

基于3D分层卷积融合的多模态生理信号情绪识别

凌文芬^{1,2}, 陈思含^{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:近年来, 脑电等生理信号由于能客观体现真实情绪已逐渐成为情绪识别研究的热门对象。然而, 单模态的脑电信号存在情绪信息特征不完备问题, 多模态生理信号存在情绪信息交互不充分问题。针对这些问题, 提出基于3D分层卷积的多模态特征融合模型, 旨在充分挖掘多模态交互关系, 更准确地刻画情感信息。首先通过深度可分离卷积网络提取脑电、眼电和肌电3种模态的生理信号的多模态初级情绪特征信息, 再对得到的多模态初级情绪特征信息进行3D卷积融合操作, 实现两两模态间的局部交互以及所有模态间的全局交互, 获取包含不同生理信号情绪特征的多模态融合特征。实验结果表明, 提出的模型在DEAP数据集的效价、唤醒度的二分类和四分类任务中达到了98%的平均准确率。

关键词: 生理信号; 情绪识别; 3D分层卷积; 多模态交互

中图分类号: TP18, TN911.7, R318

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202108

Multi-modal physiological signal emotion recognition based on 3D hierarchical convolution fusion

LING Wenfen^{1,2}, CHEN Sihan^{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: In recent years, physiological signals such as electroencephalography (EEG) have gradually become popular objects of emotion recognition research because they can objectively reflect true emotions. However, the single-modal EEG signal has the problem of incomplete emotional information representation, and the multi-modal physiological signal has the problem of insufficient emotional information interaction. Therefore, a 3D hierarchical convolutional fusion model was proposed, which aimed to fully explore multi-modal interaction relationships and more accurately describe emotional information. The method first extracted the primary emotional representation information of EEG, electro-oculogram (EOG) and electromyography (EMG) by depthwise separable convolution network, and then performed 3D convolution fusion operation on the obtained multi-modal primary emotional representation information to realize the pairwise mode local interactions between states and global interactions among all modalities, so as to obtain multi-modal fusion representations containing emotional characteristics of different physiological signals. The results show that the accuracy in the valence and arousal of the two-class and four-class tasks on DEAP dataset are both 98% by the proposed model.

Key words: physiological signal, emotion recognition, 3D hierarchical convolutional, multi-modal interaction

收稿日期: 2021-01-09; 修回日期: 2021-02-05

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.U1909202, No.U20B2074, 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.U1909202, No.U20B2074, 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.2020E10010)

1 引言

随着人工智能技术的发展和互联网的普及，人们对人机情感交互的需求在生产生活中的各个领域中日显。其中，如何准确地识别情绪一直是热门研究方向。生理信号是对人体中枢神经系统和自主神经系统的反应，这些信号的变化能体现不同的情绪。与其他情感载体（如姿势、面部表情或语音等）能刻意伪装不同，生理信号这种情绪体现方式最大的特点在于中枢神经系统和自主神经系统在很大程度上是非自愿激活的，因此不容易控制，更具客观真实性。因此，越来越多的国内外专家学者投入基于生理信号的情绪识别研究领域。特别是随着脑科学发展，以脑电图（electroencephalography, EEG）为主要媒介的情绪识别研究越来越受到关注，且相关的研究成果已被应用到许多领域，如人机交互系统^[1-2]、安全驾驶^[3]、医疗保健，尤其是心理健康监测^[4]和游戏娱乐^[5]等。在生理信号中，除常见的脑电图外，心电图（electrocardiogram, ECG）、眼电图（electro-oculogram, EOG）、体温（body temperature）、肌电图（electromyogram, EMG）、皮肤电反应（galvanic skin response, GSR）和呼吸（respiration, RESP）^[6]等外围生理信号也可由外界刺激诱发产生，并携带情绪信息。例如，强烈的情绪会刺激人体的交感神经系统，导致汗腺分泌更多的汗液，皮肤阻抗变小，GSR发生变化，同时伴随着相关的面部表情变化，EOG和EMG也随之有着相应的反应。

因此，随着研究的推进，也有一些学者尝试通过将多种生理信号融合的深度学习方法来提高情绪识别的准确率。传统的基于脑电情绪的研究通常需要人工提取特征并选择适合分类器来分类。随着自动学习数据分布和特征学习融入建模过程中的优势被发掘，深度学习不仅在机器视觉领域取得了巨大突破性进展，其他领域也逐渐运用深度学习的方法来解决一些问题。基于深度学习的研究方法已在脑电信号的情绪识别研究中取得了较大的成功，并已逐渐成为研究生理信号情绪识别的主流方法之一。

根据融合的时间先后，多模态融合可以被分为早期信号层融合、中期特征层融合和后期决策层融合3个阶段，如图1所示。

早期信号层融合面向信号本身，将多模态信号经过简单预处理后放入模型，提取多模态特征，通常用于时间上高度同步的信号间的融合，如视频和音频信号的融合^[7]。

中期特征层融合面向特征，通常是在得到单模态特征以后，通过各种手段实现模态特征间的交互，从而实现多模态信息之间的互补增强。Hassan M M等人^[8]首先利用深度信念网络（deep belief net, DBN）从皮肤电活性（electro-dermal activity, EDA）、光电容积描记图（photoplethysmogram, PPG）和颧肌肌电图（zygomaticus electromyography, zEMG）的融合观测中提取深度级别特征，然后获取EDA、PPG和zEMG的统计特征，并与深度信念网络特征结合构成特征融合向量，最后将准备好的特征向量

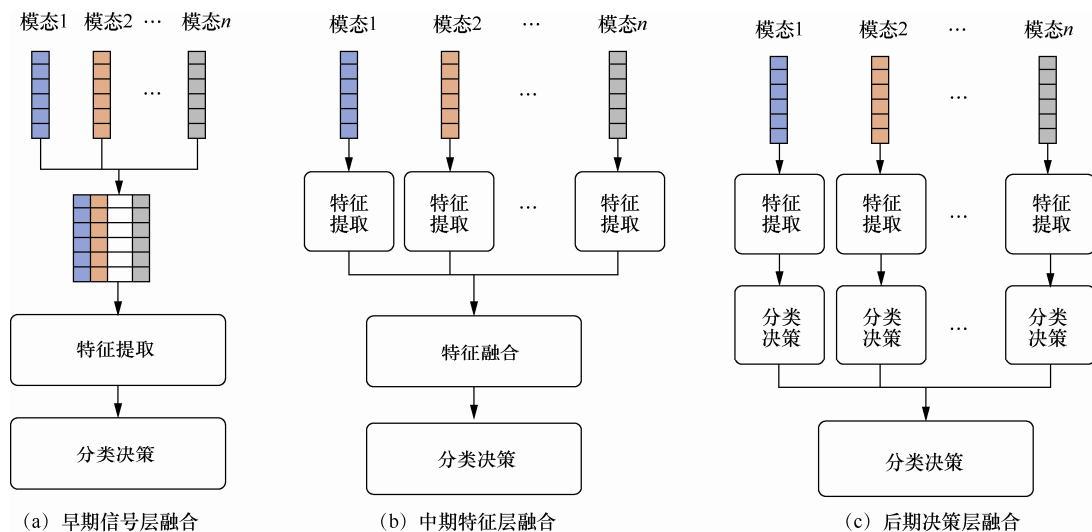


图1 特征融合阶段示意图

用于对5种基本情绪的分类。Kwon Y H等人^[9]提出一种卷积神经网络(convolutional neural network, CNN)结构,提取脑电信号情绪特征,并和GSR特征结合构成多模态情绪特征,用于情绪识别。Liao J X等人^[10]提出用卷积递归神经网络的方法提取多模态生理信号的情绪信息。该方法首先根据空间位置对脑电通道进行二维表示,然后和脑电时序信息一起输入卷积神经网络,提取脑电信号情绪信息,接着利用长短期记忆(long short term memory, LSTM)网络提取外围生理信号情绪信息,最后将两种特征同维拼接,用于最后的分类。Qiu J L等人^[11]采用深度相关分析(deep canonical correlation analysis, DCCA)共同学习两个模态的非线性变换的参数,以获取脑电信号特征和眼动信号特征的最大相关性,使得提取的两种特征更具相关性,最终提高融合后特征的分类性能。

后期决策层融合面向分类结果,利用不同模态得到的分类结果,综合考虑多个分类器的结果,在决策阶段实现多模态特征的协同作用。Zhao Y X等人^[12]根据多通道脑电数据的通道空间位置,将输入脑电预处理成三维数据,即二维空间位置信息和一维时序信息,采用三维卷积构建脑电信号情绪特征的提取模型。对于其他生理信号来说,该方法用一维卷积构建情绪特征提取模型,并在决策层融合多模态特征的情绪特征。Huang H P等人^[13]提出一种集成卷积神经网络模型,他们首先设计5个卷积网络,然后采用多元投票策略建立集成模型,最终识别情绪类别。

在多模态融合的3个阶段中,早期信号层融合模型的模态交互和特征提取同时进行,能较充分地考虑不同模态间的交互信息,且在一定程度上简化了模型,但是由于处理的是早期信号数据,存在较多冗余信息的干扰,因此在此类模型中,较难提取有效的模态特征。中期特征层融合模型由于输入为

单独提取的模态特征,减少了信息冗余带来的干扰。同时,通过模型实现不同模态间的局部交互,从而得到多种模态信息的融合特征,用于最后的决策。由于特征层面的融合模型在多模态交互融合中的优势,许多研究者大多选择特征层面的多模态融合模型。后期决策层融合模型只在决策时考虑模态间的相关性,没有充分利用不同模态特征间的局部交互信息。在上述特征层面的融合方法中,不同模态的特征通常进行同维拼接,即结合后的多模态特征维度并没有增加并且通常是一维特征,在后续的融合中不利于模态间的局部交互。因此,本文提出基于3D分层卷积的多模态特征融合模型,将提取的多模态情绪特征进行增维拼接后,利用3D分层卷积,获取不同模态间的两两局部交互信息和全部模态间的全局交互信息,得到融合多种模态情绪信息的多模态融合情绪特征,并用于分类。

2 基于3D分层卷积融合的多模态生理信号情绪识别方法

本文提出基于3D分层卷积的多模态特征融合模型,用于对多模态生理信号进行情绪识别。首先将选取的多模态生理信号经过一定的预处理后,分别放入深度可分离卷积(depthwise separable convolution)网络模型中,提取模态内的初级情绪特征,然后将多个模态情绪特征进行增维拼接,放入3D分层卷积融合模型中,实现两两模态间的局部交互以及所有模态间的全局交互,获取包含不同生理信号情绪特征的多模态融合特征,最后用于情绪的识别和分类。多模态生理信号的情绪识别模型示意图如图2所示。

2.1 基于深度可分离卷积网络的特征提取模型

生理信号数据采集条件苛刻、成本高,在数据量级上,不能与图像数据相提并论。然而卷积神经

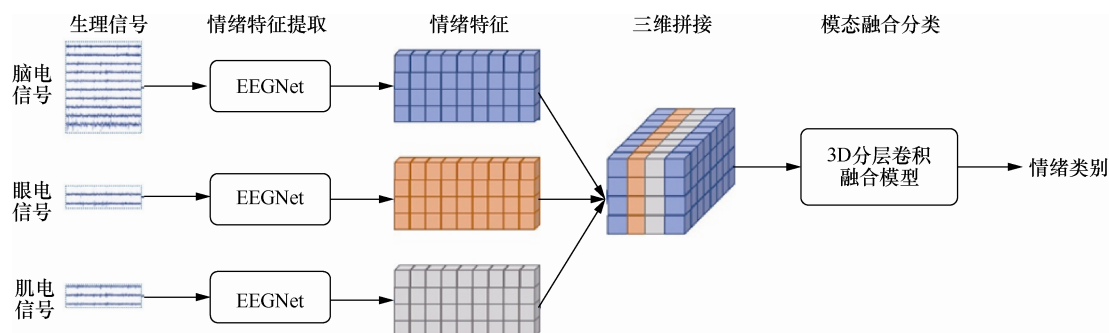


图2 多模态生理信号的情绪识别模型示意图

网络是基于数据驱动的，得到一个可靠模型需要训练大量的数据，当遇到样本量较小的数据集时，模型容易过度拟合训练集数据特征，泛化性较差，使得训练集和测试集在模型上的表现差异较大，产生过拟合现象。因此，适用于原始生理信号数据情绪特征的卷积网络模型应该是能获取生理信号的时序特征和空间特征，同时参数量少，不易出现过拟合的模型。

在常规卷积操作中，若输入数据是多通道的，如 $7 \times 7 \times 3$ 的尺寸，其中 7×7 为一个通道的数据尺寸，该输入数据有 3 个通道，则卷积核的尺寸可以为 $5 \times 5 \times 3$ ，其中卷积核的通道数和输入数据的通道数一致。卷积核的每个通道分别与输入数据的相应通道进行卷积操作，最后将不同通道在相同位置的值求和，得到一个特征图。所需参数量 $P = 5 \times 5 \times 3 \times t$ ，其中 t 为卷积后输出的特征图个数，即经过一次卷积后得到的通道数。普通卷积过程如图 3 所示。

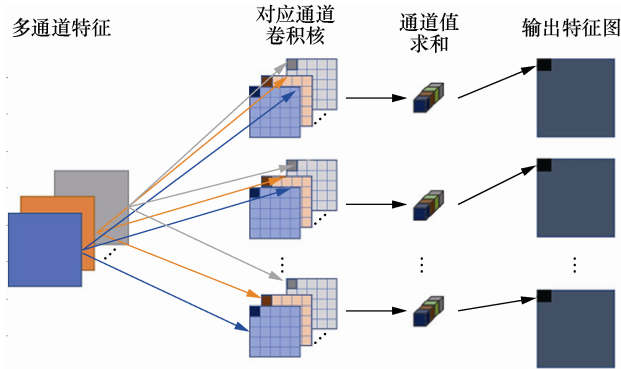


图3 普通卷积过程

深度可分离卷积是轻量级网络 MobileNets^[14]和 Xception^[15]中的精华部分。深度可分离卷积由深度卷积（depthwise convolution）和点卷积（pointwise convolution）构成，如图 4 所示。深度卷积又被称为逐通道卷积，依然以 $7 \times 7 \times 3$ 的多通道输入数据为例，逐通道卷积的卷积核为 $5 \times 5 \times 3$ ，卷积核的每个通道分别与输入数据的相应通道进行卷积操作，不同通道在相同位置的值不做求和处理，因此逐通道卷积输出的特征图数量和输入数据的通道相同，且每个特征图仅包含独立通道的信息，与其他通道没有信息交互。所需的参数量 $P_d = 5 \times 5 \times 3$ 。深度可分离卷积是在深度卷积后连接点卷积实现的。点卷积的卷积核为 $1 \times 1 \times 3$ ，其中 3 为深度卷积输出的特征通道数，在不同的特征图中，实现相同位置不同通道间的信息交互，弥补深度卷积没有通道信息交互的

缺点，若指定进行点卷积后输出 t 个特征图，则点卷积所需的参数量为 $P_p = 1 \times 1 \times 3 \times t$ ，最后输出的特征尺寸和常规卷积相同，最终深度可分离卷积的参数量 $P_{all} = P_d + P_p = 5 \times 5 \times 3 + 1 \times 1 \times 3 \times t$ ，在大部分应用场景下比常规卷积的 $5 \times 5 \times 3 \times t$ 参数量要少得多。

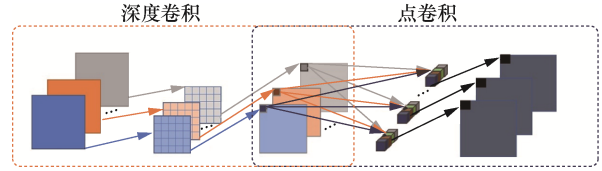


图4 深度可分离卷积

因此在提取多通道生理信号数据的情绪特征时，考虑到生理信号样本数量级小等现实情况，选择由深度可分离卷积网络构成的生理信号情绪特征提取模型 EEGNet^[16]为本文生理信号情绪特征的提取模型。模型如图 5 所示，整个模型分为两大块，3 个小部分。

第一块为普通卷积操作，目的是提取生理信号的时序特征。采用大小为 $1 \times k$ 的卷积核在每个通道上分别对不同时刻的数据进行卷积操作，卷积层的步长为 1，得到每个通道上对应的时序特征，输出的特征图大小为 $\text{channels} \times \text{samples} \times N$ ，其中 channels 为输入生理信号的导联数， samples 为输入信号的样本点数， N 为最后输出的特征的通道数， k 设置为 samples 的 $1/2$ 。

第二块为深度可分离卷积，分为深度卷积和可分离卷积两部分，主要获取生理信号空间和特征通道间的交互信息。深度卷积部分，对逐个样本点的所有导联进行卷积操作，卷积核大小为 $\text{channels} \times 1$ ，步长为 1，获取多导联信号之间的空间相关特征，对每个通道进行 M 次深度卷积，输出得到 M 个 $1 \times \text{samples} \times N$ 的深度卷积特征。可分离卷积部分，在深度卷积的基础上进行逐个样本点通道卷积，卷积核为 $1 \times 1 \times S$ ，其中 S 为 M 和 N 的乘积，步长为 1，获取所有特征图之间的交互特征，输出 $1 \times \text{samples} \times S$ 的特征图。

在模型中每层卷积后添加批归一化层 batchnorm ，使得每一次输入新的网络层中的数据具有相同的分布。在深度卷积和可分离卷积后分别添加平均池化层和随机失活（dropout）层，抽取特征和防止过拟合，使网络更具鲁棒性和泛化能力。

本文将可分离卷积的输出特征作为每个模态的情绪特征，用于后续的多模态融合模型中。

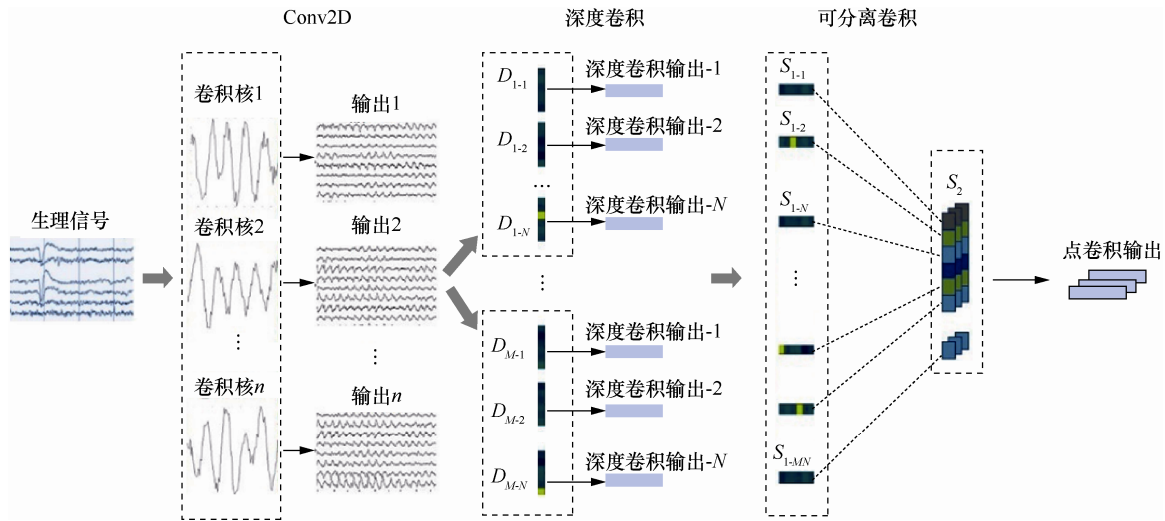


图5 EEGNet 模型的网路结构

2.2 基于 3D 分层卷积的多模态特征融合模型

通过深度可分离卷积模型, 本文可以获得不同模态的生理信号的情绪特征。然而, 在情绪诱发过程中, 单一的生理信号体现的情绪信息不是很完备, 且单模态受采集环境及被试个体差异的影响较大, 单模态生理信号的情绪特征鲁棒性不高, 因此, 本文构建了一个深度为 4 层的基于 3D 分层卷积的多模态特征融合模型, 如图 6 所示。

第一层为输入层: 由深度可分离卷积模型可得到多个模态间相互独立的情绪特征, 且每个形状都为 $1 \times \text{dim} \times F$, 其中, dim 为特征的维度, F 为特征的通道数。在输入模型前, 将不同的模态特征在第 0 维进行拼接, 形成 $I \times 1 \times \text{dim} \times F$ 的多模态三维特征矩阵, 通过维度转换, 最终以 $I \times \text{dim} \times F \times 1$ 的形状输入模型, 其中 I 为拼接的模态数量, 为了让每个模态都能两两交互, 本文设 I 为 4。

第一个 3D 卷积层: 由于输入是三维特征矩阵, 且 3 个维度之间存在相关性, 因此, 本文提出用 3D

卷积实现模态之间的交互。设卷积核为 $2 \times 2 \times 3$, 步长为 $1 \times 1 \times 2$, 隐藏神经元节点数为 H_1 , 每个卷积核能同时感知不同模态、不同时刻和不同特征图 3 个相关维度的信息, 通过卷积过程, 实现 3 个维度信息之间的两两交互。

第二个 3D 卷积层: 经过第一层卷积后, 得到初步的低层次融合特征。本层设卷积核为 $3 \times 3 \times 3$, 步长为 $1 \times 1 \times 2$, 隐藏神经元节点数为 H_2 , 在低层次的局部交互信息基础上, 进一步卷积, 获取包含低层次局部信息交互信息和高层次全局抽象信息的融合特征。

输出层: 将前一层得到的全局融合特征进行扁平化处理, 然后进行分类, 最后输出情绪类别。

3 实验设置与结果分析

3.1 数据集

本文将情绪生理信号数据集 DEAP^[17] 作为实验数据。该数据集采集了 32 个被试在观看 40 段每段

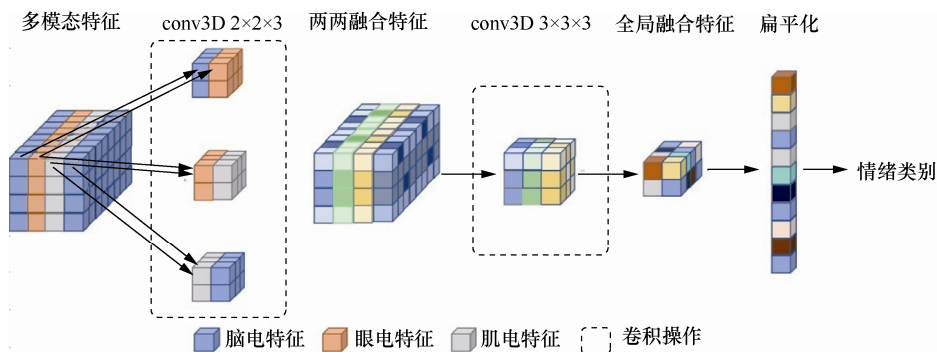


图6 基于 3D 分层卷积的多模态特征融合模型

1 min 的音乐视频后的 EEG、EOG、EMG、ECG、GSR 等生理信号。该数据集用根据 10~20 国际标准导联设计的脑电帽采集脑电信号, 这些脑电信号包含了 32 个头皮电极导联, 采样率为 512 Hz。数据集采用二维情绪模型判定情绪类别, 被试在观看完一段音乐视频后, 要求在唤醒度、效价、支配度和喜欢程度 4 个维度上, 在 1~9 范围内标出对视频的情绪感受。

数据集预处理版本包含 32 通道的脑电数据和 8 通道的外围生理信号数据。统一进行 128 Hz 的降采样处理。每个通道的数据包含 3 s 的静息态基准线信号和 60 s 的生理信号数据, 对应一个 1 min 的音乐视频。根据实验需要, 本文选取 32 通道的脑电数据, 2 通道的肌电数据和 2 通道的眼电数据作为多模态生理信号。对于所选的脑电数据, 每个通道除去静息态后, 共 $8\ 064-128\times 3=7\ 680$ 个样本点, 每个视频按照 1 s 时长进行分割后, 将同一个被试的数据按照样本量 $\times$ 通道 $\times$ 样本的格式进行拼接, 得到 $2\ 400\times 40\times 128$ 形式的生理信号数据。其中, 为弱化设备本身对信号的干扰, 每个通道的数据分别减去了前 3 s 的静息态数据的均值^[18]。在验证模型前, 数据在单个被试内进行乱序处理。

情绪评价维度主要选取唤醒度和效价两个维度, 研究情绪二分类和四分类任务, 以 5 为阈值, 将小于或等于 5 的评分设为低水平, 5 以上的评分设为高水平。其中, 二分类依次为低唤醒度 (low arousal, LA)、高唤醒度 (high arousal, HA) 和低效价 (low valence, LV)、高效价 (high valence, HV); 四分类任务由效价-唤醒度共同构成, 依次为高唤醒度高效价 (high arousal high valence, HAHV)、低唤醒度高效价 (low arousal high valence, LAHV)、低唤醒度低效价 (low arousal low valence, LALV)、高唤醒度低效价 (high arousal low valence, HALV)。

3.2 网络训练参数

本文经多次试验后选择效果最优的超参数, 其中学习率为 0.001, 优化器为 Adam, 样本批次为 128, dropout 系数为 0.5, 模型中网络层的神经元节点数见表 1。

表 1 模型中网络层的神经元节点数

模型	网络层
特征提取模型	$N=8$ 、 $M=2$ 、 $S=16$
基于 3D 分层卷积的多模态特征融合模型	$H_1=32$ 、 $H_2=64$

3.3 结果与分析

本文在 DEAP 数据集上选取脑电信号、眼电信号和肌电信号作为多模态生理信号的组合, 验证提出基于 3D 分层卷积的多模态特征融合模型。为探究多模态生理信号情绪特征融合的有效性, 本文在效价、唤醒度二分类任务和效价-唤醒度四分类任务上分别对多模态生理信号融合情绪特征和单模态脑电信号情绪特征的分类性能做对比实验, 所有被试的情绪特征在不同分类任务中的平均分类准确率见表 2。从表 2 可以看出, 单模态脑电信号情绪特征在二分类和四分类任务中的平均分类准确率分别为 94.02%、93.39%和 89.15%, 而多模态生理信号融合情绪特征在对应任务中的平均分类准确率分别为 98.31%、98.47%和 98.21%。脑电、眼电和肌电 3 种情绪特征经过 3D 分层卷积融合后的多模态生理信号融合情绪特征在二分类任务和四分类任务上的平均分类准确率都优于单模态脑电信号情绪特征, 在二分类任务中提高了 4%~5%, 在四分类任务中提高了近 10%。由此, 验证了本文提出的模型的有效性。

表 2 单模态脑电信号情绪特征和多模态生理信号融合情绪特征在不同分类任务中的平均分类准确率和标准差

特征	效价	唤醒度	效价-唤醒度
单模态脑电信号情绪特征	$0.940\ 2\pm 0.057\ 1$	$0.933\ 9\pm 0.085\ 8$	$0.891\ 5\pm 0.059\ 5$
多模态生理信号融合情绪特征	$0.983\ 1\pm 0.014\ 5$	$0.984\ 7\pm 0.017\ 8$	$0.982\ 1\pm 0.016\ 8$

此外, 本文还观察了不同被试的生理信号特征在情绪识别上的差异, 以每个被试的测试结果在效价、唤醒度和效价-唤醒度上的分类准确率为观测指标。如图 7、图 8 和图 9 所示, 在相同任务的不同被试的生理信号特征比较中, 不同被试的单模态脑电信号情绪特征分类性能差异较大。而不同被试的多模态生理信号融合情绪特征的分类性能相对稳定, 且都具有较高的分类准确率; 在被试单模态脑电信号情绪特征表示情绪不理想的情况下, 如被试 11、被试 17 或被试 22 所示, 多模态生理信号融合情绪特征依然能保持优越的情绪表示能力。在不同任务的相同被试的生理信号特征比较中, 多模态生理信号融合情绪特征在四分类任务中的分类性能提高优势比在二分类任务中更加明显。

为验证本文提出的增维拼接后的分层卷积融合的有效性, 笔者构建了同等条件下普通拼接融合的对比如实验。普通拼接融合对比实验在得到 3 种模

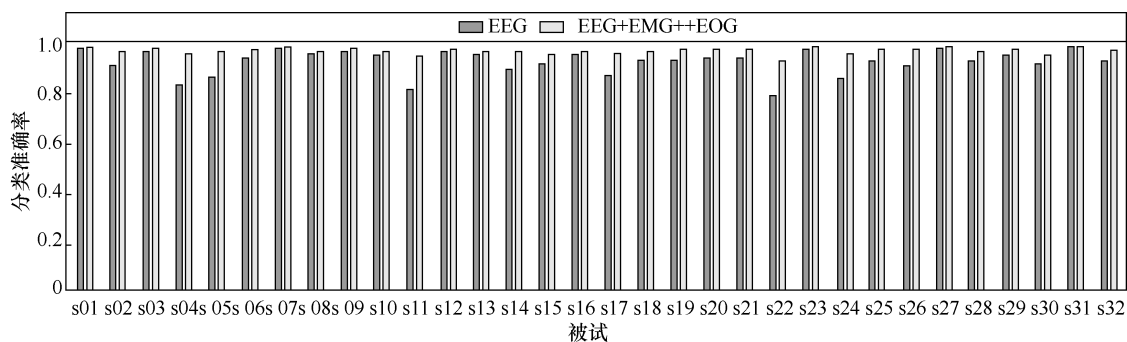


图7 在效价维度上不同被试的单模态脑电信号情绪特征和多模态生理信号融合情绪特征的分类准确率

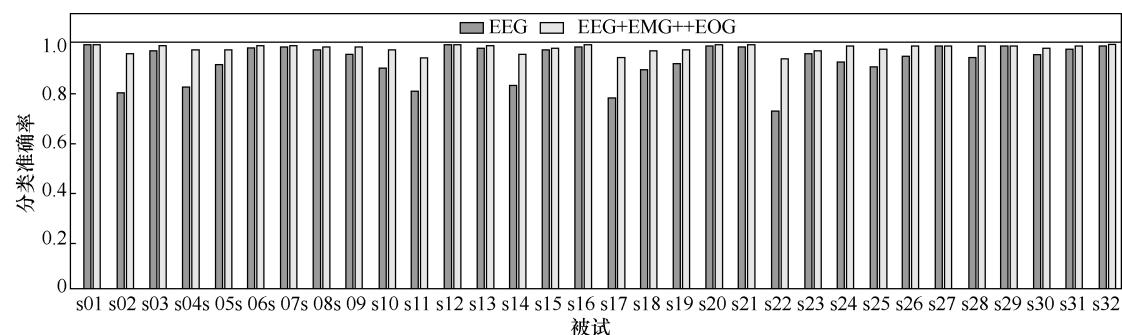


图8 在唤醒度维度上不同被试的单模态脑电信号情绪特征和多模态生理信号融合情绪特征的分类准确率

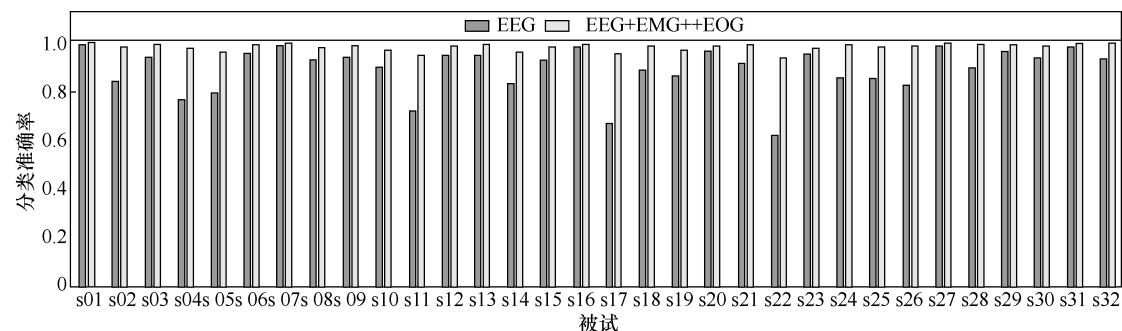


图9 在效价-唤醒度四分类任务上不同被试的单模态脑电信号情绪特征和多模态生理信号情绪融合特征的分类准确率

态的情绪特征后,在原有的维度上对特征进行同维拼接(即每种模态的情绪特征是一维的且依次在同维度合并)后,输入全连接层直接用于分类,得到普通拼接融合的分类结果。由图10的对比结果可知,本文提出的使用特征增维拼接的基于3D分层卷积的多模态特征融合模型(即图10中的3D分层卷积融合)在二分类任务和四分类任务上都比普通拼接的融合方法具有更高的分类准确率。

在研究不同融合时期对融合结果的影响上,本文设置早期融合模型作为对照组。为控制变量排除其他因素对实验结果的干扰,选用生理信号的情绪特征提取模型EEGNet作为早期多模态融合模型,将3种生理信号在通道维度拼接后,作为早期融合模型的输入数据,对比结果如图11所示。从图11可知,与早期

融合方法相比,无论是在效价和唤醒度的二分类任务上还是在效价-唤醒度构成的四分类任务上,本文提出的在中期进行融合的基于3D分层卷积的多模态特征融合方法(即图11中的3D分层卷积融合)的情绪分类性能都远高于早期融合方法。

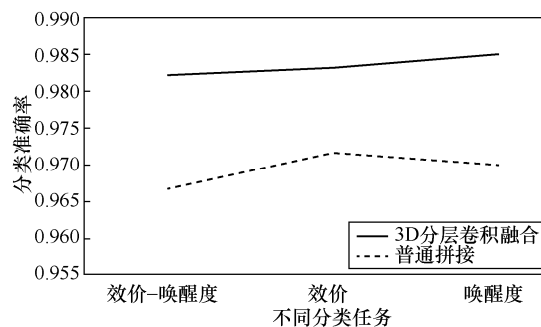


图10 多模态特征的不同融合方法的对比结果

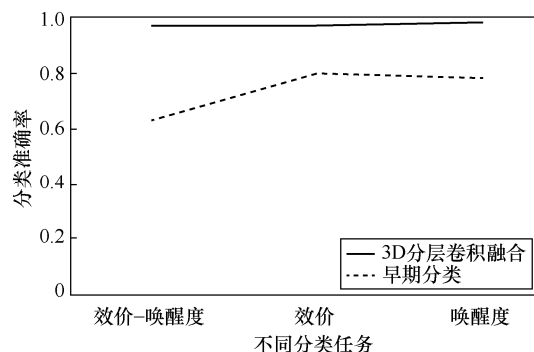


图11 多模态特征不同融合阶段的对比结果

除此之外,本文还将提出的模型和其他现有的多模态特征融合方法进行对比分析,本文方法的十折交叉验证结果见表3。由表3可知,本文提出的方法比现有的多模态特征融合方法具有更优越的分类性能,且分类准确率远高于现有方法。

4 结束语

针对单模态脑电信号包含的情绪信息不完备这一问题,本文提出在特征层面的3D分层卷积融合模型。模型首先用EEGNet分别提取脑电信号、眼电信号和肌电信号的初级情绪特征,在此基础上,对3种模态的情绪特征增维拼接,然后将拼接后的多模态特征放入3D分层卷积融合模型中,实现两两模态间的局部交互以及所有模态间的全局交互,获取包含不同生理信号情绪特征的多模态融合特征,用于情绪分类。

在DEAP数据集上初步验证了本文模型的有效性,本文模型在唤醒度的二分类任务中得到了98.47%的分类准确率,在效价的二分类任务中得到了98.31%的分类准确率,在效价-唤醒度的四分类任务中也得到了98.21%的分类准确率。在和不同对照组的实验对比中,可以得出以下结论。

① 多模态情绪特征经过特征层面的融合后得到的多模态生理信号融合情绪特征在情绪表示充分性上优于单模态脑电信号情绪特征,在不同的情

绪分类任务上都优于单模态脑电信号情绪特征。

② 本文提出的增维拼接方法和基于3D分层卷积的多模态特征融合模型提取的生理信号情绪特征融合信息能较为充分地表示情绪信息,弥补了单模态脑电信号情绪表示不充分问题,一定程度上弱化了被试间的情绪识别差异。

目前在本文的研究中,所用到的3种生理信号情绪特征的提取模型相同。然而,肌电和眼电这类外围生理信号和脑电信号有所不同,在进一步的工作中,笔者将结合外围生理信号特性构建更适合提取外围生理信号情绪特征的模型,从而在这些生理信号中提取更加精准的情绪特征,有利于在后续的多模态特征融合模型中进行更加一致的情绪特征交互,挖掘更高层次的复杂丰富的多模态特征融合信息。

参考文献:

- [1] 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.
- [2] BARTNECK C, LYONS M J. HCI and the face: towards an art of the soluble[C]//Proceedings of the 12th International Conference on Human-Computer Interaction. Berlin: Springer, 2007: 20-29.
- [3] DE NADAI S, D'INCÀ M, PARODI F, et al. Enhancing safety of transport by road by on-line monitoring of driver emotions[C]//Proceedings of the 2016 11th System of Systems Engineering Conference. Piscataway: IEEE Press, 2016: 1-4.
- [4] GUO R, LI S J, HE L, et al. Pervasive and unobtrusive emotion sensing for human mental health[C]//Proceedings of the 2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops. Piscataway: IEEE Press, 2013: 436-439.
- [5] ZHAN C, LI W, OGUNBONA P, et al. A real-time facial expression recognition system for online games[J]. International Journal of Computer Games Technology, 2008: 542918.
- [6] LIN S, XIE J Y, YANG M Y, et al. A review of emotion recognition using physiological signals[J]. Sensors, 2018, 18(7): 2074.
- [7] SNOEK C G M, WORRING M, SMEULDERS A W M. Early versus late fusion in semantic video analysis[C]//Proceedings of the 13th annual ACM International Conference on Multimedia. New York: ACM

表3 不同多模态特征融合方法在DEAP数据集上的分类准确率

方法	模态	效价	唤醒度	效价-唤醒度
CNN ^[19]	EEG+全部外围生理信号	85.5	87.3	—
DCCA ^[11]	EEG+全部外围生理信号	85.62	84.33	85.51
MM-ResLSTM ^[20]	EEG+全部外围生理信号	92.3±1.55	92.87±2.11	—
CNN+LSTM ^[10]	EEG+全部外围生理信号	91.95	93.06	—
本文方法	EEG+EOG+EMG	98.55±1.33	98.77±1.57	98.67±1.38

Press, 2005: 399-402.

- [8] HASSAN M M, ALAM M G R, UDDIN M Z, et al. Human emotion recognition using deep belief network architecture[J]. Information Fusion, 2019, 51: 10-18.
- [9] KWON Y H, SHIN S B, KIM S D. Electroencephalography based fusion two-dimensional (2D)-convolution neural networks (CNN) model for emotion recognition system[J]. Sensors, 2018, 18(5): 1383.
- [10] LIAO J X, ZHONG Q H, ZHU Y S, et al. Multimodal physiological signal emotion recognition based on convolutional recurrent neural network[J]. IOP Conference Series Materials Science and Engineering, 2020, 782: 032005.
- [11] QIU L, LIU W, LYU B L. Multi-view emotion recognition using deep canonical correlation analysis[C]//Proceedings of the 25th International Conference on Neural Information Processing. Cham: Springer, 2018: 221-231.
- [12] ZHAO Y X, CAO X Y, LIN J L, et al. Multimodal emotion recognition model using physiological signals[J]. arXiv preprint, 2019, arXiv:1911.12918.
- [13] HUANG H P, HU Z C, WANG W M, et al. Multimodal emotion recognition based on ensemble convolutional neural network[J]. IEEE Access, 2019, 8: 3265-3271.
- [14] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint, 2017, arXiv:1704.04861.
- [15] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 1251-1258.
- [16] LAWHERN V J, SOLON A J, WAYTOWICH N R, et al. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces[J]. Journal of Neural Engineering, 2018, 15(5): 056013.
- [17] KOELSTRA S, MUHL C, SOLEYMANI M, et al. DEAP: a database for emotion analysis; using physiological signals[J]. IEEE Transactions on Affective Computing, 2011, 3(1): 18-31.
- [18] YANG Y L, WU Q F, 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.
- [19] LIN W Q, LI C, SUN S Q. Deep convolutional neural network for emotion recognition using EEG and peripheral physiological signal[C]//Proceedings of the International Conference on Image and

Graphics. Cham: Springer, 2017: 385-394.

- [20] MA J X, TANG H, ZHENG W L, et al. Emotion recognition using multimodal residual LSTM network[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM Press, 2019: 176-183.

[作者简介]



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。



陈思含 (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。