

基于双流结构缩放和多重注意力机制的轻量级脑电情感识别方法

雷颖, 刘峰

引用本文

雷颖, 刘峰. [基于双流结构缩放和多重注意力机制的轻量级脑电情感识别方法](#)[J]. 计算机科学, 2023, 50(6A): 220300262-9.

LEI Ying, LIU Feng. [LDM-EEG:A Lightweight EEG Emotion Recognition Method Based on Dual-stream Structure Scaling and Multiple Attention Mechanisms](#) [J]. Computer Science, 2023, 50(6A): 220300262-9.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于PRF-RFECV特征优选的GA-LightGBM的网络安全态势评估](#)

Network Security Situation Assessment for GA-LightGBM Based on PRF-RFECV Feature Optimization
计算机科学, 2023, 50(6A): 220400151-6. <https://doi.org/10.11896/jsjcx.220400151>

[基于动态金字塔和子空间注意力的图像超分辨率重建网络](#)

Image Super-resolution Reconstruction Network Based on Dynamic Pyramid and Subspace Attention
计算机科学, 2022, 49(11A): 210900202-8. <https://doi.org/10.11896/jsjcx.210900202>

[基于纹理特征增强和轻量级网络的人脸防伪算法](#)

Face Anti-spoofing Algorithm Based on Texture Feature Enhancement and Light Neural Network
计算机科学, 2022, 49(6A): 390-396. <https://doi.org/10.11896/jsjcx.210600217>

[一种基于Bottleneck Transformer的轻量级微表情识别架构](#)

Lightweight Micro-expression Recognition Architecture Based on Bottleneck Transformer
计算机科学, 2022, 49(6A): 370-377. <https://doi.org/10.11896/jsjcx.210500023>

[基于改进哈希时间锁的区块链跨链资产交互协议](#)

Novel Hash-time-lock-contract Based Cross-chain Token Swap Mechanism of Blockchain
计算机科学, 2022, 49(1): 336-344. <https://doi.org/10.11896/jsjcx.210600170>

基于双流结构缩放和多重注意力机制的轻量级脑电情感识别方法

雷 颖^{1,3} 刘 峰^{1,2,3,4}

- 1 华东师范大学计算机科学与技术学院 上海 200062
 - 2 上海对外经贸大学人工智能与变革管理研究院 上海 201620
 - 3 华东师范大学上海智能教育研究院 上海 200062
 - 4 华东师范大学心理与认知科学学院上海市心理健康与危机干预重点实验室 上海 200062
- (lycherishcs@163.com)

摘 要 脑电情感识别是一个复杂程度高、信息密度大、海量数据的多通道时序信号分类问题。为在保持现有分类精度的情况下减少计算参数量,实现脑电情感识别的精度与性能最优,提出了一种基于双流结构缩放和多重注意力机制的轻量级网络(LDM-EEG)。该网络以基于脑电信号的微分熵特征构造的时域-空域图谱和频域-空域图谱作为输入,采用对称的双流结构对上述两种特征分别处理,通过节约参数的新型残差模块和网络缩放机制来实现轻量化,并利用新型的通道-时/频-空多重注意力机制和后注意力机制提升模型特征聚合能力。实验结果表明,在参数量明显减小的情况下,模型在 SEED 数据集上实现了 95.18% 的准确率,达到了领域的最优结果。进一步地,在略低于现有模型准确率的基础上,将其参数量缩减了 98%。

关键词: 脑电情感识别;时频双流;多重注意力;轻量级;结构缩放;可计算情感

中图法分类号 TP183

LDM-EEG: A Lightweight EEG Emotion Recognition Method Based on Dual-stream Structure Scaling and Multiple Attention Mechanisms

LEI Ying^{1,3} and LIU Feng^{1,2,3,4}

- 1 School of Computer Science and Technology, East China Normal University, Shanghai 200062, China
- 2 Institute of Artificial Intelligence and Change Management, Shanghai University of International Business and Economics, Shanghai 201620, China
- 3 Institute of AI for Education, East China Normal University, Shanghai 200062, China
- 4 Shanghai Key Laboratory of Mental Health and Psychological Crisis Intervention, School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China

Abstract EEG emotion recognition is a multi-channel time-series signal classification problem with high complexity, high information density and massive data. In order to achieve optimal accuracy and performance of EEG emotion recognition with fewer computational parameters while maintaining the existing classification accuracy, this paper proposes a lightweight network(LDM-EEG) based on dual-stream structural scaling and multiple attention mechanisms. The network takes the time-space and frequency-space maps constructed based on the differential entropy features of EEG signals as the input, processes the two features separately using a symmetric dual-stream structure, achieves lightweighting through a novel parameter-saving residual module and a network scaling mechanism, and enhances the model feature aggregation capability using a novel channel-time/frequency-space multiple attention mechanism and a post-attention mechanism. Experimental results show that the accuracy of the model is 95.18% with significantly reduced number of parameters, which achieves the optimal result in the domain. Further, about 98% reduction in the number of parameters has been achieved with slightly lower accuracy than the existing models.

Keywords EEG emotion recognition, Time-frequency dual streaming, Multiple attention, Lightweight, Structural scaling, Computational affection

1 引言

情感计算是人工智能分支之一,它的概念来源于 MIT 媒体实验室 Picard 教授于 1997 年提出的“计算与情感相关,产生并作用于情感”^[1],目的是使机器能够识别并响应人的情感状态。近年来,情感计算在医疗、教育、交通、军事和娱乐等领域

都取得了广泛的应用。情感可以使用行为和生理信号来表征。相比于面部表情^[2-3]、姿态^[4]、语音等行为信号,生理信号如脑电(EEG)、心电(ECG)、肌电(EMG)、呼吸、血压等,凭借其主观难掩饰性、高时间分辨率及较低成本等优点被逐步用于情感计算领域。其中,脑电是直接采集于大脑的多通道时序数据,是脑神经细胞电生理活动在大脑皮层或头皮表面的总体反映,

基金项目:上海市科技计划项目;中央高校基本科研业务费专项资金(20DZ2260300)

This work was supported by the Research Project of Shanghai Science and Technology Commission and Fundamental Research Funds for the Central Universities of Ministry of Education of China(20DZ2260300).

通信作者:刘峰(lsttoy@163.com)

包含了丰富的情感相关信息。研究表明,不同情绪的脑电信号在时间、频率和空间位置(通道)上存在较大差异,例如,积极情绪下大脑和颞叶区域的激活程度高于消极情绪,消极情绪使得低频段比高频段更加活跃等等^[5]。由此可以看出,脑电可以客观反映个体的不同情绪。因此,近年来脑电情感识别受到了越来越多研究者的关注,成为情感的重要表征之一。

自从 AlexNet^[6] 赢得 ImageNet 挑战,深度学习算法便逐渐得到广泛应用。近年来,凭借着其对特征优秀的自动提取能力,深度学习算法也被应用于脑电情感计算领域。总的趋势是设计更深更复杂的网络,以实现更高的精度,例如 DBN^[5], DGCNN^[7], RGNN^[8] 和 SST-EmotionNet^[9] 等。但更高的准确率并不意味着网络在规模和速度方面的高效。这些深度学习模型往往在训练和推理时更加耗时耗力,对硬件设备提出了很高的要求,同时也限制了其在嵌入式设备和移动端的部署。

为了实现脑电情感识别神经网络的分类精度和参数量的平衡,本文提出了一种全新的基于双流结构缩放和多重注意力机制的轻量级脑电情感识别模型 LDM-EEG。这一模型可有效解决脑电情感识别的问题。本文的主要贡献如下:

(1)在公开数据集上就准确率和参数量这两个指标进行了测试,该模型精度超越现有的主流模型,并且参数量达到最低。同时通过模型缩放,在准确率略低于现有最优结果的情况下,我们的模型减少了 98% 的参数量,提升了模型训练效率。

(2)提出了新型的节省参数的双流可缩放残差模块,使用深度可分离卷积和相同的双流结构精简网络设计,可以有效避免过拟合并明显减少参数量。

(3)提出了新型多重注意力机制,并最早在脑电领域探究了注意力模块的数量和位置对分类精度的影响,有效证明了多重注意力机制和后注意力机制的有效性。

2 相关工作

2.1 脑电情感识别

脑电情感识别的经典步骤一般分为特征提取和分类。常用的特征有时域特征和频域特征。时域特征有事件相关电位(Event Related Potentials, ERP)^[10]、Hjorth 特征^[11]、非平稳指数(Non-Stationary Index, NSI)^[12]等等,频域特征有波段功率(Band Power)^[10-13]、高阶谱(Higher Order Spectra, HOS)^[14]等。其中,微分熵(Differential Entropy, DE)^[15-16]是最常用的脑电情感分类的特征。Zheng 等通过构建深度置信网络,首次证明了 DE 特征是一种准确且稳定的分类特征^[16]。由于电极通道与被试头部采集电极一一对应,因此通道间存在着一定的空间位置关系。相较于简单的堆叠通道信息,构建通道间的拓扑结构或者转换为 2D 形式,可以更好地捕获通道间的关联以及空间位置信息。Song 等通过构建脑电通道间的拓扑关系,更好地提取了通道间的空间位置信息^[7]。Jung 等的工作将脑电信号转化为基于图像表示,取得了不错的分类效果^[17]。

早期的脑电情感分类常用传统的机器学习算法完成,比如最邻近分类器(K-Nearest Neighbor, KNN)^[18]、支持向量机(Support Vector Machine, SVM)^[19]等等。不过这些传统的机器学习算法往往严重依赖于专家的知识 and 经验来设计或者选择特征,且精度也十分有限。而近年来随着大规模可用

脑电数据集的产生以及图形处理单元(Graphic Processing Unit, GPU)的快速发展,脑电情感分类主要采用深度学习完成。这些算法包括擅长提取空间信息的卷积神经网络(Convolutional Neural Networks, CNN)^[20]和图神经网络(Graph Neural Networks, GNN)^[21],以及擅长提取时域信息的循环神经网络(Recurrent Neural Networks, RNN)^[22]等。例如 Zheng 等利用深度信念网络 DBN^[5,16]从脑电中提取高层次表征, Song 等提出动态图卷积神经网络 DGCNN 用于脑电情感识别^[7], Li 等提出一种分层时空神经网络 R2G-STNN 用于学习脑电情绪识别的区域和全局时空特征^[23]。

以上的特征和模型往往只考虑到了时频空三域中的一种或者两种的组合,不能充分对这 3 种特征信息进行提取。而 Jia 等的工作 SST-EmotionNet^[9]使用 DE 作为脑电特征,利用通道的空间位置信息构建 3D 表示并输入到时空-频空双流网络中,充分利用了时频空三域的信息,取得了较好的情感分类效果。本文也是基于此文做的改进工作。

2.2 注意力机制

注意力机制(Attention Mechanism)是一种通过强化重要信息抑制非重要信息提升神经网络性能的一种方法,在自然语言处理^[24]、图像识别^[25]、语音识别^[26]等领域均有广泛的应用。注意力机制本质上和人类对外界事物的观察机制类似。人类在观察事物的时候,首先会比较倾向于关注事物某些重要的局部信息,然后再把不同区域的信息组合,从而形成对事物的总体印象。类似地,注意力机制可以帮助模型对输入 X 的每个部分赋予不同的权重,抽取更加关键且重要的信息,使得模型做出更加准确的判断。

近几年关于注意力机制的研究工作,大多数是使用掩码(Mask)^[27]形成注意力机制。具体来说,就是利用相关特征图学习权重分布,再把学习出来的权重分配到原特征图上。比如, SENet^[28]通过 Squeeze 和 Excitation 模块在特征通道上加入注意力机制, CBAM^[29]结合了特征通道和特征空间两个维度的注意力机制, SKNet^[30]对不同卷积核大小加入注意力机制, FcaNet^[31]提出多光谱通道注意力机制来探讨使用不同频域分量及其组合的影响等。

由于注意力机制本身存在一定的计算和存储的开销,所以选择合理数量的注意力机制及其位置也是一个重要的问题。PP-LCNet^[32]对 SE 注意力机制^[28]分别加在模型前中后及其组合的效果进行了比较,发现注意力模块靠后在同等延迟的情况下比注意力模块较靠前和中间情况能取得更高的准确率, BotNet^[33]和 offTANet^[2]对此进行了一定的实践。

脑电数据是多通道多频段的时序数据,不同通道、频段、时间戳对情感的识别具有一定的差异性^[4],因此在脑电情感识别领域进行注意力的研究很有意义。Jia^[9]在模型中加入时/频-空注意力机制,取得了不错的分类效果。Tao^[34]在模型中加入通道注意力机制,提高了脑电情感识别的分类准确率。Chai^[35]在模型中加入通道和频率注意力,在 DEAP 和 DREAMER 数据集上取得了不错的分类效果。但已有的工作都没有综合探究过通道-时/频-空多重注意力机制,且都是统一在网络的所有模块中加上注意力机制。本文尝试对通道-时/频-空多重注意力机制及其模块的数量和位置进行探究。

2.3 网络轻量化

深度学习模型虽然带来了分类准确率的提升,但参数量

往往较大,训练和推理起来更加耗时耗力。为实现网络精度和开销的平衡,通常采取以精度换取效率来减少模型参数数量的方法,包括模型剪枝^[36]和蒸馏^[37]等。手工设计轻量化网络是这些方法之一,比如 MobileNet 系列^[38-40]主要采用了深度可分离卷积轻量化网络,ShuffleNet 系列^[41-42]主要采用分组卷积的思想轻量化网络等等。此外,神经架构搜索^[43-46]在设计高效轻量网络方面变得越来越流行,通过广泛调整网络宽度、深度、卷积内核类型和大小,实现了比手工设计网络更高的效率。

用于脑电情感分类的已有深度学习模型虽然取得了较为不错的准确率,但很少有考虑参数数量的问题。本文借鉴轻量级网络 MobileNet^[39]的思想设计主干网络,并参考 EfficientNet^[46]的思想对网络的参数进行缩放,以此来实现准确率和参数数量的平衡。用轻量级网络做脑电情感分类的任务,不仅能够加速训练和推理,还有利于设备的端模型部署,提高实时性。

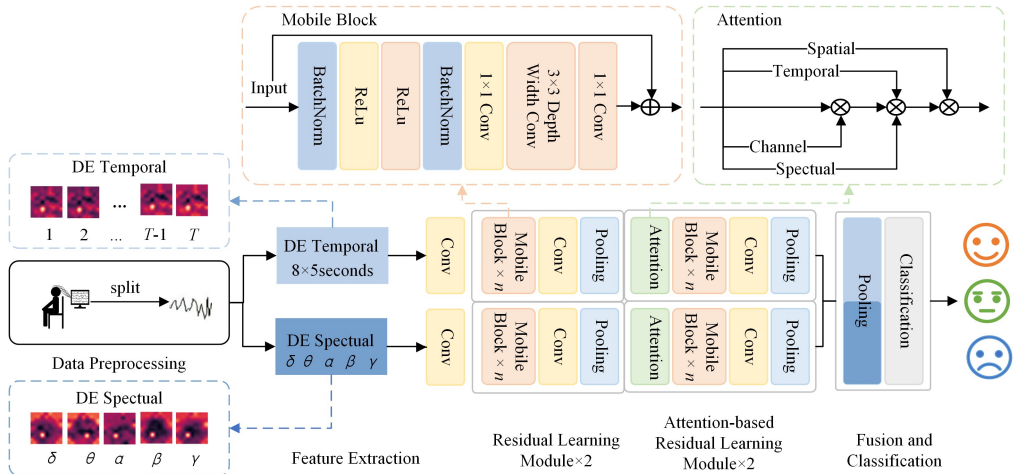


图1 基于注意力机制的时频双流残差学习情感分类模型框架

Fig.1 Framework of attention-based mechanism for time-frequency dual-stream residual learning emotion classification model

3.2 脑电信号的时频 3D 表示

经过数据预处理和切分后,一个数据样本的维度为 62×200 ,其中 62 是通道数,200 为脑电信号的采样频率。此后,数据经过时频域不同方式的特征提取过程,分别输入到时域网络分支和频域网络分支中,具体细节描述如下。

频域分支的特征提取过程中,需要充分利用不同频段的脑电信息。首先将脑电信号按照频率划分为 δ 频段^[48]、 θ 频段^[49]、 α 频段^[50]、 β 频段^[51]和 γ 频段^[52]。第 b 个频段的信号可以表示为 $F_b = (f_b^1, f_b^2, \dots, f_b^{E-1}, f_b^E)$,其中 $b \in \{\delta, \theta, \alpha, \beta, \gamma\}$, E 表示通道数, f_b 表示具有 200 个采样点的单通道时序脑电信号。然后,对每个频段的数据计算出所有通道的 DE 特征。DE 特征由香农信息熵推导出,如式(1)所示。

$$DE = - \int_{-\infty}^{+\infty} f(x) \log(f(x)) dx \quad (1)$$

其中, $f(x)$ 为连续信息的概率信息的概率密度函数。当特定长度的脑电信号 x 近似服从高斯分布 $N(\mu, \sigma^2)$ 时, $f(x)$ 可以表示为:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2)$$

将式(2)代入式(1)中,可以得到 DE 特征公式化简后的结果,如式(3)所示。

3 方法

3.1 模型概述

本文方法的总体框架如图 1 所示。该框架主要分为两部分,一部分是脑电信号的预处理和特征提取模块,另一部分是基于注意力机制的时频双流残差学习情感分类模块。脑电信号的预处理部分包括原始数据的切分和构建脑电时频域的 3D 表示模块。情感分类模块包括对称的时频双流网络结构,分别输入的是脑电信号的时域三维表示和频域三维表示。与 ResNeXt^[47]的设计类似,本网络的时频双流采用相同的拓扑结构,这一设计的目的是防止模型过拟合,提高模型在不同任务不同数据集上的适应性。其中,每个分支包括 4 个阶段的残差学习模块,前 2 个模块不包括注意力机制,后 2 个模块包括注意力机制。最后,网络融合时频双流特征来进行情感分类。

$$DE = \frac{1}{2} \log 2\pi \sigma^2 \quad (3)$$

此时,我们已经将每个通道时序数据转化为 DE 特征。然后,对于每个频段 62 通道的 DE 特征数据,将其按照通道间的相对空间位置关系(如图 2 所示)转化为 9×9 的 2 维数据的形式(如图 3 所示),其中没有数据的位置通过补零来处理。为了使得信息更加精细化,使用三次样条插值(INTER_CUBIC)将 2 维 9×9 的单频段的脑电信号转换为 32×32 维度的 2D 形式。由于共有 5 个频段,最终得到 3D 数据表示,如图 4 所示。

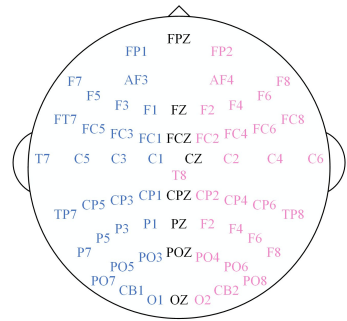


图2 国际 10-20 系统 62 通道的脑电图帽示意图¹⁾

Fig.2 EEG cap according to international 10-20 system for 62 channels

¹⁾ <https://bcmi.sjtu.edu.cn/~seed/seed.html>

			FP1	FPZ	FP2			
			AF3		AF4			
F7	F5	F3	F1	FZ	F2	F4	F6	F8
FT7	FC5	FC3	FC1	FCZ	FC2	FC4	FC6	FT8
T7	C5	C3	C1	CZ	C2	C4	C6	T8
TP7	CP5	CP3	CP1	CPZ	CP2	CP4	CP6	TP8
P7	P5	P3	P1	PZ	P2	P4	P6	P8
	PO7	PO5	PO3	POZ	PO4	PO6	PO8	
		CB1	O1	OZ	O2	CB2		

图3 转换后的2维脑电表示示意图

Fig. 3 Converted 2D EEG representation

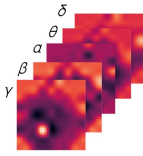


图4 脑电信号在频域上的3D特征图

Fig. 4 3D representation of EEG signal in frequency domain

同理,时域分支的特征提取过程中,需要充分利用不同时间戳的脑电信息。首先将脑电信号按照一定的采样点数量划分为 T 组。第 t 个时间戳的信号可以表示为 $F_t = (f_t^1, f_t^2, \dots, f_t^{E-1}, f_t^E)$, 其中 $t \in \{1, 2, \dots, T-1, T\}$, E 表示通道数, f_t 表示具有 $\frac{200}{T}$ 个采样点的单通道脑电信号。然后对每个时间戳的所有通道数据计算其 DE 特征,再将每个时间戳下的通道 DE 特征数据按照与频域相同的方法转换为 32×32 的 2D 形式。由于共有 T 个时间戳,最终得到 3D 数据表示,如图 5 所示。

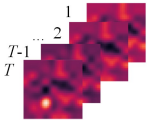


图5 脑电信号在时域上的3D特征图

Fig. 5 3D representation of EEG signal in time domain

3.3 注意力机制

不同网络通道、时间戳、频段和大脑区域上不同情感的激活程度均有一定的差异^[5],因此我们设计了一种通道-时/频-空域的多重注意力机制(如图 6 所示),从而自适应地关注到识别不同情感的特定模式。

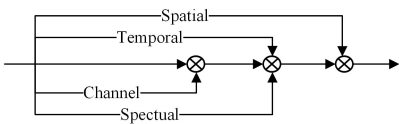


图6 通道-时/频-空域三重注意力机制

Fig. 6 Channel-time/frequency-space triple attention mechanism

具体来说,输入网络的脑电信号可以表示为 $F \in R^{c \times s \times h \times w}$ 或者 $F \in R^{c \times t \times h \times w}$,前者表示输入频域网络流的脑电信号,后者表示输入时域网络流的脑电信号,其中 c 表示通道数, s 表示频段数(初始值为 5), t 表示时间戳个数(初始值为 8), h 和 w 分别表示 2 维脑电信号的长和宽(初始值都为 32)。

3.3.1 通道注意力

通道注意力模块借鉴 SENet^[28] 的思想,即通过学习的

方式自动获取每个特征通道的重要程度,然后依照这个重要程度去提升有用通道并抑制对当前任务用处不大的通道。

具体来说,依次对输入特征图 F 的 w 和 h 维度取平均来聚合空间上的信息,从而生成 $F \in R^{c \times s/t \times 1 \times 1}$,然后再对 s/t 维度取平均聚合频段或者时间戳上的信息,即 $F \in R^{c \times 1 \times 1 \times 1}$ 。最后,通道注意矩阵通过一个带有 softmax 激活功能的全连接层实现。

$$\begin{cases} F_{c,s/t,h} = \text{GAP}_w(F_{c,s/t,h,w}) \\ F_{c,s/t} = \text{GAP}_h(F_{c,s/t,h}) \\ F_c = \text{GAP}_{s/t}(F_{c,s/t}) \\ M_{\text{elect}} = \text{softmax}(F_c) \end{cases} \quad (4)$$

其中, $F_{c,s/t,h}$, $F_{c,s/t}$, F_c 表示依次聚合后的信息, GAP_w , GAP_h , $\text{GAP}_{s/t}$ 表示对相应维度取平均, M_{elect} 表示通道的注意力图。最后,将生成的注意力图 M_{elect} 和输入特征图 F 做乘法,得到最终生成的带有通道注意力机制的特征图 F' ,如式(5)所示。

$$F' = M_{\text{elect}} \odot F \quad (5)$$

3.3.2 频率注意力

不同时间戳上被试表达情绪的强弱是不同的,此外,不同频段对识别不同情绪的重要性也是不同的。为了取得更准确的分类效果,需要对时间戳和频段的重要程度进行自适应的学习,因此我们设计了时/频注意力。

类似地,首先依次对输入特征图 F 的 w 和 h 取平均来聚合空间上的信息,从而生成 $F \in R^{c \times s/t \times 1 \times 1}$,然后再对通道 c 维度上的信息进行平均聚合,得到通道上的信息 $F \in R^{1 \times s/t \times 1 \times 1}$ 。最后,时/频注意矩阵通过一个带有 softmax 激活功能的全连接层实现,如式(6)所示。

$$\begin{cases} F_{c,s/t,h} = \text{GAP}_w(F_{c,s/t,h,w}) \\ F_{c,s/t} = \text{GAP}_h(F_{c,s/t,h}) \\ F_{s/t} = \text{GAP}_c(F_{c,s/t}) \\ M_{s/t} = \text{softmax}(F_{s/t}) \end{cases} \quad (6)$$

其中, $F_{c,s/t,h}$ 和 $F_{c,s/t}$, $F_{s/t}$ 表示依次聚合后的信息, GAP_w 和 GAP_h , GAP_c 表示对相应维度取平均, $M_{s/t}$ 表示时/频域的注意力图。最后,将生成的时/频注意力图 $M_{s/t}$ 和输入特征图 F' 做乘法,得到最终生成的带有时/频注意力机制的特征图 F'' ,如式(7)所示。

$$F'' = M_{s/t} \odot F' \quad (7)$$

3.3.3 空间注意力

在不同的情绪状态下,大脑不同区域的激活程度不同。对此,我们提出了空间注意力机制,进一步提高分类的准确性。

与前面两类注意力的方法相同,首先依次对输入特征图 F 上的 s/t 求平均来聚合时/频域上的信息 $F \in R^{c \times 1 \times h \times w}$,然后对通道维度 c 求平均来聚合通道上的信息 $F \in R^{1 \times 1 \times h \times w}$ 。最后的空间注意矩阵同样通过一个带有 softmax 激活功能的全连接层实现,如式(8)所示。

$$\begin{cases} F_{c,h,w} = \text{GAP}_{s/t}(F_{c,s/t,h,w}) \\ F_{h,w} = \text{GAP}_c(F_{c,h,w}) \\ M_{h,w} = \text{softmax}(F_{h,w}) \end{cases} \quad (8)$$

其中, $F_{c,h,w}$ 和 $F_{h,w}$ 表示依次聚合后的信息, $\text{GAP}_{s/t}$ 和 GAP_c 表示对相应的维度取平均, $M_{h,w}$ 表示生成的空间注意力图。最后,将生成的空间注意力图和输入特征图 F'' 做乘法,得到最终生成带有空间注意力机制的特征图 F''' ,如式(9)所示。

$$F''' = M_{h,w} \odot F'' \quad (9)$$

其中, F''' 是依次经过通道注意力、时/频注意力和空间注意力

机制的最终输出。

3.3.4 后注意力机制

注意力机制能够帮助网络关注到更加有效的信息,但也增加了网络训练和推理的时间,所以在网络中所有的阶段都加上注意力模块显然不是一个最好的选择。对此,我们设计实验来探究注意力模块在网络中的不同相对位置的有效性,最后观察到在4个残差学习模块中,后两个模块带有注意力机制的模型效果比所有模块都带有注意力机制的好,同时参数量也较小。

3.4 残差学习模块

残差学习模块(Residual Learning Module)是双流神经网络的重要组成部分,每个分支网络由一个卷积和4个残差学习模块堆叠而成。其中,依据后注意力机制,只在后2个残差学习模块加上了多重注意力模块,而这种基于注意力机制的残差学习模块除了注意力模块之外的其他结构与不带注意力机制的残差学习模块相同。具体来说,一个残差学习模块的基本构成是 Mobile Block(如图7所示)、卷积、池化以及可能的注意力模块。其中 Mobile Block 由多个相同的残差结构堆叠而成,残差结构的数量被称为深度缩放系数,其大小可以调整,这里取值为2。

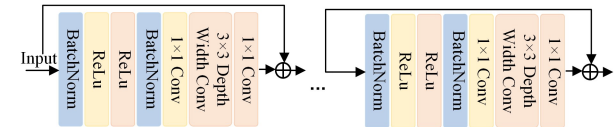


图7 一种基于倒置残差和深度可分离卷积的可缩放 Mobile Block
Fig. 7 Scalable Mobile Block based on inverted residuals and depthwise convolution

以网络中频域分支的第一个残差学习模块为例,列出输入图像的分辨率和特征通道等详细信息,如表1所列。其中,Stage1是残差学习模块前的3D卷积模块,Stage2和Stage3是包含两个堆叠的残差结构的 Mobile Block 模块,Stage4和Stage5分别是3D卷积和平均池化模块。

Mobile Block堆叠的残差结构充分利用了特征的信息,同时解决了深度神经网络难以训练的问题^[53],其基本原理如式(10)所示。

$$\begin{cases} y_l = h(x_l) + F(x_l, W_l) \\ x_{l+1} = f(y_l) \end{cases} \quad (10)$$

其中, x_l 和 x_{l+1} 是第 l 个残差单元的输入和输出; $h(x_l)$ 是恒等函数; F 是残差函数,表示学习到的残差; f 是ReLU激活函数。

表1 残差学习模块的内部结构

Table 1 Internal structure of residual learning module				
Stage	Layer	Operation	Resolution	Channels
1	—	Conv3D	$5 \times 32 \times 32$	4
	1	BatchNorm ReLU	$5 \times 16 \times 16$	4
	2	Conv3D	$5 \times 16 \times 16$	32
	3	BatchNorm ReLU	$5 \times 16 \times 16$	32
	4	DW Conv3D	$5 \times 16 \times 16$	32
2	5	Conv3D	$5 \times 16 \times 16$	4
	1	BatchNorm ReLU	$5 \times 16 \times 16$	8
	2	Conv3D	$5 \times 16 \times 16$	64
	3	BatchNorm ReLU	$5 \times 16 \times 16$	64
	4	DW Conv3D	$5 \times 16 \times 16$	64
3	5	Conv3D	$5 \times 16 \times 16$	8
	1	BatchNorm ReLU	$5 \times 16 \times 16$	8
	2	Conv3D	$5 \times 16 \times 16$	8
	3	BatchNorm ReLU	$5 \times 16 \times 16$	8
	4	Conv3D	$5 \times 16 \times 16$	8
4	—	AvgPooling	$5 \times 8 \times 8$	8

对于残差结构内部,借鉴 MobileNet^[38]的思想,使用深度可分离卷积代替普通的3D卷积,在保证网络分类效果的前提下,使得网络尽可能轻量化。3D卷积一般有3种实现方式,分别是普通3D卷积、Pseudo 3D卷积和3D深度可分离卷积(3D Depthwise Convolution)^[54]。Pseudo 3D卷积基本思想是利用一个核大小为 $1 \times 3 \times 3$ 的2D空间卷积和核大小为 $3 \times 1 \times 1$ 的1D时域卷积来模拟常用的核大小 $3 \times 3 \times 3$ 的3D卷积,如图8所示。而深度可分离卷积是指把普通的3D卷积拆分为逐通道卷积(Depthwise Convolution)和逐点卷积(Pointwise Convolution)。逐通道卷积指的是给每一个输入通道单独使用一个卷积核去进行卷积。逐点卷积类似于普通卷积,只是卷积核的大小为 1×1 ,如图9所示。

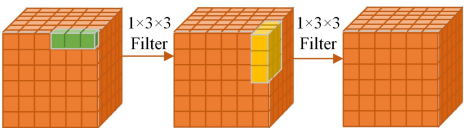


图8 Pseudo 3D卷积
Fig. 8 Pseudo 3D convolution

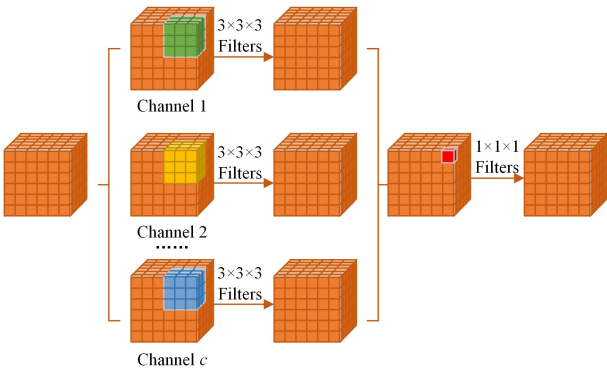


图9 3D深度可分离卷积
Fig. 9 3D depthwise convolution

从参数量的角度进行比较,它们三者的大小关系往往是:普通3D卷积 > Pseudo 3D卷积 > 3D深度可分离卷积。例如,对于形状为 l, w, h 的3D输入数据,拟采用卷积核大小为 k 的3D卷积进行处理。普通3D卷积的参数数量为 $P_{\text{Conv}} = k \times k \times k \times c_b \times c_a \times l_F \times w_F \times h_F$;Peseudo 3D卷积的参数数量为 $P_{\text{P3DConv}} = k \times k \times k \times c_b \times c_b \times l_F \times w_F \times h_F + k \times c_b \times c_a \times l_F \times w_F \times h_F$;3D深度可分离卷积的参数数量为 $P_{\text{DWConv}} = k \times k \times k \times c_b \times l_F \times w_F \times h_F + c_b \times c_a \times l_F \times w_F \times h_F$ 。它们之间的比较如式(11)所示:

$$\begin{aligned} \frac{P_{\text{DWConv}}}{P_{\text{Conv}}} &= \frac{1}{c_a} + \frac{1}{k^3} \\ \frac{P_{\text{DWConv}}}{P_{\text{P3DConv}}} &= \frac{k^3 + c_a}{k^2 \times c_b + k \times c_a} \approx \frac{k}{c_b} \end{aligned} \quad (11)$$

其中, c_b 是输入数据的通道数, c_a 是数据处理过后的通道数, l_F, w_F, h_F 分别是输入数据 F 的长、宽、高。由于以上变量都大于或等于1,可以看到,深度可分离卷积显然小于普通卷积的参数量。一般来说,卷积核的大小 k 也小于输出通道数,所以深度可分离卷积的参数量小于Pseudo 3D卷积的参数量。关于这3种卷积详细的参数量比较,以及分类精度数据,将在消融实验部分进行阐述。

此外,考虑到深度可分离卷积提取到的特征图数量和输入通道数密切相关,为了使得提取到的信息尽可能丰富,采用

倒置残差(Inverted Residual)的策略^[39],即先利用 1×1 卷积使得输入通道数扩大到原来的8倍,经过 3×3 卷积处理后,再通过 1×1 卷积将通道数进行缩小。而普通的残差结构通道数的变化部分是缩小,目的是节省计算量,但在使用深度可分离卷积这样轻量级卷积的情况下,节省参数意义不大,相反会丢失一些有用的信息。此外,Mobile Block后连接着一层卷积层和池化层用于进一步聚合特征信息。

3.5 模型缩放

为了更好地达到参数量和准确率的平衡,借鉴 Efficient-Net^[46]的思想,尝试对网络的一些参数进行缩放。Efficient-Net 选用的缩放参数为输入图片的分辨率(Resolution)、网络的宽度(Width)、网络的深度(Hight),这里的宽度指的是通道的数量,深度指的是网络子模块的堆叠次数。由于脑电输入的2D形式数据的分辨率有限,所以我们可以选择宽度和深度这两个参数进行探究,详细情况将在消融实验部分进行阐述。

3.6 特征融合和分类

网络的双流结构分别提取了脑电的时-空特征和频-空特征,为了获取更高的分类精度,我们将两者融合在一起进行分类操作。在进行特征融合之前,首先需要对各自的双流特征进行自适应池化处理,使得3D特征图的大小转换为 $1\times 1\times 1$ 。然后,在通道维度上,将下采样后的时-空和频-空特征进行拼接,进而送入分类器中进行处理。由于输入的特征维度较大,分类器由两层全连接层构成。此外,我们在两层全连接层中间添加了一个 dropout 操作,从而防止过拟合现象的发生。在模型中,我们采用 cross entropy 作为损失函数来量化预测分布和真实分布之间的差异,如式(12)所示。

$$Loss = - \sum_{c=1}^M p_c \log(q_c) \tag{12}$$

其中, Loss 表示损失, M 表示类别总数, p_c 表示真实分布, q_c 表示模型得到的预测分布。

4 实验

4.1 数据集

在 SEED 数据集^[5,15]上进行实验,来验证所提出方法的有效性。SEED 数据集包含3种情感,分别是积极、中性和消极。这些情感由15个包含3种对应情绪的视频剪辑来激发产生,每段视频的时长设置在不容易让被试疲劳的4min,同时每个视频材料容易被识别且只激发一种期望的目标情绪。在收集数据的过程中,共有15名被试参与,每名被试共有3轮实验,每轮实验需要观看15个视频剪辑。具体来说,观看每个视频剪辑的过程分为4个阶段,包括5s的开始提示,4min的视频时间,45s的自我评估和15s的休息时间。其中,相同情感的视频不会连续显示。在自我评估中,被试要求立即完成问卷,报告他们对每个视频的情绪反应。对于采集到的原始脑电信号,我们需要进行一定的预处理。实验过程中采用国际10-20系统62通道采样频率为1000Hz的脑电设备记录原始脑电信号。在SEED数据集中,提取每段视频剪辑时长对应的数据作为采集到的原始脑电信号,共 $15\times 3\times 15$ 段,其中15,3,15分别表示被试数目、实验轮数和视频数目。同时为了尽可能去除原始脑电信号中的噪声和伪迹,信号被降采样到200Hz,并利用0~75Hz的带通滤波器进行

过滤。最后,脑电信号被划分为200ms的样本,通过傅里叶变换等一系列操作对其提取DE特征,作为模型最终的输入。

4.2 实验配置

本文提出的方法在 NVDIDA RTX 8000 显卡上完成训练和测试。使用 Adam 优化器来优化损失函数,并训练121个轮次。为了兼顾训练效率和训练后期的稳定性,将学习率设置为 5×10^{-4} ,且分别在训练轮次为30,60和90处设置了衰减系数为0.7的学习率衰减,即每到一定的轮次,学习率就变为原来的0.7倍。在每个被试的数据上分别进行实验,每个样本数据随机划分为训练集和测试集,它们的比值是6:4。模型的设置中,一个残差学习模块中的 Mobile Block 数目以及网络流中第一个卷积的输出通道处是可以调节的。所有实验都基于深度学习框架 Pytorch 完成。模型代码可见 <https://github.com/ECNU-Cross-Innovation-Lab/EEG-Emotion>。

4.3 结果分析比较

本文将所提工作和 SEED 数据集上截至2021年的最新工作进行性能比较。参与比较的方法包括:Suyken 等在1999年提出的最小二乘支持向量机模型 SVM^[55];Zheng 等在2016年提出的基于群体稀疏典型相关分析的多通道脑电情感识别模型 GSCCA^[56];Zheng 等在2015年研究关键频段和通道的深度信念网络 DBN^[5];Song 等在2018年对多通道脑电数据进行情感分类的动态图卷积网络 DGCNN^[7];Li 等在2018年提出的双半球域对抗性神经网络 BiDANN^[57];Li 等在2019年提出的从局部到全局的分层特征学习空时神经网络模型 R2G-STNN^[23],以及学习两个半球之间不对称性差异的双半球差异模型 BiHDM^[58];Zhong 等在2019年寻找不同脑区的局部和全局关系时提出的正则化图神经网络 RGNN^[8];Jia 等在2020年提出的基于时空注意力的3维密集型神经网络 SST-EmotionNet^[9]等。

由于本文借鉴了最佳模型 SST-EmotionNet 的一些思想,这里依据论文和它提供 Tensorflow 版本代码的参数设置对其模型的 Pytorch 版本进行了复现,训练部分的优化器、学习率、训练轮次等设置和本模型的实验设置一致。

首先,在模型的分类准确率上进行了比较,如表2所列。这里模型 Mobile Block 的数目设置为2,即一个残差学习模块中包含两个 Mobile Block。这种设置下,模型取得的分类准确率较高。

表2 最新模型在 SEED 数据集上的性能比较

Table 2 Performance comparison of the latest models on SEED dataset

Model	(单位:%)	
	ACC	STD
SVM ^[53]	83.99	9.72
GSCCA ^[54]	82.96	9.95
DBN ^[4]	86.08	8.34
DGCNN ^[6]	90.40	8.49
BiDANN ^[55]	92.38	7.04
BiHDM ^[56]	93.12	6.06
R2G-STNN ^[22]	93.38	5.96
RGNN ^[7]	94.24	5.95
SST-EmotionNet* ^[8]	94.52	2.21
LDM-EEG(Ours)	95.18	2.13

此外,尝试对深度和宽度变量进行缩放。将模型 Mobile

Block 数目 n 设置为 1 和 2,卷积的输出通道数 c 设置为 2,4,8,并与 SST-EmotionNet 的准确率和参数量进行联合比较,如表 3 所列,对应的散点图如图 10 所示。

表 3 模型在准确率和参数量方面的联合比较

Table 3 Joint comparison of models in terms of accuracy and number of parameters

Model	ACC/%	Param/M
SST-EmotionNet[8]	94.52	2.29
LDM-EEG($n=1, c=2$)	90.97	0.01
LDM-EEG($n=1, c=4$)	93.00	0.02
LDM-EEG($n=1, c=8$)	94.12	0.05
LDM-EEG($n=2, c=2$)	94.73	0.25
LDM-EEG($n=2, c=4$)	95.18	0.90
LDM-EEG($n=2, c=8$)	95.16	3.41

表 3 表明,相同宽度设置的情况下,深度为 2 的模型准确率普遍高于深度为 1 的模型,相同深度设置的情况下,准确率大致随着宽度的增加而增加。这个现象可以解释为,模型可以通过增加参数量的方式来换取模型精度的提升。

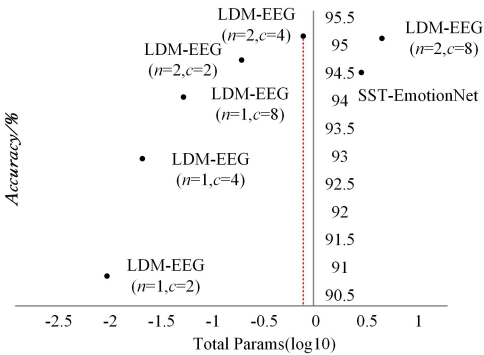


图 10 模型在准确率和参数量方面的比较散点图

Fig. 10 Scatter plot of model comparisons in terms of accuracy and number of parameters

此外,可以注意到固定深度的情况下,深度较小($n=1$)时增加模型宽度对于提升模型精度的效果明显,且能够保持模型的参数量较小,比如 LDM-EEG($n=1, c=8$)组在准确率稍次于当前最优模型 SST-EmotionNet 的情况下参数量约为 SST-EmotionNet 的 1/50,很好地达到了模型参数量和精度的平衡;而深度较大($n=2$)的情况下, $c=8$ 模型的参数量比 $c=4$ 模型的参数量显著增加,但模型精度却大致相同甚至精度略低。因此,这种结果给了我们两点启发:1)可以通过适当降低准确率来换取模型的轻量与高效,特别是在移动端部署的场景;2)在模型达到饱和的情况下,一味增加模型参数量并不能带来模型精度的显著提高,甚至会适得其反。

4.4 消融实验

为了验证模型设计的有效性,设计了以下相关对比及消融实验。

4.4.1 注意力机制

模型采用通道-时/频-空多重注意力机制,在 Jia 等[8]的工作中,时/频注意力机制和空域注意力机制已经得到了有效的验证,这里着重验证通道注意力机制的有效性,实验结果如表 4 所列,其中 NC 代表没有通道注意力,C 代表有通道注意力。可以看到,在时/频-空注意力机制的基础上,加上通道注意力机制带来了一定的性能提升,验证了此设计的有效性。

表 4 通道注意力机制的消融实验结果

Table 4 Results of ablation experiments of channel attention mechanism

Model	ACC	STD
LDM-EEG(NC)	95.03	2.22
LDM-EEG(C)	95.18	2.13

此外,我们还对后注意力机制进行了验证。在其余设置最优的基础上,分别设计了所有残差学习的 4 个阶段都带有多重注意力机制的实验(如表 5 中的 A(1111)所列),以及只有后 2 个阶段带有多重注意力机制的实验(如表 5 中的 A(0011)所列),并对它们的准确率和参数量进行了比较。从表 5 中可以观察到,当去除前 2 个阶段的注意力时,除了参数量显著减小外,准确率也得到了不少的提升。一种可能的解释为:当信息未得到一定的聚合时,前置注意力机制容易学到一些效果不明显的注意力,从而降低模型分类的效果;而当信息经过一定的聚合后,使用后置的注意力机制更加容易捕获信息的重要区域,从而提升分类效果。

表 5 后注意力机制的消融实验结果

Table 5 Results of ablation experiments of post-attention mechanism

Model	ACC/%	Param/M
LDM-EEG($n=1, c=4, A=1111$)	90.09	0.16
LDM-EEG($n=1, c=4, A=0011$)	93.00	0.02
LDM-EEG($n=2, c=4, A=1111$)	93.90	1.04
LDM-EEG($n=2, c=4, A=0011$)	95.18	0.90

4.4.2 3D 卷积的类型

3D 卷积分为普通 3D 卷积、Pseudo 3D 卷积以及深度可分离 3D 卷积[56]。由于普通 3D 卷积的参数量过于庞大,一般不被采用,而使用较为轻量的 Pseudo 3D 卷积以及深度可分离 3D 卷积,如本文采用的是深度可分离 3D 卷积,Jia 采用的是 Pseudo 3D 卷积。为了比较这两个轻量级 3D 卷积的性能,在本文提出的最优框架($n=2, c=4$)下设置了一组实验来探究它们的性能,并记录下对应的分类准确率和参数量,如表 6 所列。

表 6 3D 卷积类型的消融实验结果

Table 6 Results of ablation of 3D convolution types

Model	ACC/%	Param/M
LDM-EEG(Pseudo)	94.76	2.10
LDM-EEG(DW)	95.18	0.90

从表 5 中可以观察到,深度可分离卷积不仅带来了参数量的显著降低,同时也提升了模型的分类准确率,因此验证了该设计的有效性。

结束语 本文提出了一种基于双流结构缩放和多重注意力机制的轻量级架构,有效解决了脑电情感识别的问题。在此基于 MobileNet[41]设计的架构中,时频双流架构、注意力机制和深度可分离 3D 卷积发挥了重要的作用,不仅提升了模型的分类准确率,而且明显降低了模型的参数量。本文还进行了一系列的消融实验,保证了模型中的多重注意力机制、后注意力机制以及深度可分离 3D 卷积[54]等关键设计的有效性。

然而,本文也存在一定的局限性。为了实现模型参数量和模型精度的平衡,我们在深度和宽度方面进行了模型参数的缩放,但手工参与选择仍占据主导。因此在未来的工作中,我们将使用神经架构搜索如 EfficientNet[44]的思想,以更好地使模型自动达到分类精度与参数量的平衡。此外,在参数

量的比较方面,脑电领域很少有公开完整源代码和详细参数配置的工作,因此难以和所有模型的参数量一一进行比较。针对这一局限性,我们在后续的工作中将进行更多的实验来持续完善。

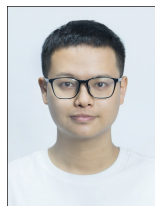
参 考 文 献

- [1] PICARD R W. Affective Computing for HCI[C]// Proceedings of the HCI International 1999-Proceedings of the 8th International Conference on Human-Computer Interaction. Munich, Germany, 1999.
- [2] ZHANG J, LIU F, ZHOU A. Off-TANet: a Lightweight Neural Micro-Expression Recognizer with Optical Flow Features and Integrated Attention Mechanism[C]// Pacific Rim International Conference on Artificial Intelligence. Cham: Springer, 2021: 266-279.
- [3] LIU F, WANG H Y, SHEN S Y, et al. OPO-FCM: A Computational Affection Based OCC-PAD-OCEAN Federation Cognitive Modeling Approach[J]. IEEE Transactions on Computational Social Systems, 2022.
- [4] FU Z, LIU F, ZHANG J, et al. SAGN: Semantic Adaptive Graph Network for Skeleton-Based Human Action Recognition[C]// Proceedings of the 2021 International Conference on Multimedia Retrieval. 2021: 110-117.
- [5] ZHENG W L, LU B L. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks[J]. IEEE Transactions on Autonomous Mental Development, 2015, 7(3): 162-175.
- [6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25.
- [7] SONG T, ZHENG W, SONG P, et al. EEG emotion recognition using dynamical graph convolutional neural networks[J]. IEEE Transactions on Affective Computing, 2018, 11(3): 532-541.
- [8] ZHONG P, WANG D, MIAO C. EEG-based emotion recognition using regularized graph neural networks[J]. arXiv:1907.07835, 2019.
- [9] JIA Z, LIN Y, CAI X, et al. Sst-emotionnet: Spatial-spectral-temporal based attention 3d dense network for eeg emotion recognition[C]// Proceedings of the 28th ACM International Conference on Multimedia. 2020: 2909-2917.
- [10] FRANTZIDIS C A, BRATSAS C, PAPADELIS C L, et al. Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli[J]. IEEE Transactions on Information Technology in Biomedicine, 2010, 14(3): 589-597.
- [11] HJORTH B. EEG analysis based on time domain properties[J]. Electroencephalography and Clinical Neurophysiology, 1970, 29(3): 306-310.
- [12] KROUPI E, YAZDANI A, EBRAHIMI T. EEG correlates of different emotional states elicited during watching music videos [C]// International Conference on Affective Computing and Intelligent Interaction. Berlin: Springer, 2011: 457-466.
- [13] WANG X W, NIE D, LU B L. EEG-based emotion recognition using frequency domain features and support vector machines [C]// International Conference on Neural Information Processing. Berlin: Springer, 2011: 734-743.
- [14] HOSSEINI S A, KHALILZADEH M A, NAGHIBI-SISTANI M B, et al. Higher order spectra analysis of EEG signals in emotional stress states[C]// 2010 Second International Conference on Information Technology and Computer Science. IEEE, 2010: 60-63.
- [15] DUAN R N, ZHU J Y, LU B L. Differential entropy feature for EEG-based emotion classification[C]// 2013 6th International IEEE/EMBS Conference on Neural Engineering(NER). IEEE, 2013: 81-84.
- [16] ZHENG W L, ZHU J Y, PENG Y, et al. EEG-based emotion classification using deep belief networks[C]// 2014 IEEE International Conference on Multimedia and EXPO(ICME). IEEE, 2014: 1-6.
- [17] SIDDHARTH S, JUNG T P, SEJNOWSKI T J. Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing[J]. arXiv:1905.07039, 2019.
- [18] BAHARI F, JANGHORANI A. Eeg-based emotion recognition using recurrence plot analysis and k nearest neighbor classifier[C]// 2013 20th Iranian Conference on Biomedical Engineering(ICBME). IEEE, 2013: 228-233.
- [19] WANG X W, NIE D, LU B L. EEG-based emotion recognition using frequency domain features and support vector machines [C]// International Conference on Neural Information Processing. Berlin: Springer, 2011: 734-743.
- [20] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]// European Conference on Computer Vision. Cham: Springer, 2014: 818-833.
- [21] WU Z, PAN S, CHEN F, et al. A comprehensive survey on graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(1): 4-24.
- [22] GRAVES A, MOHAMED A R, HINTON G. Speech recognition with deep recurrent neural networks[C]// 1988 International Conference on Acoustics, Speech, and Signal Processing (ICASSP-88). IEEE, 2013.
- [23] LI Y, ZHENG W, WANG L, et al. From regional to global brain: A novel hierarchical spatial-temporal neural network model for EEG emotion recognition[J]. IEEE Transactions on Affective Computing, 2022, 13(2): 568-578.
- [24] SOOD E, TANNERT S, MÜLLER P, et al. Improving natural language processing tasks with human gaze-guided neural attention[J]. Advances in Neural Information Processing Systems, 2020, 33: 6327-6341.
- [25] SUN Y, FISHER R. Object-based visual attention for computer vision[J]. Artificial intelligence, 2003, 146(1): 77-123.
- [26] ZHOU P, YANG W, CHEN W, et al. Modality attention for end-to-end audio-visual speech recognition[C]// 2019 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP 2019). IEEE, 2019: 6565-6569.
- [27] PANG Y, XIE J, KHAN M H, et al. Mask-guided attention network for occluded pedestrian detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 4967-4975.
- [28] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7132-7141.
- [29] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]// Proceedings of the European Conference on Computer Vision(ECCV). 2018: 3-19.
- [30] LI X, WANG W, HU X, et al. Selective kernel networks[C]//

- Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019:510-519.
- [31] LIN Z,ZHANG Z,CHEN L Z,et al. Interactive image segmentation with first click attention[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020:13339-13348.
- [32] CUI C,GAO T,WEI S,et al. PP-LCNet: A Lightweight CPU Convolutional Neural Network[J]. arXiv:2109.15099,2021.
- [33] SRINIVAS A,LIN T Y,PARMAR N,et al. Bottleneck transformers for visual recognition[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021:16519-16529.
- [34] TAO W,LI C,SONG R,et al. EEG-based emotion recognition via channel-wise attention and self attention[J]. IEEE Transactions on Affective Computing, 2023,14(1):382-293.
- [35] CHAI B,LI D D,WANG Z,et al. EEG Emotion Recognition Based on Frequency and Convolutional Attention[J]. Computer Science, 2021,48(12):312-318.
- [36] ZHU M,GUPTA S. To prune,or not to prune:exploring the efficacy of pruning for model compression[J]. arXiv:1710.01878, 2017.
- [37] HINTON G,VINYALS O,DEAN J. Distilling the knowledge in a neural network[J]. arXiv:1503.02531,2015.
- [38] HOWARD A G,ZHU M,CHEN B,et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv:1704.04861,2017.
- [39] SANDLER M,HOWARD A,ZHU M,et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018:4510-4520.
- [40] HOWARD A,SANDLER M,CHU G,et al. Searching for mobilenetv3[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019:1314-1324.
- [41] ZHANG X,ZHOU X,LIN M,et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018:6848-6856.
- [42] MA N,ZHANG X,ZHENG H T,et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]// Proceedings of the European Conference on Computer Vision(ECCV), 2018: 116-131.
- [43] ELSKEN T,METZEN J H,HUTTER F. Neural architecture search: A survey[J]. The Journal of Machine Learning Research, 2019,20(1):1997-2017.
- [44] CAI H,CHEN T,ZHANG W,et al. Efficient architecture search by network transformation[C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2018.
- [45] LIU H,SIMONYAN K,YANG Y. Darts: Differentiable architecture search[J]. arXiv:1806.09055,2018.
- [46] TAN M,LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]// International Conference on Machine Learning, PMLR, 2019:6105-6114.
- [47] XIE S,GIRSHICK R,DOLLÁR P,et al. Aggregated residual transformations for deep neural networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:1492-1500.
- [48] KNYAZEV G G. EEG delta oscillations as a correlate of basic homeostatic and motivational processes[J]. Neuroscience & Biobehavioral Reviews, 2012,36(1):677-695.
- [49] MUSAEUS C S,ENGEDAL K,HØGH P,et al. EEG theta power is an early marker of cognitive decline in dementia due to Alzheimer's disease[J]. Journal of Alzheimer's Disease, 2018, 64(4):1359-1371.
- [50] FINK A,ROMINGER C,BENEDEK M,et al. EEG alpha activity during imagining creative moves in soccer decision-making situations[J]. Neuropsychologia, 2018,114:118-124.
- [51] GOLA M,MAGNUSKI M,SZUMSKA I,et al. EEG beta band activity is related to attention and attentional deficits in the visual performance of elderly subjects[J]. International Journal of Psychophysiology, 2013,89(3):334-341.
- [52] MAFFEI A,SPIRONELLI C,ANGRILLI A. Affective and cortical EEG gamma responses to emotional movies in women with high vs low traits of empathy[J]. Neuropsychologia, 2019,133: 107175.
- [53] HE K,ZHANG X,REN S,et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016:770-778.
- [54] YE R,LIU F,ZHANG L. 3D depthwise convolution: reducing model parameters in 3D vision tasks[C]// Canadian Conference on Artificial Intelligence. Cham:Springer, 2019:186-199.
- [55] SUYKENS J A K,VANDEWALLE J. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999, 9(3):293-300.
- [56] ZHENG W. Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis[J]. IEEE Transactions on Cognitive and Developmental Systems, 2016,9(3):281-290.
- [57] LI Y,ZHENG W,CUI Z,et al. A Novel Neural Network Model based on Cerebral Hemispheric Asymmetry for EEG Emotion Recognition[C]//IJCAI, 2018:1561-1567.
- [58] LI Y,WANG L,ZHENG W,et al. A novel bi-hemispheric discrepancy model for eeg emotion recognition[J]. IEEE Transactions on Cognitive and Developmental Systems, 2020, 13(2): 354-367.



LEI Ying, born in 2000, is a student member of China Computer Federation. Her main research interests include affective computing, EEG emotion recognition and deep learning.



LIU Feng, born in 1988. Ph.D candidate, engineer, is a senior member of China Computer Federation. His main research interests include deep learning, computational affection and block-chain technology.