

基于脑机协同智能的情绪识别

刘栋军^{1,2}, 王宇涵^{1,2}, 凌文芬^{1,2}, 彭勇^{1,2}, 孔万增^{1,2}

(1. 杭州电子科技大学计算机学院, 浙江 杭州 310018;
2. 浙江省脑机协同智能重点实验室, 浙江 杭州 310018)

摘要: 面部表情识别是一种直接、有效的情绪识别模式。机器学习可以对图像表情进行形式化表征, 但由于缺乏大脑的认知表征能力, 在小样本数据集或复杂表情(伪装)数据集上的识别性能并不理想。针对此问题, 将机器人工智能的形式化表征与人脑通用智能的情感认知能力结合, 提出一种基于脑机协同智能的情绪识别方法。首先, 从脑电图信号中提取脑电情感特征, 以获取大脑对情绪的认知表征。其次, 从情感图像中提取图像的视觉特征, 以获取机器对情绪的形式化表征。为了增强机器模型的泛化能力, 在特征学习中引入样本间的迁移适配。在得到图像视觉特征和脑电情感特征后, 采用随机森林回归模型训练得到图像视觉特征与脑电情感特征之间的脑机映射关系。测试图像的图像视觉特征经过脑机映射关系产生虚拟脑电情感特征, 然后将虚拟脑电情感特征与图像视觉特征进行融合, 用于情绪识别。该方法已经在中国面孔表情图片系统上进行了验证, 发现对7种情绪的平均识别准确率为88.51%, 相比单纯基于图像的方法, 提升了3%~5%。

关键词: 情绪识别; 脑电图信号; 脑机协同智能; 深度学习

中图分类号: TP18, TN911.7, R318

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202107

Emotion recognition based on brain and machine collaborative intelligence

LIU Dongjun^{1,2}, WANG Yuhan^{1,2}, LING Wenfen^{1,2}, PENG Yong^{1,2}, KONG Wanzeng^{1,2}

1. College of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China
2. Key Laboratory of Brain Machine Collaborative Intelligence of Zhejiang Province, Hangzhou 310018, China

Abstract: Emotion recognition is a direct and effective mode of emotion recognition. Machine learning relies on the formal representation of image expressions, lacks the cognitive representation ability of the brain, and has poor recognition performance on small sample data sets or complex expression (camouflage) data sets. To this end, the formal representation of machine artificial intelligence was combined with the emotional cognitive ability of human brain general intelligence, and a brain-machine collaborative intelligence emotion recognition method was proposed. Firstly, electroencephalogram (EEG) emotional features were extracted from EEG to obtain the brain's cognitive representation of emotions. Secondly, the visual features of the image were extracted from the emotional image to obtain the machine's formal representation of the emotion. In order to enhance the generalization ability of the machine model, the transfer adaptation between samples was introduced in the feature learning. After obtaining image visual features and EEG emotional fea-

收稿日期: 2021-01-10; 修回日期: 2021-02-08

通信作者: 孔万增, kongwanzeng@hdu.edu.cn

基金项目: 国家重点研发计划基金资助项目(No.2017YFE0116800); 国家自然科学基金资助项目(No.U20B2074, No.U1909202, No.61971173); 浙江省科技计划项目(No.2018C04012); 浙江省“脑机协同智能”重点实验室开放基金项目(No.20200E10010)

Foundation Items: The National Key Research and Development Program of China (No.2017YFE0116800), The National Natural Science Foundation of China (No.U20B2074, No.U1909202, No.61971173), Science and Technology Program of Zhejiang Province (No.2018C04012), Open Fund of Key Laboratory of Brain Machine Collaborative Intelligence of Zhejiang Province (No.20200E10010)

tures, the random forest regression model was trained to obtain the brain-machine mapping relationship between image visual features and EEG emotional features. The visual features of the test image were generated through the brain-machine mapping relationship to generate virtual EEG emotional features, and then the virtual EEG emotional features and image visual features were fused for emotion recognition. This method has been verified on the Chinese facial affective picture system (CFAPS) and found that the average recognition accuracy of the seven emotions is 88.51%, which is 3%~5% higher than the image-based method.

Key words: emotion recognition, EEG signal, brain-machine collaborative intelligence, deep learning

1 引言

情绪是一种综合人的感觉、思想和行为的状态,包括人对外界刺激或自身的心理反应,也包括伴随这种心理反应的生理反应^[1]。在人机交互系统中^[2],识别情绪状态可以让人和机器的互动变得更加友好和自然。因此,情绪分析和识别是涉及神经科学、心理学、认知科学、计算机科学和人工智能等领域的一项重要的交叉学科研究课题。通常使用面部表情、身体手势和声音等方法来推断一个人的情绪状态。根据相关调查^[3],言语成分传达了人类交流的三分之一,非言语成分传达了人类交流的三分之二。其中,非言语成分中更易于访问和处理的面部表情通过承载情感含义,成为日常交流的主要信息来源之一。因此,关于通过面部表情来识别情绪的研究在过去几十年中已经引起了很多关注。

人的大脑具有对情绪的认知表征能力,对情绪识别具有明显的优势。脑电图(electroencephalogram, EEG)是一种使用电生理指标记录大脑活动的方法,它记录了大脑活动时的电波变化,是脑神经细胞的电生理活动在大脑皮层或头皮表面的直接反映^[4]。EEG具有时间分辨率高(亚毫秒)、设备和实验成本低等特点。关于EEG情绪识别的研究主要有3类:一是针对脑电图信号,利用信号分析方法进行情绪识别,例如陈景霞等人^[5]提取功率频谱密度(power spectral density, PSD)特征及微分偏侧与因果差异特征,用于情绪识别,提高了用脑电图信号做情绪识别的准确率和稳定性;二是针对脑电图信号,利用传统机器学习方法进行情绪识别,例如Nie D等人^[6]通过电影视频片段来诱发受试者的情绪,获取受试者对应的脑电图信号,对积极、消极和中性3种情绪进行识别,从EEG数据中提取特征,使用支持向量机(support vector machine, SVM)分类方法可以达到平均识别准确率为87.53%的效果;三是针对脑电图信号,利用深度学习方

进行情绪识别,例如Yang Y等人^[7]使用卷积神经网络(convolutional neural network, CNN)和递归神经网络(recursive neural network, RNN)获取EEG中的情感特征,模型在效价(valence)和唤醒度(arousal)分类任务上的平均识别准确率分别为90.80%和91.03%。RNN对于序列化的数据具有良好的处理性能,该网络会记忆之前的信息,并利用之前的信息影响后面节点的输出^[8],适用于处理带有时间序列的脑电图信号。

从利用信号分析进行分类到利用深度学习进行分类,使用EEG做情绪识别的研究方法的整体性能在不断提升。基于脑电图信号做情绪识别任务需要大脑参与,每次进行识别任务需要重新采集脑电图信号,涉及多次复杂烦琐的数据预处理过程。

机器学习可以对图像表情进行形式化表征,可被用来处理情绪识别任务。随着计算机视觉领域的发展,用于情绪识别的深度学习算法日渐成熟。这些算法大部分被应用于特征提取、分类和识别等任务,例如, CNN的主要优点是能够直接从输入图像进行端到端的学习^[9]。基于深度学习的面部表情识别主要包含3个部分^[10]:一是人脸信息获取预处理,主要包含人脸检测、人脸对齐、数据增强以及人脸归一化4个过程;二是特征学习,主要基于卷积神经网络进行特征学习;三是情绪识别,基于深度学习得到的特征可以直接用于预测概率,也可以使用支持向量机等浅层分类器进行分类。Breuer R等人^[11]应用CNN构建模型,从而学习各种面部表情数据集,在Cohn-Kanade扩展数据集上取得98.62%的识别准确率。Mollahosseini A等人^[12]设计了一种深度神经网络架构,该网络模型在7个公开的面部表情数据库上进行了试验,在MultiPIE多视点人脸图像数据集上有98.7%的识别准确率。

从使用深度神经网络做情绪识别的研究可以看出,深度神经网络可以从图像中学到形式化表征,可以被用来处理情绪识别任务。但是这种方法存在需要大量标记数据来训练的问题,在小样本数

数据集上的识别性能并不理想。

脑机协同是人工智能时代实现更强大的智能形态的重要途径。例如王云^[13]提出的基于 EEG 的脑机接口技术和计算机视觉技术结合的方法可克服计算机视觉在光照、姿态和表情上的缺陷,实现较为快速、准确的人脸检索。人的大脑具有对情绪的认知表征能力,对情绪识别具有明显的优势,但需要人的参与,且涉及复杂的数据采集过程。机器学习可以从图像中学到形式化表征,可以被用来处理情绪识别任务,但需要建立在大量数据集的基础上。脑机协同不仅可以保留大脑和机器的优点,还可以解决两者存在的问题,实现脑机协同工作。因此本文提出基于脑机协同智能的情绪识别方法,将人的高级情感认知能力和机器学习得到的形式表征结合,以增强情绪识别的性能。在利用机器获取情感图像视觉特征时引入样本迁移学习的思想,将领域适配嵌入特征表示的学习过程,实现领域适配和特征学习训练过程的整合,进而建立机器学习的图像视觉特征与包含人脑高级情感的脑电情感特征之间的映射关系。再次输入新的情绪图像得到的图像视觉特征,经过映射得到虚拟脑电情感特征,虚拟脑电情感特征再与图像视觉特征融合并用于分类,实现情绪识别。

本文的整体结构如下:第1节介绍研究背景以及相关工作;第2节主要讲解实验的设计以及数据集;第3节介绍基于脑机协同智能的情绪识别方法,包含脑电情感特征提取方法、基于深度自适应网络(deep adaptation network, DAN)深度迁移学习的情感图像视觉特征提取方法以及脑机映射回归与融合方法;第4节对实验结果以及性能进行分析;第5节对全文进行总结,并展望未来。

2 实验设计与数据获取

在训练机器的学习模型时,大数据往往是机器进行学习和训练的关键。因此本文在选取实验数据时从该角度切入,仅选取 870 张小样本情绪图像数据,将脑电图信号得到的脑电情感特征与机器学习

得到的图像视觉特征结合,实现脑机协同工作。该实验涉及的所有面部情感图像来源于中国面孔表情图片系统(Chinese facial affective picture system, CFAPS)^[14],系统中的情感图片主要包含 7 类基础表情,共收集了 870 张情感面孔图片。其中,74 张表示愤怒、47 张表示厌恶、64 张表示恐惧、95 张表示悲伤、120 张表示惊奇、222 张表示中立、248 张表示快乐。为了引入人的认知能力,在传统机器进行图像识别的基础上增加了图像诱发的脑电图特征,以此获得高级的情绪表示。由于获取脑电情感特征需要对受试者进行情绪图像的刺激实验,实验设计在 E-pirme 完成,图像按类别顺序显示,每个图像以类别为顺序显示 0.5 s,然后在不同的情绪之间显示 10 s 的黑色图像。受试者需要头戴脑电图信号记录设备观看屏幕上依据实验设计展现的情感图像。实验中的脑电图信号记录设备使用的是由德国 Brain Products 生产的 68 通道采集系统,具有 62 个 EEG 通道。通过设置参数,消除 51 Hz 的外围干扰,采样率为 1 000 Hz。实验流程如图 1 所示。

该实验记录了 6 个受试者通过观察情绪图像产生的 EEG 数据,他们具有相同的教育背景,年龄为 23~25 岁,男女比例为 1:1。因为进行 EEG 数据采集实验需要受试者具有相关的知识背景,所以本文选取的受试者均为脑机接口与情绪识别领域的学习和研究人员。为了避免实验引入负样本而影响实验结果,本文尽可能选取对情绪图像肉眼分类准确的受试者参与实验。在实验之前,本文进行离线调查,检验受试者对情绪图像肉眼分类的准确率,方法如下:受试者直接观看电脑显示出的情绪图像,根据他们的判断勾选情绪所属类别,将他们的答案与真实标签进行对比并计算准确率,根据结果选出情绪图像肉眼分类准确的 6 名受试者作为本次实验的受试对象。最终得到 870 段 62 通道脑电图信号。在获得脑电图信号后,本实验对脑电图信号进行了预处理操作,以提高脑电图信号的信噪比并消除伪迹。预处理操作主要包括从原始数据中去除无效片段,去除由眼球运动、眨眼和肌肉抖动引

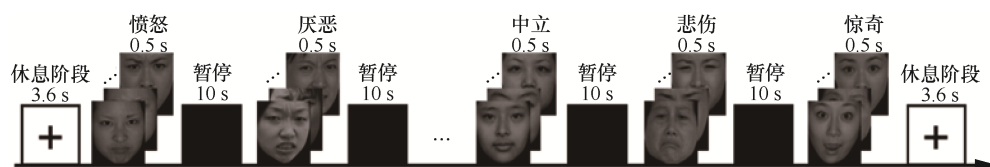


图1 针对基于 EEG 的情绪识别实验设计的实验流程

起的一些伪像,并使用 Butterworth 滤波器对信号的频率进行 1~75 Hz 滤波处理。

3 脑机协同智能

3.1 脑机协同情绪识别

本文提出一种基于脑机协同智能的情绪识别方法,整体框架如图2所示。分为3个模块,分别为脑电情感特征提取、深度域适配图像视觉特征提取及脑机映射回归与融合。脑电情感特征提取模块由门控递归单元(gated recurrent unit, GRU)网络完成,通过情绪图像诱发可获得受试者的脑电图信号,经过网络学习后可以获取更高级的脑电情感特征;深度域适配图像视觉特征提取模块由经过训练样本和测试样本迁移适配后的残差网络(residual network, ResNet)得到图像视觉特征;脑机映射回归与融合模块由前两个模块得到的脑电情感特征以及图像视觉特征共同训练随机森林回归模型,使该模型捕捉到 EEG 高级情感表征,从而具有生成虚拟脑电情感特征的能力。测试样本经过具有迁移能力的网络得到测试样本图像视觉特征后,再次经过训练好的回归模型后可以获得虚拟脑电情感特征。将虚拟脑电情感特征与图像视觉特征融合即可得到融合特征,融合特征经过已训练好的分类器即可实现对情绪图像的分类。引入迁移学习的方法可以拉近训练样本和测试样本的分布,提高回归模型的泛化能力。脑机协同智能的方法不仅可以脱离对大脑的依赖,而且具备大脑的认知表征能力,在情绪识别领域获得了较好的效果。

3.2 脑电情感特征提取

在基于 EEG 的情绪识别研究中,特征提取是至关重要的环节,只有提取出真正与情绪相关的特征,才能为后续的分类提供保证。常见的脑电情感特征主要分为3类:时域特征^[15]、频域特征^[16]和时频特征^[17]。但是这些方法简单地串联了每个通道的一维特征,忽略了通道在时间维度上的关系。近年来,具有复杂的递归隐藏单元(如长短期记忆(long short term memory, LSTM)网络和门控递归单元)的 RNN 已成为对 EEG 进行建模的主流选择^[18]。RNN 具有一定的记忆功能,因此与序列密切相关,可以被用来解决很多与时间相关的问题,如语音识别、语言模型、机器翻译等。但是它并不能很好地处理长时依赖问题。LSTM^[19]和 GRU^[20]旨在解决 RNN 长期记忆和反向传播中的梯度消失和梯度爆炸等问题。相比 LSTM,使用 GRU 能够达到相当的效果,并且参数更少、更易于训练,在很大程度上提高了训练效率,在小型数据集上更易收敛,因此本文采用 GRU 网络提取 EEG 情感特征。给定一系列输入 $x_1, x_2, \dots, x_n$ 和隐藏层状态 $h_1, h_2, \dots, h_n$, 其中 x_t 是时间为 t 时的输入,GRU 设计了两个门,即更新门 z_t 和重置门 r_t , $\tilde{h}$ 表示激活层状态。所有这些参数和变量必须满足以下计算式:

$$r_t = \sigma(w_r[h_{t-1}, x_t]) \quad (1)$$

$$z_t = \sigma(w_z[h_{t-1}, x_t]) \quad (2)$$

$$\tilde{h} = \tanh(w_h[r_t h_{t-1}, x_t]) \quad (3)$$

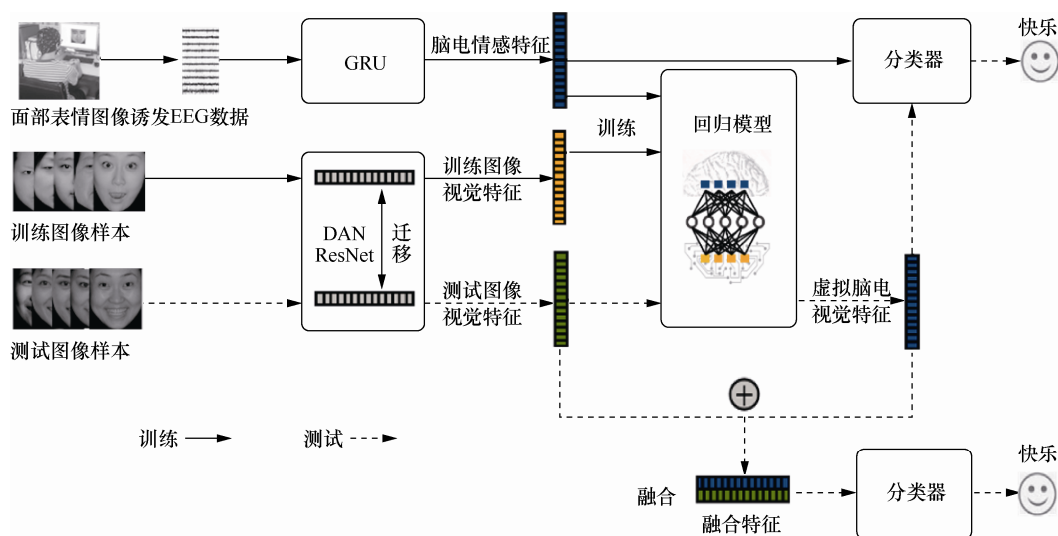


图2 基于脑机协同智能的情绪识别方法的框架

$$h_t = (1 - z_t)h_t + z_t\tilde{h}_{t-1} \quad (4)$$

其中, x_t 是时间 t 时的输入, σ 是激活函数, $[]$ 表示向量连接, w_r 、 w_z 、 w_h 是权重矩阵。更新门控制有多少前一时刻的状态信息被带入当前状态中, 更新门的值越大说明前一时刻的状态信息被带入得越多。重置门控制有多少前一时刻的状态信息被写入当前的候选集上, 重置门越小, 前一状态的状态信息被写入得越少。

本文通过设计一个堆叠的 GRU 模型从 EEG 中提取情感特征, 输入形状为 62×500 , 62 个通道每毫秒各产生一个信号, 每张图片显示 500 ms, 因此 62 个通道中每个通道都在 500 ms (即一个图片时间内) 产生 500 个信号。第一层使用包含 256 个神经元节点的多对多模型, 第二层使用包含 128 个神经元节点的多对一模型, 第三层采用 softmax 函数预测分类概率, 本文将第二层 GRU 输出的 128 维特征向量作为提取的 EEG 情感特征。GRU 基础结构如图 3 所示。

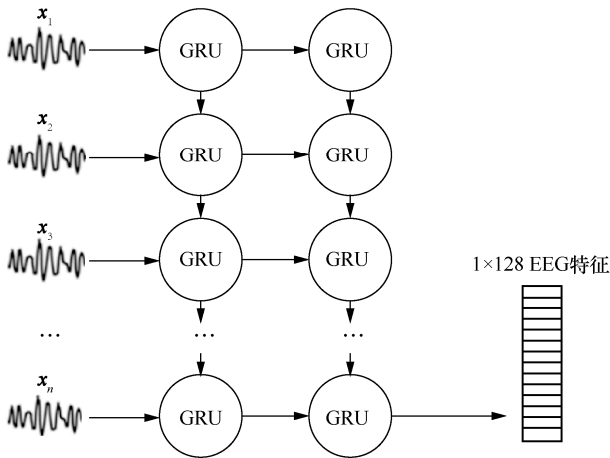


图3 GRU 基础结构

3.3 基于 DAN 深度迁移学习的图像视觉特征提取

迁移学习分为传统迁移和深度迁移两种。传统迁移方法多在特征层面做数据分布的拉近处理。深度迁移方法的优势在于在特征提取过程中引入了迁移思想, 将多核最大平均差异 (maximum mean discrepancy, MMD) 的参数学习融入深度网络的训练中, 并不断计算源域与目标域的距离, 并将其加入损失函数, 相比传统迁移方法可以达到更好的效果。DAN 是一种深度迁移神经网络, 可以有效地解决域适应问题^[21], 其核心思想在于采用多层适配与多核 MMD 度量^[22], 多层适配被用来计算源域和目标域的距离, 其中 $k(\cdot)$ 是映射, 用于把原变量映射

到高维空间中。X、Y 表示两种分布的样本, m 、 n 分别表示两种分布的样本数量, $\mathcal{F}$ 表示映射函数集, 如式 (5) 所示:

$$\text{MMD}[\mathcal{F}, X, Y] = \left[\frac{1}{m^2} \sum_{i,j=1}^m k(x_i, x_j) - \frac{2}{mn} \sum_{i,j=1}^{m,n} k(x_i, y_j) + \frac{1}{n^2} \sum_{i,j=1}^n k(y_i, y_j) \right]^{\frac{1}{2}} \quad (5)$$

本文的研究目标是找到更好的图像视觉特征到脑电情感特征的映射关系, 以提高回归后的识别准确率。本文假设模型的回归性能受到训练样本与测试样本之间的特征分布差异的影响, 鉴于回归器的作用是寻找图像视觉特征和脑电情感特征之间的映射关系, 本文合理推断, 对于具有训练样本特征空间特性的数据集来说, 训练好的回归器将有更好的回归效果。因此引入训练样本和测试样本之间的迁移, 以拉近二者的特征分布。如图 4 所示, 训练样本与测试样本的分布更接近。经实验验证, 与未使用 DAN 深度迁移方法得到的图像视觉特征相比, 将使用 DAN 深度迁移方法得到的测试样本图像视觉特征输入训练好的回归模型, 可以获得更好的回归效果, 这说明加入深度迁移方法可提高回归模型的回归性能, 进而提高虚拟脑电情感特征的识别准确率。

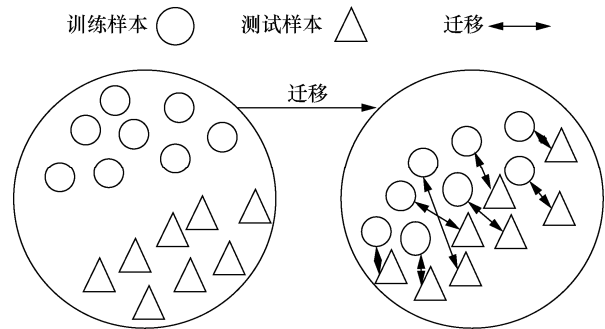


图4 训练样本与测试样本的迁移适配

本文实验主要采取的数据集为 CFAPS 中的 870 张二维灰度图像, 在实验中引入 DAN 深度迁移方法, 提取特征所采用的网络为 ResNet^[23]。本文将原始图像的大小调整为 227×227 像素, 作为 ResNet 网络的输入, 870 张情绪图像按照 9:1 的比例分为训练样本和测试样本。DAN 的思想在于利用迁移适配拉近样本间的距离, 并将其加入损失函数中。本文构建了最终损失函数中, 它由两部分构成: 交叉熵损失函数和自适应损失函数。最终损失

函数表示为 $L = L_p + \lambda L_d$ ，其中 λ 是超参数、 L_p 表示交叉熵损失函数、 L_d 表示自适应损失函数。交叉熵损失函数 L_p 是传统的损失函数，它被用来度量预测值和真实值的差异。本文的目标是使训练集的特征分布和测试集的特征分布更相近，这将通过自适应损失函数 L_d 实现。DAN 在深度学习网络的某些层加入自适应层。自适应层采用表征效果更好的多核 MMD，衡量样本数据之间的分布距离。将多核 MMD 的参数学习融合到深度网络的训练中，通过自适应损失函数实现不会额外增加训练操作和训练时间。最终损失函数通过反向传播梯度，不断更新优化器的参数，在这个过程中，不仅可以使预测标签和真实标签的分布距离更近，同时训练样本和测试样本的特征空间分布距离也更加相近。实现方法如图 5 所示。其中，Loss1 表示交叉熵损失函数、Loss2 表示自适应损失函数、Loss 表示最终损失函数、 λ 是超参数。

本文将利用 DAN 深度迁移进行训练样本和测试样本迁移适配得到的 ResNet 称为 DAN-ResNet。计算训练样本图像视觉特征与测试样本图像视觉特征的 MMD 距离，通过损失函数不断拉近二者的特征分布距离。在实验展示部分可以看到，在将平均脑电情感特征作为回归目标的情况下，经过 ResNet 得到的图像视觉特征回归生成的虚拟脑电情感特征识别准确率为 83.91%，经过 DAN-ResNet 得到的图像视觉特征回归生成的虚拟脑电情感特征识别准确率提升到 87.36%。可以得出，基于 ResNet 的 DAN 深度迁移具有良好的效果。

3.4 脑机映射回归与融合

脑机映射即寻找从大脑得到的认知表征与机器学习到的形式表征之间的映射关系。本文采用的模型为随机森林回归器，针对测试样本，机器学习得到的图像视觉特征经过此映射关系可以得到虚拟脑电情感特征。将虚拟脑电情感特征与图像视觉特征进行融合得到的融合特征可用于情绪识别。脑机映射回归与融合模型如图 6 所示，模型分为训练和测试两部分。在训练部分中，将脑电情感特征与图像视觉特征共同送入随机森林回归器进行回归训练，在此过程中寻找认知表征与形式表征之间的映射关系。在测试部分，测试样本经过 DAN-ResNet 得到的图像视觉特征经过回归映射，即可生成虚拟脑电情感特征，将虚拟脑电情感特征与图像视觉特征融合后得到的融合特征再送入分类器，实现对情绪图像的分类识别。

为了找到合适的映射关系，获取具有良好拟合性能的虚拟脑电情感特征，本文尝试使用一些主流的回归模型^[24]，例如随机森林（random forest, RF）、极端随机树（extratree, ET）和岭回归（kernel ridge regression, KRR）。采用多元回归输入输出模型，利用已提取的脑电情感特征与图像视觉特征对回归器进行训练，采用测试图像做预测，输出大小为 128 维的虚拟脑电情感特征。本文使用所有受试者的平均脑电情感特征和效果最好的第 5 名受试者的脑电情感特征来测试每个回归模型的性能，以 R^2 评估回归效果， R^2 反映因变量的全部变动能通过回归关系被自变量解释的比例值，如式（6）所示。

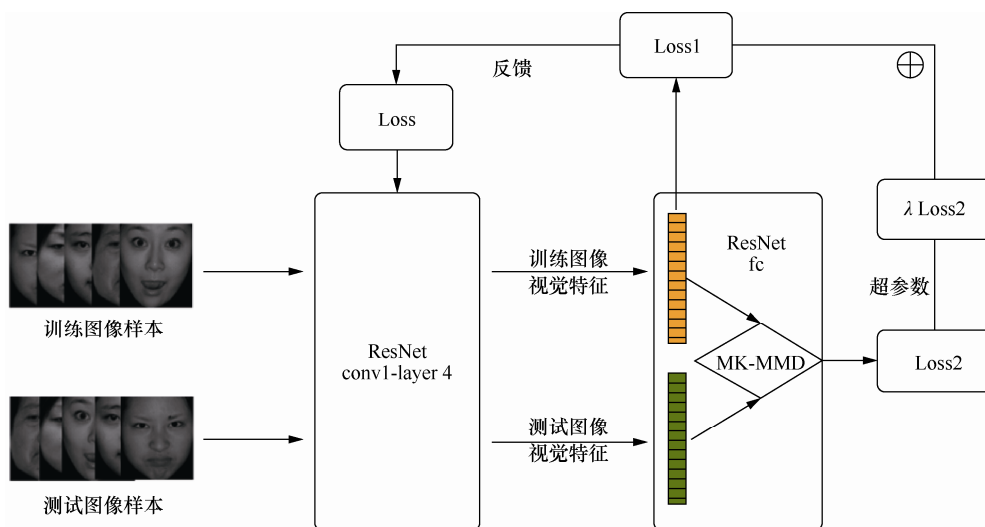


图5 图像视觉特征提取示意图

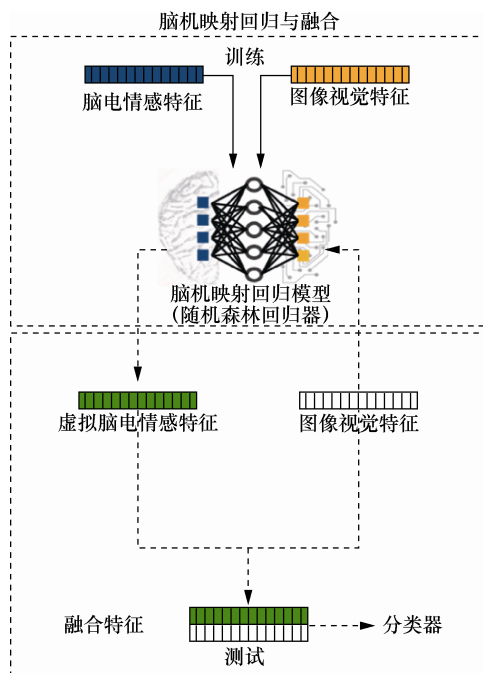


图6 脑机映射回归与融合模型

$$R^2(r_{\text{eeg}}, \hat{r}_{\text{eeg}}) = 1 - \frac{\sum_{n=1}^N (r_n - \hat{r}_n)^2}{\sum_{n=1}^N (r_n - \bar{r})^2} \quad (6)$$

其中, N 表示样本数量, r_{eeg} 表示脑电情感特征, $\hat{r}_{\text{eeg}}$ 表示虚拟脑电情感特征, r_n 表示样本, $\hat{r}_n$ 表示预测值, $\bar{r}$ 表示样本均值。计算虚拟脑电情感特征与脑电情感特征的 R^2 数据, 值越高表示两者的相关性越大, 回归器的性能越好。由 R^2 指标分析得到, 随机森林回归器具有更好的回归性能, 因此最终本文选择了随机森林回归器。采用 k -近邻 (k -nearest neighbor, KNN) 分类器对回归生成的虚拟脑电情感特征进行情绪识别, 实现脑机协同, 将人类在认知领域的识别功能加入机器, 机器学习到人类认知后进行高级分类。为了提高识别准确率, 本文还采取了融合策略, 将经过回归生成的虚拟脑电情感特征与图像视觉特征进行融合, 得到融合特征, 并训练新的分类器进行分类。相比于采用虚拟脑电情感特征进行情绪识别, 采用融合特征进行情绪识别的识别准确率提升 1% 左右。

4 实验结果与性能分析

为了减小脑电实验中的个体差异性的影响, 对 6 名受试者得到的脑电情感特征的值取平均, 得到平均脑电情感特征。在 6 名受试者中, 第 5 名受试

者的脑电情感特征用于情绪识别的效果最好, 本文单独将第 5 名受试者的脑电情感特征称为最佳脑电情感特征。在本文的实验中, 平均脑电情感特征对应的虚拟脑电情感特征是由图像视觉特征与平均脑电情感特征经回归生成的; 最佳脑电情感特征对应的虚拟脑电情感特征是由图像视觉特征与最佳脑电情感特征经回归生成的。

表 1 展示了实验结果的总体分析, 即经脑机映射回归得到的虚拟脑电情感特征对 7 种情绪图像进行分类识别的准确率 (accuracy)、精确率 (precision) 和召回率 (recall), 可以看出, 3 种评价指标得到的结果较为相近。另外, 本文采用十折交叉验证, 针对虚拟脑电情感特征进行分类, 平均识别准确率为 83.91%, 平均精确率为 85.33%, 平均召回率为 83.91%。

表 1 实验结果分析

指标	平均脑电情感特征	最佳脑电情感特征
accuracy	87.36%	87.36%
precision	88.48%	88.01%
recall	87.35%	87.36%

表 2 展示了 3 种回归模型的 R^2 得分。由表 2 可以看出, 无论是以平均脑电情感特征还是最佳脑电情感特征为回归对象, 随机森林的 R^2 得分最高。因此本文最终决定将随机森林回归器作为脑电情感特征和图像视觉特征的映射模型。

表 2 3 种回归模型的 R^2 系数

回归对象	随机森林	极端随机树	岭核回归
平均脑电情感特征	0.227 0	-0.297 3	0.215 1
最佳脑电情感特征	0.451 4	0.101 6	0.447 5

表 3 展示了加入迁移学习前后虚拟脑电情感特征用于情绪识别的准确率对比, 样本迁移表示训练样本与测试样本之间的迁移, $\times$ 表示训练样本和测试样本之间无迁移, $\sqrt$ 表示训练样本和测试样本之间有迁移。由表 3 可以看出, 引入训练样本与测试样本的迁移之后, 虚拟脑电情感特征的识别准确率有明显的提高。由于个体差异性, 相比平均脑电情感特征, 最佳脑电情感特征作为回归对象可以取得更好的效果。由此可以证明本文引入迁移学习可以提升回归器的回归性能, 得到更加优质的虚拟脑电情感特征, 从而提升情绪识别准确率。

表3 加入迁移学习前后虚拟脑电情感特征识别准确率

样本迁移	平均脑电情感特征	最佳脑电情感特征
×	83.91%	86.21%
√	87.36%	87.36%

表4展示了经过3种网络提取的图像视觉特征分别与平均脑电情感特征、最佳脑电情感特征结合,经过回归生成对应的虚拟脑电情感特征对情绪进行识别的准确率。由于本文实验选取的是小样本数据集,在本文选取的几种深度学习网络中分类效果不理想。因此,在小样本数据集的基础上,本文在特征提取过程中引入样本间的迁移学习,结合诱发的脑电图信号来提升识别准确率。经过对比可以看出,本文基于迁移学习的ResNet网络提取的图像视觉特征分类效果更好,利用训练图像视觉特征与脑电情感特征训练回归器,发现测试图像视觉特征经过回归器后得到的虚拟脑电情感特征识别准确率比单纯图像视觉特征识别准确率提高3.45%。

表4 3种网络提取图像视觉特征回归后的虚拟脑电情感特征识别准确率

方法	图像视觉特征	虚拟脑电情感特征	
		平均脑电情感特征	最佳脑电情感特征
LeNet	72.09%	75.86%	70.57%
AlexNet	82.76%	85.06%	85.06%
ResNet	83.91%	87.36%	87.36%

表5展示了虚拟脑电情感特征与图像视觉特征融合和不融合的识别准确率对比。其中,特征融合表示虚拟脑电情感特征融合图像视觉特征,×表示虚拟脑电情感特征没有融合图像视觉特征,√表示虚拟脑电情感特征融合了图像视觉特征。可以看出,相比虚拟脑电情感特征,融合之后的融合特征具有更好的情绪识别效果,最终本文得到的情绪识别准确率为88.51%。

表6展示了不同样本划分对比实验的结果。将870张情绪图像分别按照5:5、7:3、9:1划分训练样本和测试样本。从经过迁移学习得到的图像视觉特征与虚拟脑电情感特征对应的识别准确率可以看出,在不同的样本划分比例下,图像视觉特征经过脑机映射回归后生成的虚拟脑电情感特征的识别准确率比图像视觉特征识别准确率分别高1.81%、

3.45%、3.45%。另外,从实验结果可以看出,随着训练样本数的增加,情绪的识别准确率得到提升,在只有870张图片的情况下,情绪的识别准确率已经提升到87.36%,这说明即使是小样本数据,本文的方法依然会有较好的效果。最后本文以9:1划分训练样本和测试样本。

表5 虚拟脑电情感特征与图像视觉特征融合和不融合的识别准确率

特征融合	平均脑电情感特征	最佳脑电情感特征
×	87.36%	87.36%
√	88.51%	88.51%

表6 不同样本划分比例对应的识别准确率

样本划分比例 (训练样本:测试样本)	图像 视觉特征	虚拟脑电 情感特征
5:5	72.90%	74.71%
7:3	75.86%	79.31%
9:1	83.91%	87.36%

如图7所示,本文在实验中分别对4种特征进行情绪识别,并将结果进行对比,图7展示了对图像视觉特征(图像)、经过回归生成的虚拟脑电情感特征(回归)、图像视觉特征与虚拟脑电情感特征融合得到的融合特征(融合)以及脑电情感特征(脑电)进行情绪识别的准确率。相比于图像视觉特征,经过回归生成的虚拟脑电情感特征的识别准确率提升了4%左右。相比于经过回归生成的虚拟脑电情感特征,融合特征的识别准确率提升了1%左右,对于情绪识别具有更好的效果。

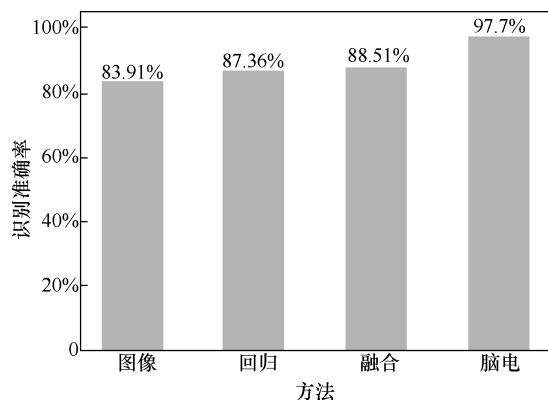


图7 不同方法对应的识别准确率

基于本文提出的基于脑机协同智能的情绪识别方法,直接将测试样本的图像视觉特征输入训练

好的回归模型中可以得到虚拟脑电情感特征,为了让虚拟脑电情感特征的表述更加形象和具体,本文将第5名受试者的测试样本的脑电情感特征数据分布与测试样本得到的虚拟脑电情感特征数据分布进行可视化。图8(a)为128维脑电情感特征数据分布,图8(b)为经回归器生成的128维虚拟脑电情感特征数据分布。可以看出,经过本文的方法得到虚拟脑电情感特征的数据分布与原脑电情感特征的数据分布较为相近,证实了虚拟脑电情感特征可以代替脑电情感特征进行情绪识别的可行性,同时也验证了实验数据的有效性。

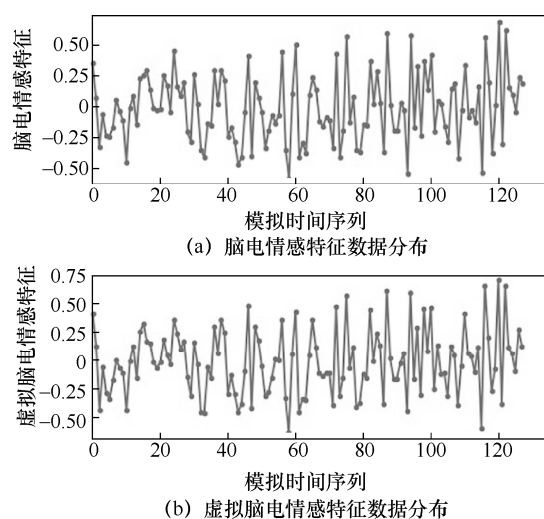


图8 脑电情感特征与虚拟脑电情感特征数据模拟分布

图9展示了DAN-ResNet训练过程中的损失和识别准确率的变化。从图9可以看出,当训练次数达到100次左右时,损失达到收敛,识别准确率稳步上升。

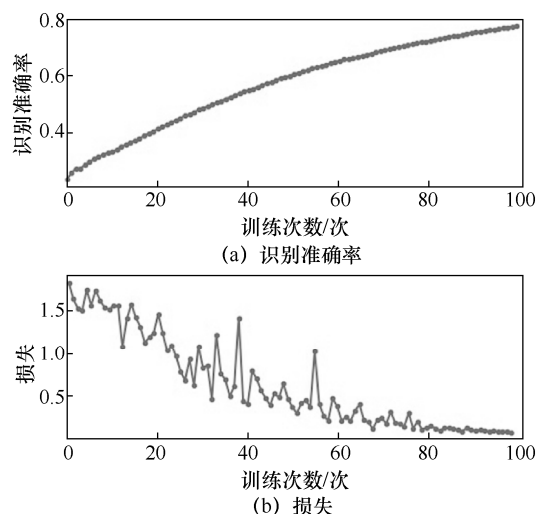


图9 DAN-ResNet 训练过程中识别准确率和损失

如图10所示,本文对比了回归前的图像视觉特征与经过回归生成的虚拟脑电情感特征对7种情绪进行识别的准确率。其中,横坐标为识别准确率,纵坐标为7种情绪类别。由图10可以看出,本文提出的基于脑机协同智能的情绪识别方法,即利用回归生成的虚拟脑电情感特征识别7种情绪,对于快乐、中立、悲伤和愤怒这4种情绪的判别较为准确;对于恐惧情绪的识别效果较弱;对于厌恶和惊奇两种情绪,相比于基于图像视觉特征的方法,识别准确率提升较多,由此可以推断出对于这两种情绪,机器更容易出现误判的情况。而经过回归加入了大脑的认知信息后,识别准确率有所提升,因此本文合理推断,相比于机器,大脑更易准确判别出两个表情背后所代表的情绪。

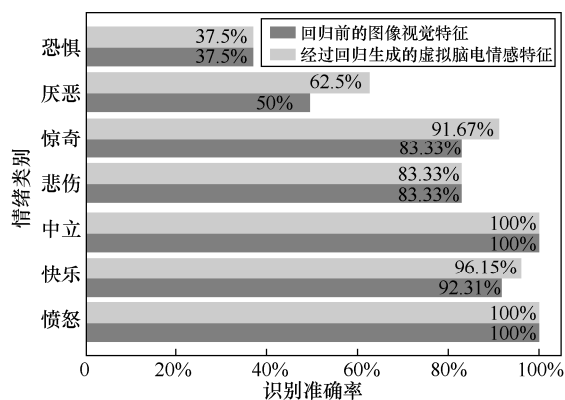


图10 回归前后对7种情绪的识别准确率

图11展示了利用t-SNE对7种情绪分别在图像视觉特征和经过回归生成的虚拟脑电情感特征下的聚类效果。图11(a)为回归前基于图像视觉特征的情绪聚类效果,可以清晰地看出,机器识别容易将厌恶与悲伤两种情绪混淆,同时恐惧与惊奇也存在聚类结果明显重叠的现象。图11(b)为回归后基于虚拟脑电情感特征的情绪聚类效果,可以看出经过回归之后,7种情绪被分为7种不同颜色,且无重叠混杂现象。由此验证了本文提出的基于脑机协同智能的情绪识别方法的有效性。

5 结束语

情绪识别作为人机交互的一项重要内容,近年来由于深度学习算法的大热而取得了前所未有的进展。EEG可以被用于评估和了解大脑中与心理、生理状态相关的变化,作为情绪状态的识别指标具有明显的优势。许多应用需要能够将知识迁移到新

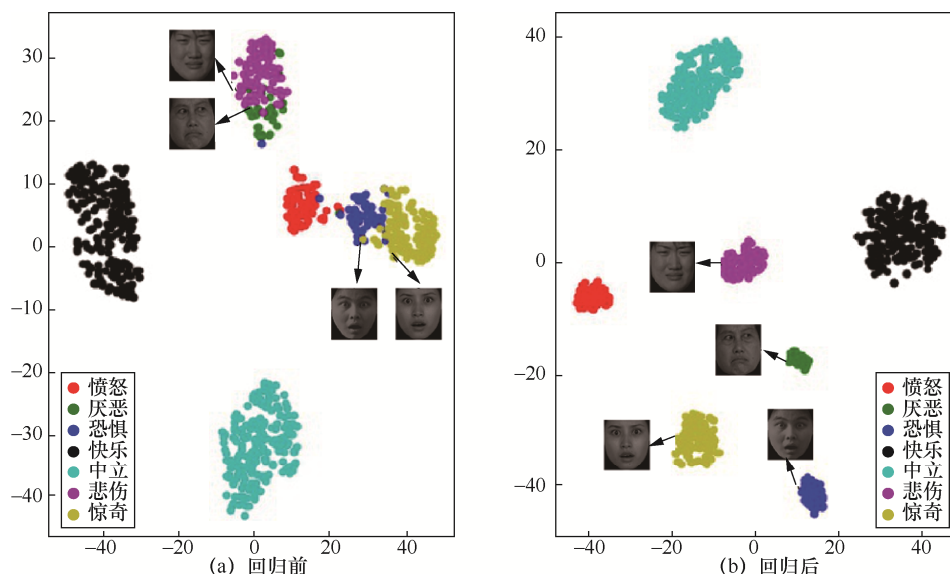


图 11 7 种情绪回归前后 t-SNE 聚类效果

的任务和新的域中的模型中, 迁移学习将会成为人工智能的一个重要助推力。本文提出了一种新颖的基于迁移学习脑机协作进行情绪识别的方法。首先, 利用 GRU 获取脑电情感特征; 其次, 采用经过训练样本和测试样本迁移适配后的 ResNet 提取图像视觉特征; 再次, 将脑电情感特征与图像视觉特征输入随机森林回归器进行训练, 并根据再次输入的图像视觉特征生成虚拟脑电情感特征; 最后, 进一步将虚拟脑电情感特征与图像视觉特征融合得到融合特征, 再对情绪进行分类与识别。结果显示, 虚拟脑电情感特征比图像视觉特征具有更好的识别效果, 这意味着机器可以学习到人类大脑的高级认知, 以提高识别情感的能力。在未来的工作中, 将尝试发现脑电情感特征和图像视觉特征之间更多的映射关系以及脑机协同更多的可能性, 继续探索机器人人工智能的形式化表征与人脑通用智能的认知表征之间的异构迁移, 寻找更多的应用场景。

参考文献:

- [1] 聂聃, 王晓韡, 段若男, 等. 基于脑电的情绪识别研究综述[J]. 中国生物医学工程学报, 2012, 31(4): 595-606.
- [2] DORNAIKA F, RADUCANU B. Efficient facial expression recognition for human robot interaction[C]//Proceedings of the 9th International Work-Conference on Artificial Neural Networks. Berlin: Springer, 2007: 700-708.
- [3] HARPER R G, WIENS A N, MATARAZZO J D. Nonverbal communication: the state of the art[M]. [S.l.]: John Wiley & Sons, 1978.
- [4] 尹旺, 李惠媛. 深度学习在脑电情感识别方面的应用研究进展[J]. 计算机时代, 2020(8): 14-17.
- [5] YIN W, LI H Y. The application research progress of deep learning in EEG emotion recognition[J]. Computer Era, 2020(8): 14-17.
- [6] 陈景霞, 郑茹, 张鹏伟, 等. 基于数据空间自适应与共空间模式的脑电情感分类[J]. 计算机工程, 2019, 45(4): 296-301.
- [7] CHEN J X, ZHENG R, ZHANG P W, et al. EEG sentiment classification based on data space adaptation and co-space mode[J]. Computer Engineering, 2019, 45(4): 296-301.
- [8] NIE D, WANG X W, SHI L C, et al. EEG-based emotion recognition during watching movies[C]//Proceedings of the 2011 5th International IEEE/EMBS Conference on Neural Engineering. Piscataway: IEEE Press, 2011: 667-670.
- [9] YANG Y, WU Q, QIU M, et al. Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network[C]//Proceedings of the 2018 International Joint Conference on Neural Networks. Piscataway: IEEE Press, 2018: 1-7.
- [10] LIPTON Z C, BERKOWITZ J, ELKAN C. A critical review of recurrent neural networks for sequence learning[J]. arXiv preprint, 2015, arXiv:1506.00019.
- [11] WALECKI R, PAVLOVIC V, SCHULLER B, et al. Deep structured learning for facial action unit intensity estimation[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 3405-3414.
- [12] LI S, DENG W. Deep facial expression recognition: a survey[J]. IEEE Transactions on Affective Computing, 2018.
- [13] BREUER R, KIMMEL R. A deep learning perspective on the origin of facial expressions[J]. arXiv preprint, 2017, arXiv:1705.01842.
- [14] MOLLAHOSSEINI A, CHAN D, MAHOOR M H. Going deeper in facial expression recognition using deep neural networks[C]//Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision. Piscataway: IEEE Press, 2016: 1-10.

- [13] 王云. 基于 EEG 与计算机视觉的人脸检索闭环系统[D]. 杭州: 浙江大学, 2016.
WANG Y. Face retrieval closed-loop system based on EEG and computer vision[D]. Hangzhou: Zhejiang University, 2016.
- [14] LUO Y J, HUANG Y X, LI X Y, et al. Effects of emotion on cognitive processing: series of event-related potentials study[J]. Advances in Psychological Science, 2006, 14(4): 505.
- [15] KHALILI Z, MORADI M H. Emotion recognition system using brain and peripheral signals: using correlation dimension to improve the results of EEG[C]//Proceedings of the 2009 International Joint Conference on Neural Networks. Piscataway: IEEE Press, 2009: 1571-1575.
- [16] ZOURIDAKIS G, PATIDAR U, PADHYE N S, et al. Spectral power of brain activity associated with emotion—a pilot MEG study[C]//Proceedings of the 17th International Conference on Biomagnetism Advances in Biomagnetism-Biomag 2010. Berlin: Springer, 2010: 354-357.
- [17] LIN Y P, WANG C H, JUNG T P, et al. EEG-based emotion recognition in music listening[J]. IEEE Transactions on Biomedical Engineering, 2010, 57(7): 1798-1806.
- [18] LI X, SONG D W, ZHANG P, et al. Emotion recognition from multi-channel EEG data through convolutional recurrent neural network[C]//Proceedings of the 2016 IEEE International Conference on Bioinformatics and Biomedicine. Piscataway: IEEE Press, 2016: 352-359.
- [19] SCHMIDHUBER J, HOCHREITER S. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [20] CHO K, VAN MERRIËNBOER B, BAHDANAU D, et al. On the properties of neural machine translation: encoder-decoder approaches[J]. arXiv preprint, 2014, arXiv:1409.1259.
- [21] LONG M, CAO Y, WANG J, et al. Learning transferable features with deep adaptation networks[C]//Proceedings of the 32nd International Conference on Machine Learning. [S.l.:s.n.], 2015: 97-105.
- [22] GRETTON A, BORGWARDT K M, RASCH M J, et al. A kernel two-sample test[J]. The Journal of Machine Learning Research, 2012, 13(1): 723-773.
- [23] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.
- [24] LING W F, KONG W Z, LONG Y F, et al. Facial emotion recognition based on brain and machine collaborative intelligence[C]//Proceedings of the 2019 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications. Piscataway: IEEE Press, 2019: 1-5.

[作者简介]



刘栋军(1997-), 男, 杭州电子科技大学计算机学院硕士生, 主要研究方向为脑机协同智能、情绪识别。



王宇涵(1997-), 女, 杭州电子科技大学计算机学院硕士生, 主要研究方向为脑机协同智能、情绪识别。



凌文芬(1995-), 女, 杭州电子科技大学计算机学院硕士生, 主要研究方向为情绪生理信号、特征提取、情绪识别。



彭勇(1985-), 男, 博士, 杭州电子科技大学计算机学院副教授, 主要研究方向为机器学习、模式识别与脑机交互算法及应用。



孔万增(1980-), 男, 博士, 杭州电子科技大学计算机学院认知与智能计算研究所教授、博士生导师, 杭州电子科技大学研究生院副院长、浙江省脑机协同智能重点实验室主任, 主要研究方向为人工智能与模式识别、嵌入式可穿戴计算、脑机交互与认知计算等。