基于多源域对抗迁移学习的可穿戴情绪识别技术

邹永攀[1)] 王丹阳[1)] 王 丹[1)] 郑灿林[1)] 宋奇峰[1)] 朱毓正[1)]
范长河[2)] 伍楷舜[3)]

[1)](深圳大学计算机与软件学院物联网研究中心 广东 深圳 518060)
[2)](广东省第二人民医院心理精神科 广州 510317)
[3)](香港科技大学(广州)信息枢纽 广州 511453)

摘 要 情绪影响身心健康及认知功能等,因而在人们的生活中扮演着重要角色.自动情绪识别有助于预警心理疾病和探索行为机制,具有巨大的研究与应用价值.在过去十余年中,研究者们提出了各种情绪识别方法,但均存在不同方面的不足:基于脑电图(Electroencephalography,EEG)信号的方法需采用专业、昂贵且不易操作的脑电仪;基于视觉和语音的方法存在隐私泄露的风险;基于手机使用模式分析的方法其可靠性和准确性有待提高等.本文利用生理信号如呼吸音、心跳音及脉搏等与情绪的潜在关联性,创新性地提出基于低成本、普适易用可穿戴硬件的情绪识别技术,借助多模态数据融合对不同类型数据进行有效利用,既减少了数据冗余又有效提升了系统性能.此外,在保证良好识别准确率的前提下,为提升情绪识别模型对不同用户的泛化性、最大化降低新用户的使用成本,本文提出了基于多源域对抗思想的情绪识别模型,借助少量来自新用户的无标签数据实现模型的无监督迁移,再辅之以极少量有标签数据微调分类器参数可进一步提升情绪识别准确率.为验证所提情绪识别方法的有效性,本文设计并实现了一套融合麦克风与光电容积脉搏波(Photoplethysmography, PPG)传感器以测量人体心跳音、呼吸音及脉搏等生理指征的可穿戴系统.基于此系统,本文在不同设置下开展了大量实验并对不同影响因素进行了评估.实验结果表明:对于四类基本情绪,本文所提方法单被试识别准确率可达95.0%,跨被试识别准确率为62.5%,比基准方法提升了5.3%.结合有监督小样本参数微调,识别准确率可进一步提高至81.1%,比基准方法提高了12.4%.上述结果验证了本文所提方法的可行性,为泛在情绪识别研究做出了崭新的探索.

关键词 可穿戴设备;情绪识别;多模态数据;迁移学习;域迁移;生成对抗学习
中图法分类号 TP18 **DOI号** 10.11897/SP.J.1016.2024.00266

Research on Wearable Emotion Recognition Based on Multi-Source Domain Adversarial Transfer Learning

ZOU Yong-Pan[1)] WANG Dan-Yang[1)] WANG Dan[1)] ZHENG Can-Lin[1)] SONG Qi-Feng[1)]
ZHU Yu-Zheng[1)] FAN Chang-He[2)] WU Kai-Shun[3)]

[1)](*The IoT Research Center, College of Computer Science and Software Engineering, ShenzhenUniversity, Shenzhen, Guangdong* 518060)
[2)](*Department of Psychiatry, Guangdong Second Provincial General Hospital, Guangzhou* 510317)
[3)](*Information Hub, The Hong Kong University of Science and Technology (guangzhou), Guangzhou* 511453)

Abstract Emotions can profoundly impact both human's overall well-being and cognitive

收稿日期:2022-11-08;在线发布日期:2023-09-13. 本课题得到国家自然科学基金面上项目(62172286)、国家自然科学基金联合重点项目(U2001207)、广东省自然科学基金面上项目(2022A1515011509)、广州市科技计划项目(202102010115)、广东省颐养健康慈善基金会项目(JZ2022001-3)以及腾讯"犀牛鸟"-深圳大学青年教师科研基金项目资助.**邹永攀**,博士,副教授,中国计算机学会(CCF)会员,主要研究领域为移动计算、普适计算和人机交互. E-mail: yongpan@szu.edu.cn. **王丹阳**,硕士,主要研究领域为情感计算. **王 丹**,女,1996年生,硕士,主要研究领域为情感计算. **郑灿林**,硕士,主要研究领域为智能感知与人机交互. **宋奇峰**,男,1999年生,硕士研究生,主要研究领域为情感计算. **朱毓正**,硕士研究生,主要研究领域为智能感知与人机交互. **范长河**,博士,主任医师,主要研究领域为生物精神病学和行为医学. **伍楷舜**(通信作者),博士,教授,中国计算机学会(CCF)会员,IEEE Fellow,主要研究领域为移动计算、人工智能、物联网. E-mail: wu@szu.edu.cn.

function. As a result, they are of paramount significance in the realm of human life especially in modern society with increasing pressures. Automatic emotion recognition contributes to early warning of psychological disorders and the exploration of behavioral mechanisms, holding immense research and practical value. Over the past decade, researchers have proposed various kinds of methods for automatic emotion recognition based on different sensing mechanisms. Nevertheless, each of them exhibits deficiencies in different aspects. For example, the methods based on electroencephalogram (EEG) signals require the use of specialized, costly, and challenging-to-operate EEG devices; the methods relying on visual and speech cues carry privacy risks; and the methods based on the analysis of mobile phone usage pattern need improvement in terms of reliability and accuracy. Considering the above, this paper proposes a novel approach to automatic emotion recognition that utilizes low-cost, readily available, and easy-to-use wearable hardware. In a detail, this paper makes use of the potential correlations between physiological signals, namely, breathing and heartbeat sounds, and pulse with human emotions. By employing data fusion across multiple sensing modalities, this work effectively harnesses diverse information types, reducing data redundancy, and substantially improving the system performance at the same time. Furthermore, while ensuring a high recognition accuracy, this paper also proposes an emotion recognition model based on a multi-source domain adversarial approach which aims to enhance the generalization of emotion recognition across diverse users and minimize the cost for unseen users. Our method first leverages a small amount of unlabeled data from unseen users to achieve quick model adaptation in an unsupervised approach, and then fine-tune the classifier's parameters with a minimal amount of labeled data to further improve emotion recognition accuracy. To validate the effectiveness of our proposed emotion recognition method, this paper designs and implements a wearable system that integrate two microphones and photoplethysmography (PPG) sensors to measure physiological signs. Among them, the two microphones are equipped in a smartglasses and earphone to collect sounds produced by heartbeats and breathing, respectively; the two PPG sensors are embedded in the smartglasses and a smartwatch to measure the blood pulses in the head and wrist, respectively. Based on this wearable system, we have conducted extensive experiments in diverse settings with thirty participants aged from 17 to 30 years old. We have also carried an assessment of the impact of different environmental factors such as noise, hardware, and wearing positions to evaluate the robustness of our emotion recognition system. The experimental results demonstrate that for the four basic emotions, the proposed method achieves an average recognition accuracy of 95.0%in the subject-dependent cases, and an average accuracy of 62.5% in the cross-subject cases after using multi-source domain adversarial transfer learning, with a 5.3% improvement over the baseline methods. When combined with supervised fine-tuning with few shots, the recognition accuracy further increases to 81.1%, surpassing the baseline methods by 12.4%. These findings affirm the feasibility of the proposed method and offer a fresh perspective for ubiquitous emotion recognition research.

Keywords wearable devices; emotion recognition; multimodal data; transfer learning; domain adaptation; generative-adversarial learning

1 引 言

如今，快节奏的生活使得人们的心理压力日渐增大. 越来越多的人出现情绪不稳定、长期低落等不良症状，甚至患上焦虑症、抑郁症等心理疾病. 2019～2020年度心理健康报告显示，我国有高达31.1%的大学生有抑郁或焦虑倾向[1]. 自动情绪

识别技术能有效地帮助人们自我调节，辅助研究者探究心理疾病的内在机理与开展相应的治疗．目前学界对于情绪的分类主要存在两派观点：一种认为情绪是离散的，可被分为若干类别．比如美国心理学家R. Plutchik就认为人的基本情绪共有八种（生气、厌恶、恐惧、悲伤、期盼、快乐、惊讶和信任），其他情绪则是这些基本情绪复合叠加所派生出来的[2]．第二种观点则认为情绪是连续的，可以通过多维度进行分析．比如心理学家James A. Russell于1980年提出的二维空间情绪模型VA（Valence-Arousal）就采用愉悦度（Valence）和激活度（Arousal）分别描述情绪的极性和度量情绪的强度[3]．

近年来，自动情绪识别的相关研究逐渐涌现．一些工作利用摄像头捕捉面部表情[4-6]，或躯干的姿态、步态等体势语言[7-9]进行情绪识别．然而这类方法存在隐私泄露、设备位置受限、易受外界环境干扰等缺陷．还有一些工作利用语音进行情绪识别，通过提取语音中的语义、语音或韵律信息进行情绪推断[10-12]．但该方法存在隐私泄露、抗噪性差、不适合连续监测等缺点．此外，基于各种生理信号如脑电图（Electroencephalography，EEG）[13-20]、肌电图（Electromyography，EMG）[21]、心电图（Electrocardiogram，ECG）[22]、皮肤电反应（Galvanic Skin Response，GSR）[23]等进行情绪识别的研究工作也不断出现．此类技术准确率较高，但所需设备价格昂贵、操作复杂、无法做到对情绪连续捕获，对专业技能的依赖性较高．还有一些研究者使用上述不同模态的数据相互结合进行情绪推断[24-30]，相应方法在具有各种单模态方法优点的同时也兼具其不足之处．

智能可穿戴设备配备了类型丰富的传感器，能够检测人的行为活动、监测生理指标、捕捉图像及音视频等．借助此类设备，多模态感知数据可以以非侵入的方式被便捷地连续采集．同时，上述信息与个体的心理状态也存在着内在关联性，如若能借此实现对用户情绪的感知，将极大地提升自动情绪识别的普适性和易用性．其中，“普适性”与“易用性”指的是：相较于基于专业仪器如脑电仪、心电仪等的识别技术，本文所提出的方法仅仅依赖于低成本的可穿戴设备，降低了对设备和专业技能的要求．此外，可穿戴设备所感知的数据可解释性差、隐私性好、有利于避免隐私泄露．基于以上思考，本文提出利用低成本可穿戴传感器获取用户的多模态数据，并进而对其情绪状态进行分析与识别．

然而，要实现以上想法需克服如下的挑战：一方面，相较于脑电、心电、眼动等信号，可穿戴设备的感知信号精度和信噪比均更低，其与用户情绪的关联性也更加隐蔽复杂、难以挖掘．因此，如何充分利用不同模态的感知数据使其相互补充，提取能对情绪进行有效表征的特征，是本文需要解决的一个挑战．另一方面，相较于手势、行为活动等表观情境，用户的心理、认知等深层情境存在着更为明显的个体差异性，且与诸多因素有着复杂的耦合关系．由此导致在跨被试场景中，直接使用基于其他被试者（也称为源域用户）的数据所训练好的模型对新用户（也称为目标域用户）进行测试时，其识别准确率往往大幅下降．已有情绪识别的研究工作或对此问题缺乏考虑，没有做出相应的算法设计；或针对脑电、心电等信号设计方法，并不适用于可穿戴感知数据．因此，如何在尽可能降低新用户样本采集负担的前提下，提升情绪识别模型在跨被试场景下的准确率是本文需要解决的另一挑战．

针对以上挑战，本文提出对不同模态数据在数据级别融合以获得对情绪的高效表征的策略．同时，本文还提出基于多源域对抗迁移学习的思想，通过向训练模型中输入目标域的无标签数据，来减小经过特征提取器后源域和目标域数据分布之间的差异，使得在源域训练出的分类器也适用于目标域．进一步地，仅需要引入少量有标签目标域数据对模型进行个性化微调，即可实现目标域新用户识别准确率的大幅提升．本文设计并实现了一套融合麦克风和PPG传感器，能获取用户心跳、呼吸、脉搏等信息，并进而实现情绪识别的可穿戴系统．实验结果表明：本文所提出的识别模型对于四种基本情绪的识别准确率在单被试场景下可达95.0%，跨被试场景下可达62.5%，加入来自目标用户的少量带标签样本对模型进行微调后，跨被试准确率可提升至81.1%．

总结而言，本文主要的创新和贡献包括：

（1）基于生理信号如呼吸音、心跳音和脉搏等与情绪的潜在关联性，本文创新性地提出了基于低成本、普适易用的可穿戴设备所获取的多模态感知数据进行情绪识别的方法，提升了自动情绪识别技术的普适性．

（2）本文设计了一种基于多源域对抗学习的情绪识别模型，包含基于无监督和基于半监督迁移学习的情绪识别方法，在保证良好识别准确率的前提下，最大化地减少了识别模型对被测主体的依赖性，

提升了系统的易用性.

(3)本文实现了一套融合多种传感器的可穿戴情感识别系统,并在不同场景下开展了大量实验以验证其对于四种基本情绪的识别性能.结果表明:本文所提出的情绪识别系统无论是在单被试还是跨被试场景下均能取得良好的识别准确率.

2　相关工作

按照感知媒介划分,已有的情绪识别工作主要可分为基于肢体语言、语音信号、生理信号以及多模态感知数据等.

肢体语言包括面部表达、姿势体态、步态等,是一个人情绪和内心状态的重要反映.基于面部表达的技术是通过提取人在不同情绪状态下的面部特征进行情绪识别.印度理工学院的Mehendale教授等人[4]提出FERC框架使用全面部图像进行情绪识别.佐治亚理工学院的Hickson等人[5]使用摄像机对眼部追踪进行情绪识别.南京大学许封元教授等人[6]利用头戴式摄像头拍摄单眼图像从而进行情绪识别.该技术能以12.8帧的速度连续识别7种情绪并取得72.2%的识别准确率.基于人体姿态、步态的技术则主要借助摄像头等设备进行人体检测及骨架追踪,提取不同情绪状态下肢体静态或者动态的典型特征进行情绪识别.印度理工学院Saha教授等人[7]使用Kinect追踪肢体手势进行情绪识别.香港科技大学的Chiu等人[31]根据捕获的视频帧进行姿势估计和提取运动特征,利用监督学习模型对五种情绪进行分类.

基于语音信号的情绪识别技术可以细分为基于语音声学特征和基于语音语义内容两类.基于语音声学特征的技术主要通过提取说话者的语音信号中一些常见声学特征如频谱、倒频谱、梅尔倒谱系数(Mel-Frequency Cepstral Coefficients, MFCC)、能量、过零率等进行情绪识别.弗吉尼亚大学的Salekin等人[32]对说话者的声音进行分析以检测社交焦虑和抑郁症.基于语音内容进行情感分析是自然语言处理中的热点问题,主要通过将语音转为文字并分析情绪特征词的极性来进行情绪推断.还有一些研究同时将语音声学特征与语义信息结合以进行情绪推断,如滴滴AI研究院的Xu等人[12]联合采用MFCC特征和语义信息进行情绪识别.

由于人类的情绪变化通常会引起生理状况的改变,相比于肢体语言和语音信号,生理信号更能反映真实的情绪.兰州大学的胡斌教授等人[14]提出了一种使用GBDT分类器的新融合方法,利用额叶脑电图的双通道信号进行情绪分类,在DEAP数据库[33]中达到约75%的分类正确率.东南大学的郑文明教授等人[15]提出可转移注意力神经网络,通过局部和全局注意力机制自适应突出可转移EEG样本,在SEED数据集[34]上跨被试准确率可达84.4%.新加坡南洋理工大学李默教授等人[13]定制了一款装配有脑电传感器的智能眼镜来进行日常情绪日志的收集.华南理工大学舒琳教授等人[22]提出了一种基于ECG的新型实时检测方法,通过提取合适的特征进行融合以训练SVM分类器,取得了79.5%的分类准确率.美国麻省理工学院的Katabi教授等人[35]开创性地使用无线射频信号以提取人的呼吸及心跳信息以进行情绪推断.

基于多模态感知信号的情绪识别技术是通过融合上述两种或多种模态的信号进行协同合作以提升分类性能.清华大学的王雪教授等人[25]提出基于脑电、脉搏以及血氧信号并利用栈式自编码器进行情绪识别.上海交通大学的吕宝粮教授等人[24]提出了异质迁移学习以提升跨被试情绪识别准确率.新被试者只需提供眼动追踪数据即可与其他被试者的脑电信号进行信息迁移,跨被试准确率可达69.7%.美国斯坦福大学的Li教授等人[27]提出结合3D面部表情和语音进行抑郁症严重程度测量的算法.韩国世宗大学的Shin等人[29]开发了一个复杂的生物信号情感识别系统,通过将自主神经系统的ECG与EEG的相对功率值(θ、α、β和γ)按比率混合,实现对六种情绪的识别.还有很多研究人员通过结合多种生理信号如ECG与GSR[28],EEG与EOG[30]来进行情绪推断.此外,研究者还探索了基于生理信号与表情信息(图像或视频)相融合的情绪识别方法.Huang等人[36]和Zhong等人[37]将EEG信号或生理信号作为面部表情的补充信息以实现多模态情绪识别.

随着迁移学习的发展,其在情绪识别中的应用也逐渐被重视.当源域与目标域数据分布相似但不同时,可借助迁移学习提升模型在目标域中的识别性能.Ganin等人[38]提出了域对抗神经网络(DANN).该方法受域适应理论启发,要实现有效的域转移,必须基于不能区分训练(源)域和测试(目标)域的特征进行预测.Jiang等人[39]基于DANN思想设计了一种基于无线信号的跨环境人类行为识别模型.与之不同的是,本文根据情绪识别应用中感知信号与任

务的独特属性，对DANN框架中的特征提取器、鉴别器以及损失函数均做了针对性的设计．具体地，相比于无线信号与行为活动之间存在的显性关联性（如信号相位、幅度或者频率改变），可穿戴感知数据与人类情绪之间的关联更为隐蔽．因此，不同于该文使用三层堆叠CNN作为特征提取器，本文选择层数更深、表征能力更强的ResNet18以深入挖掘感知信号与情绪状态之间的潜在关系．其次，由于上述特点，该文所用对抗网络的训练方式在本任务中很难收敛且效果不佳．为解决此问题，本文在特征提取器与用户鉴别器之间增加了梯度反转层，从而使二者更好地对抗学习．最后，本文还设计了与情绪识别任务更匹配的损失函数，使得模型在训练中能将学习重心转移到情绪分类任务上．尽管上述模型设计思想已在文献中有所体现，本文创新性地将其运用于情绪识别任务中并充分考虑感知数据与任务的独特属性而对模型做出针对性设计，使模型取得了良好的识别性能．Yang等人[40]提出双半球域对抗神经网络，借助生成对抗训练得到与域无关的特征．Luo等人[41]构建了Wasserstein生成对抗网络域自适应框架，在预训练中将源域和目标域映射到公共空间，在对抗训练中缩小源域和目标域在公共特征空间上的差距．Li等人[42]提出一种领域自适应方法，通过最小化源域分类误差，同时使源域和目标域潜在表示相近来优化模型．该方法在SEED数据中的跨对象和跨场景准确率分别为86.7%和91.2%．Wang等人[43]提出了少标签对抗域自适应的新方法来进行跨对象EEG情感分类．Ding等人[19]基于DANN并融入任务特定领域思想，可同时适应域之间的条件分布和类之间的分类边界，从而增强了对新阶段情绪的预测效果．

总结而言，尽管情绪识别技术正蓬勃发展，但是仍存在以下不足：基于摄像机、语音等媒介的情绪识别技术依赖于包含丰富私密信息的感知数据，容易造成隐私泄露．EEG、EMG、ECG等专业生理数据监测仪价格昂贵且笨重、便携性差、使用场景及用户不具普适性．同时，基于图像、语音、文本等的情绪感知方式，受个人主观因素及个体差异性影响较大，且易产生情绪隐藏现象．针对当前情绪识别研究的不足，本文开展了使用低成本普适可穿戴设备所感知的多源数据进行情绪识别的研究，实现基于多模态泛在感知数据的情绪识别系统．进一步，通过信号处理与多模态融合技术对不同类型数据进行有效融合，设计了一种基于多源域对抗迁移学习的情绪识别模型，在保证良好的识别性能的前提下最大化减少对被测主体的依赖性和新用户的学习成本．

3 系统设计

图1展示了本文所提出的基于普适智能可穿戴设备进行情绪识别的总体系统框架．其主要包含四个部分：数据采集、数据预处理、单被试情绪识别模型以及跨被试情绪识别模型．对于数据采集部分，则是通过四种常见的可穿戴智能设备来采集五通道多模态数据；数据预处理部分依次进行原始信号切分、滤波、归一化以及信号的时频特征提取等操作；对于单被试情绪识别模型部分，使用多模态融合策略对单个用户的生理信号数据进行训练与测试来完成情绪识别任务；对于跨被试情绪识别模型部分，主要通过域对抗迁移学习的思想在跨用户的情况下情绪识别模型也能表现出良好的识别效果．此外，该系统主要识别四种基本情绪，分别为：Happy（开心）、Neutral（中立）、Sad（悲伤）以及Mixed（负面情绪综合，包含恐惧、厌恶和愤怒）．由于这些负面情绪非常相似[44]，因此在本工作中将它们统一划分为Mixed类别．下面将对系统的四个部分进行详细阐述．

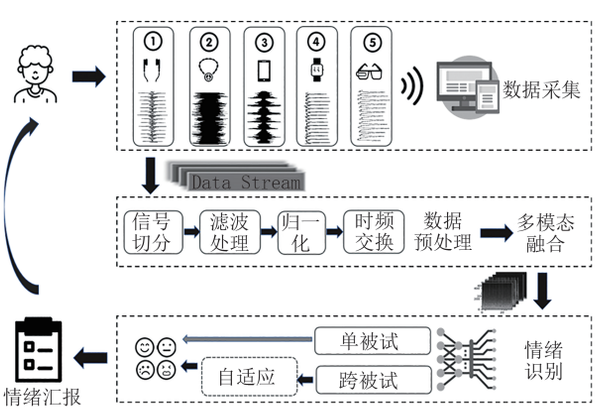


图1 基于多源可穿戴感知数据的情绪识别系统框架图

3.1 数据采集

本文所提出的情绪识别系统中采用四种常见可穿戴智能设备，分别为智能腕带、智能耳机、智能眼镜和智能项链．它们分别佩戴在用户的腕部、耳部、头部和颈部，获取用户脉搏PPG信号、耳内体声信号、耳部PPG信号及鼻部呼吸音、喉部气管音．四个智能终端设备的穿戴位置与相应数据类别如图2所示．

图中位置1处是一个入耳式智能耳机（简记为

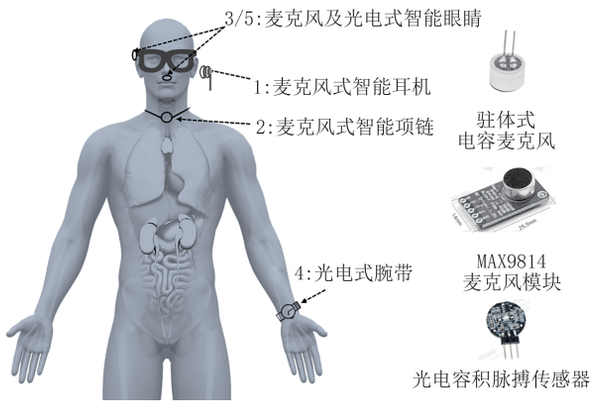


图2　可穿戴设备硬件系统及佩戴位置图

Mic-Ear)，内置一个带有信号放大芯片的微型麦克风用于采集耳道传来的心跳音信号. 该耳机与外耳道完全贴合形成一个共振空腔，能够放大心脏跳动产生的声信号. 佩戴在位置2和位置3的分别是内置有MAX9814电容麦克风的智能项链和智能眼镜，用于收集喉部气管音与鼻部呼吸音等体声信号，依次简记为Mic-Throat和Mic-Nose. 而佩戴在位置4和位置5的传感器是光电容积脉搏传感器，可以监测到佩戴者的心跳，分别简记为PPG-Hand和PPG-Head. 整个系统使用STM32单片机作为处理单元，负责控制传感器以1000 Hz的采样率采集数据，并通过无线蓝牙模块发送至手机端进行处理，由手机端进一步自动实时地进行情绪推测. 总结而言，本文所利用的生理信号可以分为以下三类：

(1)心跳音信号

心脏跳动引起压力变化，带动耳内结构如鼓膜及耳道内的空气进行振动，由此产生的声信号在密闭共振腔中放大并最终被耳机捕捉到. 而人体情绪状态与心跳活动紧密相关，并最终影响心跳音信号.

(2)呼吸音信号

人体呼吸时，气流经鼻腔进入喉管，由于摩擦在相应部位产生气流声信号. 而不同情绪状态下人体的呼吸气流速度、频率存在差别，从而导致产生的声信号有差异.

(3)脉搏光电信号

人体血液含氧量随着心脏跳动而周期性变化，进而改变血液对不同波长可见光的吸收率，使得光强接收器的信号强度随之变化. 脉搏光电信号反映了血氧随心跳而变化的情况.

需要说明是，本文之所以选择上述信号类别进行情绪识别是基于两方面的考虑. 其一，目前已有很多研究者从生理学角度证实了呼吸、心跳及其他体声信号与情绪的关联性. 如日本生理学家Ikuo Homma[45]从大脑神经学的视角对呼吸与情绪的关系进行了探讨，指出自主呼吸不仅由代谢需求控制且不断对情绪变化做出反应. 瑞典心理学家Kenneth Hugdahl[46]探讨了情绪与心率、血压的关系. 瑞士心理学家Sylvia D. Kreibig[47]从自主神经系统活动的角度探究情绪反应中心血管特征的变化，证明了情绪对心血管的影响. 此外，美国华盛顿大学组织行为学教授Hillary Anger Elfenbein[48]在其相关研究中也表明非语言发声是一种丰富而微妙的情感信号来源. 其二，商用智能设备能够便捷地获取上述信号，从而保证了情绪识别系统的普适性和易用性.

3.2　数据预处理

3.2.1　生理信号预处理

传感器采集得到的五通道数据源源不断地由串口或蓝牙发送至PC或手机进行预处理. 首先，对连续的信号流以长度为15 s的窗口进行切分，所得片段作为情绪识别的基本单位. 这是因为根据德国奥格斯堡大学Jonghwa Kim教授等人的研究[49-50]，情绪展示的窗口长度应根据不同感知方式而改变，2 s～6 s用于讲话，3 s～15 s用于生物信号. 其后，考虑到传感器器件的直流漂移，因此将各通道信号减去其各自均值以消除直流干扰. 此外，由于受附近电力线的干扰，采集到的信号含有50 Hz分量和相应的奇次谐波. 为减少这些工频干扰，首先对得到的各通道信号分别进行快速傅里叶变换，将不同频率干扰分量所对应的系数设为零，然后通过快速傅里叶逆变换将信号变换回时域序列.

根据传感器类型与采集位置的不同，本文采用不同的降噪方法处理不同模态的感知数据. 根据频率分析可以发现：声学耳机得到的信号(记为Mic-Ear)频率范围在0.8 Hz到300 Hz之间；智能项链与智能眼镜所得到的声学信号(记为Mic-Throat和Mic-Nose)的频率分布在0.8 Hz到500 Hz之间；而两种PPG信号(记为PPG-Hand和PPG-Head)的频率主要分布在12 Hz以内. 为了消除因传感器与皮肤之间的接触状态所带来的数据漂移，首先对信号中小于0.5 Hz的频段进行截断，仅保留0.5 Hz至12 Hz的频带. 基于以上分析，本工作选择巴特沃斯带通滤波器来滤除带外噪音，并根据不同模态信号的频率分布设置其带通频率. 滤波器采样率为1000 Hz，滤波器的脉冲响应类型使用的是IIR.

紧接着，为消除信号强度大小对于模型训练的影响，本文采用如下式所示的软归一化处理：

$$x_{norm}=2\frac{x-Q_{0.05}(x)}{Q_{0.95}(x)-Q_{0.05}(x)}-1 \tag{1}$$

其中x_{norm}表示归一化后的值，$Q_{0.05}(x)$和$Q_{0.95}(x)$分别表示信号值第5和第95分位数．之后进行异常值截断，最后使用短时傅里叶变换(Short Time Fourier Transform，STFT)提取每个通道的时频图．

3.2.2 情绪标签预处理

情绪标签的准确性对于情绪识别系统而言至关重要，也是相关研究中的难点问题．以往的研究工作往往默认被试者自报告类别标签是准确的并将其作为数据的真实标签[51]，或者直接将刺激材料的情绪标签设定为实验数据的标签[17-20,52]．但这两种标签处理方式均存在缺陷．其一，被试者自报告情绪标签往往并不准确．其原因是多方面的，主要包括：被试者在观看刺激材料时可能分心而报告了错误的标签；被试者对于情绪类别的划分标准不清楚；被试者事后报告标签时存在记忆偏差等．其二，由于语言、文化及个体差异，被试者观看刺激材料时所产生的情绪与该材料的情绪标签并不一致．因此，尽管直接使用刺激材料的标签作为数据真实标签可在一定程度上避免被试者自报告标签时的主观偏差，但也会导致数据真实标签存在较大误差．

针对此问题，本文设计了一种基于自报告标签和刺激材料标签的数据真实标签生成方法．首先，本工作根据罗伯特•普拉切克[2]所提出的离散情绪模型，结合情绪刺激材料数据集FilmStim[53]所提供的6种基本情绪，将中立(Neutral)、快乐(Happiness)、悲伤(Sadness)、恐惧(Fear)、愤怒(Anger)、厌恶(Disgust)作为目标情绪进行情绪诱发．考虑到恐惧、愤怒以及厌恶这三种基本情绪非常相似，均为消极价效且具有较高唤醒度的情绪．从情绪维度模型的量化角度出发，本工作对这三种消极情绪进行合并，归类为同一种情绪类别-混合情绪(Mixed)．此种简并处理有助于保证标签的可靠性．其次，本文设计了一种基于投票机制的数据真实标签的生成策略．其基本原理为：首先，对于每一个情绪刺激材料，选择超过一半的被试者所报告的情绪类别作为其所诱发的目标情绪．该操作有助于降低因语言、文化等差异造成的刺激材料情绪标签与被试者产生的真实情绪之间的偏差，提升二者的一致性．随后，仅挑选被试者自报告标签类别与目标情绪一致的信号作为有效数据，以此减少因自报告标签错误带来的误差．此外，除中立情绪外，对于其他情绪类别数据，本文仅选取愉悦度Pleasure大于6或小于4的实验样本作为有效数据以确保刺激材料成功地诱发了被试人员的情绪，避免噪声样本的干扰．

3.3 基于多模态融合的单被试情绪识别

如前所述，本工作所利用的感知数据与情绪之间关联性隐蔽，需要设计有效的融合策略以充分利用不同模态的数据．下面将对此做出详细的介绍．

3.3.1 数据级融合情绪识别

为利于系统在移动端部署，本文采用较轻量化的深度残差网络(Deep Residual Network，ResNet)[54]ResNet18作为骨干网络来进行情绪分类．本小节介绍基于ResNet18的数据级融合策略进行情绪识别的性能．对所收集的生理信号预处理之后，将得到的不同通道多模态时频谱数据直接进行集成，并送入ResNet18中进行情绪分类，使网络自动学习与利用不同模态的低层数据之间的相关性和相互作用．所输入的时频矩阵及网络结构如表1所示．同时，该模型使用如下所示的交叉熵函数作为情绪分类的损失函数：

$$Loss=-\sum_{i=0}^{C-1}\left(y_i\log\left(p_i\right)\right) \tag{2}$$

其中$p=[p_0,\cdots,p_{C-1}]$是一个概率分布；$y=[y_0,\cdots,y_{C-1}]$表示当样本属于第i类时$y_i=1$，否则$y_i=0$；C表示类别数．

表1 数据级融合时频矩阵及ResNet18网络结构

Layer_name	Output size	Operator
Input	257×126×5	Input STFT spectrums×5
conv1	129×63×64	Conv2d: 7×7，64，stride 2，padding 3
conv2_x	65×32×64	3×3 max pool，stride 2 $\begin{bmatrix}3\times3,64\\3\times3,64\end{bmatrix}\times2$
conv3_x	33×16×128	$\begin{bmatrix}3\times3,128\\3\times3,128\end{bmatrix}\times2$
conv4_x	17×8×256	$\begin{bmatrix}3\times3,256\\3\times3,256\end{bmatrix}\times2$
conv5_x	9×4×512	$\begin{bmatrix}3\times3,512\\3\times3,512\end{bmatrix}\times2$
average pool	1×1×512	AdaptiveAvgPool2d

3.3.2 特征级融合情绪识别

上述数据级融合策略对融合后的数据进行统一建模，忽略了不同模态数据的特性、情绪表征能力以及信号质量差异等，在融合过程中各通道的权重相

同. 因此，本小节引入SENet(Squeeze-and-Excitation Networks)[34]网络，通过学习自动获取各通道的重要程度，并将重要程度作用到每个通道上来提升网络的能力. SENet主要手段是Squeeze和Excitation，它们的功能如下：

(1)Squeeze：由全局平均池化来实现，以空间维度进行特征压缩，对每通道的二维特征提取一个实数，表征着该通道上的全局分布.

(2)Excitation：可以理解为一个两层的网络，通过两个全连接层去学习通道间的相关性，学习到的权重表征着每个通道的重要性，随后以乘法加权到之前的特征上.

基于SENet网络结构的思想，对采集的信号进行上述预处理，然后将得到的五通道时频谱按Mic-Ear、(Mic-Throat，Mic-Nose)、(PPG-Hand，PPG-Head)三种信号模态送入SENet+ResNet的结构中. 具体的网络结构如图3所示. 其中，将去掉全连接层(FC)的ResNet18作为基础特征提取器，三种信号模态分别独立进行特征提取，并通过Squeeze和Excitation结构自动获取每通道的重要程度，并将学习到的重要程度作用到原通道特征，以赋予不同特征通道不同的权重. 然后将从这三个通道特征提取器获得的特征连接起来，进行加权融合，形成一个单一的特征向量，并将其馈送到FC层，使用交叉熵作为情绪分类损失函数，最终实现多标签分类.

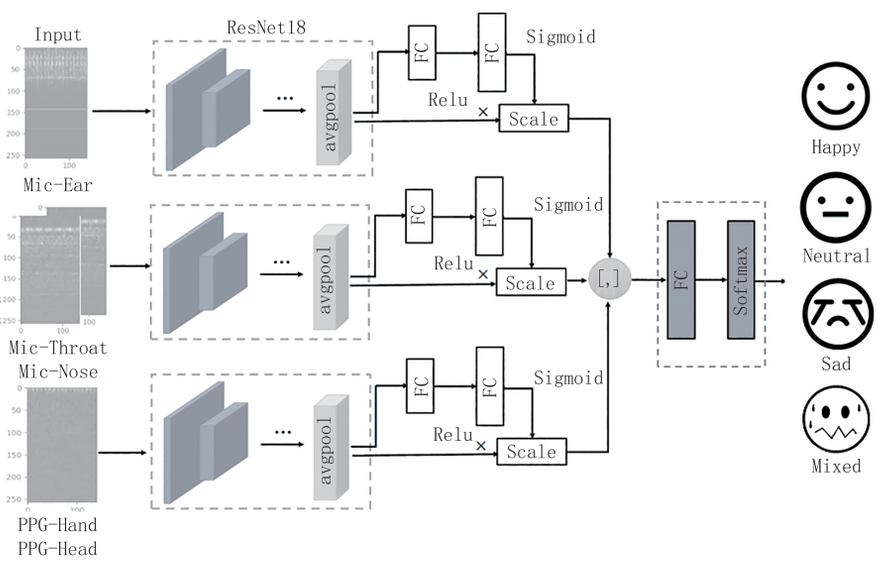


图3　特征级融合情绪识别框架

3.4　基于域对抗迁移学习的跨被试情绪识别

在深度学习领域，实现跨域任务的常见方法之一是使用目标域的带标签数据对已训练好的模型进行重训练. 但是，这种方案要求新域提供高质量的带标签数据. 这一要求对于本文所提出的情绪识别任务而言充满挑战，因为这需要用户有意识地如实表现出几种情绪. 针对此难点，本文根据无监督对抗领域自适应思想，提出了多源对抗迁移学习情绪识别模型(MultiSource-Dann)，即只使用目标用户的少量无标注数据，用于提取目标域与源域之间的共性特征，提升模型在目标域上的识别效果. 图4展示了该情绪识别网络的结构，主要包含特征提取器、情绪鉴别器和用户鉴别器三个部分. 各部分的具体功能如下：

(1)特征提取器：该模块用于提取原始数据的深层特征. 在训练阶段，提取的深层特征将同时传入情绪鉴别器和用户鉴别器. 在前者的辅助下，提取出更具情绪区分度的特征，从而更准确地识别出具体的情绪. 同时，在与后者的对抗中提取出用户不变性特征，以迷惑用户鉴别器，使其无法鉴别出样本来自源域还是目标域. 在测试阶段，提取的深层特征将只传入情绪鉴别器，以得到最终结果.

(2)情绪鉴别器：其最终的任务分类器，用于识别具体的情绪. 在训练过程中，该模块会辅助特征提取器提取出更具情绪区分度的特征.

(3)用户鉴别器：用户鉴别器只在训练阶段作用，其任务是识别样本的域标签. 与特征提取器形成对抗关系. 用户鉴别器需要提高自己的性能，从特征提取器提取出来的用户无关特征中识别出具体的用户. 而特征提取器则需要提升其性能，以提取与用户更无关的特征，以迷惑用户鉴别器.

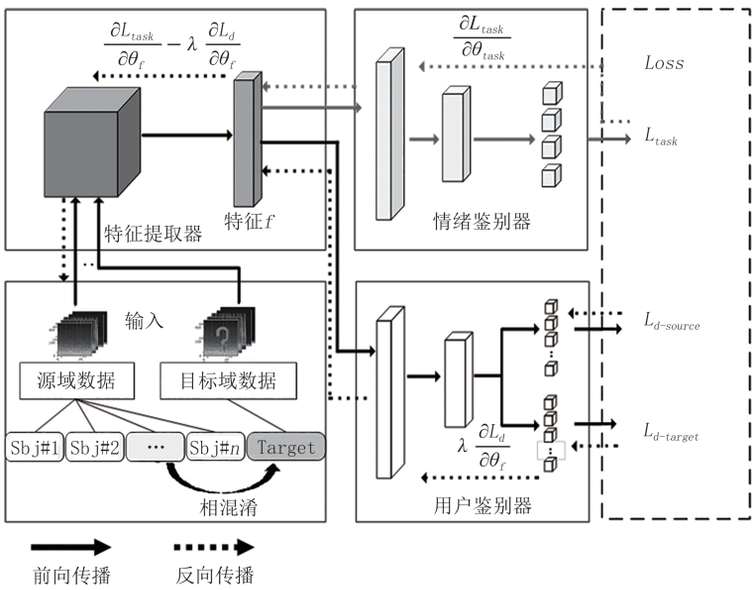


图4 基于多源域对抗迁移学习的情绪识别结构示意图

由于用户鉴别器与特征提取器是形成对抗关系，故而模型在特征提取器和用户鉴别器之间加入了一个梯度反转层$R_\lambda(x)$(Gradient Reversal Layer，GRL)，在反向传播中，梯度在经过梯度反转层后，会实现梯度取反，从而实现用户鉴别器和特征提取器的对抗学习. 其具体实现如式4所示：

$$R_\lambda(x)=x \tag{3}$$

$$\frac{\mathrm{d}R_\lambda(x)}{\mathrm{d}x}=-\lambda I \tag{4}$$

其中I为一个单位阵. λ为一个可调参数，取值区间为(0,1].

为实现在训练过程中模型逐步将学习重心移动到情绪分类任务上，故而此处我们对参数λ进行特殊设置，使得其在训练过程中逐渐增大，其变化公式如下，其中p表示当前迭代次数与总迭代次数的比率，γ为常数.

$$\lambda=\frac{2}{1+e^{-\gamma p}}-1 \tag{5}$$

整个网络的损失函数为：

$$Loss=L_{task}-\lambda\left(L_{d-source}+L_{d-target}\right) \tag{6}$$

其中L_{task}和$L_{d-source}$分别为非目标用户域情绪鉴别器的损失和用户鉴别器的损失，$L_{d-target}$为目标用户域用户鉴别器的损失.

使用交叉熵作为情绪鉴别器的损失函数，使用负对数似然损失(Negative Log Likelihood Loss，NLLLoss)作为用户鉴别器的损失函数.

本小节通过引入经典的无监督领域对抗自适应思想，提出了多源域对抗领域迁移学习模型，将单源域对抗模式改进为多源域对抗模式，同时设计了参数λ的自动调整方案，使得模型更加契合本文的任务. 网络训练时，网络的输入为带标签的源域数据集与不带标签的目标域数据集，以及源域与目标域数据集的域标签. 训练过程中，模型对来自非目标用户的有标签数据不断最小化情绪鉴别器的损失，而对来自非目标用户的有标签数据和目标用户域的无标签数据则不断最大化用户鉴别器的损失，从而使得情绪鉴别器与用户鉴别器在训练过程中相互对抗实现情绪分类与用户域分类损失之间的相互平衡. 通过这种域对抗方式，缩小源域数据和目标域数据经过特征提取器提取出的特征之间的差距，使得所提取的特征在具有情绪分类可识别性的同时具有用户域不变性，从而达到域迁移的目的.

3.5 监督式样本反馈参数调整

在上述模型的基础上，本文还进一步提出通过引入少量有标签的目标域数据来对模型参数进行微调，以实现情绪识别系统的个性化定制，进而提升跨被试测试的准确率，在尽可能减少新用户学习成本的同时保证较好的识别性能. 具体而言，首先在仅拥有源域数据及目标域无标签数据的情况下进行多源域对抗无监督迁移学习，得到情绪识别预训练模型，并保存特征提取器的网络参数. 其后，使用已保存的网络参数初始化网络，并冻结除分类器以外的所有参数，使用少量目标域带标签数据对分类器进行微调重训练. 其核心思想在于：通过引入少量目标域带标签数据对分类器参数进行调整，能够有效地引导模型快速收敛到目标域中，从而学习到目标域的数据分布.

4　系统实现与评估

4.1　实验设置与系统搭建

为有效获得用户情绪样本，本文采用DEAP数据库[55]、FilmStim数据库[53]、IADS数据库[56]、SEED数据库[57]以及EMDB数据库[58]作为情绪激发材料．为避免因语言障碍而削弱刺激材料对情绪的诱发力度，我们为数据库中的外语材料均加上字幕．此外，考虑到文化差异，本文还通过投稿及评分排序选取了一部分网络上的短视频片段，比如生离死别、娱乐搞笑片段等．我们从上述四个情绪诱发材料数据库中分别选取了4、54、30、15和50段材料．同时，从网络短视频中选取了20段材料作为情绪诱发的补充材料．因此，实验中用于诱发情绪的音视频材料总数为173段．同时，在选取上述材料时，我们保持四类情绪的诱发材料数量大体均等．

本文的实验均在常见办公环境（噪声大小为40 dB～55 dB）中开展．受试者共有30人，年龄分布在17至30岁之间．为了便于实验者观看刺激材料并及时报告情绪状态，我们还开发一款如图5所示的基于Windows系统的情绪数据采集软件．实验过程中，受试者一只耳朵佩戴本文所设计的可穿戴系统用以收集多模态信号，另一只耳朵则佩戴常规耳机收听用于情绪诱发的音视频材料．在此种设置中，音视频材料的声音并不会对感知信号造成干扰，保证了信号质量．在一次实验中随机筛选100个刺激材料，分节播放给被试人员，每小节播放一个刺激材料给被试人员．在播放每节的刺激材料之前，被试人员有1 min的缓冲时间去平复心情．心情调节时间可个性化延长至被试人员已恢复平静，进而循环，最后进行情绪自报告．自报告的内容包括使用自我情绪评定量表（Self-AssessmentManikin，SAM）[59]对愉悦度、激活度、支配度进行1～9的打分，以及对熟悉程度和喜欢程度（1～5）打分．此外，被试人员还被要求报告离散情绪．由于个体差异性，即不同被试人员对于同类刺激材料产生的情绪状态并不同，因此最终整个实验所收集的样本总量为13 918个，个体样本量均值为463.9，样本量的标准差为121.5.

4.2　评价指标

考虑到不同情绪类别的样本量不均衡，本文采用宏平均（Macro-average）方法计算宏精确率（Macro-Precision）、宏召回率（Macro-Recall）、宏F1值（Macro-F1 Score）、以及准确率（Accuracy）作为评价指标．

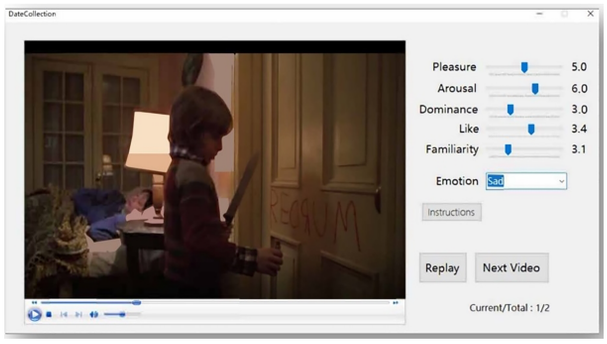


图5　基于Windows系统的情绪数据采集软件

4.3　对比模型

4.3.1　基于基础深度学习模型的情绪识别

在跨被试情绪识别中，最基本的做法就是保持其他数据处理方式一致，包括信号预处理、数据融合等，直接使用RestNet18进行留一用户交叉验证（leave-one-user-out）．由于此网络只需完成源域用户情绪分类这一单一任务，为叙述方便后续将此方法简记为Single-Task，并将其作为一种基准方法．

4.3.2　基于对抗式双任务学习的情绪识别

受多源域对抗迁移学习的启发，在不加入目标域无标签数据信息的情况下，仅使用多个源域有标签数据进行训练对抗学习．在构建网络中，用户鉴别器仅对来自多个非目标用户域的数据进行多用户类别判别，尽可能分出数据来自哪个用户，而不加入任何目标域的数据．此时，网络的损失函数为：

$$Loss = L_{task} - \lambda L_{d-source} \tag{7}$$

网络训练时对来自多个非目标用户域的有标签数据不断最小化情绪鉴别器的损失，并不断最大化其用户鉴别器的损失，通过这种域对抗方式缩小多个源域在特征提取器所提取出的特征之间的差距，增强特征提取器的泛化能力，从而达到域迁移的目的．此模型被称为对抗式双任务学习模型，简记为Dual-Task.

4.4　实验分析结果

4.4.1　多模态信号融合策略

对每个被试人员采用十折交叉验证进行单被试的情绪分类实验，即分别使用本人数据进行模型训练与测试．在进行交叉验证的过程中，采用分层抽样进行数据划分以保持数据分布的一致性．分别采用数据级融合和特征级融合策略进行情绪识别分类实验，得到了如表2所示的实验结果．

从平均精确率、召回率、F1值、准确率等性能指

表2 不同融合策略性能比较

参数 \ 模型	数据级融合	特征级融合
Precision (%)	94.58	94.99
Recall (%)	94.45	94.53
F1 score (%)	94.51	**94.75**
Accuracy(%)	94.99	**95.19**
训练时占用GPU(Mb)	**2685**	4469
训练时每次迭代耗时(s)	**4.13**	8.51
模型参数量(Mb)	**11.18**	33.62

标来看，特征级融合的分类性能略优于数据级融合（各项指标约高0.2%）. 由此可见，虽然数据级融合学习了模态之间底层的相关性和相互作用，但特征级的融合策略可对每个模态单独建模，并赋予不同模态的通道不同的权重，也学习了不同模态特征之间的相关性和相互作用. 另一方面，从模型训练所占用的GPU内存来看，特征级融合是数据级融合的1.7倍；从模型参数量来看，特征级融合是数据级融合的3倍；从模型推理时间来看，特征级融合是数据级融合的2.1倍. 因此，在分类性能相差较小的情况下，考虑到系统实际部署时的运行性能及移动端用户体验，本文最终推荐采用数据级融合策略.

4.4.2 不同模型的对比

本文还对比了不同模型在跨被试场景下的情绪识别性能. 除了4.3节中所述的Single-Task和Dual-Task两种基准模型外，本文还对比了其他五种新颖经典的域适应方法，包括MCD[60]、CLAN[61]、DFA[62]、CGDM[63]以及EI[39]. 为保证对比实验的公平性，我们采用了参考文献[60-63]中所提供的开源代码，只将模型的特征提取器替换为本文所使用的ResNet18并相应修改输入输出层，而对模型其他结构不做任何改动. 而由于EI[39]没有提供开源代码，本文只能根据论文叙述复现该模型并将两个限制条件根据本文任务做出修改. 从所得实验结果图6可以看到，在所对比的几种模型中，MultiSource-Dann表现最佳，其准确率和F1值分别为62.5%和60.4%，分别高于次优模型CALN的59.6%和57.3%. 而相较于基准模型Single-Task和Dual-Task，本文所提出的模型在准确率和F1值分别提升了5.2%和3.5%，以及5.4%和4%. 其原因主要在于：首先，已有域迁移工作MCD[60]、CLAN[61]、DFA[62]和CGDM[63]在设计网络结构和优化目标时是基于图像数据的特点，而本文的输入数据类型是时频谱图，与图像数据存在本质的差异. 其二，尽管EI[39]也利用了多源域对抗学习的思想，但如前所述，本文的网络结构设计与之有着很大不同，主要体现在特征提取器、梯度反转层以及损失函数设计. 其原因在于：相比于行为识别，人类情绪状态与感知信号存在更为隐蔽深层的关联性，且情绪体验的个体差异性更大，模型域适应难度更高. 上述结果表明：MultiSource-Dann在跨用户情绪识别时具备更好的泛化能力，且由于模型训练时目标域无标签数据的加入，进一步减小了源域和目标域数据在特征空间中的分布差异，使得模型的迁移能力更强.

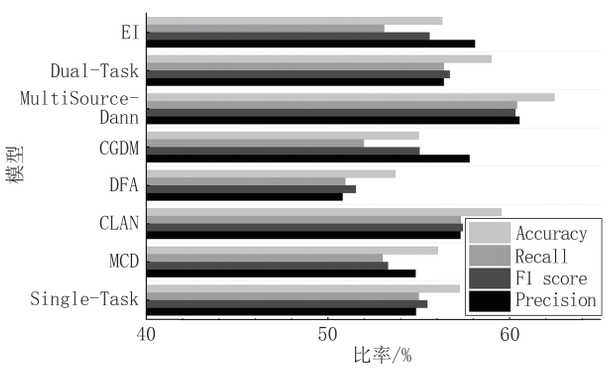


图6 不同模型的跨用户情绪分类性能

4.4.3 特征提取器的影响

在MultiSource-Dann跨被试模型框架的基础上，本文还对比了不同特征提取器的性能. 考虑到模型需在移动端部署，本文选择了较轻量化的深度神经网络作为特征提取器进行测试，包括MobileNet-V2[64]、MobileNet-V3[65]、ShuffleNet-V2[66]、SqueezeNet[67]、长短时记忆网络LSTM[68]、GhostNet[69]和深度残差网络ResNet18. 除LSTM的输入为五通道时间序列外，其他网络的输入均为时频谱图. 所得实验结果如图7所示. 可以看到在所对比的网络结构中，ResNet18表现最优. 其精确率、召回率、F1值和准确率分别为60.6%、60.3%、60.4%和62.5%. 相比于时序模型LSTM，ResNet18表现更优的原因可能是：情绪对于不同通道生理信号的影响主要表现在时频域中，也就是对信号在时域和频域两个维度上的能量分布产生影响. 不同的情绪类别表现为信号在时频域能量分布模式的差异. 因此，本文最终选取ResNet18作为情绪识别模型的特征提取器.

4.4.4 情绪类型的混淆情况

为了分析不同情绪类别之间的混淆情况以便更好地改进系统，本文分别统计了单被试和跨被试场景下Single-Task、Dual-Task及MultiSource-Dann

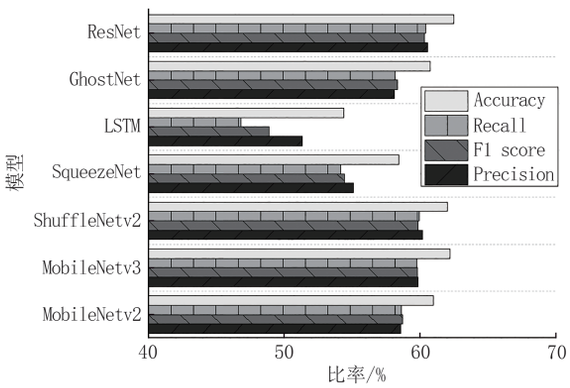


图7　采用不同特征提取器时的分类性能

方法的实验结果，得到如图8所示的情绪识别混淆矩阵．可以看到，无论是单被试还是跨被试情绪识别，其他情绪均更容易与中立情绪相互混淆．这可能是由于中立情绪的唤起强度较低，数据模式表现不明显．而Happy与Sad也较为容易相混淆．这可能是个体差异性和所唤醒的情绪强度导致．Mixed情绪准确率是最高的．这是由于刺激材料中引起Mixed情绪的素材唤醒度较高．此外，当使用MultiSource-Dann模型时，除Mixed类情绪的准确率稍有下降外，其他情绪的准确率均得到提高．

4.4.5　单通道及通道组合的影响

基于MultiSource-Dann跨被试模型，本文对单一通道及不同通道组合的情绪识别性能进行了对比实验．其中，ch0表示Mic-Ear通道，ch1表示Mic-Throat通道，ch2表示Mic-Nose通道，ch3表示PPG-Hand通道，ch4表示PPG-Head通道，ch01表示使用这两个通道的组合，即使用Mic-Ear通道和Mic-Throat通道组成的两通道进行情绪识别．从传感器性质角度来看，ch0、ch1与ch2为同种传感单元，ch3与ch4为同种传感单元．我们将所有传感单元组合进行实验后发现，同种传感单元组合效果差异不大．因此经筛选后，图中列举出具有代表性以及效果良好的组合情况．如图9所示，五个单一通道信号的情绪识别准确率有较大差异．其中，两个呼吸通道即ch1和ch2的效果最佳．ch2的信号具有最强的跨被试情绪表达能力．其情绪分类的精确率、召回率、F1值和正确率分别可达52.7%、52.2%、52.3%和55.1%，而ch1性能稍弱于ch2．这是因为呼吸所蕴含的情绪特征最为明显，如当用户开心的时候，鼻腔呼吸的速度会加快；伤心时鼻腔呼吸亦会加速，但与开心情绪相比稍弱一些．

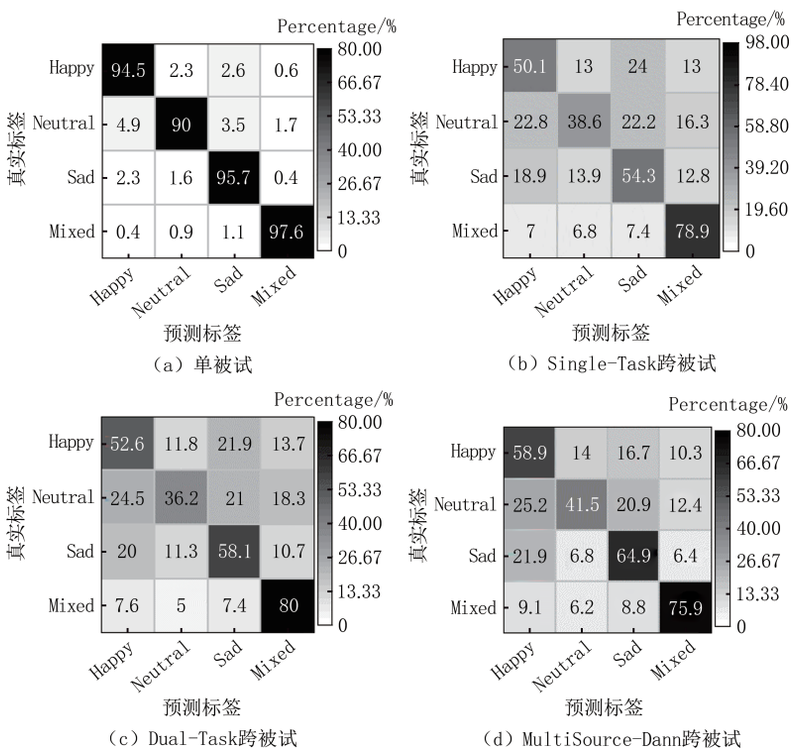


图8　不同模型的情绪分类混淆矩阵

同时，从图9所示的结果还可以看出，多通道组合的情绪识别性能性能普遍优于单通道．这表明不同通道的感知数据之间可以相互补充进而提高识别性能．当全部五通道数据一起使用时，系统情绪识

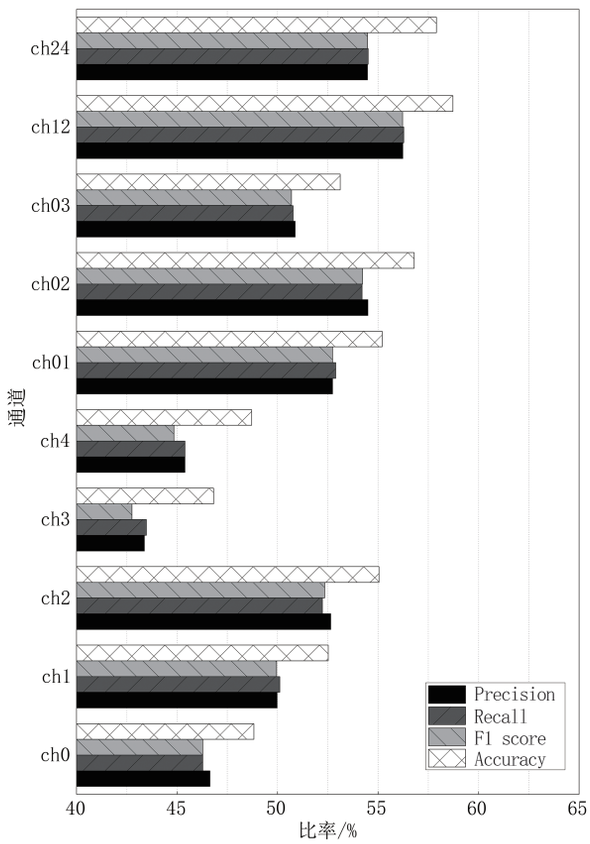


图9 单通道以及通道组合的影响实验结果

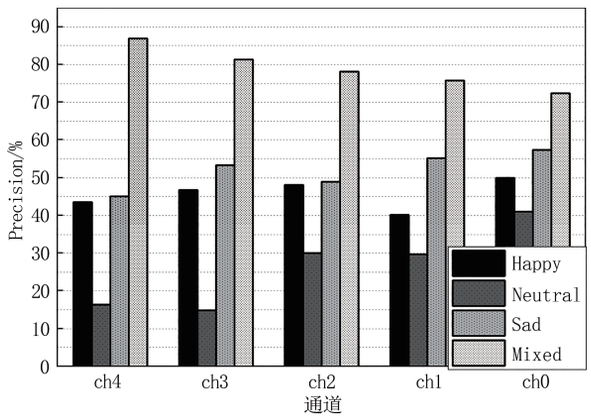


图10 单通道信号对不同类别情绪的识别效果

表3 可调参数λ的不同取值下的分类性能(%)

参数取值	Precision	Recall	F1 score	Accuracy
0.1	55.78	55.46	55.5	57.42
0.3	56.47	56.13	56.21	57.73
0.5	76.95	63.84	63.81	67.43
0.7	75.35	68.46	69.21	68.15
0.9	**81.97**	77.58	77.27	**77.94**

别的性能达到最佳.

为了分析不同通道信号对于四个类别情绪的识别效果,我们展示了如图10所示的结果.可以看出,不同的通道信号都具有一定的情绪偏向性.比如,除了ch0,其他通道信号对于Neutral类别的识别效果普遍较低.这是因为ch0在采集信号时,设备会与耳道形成一个密闭的空间,具备一定的滤噪功能,故而ch0通道的信号噪声较少,质量更高.而Neutral类别的特征是平静,所以ch0对于Neutral类别的信号识别效果会比其他通道更好一些.所有通道信号对于混合情绪均具有较好的识别效果.这是因为混合情绪往往信号变化更为明显,易于捕捉.

4.4.6 损失函数可调参数的影响

在4.3小节所述MultiSource-Dann模型设计中,为实现模型在训练过程中更专注于情绪分类的任务,我们为损失函数设计了一个可调参数λ.为了评估模型训练时该参数对于系统识别性能的影响,本文设定λ的取值分别为0.1、0.3、0.5、0.7和0.9,依次对模型进行训练和测试,得到了如表3所示的实验结果.可以看到,随着参数λ的增大,模型的识别性能也相应逐步升高.因此,为使系统达到良好的情绪识别性能,在实际模型训练过程中,我们设定随着训练进程的推进,参数λ的取值将自动地提升.

4.4.7 无标签目标域训练样本数的影响

针对MultiSource-Dann模型,本文进一步评估模型训练时使用目标域无标签样本数量对于系统识别性能的影响.本文对目标域四种情绪的无标签样本进行随机抽样,分别抽取了20个、40个、60个、80个、100个和120个样本进行实验.如图11所示,随着样本数的增加,系统的平均识别性能逐渐增加并最终趋于平稳.当反馈样本数小于等于40个时,情绪分类的F1值、准确率分别低于57%和59.8%;而当样本数达60个时,情绪分类的F1值、准确率均提高了约2%.由此可见,在模型训练过程中引入目标域无标签样本可以提高分类性能,且当样本量达到120个后,模型性能已趋于平稳.因此在后续实验中,本文默认采用从目标域数据中随机抽取120个无标签样本进行模型训练.

4.4.8 有标签目标域微调样本的影响

如3.5节所述,本文还提出了采用少量目标域带标签样本对模型分类器进行参数微调以进一步提升分类性能.为了评估反馈的带标签样本数量对于分类性能的影响,我们对四种情绪的目标域样本集进行随机抽样,分别使用采样时长为5 min、10 min、15 min、20 min、25 min和30 min的带标签样本进行模型微调.由于本文设定传感器的采样频率固定为

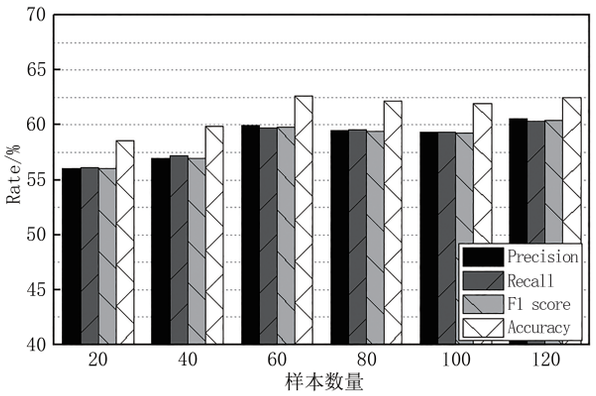
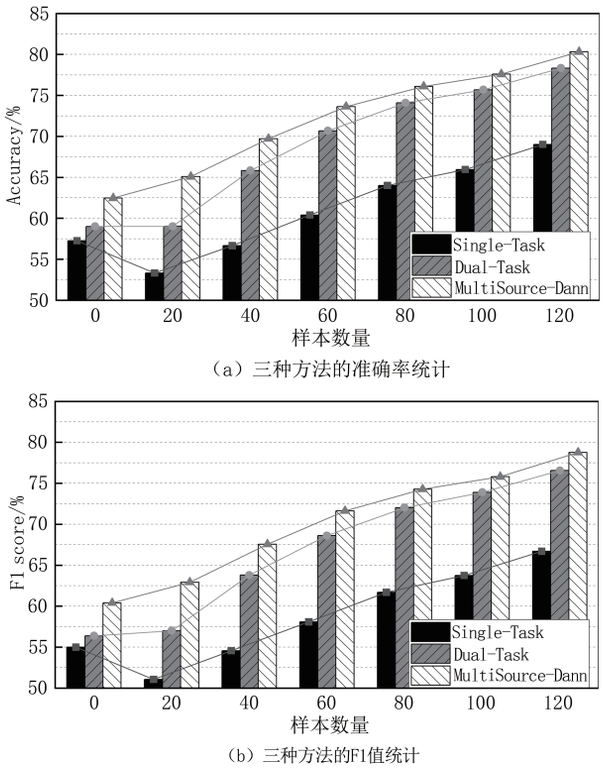


图11　MultiSource-Dann无监督训练样本数的影响

1000 Hz，并且以15 s的窗口长度对连续信号进行切分，所以相应得到的样本数量分别为20个、40个、60个、80个、100个和120个．我们使用这些样本分别对三种训练好的模型Single-Task、Dual-Task、MultiSource-Dann的分类器参数进行微调，然后进行跨被试测试．所得结果如图12所示．可以看到，无论何种分类模型，随着反馈样本数量的增加，其情绪识别的准确率和F1值也不断提升．这表明个体差异性对于情绪识别的确存在较大影响，但可以通过加入少量个体样本对模型进行微调以提升其性能．对于被测的三种模型，加入目标域样本进行参数微调后，MultiSource-Dann的平均性能提升最大，其次为Dual-Task和Single-Task．当反馈样本数量达到120个（即采样时间为30 min），MultiSource-Dann的精确率、召回率、F1值和准确率分别可达78.8%、78.8%、78.8%和81.1%．而此时基础模型Single-Task的精确率、召回率、F1值和准确率仅分别为67.1%、66.5%、66.7%和69%．对比图11所示结果，相比于仅使用目标域无标签样本进行模型训练，使用带标签样本微调模型的方法可以将MultiSource-Dann的识别准确率从62.5%进一步提升至81.1%．这是因为采用有标签样本对模型分类器参数进行微调时，其他层的参数均被冻结．这既保证非监督对抗学习训练过程中所得特征提取器依旧能够提取域无关的特征，又能引导模型快速学习目标域的分布并进行参数调整．

表4则展示了三种方法在加入120个目标用户有标签数据进行微调后的情绪分类混淆矩阵．可以看到，相较于基础模型Single-Task、MultiSource-Dann和Dual-Task对每种情绪的分类准确率均提高了约10%．其中Neutral情绪的识别准确率提升最明显．这也证实了个体差异性对情绪识别的影响



（a）三种方法的准确率统计

（b）三种方法的F1值统计

图12　自反馈的有标签样本量对识别性能的影响

是客观存在的．而在本文中，与基础模型Single-Task相比，Dual-Task和MultiSource-Dann具备更好的泛化能力，在加入少量目标用户的样本进行分类器参数微调后，能使模型更好地适配目标用户，显著提升分类性能，从而实现在保证识别准确率的前提下尽可能减少用户反馈样本负担．

表4　有监督参数微调后不同模型情绪分类的混淆矩阵

	真实标签＼预测标签	Happy	Neutral	Sad	Mixed
Single-Task自适应后	Happy	**67.7**	9.5	16.3	6.5
	Neutral	23.8	**45.9**	20.3	9.9
	Sad	16.0	7.3	**71.0**	5.6
	Mixed	7.0	4.6	7.1	**81.2**
	真实标签＼预测标签	Happy	Neutral	Sad	Mixed
Dual-Task自适应后	Happy	**76.4**	8.5	11.1	4.0
	Neutral	19.3	**61.5**	13.1	6.1
	Sad	10.0	6.8	**80.0**	3.2
	Mixed	4.2	3.7	4.0	**88.1**
	真实标签＼预测标签	Happy	Neutral	Sad	Mixed
MultiSource-Dann自适应后	Happy	**79.8**	8.7	11.1	3.1
	Neutral	18.1	**72.5**	10.7	5.4
	Sad	9.5	5.0	**82.2**	3.3
	Mixed	3.1	3.9	3.1	**89.9**

4.4.9 用户多样性的影响

由于身体机能及个人经历的差异，个体对于情感的认知与体验存在较大的个体差异. 为验证本系统对用户多样性的支持，本文对30名实验者在跨被试场景下采用Single-Task、Dual-Task、MultiSource-Dann三种模型所得的情绪分类结果进行统计，得到了如图13所示的结果. 其中，(w)标签代表对三种模型使用120个目标域样本进行分类器参数微调；而(wo)标签则表示未对模型进行参数微调. 由图可知，未采用监督式参数微调的情况下，三种方法的平均正确率分别为57.5%、59%和62.5%，标准差分别为6.3%、6.5%和9.6%. 具体到不同用户的识别准确率，相比于Single-Task模型，使用MultiSource-Dann模型时，大部分的个体准确率均有提升，并且比采用Dual-Task时的性能提升更为显著. 此外还可以看到，采用监督式参数微调后，三种方法的平均正确率分别为69.8%、79.1%和81.1%，标准差分别为5.1%、5.3%和6.2%. 可以看到，参数微调使得三种方法的平均准确率分别提高了12.3%、20%、18.7%. 相比于Single-Task，微调策略使得Dual-Task和MultiSource-Dann的分类性能提升更加显著，表明目标域样本的指导作用被进一步放大. 对于单个用户而言，对模型进行后参数微调后，Dual-Task和MultiSource-Dann的性能均显著优于Single-Task.

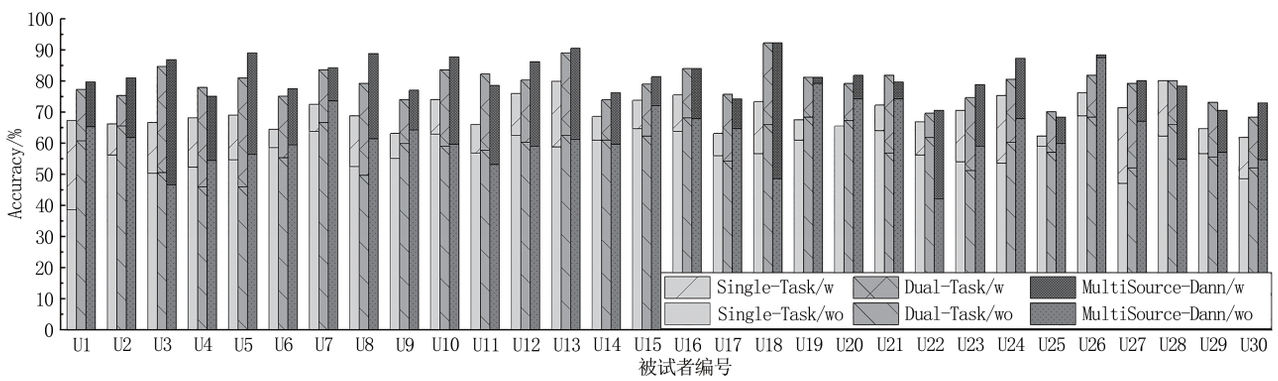


图13 三种方法跨被试用户分类性能对比统计

为进一步验证本文所提模型MultiSource-Dann对不同用户的泛化性能，除上述30名实验者外，我们还额外招募了15名年龄分布更宽广(15岁至65岁之间)、职业更多样、受教育程度差异化更显著的志愿者作为被试个体参与测试实验. 此部分实验的设置与流程与文中4.1节“实验设置与系统搭建”所述一致. 为使实验场景更加真实，我们将每人的数据采集过程分散到为期约45天的时段中，最终总共采集到约9800条样本. 其后，我们将此部分所采集到的数据作为测试集，用于测试前文基于30人的样本集而训练所得的模型MultiSource-Dann的情绪识别性能. 我们分别统计了这部分被试者的情绪识别准确率. 其结果为：当新用户提供120条样本用于模型参数微调时，新用户的平均情绪识别准确率为80.9%，标准差为8.7%. 这一结果与上述30人的实验结果基本相当. 这表明：由于采用目标域少量样本对模型参数进行快速微调的机制，本文所提方法能够很好地应对用户个体差异，保证方法的性能稳定.

4.4.10 环境因素的影响

由于本系统中采用了呼吸音和心跳音作为感知数据，本文对外界环境中的噪声干扰也进行了评估. 除前文4.1节中所述的场景(办公环境，40 dB～55 dB)外，我们还开展了系统处在其他环境噪声等级时的性能评估. 由于时间限制及人员流动的原因，我们招回原30名实验者的10人按照4.1节中的实验流程在噪声等级为55 dB～65 dB的办公室中平均每人采集了约800个样本. 其中，噪声等级的控制是通过控制播放白噪声音频文件的音量来实现的. 所采集的数据均只用于测试使用“留一用户”原则而训练好的MultiSource-Dann模型，也即保证模型测试既跨用户又跨环境. 本文分别统计了此十人在两种环境下的情绪识别结果，如表5所示. 可以看到，环境噪声的提高将会降低系统的分类性能，准确率下降约4%. 其原因在于噪声的存在将会降低麦克风所拾取到的呼吸音和心跳音的信噪比，从而影响模型最终的分类性能. 另一方面，由于多通道数据融合的原因，尽管部分模态数据被干扰，系统的分类性能也不会下降严重. 此外，鉴于日常家居、环境的背景噪声一般在55 dB以内，所以本系统依然具有较广阔的应用场景.

表5　不同环境下的分类性能(%)

	Precision	Recall	F1 score	Accuracy
40～55 dB	79.15	78.06	78.60	82.13
55～65 dB	75.34	76.30	75.82	78.32

4.4.11　硬件及其佩戴方式的影响

最后，鉴于生理信号易受感知设备及其佩戴方式的影响，本文对3.1节中所描述的可穿戴数据采集系统(记为A系统)做了如下修改：(1)考虑传感器硬件差异对信号的影响，将所有传感器替换为同类型但不同品牌与型号的产品；(2)将原位于图2所示位置2处的PPG传感器改至食指指尖；(3)将原位于图2所示位置4处的PPG腕带进行左右手腕调换. 经过以上改动后的可穿戴系统记为B系统. 如以上4.4.10节，我们招回原30名实验者的10人，让其完全按照4.1节中的实验流程，使用B系统平均每人采集约800个样本. 相似地，所采集的数据均只用于测试使用“留一用户”原则而训练好的MultiSource-Dann模型，也即保证模型测试既跨用户又跨硬件系统. 本文分别统计了此十人使用两套系统时的情绪识别结果，如表6所示. 可以看到，设备及佩戴位置的改变会对系统的识别性能造成一定的影响，准确率下降约2%. 这是因为传感器器件本身的差异会造成感知信号的不一致性，此外佩戴位置的改变也会改变感知信号. 然而，考虑到商用硬件的稳定性相较于实验室自制原型系统更优，且穿戴式设备由于形态限制，其佩戴方式基本固定，因此其影响也相对较小. 当然，本文所开展的这方面评估尚不充分，在后文第5节中有对此的进一步讨论.

表6　使用不同硬件设备时的分类性能(%)

	Precision	Recall	F1 score	Accuracy
A系统	79.15	78.06	78.60	82.13
B系统	77.38	76.47	76.92	80.06

4.4.12　相关工作对比

表7给出了本文和已有情绪识别工作在刺激材料、被测人数、感知信号、情绪类别和准确率五个维度的对比结果. 可以看出，目前基于EEG信号的单被试和跨被试情绪识别，可达到较高的识别准确率. 使用调频连续波雷达识别情绪的方法则存在依赖专用硬件且使用位置受限等不足. 而本文所提出的方法对于四种基本情绪的单被试识别准确率可达95.0%. 该结果与基于EEG信号的方法基本相当. 需要指出的是，尽管本文所提技术在跨被试情况下的准确率略低于已有基于EEG的工作，但其具有被测人员更丰富，所用设备更普适，更适宜连续监测等优势. 此外，结合本文所提出的基于多源域对抗思想的情绪识别模型，在提供少量目标域带标签样本

表7　相关工作对比

	刺激材料	被测人数	情绪类别	信号	平均识别准确率
文献[35]	自备	12	sadness, anger pleasure, joy	身体反射的调频连续波	单被试87.0% 跨被试72.3%
文献[13]	DEAP	5	fear, frustrated, sad, satisfied, pleasant, happy	EEG	跨被试57%
文献[24]	多个融合	13	负面、正面、中性	EEG 眼动轨迹	跨被试69.72%
文献[15]	SEED	15	neutral, sad, happy	EEG	单被试93.34% 跨被试84.41%
文献[17]	SEED	15	neutral, sad, happy	EEG	跨被试86.7%
文献[18]	SEED MPED	15 23	positive,neutral,negative	EEG	跨被试86.3% 跨被试74.77%
文献[19]	SEED	15	neutral, sad, happy	EEG	跨被试74.19%
文献[20]	DEAP	32	arousal, valence, dominance	EEG	跨被试66.87%～69.92%
文献[25]	日常生活	8	LVHA,NAHA, HVHA, LVNA, NVNA, HVNA, LVLA, NVLA, HVLA	EEG 脉搏 血压	跨被试 53.77%～73.48%
文献[6]	FileStim	20	fear, surprise, sadness, anger, happiness, disgust, neutrality	单眼图像	跨被试72.2%
本文	多个融合	30	neutral, happy, sad, mixed	可穿戴 生理数据	单被试95.0% 跨被试62.5% 微调后跨被试81.1%

时，本工作的情绪识别准确率可从62.5%提升至81.1%. 此时系统性能与基于EEG信号的情绪识别技术基本相当.

4.4.13 系统运行性能

最后，我们还开展了对实时系统运行性能的评估实验. 实验前，我们将实验设备（可穿戴硬件、红米K30s智能手机）充电至满电量状态. 实验开始后，实验者佩戴并启动可穿戴硬件系统使其正常工作，同时启动手机端配套APP，关闭其他所有应用和服务，持续运行进行情绪识别. 整个实验过程持续120 min，每隔10 min使用Android API获取手机的实时电量，得到测试结果如图14所示. 由图可知，本系统持续运行2 h后，智能手机的电量由100%下降至95%. 据此可估算出满电状态下该手机可支持系统运行约40 h. 同时，对硬件进行功耗测试的结果表明：传感器采样率为1000 Hz时，硬件单位时间的耗电为250 mA. 对于配备1000 mAh电池的本系统而言，其可持续采集与发送数据4 h. 考虑到在硬件设计时增加了休眠与唤醒机制，因此实际使用时硬件的续航能力将更高. 以上结果表明：当传感器的采样率设置为1000 Hz时，本系统（硬件&软件）具有良好的持续工作能力.

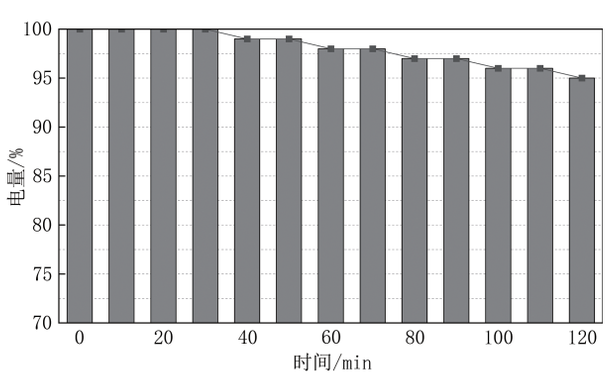


图14 系统运行时移动端电量消耗

此外，我们还评估了移动端APP对手机CPU和内存的占用情况. 测试时，关闭其他所有应用和服务，在USB调试模式下使用Android Profiler对CPU和内存占用进行持续100 s的实时监测，得到了如图15所示的结果. 可以看到，蓝牙连接稳定后，系统周期性地进行数据采集、数据预处理和情绪推断三个阶段. 应用程序的内存消耗保持在约260 MB，CPU占用率则在2%到32%之间波动. 其中，数据预处理和情绪推断阶段的CPU占用率最高，分别约为12%和25%. 然而这两个阶段的总耗时约为1.4 s，其中情绪推断的耗时更是仅约为73.2 ms. 因此，系统的总体CPU资源消耗与内存消耗是合理的.

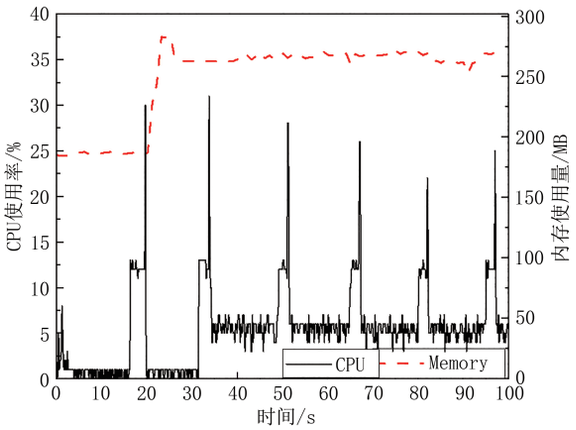


图15 系统运行时移动端CPU与内存占用

5 讨论与展望

尽管本文所提技术在普适性、易用性等方面具有显著优势，但仍然存在以下几个方面有待改进和优化：

（1）细粒度情绪识别

目前本文只研究了四种基本情绪的识别问题. 然而，人类的情绪更为精细复杂，情绪类别的划分也更为丰富多样. 此外，根据多维连续情绪理论，人类情绪也并不局限于离散的类别之分，还存在极性、强度等多个维度上的连续取值的差异. 正因如此，依据不同的情绪理论，情绪识别的任务也存在较大的差异. 如何利用普适可穿戴设备对个体情绪进行更细粒度、更精准的识别仍有待进一步探索.

（2）影响因素与系统鲁棒性

本文着重于探索利用普适可穿戴设备进行情绪识别的可行性，虽然对设备类型、佩戴位置、背景噪声等因素的影响做出了实验评估，但本实验中对这几个主要影响因素的区分仍不够细致，对于更多影响因素如用户运动状态等缺乏考虑，对于不同影响因素之间的各种组合情况也没有进行全面评估. 在后续工作中，我们将主要针对上述问题开展更全面的定量评估实验，以使该系统能够更加贴近实用场景.

（3）情绪识别模型的泛化性

受限于时间、经费等因素，本文目前的被试个体数目虽已超过现有相关研究工作，但仍然比较有限. 而本文也无法穷尽所有人员来参与测试实验以广泛地验证所提方法的泛化性. 虽然不同个体在情

绪表现上具有很强的差异性，但本文通过引入多源域对抗学习思想使得模型能够学习到数据在隐空间的特征表达，同时借助小样本微调模型参数的方法，能够让已训练好的模型快速迁移到目标用户的数据域，从而实现模型泛化能力的提升．可以预期的是，随着参与模型训练的人员数量与多样性增加，模型的特征提取能力将有效提升，进而增强其泛化性能．未来，我们将进一步增加被试者人数以更广泛地验证本文提出的情绪识别方法．

(4)应用场景的局限性

本文目前仅在实验室环境中对所提出的技术进行了验证，且要求用户处于相对静止状态．此种设置比较适合于医院、家居、办公室等应用场景．而在包含明显外界噪声和用户运动的应用场景如智慧城市等中，系统的识别性能将受到影响．在未来的工作中，我们将优化电路和外观设计，提升其抗噪水平和集成度，减小设备体积，使其能够与用户肢体更紧密地贴合，从而提升感知信号的质量．同时，未来可考虑改进信号处理、数据增强等方法，以此来提升系统的鲁棒性，以扩展情绪识别的应用场景．

6 总 结

鉴于已有情绪识别研究存在着隐私泄露、情绪隐藏、依赖专用设备、需要用户配合等不足，本文提出了基于普适可穿戴感知数据进行情绪识别的技术，利用呼吸、心跳和其他体声信号与情绪的潜在关联性，实现了对用户情绪的普适、便捷、连续监测．为了应对穿戴设备感知数据与用户情绪关联隐蔽难以挖掘的挑战，本文设计了多模态数据融合框架以提升不同模态感知信号对情绪的协同表征能力，提高情绪识别的准确性．同时，本文还设计了一种基于多源域对抗迁移学习的情绪识别模型，在保证良好识别准确率的同时，尽可能减少对被测主体的依赖性和新用户使用系统的成本．实验结果表明，本文所提出的基于可穿戴感知数据的情绪识别方法在单被试和跨被试场景下均具有良好的准确率，且优于所对比的几种方法．

参 考 文 献

[1] Chen Xue-Feng, Fu Xiao-Lan, Zhang Kan. Report on National Mental Health Development In China (2019-2020). 2021
(陈雪峰，傅小兰，张侃．中国国民心理健康发展报告(2019-2020)，2021)

[2] Plutchik R. Emotions and life: perspectives from psychology, biology, and evolution. Washington, USA: American Psychological Association, 2003

[3] Russell J A. A circumplex model of affect. Journal of Personality and Social Psychology, 1980, 39(6): 1161-1178

[4] Mehendale N. Facial emotion recognition using convolutional neural networks (FERC). SN Applied Sciences, 2020, 2(3): 1-8

[5] Hickson S, Kwatra V, Dufour N, Sud A, Essa I. Eyemotion: classifying facial expressions in VR using eye-tracking cameras//Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Waikoloa Village, Hawaii, USA, 2019: 1626-1635

[6] Wu H, Feng J, Tian X, Sun E, Liu Y, Dong B, Xu F, Zhong S. EMO: real-time emotion recognition from single-eye images for resource-constrained eyewear devices//Proceedings of the ACM International Conference on Mobile Systems, Applications, and Services. Toronto , Canada, 2020: 448-461

[7] Saha S, Datta S, Konar A, Janarthanan R. A study on emotion recognition from body gestures using Kinect sensor//Proceedings of the International Conference on Communication and Signal Processing. Bangkok, Thailand, 2014: 56-60

[8] Castellano G, Villalba S D, Camurri A. Recognising human emotions from body movement and gesture dynamics//Proceedings of the International Conference on Affective Computing and Intelligent Interaction. Berlin, Germany, 2007: 71-82

[9] Montepare J M, Goldstein S B, Clausen A. The identification of emotions from gait information. Journal of Nonverbal Behavior, 1987, 11(1): 33-42

[10] Lalitha S, Geyasruti D, Narayanan R, et al. Emotion detection using MFCC and cepstrum features. Procedia Computer Science, 2015, 70: 29-35

[11] Lalitha S, Mudupu A, Nandyala B V, et al. Speech emotion recognition using DWT//Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research. Madurai, India, 2015: 1-4

[12] Xu H, Zhang H, Han K, Wang Y, Peng Y, Li X. Learning alignment for multimodal emotion recognition from speech//Proceedings of the Annual Conference of the International Speech Communication Association.Graz, Austria,2019: 3569-3573

[13] Jiang S, Li Z, Zhou P, et al. Memento: An emotion-driven lifelogging system with wearables. ACM Transactions on Sensor Networks, 2019, 15(1): 1-23

[14] Wu S, Xu X, Shu L, Hu B. Estimation of valence of emotion using two frontal EEG channels//Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine. Kansa City, MO, USA, 2017:1127-1130

[15] Li Y, Fu B, Li F, et al. A novel transferability attention neural network model for EEG emotion recognition. Neurocomputing, 2021, 447:92-101

[16] Duan R N, Zhu J Y, Lu B L. Differential entropy feature for

EEG-based emotion classification//Proceedings of the 6th International IEEE EMBS Conference on Neural Engineering. San Diego, USA. 2013: 81-84

[17] Zhao L M, Yan X, Lu B L. Plug-and-play domain adaptation for cross-subject EEG-based emotion recognition//Proceedings of the AAAI Conference on Artificial Intelligence. Virtual event, 2021, 35(1): 863-870

[18] Song T, Liu S, Zheng W, et al. Instance-adaptive graph for EEG emotion recognition//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA, 2020, 34(3): 2701-2708

[19] DingKe-Ming, KimuraTsukasa, FukuiKen-ichi, and NumaoMasayuki. EEG emotion enhancement using task specific domain adversarial neural network//Proceedings of the International Joint Conference on Neural Networks. Shenzhen, China, 2021: 1-8

[20] Bhosale S, Chakraborty R, Kopparapu S K. Calibration free meta learning-based approach for subject independent EEG emotion recognition. Biomedical Signal Processing and Control, 2022, 72: 103289

[21] Cheng B, Liu G. Emotion recognition from surface EMG signal using wavelet transform and neural network//Proceedings of the International Conference on Bioinformatics and Biomedical Engineering. Shanghai, China , 2008: 1363-1366

[22] Cheng Z, Shu L, Xie J.,Chen C.L.P. A novel ECG-based real-time detection method of negative emotions in wearable applications//Proceedings of the International Conference on Security, Pattern Analysis, and Cybernetics, Shenzhen, China, 2017: 296-301

[23] Wu G, Liu G, Hao M. The analysis of emotion recognition from GSR based on PSO//Proceedings of the IEEE International Symposium on Intelligence Information Processing and Trusted Computing. Huanggang, China, 2010: 360-363

[24] Zheng Wei-Long, Shi Zhen-Feng, Lü Bao-Liang. Building cross-subject EEG-based affective models using heterogeneous transfer learning. Chinese Journal of Computers, 2020, 43(2): 177-189 (in Chinese)
(郑伟龙, 石振锋, 吕宝粮. 用异质迁移学习构建跨被试脑电情感模型. 计算机学报, 2020, 43(02): 177-189)

[25] Dai Yi-Xiang, Wang Xue, Dai Peng, et al. Stacked auto-encoder optimized emotion recognition in multimodal wearable biosensor network. Chinese Journal of Computers, 2017, 40(8): 1750-1763 (in Chinese)
(戴逸翔, 王雪, 戴鹏等. 面向可穿戴多模传感网络的栈式自编码器优化情绪识别. 计算机学报, 2017, 40(8): 1750-1763)

[26] Caridakis G, Castellano G, Kessous L, et al. Multimodal emotion recognition from expressive faces, body gestures andspeech//Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations. Peania, Athens, Greece, 2007: 375-388

[27] Haque A, Guo M, Miner A S, et al. Measuring depression symptom severity from spoken language and 3D facial expressions. arXiv preprint arXiv:1811.08592, 2018

[28] Das P, Khasnobish A, Tibarewala D N. Emotion recognition employing ECG and GSR signals as markers of ANS// Proceedings of the IEEE Conference on Advances in Signal Processing. Pune, India,2016: 37-42

[29] Dongmin Shin, Dongil Shin, Dongkyoo Shin. Development of emotion recognition interface using complex EEG/ECG bio-signal for interactive contents. Multimedia Tools and Applications, 2017, 76(9): 11449-11470

[30] Zheng W L, Lu B L. A multimodal approach to estimating vigilance using EEG and forehead EOG. Journal of Neural Engineering, 2017, 14(2): 026017

[31] Ware S, Yue C, Morillo R, et al. Large-scale automatic depression screening using meta-data from WiFi infrastructure. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(4): 1-27

[32] Salekin A, Eberle J W, Glenn J J, et al. A weakly supervised learning framework for detecting social anxiety and depression. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(2): 1-26

[33] Koelstra S, Muhl C, Soleymani M, et al. Deap: A database for emotion analysis using physiological signals. IEEE Transactions on Affective Computing, 2011, 3(1): 18-31

[34] Hu J, Shen L, Sun G. Squeeze-and-excitation networks// Proceedings of the IEEE/CVF Computer Vision and Pattern Recognition Conference. Salt Lake City, USA,2018:7132-7141

[35] Zhao M, Adib F, Katabi D. Emotion recognition using wireless signals//Proceedings of the ACM Annual International Conference onMobile Computing and Networking. New York, USA,2016:95-108

[36] Huang X, Kortelainen J, Zhao G, et al. Multi-modal emotion analysis from facial expressions and electroencephalogram. Computer Vision and Image Understanding, 2016, 147: 114-124

[37] Zhong B, Qin Z, Yang S, et al.,Emotion recognition with facial expressions and physiological signals//Proceedings of the IEEE Symposium Series on Computational Intelligence, Honolulu, USA, 2017:1-8

[38] Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks. Journal of Machine Learning Research, 2016, 17(59): 1-35

[39] Jiang W, Miao C, Ma F, et al. Towards environment independent device free human activity recognition// Proceedings of the ACM Annual International Conference on Mobile Computing and Networking. New Delhi, India, 2018: 289-304

[40] Li Y, Zheng W, Cui Z, et al. A Novel Neural Network Model based on Cerebral Hemispheric Asymmetry for EEG Emotion Recognition// Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. Stockholm, Sweden , 2018:1561-1567

[41] Luo Y, Zhang S Y, Zheng W L, et al. WGAN domain adaptation for EEG-based emotion recognition//Proceedings of the International Conference on Neural Information Processing System. Montréal, Canada, 2018:275-286

[42] Li J, Qiu S, Du C, et al. Domain adaptation for EEG emotion

recognition based on latent representation similarity. IEEE Transactions on Cognitive and Developmental Systems，2019，12(2)：344-353

[43] Wang Y，Liu J，Ruan Q，et al. Cross-subject EEG emotion classification based on few-label adversarial domain adaption. Expert Systems with Applications，2021，185：115581

[44] Hamann S. Mapping discrete and dimensional emotions onto the brain：controversies and consensus. Trends in Cognitive Sciences，2012，16(9)：458-466

[45] Homma I，Masaoka Y. Breathing rhythms and emotions. Experimental Physiology，2008，93(9)：1011-1021

[46] Hugdahl K. Psychophysiology：The mind-body perspective. Cambridge，MA，USA：Harvard University Press，1995

[47] Kreibig S D. Autonomic nervous system activity in emotion：A review. Biological Psychology，2010，84(3)：394-421

[48] Laukka P，Elfenbein H A，Söder N，et al. Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. Frontiers in Psychology，2013，4(353)：1-8

[49] Kim J. Robust Speech Recognition and Understanding. Wilmington，USA：Scitus Academics LLC，2017

[50] Ménard M，Richard P，Hamdi H，et al. Emotion recognition based on heart rate and skin conductance//Proceedings of the 2nd International Conference on Physiological Computing Systems. Loire Valley，France，2015：26-32

[51] He X，Zhang W. Emotion recognition by assisted learning with convolutional neural networks. Neurocomputing，2018，291：187-194

[52] Ghai M，Lal S，Duggal S，et al. Emotion recognition on speech signals using machine learning//Proceedings of the International Conference on Big Data Analytics and Computational Intelligence. Andhra Pradesh，India，2017：34-39

[53] Schaefer A，Nils F，Sanchez X，et al. Assessing the effectiveness of a large database of emotion eliciting films：a new tool for emotion researchers. Cognition and Emotion，2010，24(7)：1153-1172

[54] He K，Zhang X，Ren S，et al. Deep residual learning for image recognition//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Las Vegas，USA，2016：770-778

[55] Koelstra S，Muhl C，Soleymani M，et al. Deap：a database for emotion analysis using physiological signals. IEEE Transactions on Affective Computing，2011，3(1)：18-31

[56] Bradley M M，Lang P J. The International affective digitized sounds (2nd edition；IADS-2)：Affective Ratings of Sounds and Instruction Manual. Gainesville，USA：University of Florida，Technical report：B-3，2017

[57] Duan R N，Zhu J Y，Lu B L. Differential entropy feature for EEG-based emotion classification//Proceedings of the International IEEE/EMBS Conference on Neural Engineering. San Diego，USA，2013：81-84

[58] Carvalho S，Leite J，Galdo-Álvarez S，et al. The emotional movie database (EMDB)：a self-report and psychophysiological study. Applied Psychophysiology and Biofeedback，2012，37(4)：279-294

[59] Bradley M M，Lang P J. Measuring emotion：the self-assessment manikin and the semantic differential. Journal of Behavior Therapy and Experimental Psychiatry，1994，25(1)：49-59

[60] Saito K，Watanabe K，Ushiku Y，et al. Maximum classifier discrepancy for unsupervised domain adaptation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City，USA，2018：3723-3732

[61] Luo Y，Zheng L，Guan T，et al. Taking a closer look at domain shift：category-level adversaries for semantics consistent domain adaptation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach，California，USA，2019：2507-2516

[62] Wang J，Chen J，Lin J，et al. Discriminative feature alignment：improving transferability of unsupervised domain adaptation by Gaussian-guided latent alignment. Pattern Recognition，2021，116：107943

[63] Du Z，Li J，Su H，et al. Cross-domain gradient discrepancy minimization for unsupervised domain adaptation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual event，2021：3937-3946.

[64] Sandler M，Howard A，Zhu M，et al. Mobilenetv2：inverted residuals and linear bottlenecks//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City，USA，2018：4510-4520

[65] Howard A，Sandler M，Chu G，et al. Searching for mobilenetv3//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul，Republic of Korea，2019：1314-1324

[66] Ma N，Zhang X，Zheng H T，et al. Shufflenet v2：Practical guidelines for efficient CNN architecture design//Proceedings of the European Conference on Computer Vision. Munich，Germany，2018：116-131

[67] IandolaForrest N.，MoskewiczMatthew W.，AshrafKhalid，et al. SqueezeNet：AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. arXiv preprint arXiv：1602.07360，2016

[68] Hochreiter S，Schmidhuber J. Long short-term memory. Neural Computation，1997，9(8)：1735-1780

[69] Han K，Wang Y，Tian Q，et al. Ghostnet：More features from cheap operations//Proceedings of the IEEE/CVF Computer Vision and Pattern Recognition. Seattle，USA，2020：1580-1589

ZOU Yong-Pan, Ph. D, associate professor. His current research interests include mobile computing, ubiquitous computing, and human-computer interaction (HCI).

WANG Dan-Yang, M. S. candidate. Her research interest is affective computing

WANG Dan, M. S. candidate. Her research interest is affective computing.

ZHENG Can-Lin, M. S. candidate. His research interest is smart sensing and HCI.

SONG Qi-Feng, M. S. candidate. His research interest is affective computing.

ZHU Yu-Zheng, M. S. candidate. His research interest is smart sensing and HCI.

FAN Chang-He, Ph. D, chief physician. His research interests include biological psychiatry and behavioral medicine.

WU Kai-Shun, Ph. D, professor. His research interests include mobile computing, artificial intelligence, and IoT.

Background

The research in this work belongs to the field of affective computing focusing on automatic emotion recognition. As is well known, people are experiencing a steadily increasing psychological stress in today's fast-paced life. An ever-growing number of individuals are facing symptoms like emotional instability, persistent low mood, and even psychological disorders such as anxiety and depression. Automatic emotion re-cognition can effectively assist individuals in self-regulation and aid researchers in exploring the underlying mechanisms of psychological disorders, as well as facilitating corresponding treatments. Although researchers have proposed various kinds of methods for automatic emotion recognition based on different sensing mechanisms, they exhibit deficiencies in different aspects. Some researchers proposed electroencephalogram (EEG), electromyography(EMG), electrocardiogram (ECG), and/or galvanic skin response (GSR) -based approaches to emotion recognition. These methods can achieve high recognition accuracy but require the use of specialized, costly, and challenging-to-operate EEG devices. The computer vision-based methods relying on visual and speech cues carry privacy risks and are vulnerable external noises. The methods based on the analysis of mobile phone usage pattern need improvement in terms of reliability and accuracy. As a result, we consider a novel and challenging problem, that is, how to accomplish accurate and robust automatic emotion recognition with low-cost, readily available, and easy-to-use wearable hardware.

In response to this problem, this paper makes an attempt to design a wearable system integrated with microphones and photoplethysmography (PPG) sensors to collect body sounds produced by heartbeats and breathing, and blood pulse signals, respectively. This work makes use of the potential correlations between physiological signals, namely, breathing and heartbeat sounds, and pulse with human emotions. By employing data fusion across multiple sensing modalities, this work effectively harnesses diverse information types, reducing data redundancy, and substantially improving the system performance at the same time. Furthermore, while ensuring a high recognition accuracy, this paper also proposes an emotion recognition model based on a multi-source domain adversarial approach which aims to enhance the generalization of emotion recognition across diverse users and minimize the cost for unseen users. Extensive experiments demonstrate that our method can achieve an average recognition accuracy of 81.1% for the four basic emotions in cross-subject cases with only few shots from unseen users, surpassing the baseline methods by 12.4%.

This research is supported in part by the National Natural Science Foundation of China (No. 62172286), the Joint Funds of the National Natural Science Foundation of China (Key Program Grant No. U2001207), the Guangdong Basic and Applied Basic Research Foundation (No. 2022A1515011509), Science and Technology Program of Guangzhou (No. 202102010115), Guangdong Yiyang Healthcare Charity Foundation Program (No. JZ2022001-3), and Tencent "Rhinoceros Birds" - Scientific Research Foundation for Young Teachers of Shenzhen University.