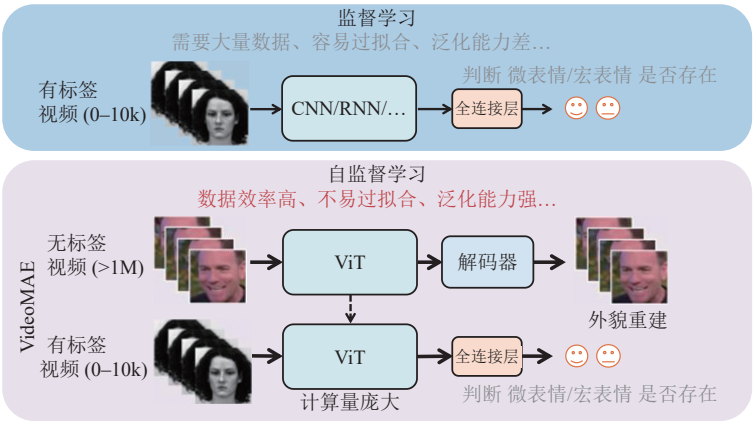



图 9 基于 VideoMAE 预训练的宏表情+微表情检测模型框架[132]

### 3.4 总 结

微表情检测作为微表情分析的关键前置环节, 核心目标是从长视频序列中精准定位微表情的发生区间或顶点帧, 区间检测聚焦微表情时段定位, 通过对时段进行特征提取, 结合时段提案与分类回归实现端到端检测, 顶点检测则侧重挖掘帧间特征差异. 由于受到 MEGC2020 的影响, 大多研究者开始同时对宏表情和微表情进行检测, 这更加贴合宏、微表情交织的现实场景, 我们将代表性方法总结到图 10 的时间轴上, 以便读者快速掌握发展的时间脉络.

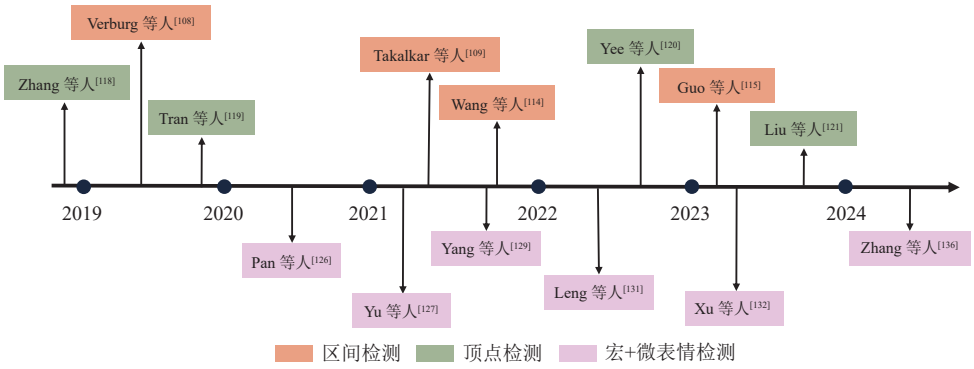



图 10 基于深度学习的微表情检测研究发展脉络

需要注意的是, MEGC 挑战赛中出现了大量优秀的基于深度学习的方法, 但这些深度学习方法在指标上与同期参赛的传统手工特征方法不分伯仲[10,104,105,135], 我们认为是过少的样本量使其无法通过简单的监督学习获得优异性能, 并且基于深度学习的微表情检测整体仍处于研究初期, 在模型架构和方法上仍没有较为成熟统一的方案, 这表示基于深度学习的微表情检测仍存在巨大发展潜力.

## 4 评估指标

### 4.1 微表情识别评估方法

微表情识别任务中常用交叉验证作为评估协议. 在交叉验证中, 数据集被划分为多个子集, 每次取不同的子集作为测试集进行评估, 其中在微表情识别中留一主体交叉验证 (leave-one-subject-out, LOSO) 和 K 折交叉验证使用最为广泛.

由于微表情具有较大的个体差异性以及较小的数据规模, LOSO 是目前微表情识别领域最为普遍使用的评估协议, 以不同被试者作为数据集划分子集的依据, 每次以某个被试者的数据作为测试集, 其余数据作为训练集, 最终统合所有结果作为最终指标, 这样做可以确保每个被试者样本均被独立的评估, 避免受试者偏差, 但这意味着每个不同的被试者都要做一次实验, 会极大增加实验的时间开销. K 折交叉验证中, 原始数据集被随机划分为 $K$ 个大小相等的子集, 依次将每个子集作为测试集, 其余部分作为训练数据. 在未来会有更大规模的数据集出现, 尤其是被试者数量多的情况下, 采用 LOSO 会导致过多的时间消耗, 这时可以采用 K 折交叉验证.

微表情识别本质上是一个多分类问题, 因此当前通常采用精度 (accuracy, $ACC$)、未加权 $F1$ (unweighted $F1$-score, $UF1$)、未加权平均召回率 (unweighted average recall, $UAR$) 进行评价, 早期的工作多以 $ACC$ 作为评估指标, 这可以直观地反映出预测的准确率, 但在样本类别严重不均衡的情况下即便全部预测成为多数类也能取得较高的指标, 不能准确反映方法效果, $UF1$ 和 $UAR$ 成为更受欢迎的评估指标, 一般情况下这 3 个指标越高代表识别效果越好. 上述评估指标计算方法如公式 (1)–(3) 所示:

$$ACC = \frac{\sum_{c} TP_c}{N} \tag{1}$$

$$UF1 = \frac{1}{C}\sum_{c}\frac{2\times TP_c}{2\times TP_c + FP_c + FN_c} \tag{2}$$

$$UAR = \frac{1}{C}\sum_{c}\frac{TP_c}{N_c} \tag{3}$$

其中, $C$ 代表多分类任务的类别数, $N_c$ 为 $C$ 类别的样本总量, $TP$ 为真正例, $FP$ 为假正例, $FN$ 为假负例.

### 4.2 微表情检测评估方法

微表情区间检测方法预测值通常是一个区间, 要判断该区间是否预测正确, 通常计算预测区间和真实区间的交并比 (intersection over union, $IoU$), 假设预测区间为 $W_{\text{spotted}}$, 真实区间为 $W_{\text{groundTruth}}$, 通过计算两者交集的长度和并集的长度之比得到 $IoU$, 具体计算方法如公式 (4) 所示:

$$IoU = \frac{W_{\text{spotted}} \cap W_{\text{groundTruth}}}{W_{\text{spotted}} \cup W_{\text{groundTruth}}} \tag{4}$$

当 $IoU$ 大于等于预设阈值 (通常为 0.5) 时, 该预测区间被判定为 $TP$, 否则判定为 $FP$, 此外若真实存在的微表情区间未被模型检测到时, 记为 $FN$. 在此基础上通过查准率 ($Precision$)、召回率 ($Recall$) 和 $F1$ 分数对方法进行评估, 通常这 3 个指标越高代表方法效果越好, 计算方法如公式 (5)–(7) 所示:

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1 = \frac{2\times Precision\times Recall}{Precision + Recall} \tag{7}$$

目前宏表情+微表情检测中需要统合宏表情检测结果和微表情检测结果来得到综合评价, 其中 $TP$ 和 $FP$ 等的计算方法与前文区间检测一致, 假设数据集中包含有 $V$ 段视频, 其中有 $M_1$ 段宏表情序列及 $M_2$ 段微表情序列, 对应有 $N_1$ 个宏观表情区间和 $N_2$ 个微表情区间, 总共有 $A_1$ 个宏观表情的真正例和 $A_2$ 个微表情的真正例. 可将数据集中的所有视频视作一个长视频, 由此可确定所有视频的整体性能, 评估过程中先分别计算宏表情检测与微表情检测的各项指标, 再综合得出整个数据集的总体评估结果.

对于宏表情, 指标计算方法如公式 (8)、(9) 所示:

$$Recall_{\mathrm{MaE}} = \frac{A_1}{M_1} \tag{8}$$

$$Precision_{\mathrm{MaE}} = \frac{A_1}{N_1} \tag{9}$$

对于微表情, 指标计算方法如公式 (10)、(11) 所示:

$$Recall_{\mathrm{ME}} = \frac{A_2}{M_2} \tag{10}$$

$$Precision_{\mathrm{ME}} = \frac{A_2}{N_2} \tag{11}$$

整体评估指标计算方法如公式 (12)、(13) 所示:

$$Recall_D = \frac{A_1 + A_2}{M_1 + M_2} \tag{12}$$

$$Precision_D = \frac{A_1 + A_2}{N_1 + N_2} \tag{13}$$

同样地, $F1$ 可以采用公式 (7) 进行计算.

## 5 算法性能

### 5.1 微表情识别

我们将前文所述的代表性微表情识别算法汇总在表 3 中, 表中使用年份升序, 列出了模型输入、使用的评估协议以及分类类别数等重要参数, 展示了不同数据集中各方法的性能.

表 3 微表情识别算法性能

| 数据集 | 方法 | 年份 | 模型输入 | 评估协议 | 类别数 | 实验指标 | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | *ACC* | *UF*1 | *UAR* |
| SMIC | LEARNet[53] | 2019 | 顶点帧 | LOSO | 2<br>5 | 0.9109<br>0.8160 | —<br>— | —<br>— |
| | Liu等人[96] | 2019 | 光流+顶点帧 | LOSO | 3 | — | 0.7461 | 0.7530 |
| | EIDNET[99] | 2020 | 序列 | LOSO | 3 | — | 0.8640 | 0.8610 |
| | AU-GCN[65] | 2021 | 起始帧+顶点帧+AU | LOSO | 3 | — | 0.7192 | 0.7215 |
| | Xia等人[100] | 2021 | 序列 | LOSO | 3 | — | 0.8730 | 0.8670 |
| | MobileViT[94] | 2022 | 序列 | LOSO | 3 | — | 0.7356 | 0.7141 |
| | CMNet[71] | 2023 | 序列 | LOSO | 3 | 0.7927 | 0.8011 | — |
| | $\mu$-BERT[82] | 2023 | 起始帧+顶点帧 | LOSO | 3 | — | 0.8550 | 0.8384 |
| | SelfME[83] | 2023 | 序列 | LOSO | 3 | — | 0.6972 | 0.7012 |
| | FRL-DGT[84] | 2023 | 光流+AU | LOSO | 3 | — | 0.7430 | 0.7490 |
| | PLMaM-Net[98] | 2024 | 序列 | LOSO | 3 | 0.8232 | 0.8221 | — |
| | HTNet[55] | 2024 | 光流+顶点帧 | LOSO | 3 | — | 0.8049 | 0.7905 |
| | MFDAN[56] | 2024 | 光流+顶点帧 | LOSO | 3 | — | 0.6815 | 0.7043 |

表 3 微表情识别算法性能 (续)

| 数据集 | 方法 | 年份 | 模型输入 | 评估协议 | 类别数 | 实验指标 | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | *ACC* | *UF*1 | *UAR* |
| SMIC | OFVIG-Net[57] | 2025 | 光流 | LOSO | 3 | — | 0.6435 | 0.6400 |
| | MPFNet[72] | 2025 | 起始帧+顶点帧 | LOSO | 3 | 0.7860 | 0.7870 | 0.7830 |
| | SODA4MER[85] | 2025 | 序列 | LOSO | 3 | — | 0.8855 | 0.8881 |
| CASME II | Liu等人[96] | 2019 | 光流+顶点帧 | LOSO | 3 | — | 0.8293 | 0.8209 |
| | LEARNet[53] | 2019 | 顶点帧 | LOSO | 7 | 0.7657 | — | — |
| | MER-GCN[62] | 2020 | 序列 | k-fold | 7 | 0.5882 | — | — |
| | | | | LOSO | | 0.4271 | — | — |
| | AU-GACN[63] | 2020 | 序列 | LOSO | 3 | 0.7120 | 0.3550 | — |
| | | | | | 5 | 0.5610 | 0.3940 | — |
| | Graph-TCN[40] | 2020 | 顶点帧 | LOSO | 5 | 0.7398 | 0.7246 | — |
| | EIDNET[99] | 2020 | 序列 | LOSO | 3 | — | 0.8700 | 0.8720 |
| | AU-GCN[65] | 2021 | 起始帧+顶点帧+AU | LOSO | 3 | — | 0.8798 | 0.8710 |
| | | | | | 4 | 0.8080 | 0.7871 | — |
| | | | | | 5 | 0.7427 | 0.7047 | — |
| | Xia等人[100] | 2021 | 序列 | LOSO | 3 | — | 0.8810 | 0.8810 |
| | MMNet[54] | 2022 | 起始帧+顶点帧 | LOSO | 3 | 0.9551 | 0.9494 | — |
| | | | | | 5 | 0.8835 | 0.8676 | — |
| | Song等人[70] | 2022 | 序列 | LOSO | 5 | 0.7782 | — | — |
| | MER-Supcon[68] | 2022 | 起始帧+顶点帧+结束帧 | LOSO | 3 | 0.8965 | 0.8806 | — |
| | | | | | 5 | 0.7358 | 0.7286 | — |
| | MobileViT[94] | 2022 | 序列 | LOSO | 3 | — | 0.6997 | 0.7251 |
| | CMNet[71] | 2023 | 序列 | LOSO | 5 | 0.7805 | 0.7399 | — |
| | $\mu$-BERT[82] | 2023 | 起始帧+顶点帧 | LOSO | 3 | — | 0.9034 | 0.8914 |
| | | | | | 5 | — | 0.8553 | 0.8348 |
| | SelfME[83] | 2023 | 序列 | LOSO | 3 | — | 0.9078 | 0.9290 |
| | FRL-DGT[84] | 2023 | 光流+AU | LOSO | 3 | — | 0.9190 | 0.9030 |
| | PLMaM-Net[98] | 2024 | 序列 | LOSO | 5 | 0.8252 | 0.7998 | — |
| | HTNet[55] | 2024 | 光流+顶点帧 | LOSO | 3 | — | 0.9532 | 0.9516 |
| | MFDAN[56] | 2024 | 光流+顶点帧 | LOSO | 3 | — | 0.9134 | 0.9326 |
| | SPCL-MER[69] | 2024 | 光流 | LOSO | 3 | 0.9452 | 0.9204 | 0.9312 |
| | OFVIG-Net[57] | 2025 | 光流 | LOSO | 3 | — | 0.7129 | 0.7195 |
| | MPFNet[72] | 2025 | 起始帧+顶点帧 | LOSO | 3 | — | 0.8790 | 0.8950 |
| | | | | | 5 | 0.8200 | 0.8190 | — |
| | SODA4MER[85] | 2025 | 序列 | LOSO | 3 | — | 0.8870 | 0.8809 |
| | | | | | 5 | 0.8418 | 0.8141 | — |
| CAS(ME)$^3$ | $\mu$-BERT[82] | 2023 | 起始帧+顶点帧 | LOSO | 3 | — | 0.5604 | 0.6125 |
| | | | | | 4 | — | 0.4718 | 0.4913 |
| | | | | | 7 | — | 0.3264 | 0.3254 |
| | Lite-Point-GCN[58] | 2025 | 起始帧+顶点帧+深度 | LOSO | 3 | — | 0.6819 | 0.7412 |
| | | | | | 4 | — | 0.4764 | 0.5366 |
| | | | | | 7 | — | 0.3564 | 0.4159 |
| | HTNet[55] | 2024 | 光流+顶点帧 | LOSO | 3 | — | 0.5767 | 0.5415 |

表 3 微表情识别算法性能 (续)

| 数据集 | 方法 | 年份 | 模型输入 | 评估协议 | 类别数 | 实验指标 | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | *ACC* | *UF*1 | *UAR* |
| SAMM | Liu等人[96] | 2019 | 光流+顶点帧 | LOSO | 3 | — | 0.7754 | 0.7152 |
| | AU-GACN[63] | 2020 | 序列 | LOSO | 3 | 0.7020 | 0.4330 | — |
| | | | | | 5 | 0.5230 | 0.3570 | — |
| | Graph-TCN[40] | 2020 | 顶点帧 | LOSO | 4 | 0.8050 | 0.7657 | — |
| | | | | | 5 | 0.7500 | 0.6985 | — |
| | EIDNET[99] | 2020 | 序列 | LOSO | 3 | — | 0.8250 | 0.8190 |
| | AU-GCN[65] | 2021 | 起始帧+顶点帧+AU | LOSO | 3 | — | 0.7751 | 0.7890 |
| | | | | | 4 | 0.8239 | 0.7735 | — |
| | | | | | 5 | 0.7426 | 0.7045 | — |
| | Xia等人[100] | 2021 | 序列 | LOSO | 3 | — | 0.8960 | 0.8840 |
| | MMNet[54] | 2022 | 起始帧-顶点帧 | LOSO | 3 | 0.9022 | 0.8391 | — |
| | | | | | 5 | 0.8014 | 0.7291 | — |
| | MER-Supcon[68] | 2022 | 起始帧+顶点帧+结束帧 | LOSO | 3 | 0.8120 | 0.7125 | — |
| | | | | | 5 | 0.6765 | 0.6251 | — |
| | MobileViT[94] | 2022 | 序列 | LOSO | 3 | — | 0.6781 | 0.7428 |
| | CMNet[71] | 2023 | 序列 | LOSO | 5 | 0.7868 | 0.7719 | — |
| | $\mu$-BERT[82] | 2023 | 起始帧+顶点帧 | LOSO | 5 | — | 0.8386 | 0.8475 |
| | FRL-DGT[84] | 2023 | 光流+AU | LOSO | 3 | — | 0.7720 | 0.7580 |
| | PLMaM-Net[98] | 2024 | 序列 | LOSO | 5 | 0.7647 | 0.6961 | — |
| | HTNet[55] | 2024 | 光流+顶点帧 | LOSO | 3 | — | 0.8131 | 0.8124 |
| | MFDAN[56] | 2024 | 光流+顶点帧 | LOSO | 3 | — | 0.7871 | 0.8196 |
| | SPCL-MER[69] | 2024 | 光流 | LOSO | 3 | 0.8571 | 0.7642 | 0.7794 |
| | OFVIG-Net[57] | 2025 | 光流 | LOSO | 3 | — | 0.6066 | 0.5787 |
| | MPFNet[72] | 2025 | 起始帧+顶点帧 | LOSO | 3 | — | 0.7910 | 0.8260 |
| | | | | | 5 | 0.7040 | 0.6950 | — |
| | SODA4MER[85] | 2025 | 序列 | LOSO | 5 | 0.8030 | 0.7893 | — |

从表 3 中可以观察到, 受到 MEGC 2019 的影响, 目前微表情识别领域研究主要采用的数据集为 SMIC、CASME II 和 SAMM, 少数工作会采用 CAS(ME)$^3$ 数据集进行验证, 其中仅有 Lite-Point-GCN[58]利用到了 CAS(ME)$^3$ 提供的深度图像信息, 而 $\mu$-BERT[82]则仅是利用 CAS(ME)$^3$ 中大量无标签的长视频数据辅助进行自监督学习, 未用到多模态数据, 4DME 和 CAS(ME)$^3$ 等数据集中的多模态数据目前还未被充分利用, 究其原因, 我们认为目前即便是在性能表现最好的 CASME II 数据集上的三分类任务, 指标也仅达到 0.95 左右, 更具挑战性的多模态数据仍未成为大多数研究者的第 1 选择; 目前的微表情识别工作基本都采用 LOSO 作为评估协议, 并且大多数工作在三分类任务上验证性能. 值得注意的是, 微表情识别在 SMIC、CASME II 和 SAMM 这 3 个数据集上的三分类任务指标提升空间已经有限, 各项指标基本都达到 0.85 左右及以上, 后续的工作应该专注于更细致的类别划分以及更具挑战难度的数据集 (多模态数据、真实场景数据); 从表 3 还可以看出, 目前大多数模型输入依赖于光流和顶点帧, 在实际场景应用时会极大的限制微表情识别方法的泛用性, 在未来的研究中可以对不依赖于光流、顶点帧及复杂预处理过程的微表情识别方法进行深入研究, 以开发更具泛用性及实际应用价值的微表情识别算法.

### 5.2 微表情检测

我们将前文所述的代表性微表情检测算法汇总在表 4 中, 由于微表情顶点检测相关工作数量过少, 并且评价方法各有差异, 因此仅汇总了区间检测和宏表情+微表情检测相关算法性能, 其中宏表情+微表情检测的指标结果是宏表情和微表情检测的联合指标.

表 4 微表情检测算法性能

| 数据集 | 方法 | 年份 | 检测目标 | 实验指标 | | |
|---|---|---|---|---|---|---|
| | | | | *F*1 | *Precision* | *Recall* |
| CAS(ME)$^2$ | LBCNN[126] | 2020 | 宏+微表情 | 0.0595 | 0.0386 | 0.1301 |
| | MESNet[114] | 2021 | 微表情 | 0.0260 | 0.0130 | 0.1800 |
| | LSSNet[127] | 2021 | 宏+微表情 | 0.0850 | — | — |
| | Yang等人[129] | 2021 | 宏+微表情 | 0.0339 | — | — |
| | ABPN[131] | 2022 | 宏+微表情 | 0.1590 | — | — |
| | Guo等人[115] | 2023 | 微表情 | 0.1373 | — | — |
| | WCMN[116] | 2024 | 微表情 | 0.1831 | — | — |
| CAS(ME)$^3$ | ABPN[131] | 2022 | 宏+微表情 | 0.3529 | — | — |
| | Xu等人[132] | 2023 | 宏+微表情 | 0.2100 | 0.2700 | 0.1700 |
| | Zhang等人[136] | 2024 | 宏+微表情 | 0.2500 | 0.3400 | 0.2000 |
| SAMM LONG | Verburg等人[108] | 2019 | 微表情 | 0.0821 | 0.0450 | 0.4654 |
| | LBCNN[126] | 2020 | 宏+微表情 | 0.0813 | 0.0681 | 0.1010 |
| | MESNet[114] | 2021 | 微表情 | 0.0490 | 0.0280 | 0.2300 |
| | LSSNet[127] | 2021 | 宏+微表情 | 0.2180 | — | — |
| | Yang等人[129] | 2021 | 宏+微表情 | 0.1155 | — | — |
| | ABPN[131] | 2022 | 宏+微表情 | 0.1689 | — | — |
| | Guo等人[115] | 2023 | 微表情 | 0.2770 | — | — |
| | Xu等人[132] | 2023 | 宏+微表情 | 0.2200 | 0.2800 | 0.1900 |
| | WCMN[116] | 2024 | 微表情 | 0.2330 | — | — |
| | Zhang等人[136] | 2024 | 宏+微表情 | 0.3500 | 0.3800 | 0.3300 |

宏表情+微表情检测常用 CAS(ME)$^2$、SAMM LONG 等包含大量已标注长视频片段的数据集进行训练, 但测试数据集的选取通常依据当年的 MEGC 挑战赛标准, 例如 MEGC2022 采用了包含 SAMM 和 CAS(ME)$^3$ 的各 5 段视频数据, MEGC2023 则采用了 SAMM 数据集中的 10 段长视频以及 CAS(ME)$^3$ 中裁剪出的 20 个视频片段, 这些测试数据在此前均未公开, 与前文中介绍的数据集不完全一致, 并且大多工作都会采用 *F*1 指标进行算法评估; 此外表 4 中还可以看出, 基于深度学习的微表情检测方法在指标上呈快速发展趋势, 但仍还有很大进步空间, 目前最先进的算法性能在 *F*1 指标上也仅能达到 0.35 左右, 这距离需要提供高可信度微表情检测区间结果的真实场景应用还有较大距离, 并且目前取得最佳性能的深度学习算法[136]仍然采用了基于光流的手工特征方法对结果区间进行联合修正, 并未全部依赖深度学习方法, 深度学习在微表情检测工作上的应用有待进一步开发.

## 6 微表情分析未来研究展望

基于深度学习的微表情分析方法已经取得巨大突破, 但仍存在大量可探索的研究方向.

(1) 大规模高质量数据集开发

深度学习的成功离不开大量高质数据的驱动, 尽管目前已有很多微表情数据集被开源用于研究, 但数据集规模小、类别不均衡、标签主观等问题始终存在, 仅依靠目前的数据量很难学习到真正有意义的微表情特征, 模型难以摆脱过拟合并拥有真正强大的泛化能力. 微表情数据集的建立从场景设计、采集和标注都需要专业的人员耗费大量时间精力, 这极大地限制了微表情数据生态发展, 未来可以将自动化的微表情检测技术结合到微表情数据采集以提高微表情数据集开发效率, 或是引入脑电、面周肌电等信号辅助定位微表情发生时刻, 目前绝大部分数据集仍采用高压抑制的诱发范式进行采集, 个别数据集采集自真实场景但也面临数据质量低的问题, 开发可控环境下的更高生态效度的微表情诱发范式也可以为推动微表情分析领域做出极大贡献. 微表情分析的最终目标是发

掘出目标人物的潜在情感, 但这一目标的达成并不一定仅通过面部表情实现, 真实的场景中有很多带有情感线索的信息, 例如微姿态、心率、体温等, 在实际应用中也有许多无感采集设备可以简单获取这些生理指标, 因此有必要开发蕴含更多模态信息的微表情数据集. 此外, 目前的微表情数据集多采集于 20–30 岁的年轻人, 并且大多为亚洲人, 未来的微表情数据集应该在性别、人种、年龄等属性上更加多样性, 以防止数据偏差的影响.

(2) 小样本学习方法

深度学习在数据密集型应用中取得了巨大成功, 但当数据集较小时, 其性能往往会受到限制, 微表情分析领域就面临样本量小的困境. 微表情数据集的建立是一件耗时耗力的工作, 短期之内不能完全依赖于等待大规模高质量的微表情数据集诞生, 小样本学习能够借助先验知识快速泛化至仅包含少量有监督信息样本的新任务, 在许多领域取得了卓越进展[138], 微表情分析可以借助小样本学习的思想和成功经验解决样本数量不足、标签不均衡等问题, 另外需要着重设计让模型能够从少量样本中就学习到关注细微运动变化的能力.

(3) 更泛用的微表情分析方法

由于微表情的细微、短暂等特性, 微表情分析极为依赖数据预处理, 合适的数据预处理方式对效果的影响有时甚至超过了方法改进的效果, 且繁琐的预处理步骤只会降低实际应用的效率, 同时这也间接导致了方法开源难度较大、实验复现困难, 极度不利于领域发展, 如何摆脱对数据预处理的依赖、提出对数据变化具有鲁棒性的方法值得研究. 其中, 微表情识别还存在特殊帧的依赖, 大部分识别工作都依赖顶点帧的标注, 但据我们所知, 近几年顶点检测的工作数量极少, 绝大部分微表情检测工作都专注于区间检测, 并且在实际应用时, 一个完整的微表情分析过程需要先进性检测再进行识别, 这就同时受到了检测和识别算法性能的影响, 每多引入一个额外环节都会加倍的影响整体系统性能, 因此不依赖于顶点帧的微表情识别方法值得深入研究.

(4) 检测和识别统一方法研究

单独的研究检测或识别算法始终无法满足落地应用需求, 并且两个环节串行会导致性能损失成倍增加, 将检测和识别统一在同一个模型框架下进行是一个可行的未来研究方向. 目标检测在给出目标框预测的同时进行识别分类, 这种端到端的统一架构有着非常出色的落地应用能力, 这也是微表情分析有意义的未来研究方向之一. 需要注意的是, 将识别和检测放在同一个框架下进行时需要设计的模型可以充分利用到检测和识别所需要的共性特征, 而不是简单的检测和识别任务堆叠.

(5) 更好的评估协议

为了充分利用现有样本, 微表情识别普遍采用 LOSO 作为评估协议计算指标, 但我们注意到目前的工作通常选用每个 LOSO 轮中效果最好 epoch 的指标作为结果, 但由于微表情识别领域过小的数据集规模, 有些 LOSO 轮中测试集可能仅包含一两个样本或者均为同一个类别, 这会导致较大的随机性, 例如有时网络初始化参数较为随机时就能侥幸将所有样本分类正确, 这种随机性也导致了对开源工作复现难度加大. 随着数据集规模的扩大尤其是数据集中每一个被试者都拥有更大的样本量及更均衡的样本类别时该问题自然消失, 但构建一个大规模的微表情数据集困难重重, 若不依赖数据规模来改善该情况则需要制定更合理的评估协议. 新的评估方法应该尽量排除这种随机性, 可以真正客观的比较方法优劣并且方便复现开源算法.

(6) 合成微表情数据的潜在价值挖掘

由于微表情的一些固有特性, 相较于其他领域, 非真实的微表情数据更加难以被研究人员所接受, 目前仍没有针对微表情的合成/生成/伪造数据集诞生, 少数工作采用微表情生成来扩展原有数据集[63,139], 以提高微表情检测和识别的表现. 但随着近年来生成技术不断发展, 以及一些其他领域在合成/生成/伪造数据上取得的成功[140,141], 我们认为这类数据在微表情分析中的潜在价值仍有待挖掘, 包括但不限于研究如何构建兼具真实性与多样性的微表情生成框架以及如何合理的利用这类非真实数据辅助识别和检测任务.

## 7 总 结

面部表情作为一种基本的交流方式, 是情感表达与认知的重要途径, 而微表情更是反映了深层次的真实情感. 由于微表情细微、短暂等特性, 导致微表情分析困难重重, 基于深度学习的微表情分析技术在近 5 年左右才开始较为快速的发展并已经成为目前的研究主流, 本文全面总结并介绍目前开源可供研究的主流微表情数据集、数据集制作方式以及常用的数据预处理方法; 分别详细总结了微表情识别和微表情检测这两个微表情分析重要研究分支的基于深度学习的代表性工作, 讨论了其中一些工作的优势和局限, 并且还对这些方法的性能指标进行汇总; 最后我们对微表情分析领域目前的发展现状进行总结并且对未来可能的研究方向进行展望, 期待能有更多研究人员参与到微表情分析工作中, 希望本文能为微表情分析未来研究方向提供一定参考价值.

## References


[1] Corneanu CA, Simón MO, Cohn JF, Guerrero SE. Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2016, 38(8): 1548–1568. [doi: 10.1109/TPAMI.2016.2515606]

[2] Ekman P. Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life. New York: Henry Holt and Company, 2003.

[3] Ekman P, Friesen WV. Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 1971, 17(2): 124–129. [doi: 10.1037/h0030377]

[4] Ekman P. Lie catching and microexpressions. In: Martin C, ed. The Philosophy of Deception. Oxford: Oxford University Press, 2009. 118–136. [doi: 10.1093/acprof:oso/9780195327939.003.0008]

[5] Yan WJ, Wu Q, Liang J, Chen YH, Fu XL. How fast are the leaked facial expressions: The duration of micro-expressions. Journal of Nonverbal Behavior, 2013, 37(4): 217–230. [doi: 10.1007/s10919-013-0159-8]

[6] Ekman P. Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (Revised Edition). New York: W. W. Norton & Company, 2009.

[7] Porter S, ten Brinke L. Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. Psychological science, 2008, 19(5): 508–514. [doi: 10.1111/j.1467-9280.2008.02116.x]

[8] Endres J, Laidlaw A. Micro-expression recognition training in medical students: A pilot study. BMC Medical Education, 2009, 9(1): 47. [doi: 10.1186/1472-6920-9-47]

[9] O'Sullivan M, Frank MG, Hurley CM, Tiwana J. Police lie detection accuracy: The effect of lie scenario. Law and Human Behavior, 2009, 33(6): 530–538. [doi: 10.1007/s10979-008-9166-4]

[10] Qin WF, Zou BC, Li X, Wang WP, Ma HM. Micro-expression spotting with face alignment and optical flow. In: Proc. of the 31st ACM Int'l Conf. on Multimedia. Ottawa: ACM, 2023. 9501–9505. [doi: 10.1145/3581783.3612853]

[11] Yu J, Cai ZP, Du SS, Shen XX, Wang L, Gao F. Efficient micro-expression spotting based on main directional mean optical flow feature. In: Proc. of the 31st ACM Int'l Conf. on Multimedia. Ottawa: ACM, 2023. 9541–9545. [doi: 10.1145/3581783.3612861]

[12] Frank M, Herbasz M, Sinuk K, Keller A, Nolan C. I see how you feel: Training laypeople and professionals to recognize fleeting emotions. In: Proc. of the 2009 Annual Meeting of the Int'l Communication Association. Sheraton New York, 2009. 1–35.

[13] Nag S, Bhunia AK, Konwer A, Roy PP. Facial micro-expression spotting and recognition using time contrasted feature with visual memory. In: Proc. of the 2019 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing. Brighton: IEEE, 2019. 2022–2026. [doi: 10.1109/ICASSP.2019.8683737]

[14] Li YT, Wei JS, Liu Y, Kauttonen J, Zhao GY. Deep learning for micro-expression recognition: A survey. IEEE Trans. on Affective Computing, 2022, 13(4): 2028–2046. [doi: 10.1109/TAFFC.2022.3205170]

[15] Zhang H, Yin L, Zhang HL. A review of micro-expression spotting: Methods and challenges. Multimedia Systems, 2023, 29(4): 1897–1915. [doi: 10.1007/s00530-023-01076-z]

[16] Ben XY, Ren Y, Zhang JP, Wang SJ, Kpalma K, Meng WX, Liu YJ. Video-based facial micro-expression analysis: A survey of datasets, features and algorithms. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2022, 44(9): 5826–5846. [doi: 10.1109/TPAMI.2021.3067464]

[17] Zhao GY, Li XB, Li YT, Pietikäinen M. Facial micro-expressions: An overview. Proc. of the IEEE, 2023, 111(10): 1215–1235. [doi: 10.1109/JPROC.2023.3275192]

[18] Ren SQ, He KM, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. on

Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]

[19] Diwan T, Anirudh G, Tembhurne JV. Object detection using YOLO: Challenges, architectural successors, datasets and applications. Multimedia Tools and Applications, 2023, 82(6): 9243–9275. [doi: 10.1007/s11042-022-13644-y]

[20] Sun FM, Hu XH, Wu JY, Sun J, Wang FS. RGB-D salient object detection based on cross-modal interactive fusion and global awareness. Ruan Jian Xue Bao/Journal of Software, 2024, 35(4): 1899–1913 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6833.htm [doi: 10.13328/j.cnki.jos.006833]

[21] Kirillov A, Mintun E, Ravi N, Mao HZ, Rolland C, Gustafson L, Xiao TT, Whitehead S, Berg AC, Lo WY, Dollár P, Girshick R. Segment anything. In: Proc. of the 2023 IEEE/CVF Int'l Conf. on Computer Vision. Paris: IEEE, 2023. 3992–4003. [doi: 10.1109/ICCV51070.2023.00371]

[22] Ma J, He YT, Li FF, Han L, You CY, Wang B. Segment anything in medical images. Nature Communications, 2024, 15(1): 654. [doi: 10.1038/s41467-024-44824-z]

[23] Zhang ZM, Guo Y, Ma CX, Deng XM, Wang HA. GT-4S: Graph Transformer for scene sketch semantic segmentation. Ruan Jian Xue Bao/Journal of Software, 2025, 36(3): 1375–1389 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/7155.htm [doi: 10.13328/j.cnki.jos.007155]

[24] Polikovsky S, Kameda Y, Ohta Y. Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor. In: Proc. of the 3rd Int'l Conf. on Imaging for Crime Detection and Prevention. London: IET, 2009. 1–6. [doi: 10.1049/ic.2009.0244]

[25] Shreve M, Godavarthy S, Goldgof D, Sarkar S. Macro- and micro-expression spotting in long videos using spatio-temporal strain. In: Proc. of the 2011 IEEE Int'l Conf. on Automatic Face & Gesture Recognition. Santa Barbara: IEEE, 2011. 51–56. [doi: 10.1109/FG.2011.5771451]

[26] Yan WJ, Wu Q, Liu YJ, Wang SJ, Fu XL. CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces. In: Proc. of the 10th IEEE Int'l Conf. and Workshops on Automatic Face and Gesture Recognition. Shanghai: IEEE, 2013. 1–7. [doi: 10.1109/FG.2013.6553799]

[27] Yan WJ, Li XB, Wang SJ, Zhao GY, Liu YJ, Chen YH, Fu XL. CASME II: An improved spontaneous micro-expression database and the baseline evaluation. PLoS One, 2014, 9(1): e86041. [doi: 10.1371/journal.pone.0086041]

[28] Qu FB, Wang SJ, Yan WJ, Li H, Wu SH, Fu XL. CAS(ME)$^2$: A database for spontaneous macro-expression and micro-expression spotting and recognition. IEEE Trans. on Affective Computing, 2018, 9(4): 424–436. [doi: 10.1109/TAFFC.2017.2654440]

[29] Li JT, Dong ZZ, Lu SY, Wang SJ, Yan WJ, Ma YH, Liu Y, Huang CB, Fu XL. CAS(ME)$^3$: A third generation facial spontaneous micro-expression database with depth information and high ecological validity. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2023, 45(3): 2782–2800. [doi: 10.1109/TPAMI.2022.3174895]

[30] Li JT, Lu SY, Wang Y, Dong ZZ, Wang SJ, Fu XL. Could micro-expressions be quantified? Electromyography gives affirmative evidence. IEEE Trans. on Affective Computing, 2025, 16(4): 2959–2974. [doi: 10.1109/TAFFC.2025.3575127]

[31] Davison AK, Lansley C, Costen N, Tan K, Yap MH. SAMM: A spontaneous micro-facial movement dataset. IEEE Trans. on Affective Computing, 2018, 9(1): 116–129. [doi: 10.1109/TAFFC.2016.2573832]

[32] Yap CH, Kendrick C, Yap MH. SAMM long videos: A spontaneous facial micro- and macro-expressions dataset. In: Proc. of the 15th IEEE Int'l Conf. on Automatic Face and Gesture Recognition. Buenos Aires: IEEE, 2020. 771–776. [doi: 10.1109/FG47880.2020.00029]

[33] Li XB, Pfister T, Huang XH, Zhao GY, Pietikäinen M. A spontaneous micro-expression database: Inducement, collection and baseline. In: Proc. of the 10th IEEE Int'l Conf. and Workshops on Automatic Face and Gesture Recognition. Shanghai: IEEE, 2013. 1–6. [doi: 10.1109/FG.2013.6553717]

[34] Husák P, Cech J, Matas J. Spotting facial micro-expressions "in the wild". In: Proc. of the 22nd Computer Vision Winter Workshop. Retz, 2017. 1–9.

[35] Li XB, Cheng SY, Li YT, Behzad M, Shen J, Zafeiriou S, Pantic M, Zhao GY. 4DME: A spontaneous 4D micro-expression dataset with multimodalities. IEEE Trans. on Affective Computing, 2023, 14(4): 3031–3047. [doi: 10.1109/TAFFC.2022.3182342]

[36] Zhao SR, Tang HY, Mao XL, Liu SF, Zhang YM, Wang H, Xu T, Chen EH. DFME: A new benchmark for dynamic facial micro-expression recognition. IEEE Trans. on Affective Computing, 2024, 15(3): 1371–1386. [doi: 10.1109/TAFFC.2023.3341918]

[37] Ekman P, Friesen WV. Facial action coding system. Environmental Psychology & Nonverbal Behavior, 1978. [doi: 10.1037/t27734-000]

[38] Yang SR, Yang HC, Shen FR, Zhao J. Image data augmentation for deep learning: A survey. Ruan Jian Xue Bao/Journal of Software, 2025, 36(3): 1390–1412 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/7263.htm [doi: 10.13328/j.cnki.jos.007263]

[39] Wu HY, Rubinstein M, Shih E, Guttag J, Durand F, Freeman W. Eulerian video magnification for revealing subtle changes in the world. ACM Trans. on Graphics, 2012, 31(4): 65. [doi: 10.1145/2185520.2185561]

[40] Lei L, Li JF, Chen T, Li SG. A novel Graph-TCN with a graph structured representation for micro-expression recognition. In: Proc. of the 28th ACM Int'l Conf. on Multimedia. Seattle: ACM, 2020. 2237–2245. [doi: 10.1145/3394171.3413714]

[41] Oh TH, Jaroensri R, Kim C, Elgharib M, Durand F, Freeman WT, Matusik W. Learning-based video motion magnification. In: Proc. of the 15th European Conf. on Computer Vision. Munich: Springer, 2018. 663–679. [doi: 10.1007/978-3-030-01225-0_39]

[42] Lucas BD. Generalized image matching by the method of differences [Ph. D. Thesis]. Pittsburgh: Carnegie Mellon University, 1985.

[43] Farnebäck G. Two-frame motion estimation based on polynomial expansion. In: Proc. of the 13th Scandinavian Conf. on Image Analysis. Halmstad: Springer, 2003. 363–370. [doi: 10.1007/3-540-45103-X_50]

[44] Wedel A, Pock T, Zach C, Bischof H, Cremers D. An improved algorithm for TV-$L^1$ optical flow. Int'l Dagstuhl Seminar. Dagstuhl: Springer, 2009. 23–45. [doi: 10.1007/978-3-642-03061-1_2]

[45] Dosovitskiy A, Fischer P, Ilg E, Hausser P, Hazirbas C, Golkov V, Smagt P, Cremers D, Brox T. FlowNet: Learning optical flow with convolutional networks. In: Proc. of the 2015 IEEE Int'l Conf. on Computer Vision. Santiago: IEEE, 2015. 2758–2766. [doi: 10.1109/ICCV.2015.316]

[46] Liu YJ, Zhang JK, Yan WJ, Wang SJ, Zhao GY, Fu XL. A main directional mean optical flow feature for spontaneous micro-expression recognition. IEEE Trans. on Affective Computing, 2016, 7(4): 299–310. [doi: 10.1109/TAFFC.2015.2485205]

[47] Huang XH, Zhao GY. Spontaneous facial micro-expression analysis using spatiotemporal local radon-based binary pattern. In: Proc. of the 2017 Int'l Conf. on the Frontiers and Advances in Data Science (FADS). Xi'an: IEEE, 2017. 159–164. [doi: 10.1109/FADS.2017.8253219]

[48] Davison AK, Merghani W, Yap MH. Objective classes for micro-facial expression recognition. Journal of Imaging, 2018, 4(10): 119. [doi: 10.3390/jimaging4100119]

[49] Patel D, Hong XP, Zhao GY. Selective deep features for micro-expression recognition. In: Proc. of the 23rd Int'l Conf. on Pattern Recognition (ICPR). Cancun: IEEE, 2016. 2258–2263. [doi: 10.1109/ICPR.2016.7899972]

[50] Kim DH, Baddar WJ, Ro YM. Micro-expression recognition with expression-state constrained spatio-temporal feature representations. In: Proc. of the 24th ACM Int'l Conf. on Multimedia. Amsterdam: ACM, 2016. 382–386. [doi: 10.1145/2964284.2967247]

[51] Wang SJ, Li BJ, Liu YJ, Yan WJ, Ou XY, Huang XH, Xu F, Fu XL. Micro-expression recognition with small sample size by transferring long-term convolutional neural network. Neurocomputing, 2018, 312: 251–262. [doi: 10.1016/j.neucom.2018.05.107]

[52] See J, Yap MH, Li JT, Hong XP, Wang SJ. MEGC 2019—The second facial micro-expressions grand challenge. In: Proc. of the 14th IEEE Int'l Conf. on Automatic Face & Gesture Recognition. Lille: IEEE, 2019. 1–5. [doi: 10.1109/FG.2019.8756611]

[53] Verma M, Vipparthi SK, Singh G, Murala S. LEARNet: Dynamic imaging network for micro expression recognition. IEEE Trans. on Image Processing, 2020, 29: 1618–1627. [doi: 10.1109/TIP.2019.2912358]

[54] Li HT, Sui MZ, Zhu ZQ, Zhao F. MMNet: Muscle motion-guided network for micro-expression recognition. arXiv:2201.05297, 2022.

[55] Wang ZF, Zhang KH, Luo WH, Sankaranarayana R. Htnet for micro-expression recognition. Neurocomputing, 2024, 602: 128196. [doi: 10.1016/j.neucom.2024.128196]

[56] Cai WH, Zhao JL, Yi R, Yu MJ, Duan FQ, Pan ZK, Liu YJ. MFDAN: Multi-level flow-driven attention network for micro-expression recognition. IEEE Trans. on Circuits and Systems for Video Technology, 2024, 34(12): 12823–12836. [doi: 10.1109/TCSVT.2024.3437481]

[57] Zhang LJ, Zhang YF, Sun XZ, Tang WC, Wang XM, Li ZS. Micro-expression recognition based on direct learning of graph structure. Neurocomputing, 2025, 619: 129135. [doi: 10.1016/j.neucom.2024.129135]

[58] Zhang R, Yin JQ, Qi C, Dang YH, Wang ZH, Zhang ZC, Liu HP. Facial 3D regional structural motion representation using lightweight point cloud networks for micro-expression recognition. IEEE Trans. on Affective Computing, 2025. [doi: 10.1109/TAFFC.2025.3535569]

[59] Lu LP, Tavabi L, Soleymani M. Self-supervised learning for facial action unit recognition through temporal consistency. In: Proc. of the 2020 British Machine Vision Conference. Manchester: BMVA, 2020. [doi: 10.5244/c.34.180]

[60] Ma C, Chen L, Yong JH. AU R-CNN: Encoding expert prior knowledge into R-CNN for action unit detection. Neurocomputing, 2019, 355: 35–47. [doi: 10.1016/j.neucom.2019.03.082]

[61] Li Y, Zeng JB, Shan SG. Learning representations for facial actions from unlabeled videos. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2022, 44(1): 302–317. [doi: 10.1109/TPAMI.2020.3011063]

[62] Lo L, Xie HX, Shuai HH, Cheng WH. MER-GCN: Micro-expression recognition based on relation modeling with graph convolutional networks. In: Proc. of the 2020 IEEE Conf. on Multimedia Information Processing and Retrieval (MIPR). Shenzhen: IEEE, 2020. 79–84. [doi: 10.1109/MIPR49039.2020.00023]

[63] Xie HX, Lo L, Shuai HH, Cheng WH. Au-assisted graph attention convolutional network for micro-expression recognition. In: Proc. of the 28th ACM Int'l Conf. on Multimedia. Seattle: ACM, 2020. 2871–2880. [doi: 10.1145/3394171.3414012]

[64] Wang L, Huang PY, Cai WY, Liu XY. Micro-expression recognition by fusing action unit detection and spatio-temporal features. In:

Proc. of the 2024 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP 2024). Seoul: IEEE, 2024. 5595–5599. [doi: 10.1109/ICASSP48485.2024.10446702]

[65] Lei L, Chen T, Li SG, Li JF. Micro-expression recognition based on facial graph representation learning and facial action unit fusion. In: Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 1571–1580. [doi: 10.1109/CVPRW53098.2021.00173]

[66] Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. In: Proc. of the 37th Int'l Conf. on Machine Learning. Vienna: PMLR, 2020. 1597–1607.

[67] Khosla P, Teterwak P, Wang C, Sarna A, Tian YL, Isola P, Maschinot A, Liu C, Krishnan D. Supervised contrastive learning. In: Proc. of the 34th Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020. 18661–18673.

[68] Zhi RC, Hu J, Wan F. Micro-expression recognition with supervised contrastive learning. Pattern Recognition Letters, 2022, 163: 25–31. [doi: 10.1016/j.patrec.2022.09.006]

[69] Fang XQ, Wu QF, Cao L. SPCL-MER: Supervised prototypical contrastive learning for micro-expression recognition. In: Proc. of the 2024 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP 2024). Seoul: IEEE, 2024. 5690–5694. [doi: 10.1109/ICASSP48485.2024.10448102]

[70] Song YX, Wang JZ, Wu TB, Huang ZC, Xiao J. Micro-expression recognition based on attribute information embedding and cross-modal contrastive learning. In: Proc. of the 2022 Int'l Joint Conf. on Neural Networks (IJCNN 2022). Padua: IEEE, 2022. 1–7. [doi: 10.1109/IJCNN55064.2022.9892347]

[71] Wei MT, Jiang XX, Zheng WM, Zong Y, Lu C, Liu JT. CMNet: Contrastive magnification network for micro-expression recognition. In: Proc. of the 37th AAAI Conf. on Artificial Intelligence. Washington: AAAI Press, 2023. 119–127. [doi: 10.1609/aaai.v37i1.25083]

[72] Ma C, Zhao SK, Pei Y, Xie L, Yin EW, Yan Y. A multi-prior fusion network for video-based micro-expression recognition. In: Proc. of the 2025 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP 2025). Hyderabad: IEEE, 2025. 1–5. [doi: 10.1109/ICASSP49660.2025.10889587]

[73] Hospedales TM, Antoniou A, Micaelli P, Storkey AJ. Meta-learning in neural networks: A survey. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2022, 44(9): 5149–5169. [doi: 10.1109/TPAMI.2021.3079209]

[74] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.

[75] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional Transformers for language understanding. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Vol. 1: Long and Short Papers). Minneapolis: ACL, 2019. 4171–4186. [doi: 10.18653/v1/N19-1423]

[76] Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. 2018. https://api.semanticscholar.org/CorpusID:49313245

[77] Gui T, Xi ZH, Zheng R, Liu Q, Ma RT, Wu T, Bao R, Zhang Q. Recent researches of robustness in natural language processin based on deep neural network. Chinese Journal of Computers, 2024, 47(1): 90–112 (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2024.00090]

[78] Lu JS, Batra D, Parikh D, Lee S. ViLBERT: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. In: Proc. of the 33rd Conf. on Neural Information Processing Systems. Vancouver, 2019.

[79] Bao HB, Dong L, Piao SH, Wei FR. BEiT: BERT pre-training of image Transformers. arXiv:2106.08254, 2022.

[80] Li YH, Fan HQ, Hu RH, Feichtenhofer C, He KM. Scaling language-image pre-training via masking. In: Proc. of the 2023 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 23390–23400. [doi: 10.1109/CVPR52729.2023.02240]

[81] Shi ZN, Chen HP, Zhang D, Shen XJ. Pre-training-driven multimodal boundary-aware vision Transformer. Ruan Jian Xue Bao/Journal of Software, 2023, 34(5): 2051–2067 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6768.htm [doi: 10.13328/j.cnki.jos.006768]

[82] Nguyen XB, Duong CN, Li X, Gauch S, Seo HS, Luu K. Micron-BERT: BERT-based facial micro-expression recognition. In: Proc. of the 2023 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 1482–1492. [doi: 10.1109/CVPR52729.2023.00149]

[83] Fan XQ, Chen XL, Jiang MJ, Shahid AR, Yan H. SelfME: Self-supervised motion learning for micro-expression recognition. In: Proc. of the 2023 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 13834–13843. [doi: 10.1109/CVPR52729.2023.01329]

[84] Zhai ZJ, Zhao JH, Long CJ, Xu WJ, He SJ, Zhao HJ. Feature representation learning with adaptive displacement generation and Transformer fusion for micro-expression recognition. In: Proc. of the 2023 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 22086–22095. [doi: 10.1109/CVPR52729.2023.02115]

[85] Zhang BH, Wang XJ, Wang CB, He GQ. Dynamic stereotype theory induced micro-expression recognition with oriented deformation.

In: Proc. of the 2025 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Nashville: IEEE, 2025. 10701–10711. [doi: 10.1109/CVPR52734.2025.01000]

[86] Windholz G. Pavlov's conceptualization of the dynamic stereotype in the theory of higher nervous activity. The American Journal of Psychology, 1996, 109(2): 287–295. [doi: 10.2307/1423277]

[87] Georghiades AS, Belhumeur PN, Kriegman DJ. From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2001, 23(6): 643–660. [doi: 10.1109/34.927464]

[88] Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I. The extended cohn-kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: Proc. of the 2010 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops. San Francisco: IEEE, 2010. 94–101. [doi: 10.1109/CVPRW.2010.5543262]

[89] Zhao GY, Huang XH, Taini M, Li SZ, PietikäInen M. Facial expression recognition from near-infrared videos. Image and Vision Computing, 2011, 29(9): 607–619. [doi: 10.1016/j.imavis.2011.07.002]

[90] Lyons M, Akamatsu S, Kamachi M, Gyoba J. Coding facial expressions with gabor wavelets. In: Proc. of the 3rd IEEE Int'l Conf. on Automatic Face and Gesture Recognition. Nara: IEEE, 1998. 200–205. [doi: 10.1109/AFGR.1998.670949]

[91] Aifanti N, Papachristou C, Delopoulos A. The MUG facial expression database. In: Proc. of the 11th Int'l Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 10. Desenzano del Garda: IEEE, 2010. 1–4.

[92] Li S, Deng WH, Du JP. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 2584–2593. [doi: 10.1109/CVPR.2017.277]

[93] Deng J, Dong W, Socher R, Li LJ, Li K, Li FF. ImageNet: A large-scale hierarchical image database. In: Proc. of the 2009 IEEE Conf. on Computer Vision and Pattern Recognition. Miami: IEEE, 2009. 248–255. [doi: 10.1109/CVPR.2009.5206848]

[94] Liu YJ, Li YG, Yi XH, Hu ZJ, Zhang HY, Liu YZ. Lightweight ViT model for micro-expression recognition enhanced by transfer learning. Frontiers in Neurorobotics, 2022, 16: 922761. [doi: 10.3389/fnbot.2022.922761]

[95] Mehta S, Rastegari M. MobileViT: Light-weight, general-purpose, and mobile-friendly vision Transformer. arXiv:2110.02178, 2022.

[96] Liu YC, Du HM, Zheng L, Gedeon T. A neural micro-expression recognizer. In: Proc. of the 14th IEEE Int'l Conf. on Automatic Face & Gesture Recognition (FG 2019). Lille: IEEE, 2019. 1–4. [doi: 10.1109/FG.2019.8756583]

[97] Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, Marchand M, Lempitsky V. Domain-adversarial training of neural networks. The Journal of Machine Learning Research, 2016, 17(1): 1–35.

[98] Wang FF, Zong Y, Zhu J, Wei MT, Xu XL, Lu C, Zheng WM. Progressively learning from macro-expressions for micro-expression recognition. In: Proc. of the 2024 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP 2024). Seoul: IEEE, 2024. 4390–4394. [doi: 10.1109/ICASSP48485.2024.10446028]

[99] Xia B, Wang WK, Wang SF, Chen EH. Learning from macro-expression: A micro-expression recognition framework. In: Proc. of the 28th ACM Int'l Conf. on Multimedia. Seattle: ACM, 2020. 2936–2944. [doi: 10.1145/3394171.3413774]

[100] Xia B, Wang SF. Micro-expression recognition enhanced by macro-expression from spatial-temporal domain. In: Proc. of the 30th Int'l Joint Conf. on Artificial Intelligence. Montreal, 2021. 1186–1193. [doi: 10.24963/ijcai.2021/164]

[101] Li XB, Hong XP, Moilanen A, Huang XH, Pfister T, Zhao GY, Pietikäinen M. Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods. IEEE Trans. on Affective Computing, 2018, 9(4): 563–577. [doi: 10.1109/TAFFC.2017.2667642]

[102] Moilanen A, Zhao GY, Pietikäinen M. Spotting rapid facial movements from videos using appearance-based feature difference analysis. In: Proc. of the 22nd Int'l Conf. on Pattern Recognition. Stockholm: IEEE, 2014. 1722–1727. [doi: 10.1109/ICPR.2014.303]

[103] He Y, Wang SJ, Li JT, Yap MH. Spotting macro-and micro-expression intervals in long video sequences. In: Proc. of the 15th IEEE Int'l Conf. on Automatic Face and Gesture Recognition. Buenos Aires: IEEE, 2020. 742–748. [doi: 10.1109/FG47880.2020.00036]

[104] Ye YH. Research on micro-expression spotting method based on optical flow features. In: Proc. of the 29th ACM Int'l Conf. on Multimedia. Virtual Event: ACM, 2021. 4803–4807. [doi: 10.1145/3474085.3479225]

[105] Zhang LW, Li JT, Wang SJ, Duan XH, Yan WJ, Xie HY, Huang SC. Spatio-temporal fusion for macro- and micro-expression spotting in long video sequences. In: Proc. of the 15th IEEE Int'l Conf. on Automatic Face and Gesture Recognition. Buenos Aires: IEEE, 2020. 734–741. [doi: 10.1109/FG47880.2020.00037]

[106] Davison A, Merghani W, Lansley C, Ng CC, Yap MH. Objective micro-facial movement detection using FACS-based regions and baseline evaluation. In: Proc. of the 13th IEEE Int'l Conf. on Automatic Face & Gesture Recognition. Xi'an: IEEE, 2018. 642–649. [doi: 10.1109/FG.2018.00101]

[107] Li JT, Wang SJ, Yap MH, See J, Hong XP, Li XB. MEGC2020—The third facial micro-expression grand challenge. In: Proc. of the 15th IEEE Int'l Conf. on Automatic Face and Gesture Recognition. Buenos Aires: IEEE, 2020. 777–780. [doi: 10.1109/FG47880.2020.00035]

[108] Verburg M, Menkovski V. Micro-expression detection in long videos using optical flow and recurrent neural networks. In: Proc. of the 14th IEEE Int'l Conf. on Automatic Face & Gesture Recognition. Lille: IEEE, 2019. 1–6. [doi: 10.1109/FG.2019.8756588]

[109] Takalkar MA, Thuseethan S, Rajasegarar S, Chaczko Z, Xu M, Yearwood J. LGAttNet: Automatic micro-expression detection using dual-stream local and global attentions. Knowledge-based Systems, 2021, 212: 106566. [doi: 10.1016/j.knosys.2020.106566]

[110] Uijlings JRR, Van De Sande KEA, Gevers T, Smeulders AWM. Selective search for object recognition. Int'l Journal of Computer Vision, 2013, 104(2): 154–171. [doi: 10.1007/s11263-013-0620-5]

[111] Girshick R. Fast R-CNN. In: Proc. of the 2015 IEEE Int'l Conf. on Computer Vision. Santiago: IEEE, 2015. 1440–1448. [doi: 10.1109/ICCV.2015.169]

[112] Zeng RH, Huang WB, Gan C, Tan MK, Rong Y, Zhao PL, Huang JZ. Graph convolutional networks for temporal action localization. In: Proc. of the 2019 IEEE/CVF Int'l Conf. on Computer Vision. Seoul: IEEE, 2019. 7093–7102. [doi: 10.1109/ICCV.2019.00719]

[113] Zeng RH, Gan C, Chen PH, Huang WB, Wu QY, Tan MK. Breaking winner-takes-all: Iterative-winners-out networks for weakly supervised temporal action localization. IEEE Trans. on Image Processing, 2019, 28(12): 5797–5808. [doi: 10.1109/TIP.2019.2922108]

[114] Wang SJ, He Y, Li JT, Fu XL. MESNet: A convolutional neural network for spotting multi-scale micro-expression intervals in long videos. IEEE Trans. on Image Processing, 2021, 30: 3956–3969. [doi: 10.1109/TIP.2021.3064258]

[115] Guo XP, Zhang XB, Li L, Xia ZQ. Micro-expression spotting with multi-scale local Transformer in long videos. Pattern Recognition Letters, 2023, 168: 146–152. [doi: 10.1016/j.patrec.2023.03.012]

[116] Zhou JX, Wu Y. Micro-expression spotting with a novel wavelet convolution magnification network in long videos. Pattern Recognition Letters, 2024, 178: 130–137. [doi: 10.1016/j.patrec.2024.01.004]

[117] Li QF, Shen LL, Guo S, Lai ZH. Wavelet integrated CNNs for noise-robust image classification. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 7243–7252. [doi: 10.1109/CVPR42600.2020.00727]

[118] Zhang ZH, Chen T, Meng HY, Liu GY, Fu XL. SMEConvNet: A convolutional neural network for spotting spontaneous facial micro-expression from long videos. IEEE Access, 2018, 6: 71143–71151. [doi: 10.1109/ACCESS.2018.2879485]

[119] Tran TK, Vo QN, Hong XP, Zhao GY. Dense prediction for micro-expression spotting based on deep sequence model. Electronic Imaging, 2019, 31(8): 401-1–401-5. [doi: 10.2352/ISSN.2470-1173.2019.8.IMAWM-401]

[120] Yee NL, Zulkifley MA, Saputro AH, Saputro AH, Abdani SR. Apex frame spotting using attention networks for micro-expression recognition system. Computers, Materials and Continua, 2022, 73(3): 5331–5348. [doi: 10.32604/cmc.2022.028801]

[121] Liu BT, Zhang ZY, Zhou J, Wang HP, Chen T. Long video micro-expression spotting based on OCC theory. In: Proc. of the 2023 IEEE Int'l Conf. on Bioinformatics and Biomedicine (BIBM). Istanbul: IEEE, 2023. 2101–2106. [doi: 10.1109/BIBM58861.2023.10385507]

[122] Li JT, Yap MH, Cheng WH, See J, Hong XP, Li XB, Wang SJ. FME'21: 1st workshop on facial micro-expression: Advanced techniques for facial expressions generation and spotting. In: Proc. of the 29th ACM Int'l Conf. on Multimedia. ACM, 2021. 5700–5701. [doi: 10.1145/3474085.3478579]

[123] Li JT, Yap MH, Cheng WH, See J, Hong XP, Li XB, Wang SJ, Davison AK, Li YT, Dong ZZ. MEGC2022: ACM multimedia 2022 micro-expression grand challenge. In: Proc. of the 30th ACM Int'l Conf. on Multimedia. Lisboa: ACM, 2022. 7170–7174. [doi: 10.1145/3503161.3551601]

[124] Davison AK, Li JT, Yap MH, See J, Cheng WH, Li XB, Hong XP, Wang SJ. MEGC2023: ACM multimedia 2023 ME grand challenge. In: Proc. of the 31st ACM Int'l Conf. on Multimedia. Ottawa: ACM, 2023. 9625–9629. [doi: 10.1145/3581783.3612833]

[125] See J, Li JT, Davison AK, Liong GB, Yap MH, Cheng WH, Li XB, Hong XP, Wang SJ. MEGC2024: ACM multimedia 2024 facial micro-expression grand challenge. In: Proc. of the 32nd ACM Int'l Conf. on Multimedia. Melbourne: ACM, 2024. 11482–11483. [doi: 10.1145/3664647.3689140]

[126] Pan H, Xie L, Wang ZL. Local bilinear convolutional neural network for spotting macro- and micro-expression intervals in long video sequences. In: Proc. of the 15th IEEE Int'l Conf. on Automatic Face and Gesture Recognition. Buenos Aires: IEEE, 2020. 749–753. [doi: 10.1109/FG47880.2020.00052]

[127] Yu WW, Jiang JW, Li YJ. LSSNet: A two-stream convolutional neural network for spotting macro- and micro-expression in long videos. In: Proc. of the 29th ACM Int'l Conf. on Multimedia. ACM, 2021. 4745–4749. [doi: 10.1145/3474085.3479215]

[128] Carreira J, Zisserman A. Quo vadis, action recognition? A new model and the kinetics dataset. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 4724–4733. [doi: 10.1109/CVPR.2017.502]

[129] Yang B, Wu JM, Zhou ZG, Komiya M, Kishimoto K, Xu JF, Nonaka K, Horiuchi T, Komorita S, Hattori G, Naito S, Takishima Y. Facial action unit-based deep learning framework for spotting macro- and micro-expressions in long video sequences. In: Proc. of the 29th ACM Int'l Conf. on Multimedia. ACM, 2021. 4794–4798. [doi: 10.1145/3474085.3479209]

[130] Baltrusaitis T, Zadeh A, Lim YC, Morency LP. OpenFace 2.0: Facial behavior analysis toolkit. In: Proc. of the 13th IEEE Int'l Conf. on

Automatic Face & Gesture Recognition. Xi'an: IEEE, 2018. 59–66. [doi: 10.1109/FG.2018.00019]

[131] Leng WB, Zhao SR, Zhang YM, Liu SF, Mao XL, Wang H, Xu T, Chen EH. ABPN: Apex and boundary perception network for micro- and macro-expression spotting. In: Proc. of the 30th ACM Int'l Conf. on Multimedia. Lisboa: ACM, 2022. 7160–7164. [doi: 10.1145/3503161.3551599]

[132] Xu K, Chen K, Sun LC, Lian Z, Liu B, Chen G, Sun HY, Xu MY, Tao JH. Integrating VideoMAE based model and optical flow for micro- and macro-expression spotting. In: Proc. of the 31st ACM Int'l Conf. on Multimedia. Ottawa: ACM, 2023. 9576–9580. [doi: 10.1145/3581783.3612868]

[133] Chung JS, Nagrani A, Zisserman A. VoxCeleb2: Deep speaker recognition. arXiv:1806.05622, 2018

[134] Tong Z, Song YB, Wang J, Wang LM. VideoMAE: Masked autoencoders are data-efficient learners for self-supervised video pre-training. In: Proc. of the 36th Int'l Conf. on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2022. 10078–10093.

[135] Yu J, Cai ZP, Liu ZP, Xie GC, He P. Facial expression spotting based on optical flow features. In: Proc. of the 30th ACM Int'l Conf. on Multimedia. Lisboa: ACM, 2022. 7205–7209. [doi: 10.1145/3503161.3551608]

[136] Zhang ZY, Zhao SR, Mao XL, Liu SF, Wang H, Xu T, Chen EH. A multi-scale feature learning network with optical flow correction for micro- and macro-expression spotting. In: Proc. of the 32nd ACM Int'l Conf. on Multimedia. Melbourne: ACM, 2024. 11497–11502. [doi: 10.1145/3664647.3689143]

[137] Bodla N, Singh B, Chellappa R, Davis LS. Soft-NMS—Improving object detection with one line of code. In: Proc. of the 2017 IEEE Int'l Conf. on Computer Vision. Venice: IEEE, 2017. 5562–5570. [doi: 10.1109/ICCV.2017.593]

[138] Wang YQ, Yao QM, Kwok JT, Ni LM. Generalizing from a few examples: A survey on few-shot learning. ACM Computing Surveys, 2020, 53(3): 63. [doi: 10.1145/3386252]

[139] Zhao SR, Yin SK, Tang HY, Jin RJ, Xu YF, Xu T, Chen EH. Fine-grained micro-expression generation based on thin-plate spline and relative au constraint. In: Proc. of the 30th ACM Int'l Conf. on Multimedia. Lisboa: ACM, 2022. 7150–7154. [doi: 10.1145/3503161.3551597]

[140] Zhang TY, Xie LX, Wei LH, Zhuang ZJ, Zhang YF, Li B, Tian Q. UnrealPerson: An adaptive pipeline towards costless person re-identification. In: Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 11501–11510. [doi: 10.1109/CVPR46437.2021.01134]

[141] Wu HY, Singh J, Tian SC, Zheng L, Bowyer KW. Vec2Face: Scaling face dataset generation with loosely constrained vectors. arXiv:2409.02979, 2025.

## 附中文参考文献

[20] 孙福明, 胡锡航, 武景宇, 孙静, 王法胜. 跨模态交互融合与全局感知的 RGB-D 显著性目标检测. 软件学报, 2024, 35(4): 1899–1913. http://www.jos.org.cn/1000-9825/6833.htm [doi: 10.13328/j.cnki.jos.006833]

[23] 张拯明, 郭燕, 马翠霞, 邓小明, 王宏安. GT-4S: 基于图 Transformer 的场景草图语义分割. 软件学报, 2025, 36(3): 1375–1389. http://www.jos.org.cn/1000-9825/7155.htm [doi: 10.13328/j.cnki.jos.007155]

[38] 杨锁荣, 杨洪朝, 申富饶, 赵健. 面向深度学习的图像数据增强综述. 软件学报, 2025, 36(3): 1390–1412. http://www.jos.org.cn/1000-9825/7263.htm [doi: 10.13328/j.cnki.jos.007263]

[77] 桂韬, 奚志恒, 郑锐, 刘勤, 马若恬, 伍婷, 包容, 张奇. 基于深度学习的自然语言处理鲁棒性研究综述. 计算机学报, 2024, 47(1): 90–112. [doi: 10.11897/SP.J.1016.2024.00090]

[81] 石泽男, 陈海鹏, 张冬, 申铉京. 预训练驱动的多模态边界感知视觉 Transformer. 软件学报, 2023, 34(5): 2051–2067. http://www.jos.org.cn/1000-9825/6768.htm [doi: 10.13328/j.cnki.jos.006768]

## 作者简介

**胡晨星**, 博士生, CCF 学生会员, 主要研究领域为微表情分析, 计算机视觉, 智能教育.

**苗启广**, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为智能图像处理, 模式识别, 高性能计算.

**刘璨宇**, 硕士生, 主要研究领域为智能人机交互, 强化学习.

**刘旭杰**, 博士生, CCF 学生会员, 主要研究领域为多模态大模型, 自然语言处理, 深度学习.

**刘如意**, 博士, 副教授, CCF 高级会员, 主要研究领域为遥感图像分析和处理, 计算机视觉, 智能教育.

**王泉**, 博士, 教授, CCF 会士, 主要研究领域为嵌入式系统与安全, 人机交互, 可穿戴计算.

**杨宗凯**, 博士, 教授, 博士生导师, 主要研究领域为教育技术, 人工智能.