

人脸微表情分析方法综述

于明¹, 钟元想¹, 王岩²

(1. 河北工业大学 人工智能与数据科学学院, 天津 300401; 2. 天津商业大学 信息工程学院, 天津 300134)

摘要: 微表情分析在医学、公共安全、商业谈判等领域得到广泛应用并备受关注。微表情运动幅度小、变化快, 导致人工分析难度较大, 开发一个可靠的自动化微表情分析系统非常有必要。随着计算机视觉技术的发展, 研究人员能够结合相关算法捕捉微表情运动变化特征以用于微表情分析。阐述微表情分析的发展历程和现状, 从多个角度对微表情分析的两大分支, 即微表情检测方法和微表情识别方法进行总结。整理现有微表情数据集以及微表情分析流程中常用的面部图像预处理方法。根据特征提取方式的不同, 从基于时间特征、基于特征变化和基于深度特征这3个方面对微表情检测方法进行阐述。将微表情识别方法归纳为基于纹理特征和基于光流特征的传统机器学习方法以及深度学习方法, 其中, 基于深度学习的微表情识别包括基于运动单元、基于关键帧和基于迁移学习的方法。通过不同实验指标对以上方法进行分析和比较, 在此基础上, 探讨当前微表情分析中存在的问题和挑战并展望该领域未来的发展方向。

关键词: 微表情分析; 计算机视觉; 微表情检测; 微表情识别; 深度学习

开放科学(资源服务)标志码(OSID):



中文引用格式: 于明, 钟元想, 王岩. 人脸微表情分析方法综述[J]. 计算机工程, 2023, 49(2): 1-14.

英文引用格式: YU M, ZHONG Y X, WANG Y. A survey of facial micro-expression analysis methods [J]. Computer Engineering, 2023, 49(2): 1-14.

A Survey of Facial Micro-Expression Analysis Methods

YU Ming¹, ZHONG Yuanxiang¹, WANG Yan²

(1. School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China;

2. School of Information Engineering, Tianjin University of Commerce, Tianjin 300134, China)

[Abstract] Micro-expression analysis has been widely used in the medical industry, public safety, business disputes and other fields and has attracted much attention. Manual analysis is challenging because the intensity of micro-expressions is relatively modest and they fluctuate quickly. Therefore, automated methods must be developed to perform micro-expression analysis. Driven by advancements in learning models, many works on computer vision have considered the use of subtle facial movements to investigate micro-expressions. This study thoroughly describes the evolution of micro-expression analysis. The two key issues involved, including detection and identification methods, are summarized rigorously from various perspectives in light of the current state of the art. First, existing micro-expression datasets and facial image preprocessing methods that are commonly used in micro-expression analysis are summarized and introduced. Then, according to different feature extraction methods, micro-expression detection methods are divided into those that consider temporal features, feature change, and depth features. Micro-expression recognition techniques are largely based on texture features, optical flow features of traditional machine learning methods, and deep learning methods. Deep learning methods have been developed for micro-expression recognition based on motion units, key frames, and transfer learning methods. Thus, several experimental indicators are analyzed and compared. In conclusion, the current problems and challenges in micro-expression analysis are discussed and some possible directions for future development are suggested.

[Key words] micro-expression analysis; computer vision; micro-expression detection; micro-expression recognition; deep learning

DOI: 10.19678/j.issn.1000-3428.0065790

基金项目: 河北省自然科学基金(F2019202381, F2019202464, F2020202025, F2021202030)。

作者简介: 于明(1964—), 男, 教授、博士, 主研方向为图像处理、模式识别、智能感知与优化算法; 钟元想, 硕士研究生; 王岩, 讲师、博士。

收稿日期: 2022-09-19 修回日期: 2022-11-18 E-mail: yuming@hebut.edu.cn

0 概述

面部表情、肢体语言、语音等能够反映人类的真实情感,其中,面部表情是人类情感的重要表达方式之一。很多学者利用计算机视觉技术对面部表情进行分析,进而将其应用到医疗服务^[1]、学生课堂质量检测^[2]、测谎^[3]等人机交互系统中。

面部表情分为宏表情和微表情。宏表情是主动发生的,可以伪造,有着较大的运动幅度。微表情持续时间较短,运动幅度较小,是下意识发生的,表示一个人在尝试抑制真实情感时的一种情感泄露,其可以揭示人类的真实情感,因此,分析微表情能够发现潜在的心理活动。

2003年,EKMAN开发了微表情培训工具(Micro-Expression Training Tool, METT),利用该工具可以进行微表情识别^[4]。然而,这种方式耗费时间过长,且只能取得50%左右的准确率。因此,众多学者尝试开发基于计算机视觉技术的微表情自动分析系统。在研究初期,主要采用结合机器学习的多阶段训练方式,先提取具有判别性的微表情运动特征,再训练一个可靠的分类器。芬兰奥卢大学的赵国英研究团队^[5-7]采用局部二值模式(Local Binary Pattern, LBP)以及扩展算法捕捉面部纹理信息。中国科学院心理所的王甦菁团队提出主方向平均光流(Main Directional Mean Optical-flow, MDMO)^[8]、张量独立的颜色空间^[9]等方法;山东大学的贾颀焯团队针对高维张量数据不易提取判别信息的问题,研发基于张量的最大边缘投影用于微表情识别^[10]。

随着卷积神经网络(Convolutional Neural Networks, CNN)在各个领域的广泛应用,许多研究人员将CNN与长短期记忆(Long Short-Term Memory, LSTM)网络相结合,研发端到端的微表情分析系统。KHOR等^[11]将CNN与LSTM相结合提取微表情时空特征,但是该

方法先提取空间特征再提取时序特征。LO等^[12]采用3DCNN网络表达空间信息与时间信息之间的关系。近年来,Transformer因其能够发现图像全局依赖关系而在相关领域取得了良好表现,ZHOU等^[13]尝试利用Transformer的这一优势,挖掘微表情面部全局运动关联。

微表情分析包括检测和识别两个部分。微表情检测是在微表情视频中判断是否存在微表情并定位时间位置。微表情识别是对存在微表情的图像或视频进行分类。本文分析微表情数据集和面部图像序列预处理方法,总结近年来微表情检测和微表情识别方面的研究成果。首先分析基于传统机器学习和基于深度学习的微表情检测方法,其中,将机器学习方法细分为基于时间特征和基于特征变化的方法;然后比较基于纹理特征、基于光流特征和基于深度学习的微表情识别方法,在深度学习微表情识别方法中,从面部区域运动、关键帧、样本量等角度进行区分;最后通过多个实验指标比较各类方法的特点,并对该领域未来的发展方向进行展望。

1 微表情数据集

微表情分析技术的发展依赖于完善的数据集,现有微表情数据集可分为两类:一类是非自发的微表情数据集,包括美国南佛罗里达大学创建的USF-HD^[14]、日本筑波大学创建的Polikovsky数据集^[15];另一类是自发的微表情数据集,包括芬兰奥卢大学创建的SMIC^[16]、中国科学院创建的CASME^[17]、CASMEII^[18]、CAS(ME)^{2[19]}、CAS(ME)^{3[20]}、英国曼彻斯特城市大学创建的SAMM^[21]、山东大学创建的MMEW^[22]、加拿大PORTER团队创建的York DDT^[23]。表1对上述数据集进行分析与归纳,其中,FACS(Facial Action Coding System)表示面部动作编码系统。

表1 微表情数据集

Table 1 Micro-expression datasets

数据集	子集	样本数	帧率/(帧·s ⁻¹)	表情类别	标注方式
USF-HD	—	100	29.7	—	微/非微
Polikovsky	—	42	200.0	—	FACS
York DDT	—	18	25.0	—	情绪
SMIC	HS	164	100.0	—	—
	VIS	71	25.0	3	情绪
	NIR	71	25.0	—	—
CASME	—	195	60.0	8	情绪+FACS
CASMEII	—	247	200.0	5	情绪+FACS
CAS(ME) ²	—	57	30.0	4	情绪+ FACS+视频类型
CAS(ME) ³	Part A	943	—	—	—
	Part B	—	30.0	7	情绪+ FACS
	Part C	166	—	—	—
SAMM	—	159	200.0	8	情绪+ FACS
MMEW	—	300	90.0	7	情绪+ FACS

在上述数据集中,USF-HD、Polikovsky和York DDT并未公开。USF-HD中的受试者被要求同时表现宏表情和微表情,主要用于区分宏表情和微表情。

Polikovsky中的受试者被要求模拟微表情动作,主要用于分析微表情所处阶段。York DDT由自发微表情组成,主要用于测谎。

经典的自发微表情数据集有SMIC、CASME、CASMEII、SAMM、CAS(ME)²和MMEW。SMIC有HS、VIS和NIR这3个子集,分别用不同类型的摄像机捕捉微表情,受试者由8个高加索人和8个亚洲人组成,一共有164个微表情片段,包含3个微表情类别。CASME、CASMEII和CAS(ME)²采用相同的实验方案。CASMEII是CASME的扩展版本,分别有247和195个样本;CASMEII的分辨率为280×340像素,有5类微表情;CASME的分辨率为150×190像素,有8类微表情。CAS(ME)²收集宏表情和微表情的数据,微表情分为4类,判定该数据集中表情持续时间小于0.5 s的是微表情,否则为宏表情。SAMM包含159个样本,这些样本由高速摄像机进行记录,分辨率为2 040×1 088像素,有7个微表情类别。MMEW包含300个样本,有7个微表情类别,该数据集中每个样本都由专家标定起始帧、顶点帧和终止帧,同时提供AU标注,分辨率为1 920×1 080像素,相比于其他数据集具有更详细的情感分类。

考虑到深度信息对视觉感知的影响,CAS(ME)³对深度信息进行采集,可以更加敏锐地捕捉人脸信息的变化,该数据集包含7个微表情类别,分为Part A、Part B、Part C三部分,Part A和Part B采用观看视频时保持中性表情的方式诱导微表情,Part A中包含943个有标签微表情样本,Part B中包含1 508个无标签样本,Part C中包含166个样本,其采用模拟犯罪的诱发范式诱导微表情,得到了高生态效度的数据集,为在真实环境中进行微表情分析提供了基础。

数据集中的评估方式包括单一数据集验证方法和跨数据集验证方法。在单一数据集验证方法中,采用留一法、留一视频法和k折交叉验证法将微表情数据集划分为训练集和测试集。在跨数据集验证方法中,

采用复合数据集验证法和保留数据集验证法,其中,复合数据集验证法将多个数据集进行合并再利用留一法划分训练集和测试集,保留数据集验证法将一个数据集作为训练集,另外一个数据集作为测试集。

通常采用的实验指标包括准确率、精确率、未加权F1分数、加权F1分数、未加权平均召回率和加权平均召回率。

2 面部图像序列预处理

原始图像序列是在一定的环境背景下拍摄而成的,不可避免地存在噪声,对图像序列进行预处理,消除无关信息,保留面部关键信息,有利于特征提取和微表情分类。预处理的关键技术主要包括人脸检测、人脸裁剪和对齐、帧归一化、面部图像分块。

人脸检测方法有基于方向梯度直方图(Histogram of Oriented Gradients, HOG)特征的人脸检测法^[24]、基于特征金字塔网络的方法^[25]等;人脸裁剪和对齐方法有活动形状模型(Active Shape Models, ASM)^[26]、约束局部模型(Constraint Local Model, CLM)^[27]等;帧归一化方法有时域插值模型等;面部图像分块方法有象限分割、网格分割、德劳奈三角测量、基于FACS的分割等^[28]。

3 微表情检测

微表情检测的目标是判断视频中是否存在微表情,若存在,则对微表情进行定位,找到微表情起始点、顶点、终止点。起始点是微表情运动的开始,顶点是微表情肌肉运动幅度达到最大值的时间点,终止点是微表情结束的时间点。微表情序列检测方法可以分为基于传统机器学习特征的检测方法和基于深度特征的检测方法两类。表2和表3分别对两种检测方法进行总结。

表2 基于传统机器学习特征的微表情检测方法

Table 2 Micro-expression detection methods based on traditional machine learning features

分类	方法来源	特征	解决问题	优点	缺点
基于时间特征的方法	文献[14,29]	光学应变	区分宏、微表情		
	文献[30]	光流场	检测起始点、顶点、终止点		
	文献[31]	HOOF	检测顶点		
	文献[32]	结合幅值与角度的光流特征	检测起始点、顶点、终止点	能够捕捉细微的表情变化	需要人工设定阈值
	文献[33]	MDMD	在长视频中检测面部运动		
	文献[34]	MDMD	检测微表情区间		
	文献[35]	结合幅值与角度的光流特征	检测起始点、顶点、终止点	降低噪声影响、鲁棒性强	滤波器可靠性较差
	文献[36]	密集光流	检测微表情区间		需要人工设定阈值
基于特征变化的方法	文献[15,37]	HOG	检测起始点、顶点、终止点	设计简洁	不适用于自发微表情
	文献[38]	HOG			
	文献[5]	LBP	检测微表情区间		
	文献[39]	几何形变特征		能够提取有效的运动信息	需要人工设定阈值
	文献[40]	LBP、HOOF	检测起始点、顶点、终止点		
	文献[41]	CLM、LBP			
	文献[42]	LBP、CLM、OS	检测顶点	无须人工设置阈值或参数	只能检测顶点
	文献[43]	LBP、MDMO			

表3 基于深度特征的微表情检测方法

Table 3 Micro-expression detection methods based on depth features

方法来源	特征	解决问题	优点	缺点
文献[44]	CNN	检测顶点	突出关键运动信息	只检测顶点
文献[45]	手工特征+CNN	检测顶点	突出关键运动信息	只检测顶点
文献[46]	光流+LSTM	得到时间特征	计算复杂度低	滤波器不稳定
文献[47]	全局+局部 CNN	检测单帧图片	能够捕捉微表情细微的运动、	耗时、只能检测单帧图片
文献[48]	光流+CNN	检测顶点	突出关键运动信息	只检测顶点
文献[49]	2DCNN+1DCNN	得到时间特征	得到时空特征	过拟合
文献[50]	光流+CNN	检测顶点、得到时间特征	得到时空特征	耗时
文献[51]	CNN	检测顶点	局部+全局特征、鲁棒性强	耗时
文献[52]	手工特征+CNN	得到时间特征	降低噪声影响	滤波器不稳定
文献[13]	CNN	检测顶点	突出帧间相关性	耗时、需要人工设定阈值

3.1 基于传统机器学习特征的检测方法

在基于传统机器学习特征的检测方法中,一类方法在时间维度上检测微表情,另一类方法通过计算特征变化来检测微表情。

3.1.1 基于时间特征的微表情检测方法

SHREVE等^[14,29]利用中心差分法计算下巴、嘴、脸颊、前额这些区域的光流场,得到面部产生的运动强度,以此来检测微表情。该方法可以判断视频序列中是否存在微表情,但是需要通过人工设定阈值,在训练数据较少的情况下,无法保证所选择阈值的可靠性。对应区域点的运动向量可以用 $\mathbf{u}=[u, v]^T$ 表示,有限应变张量 $\boldsymbol{\varepsilon}$ 可以用式(1)表示:

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{xx} = \frac{\partial u}{\partial x} & \varepsilon_{xy} = \frac{1}{2} \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \\ \varepsilon_{yx} = \frac{1}{2} \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right) & \varepsilon_{yy} = \frac{\partial v}{\partial y} \end{bmatrix} \quad (1)$$

其中: \mathbf{u} 、 \mathbf{v} 表示每个点的运动向量; ε_{xx} 和 ε_{yy} 表示法向应变分量; ε_{xy} 和 ε_{yx} 表示切向应变分量。

PATEL等^[30]利用运动特征获取方向连续性进行微表情检测,先将判别响应图拟合模型^[17]结合FACS定位人脸关键点,并将关键点分为多个区域,再计算每个区域的平均运动,最后通过累加每个区域的运动幅值找到相应区域的运动顶点。通过计算平均运动并设定阈值,减少头部运动、眨眼等动作对微表情检测的影响。然而,不同区域之间存在隐藏关联,该方法仅依靠单一区域检测微表情,无法充分体现微表情的区域运动关联。

MA等^[31]统计运动单元发生的区域,接着根据运动出现的频率选择感兴趣区域(Regions of Interest, ROI),最后提取光流场,计算定向光流直方图(Histogram of Oriented Optical Flow, HOOF),自动识别顶点帧。该方法可以有效提高顶点帧检测的准确性,但其忽视了角度信息。GUO等^[32]提出结合幅值与角度的光流特征提取方法,选择眼睑、眉毛、嘴角鼻子两侧这4个ROI提取光流向量,获取可靠的运动信息,每一帧的光

流幅值 A_i 计算如式(2)所示,光流角度信息 θ_i 使用反三角函数进行计算,如式(3)所示:

$$A_i = \sqrt{p_i^2 + r_i^2} \quad (2)$$

$$\theta_i = \begin{cases} \arctan \left| \frac{r_i}{p_i} \right|, & p_i > 0, r_i > 0 \\ \frac{\pi}{2} + \arctan \left| \frac{r_i}{p_i} \right|, & p_i < 0, r_i > 0 \\ \pi + \arctan \left| \frac{r_i}{p_i} \right|, & p_i < 0, r_i < 0 \\ \frac{3\pi}{2} + \arctan \left| \frac{r_i}{p_i} \right|, & p_i > 0, r_i < 0 \end{cases} \quad (3)$$

其中: i 表示当前帧数; p_i 表示水平分量; r_i 表示垂直分量。

WANG^[33]使用光流在主方向最大差异(Main Directional Maximal Difference, MDMD)进行微表情定位,获得了更加可靠的运动特征。在输入的视频序列中,设定当前帧为 F_i , F_i 的前 k 帧为 F_{i-k} , F_i 的后 k 帧为 F_{i+k} 。将面部分为不同的区域,计算 F_{i-k} 帧与 F_i 帧、 F_{i-k} 帧与 F_{i+k} 帧的鲁棒局部光流,将光流矢量个数最多的方向定义为主方向,并根据主方向计算所有区域中的差值并进行降序排列,选择前1/3作为 F_i 帧的特征,最后根据运动方向检测微表情,检测过程如图1所示。MDMD虽然在检测长视频中的微表情时能够取得很好的效果,但是不易选择合适的 k 值, k 值过大,对光流的计算会有影响, k 值过小,则主方向上的差异也会很小,另外,该方法只能检测单一帧的面部运动。在MDMD方法的基础上,HE等^[34]增加一个后处理步骤,将原先输出帧的相邻帧都视为微表情,形成一个间隔,但是该方法同样需要手动设置参数。为了消除头部摆动造成的影响,ZHANG等^[35]从鼻子区域提取平均光流和局部运动矢量检测微表情,此外还采用多尺度滤波器提高微表情检测性能。HE等^[36]选取14个ROI捕获细微面部运动,引入密集光流来估计ROI的局部运动,结合时域变化曲线的峰值检测方法精确定位运动间隔。

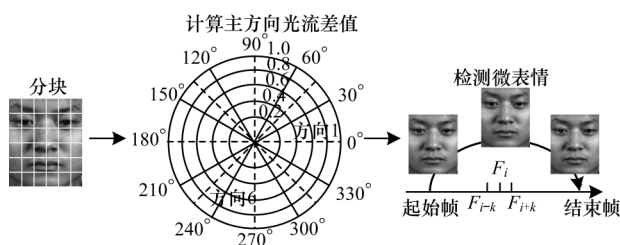


图1 MDMD方法检测流程

Fig.1 Detection procedure of MDMD method

3.1.2 基于特征变化的微表情检测方法

POLIKOVSKY等^[15,37]将人脸图像划分为12个ROI区域,计算区域中的HOG特征,通过k-means算法判断面部肌肉运动幅度的变化区间。该方法设计简洁,但是不适用于自发的微表情。DAVISON等^[38]使用HOG特征关注人脸运动,但是没有充分利用角度信息。

MOILANEN等^[5]将人脸图像分割为36个区域,计算每个区域的LBP直方图特征,观察相同的区域中序列帧的特征变化情况。此外,为了观察当前帧在图像序列中的变化强度,计算与当前帧间隔 k 帧的前后两帧的特征平均值,最后得到当前帧与平均值的卡方距离。XIA^[39]基于ASM和尺度不变特征变换描述符提取人脸关键点,接着使用Procrustes分析方法将关键点进行对齐,消除头部运动和光照变化对微表情的影响。此外,采用随机漫步模型计算帧间的变形相关性,获取转移概率。最后,设定阈值进行微表情检测。LI等^[40]利用Kanade-Lucas-Tomasi算法跟踪图像序列中的3个点,分别为2个内眼角和1个鼻脊点,根据3个点对图像序列进行校正,接着将人脸图像分为36个图像区域,计算每个区域中的LBP特征和HOF特征,以此计算每帧的特征差分值,最后根据设定的阈值检测微表情的起始点、顶点、终止点。虽然文献^[5,39-40]方法均能取得较好的实验结果,但是不易确定可靠的阈值或参数,在实际应用中有很多不确定性。

YAN等^[41]、LIONG等^[42]和HAN等^[43]分别提出定位微表情顶点的方法。文献^[41]方法使用CLM定位人脸关键点,计算图像序列与第一帧的相关性,最后根据峰值定位顶点帧。文献^[42]方法根据运动单元和人脸关键点对人脸区域进行划分,然后采用LBP、CLM、光学应变(Optical Strain, OS)提取特征,最后根据二叉搜索策略定位顶点帧。上述两种方法不需要人工设定阈值,但是只能定位顶点帧。文献^[43]提出协同特征差异算法,利用LBP特征和MDMO特征作为互补特征,将人脸划分为ROI区域,并给不同的区域分配不同的权重,从而突出关键区域,最后计算区域的特征差以定

位顶点帧。

基于机器学习的微表情检测方法能够有效提取更多的面部运动变化细节信息,基于特征变化的方法大多依赖于设定的最佳上下限阈值,上限阈值能够区分宏微表情,下限阈值能够定义微表情的最小运动振幅,但是阈值的设定具有不可靠性,面部眨眼等噪声动作也会对阈值判定造成影响。基于时间特征的方法计算复杂度高,无法满足实时检测的需求。

3.2 基于深度特征的微表情检测方法

ZHANG等^[44]首次将深度学习方法用于微表情检测,利用顶点帧的相邻帧扩充数据集,然后使用SMEConvNet网络提取特征,最后使用滑动窗口对特征矩阵进行处理,从而定位顶点帧。相较传统机器学习方法,其检测效果得到提升,但是只能检测顶点帧。TRAN等^[45]提出基于深度序列模型的微表情检测方法,采用机器学习方法提取视频序列中每个位置的时空特征,接着使用LSTM网络预测顶点帧,其改善了由帧间距离导致的网络鲁棒性差问题。DING等^[46]利用滑动窗口将微表情长视频片段分割成几个短视频,将光流与LSTM相结合,通过改进的低复杂度的光流算法提取特征曲线,接着使用LSTM预测微表情的发生,从而达到实时检测微表情的目的,但是该方法在划分短视频时采用固定大小的滑动窗口,不能很好地反映滑动窗口生成的候选片段属于微表情的程度,导致起始帧和终止帧出现定位不准确的情况。

为了解决低强度微表情对微表情定位的影响,PAN等^[47]提出一种局部双线性卷积神经网络,将微表情定位问题转变为细粒度图像分类问题,使用MDMD获取面部局部区域,分别为左眉毛、右眉毛、鼻子和嘴,再将人脸图像和这4个区域输入到网络中,分别获取全局特征和局部特征,最后融合全局特征和局部特征进行微表情检测。该方法可以捕捉微表情细微的运动,但是只能检测单帧图片。同样地,LIONG等^[48]提出的多流浅层网络也只能检测单帧微表情。WANG等^[49]提出微表情定位网络(Micro-Expression Spotting Network, MESNet),用于在长视频中定位微表情序列,其网络结构如图2所示。首先使用二维卷积网络提取空间特征,接着使用一维卷积提取时间特征,根据剪辑建议网络找出微表情序列的时间位置,最后通过分类网络判定剪辑的视频是否属于微表情。该方法具有较好的检测效果,但是模型包含了时空卷积网络模块、剪辑模块和分类回归模块,庞大的参数量导致网络效率较低,无法达到实时检测的效果。

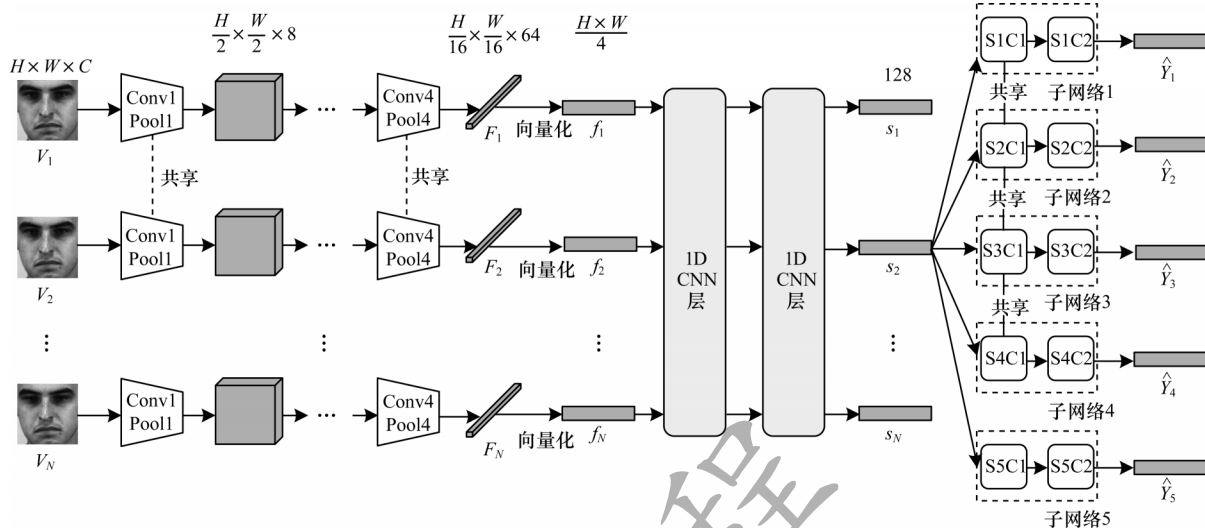


图2 MESNet网络结构

Fig.2 MESNet network structure

与文献[48]方法和文献[49]方法不同, XUE等^[50]提出两阶段的微表情定位网络, 将顶点帧定位和微表情间隔相结合, 将光流的水平分量、垂直分量以及光学应变输入到第一阶段的三流注意力网络中, 提取时空特征, 定位顶点帧。TAKALKAR等^[51]将双重注意力网络与二维卷积网络相结合, 提出一种基于双重注意力网络结构, 使用局部注意力模块关注特定区域, 同时构建全局注意力模块关注全局面部运动信息, 模型在具有较高检测效率的同时取得了较好的鲁棒性。为了降低眨眼和光照的影响, GUPTA^[52]将改进的特征编码与多尺度滤波器卷积网络相结合, 增加眉毛区域时间变形的可信度, 同时也更好地探索序列之间的时序关联, 提高了网络效率。但是, 该方法依赖精准的预处理方法对眉毛和嘴巴进行对齐和定位, 需要更多的预处理时间。ZHOU等^[13]提出一种基于双向Transformer的微表情检测网络, 首先利用候选片段生成模块生成候选段, 接着由时空特征提取模块将候选段划分为更小的时隙, 最后通过分组模块合并片段, 定位微表情的起始帧和顶点帧。该方法很好地利用了时序信息, 关注了帧间的相关性。

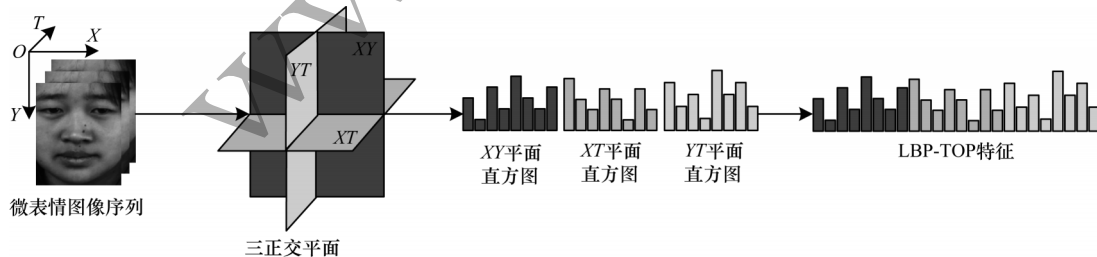


图3 LBP-TOP微表情识别流程

Fig.3 Micro-expression recognition procedure of LBP-TOP

LBP-TOP在XY、XT、YT这3个正交平面上提取LBP特征, 并将得到的结果进行拼接作为最终的LBP-TOP特征, 该方法能够有效捕捉时域信息, 但是需要计算3个

基于深度学习的微表情检测方法相较于机器学习方法, 具有更高的检测效率, 但是存在耗时、不稳定等问题。由于微表情检测模型需要提供判别模型以及剪辑模型, 存在庞大的计算量, 计算复杂度随之上升, 在部分小型网络中, 通常采用滤波器过滤峰值曲线, 但是滤波器的不稳定性会给检测结果造成影响。

4 微表情识别方法

微表情识别是指对检测到的微表情进行识别并分类, 根据特征提取的方式可以将微表情识别方法分为基于手工机器学习特征的识别方法和基于深度学习特征的识别方法。

4.1 基于手工机器学习特征的识别方法

基于手工机器学习特征的识别方法又可以分为基于纹理特征和基于光流特征的识别方法。

4.1.1 基于纹理特征的识别方法

微表情包含大量有效的时序信息, PFISTER等^[53]为了发现微表情中的运动时空关系, 利用三维正交平面的局部二值模式(Local Binary Pattern from Three Orthogonal Planes, LBP-TOP)提取特征, 其微表情识别流程如图3所示。

平面的特征, 其特征维度高, 只考虑2个像素之间的特征差异, 未考虑其他有用信息。为了解决该问题, WANG等^[54]使用中心点上3条相交线的六交点减少冗余信息,

提出六交点局部二值模式(Local Binary Pattern with Six Intersection Points, LBP-SIP)算法,该算法可以减少很多冗余信息,且计算速度快,但是其存在鲁棒性差的问题。HUANG等^[6]为了能够捕捉水平和垂直积分投影的外观和运动变化,提出积分投影时空局部二值模式(Spatio-Temporal Local Binary Pattern with Integral Projection, STLBP-IP)算法,该算法能够有效保留人脸图像的形状特征,所采用的积分投影能够更好地发现同类情绪的统计分布,但是其只关注类内信息的相似性,忽视了样本间的判别信息。之后, HUANG等^[7]提出时空完全局部量化模式(Spatio-Temporal Completed Local Quantized Patterns, STCLQP)算法,该算法利用3种有用的信息,包括基于符号、基于大小和基于方向的像素差,同时设计一种时空域的码本,在码本的基础上提取时空特征,这使得局部模式更具判别性。

融合纹理和形状信息进行识别的效果优于仅使用外观信息进行识别,为此, HUANG等^[55]提出判别性时空局部二值模式算子(Discriminative Spatio-Temporal Local Binary Pattern with Revisited Integral Projection, DISTLBP-RIP)方法,以用于微表情识别,该方法将形状属性与动态纹理信息相结合,获得更具判别性的特征。当微表情发生时,面部肌肉在斜方向上会发生偏移,但LBP-TOP只能捕捉水平和垂直方向上的信息,因此, WEI等^[56]提出五相交局部二值模式(Local Binary Pattern from Five Intersecting Planes, LBP-FIP)方法,先提取在偏离 X 或 Y 方向 45° 的斜方向上8个顶点的LBP特征,获得八顶点局部二值模式特征(Eight Vertices Local Binary Pattern features, EVLBP),然后与LBP-TOP提取的水平和垂直方向上的特征进行级联。

HOG特征可以表示图像的纹理信息, LI等^[40]利用HOG及其扩展算法作为特征描述符,提取面部结构信息,引入图像梯度直方图(Histograms of Image Gradient Orientation, HIGO),改变投票策略,降低光照对微表情识别的影响。相较LBP相关方法,该算法更加适合在彩色视频中进行微表情识别。WEI等^[57]提出单方向梯度直方图(Histogram of Single Direction Gradient, HSDG)算法,该算法在某一运动方向上提取梯度值,简化HOG,最后将HSDG与LBP-TOP特征进行级联,得到具有外观纹理和运动信息的LBP-SDG(LBP with Single Direction Gradient)特征,其捕捉到了最有利于微表情识别的运动特征,但是选取18个方向进行测试,计算复杂度过高。

4.1.2 基于光流特征的识别方法

光流特征可以有效提取微表情的时间信息。LIU等^[8]提出MDMO特征用于微表情识别,将人脸划分为36个感兴趣区域,减少了与表情无关的冗余信息。在各个区域中计算HOOF直方图,将直方图特征的最大值索引作为主方向,最后计算主方向上向量特征的均值从而进行微表情识别,其识别流程如图4所示。

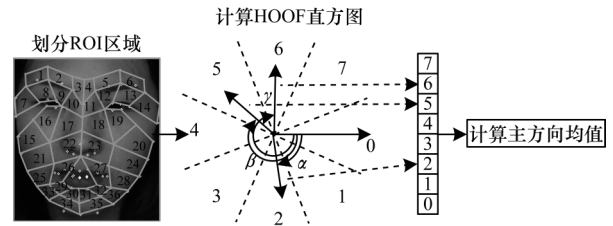


图4 MDMO微表情识别流程

Fig.4 Micro-expression recognition procedure of MDMO

MDMO考虑到了各个区域的运动信息和空间位置,易于实现且特征维数小,但是容易丢失特征空间中固有的底层流形结构,采用预定义的AU关系进行建模,导致模型泛化能力较差。LIU等^[58]在MDMO的基础上提出Sparse MDMO,能够有效揭示底层流形结构,比MDMO特征具有更强的判别能力。

为降低小强度噪声光流的影响, LIONG^[59]提出双加权定向光流(Bi-Weighted Oriented Optical Flow, Bi-WOOF),首先估计顶点帧和起始帧之间的水平和垂直光流分量,然后根据这两个分量计算每个像素点的方向、幅值和光学应变,最后通过局部加权的幅值和全局加权的的光学应变获得基于方向的Bi-WOOF直方图。然而,利用光流直方图作为分类器的特征向量时,光流直方图的轻微偏移会增大图像之间的欧氏距离,放大两个图像之间的差异,从而影响识别效果。HAPPY等^[60]提出光流方向的模糊直方图(Fuzzy Histogram of Optical Flow Orientations, FHOFO),其能忽略无关运动,对表情强度和光照的变化不敏感。WANG^[61]认为基于运动强度的权值容易受到图像噪声的影响,因此,提出一种新的基于光流时间累积的加权特征提取方法MINOF(Motion Intensities of Neighboring Optical Flows):首先计算微表情视频中的光流时间累积,以降低图像噪声的影响;接着根据光流累积大小计算各区域的运动强度,获取每个区域的权重;最后将局部特征和权重相乘生成全局特征,有效去除噪声运动的权重,提高权值的有效性。

基于手工机器学习特征的微表情识别方法关注像素点的变化情况,可以保留更多的信息,对不同的面部表情具有很好的灵活性,能够忽略光照强度的影响,但是其特征维度高,计算复杂度高,需要结合特征选择算法过滤无效信息,以降低计算复杂度。MDMO、Bi-WOOF等方法依赖于面部对齐效果,无法实现良好的微表情识别效果。上述方法的实验指标对比如表4所示,其中,ACC(Accuracy)表示准确率, F1表示F1分数。从表4可以看出:LBP-TOP、STCLQP、Sparse MDMO在SMIC数据集上分别取得了0.542、0.640、0.705的准确率,这是因为SMIC数据集只有3个情绪类别,同时采集微表情所用的近红外摄像机也减少了光照对微表情的影响;LBP-TOP、STCLQP、Sparse MDMO在CASMEII数据集上分别取得了0.464、0.584、0.670的准确率,这是由于CASMEII数据集情绪类别较完备,样本分布更加均衡。

表4 基于手工机器学习特征的微表情识别方法

Table 4 Micro-expression recognition methods based on manual machine learning features

分类	方法	数据集	实验结果	
			ACC	F1
基于纹理特征的识别方法	LBP-TOP ^[53]	CASMEII	0.464	—
		SMIC	0.542	—
	LBP-SIP ^[54]	CASMEII	0.466	—
		SMIC	0.445	—
	STLBP-IP ^[6]	CASMEII	0.595	—
		SMIC	0.579	—
	STCLQP ^[7]	CASMEII	0.584	—
		SMIC	0.640	—
	DISTLBP-RIP ^[55]	CASMEII	0.648	—
		SMIC	0.634	—
	LBP-FIP ^[56]	CASMEII	0.692	—
		SMIC	0.679	—
LBP-SDG ^[57]	CASMEII	0.713	—	
	SMIC	0.697	—	
基于光流特征的识别方法	MDMO ^[8]	CASMEII (4类)	0.572	—
		SAMM	0.500	—
	Sparse MDMO ^[58]	CASMEII	0.670	0.691
		SMIC	0.705	0.704
	Bi-WOOF ^[59]	CASMEII	0.589	0.610
		SMIC	0.620	—
	FHOFO ^[60]	CASMEII	0.559	0.520
SMIC		0.512	0.518	
MINOF ^[61]	CASMEII	0.629	—	

4.2 基于深度学习的识别方法

本节从AU、关键帧、迁移学习三方面对基于深度学习的微表情识别方法展开讨论。

4.2.1 基于AU的识别方法

在微表情运动中,分析AU能够发现潜在面部运动和情绪之间的关系。FACS对面部行为进行编码,每个编码表示一个AU,如AU1代表内部眉毛抬起、AU6代表脸颊抬起等。不同AU组合所表示的情绪如表5所示。基于图的学习算法可以在非欧氏数据中发现每个对象节点之间的关系,AU通常与图卷积(Graph Convolutional Network, GCN)相结合。基于AU的微表情识别方法实验指标对比如表6所示,其中,CASMEII和SAMM默认为五分类。

表5 情绪与AU的关系

Table 5 The relationship between emotion and AU

情绪	AU
开心	AU6+AU12
蔑视	AU12+AU14
悲伤	AU1+AU4+AU15
厌恶	AU9+AU15+AU16
生气	AU4+AU5+AU7+AU23
惊讶	AU1+AU2+AU5+AU26
恐惧	AU1+AU2+AU4+AU5+AU7+AU20+AU26

表6 基于AU的微表情识别方法性能对比

Table 6 Performance comparison of micro-expression recognition methods based on AU

年份	方法	CASMEII		SAMM	
		ACC	F1	ACC	F1
2020	MER-GCN ^[12]	0.427	—	—	—
2020	Graph-TCN ^[62]	0.740	0.725	0.750	0.700
2021	Graph-GCN ^[63]	0.743	0.705	0.743	0.705
2021	STA-GCN ^[64]	0.761 (4类)	0.710 (4类)	—	—
2022	DEF-Net ^[65]	0.636	—	—	—
2022	MER-AMRE ^[66]	0.650	—	0.533	—
2022	MFAPLF ^[67]	0.801	0.788	—	—

WANG等^[9]基于AU定义16个ROI,提取每个ROI的特征,但预定义的AU规则会导致泛化能力有限。LO等^[12]利用GCN发现AU之间的依赖关系,提出MER-GCN网络用于微表情识别,这是第一个基于GCN的端到端微表情识别网络,其结构如图5所示。

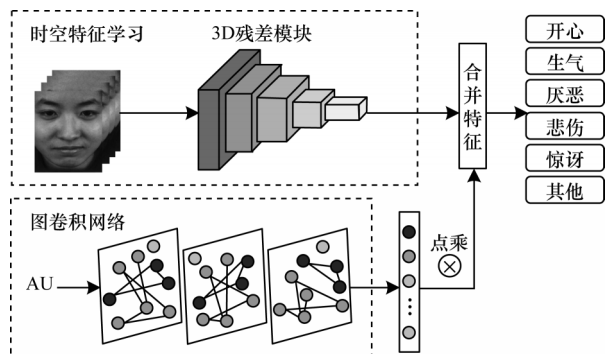


图 5 MER-GCN 网络结构

Fig.5 MER-GCN network structure

MER-GCN将每对AU的共现作为相关关系构建邻接矩阵,然后通过GCN获取不同AU之间的隐藏关系,最后将其与3DCNN残差模块提取的特征进行融合,以完成微表情识别。该方法关注了局部动作变化,但是3DCNN具有较大的参数量,同时AU关联矩阵依靠共现作为相互关系,极易受到噪声动作的影响。LEI等^[62]设计双流图时序卷积网络(Graph-TCN),通过获取微表情局部肌肉的运动特征进行微表情识别。首先选择眉毛和嘴巴区域的28个人脸关键点,提取这些关键点7×7的特征矩阵,并压缩成长度为49的特征向量,以此为基础构建图结构并输入到双通道网络中,最后将提取的节点特征和边缘特征进行融合以完成微表情识别。Graph-TCN能够更好地分析微表情的局部运动信息,具有较好的判别性,但其利用固定的扩张因子学习边缘特征,只能学习到固定区域的关联变化。LEI等^[63]提出双流网络Graph-GCN,包括图学习流和AU学习流。AU学习模型选择与眉毛和嘴巴相关的9个AU,利用AU的共现关系构建邻接矩阵,采用单词嵌入的方法构建节点矩阵,将这两个矩阵输入到双层GCN中提取特征,最后融合图学习流的特征进行分类。ZHAO等^[64]提出时空AU图卷积网络(Spatio-Temporal AU Graph Convolutional Network, STA-GCN),先利用3DCNN提取AU相关区域的时空运动信息,再通过GCN捕捉AU的依赖关系,最后与全

脸特征相乘得到激活特征进行微表情识别。SUN等^[65]提出一个双分支融合的微表情识别框架(Dual Expression Fusion micro-expression recognition framework, DEF-Net),使用OpenFace模型分析AU的发生,将深度网络提取的人脸特征与AU进行级联,捕获细微面部运动以完成微表情识别。WANG等^[66]将AU与人脸关键点相结合,构建眼睛、鼻子、脸颊和嘴巴等4个感兴趣区域,接着将相应区域进行加权,最后结合所提出的MER-AMRE(MER framework with Attention Mechanism and Region Enhancement)网络提取特征,提高了网络提取局部运动信息的能力。CEN等^[67]为了挖掘面部表情与AU之间的关联,将微表情视频分割为多个相邻视频片段,揭示三维邻域的时空特征变化情况,最后结合所提出的多任务面部活动模式学习框架(Multi-task Facial Activity Patterns Learning Framework, MFAPLF),促进同类微表情聚合。

4.2.2 基于关键帧的识别方法

文献[59,68]使用单一顶点帧进行微表情识别,大幅减少了输入帧的冗余信息,降低了特征的计算复杂度。LI等^[69]基于VGG-16 CNN架构提出一个基于局部和全局信息的学习模型(LGCcon),分别提取全局和局部特征。先利用欧拉运动放大方法将顶点帧的细微运动进行放大,再采用全局信息流提取整个面部图像的上下文信息,利用局部信息流将面部图像划分为子区域,寻找贡献最大的局部区域并提取特征,最后将局部和全局特征进行融合以完成微表情识别。实验结果表明,与完整的序列相比,单一顶点帧也能获得较好的性能,但是参数量相比于单分支结构更大,模型能够获取关联特征,但需要对全局和局部定义多重约束条件,同时需要引入多个损失函数来区分类内和类间特征,模型复杂度过高。

起始帧和顶点帧之间的运动变化在微表情识别中具有较强的判别性,文献[70-72]方法使用起始帧和顶点帧的光流提取面部的运动特征。LEI等^[63]抽取起始帧和顶点帧作为输入,设计Graph-GCN网络,先利用MagNet^[73]对顶点帧进行放大,接着将图学习模型和AUs学习模型提取的特征进行融合以完成微表情识别,其网络结构如图6所示。

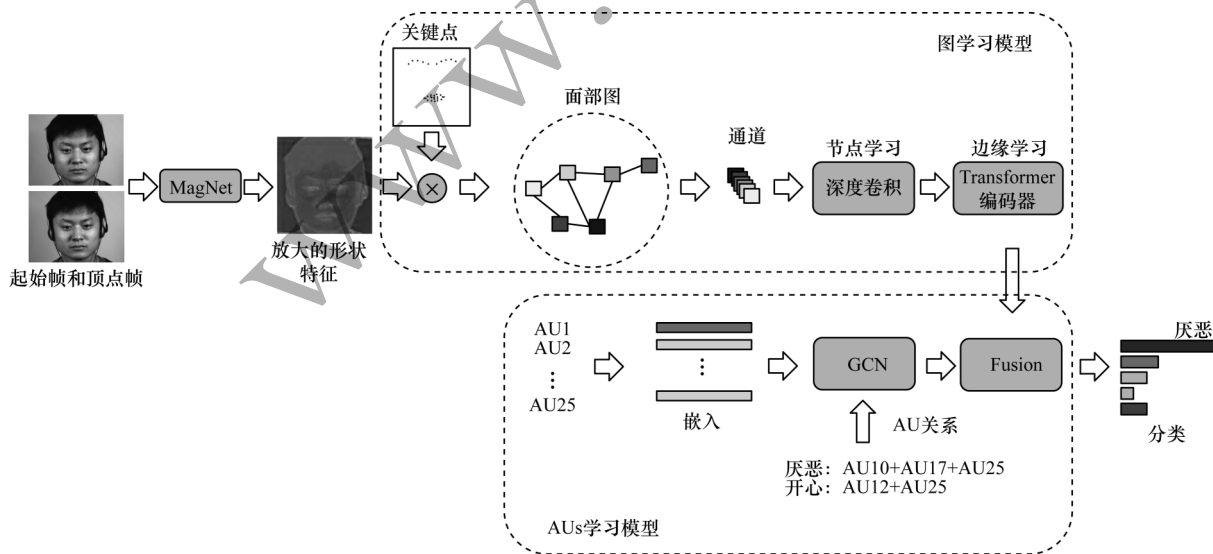


图 6 Graph-GCN 网络结构

Fig.6 Graph-GCN network structure

Graph-GCN网络可以学习到节点间的信息关联,但采用预定义的AU信息,限制了模型的学习能力。其次,Transformer层数缩减为原始模型的一半,但依然存在较大的参数量。最后,双分支融合模块仅采用简单的串联操作,增加了冗余信息。ZHAO等^[71]采用基于L1的总变差光流算法计算起始帧与顶点帧之间的运动信息,并将得到的光流特征反馈给后续的深度网络。GAN等^[74]提出基于顶点帧的光流特征网络(Optical Flow Features from Apex frame Network, OFF-ApexNet),采用起始帧和顶点帧来提取两个方向的光流,表示微表情的运动细节,输入到深度网络中进行特征增强,从而获得更具判别性的特征以完成微表情识别。文献[75]和文献[76]也采用类似的方法,将起始帧和顶点帧的

光流作为网络输入,在降低模型输入复杂性的同时又能保持较好的识别效果。CHEN等^[77]提出分块卷积网络(Block Division Convolutional Network, BDCNN),计算起始帧和顶点帧4个光流特征,接着将这4个光流图进行分块,并对这些小块进行卷积和池化操作,最后将4个光流提取的特征向量进行特征级联,以完成微表情识别。

基于关键帧的微表情识别方法能够在降低特征维度的同时保持良好的识别效果,但是会丢失较多的时序信息。各方法的实验指标对比如表7所示,其中,UAR(Unweighted Average Recall)表示未加权平均召回率,标有“*”代表三分类,SMIC为三分类。

表7 基于关键帧的微表情识别方法性能对比

Table 7 Performance comparison of micro-expression recognition methods based on key frame

特征	关键帧	数据集	实验结果		
			ACC	F1	UAR
LGCcon ^[69]	顶点帧	CASMEII	0.650	0.640	—
		CASME	0.608	0.600	—
		SAMM	0.409	0.340	—
		SMIC	0.634	0.620	—
光流+CNN ^[70]	起始帧	CASMEII*	—	0.838	0.869
	顶点帧	SAMM*	—	0.659	0.681
		SMIC	—	0.680	0.701
光流+CNN ^[72]	起始帧	CASMEII*	0.876	—	0.839
	顶点帧	SAMM*	0.684	—	-0.653
		SMIC*	0.628	—	0.601
OFF-ApexNet ^[74]	起始帧	CASMEII*	0.883	0.870	—
	顶点帧	SAMM*	0.681	0.542	—
		SMIC	0.677	0.671	—
RCN ^[75]	起始帧	CASMEII	—	—	0.812
	顶点帧	SAMM	—	—	0.672
		SMIC	—	—	0.644
JFM ^[76]	起始帧	CASMEII*	—	—	0.887
	顶点帧	SAMM*	—	—	0.731
		SMIC	—	—	0.800
BDCNN ^[77]	起始帧	CASMEII	—	—	0.697
	顶点帧	SAMM	—	—	0.660
		SMIC	—	—	0.724

4.2.3 基于迁移学习的识别方法

迁移学习方法通常将知识从具有大样本容量的源域迁移到目标域中。现有的微表情数据集较小,而宏表情数据集包含大量的训练样本。考虑到宏表情与微表情存在相关性,通常会结合宏表情数据集进行训练来提高微表情的识别性能。

文献[68,78]方法利用迁移学习的思想,结合宏表情数据集提高微表情识别性能。文献[79]方法使用3种有效的二值面部描述子提取特征,将这些特征输入到源域和目标域共享的公共子空间中进行学习。文献[80]方法引入表情身份分离网络(Expression-Identity Disentangle Network, EIDNet)作为特征提取器,分别用宏表情和微表情数据集对两个EIDNet进行训练,分别

命名为MacroNet和MicroNet,固定MacroNet并对MicroNet进行微调,MicroNet可以从宏表情样本中学习共享知识,并采用不等式正则化损失使MicroNet的输出收敛于MacroNet的输出。EIDNet利用从宏表情样本学习到的知识指导微表情网络的调整,提高了微表情的识别性能,但整个网络模型包含多个模块,模型设计较为复杂,且训练过程较为繁琐,很难满足实时性的需求。

文献[81]方法在CK+、Oulu-CASIA NIR&VIS、Jaffe和MUGFE这4个宏表情数据集^[82]上进行训练,接着利用微表情数据集对网络进行微调,然而该网络仅在注意力方面做出了改进,且在经过宏表情预训练后再对微表情进行训练,不能很好地突出宏微表

情之间的关联。TANG等^[83]首次将迁移学习和GCN相结合应用于微表情识别领域,提出迁移双流随机图卷积网络(Transferring Dual Stochastic Graph Convolutional Network, TDSGCN),利用宏表情数据集对源网络进行训练,然后用微表情数据集对网络

进行微调,最后完成微表情识别。ZHANG等^[84]提出的运动放大特征关系网络利用迁移的ResNet50来提取全局特征,在一定程度上避免了过拟合问题。

基于迁移学习的微表情识别方法性能对比如表8所示。

表8 基于迁移学习的微表情识别方法性能对比

Table 8 Performance comparison of micro-expression recognition methods based on transfer learning

特征	宏表情数据集	微表情数据集	实验结果	
			ACC	F1
ResNet ^[68]	BU-3DFE	CASMEII(3类)	—	0.761
		SAMM(3类)	—	0.448
		SMIC(3类)	—	0.551
3DCNN ^[78]	CK+	CASMEII	0.676	—
DCP-TOP & HWP-TOP ^[79]	CK+	CASMEII	0.607	—
EIDNet ^[80]	CK+	CASMEII	0.756	0.701
	MMI	SAMM	0.741	0.736
	Oulu-CASIA	SMIC	0.768	0.744
ResNet ^[81]	CK+	CASMEII	0.659	0.539
	Oulu-CASIA	SAMM	0.485	0.402
	Jaffe	SMIC	0.494	0.496
	MUGFE	SMIC	0.494	0.496
TDSGCN ^[83]	MMEW(Macro)	MMEW	0.727	—
	CK+	SAMM	0.750	—

基于深度学习的微表情识别方法由多段训练发展到如今的端到端训练,可以提取微表情更深层次的信息,同时也避免了复杂的手工特征描述,但深度网络需要依靠一个更大更真实的微表情数据集才能获得较好的性能。Graph-TCN、LGCcon、Graph-GCN等方法关注局部区域运动信息以挖掘微表情信息,这一点与手工特征方法中依赖面部网格分块突出局部信息的思想一致。在深度学习方法中,Graph-TCN、Graph-GCN、OFF-ApexNet在CASMEII数据集上分别取得了0.740、0.743、0.883的准确率,这是因为CASMEII数据集中有AU标注,能够突出运动区域;Graph-TCN、Graph-GCN、OFF-ApexNet在SAMM数据集上分别取得了0.750、0.743、0.681的准确率,这是由于SAMM具有较为完备的情绪分类,同时也有较高的帧率。但是,这一类方法依然受限于现有微表情数据集的数据量,在现实环境中泛化能力较差。

5 总结与展望

微表情检测和识别对人机交互的发展具有重大意义,拥有广阔的应用前景,未来会有更多的研究人员从各个角度对微表情展开研究。但是,目前微表情检测和识别中还有一些亟待解决的问题,针对这些问题,本文总结了该领域未来可能的发展方向,具体如下:

1)生成高质量微表情数据。为了弥补微表情数据量有限和部分微表情图像质量过低的问题,已有方法开始尝试采用生成对抗网络生成微表情数据,但是在该过程中存在两个问题:一是生成的微表情图像对网络性能提升效果有限,说明生成的AU关系不够准确,同时在生成新的图像时会引入噪声,从而

对分类器造成影响,在未来的工作中,利用GAN生成高质量的微表情图像也是一个值得研究的问题;二是在FACS中定义了微表情与面部动作的关联,已有工作利用这一关联作为先验知识解决微表情分析问题,事实上,这一关联也直观解释了微表情发生机理,但现有方法采用的GAN并没有与之相结合,因此,如何判定生成的图像是否符合微表情发生机理,还没有一个规范的评价标准,这也是未来的一个工作方向。

2)多模态微表情分析。目前微表情分析算法仅依靠单一面部表情进行分析,但在实际生活中,微表情的发生常伴随着肢体语言以及语音等状态的变化。在未来的研究中,需要充分利用肢体语言等相关信息,充分体现微表情的发生机理,揭示多模态数据的内在关联。

3)采用其他学习策略。已有方法都是采用有监督的方式训练网络,对数据量有一定的依赖性。在今后的研究中,可以考虑采用自监督的方式,减少网络对数据量的依赖,提升所学习到特征的适应性。另外,现有的微表情数据库存在类别不平衡的情况,这包括数量不平衡和识别难度不平衡,今后除了关注样本数量外,还需要结合样本的识别难度来动态调整学习过程。

6 结束语

与人工分析微表情相比,基于计算机视觉的微表情分析方法具有较大优势,本文对基于传统机器学习与基于深度学习的微表情检测与识别方法进行了分析和总结。通过对这些方法的分析比较发现,基于机器学习的方法关注像素点的变化情况,能够保留更多的有效信息,鲁棒性较高,并且可以满足实时

性的需求,但是该类方法在预处理阶段将面部图像划分为固定网格,难以体现面部运动的发生机理,存在精度低、过程繁杂等问题。相较机器学习方法,深度学习方法具有精度高、易于捕捉上下文信息的优点,但是存在容易过拟合、耗时、依赖数据量等缺点。在未来,微表情分析会有更多的应用需求,对性能也会有更高的要求,下一步将对基于深度特征的微表情分析方法进行深入探究,以开发具有高性能、高精度和强鲁棒性的系统。

参考文献

- [1] EKMAN P, FRIESEN W V. Nonverbal leakage and clues to deception[J]. *Psychiatry*, 1969, 32(1): 88-106.
- [2] WHITEHILL J, SERPELL Z, LIN Y C, et al. The faces of engagement: automatic recognition of student engagement from facial expressions[J]. *IEEE Transactions on Affective Computing*, 2014, 5(1): 86-98.
- [3] EKMAN P, O'SULLIVAN M. From flawed self-assessment to blatant whoppers: the utility of voluntary and involuntary behavior in detecting deception[J]. *Behavioral Sciences & the Law*, 2006, 24(5): 673-686.
- [4] ZHANG M, FU Q, CHEN Y H, et al. Emotional context influences micro-expression recognition[J]. *PLoS One*, 2014, 9(4): e95018.
- [5] MOILANEN A, ZHAO G Y, PIETIKÄINEN M. Spotting rapid facial movements from videos using appearance-based feature difference analysis[C]//*Proceedings of the 22nd International Conference on Pattern Recognition*. Washington D. C., USA: IEEE Press, 2014: 1722-1727.
- [6] HUANG X H, WANG S J, ZHAO G Y, et al. Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection[C]//*Proceedings of IEEE International Conference on Computer Vision Workshop*. Washington D. C., USA: IEEE Press, 2015: 1-9.
- [7] HUANG X H. Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns[J]. *Neurocomputing*, 2016, 175: 564-578.
- [8] LIU Y J, ZHANG J K, YAN W J, et al. A main directional mean optical flow feature for spontaneous micro-expression recognition[J]. *IEEE Transactions on Affective Computing*, 2016, 7(4): 299-310.
- [9] WANG S J, YAN W J, LI X B, et al. Micro-expression recognition using color spaces[J]. *IEEE Transactions on Image Processing*, 2015, 24(12): 6034-6047.
- [10] BEN X Y, ZHANG P, YAN R, et al. Gait recognition and micro-expression recognition based on maximum margin projection with tensor representation[J]. *Neural Computing and Applications*, 2016, 27(8): 2629-2646.
- [11] KHOR H Q, SEE J, PHAN R C W, et al. Enriched long-term recurrent convolutional network for facial micro-expression recognition[C]//*Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition*. Washington D. C., USA: IEEE Press, 2018: 667-674.
- [12] LO L, XIE H X, SHUAI H H, et al. MER-GCN: micro-expression recognition based on relation modeling with graph convolutional networks[C]//*Proceedings of IEEE Conference on Multimedia Information Processing and Retrieval*. Washington D. C., USA: IEEE Press, 2020: 79-84.
- [13] ZHOU Y, SONG Y, CHEN L, et al. A novel micro-expression detection algorithm based on BERT and 3DCNN[J]. *Image and Vision Computing*, 2022, 119: 104378.
- [14] SHREVE M, GODAVARTHY S, GOLDFOLD, et al. Macro and micro-expression spotting in long videos using spatio-temporal strain[C]//*Proceedings of IEEE International Conference on Automatic Face & Gesture Recognition*. Washington D. C., USA: IEEE Press, 2011: 51-56.
- [15] POLIKOVSKY S, KAMEDA Y, OHTA Y. Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor[J]. *Journal of Nonverbal Behavior*, 2009, 33(1): 59-69.
- [16] LI X B, PFISTER T, HUANG X H, et al. A spontaneous micro-expression database: inducement, collection and baseline[C]//*Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*. Washington D. C., USA: IEEE Press, 2013: 1-6.
- [17] ASTHANA A, ZAFEIRIOU S, CHENG S Y, et al. Robust discriminative response map fitting with constrained local models[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2013: 3444-3451.
- [18] YAN W J, LI X, WANG S J, et al. CASMEII: an improved spontaneous micro-expression database and the baseline evaluation[J]. *PLoS One*, 2014, 9(1): e86041.
- [19] QU F B, WANG S J, YAN W J, et al. CAS(ME)²: a database for spontaneous macro-expression and micro-expression spotting and recognition[J]. *IEEE Transactions on Affective Computing*, 2018, 9(4): 424-436.
- [20] LI J T, DONG Z Z, LU S Y, et al. CAS(ME)³: a third generation facial spontaneous micro-expression database with depth information and high ecological validity[EB/OL]. [2022-08-05]. <https://pubmed.ncbi.nlm.nih.gov/35560102/>.
- [21] DAVISON J A K, LANSLEY C, COSTEN N, et al. SAMM: a spontaneous micro-facial movement dataset[J]. *IEEE Transactions on Affective Computing*, 2018, 9(1): 116-129.
- [22] BEN X Y, REN Y, ZHANG J P, et al. Video-based facial micro-expression analysis: a survey of datasets, features and algorithms[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(9): 5826-5846.
- [23] WARREN G, SCHERTLER E, BULL P. Detecting deception from emotional and unemotional cues[J]. *Journal of Nonverbal Behavior*, 2009, 33(1): 59-69.
- [24] DADI H S, MOHAN PILLUTLA G K. Improved face recognition rate using HOG features and SVM classifier[J]. *IOSR Journal of Electronics and Communication Engineering*, 2016, 11(4): 34-44.
- [25] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2017: 936-944.
- [26] COOTES T F. Active shape models-their training and application[J]. *Computer Vision and Image Understanding*, 1995, 61(1): 38-59.
- [27] LUCEY S, WANG Y, COX M, et al. Efficient constrained local model fitting for non-rigid face alignment[J]. *Image and Vision Computing*, 2009, 27(12): 1804-1813.
- [28] MERGHANI W, YAP M H. Adaptive mask for region-based facial micro-expression recognition[C]//*Proceedings of 2020 IEEE International Conference on Automatic Face*

- and Gesture Recognition. Washington D. C. , USA ; IEEE Press, 2020; 765-770.
- [29] SHREVE M, GODAVARTHY S, MANOHAR V, et al. Towards macro- and micro-expression spotting in video using strain patterns [C]//Proceedings of Workshop on Applications of Computer Vision. Washington D. C. , USA ; IEEE Press, 2009; 1-6.
- [30] PATEL D, ZHAO G, PIETIK INEN M. Spatiotemporal integration of optical flow vectors for micro-expression detection [C]//Proceedings of International Conference on Advanced Concepts for Intelligent Vision Systems. Berlin, Germany ; Springer, 2015; 369-380.
- [31] MA H Y, AN G Y, WU S J, et al. A Region Histogram of Oriented Optical Flow (RHOOF) feature for apex frame spotting in micro-expression [C]//Proceedings of International Symposium on Intelligent Signal Processing and Communication Systems. Washington D. C. , USA ; IEEE Press, 2017; 281-286.
- [32] GUO Y F, LI B, BEN X Y, et al. A magnitude and angle combined optical flow feature for microexpression spotting [J]. IEEE MultiMedia, 2021, 28(2) ; 29-39.
- [33] WANG S J. A main directional maximal difference analysis for spotting facial movements from long-term videos [J]. Neurocomputing, 2017, 230; 382-389.
- [34] HE Y, WANG S J, LI J T, et al. Spotting macro- and micro-expression intervals in long video sequences [C]// Proceedings of the 15th IEEE International Conference on Automatic Face and Gesture Recognition. Washington D. C. , USA ; IEEE Press, 2020; 742-748.
- [35] ZHANG L W, LI J T, WANG S J, et al. Spatio-temporal fusion for macro- and micro-expression spotting in long video sequences [C]//Proceedings of the 15th IEEE International Conference on Automatic Face and Gesture Recognition. Washington D. C. , USA ; IEEE Press, 2020; 734-741.
- [36] HE Y, XU Z, MA L, et al. Micro-expression spotting based on optical flow features [J]. Pattern Recognition Letters, 2022, 163; 57-64.
- [37] POLIKOVSKY S, KAMEDA Y, OHTA Y. Detection and measurement of facial micro-expression characteristics for psychological analysis [J]. Kameda's Publication, 2010, 110; 57-64.
- [38] DAVISON A K, YAP M H, LANSLEY C. Micro-facial movement detection using individualised baselines and histogram-based descriptors [C]//Proceedings of 2015 IEEE International Conference on Systems, Man, and Cybernetics. Washington D. C. , USA ; IEEE Press, 2015; 1864-1869.
- [39] XIA Z Q. Spontaneous micro-expression spotting via geometric deformation modeling [J]. Computer Vision and Image Understanding, 2016, 147; 87-94.
- [40] LI X B, HONG X P, MOILANEN A, et al. Towards reading hidden emotions: a comparative study of spontaneous micro-expression spotting and recognition methods [J]. IEEE Transactions on Affective Computing, 2018, 9(4); 563-577.
- [41] YAN W J, WANG S J, CHEN Y H, et al. Quantifying micro-expressions with constraint local model and local binary pattern [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany ; Springer, 2015; 296-305.
- [42] LIONG S T, SEE J, WONG K, et al. Automatic apex frame spotting in micro-expression database [C]//Proceedings of the 3rd IAPR Asian Conference on Pattern Recognition. Washington D. C. , USA ; IEEE Press, 2015; 665-669.
- [43] HAN Y H, LI B J, LAI Y K, et al. CFD: a collaborative feature difference method for spontaneous micro-expression spotting [C]//Proceedings of the 25th IEEE International Conference on Image Processing. Washington D. C. , USA ; IEEE Press, 2018; 1942-1946.
- [44] ZHANG Z H, CHEN T, MENG H Y, et al. SMEConvNet: a convolutional neural network for spotting spontaneous facial micro-expression from long videos [J]. IEEE Access, 2018, 6; 71143-71151.
- [45] TRAN T K, VO Q N, HONG X P, et al. Dense prediction for micro-expression spotting based on deep sequence model [J]. Electronic Imaging, 2019, 31(8) ; 1-6.
- [46] DING J, TIAN Z, LÜ X, et al. Real-time micro-expression detection in unlabeled long videos using optical flow and LSTM neural network [C]//Proceedings of International Conference on Computer Analysis of Images and Patterns. Berlin, Germany ; Springer, 2019; 622-634.
- [47] PAN H, XIE L, WANG Z L. Local bilinear convolutional neural network for spotting macro- and micro-expression intervals in long video sequences [C]//Proceedings of 2020 IEEE International Conference on Automatic Face and Gesture Recognition. Washington D. C. , USA ; IEEE Press, 2020; 749-753.
- [48] LIONG G B, SEE J, WONG L K. Shallow optical flow three-stream CNN for macro- and micro-expression spotting from long videos [C]//Proceedings of IEEE International Conference on Image Processing. Washington D. C. , USA ; IEEE Press, 2021; 2643-2647.
- [49] WANG S J, HE Y, LI J T, et al. MESNet: a convolutional neural network for spotting multi-scale micro-expression intervals in long videos [J]. IEEE Transactions on Image Processing, 2021, 30; 3956-3969.
- [50] XUE L Y, ZHU T, HAO J S. A two-stage deep neural network for macro- and micro-expression spotting from long-term videos [C]//Proceedings of the 14th International Symposium on Computational Intelligence and Design. Washington D. C. , USA ; IEEE Press, 2021; 282-286.
- [51] TAKALKAR M A. LGAttNet: automatic micro-expression detection using dual-stream local and global attentions [J]. Knowledge-Based Systems, 2021, 212; 106566.
- [52] GUPTA P. PERSIST: improving micro-expression spotting using better feature encodings and multi-scale Gaussian TCN [J]. Applied Intelligence, 2022, 10; 1-15.
- [53] PFISTER T, LI X B, ZHAO G Y, et al. Recognising spontaneous facial micro-expressions [C]//Proceedings of International Conference on Computer Vision. Washington D. C. , USA ; IEEE Press, 2011; 1449-1456.
- [54] WANG Y D, SEE J, PHAN R C W, et al. LBP with six intersection points: reducing redundant information in LBP-TOP for micro-expression recognition [C]//Proceedings of Asian Conference on Computer Vision. Berlin, Germany ; Springer, 2015; 525-537.
- [55] HUANG X H, WANG S J, LIU X, et al. Discriminative spatiotemporal local binary pattern with revisited integral projection for spontaneous facial micro-expression recognition [J]. IEEE Transactions on Affective Computing, 2019, 10(1) ; 32-47.
- [56] WEI J S, LU G M, YAN J J, et al. Micro-expression

- recognition using local binary pattern from five intersecting planes [J]. *Multimedia Tools and Applications*, 2022, 81(15):20643-20668.
- [57] WEI J, LU G, YAN J. A comparative study on movement feature in different directions for micro-expression recognition[J]. *Neurocomputing*, 2021, 449:159-171.
- [58] LIU Y J, LI B J, LAI Y K. Sparse MDMO: learning a discriminative feature for micro-expression recognition[J]. *IEEE Transactions on Affective Computing*, 2021, 12(1):254-261.
- [59] LIONG S T. Less is more: micro-expression recognition from video using apex frame[J]. *Signal Processing: Image Communication*, 2018, 62:82-92.
- [60] HAPPY S L, ROUSTRAY A. Fuzzy histogram of optical flow orientations for micro-expression recognition [J]. *IEEE Transactions on Affective Computing*, 2019, 10(3):394-406.
- [61] WANG L. A weighted feature extraction method based on temporal accumulation of optical flow for micro-expression recognition[J]. *Signal Processing: Image Communication*, 2019, 78:246-253.
- [62] LEI L, LI J F, CHEN T, et al. A novel Graph-TCN with a graph structured representation for micro-expression recognition[C]//*Proceedings of the 28th ACM International Conference on Multimedia*. New York, USA: ACM Press, 2020:2237-2245.
- [63] LEI L, CHEN T, LI S G, et al. Micro-expression recognition based on facial graph representation learning and facial action unit fusion[C]//*Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Washington D. C. , USA: IEEE Press, 2021:1571-1580.
- [64] ZHAO X, MA H, WANG R. STA-GCN: spatio-temporal AU graph convolution network for facial micro-expression recognition[C]//*Proceedings of Chinese Conference on Pattern Recognition and Computer Vision*. Berlin, Germany: Springer, 2021:80-91.
- [65] SUN X, LANG J J, ZHENG S X, et al. Dual expression fusion: a universal microexpression recognition framework [EB/OL]. [2022-08-05]. <https://www.computer.org/csdl/magazine/mu/5555/01/09817631/1EOA8CImZYQ>.
- [66] WANG Y, ZHENG S X, SUN X, et al. Micro-expression recognition with attention mechanism and region enhancement [EB/OL]. [2022-08-05]. <https://link.springer.com/article/10.1007/s00530-022-00934-6>.
- [67] CEN S, YU Y, YAN G, et al. Multi-task facial activity patterns learning for micro-expression recognition using joint temporal local cube binary pattern [J]. *Signal Processing: Image Communication*, 2022, 103:116616.
- [68] ZHOU L, MAO Q R, XUE L Y. Cross-database micro-expression recognition: a style aggregated and attention transfer approach[C]//*Proceedings of IEEE International Conference on Multimedia & Expo Workshops*. Washington D. C. , USA: IEEE Press, 2019:102-107.
- [69] LI Y T, HUANG X H, ZHAO G Y. Joint local and global information learning with single apex frame detection for micro-expression recognition [J]. *IEEE Transactions on Image Processing*, 2021, 30:249-263.
- [70] LIONG S T, GAN Y S, SEE J, et al. Shallow Triple Stream Three-dimensional CNN (STSTNet) for micro-expression recognition[C]//*Proceedings of the 14th IEEE International Conference on Automatic Face & Gesture Recognition*. Washington D. C. , USA: IEEE Press, 2019:1-5.
- [71] ZHAO Y, XU J. A convolutional neural network for compound micro-expression recognition[J]. *Sensors (Basel)*, 2019, 19(24):E5553.
- [72] 刘德志, 梁正友, 孙宇. 结合空间注意力机制与光流特征的微表情识别方法[J]. *计算机辅助设计与图形学学报*, 2021, 33(10):1541-1552.
- LIU D Z, LIANG Z Y, SUN Y. Micro-expression recognition method based on spatial attention mechanism and optical flow features[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2021, 33(10):1541-1552. (in Chinese)
- [73] OH T H, JAROENSRI R, KIM C, et al. Learning-based video motion magnification[C]//*Proceedings of European Conference on Computer Vision*. Berlin, Germany: Springer, 2018:633-648.
- [74] GAN Y S. OFF-ApexNet on micro-expression recognition system[J]. *Signal Processing: Image Communication*, 2019, 74:129-139.
- [75] XIA Z Q, PENG W, KHOR H Q, et al. Revealing the invisible with model and data shrinking for composite-database micro-expression recognition [J]. *IEEE Transactions on Image Processing*, 2020, 29:8590-8605.
- [76] LI X Y, WEI G S, WANG J, et al. Multi-scale joint feature network for micro-expression recognition [J]. *Computational Visual Media*, 2021, 7(3):407-417.
- [77] CHEN B, LIU K H, XU Y, et al. Block division convolutional network with implicit deep features augmentation for micro-expression recognition [EB/OL]. [2022-08-05]. <https://ieeexplore.ieee.org/document/9676412>.
- [78] 梁正友, 刘德志, 孙宇. 结合迁移学习与可分离三维卷积的微表情识别方法[J]. *计算机工程*, 2022, 48(1):228-235.
- LIANG Z Y, LIU D Z, SUN Y. Micro-expression recognition method combining transfer learning and separable 3D convolution [J]. *Computer Engineering*, 2022, 48(1):228-235. (in Chinese)
- [79] BEN X Y. Learning effective binary descriptors for micro-expression recognition transferred by macro-information [J]. *Pattern Recognition Letters*, 2018, 107:50-58.
- [80] XIA B, WANG W K, WANG S F, et al. Learning from macro-expression: a micro-expression recognition framework[C]//*Proceedings of the 28th ACM International Conference on Multimedia*. New York, USA: ACM Press, 2020:2936-2944.
- [81] WANG C Y. Micro-attention for micro-expression recognition [J]. *Neurocomputing*, 2020, 410:354-362.
- [82] WEBER R, LI J, SOLADI C, et al. A survey on databases of facial macro-expression and micro-expression [C]//*Proceedings of International Joint Conference on Computer Vision, Imaging and Computer Graphics*. Berlin, Germany: Springer, 2018:298-325.
- [83] TANG H, CHAI L, LU W. Transferring dual stochastic graph convolutional network for facial micro-expression recognition [EB/OL]. [2022-08-05]. <https://doi.org/10.48550/arXiv.2203.05208>.
- [84] ZHANG J, YAN B Y, DU X H, et al. Motion magnification multi-feature relation network for facial microexpression recognition [J]. *Complex & Intelligent Systems*, 2022, 8(4):3363-3376.