

曲阜师范大学

本科生毕业论文（设计）



题 目 基于深度学习模型的脑电情绪认知算法研究

姓 名 戚嘉禾 学号 2019414042

院 系 网络空间安全学院

专 业 软件工程

指导教师 高 鹏 职 称 副教授

2023 年 5 月 20 日

曲阜师范大学教务处制

目 录

摘要	1
关键词	1
Abstract	1
Key words	1
1 绪论	2
1.1 选题背景及研究意义	2
1.2 研究现状	2
1.3 主要研究内容	3
1.4 论文组织结构	3
2 相关背景知识介绍	3
2.1 情绪的相关背景知识	3
2.1.1 情绪的定义	3
2.1.2 情绪的分类	3
2.1.3 情绪的诱发	4
2.2 脑电的相关背景知识	4
2.2.1 脑电信号介绍	4
2.2.2 脑电信号的预处理	4
2.2.3 脑电信号的特征提取	5
2.3 深度学习相关背景知识	6
2.3.1 深度学习简介	6
2.3.2 人工神经网络	6
2.4 基于深度学习的脑电情绪认知	6
2.5 实验数据集	6
3 基于三维特征卷积循环神经网络的脑电情绪认知算法	7
3.1 方案架构设计	7
3.2 数据预处理和特征提取	7
3.3 三维结构时间切片输入的卷积循环神经网络	8
3.3.1 三维结构的时间切片构造	8
3.3.2 频率和空间特征学习	9
3.3.3 时间特征学习	9
3.3.4 分类器	9
3.3.5 基于 EEG 的情绪识别架构	10
4 实验设置与结果分析	10
4.1 数据训练环境与测试环境	10
4.2 训练集和测试集	11
4.3 实验设置及模型训练	11
4.4 实验结果	11
4.5 实验比较	13
5 总结与展望	14

5.1 总结 14

5.2 展望 14

致谢 14

参考文献 14

基于深度学习模型的脑电情绪认知算法研究

软件工程专业学生 戚嘉禾

指导老师 高 鹏

摘要：情绪作为人脑的高级功能，在人际交往和医学研究中发挥着重要作用。情绪认知是情绪相关研究中非常重要的一个环节，其中基于脑电信号和深度学习的情绪认知具有较高的客观准确性和极强的学习泛化能力，对于构建更加智能的人机交互和医疗诊断系统具有重要意义。针对多通道脑电信号包含的频率、空间和时间信息，采用了一种三维特征输入卷积循环神经网络，将不同通道的微分熵特征转换为三维结构来训练深度模型，融合卷积神经网络和长短期记忆单元从三维输入的每个时间切片中学习频率和空间信息，在 DEAP 数据集上取得了较好的情绪识别效果。

关键词：深度学习 脑电信号 情绪认知

Research on EEG Emotional Cognition Algorithms based on Deep Learning Models

Student Majoring in Software Engineering Qi Jia-He

Tutor Gao Peng

Abstract: As a higher function of human brain, emotion plays an important role in interpersonal communication and medical research. Emotional cognition is a very important link in emotion-related research, among which emotional cognition based on EEG and deep learning has high objective accuracy and strong learning generalization ability, which is of great significance for the construction of a more intelligent human-computer interaction and medical diagnosis system. Aiming at the frequency, space and time information contained in multi-channel EEG signals, a three-dimensional feature input convolutional cyclic neural network is adopted to convert the differential entropy features of different channels into three-dimensional structures to train the depth model. Convolutional neural network and long and short term memory units are integrated to learn the frequency and spatial information from each time slice of three-dimensional input. Good emotion recognition results are obtained on DEAP data set.

Key words: Deep learning; EEG signals; Emotional cognition;

1 绪论

1.1 选题背景及研究意义

情绪是人生理活动的重要组成部分，在人与人之间的互动、决策和学习等日常活动中十分重要。更好地认知的情绪在医疗辅助、教育、交通安全等方面有很大的应用价值，同时对构建更智能更人性化的人机交互系统具有重要的意义，因此用于情绪认知的情感计算越来越受到人们的关注和认可。情感计算是利用传感器采集由情绪引发的生理信号或非生理信号，将其数字化后利用人工智能和数据处理技术对其建模得到情绪模型，进而将模型应用于人机交互系统中对人类的情绪进行感知、识别和理解。

情感计算实现过程中需要获取的生理信号包括心率、皮肤阻抗和脑电（Electroencephalogram, EEG）信号等，对非生理信号通常利用摄像机、麦克风等设备获取面部表情、语音语调等。其中，表情、语音语调等非生理信号容易获取但更容易被伪装，对情感的反映真实性较低。而且对于一些患有特殊疾病的人群，非生理信号较难采集，不易实现对情绪的认知。而诸如心率、呼吸等生理信号虽不易伪造，但由于准确性低，在实验中很少使用。而基于中枢神经系统的脑电信号具有客观真实性等特点，更能直接反映出大脑的电生理活动情况，从而可以实现通过分析大脑发出的不同信号来识别得到相应的情绪，识别率准确相较之前几种信号更高，在情绪识别和认知方面得到了越来越多的关注和应用。

1.2 研究现状

情绪认知的研究主要经历了三个发展阶段：

第一阶段是使用非生理信号（包括语音、手势和面部表情）进行情绪认知和分析。起初研究人员先用面部表情来识别情绪，但由于单一情绪识别成功率较低，又进一步采用复合情绪来识别情绪。后来，部分研究人员使用语音来识别情感，初始时他们使用了各种手工设计的特征^[1]，但此类特征的表征能力较弱，不足以识别一个人的情绪。为了解决这个问题，人们渐渐采用深度学习的方法，Abdel-Hamid 等人^[2]将卷积神经网络（Convolutional Neural Network, CNN）应用于语音情感识别；孙晓等人^[3]为了进一步提高深度学习在面部表情分类中的鲁棒性，在深度 CNN 中引入感兴趣区域和 k 近邻（ k -Nearest Neighbors, kNN），提取情感特征的同时解决了深度网络模型泛化能力较差的问题。

第二阶段是利用人体的生理信号（包括眼电、肌电、皮肤电、EEG 信号等）来完成情绪识别任务。Torres 等人^[4]利用核主成分分析（Kernel Principal Component Analysis, K-PCA）对 EEG 信号进行预处理，并使用 k 近邻和含有多种不同形式核函数的支持向量机根据标签类型进行分类。但在数据量较大和分类标签种类较多的情况时，该算法对于情绪识别的正确率不高。车敏诗等人^[5]构建了 CHAID（Chi-square Automatic Interaction Detection）决策树对皮肤电、心电和呼吸信号的三种混沌特征进行优化，最终对喜、怒、哀三种情绪的识别率分别达到了 88%、100%和 100%。

第三阶段是采用多模态方法来识别情绪。研究人员提出了多模态情绪识别进行特征提取，并使用外周生理信号辅助 EEG 信号，充分考虑了人体的各类生理影响。Wang 等人^[6]利用基于纯注意力机制的情绪转换器融合（ETF）对多通道 EEG 信号和外周眼动信号的互补信息进行分析，显著提高了辨别愤怒、惊讶和中性情绪的表现。方伟杰等人^[7]针对语音、EEG 信号和面部表情分别设计了轻量化 CNN、树状 LSTM 和 GhostNet 进行特征学习，并融合多模态情绪识别结果综合判断人的情绪，有效地提高了情绪识别准确率。

1.3 主要研究内容

本文主要研究基于深度学习与脑电信号的情绪识别。采用了一种深度学习模型应用于脑电信号的情绪认知，总结得到具有较好情绪认知效果的模型，并对部分深度学习模型进行对比验证，以体现本文模型的识别优势。具体内容如下：

(1) 为了更好地学习脑电信号的频率、时间和空间结构信息，构建了具有时间信息的三维特征结构，包含从四个频段 (α 、 β 、 γ 、 θ) 中提取的频率信息，还通过二维图保留了电极的空间拓扑结构，从而保留了多通道的空间信息。此外，包含了数秒连续的微分熵特征保留时间信息，可以捕捉到情绪的动态内容。

(2) 将 CNN 与长短期记忆网络 (Long Short-Term Memory, LSTM) 深度融合。首先利用 CNN 从三维脑电结构的二维图中提取高级空间特征，再利用 LSTM 从含有空间特征的卷积层输出中学习时间信息，最后利用 LSTM 模块的输出结果进行分类。

(3) 通过对各种方法的对比实验，验证了所提出深度学习网络的有效性，并且也验证了卷积循环的融合网络结构相比于传统的网络模型更适合情绪分析。

1.4 论文组织结构

论文组织结构如下：

第 1 章为绪论，阐述了课题背景与研究意义；第 2 章介绍了相关背景知识，主要包括情绪、脑电信号以及深度学习的相关知识，并说明了本文采用的数据集；第 3 章针对情绪认知提出了应用三维特征输入的卷积循环神经网络进行情绪识别；第 4 章介绍了实验的设置，并对实验结果进行了分析；最后总结了论文的研究内容与成果，并对未来基于深度学习与 EEG 信号的情绪认知工作进行了展望。

2 相关背景知识介绍

2.1 情绪的相关背景知识

2.1.1 情绪的定义

自然界中每一个生物都有情绪，但一直以来对情绪的研究存在非常多的声音，很难给出一个单一的定义。最早的情绪理论由美国心理学之父 William 在 1884 年提出，他对情绪的解释是人体对内在变化产生的感受，而对这些生理变化的反应是导致对情绪认识的原因，例如面部表情、肌肉活动和身体的搏动。到了 20 世纪 50 年代，对情绪的认知开始有了理论知识的指导和规定，最具代表性的是 Schachter 和 Singer 提出的情绪双因素理论^[8]，他们认为情绪产生必须满足两个条件，即生理唤醒和认知唤醒，而认知唤醒是该理论的核心。现在，多数研究将情绪理解为是对外部事件产生的一种内部主观体验，通常会伴随着面部表情或言语方式等外部表现。但很多时候隐藏的情绪不容易被发现，这便是只有内在心理状态的变化。

尽管当前学术界没有对情绪形成较为系统的定义，但情绪的基本组成受到研究者们的一致认可，其主要包括三个要素：一是情绪产生时会伴随一些生理变化；二是情绪受主观意识控制；三是情绪的表达实际上是对已发生事件的一种自我评价，即使面对同样的事情，不同人感受和表达情绪的方式也不同。

2.1.2 情绪的分类

虽然对情绪的定义没有一个公认的结论，但一些研究人员提出了情绪空间模型的概念，可以更好地用于情绪研究和实现情绪可视化。情绪的空间模型是指预先定义一个空间，并将情绪分解为几个不同的维度，再将不同类型的情绪以某种方式映射到空间中的特定点上，实现了用空间中的点来唯一的表示不同的情绪。当前应用最多的情绪模型是

由 Russell 提出的效价—唤醒度二维情绪分类模型^[9]，如图 2-1 所示。横轴表示情绪效价（Valence），对应情绪的愉悦程度，从高到低代表了情绪愉快到不愉快的变化；纵轴表示情绪唤醒度（Arousal），对应状态的兴奋程度，从高到低代表从兴奋到平静的变化。因此，不同的情绪都可以根据其愉悦和唤醒的程度映射到二维情绪空间中相应位置，并且可以描述情绪演变的过程，可视化标记每种情绪类型。

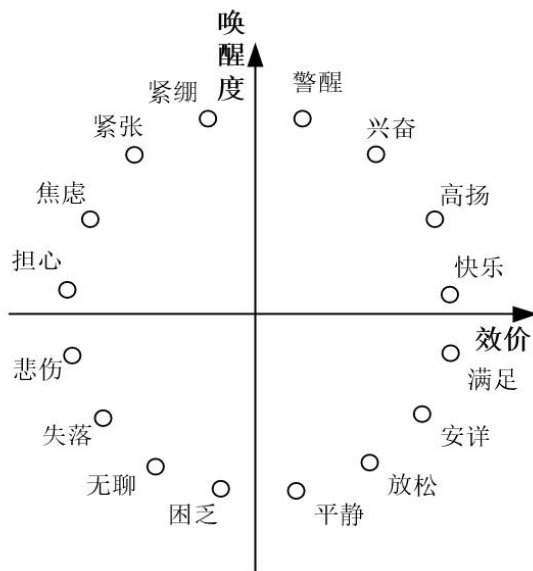


图 2-1 Russell 情绪分类模型

2.1.3 情绪的诱发

情绪是注意力、记忆和决策等认知活动的背景。在基于生理信号的情绪识别中，情绪的诱发很大程度上影响着认知活动。目前情感的诱发方法可以分为两类：材料诱发和情境诱发。

诱发情绪的材料主要有文字、图片、视频、音乐、气味和来自多个感官通道的材料联合。图片、视频等能通过视觉呈现出具有浓郁情绪色彩的信息，可以更容易地产生目标情绪。但是，诱发情绪的材料选择不仅取决于研究目的和情绪的分类模型，还要考虑被试的主观情绪影响。不同被试对视觉、听觉和嗅觉刺激的敏感性不同，所以情绪反应的强度和持续时间也存在差异，这些因素都影响情绪诱发的效果。

情境诱发是通过被试自行回忆或在引导下想象情境来诱发出特定的情感。其中回忆法需要被试深入地体会和感受某种情景的氛围，比想象法更能控制情绪唤起的强度和时机。近年来随着信息技术的发展，利用虚拟现实技术的情绪诱发，能帮助参与者更加身临其境的体验并产生相应的情绪，提高了情绪诱发的普适性。

2.2 脑电的相关背景知识

2.2.1 脑电信号介绍

EEG 信号是指按时间顺序，在头皮表层由一组大脑神经元自发的、有节奏的运动所产生的联合电活动^[10]。近年来，EEG 信号在神经科学、心理学和认知科学等领域得到了广泛应用，同时也被用于诊断和识别许多脑部疾病，如睡眠障碍、阿尔茨海默症、癫痫活动等，在医学研究中发挥了非常重要的作用。

2.2.2 脑电信号的预处理

EEG 信号预处理包括选择、去噪和波段分离。对于 EEG 信号的选择，部分学者指出，当一些相同的材料刺激一组被试时，该组 EEG 数据中产生较大差异反应的部分应被去除，以提高识别精度。同时有些刺激材料并不能引起绝大多数被试产生相同或相似

的反应，导致被试产生的情绪一致性较差，缺乏统计学意义，也应予以排除。

EEG 信号去噪是指去除信号中所掺杂的噪声。在情绪认知研究中，对 EEG 信号的去噪主要包括去除眼电、肌电、工频干扰等跟任务不相干的噪声伪迹^[10]，以提高状态判断准确率。多数研究工作采用人工处理去噪，即通过视觉识别异常并及时去除，但这种方法很难识别隐藏在多通道 EEG 信号中的稀疏噪声，而且人工处理依赖于研究人员的个人经验，主观性强，去噪结果的复现性较差。因此部分研究工作利用自动去除的方式，包括独立成分分析法、主成分分析法以及经验模式分解等方法。Nathan 等人^[11]采用独立成分分析法将 EEG 信号分解成多通道独立分量，通过分析每个独立分量的时空特征，能自动识别眨眼、肌肉活动等部分伪影，获得了良好的去噪效果。Guarascio 等人^[12]提出了一种将经验模态分解与自适应噪声和 Renyi 熵相结合的自适应方法，利用正则准则自动选择固有模态函数重建伪影最小化的脑电信号，能自动最大限度地减少受损脑电图信号的眼伪影噪声。

对 EEG 信号的频段分离，通常按照不同的频率范围定义 δ (1~4Hz)、 θ (4~8Hz)、 α (8~12Hz)、 β (13~30Hz) 和 γ (31~45Hz) 五个主要频段^[13]。EEG 信号不同频段与大脑的不同活动状态息息相关，比如当人心情愉悦或静思冥想时，其枕骨区脑电 α 波会显著增强，而当人紧张、情绪激动或亢奋时，原来的慢波节律可立即被 β 波替代。所以利用此特质对脑电信号进行波段分离可以进一步判断不同情绪在不同波段之间的差异，有利于提高情绪识别的准确性^[14]。基于计算机技术的 EEG 信号的波段分离方法基本上可以分为两种。文献^[15]利用构建数字滤波器的方法得到的几个不同波段的信号。文献^[16]提出利用离散小波变换将 EEG 信号分解为五个波段。实验结果证明，两种方法都能取得较好的效果。

2.2.3 脑电信号的特征提取

EEG 信号的特征提取是指从预处理后的脑电信号中提取出部分能区分大脑不同状态的特征信息，主要包括时域特征、频域特征、时频域特征和非线性特征。

(1) 时域特征

EEG 信号的时域特征是指在时域内的信息或信号统计量，主要包括 EEG 时域信号的均值、方差、偏度、峰度、Hjorth 特征等信号统计量。其中，偏度和峰度用来表示数据分布偏离正态分布的程度；Hjorth 特征是融合活动性、移动性和复杂性三个参数的统计特征，可以从多个角度对信号进行更全面的估计。EEG 信号的时域特征是几种特征中更容易理解、能直接处理的特征，在情绪识别分析中有较早的应用历史。

(2) 频域特征

EEG 信号的频域特征是将信号从时间序列转化为频率序列，然后提取相关特征分析。由于 EEG 信号通常按频率范围分为五个波段： α 、 β 、 θ 、 γ 、 δ ，因此在基于 EEG 的情绪识别中，研究人员首先将 EEG 信号映射到五个频段，然后分别获得每个频段的频域特征。常见的频域特征包括能量、功率谱密度等。其中，功率谱密度表示单位频率波所携带的功率，在情绪识别中应用更为流行^[17]。

(3) 时频特征

为了寻求更高的分类精度，研究人员致力于发现同时反映时域和频域特征的 EEG 信号。提取时频特征的常用方法主要包括短时傅里叶变换 (Short-time Fourier Transform, STFT) 和小波包变换 (Wavelet Packet Transform, WPT)。STFT 在传统的傅里叶变换中增加了一个滑动窗函数，将整个时域阶段分解为无数个小的、近似光滑的等长小阶段，提取时变信号中的频率和相位信息。WPT 是一种新的变换分析方法，它提供了一个可变的“时间-频率”窗口，通过提取 EEG 信号的高频部分并对其分解，可以更好的进行非

平稳的信号的局部特征提取。Krisnandhka 等人^[18]利用相对小波能量提取时频特征，在情绪识别中的准确度达到了 76%。

(4) 非线性特征

在 EEG 信号的特征提取中，对非线性特征的分析逐渐被广泛应用，常用的非线性特征主要包括熵特征，其中的微分熵特征（Differential Entropy, DE）通常用来表示连续随机变量的复杂性，是 EEG 情绪识别领域常用且有效的特征提取方法，其公式如下：

$$h(x) = - \int_x p(x) \log(p(x)) dx \quad (2-1)$$

式中， $p(x)$ 表示信息的概率密度函数。

值得注意的是，微分熵本身并不能反映连续信源的信息度量，只有熵差才具有信息度量的意义。通常采用微分熵的高斯分布计算形式来表示信号的特征。

2.3 深度学习相关背景知识

2.3.1 深度学习简介

深度学习（Deep Learning）是随着硬件设施的算力提高而逐渐兴起的人工智能算法，通过学习样本的内在规律让计算机具备识别和分析的能力，并使用多层网络模型精准完成图像检测、语音识别和语言翻译等通用任务，以实现人工智能的目的，目前深度学习以人工神经网络为主要参考模型。

2.3.2 人工神经网络

人工神经网络（Artificial Neural Network, ANN）也称为神经网络，是一种模仿生物神经元的数学理论。随着神经网络的发展，以由最初的全链接神经网络逐渐升级为适用于不同情况的更为复杂的网络模型，其大致可以分为三类：卷积神经网络、循环神经网络（Recurrent Neural Network, RNN）和对抗神经网络。通常的网络结构由输入层、隐藏层和输出层三个部分组成，其中存在大量的参数，而神经网络的训练过程便是将训练集数据输入，不断更新网络得到符合该场景的参数。在模型训练完成后，将需要识别或分类的数据输入网络便可输出得到我们最终所需要的结果。

2.4 基于深度学习的脑电情绪认知

随着深度学习模型在各种分类识别任务中应用取得的成功，越来越多基于深度学习的 EEG 情绪认知相继提出。相较传统的机器学习方法，深度学习能更好的理解问题并进行拟合，且适应性高、易于迁移。此外，深度学习模型的特性能够克服传统 EEG 情绪识别方法存在的一些不足，并解决任务本身存在的一些挑战。现阶段，基于深度学习与 EEG 信号的情绪识别常用的深度学习模型包括 CNN、LSTM、变分自编码器等。这类深度学习模型的应用能对原始信号或者人工提取出的特征进行进一步的抽象和深层特征提取，获得更加具有判别力的特征，也在基于 EEG 信号的情绪识别中取得了令人满意的效果。

2.5 实验数据集

本文采用 DEAP 数据集^[19]，其是由来自四所世界知名高校的研究人员为了分析人类情绪状态提出的一个多模态数据集。在实验中，研究人员要求 32 名被试（其中男性 16 人，女性 16 人）观看 40 段 1 分钟左右的音乐 MV，并收集了他们的脑电波信号。该信号由 40 个通道构成，包含 32 个 EEG 信号通道和 8 个外部生理信号通道。在观看完视频后，要求被试自我评估主观情绪体验，并根据图 2-1 中的 Russell 情绪空间模型对每段视频的唤醒度、效价、喜爱度、主导地位和熟悉程度进行分别评分。最后，将来自不同模式的分类结果进行决策融合。

3 基于三维特征卷积循环神经网络的脑电情绪认知算法

3.1 方案架构设计

由于 EEG 信号的频率、空间和时间特征对情绪识别都很重要，然而目前的研究大多只考虑这三种特征中的一种或两种。因此，为了更有效地整合 EEG 信号的频率、空间和时间信息，本文借鉴了 CNN 和 RNN 网络结构的不同特点，提出了针对 EEG 信号的卷积循环神经网络（Convolutional Recurrent Neural Network, CRNN），算法框架如图 3-1 所示。

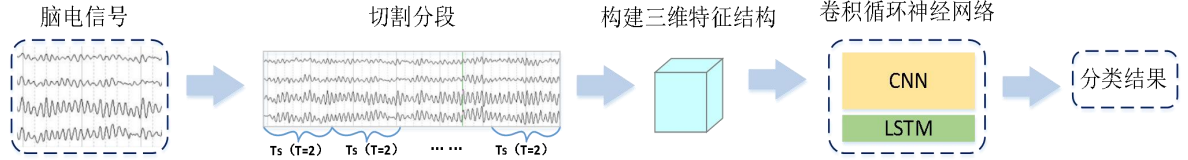


图 3-1 基于三维特征卷积循环神经网络的脑电情绪认知算法框架

3.2 数据预处理和特征提取

3.2.1 数据预处理

本文实验采用的 DEAP 数据集包含了 32 位实验参与者完成了 40 个实验的脑电信号数据，其中每位参与者的脑电数据包含 40 个通道的信息，每段脑电信号的时长为 63 秒，其中前 3 秒是试验前放松状态下的基线信号。由于该数据集提供的脑电信号数据已经进行了去噪和滤波等基本操作，所以为了增加训练数据量，并更好地反映 EEG 信号的时间和频率特征，采用对 EEG 数据进行分割的方法进一步处理数据。具体来说，将采集的 EEG 信号划分为 N 个长度为 T 秒的不重叠的片段，并为每个片段分配原始试验的情绪分类标签。图 3-2 对分割方法进行了简单的展示。

脑电信号（共60秒）

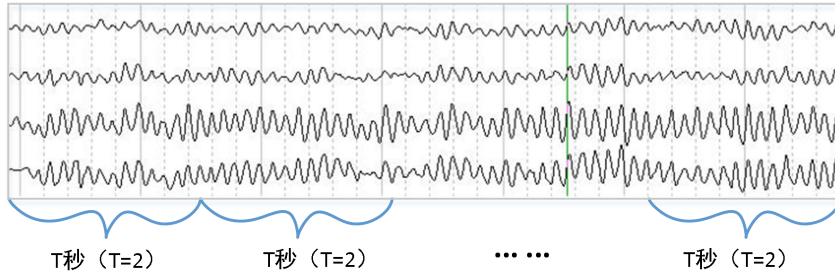


图 3-2 脑电信号分割示意图

3.2.2 脑电特征提取

在对脑电数据进行切割之后，我们从每个频段以 0.5 秒的窗口提取微分熵，用以表示 EEG 信号的复杂程度，有研究证实了其在情绪识别方面的有效性^[20]，微分熵定义如下：

$$h(x) = \int_X f_X(x) \log f_X(x) dx \quad (3-1)$$

式中，X 是随机变量， $f(x)$ 是 X 的概率密度函数。

对于高斯分布的微分熵，设随机变量服从均匀分布，概率密度可表示为：

$$f(x) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp \left[-\frac{(x-\mu)^2}{2\sigma^2} \right] \quad (3-2)$$

则高斯分布的微分熵计算表示为:

$$\begin{aligned} h(x) &= -\log_e \int_{-\infty}^{\infty} f(x) \ln f(x) = -\log_e \int_{-\infty}^{\infty} f(x) \left[-\frac{1}{2} \ln(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2} \right] dx \\ &= \frac{1}{2} \log_e [\ln(2\pi\sigma^2) + \sigma^2 E(x-\mu)^2] = \frac{1}{2} \log_e [\ln(2\pi\sigma^2) + 1] = \frac{1}{2} \log(2\pi e\sigma^2) \quad (3-3) \end{aligned}$$

式中, e 和 σ 分别是时间序列的欧拉常数和标准差。

EEG 信号不同频段的活动情况可以反映大脑的不同工作状态, 利用微分熵分析 EEG 信号在不同频段上单位频率波所携带的功率, 可以获得更精确的电波活动图像, 从而更精确地感知大脑的工作状态。脑电信号进行波段分离后可以划分为 α 、 β 、 θ 、 γ 和 δ 共 5 个频段。由于 DEAP 数据集预处理之后的数据频率范围是 4~45Hz, 所以对于预处理过后划分好的每个信号片段, 用巴特沃斯滤波器提取信号后获得 θ 、 α 、 β 和 γ 和 4 个频段。在完成信号分割和频段划分后, 数据集中每个被试信号样本的构成变为 $32 \times 40 \times 4 \times N \times T$ 长度大小的数据 (通道数 \times 被试数 \times 频段数 \times 信号数 \times 片段长度)。

对于经过分割操作后的单个样本, 分别对每一个通道的信号进行特征提取, 最后计算整段信号微分熵。将每一个子频段内微分熵连接起来的结果作为该频段的微分熵特征, 然后再将各个子频段的微分熵特征连接起来形成特征向量。最终, 在单个样本的单个通道上, 得到特定长度的微分熵特征向量。

3.3 三维结构时间切片输入的卷积循环神经网络

3.3.1 三维结构的时间切片构造

为了整合 EEG 信号的频率、空间和时间特征, 构建了包含这三种信息的三维特征结构。首先整合 EEG 信号的频率和时间特征。假设一个原始 EEG 信号片段属于和共同决定的实数集合, 其中表示采集信号所用电极数量, 表示采样率 (设为 128)。然后引入频段分离的思路, 对于每个 EEG 信号片段, 我们计算每个频段 0.5 秒窗口的微分熵特征, 并对每个微分熵特征向量进行归一化。因此, 原始 EEG 信号片段转化为用微分熵特征表示的实数集合, 由 s 、 rT 和 d 共同决定, 其中为频段数 (设为 4)。

为了保留电极位置的空间结构信息, 根据电极位置进一步将微分熵特征向量转换为紧凑的二维图。例如, 33 通道 EEG 电极分布如图 3-3 (a) 所示, 信道的二维映射如图 3-3 (b) 所示, 其中 0 表示信道信号未被使用。当所有频段的微分熵特征向量都转换为二维图后, 再将每个频段的二维图堆叠形成三维结构, 期望将不同波段的互补信息组合在一起。因此, 每一段 EEG 信号都可以转换为具有时间特征的三维微分熵特征结构, 最终表示为 $X_n \in R^{h \times w \times d \times 2T}$, $n = 1, 2, 3, \dots, N$ 。其中 h 和 w 分别是二维特征图的高度和宽度, d 为频段数, N 为总样本数。在本文中, 设置 $h = 7$ 、 $w = 7$ 、 $d = 4$ 。

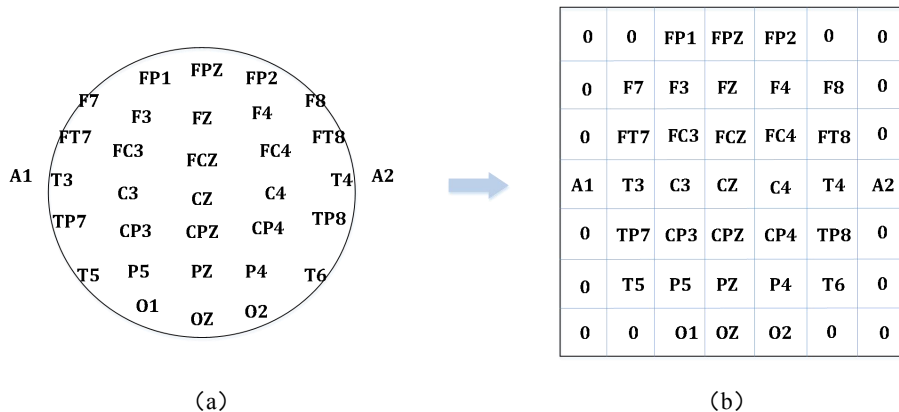


图 3-3 EEG 信号的 33 通道电极分布及其二维映射图

3.3.2 频率和空间特征学习

对于每个样本 (具有时间特征的三维结构), 通过 CNN 从它的每个信号片段中提取频率和空间信息, 如图 3-4 所示。所使用的 CNN 包含四个卷积层, 一个最大池化层和一个全连接层。具体来说, 第一卷积层 (Conv1) 的输出通道数为 64, 采用的卷积核大小为 5×5 , Stride 设为 1; 第二卷积层 (Conv2) 和第三卷积层 (Conv3) 的输出通道数分别为 128 和 256, 卷积核大小为 4×4 , Stride 设为 1; 第四卷积层 (Conv4) 的输出通道数为 64, 卷积核大小变为 1×1 , 用于融合上一卷积层的特征图。对于所有卷积层, 在卷积操作前首先完成零填充, 用 ReLU 作为激活函数。完成卷积运算后, 设置了一个最大池化层, 大小为 2×2 , 以信息丢失为代价来减少参数的数量。将最大池化层的输出映射为长度为 1024 的一个向量, 再送到具有 512 个单元的全连接层中, 输出得到由频率和空间信息表示的脑电信号数据, 即实现由 $X_n \in R^{h \times w \times d \times 2T}$ 到 $Y_n \in R^{512 \times 2T}$ 的映射转换。

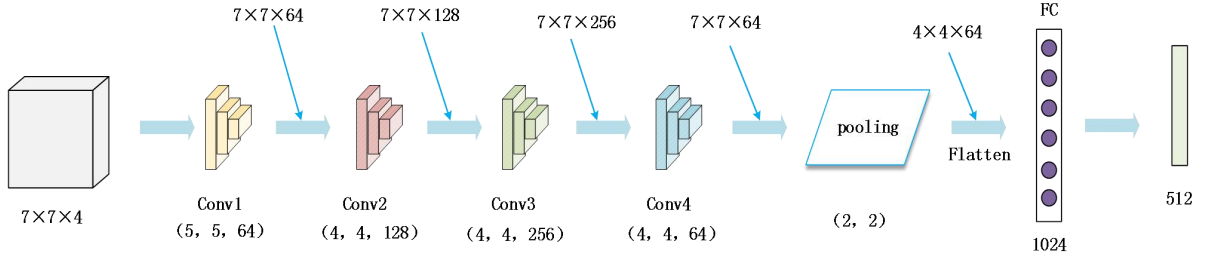


图 3-4 用于频率和空间特征学习的卷积模块结构

3.3.3 时间特征学习

由于 EEG 信号片段之间可能隐藏了额外的信息, 有助于更准确地进行情绪分类。因此, 利用 LSTM 从上文卷积单元的最终输出中提取时间信息。给定卷积单元的输出序列为 $Y_n \in R^{512 \times 2T}$, 即可以表示个连续的时间切片, 每个切片大小 $y_t \in R^{512}$ 。采用一个具有 128 个存储单元的 LSTM 来挖掘内段的时间依赖性。LSTM 包含三个门结构, 不同结构的输出和最终输出可以计算如下:

$$f_t = \sigma(W_f \cdot [h_{t-1}, y_t] + b_f) \quad (3-4)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, y_t] + b_i) \quad (3-5)$$

$$C_t = f_t * C_{t-1} + i_t * \tanh(W_c \cdot [h_{t-1}, y_t] + b_c) \quad (3-6)$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, y_t] + b_o) \quad (3-7)$$

$$h_t = O_t * \tanh(C_t) \quad (3-8)$$

$$Z_t = W_{h0} h_t + b_t \quad (3-9)$$

在经历了频率、空间、时间信息的学习后, EEG 信号的最终高级表示为一个 128 的输出 LSTM 节点 Z_n 。它集成了 T 秒的脑电图片段的频率、空间和时间线索。

3.3.4 分类器

基于最终的输出表示, 通过线性变换方法预测原始脑电信号片段所属的情绪类型标

签，可计算为：

$$O = WZ_n + b = \{O_1, O_2, \dots, O_k\} \quad (3-10)$$

式中， W 是线性变换矩阵， b 是偏差， k 是实验中划分的情感类别的个数。

最后，将输出集合输入到 softmax 分类器中进行情绪识别，可以表述为：

$$P(k|X_n) = \max \left\{ \frac{\exp(O_i)}{\sum_{j=1}^k \exp(O_j)} \mid i = 1, 2, \dots, k \right\} \quad (3-11)$$

式中， $P(k|X_n)$ 表示 EEG 信号片段 X_n 属于类别 k 的概率。

3.3.5 基于 EEG 的情绪识别架构

基于 EEG 的情绪识别框架如图 4-5 所示。设计了一个组合式深度模型，从三维特征结构中同时学习 EEG 信号的三种信息，用于情绪认知。首先，构建了一个三维特征结构，明确地组织了频率、时间和位置结构这三种信息。然后，引入了卷积循环神经网络模型，该模型由 CNN 和 LSTM 组合而成。具体来说，CNN 用于学习三维结构中每个信号片段的频率和空间表示，LSTM 将卷积单元的输出作为输入，提取片段之间的时间依赖关系，最后输入到 softmax 分类器中进行情绪识别。

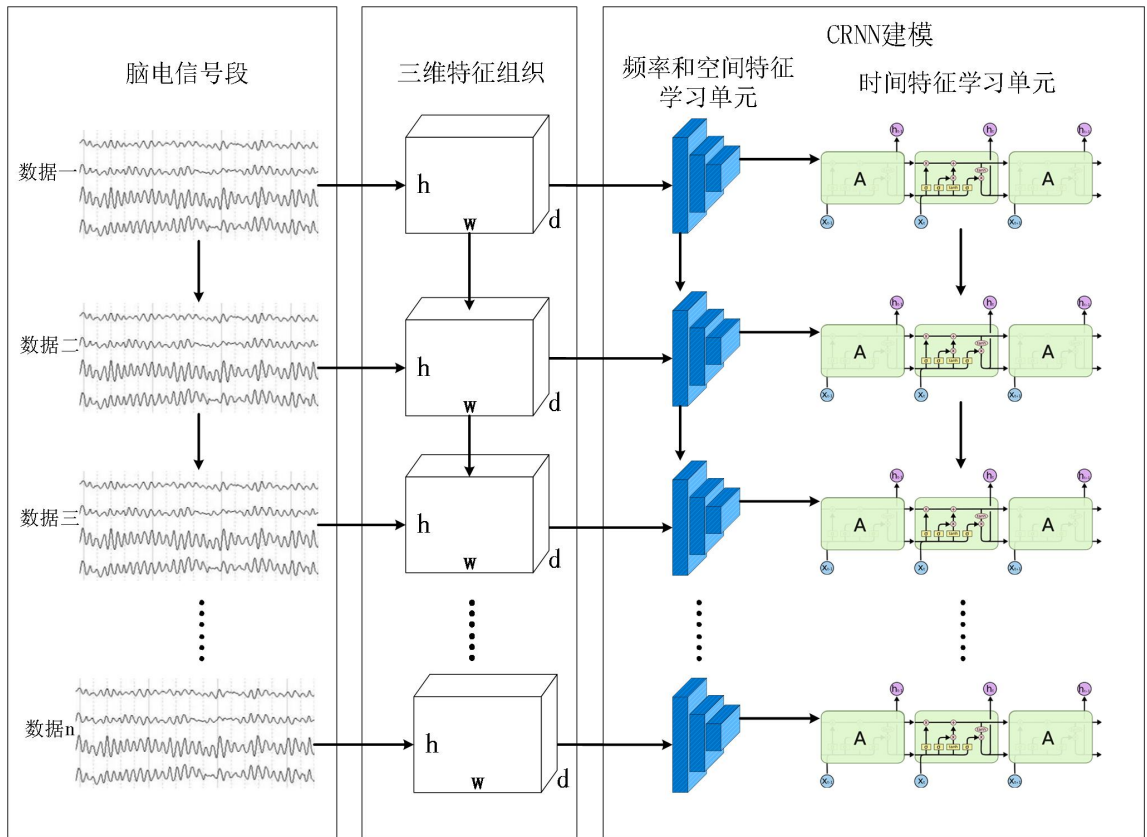


图 3-5 基于 EEG 的情绪识别架构

4 实验设置与结果分析

4.1 数据训练环境与测试环境

本模型在训练时采用 Python 语言环境以及第三方库，包括 sklearn 提供机器学习方法；Tensorflow 提供深度学习方法；Keras 库是对 Tensorflow 进一步封装，提供更高级

的和封装度更高的使用方法具体版本如表 4-1 所示。

表 4-1 训练测试环境信息

环境工具	版本
Python	3.7
Keras	2.3.1
sklearn	1.1.3
Tensorflow	2.2.0

4.2 训练集和测试集

本文采用 DEAP 数据集作为模型的训练数据集，将每名被试每次实验的前 3 秒基线信号划分为 6 段，从每段信号的四个频段中提取微分熵特征，然后将这 6 个分段在每个频段的微分熵特征平均化，计算出基线微分熵特征，最后利用基线微分熵特征和情绪微分熵特征之间的差异，获取到最终的情绪微分熵特征。同时对于情绪标签的确定，由第 2 章提到的二维情绪分类模型中唤醒度和效价的额定水平（1-9）标记 40 个片段的情绪类别，并选择 5 作为阈值，将标签分为两个二元分类问题，总共四分类。采用 Shen 等人^[21]自采集的 32 名被试 EEG 信号作为测试集，采用 32 通道的 Biosemi-ActiveTwo 设备按照国际 10-20 系统采集 EEG 信号，在采集实验过程中，被试在每次实验结束时对自己的兴奋水平和效价进行自我评估，这用于判断相应的情绪是否正确引发。

4.3 实验设置及模型训练

三维特征卷积循环模型的训练批次大小为 128，使用 Adam 进行优化，学习率设置为 0.001，最大 epoch 设置为 100。模型使用 Tensorflow 框架扩展而来的 Keras 实现，对于 EEG 信号情绪认知方法的性能评估过，我们对每个实验对象应用训练集与测试集 3:1 的比例关系进行交叉验证，其中的平均分类精度和方差代表个体的分类性能。并且为了模型在训练时的规范性，对其标签采用 one-hot 独热编码方式，它将离散分类值用二进制向量来运算，可以消除分类值的差别，更好进行分类任务。完成以上操作后开始数据训练阶段，具体训练步骤如下：

- （1）先将数据从脑电通道线性格式数据转化为二维映射图，二维映射图为 7×7 尺寸。
- （2）将唤醒度和效价标签结果进行 one-hot 编码。
- （3）构建神经网络模型，先将数据送入多层卷积层，后送入 128 次循环的 LSTM 网络，最后通过全链接和交叉熵层进行分类。
- （4）搭建上述实验网络，并采用 adam 优化器，设置学习率为 0.001，batch_size 大小为 128。
- （5）针对 32 组被试其个人的模型依次训练，每个模型 epoch 设置为 100 次，并使用交叉验证进行检验模型的准确性。

4.4 实验结果

本文一共 32 组实验数据，具体准确率如表 4-2、4-3 和 4-4 所示；32 组方差条形图如图 4-1 所示。32 组模型全部准确率平均值为 94.43%，方差为 0.0172。由 32 组平均数据准确率所组成的箱型图如 4-2 图所示，箱型图由最大值、最小值、平均值、25%值、75%值和异常值所组成，是数据展示时常用的数据可视化图像。观察本文平均准确率所生成的箱型图，可直观得知该模型在针对不同人群时具有普适性，单个异常值的存在为正常现象，且准确率达到 80%，说明模型的效果较好。

表 4-2 网络模型验证详细结果表（第 1-10 组实验）

实验编号	准确率 1(%)	准确率 2(%)	准确率 3(%)	准确率 4(%)	准确率 5(%)	平均准确率(%)	方差(10^{-2})
1	96.87	97.5	96.875	96.87	93.75	96.75	1.64
2	89.37	89.37	86.25	86.25	91.25	88.50	1.96
3	93.75	98.12	95.62	95.62	97.50	96.12	1.55
4	92.50	86.25	86.87	91.87	90.625	89.62	2.57
5	94.99	94.99	96.87	96.24	92.50	95.12	1.49
6	96.24	94.99	98.12	97.50	98.12	96.99	1.21
7	98.12	98.75	99.37	98.75	98.12	98.62	0.46
8	98.12	98.75	93.75	96.24	98.75	97.12	1.92
9	90.62	93.12	90.62	93.12	90.625	91.62	1.22
10	96.87	91.25	96.87	96.24	94.37	95.12	2.14

表 4-3 网络模型验证详细结果表（第 11-20 组实验）

实验编号	准确率 1(%)	准确率 2(%)	准确率 3(%)	准确率 4(%)	准确率 5(%)	平均准确率(%)	方差(10^{-2})
11	91.25	92.50	90.62	95.62	93.12	92.62	1.74
12	95.62	96.85	99.37	96.87	98.12	97.37	1.27
13	99.37	98.12	99.37	99.37	98.75	98.99	0.49
14	95.62	93.75	97.50	99.37	96.87	96.62	1.87
15	96.24	88.74	94.99	93.12	91.87	92.99	2.60
16	97.50	95.62	94.37	95.62	94.37	95.50	1.14
17	89.37	94.37	94.37	90.62	96.87	93.12	2.73
18	98.12	96.24	94.37	96.24	95.62	96.12	1.21
19	92.50	94.37	92.50	92.50	92.50	92.87	0.75
20	96.24	97.50	96.24	94.99	95.62	96.12	0.82

表 4-4 网络模型验证详细结果表（第 21-32 组实验）

实验编号	准确率 1(%)	准确率 2(%)	准确率 3(%)	准确率 4(%)	准确率 5(%)	平均准确率(%)	方差(10^{-2})
21	95.62	97.50	98.12	98.75	98.75	97.75	1.15
22	84.37	78.12	78.75	75.62	76.24	78.62	3.09
23	97.50	97.50	98.75	99.37	98.75	98.37	0.74
24	93.12	90.62	94.99	95.62	94.99	93.87	1.82
25	93.12	90.62	95.00	95.62	95.00	93.87	1.82
26	97.50	90.62	96.87	99.37	98.75	96.62	3.12
27	91.25	99.37	92.50	88.75	90.62	92.50	3.64
28	93.75	98.75	97.50	96.87	98.12	97.00	1.74
29	93.12	96.87	95.00	96.87	96.87	95.75	1.50
30	91.87	95.00	94.37	93.75	96.25	94.25	1.44
31	93.75	95.00	89.37	95.00	91.87	93.00	2.14
32	96.25	91.87	92.50	95.00	90.62	93.25	2.06

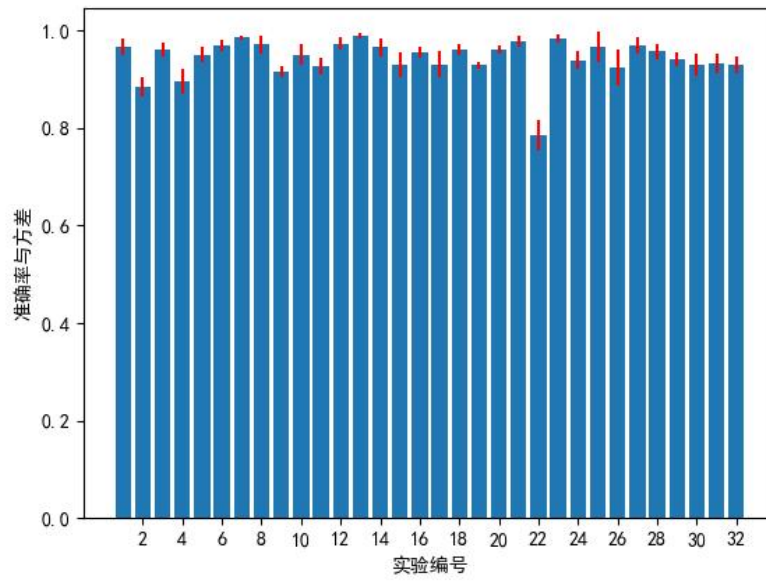


图 4-1 网络模型验证结果

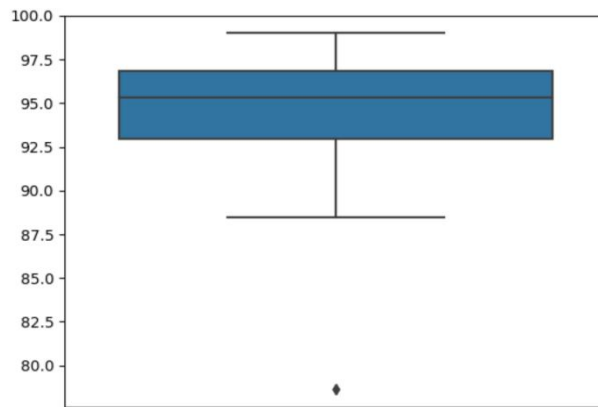


图 4-2 网络模型准确率箱型图

4.5 实验比较

本文结果与已有研究工作的结果比较如表 4-5 所示。将线性数据展开为二维数据，扩充了实验数据的维度，能使模型在更高层次上去学习数据中的信息。在比较中，我们比 CCNN^[22]方法准确率提高了 4.29%，比 PCRNN^[24]方法提高了 3.89%，验证本文模型的准确性，以及融合空间、时间与频率信息思想的可行性。

表 4-5 实验结果比较

方法	输入尺寸	输入信息	准确率 (%)
CCNN ^[22]	D×W×H	频率与空间信息	90.14
EmotionNet ^[23]	2R×W×H	空间与时间信息	90.07
PCRNN ^[24]	2R×W×H	空间与时间信息	90.54
CNN+LSTM (本文)	2T×W×H×D	空间、时间与频率信息	94.43

5 总结与展望

5.1 总结

本文提出了一种基于深度学习模型的 EEG 情绪分类方法,该方法能够更高维度综合分析脑电信号的频域、空间和时域信息,在 DEAP 数据集上取得了较好的效果。首先,将脑电数据构建为含时间线索的三维特征结构,组织 EEG 信号的频率和空间特征;其次,引入 CNN 和 LSTM 深度融合的卷积循环神经网络模型。CNN 处理频率和空间信息,LSTM 提取时间依赖关系。通过其他研究的比较,验证了模型在准确率和稳定性上有着不错的表现。

5.2 展望

本文还存在着一系列的问题,如只关注用单一的脑电生理信号来进行情感识别,没有关注到脑电与其余多种外部生理信号如皮肤电、肌电等之间的相关性,单一考虑脑电信号可能会导致情绪判断不够准确。在下一步的工作中,针对上述问题,采用多模态的脑电情绪识别,挖掘多种外部生理信号与多通道脑电信号之间的相关性,并根据提取的综合特征提高情绪识别的准确性。

致谢

回想在曲园四年来的所有相遇和经历,或是成功的喜悦或是失败的教训,或是同行的友人亦或竞争的对手,好的不好的都有其意义。在这里,是我的青春,亦是满满的收获,纵然不舍,仍心怀感激。

首先我要感谢高鹏老师,从选题的引导到成文过程都给予了我耐心的指导和莫大的帮助。也感谢传道受业解惑的各位任课老师,师恩难忘,铭刻于心。

一路走来,一路收获,这些都将成为宝贵的财富,陪伴我一生。

参考文献

- [1] El Ayadi M, Kamel M S, Karray F. Survey on speech emotion recognition: Features, classification schemes, and databases[J]. Pattern Recognition, 2011, 44(3): 572-587.
- [2] Abdel-Hamid O, Mohamed A, Jiang H, et al. Convolutional Neural Networks for Speech Recognition[J]. IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, 2014, 22(10): 1533.
- [3] 孙晓, 潘汀, 任福继. 基于 ROI-KNN 卷积神经网络的面部表情识别[J]. 自动化学报, 2016, 42(06): 883-891.
- [4] Verma G K, Tiwary U S. Affect representation and recognition in 3d continuous valence-arousal-dominance space[J]. Multimedia Tools and Applications, 2017, 76(2): 2159-2183.
- [5] 车敏诗, 聂春燕, 范如俊等. 一种基于混沌特征及优化 CHAID 决策树的情绪识别方法[J]. 计算机应用研究, 2020, 37(S2): 105-107.
- [6] Wang Y, Jiang W, Li R, et al. Emotion Transformer Fusion: Complementary Representation Properties of EEG and Eye Movements on Recognizing Anger and Surprise[C]//2021 IEEE International Conference on Bioinformatics and Biomedicine. 2021: 1575-1578.
- [7] 方伟杰, 张志航, 王恒畅等. 融合语音、脑电和人脸表情的多模态情绪识别[J]. 计算机系统应用, 2023, 32(01): 337-347.
- [8] Chachter S, Singer J E. Cognitive, social, and physiological determinants of emotional state[J]. Psychological Review, 1962, 69(5): 379.

- [9] Russell J A. A circumplex model of affect[J]. *Journal of personality and social psychology*, 1980, 39(6): 1161.
- [10] 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(04): 595-606.
- [11] Nathan K, Contreras-Vidal JL. Negligible motion artifacts in scalp electroencephalography (EEG) during treadmill walking[J]. *Frontiers in Human Neuroscience*, 2015, 9: 708.
- [12] Guarascio, Mario, Puthusserypady, et al. Automatic minimization of ocular artifacts from electroencephalogram: A novel approach by combining Complete EEMD with Adaptive Noise and Renyi's Entropy[J]. *Biomedical signal processing and control*, 2017, 36(Jul.): 63-75.
- [13] 李锦瑶, 杜肖兵, 朱志亮等. 脑电情绪识别的深度学习研究综述[J]. *软件学报*, 2023, 34(01): 255-276.
- [14] Li D, Xie L, Chai B, et al. Spatial-frequency convolutional self-attention network for EEG emotion recognition[J]. *Applied Soft Computing*, 2022, 122: 108740.
- [15] Serna J, Paternina M, Zamora-Mendez A, et al. EEG-rhythm specific Taylor–Fourier filter bank implemented with O-splines for the detection of epilepsy using EEG signals[J]. *IEEE Sensors Journal*, 2020, 20(12): 6542-6551.
- [16] Alickovic E, Kevric J, Subasi A. Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction[J]. *Biomedical signal processing and control*, 2018, 39: 94-102.
- [17] 刘鹏, 乔晓艳. 基于栈式自编码神经网络的脑电信号情绪识别[J]. *测试技术学报*, 2021, 35(2): 145-151.
- [18] Krisnandhika B, Faqih A, Pumamasari P D, et al. Emotion Recognition System Based on EEG Signals Using Relative Wavelet Energy Features and a Modified Radial Basis Function Neural Networks[C]//*International Conference on Consumer Electronics and Devices*. London, England, 2017: 50-54.
- [19] Koelstra S, Muhl C, Soleymani M, et al. DEAP: A Database for Emotion Analysis Using Physiological Signals[J]. *IEEE Transactions on Affective Computing*, 2012, 3(1): 18–31.
- [20] Ahmed M Z I, Sinha N, Phadikar S, et al. Automated feature extraction on AsMap for emotion classification using EEG[J]. *Sensors*, 2022, 22(6): 2346.
- [21] Shen F, Dai G, Lin G, et al. EEG-based emotion recognition using 4D convolutional recurrent neural network[J]. *Cognitive Neurodynamics*, 2020, 14: 815-828.
- [22] Yang Y, Wu Q, Fu Y, et al. Continuous convolutional neural network with 3D input for EEG-based emotion recognition[C]// *International Conference on Neural Information Processing*. 2018: 433–443.
- [23] Wang Y, Huang Z, Mccane B, et al. EmotionNet: A 3-D Convolutional Neural Network for EEG-based Emotion Recognition[C]// 2018: 1-7.
- [24] Yang Y, Wu Q, Qiu M, et al. Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network[C]// *International Joint Conference on Neural Networks*. 2018: 1-7.