

基于 CNN-Transformer-GCN 的脑电情绪识别研究

许 韵¹ 韩伟业¹ 秦 伟^{1*}

XU Yun HAN Weiye QIN Wei

摘 要

针对现有脑电情绪识别分析中时空信息提取不完备的问题,文章提出一种基于卷积神经网络(convolutional neural networks, CNN)、注意力机制(Transformer)和图卷积网络(graph convolution network, GCN)的组合模型用于脑电信号(electroencephalogram, EEG)解码。此模型包含时空特征提取、图结构特征提取和情绪分类模块:预处理后的脑电信号,先经 CNN 提取局部时空特征,再由 Transformer 编码器实现全局时间关注;接着通过 GCN 自适应学习通道拓扑连接,强化多通道联系表征;最后经线性映射完成情绪识别分类。在 EMO 和 SEED 数据集上验证发现, CNN-Transformer-GCN 模型平均准确率优于 EEGNet、EEG-Conformer 等模型,为脑电情绪识别研究提供了新的研究思路。

关键词

脑电信号; CNN; Transformer; 图卷积; 情绪识别

doi: 10.3969/j.issn.1672-9528.2025.08.043

0 引言

脑机接口(brain-computer interface, BCI)^[1]通过在大脑和外部设备之间建立直接通信,极大促进了人机交互的发展。目前,BCI 技术已在心理健康检测、工作负荷以及医疗康复等领域得到广泛应用^[2-5]。BCI 系统可借助多种神经影像模式实现,包括立体脑电图(SEEG)、功能性磁共振成像(fMRI)、脑电图(electroencephalogram, EEG)和功能性近红外光谱(fNIRS)。

脑电信号因其非侵入性、抗干扰性以及便捷的数据采集特性,成为情绪识别研究与应用的核心焦点。情绪识别作为 BCI 领域的重要研究方向,旨在通过对脑电信号的分析,精准解读个体的情绪状态。这对于理解人类情感、开发情感交互系统以及辅助心理健康治疗等方面具有重要意义。

早期脑电情绪识别分析常用支持向量机(SVM),它在小样本、非线性分类有优势,但处理高维复杂脑电信号时,难以适应大规模数据和复杂模型。深度学习兴起后,循环神经网络(RNN)及其变体 LSTM^[6]、门控循环单元(GRU)被应用于脑电情绪识别分析,RNN 能捕捉时间依赖关系,但存在梯度消失问题,LSTM 和 GRU 解决了该问题,不过在处理空间特征和情绪相关的复杂脑电模式时存在局限。

卷积神经网络(convolutional neural networks, CNN)能有效提取脑电信号的局部时间特征。以 EEGNet^[7]为例,其引入深度卷积和可分离卷积,在减少模型参数的同时,在多种 BCI 范式中展现出一定的跨范式泛化能力。然而,EEGNet

因卷积核局部感受野特性,主要聚焦局部时间特征,难以捕捉脑电信号中的全局长程依赖关系和整体特征。并且,在处理复杂脑电信号特征时,对于高度个性化和复杂的脑电模式,EEGNet 还需要进一步优化模型结构以提高特征提取的完整性和准确性。

为解决 EEGNet 在获取全局信息上的不足,Transformer 架构^[8]受到广泛关注。Transformer 基于自注意力机制,能对输入序列全局关注,有效学习全局信息。Song 等人^[9]提出的 EEG-Conformer 模型,结合卷积神经网络和自注意力机制,在提取 EEG 信号的局部和全局特征方面取得一定成效,提升了相关 EEG 解码任务的性能。但 Transformer 侧重时序特征,缺乏对脑电数据空间维度信息的利用,限制了其在脑电情绪识别中的表现。

图卷积网络(graph convolution network, GCN)^[10]作为一种处理图结构数据的有效工具,在脑电信号分析领域也逐渐得到应用。基于谱图卷积的一阶近似,通过对邻接矩阵进行归一化处理,实现了在图结构数据上的信息传播和特征学习。在脑电信号分析相关研究中,GCN 能够自动获取脑区通道间的联系,为脑电信号分析提供了新的视角。例如,Yin 等人^[11]提出的浅层图卷积神经网络模型(GCNN),考虑了脑电通道空间位置对情绪识别性能的影响;高越等人^[12]提出的基于时空自适应的图卷积神经网络模型(SABGCN-ST),通过自适应学习脑网络的拓扑连接矩阵,进一步提升了图网络对多通道脑电信号间联系的表征能力。然而,在脑电信号分析领域通常采用的浅层图卷积网络,由于特征传播能力有限,只能考虑到脑电信号局部脑区间的联系。而加深图卷积

1. 五邑大学 广东江门 529020

网络的层数虽可能捕捉全局脑电通道间的连接信息，但又会导致深层图卷积网络出现节点特征过度平滑的问题，使得模型无法学习到有效的情绪相关特征。

为解决上述问题，本文提出一种基于 CNN、Transformer 和 GCN 的组合模型，期望充分整合三者的优势，突破现有技术瓶颈，为脑电情绪识别研究开辟新的路径。

1 数据集和数据预处理

1.1 数据集

本文采用了两个情绪识别数据集作为实验数据。

(1) EMO 数据集是实验室自采数据集，包含 8 名被试的 EEG 数据。在本实验中，使用了 62 个 EEG 电极，电极通道符合 10-20 国际标准导联系统，采样率为 1 000 Hz。

每个被试会进行 3 次实验，每次观看不同的视频，每个视频均包含 15 个片段 (trial)，积极、中性和消极情绪各 5 个片段。在一个视频中，每个影片播放之前有 5 s 的黑屏，影片持续时间约为 3~4 min，而每个影片之后总计有 60 s 广告时间，包含 45 s 的自我评估时间和 15 s 的休息时间。为获得反馈，参与者被告知在观看每个影片后立即完成调查问卷，以报告其对每个影片的情绪反应。实验范式的详细过程如图 1 所示。

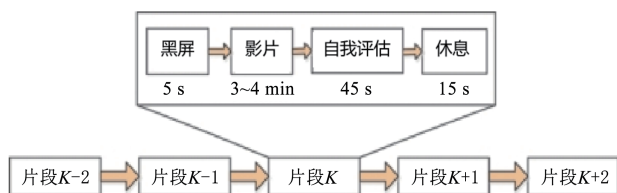


图 1 实验范式详细过程

(2) SEED 数据集^[13]由上海交通大学教授吕宝粮于 2015 年 10 月发布。该数据集包含 15 名被试的 EEG 数据，每位被试均体验了 3 类情绪类型：积极、中性和消极。每个被试进行 3 次实验，每次观看相同的视频。数据集采集的通道数为 62，采样率为 1 000 Hz。

1.2 数据预处理

按照文献 [14] 的方法，对脑电图预处理主要包括将脑电信号降采样至 200 Hz，去除基线，在 1~55 Hz 范围内进行滤波，并应用独立成分分析 (ICA) 来消除眼电图 (EOG) 伪影。进一步将 EEG 信号取 3 s 时间窗口，步长为 1 s，以评估实时 BCI 性能。

2 CNN-Transformer-GCN 模型

在脑电情绪识别分析中，同时包含局部时间特征、全局时间特征与多通道复杂联系，这些信息对理解心理状态、认知决策等大脑复杂功能至关重要。因此，建模时间和通道维

度上的依赖关系，可以更有效地提取信号的复杂时空特征，从而提升情绪识别准确率。

本文提出融合 CNN、Transformer 和 GCN 的创新组合模型，能有效应对这些挑战。如图 2 所示，本模型包括时空特征提取、图结构特征提取和分类 3 个模块。具体来说，数据经采集和预处理后，先经由 CNN 部分提取脑电信号的局部时空特征，再经 Transformer 编码器对输入序列进行全局时间关注；接着通过 GCN 自适应学习通道的拓扑连接，强化对脑电信号多通道间联系的表征；最后经线性映射进行情绪识别任务分类。

2.1 时空特征提取模块

大脑中认知过程的变化反映在不同时间戳和大脑区域的激活水平上。假设 $X \in \mathbb{R}^{C \times S}$ 是来自 EEG 的输入，其中， C 是通道的数量， S 是采样点的数量。

受 EEG-Conformer 的启发，提出了一个时空特征提取模块。具体来说，时空特征提取模块主要包括两个阶段：CNN 层和 Transformer 编码层。

在 CNN 层中，如图 2 所示，Conv2D-S 层关注时间维度，该层输入和输出通道数分别为 1、 K ，卷积核大小为 (1, 25)，步长为 (1, 1)；Conv2D-C 层关注空间维度，该层输入与输出通道数都为 K ，卷积核大小为 (C , 1)，步长为 (1, 1)；随后进行批归一化 (BatchNorm2D) 和 ELUs 激活；第三层 (AvgPool2D) 进行平均池化，以减少过拟合和复杂性，该层输入和输出通道数都为 K ，核大小为 (1, 75)，步长为 (1, 25)；最后，通过卷积 (Conv2D) 和转置 (Rearrange) 对特征进行重新排序，数据维度从 (K , 1, M) 变为 (M , K)。其中， M 表示时间点的数量， K 表示卷积核数量。

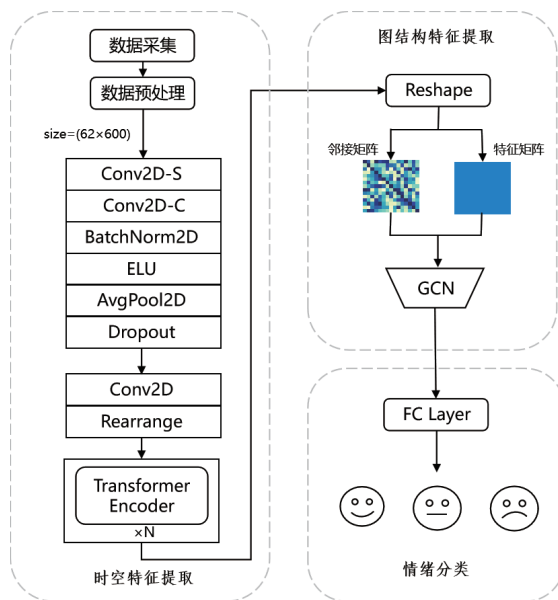


图 2 CNN-Transformer-GCN 模型的总体架构

通过上述 CNN 层, 得到 $\tilde{\mathbf{X}} \in \mathbb{R}^{M \times K}$ 作为初步特征, 输入到后续的 Transformer 编码层中。

首先将 $\tilde{\mathbf{X}}$ 嵌入到 3 个空间中, 分别记为键 $\mathbf{K} = \text{LN}(\tilde{\mathbf{X}})$ 、值 $\mathbf{V} = \text{LN}(\tilde{\mathbf{X}})$ 和查询 $\mathbf{Q}^{(l)} = \text{LN}(\tilde{\mathbf{X}})$, 而第 $l+1$ 层的查询定义为 $\mathbf{Q}^{(l+1)} = \text{LN}(\tilde{\mathbf{X}}^{(l+1)})$ 。第 l 层的 Transformer 定义用公式表示为:

$$\mathbf{U}^{(l)} = \text{softmax}(\mathbf{Q}^{(l)} \mathbf{K}^T) \mathbf{V} + \mathbf{Q}^{(l)} \quad (1)$$

$$\tilde{\mathbf{X}}^{(l+1)} = \text{FC}(\text{LN}(\mathbf{U}^{(l)})) + \mathbf{U}^{(l)} \quad (2)$$

式中: LN 表示层归一化; FC 表示前馈层; $l = 1, 2, \dots, N$, N 表示层数。同时采用了多头注意力策略来进一步增强表示的多样性。随后, 经过 Reshape 层将输出表示为特征矩阵 $\mathbf{H} = (\tilde{\mathbf{X}}^{(N+1)})^T$ 。

2.2 图结构特征提取模块

图卷积网络的具体实现如下:

对于在时空特征提取模块后获得的特征矩阵 $\mathbf{H} \in \mathbb{R}^{K \times M}$, 为每个样本动态构建了一个图结构 $\mathcal{G} = (\mathbf{A}, \mathbf{H})$ 以学习通道之间的关联。

根据文献 [15], $\mathcal{G}$ 中的邻接矩阵 $\mathbf{A}$ 由基础邻接矩阵与可训练掩码逐元素相乘后, 经 ReLU 激活函数处理得到。进一步, 对邻接矩阵进行归一化, 记为 $\tilde{\mathbf{A}}$ 。

利用增强特征矩阵 $\mathbf{H}$ 和归一化邻接矩阵 $\tilde{\mathbf{A}} \in \mathbb{R}^{K \times K}$, 将 GCN 层表示为:

$$\mathbf{Z} = \Phi_{\text{ReLU}}(\tilde{\mathbf{A}}(\mathbf{H}\mathbf{W} - \mathbf{b})) \quad (3)$$

式中: $\mathbf{W} \in \mathbb{R}^{M \times M_{\text{out}}}$ 为权值矩阵, $\mathbf{b}$ 为偏差向量。

2.3 情绪分类模块

将图结构特征提取模块的输出 $\mathbf{Z} \in \mathbb{R}^{K \times M_{\text{out}}}$ 转换为长度为 $d = K \times M_{\text{out}}$ 的向量 $\mathbf{v} \in \mathbb{R}^{1 \times d}$, 接着进行线性变换, 得到类预测:

$$\mathbf{y} = \text{softmax}(\mathbf{v}\mathbf{W}' + \mathbf{b}') \quad (4)$$

式中: $\mathbf{W}' \in \mathbb{R}^{d \times G}$, $\mathbf{y} \in \mathbb{R}^{1 \times G}$, 这里 G 为类别数。

2.4 任务损失

使用标签平滑 [16] 防止模型过拟合并提高泛化能力。假设训练过程中每个批次包含 B 个样本, 模型的预测输出为 $\hat{\mathbf{Y}} = [\hat{y}_{b,g}] \in \mathbb{R}^{B \times G}$ 。计算每个批次样本的交叉熵损失, 并取批次内的平均值:

$$\mathcal{L}_{\text{task}} = \frac{1}{B} \sum_{b=1}^B \sum_{g=1}^G -y_{b,g}^{LS} \log(\hat{y}_{b,g}) \quad (5)$$

式中: $\hat{y}_{b,g}$ 是批次中第 b 个样本在类别 g 上的预测概率; $y_{b,g}^{LS}$ 是同一样本和类别的平滑真实概率。

3 实验

3.1 实验设置

为评估提出的 CNN-Transformer-GCN, 如第 1 节所述, 使用 EMO 和 SEED 作为基准数据集。具体来说, 采用 5 折交叉验证 (5-fold-CV) 的方法来验证个体差异和模型性能。为每个被试分别构建一个模型, 每次将 1/5 的数据作为测试集, 其余数据作为训练集, 随后重复这一过程。持续进行上述操作, 直至所有被试都完成测试。报告的结果是所有被试的平均值。使用准确率 (ACC) 和 F_1 -分数 [17] 来衡量模型的性能。

为控制某些可变因素, 将批处理大小 (batch size) 设置为 60, 卷积核的数量 K 设置为 40, 训练的轮次 (epoch) 设置为 120, 采用 AdamW 优化器, 初始化的学习速率为 0.001。为加快训练和提高性能, 我们在特定时期降低了 Flooding [18] 水平, 以保持持续的正则化效应。

3.2 对比方法

为将本模型与现有的深度神经网络进行比较, 采用了以下方法对比本模型:

- (1) LSTM: 通过门控机制处理序列数据的长期依赖。
- (2) EEGNet: 包含卷积层、深度可分离的卷积和批处理归一化结构, 它既高效又轻量级。
- (3) TSCception [19]: 融合多尺度卷积和注意力机制, 从时间序列数据中提取不同层次的特征进行分类。
- (4) EEG-Conformer: 融合了卷积神经网络 (CNN) 的局部特征提取能力和 Transformer 的全局信息建模能力, 通过特定的模块设计和连接方式, 以实现 EEG 信号中有效信息的高效提取和分类等任务。

表 1 和表 2 分别展示了在 EMO 和 SEED 数据集上的实验结果。结果表明, 本模型的平均准确率和 F_1 -分数比另外 4 个模型都高, 且具有较小的方差。相比 EEGNet 模型, 本模型在两个数据集上的平均准确率分别高出 2.75% 和 3.93%; 相比 TSCception 模型, 本模型分别高 0.74% 和 8.08%; 本模型比 EEG-Conformer 模型分别高 1.28% 和 0.40%; 这个结果说明本模型在脑电情绪识别准确率上有提升。

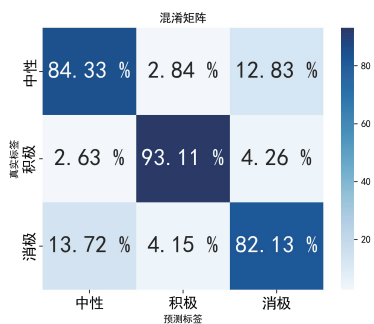
表 1 不同模型在 EMO 数据集上平均准确率和 F_1 -分数

模型	平均准确率 /%	F_1 -分数 /%
LSTM	68.07 ± 4.38	67.48
EEGNet	95.83 ± 5.55	95.67
TSCception	97.84 ± 3.39	97.84
EEG-Conformer	97.30 ± 2.16	97.28
CNN-Transformer-GCN (Ours)	98.58 ± 1.75	98.41

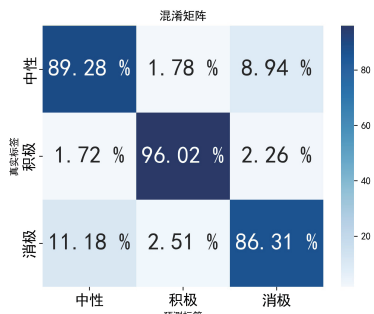
表 2 不同模型在 SEED 数据集上平均准确率和 F_1 - 分数

模型	平均准确率 /%	F_1 - 分数 /%
LSTM	50.44 ± 9.96	50.38
EEGNet	88.41 ± 5.54	88.32
TSCeption	84.26 ± 6.44	84.23
EEG-Conformer	91.94 ± 3.19	91.93
CNN-Transformer-GCN (Ours)	92.34 ± 3.72	92.33

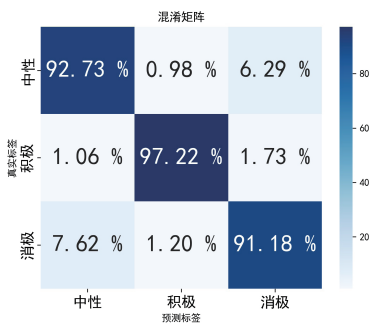
模型在 SEED 数据集上的混淆矩阵如图 3 (a)~(c) 所示, 通过对 TSCeption、EEGNet 和 CNN-Transformer-GCN (Ours) 模型的混淆矩阵进行分析可以看出, 3 个模型都有较好的识别性能。



(a) TSCeption



(b) EEGNet



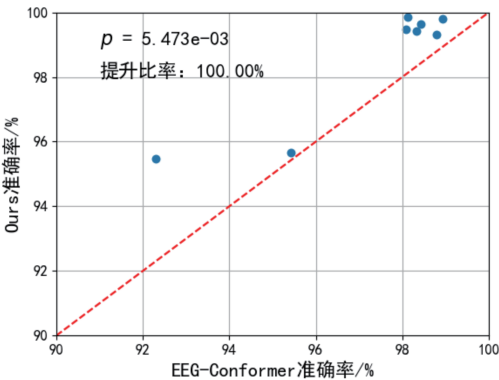
(c) CNN-Transformer-GCN (Ours)

图 3 在 SEED 数据集上的混淆矩阵

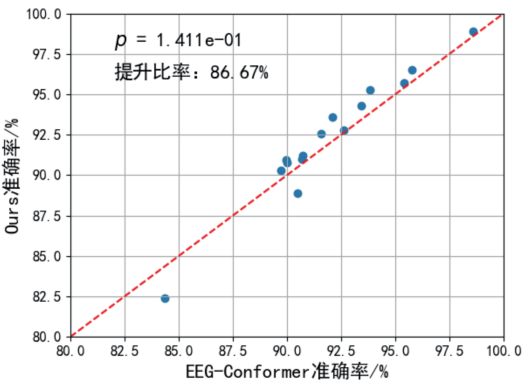
其中, 本模型对积极情绪的正确预测率为 97.22%, 高于 TSCeption 和 EEGNet 模型, 显示出更好的区分能力。对于

中性情绪, 本模型正确率最高, 为 92.73%, 整体表现稳定。在消极情绪识别上, 3 个模型结果有一定差异, 本模型表现相对较好, 为 91.18%。总体来看, 本模型在积极情绪上表现最佳, 未来可进一步优化其他类别情绪的识别准确性, 以提升模型的整体性能。

如图 4 (a)~(b) 所示, 对本模型与 EEG-Conformer 模型进行了 t 检验对比。圆点位置代表参与实验的被试, 虚线对角线上方的圆点表明这些被试运用本模型后, 分类性能有所提高。提升比率表示采用本方法后, 分类性能改善的被试数量占总被试数量的百分比, 配对 t 检验的 p - 值标注在图的左上角。在 EMO 数据集中, 本模型性能优于 EEG-Conformer 模型的被试比例为 100% ($p=5.55e-03$); 在 SEED 数据集中, 这一比例为 86.7% ($p=1.41e-01$)。这些结果表明, 本模型在情绪识别任务上相较于 EEG-Conformer 模型有显著性能提升, 体现了本模型的有效性和优越性。



(a) EMO 数据集



(b) SEED 数据集

图 4 EEG-Conformer 模型 (x 轴) 与 Ours 模型 (y 轴) 的散点图

4 结论

本文针对现有脑电情绪识别分析中时空信息提取不完备的问题, 提出基于 CNN、Transformer 和 GCN 的组合模型用于脑电情绪识别解码。通过该模型对时间和通道维度上的依

赖关系进行建模,有效提取信号的复杂时空特征,以应对脑电情绪识别处理中存在的个体差异及复杂信息提取难题。实验结果表明,该模型平均准确率在自采 EMO 数据集和公开 SEED 数据集上均优于现有模型,为脑电情绪识别研究提供新思路,对推动脑电情绪识别研究具有重要意义。

参考文献:

- [1] 朱文明,高诺.脑机接口技术研究概述[J].信息技术与信息化,2008(6):26-28.
- [2] 谢黎,史丰硕,王雨晴,等.基于通道选择的孤独症儿童 EEG 情绪测量方法研究[J].信息技术与信息化,2024(5):120-122.
- [3] 柯清超,王朋利.脑机接口技术教育应用的研究进展[J].中国电化教育,2019,10:9.
- [4] HUANG J Y, VONG C M, LI C, et al. HSA-Former: hierarchical spatial aggregation transformer for EEG-based emotion recognition[J]. IEEE transactions on computational social systems, 2025:1-12.
- [5] CHEN G J, LIU Y, ZHAN XY, et al. EEG-fNIRS-based emotion recognition using graph convolution and capsule attention network[J]. Brain sciences, 2024,14(8):820.
- [6] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural computation, 1997,9(8):1735-1780.
- [7] 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):56013.
- [8] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. NewYork:ACM,2017:6000-6010.
- [9] SONG Y H, ZHENG Q Q, LIU B C, et al. EEG Conformer: convolutional transformer for EEG decoding and visualization[J]. IEEE transactions on neural systems and rehabilitation engineering, 2022,31:710-719.
- [10] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[EB/OL].(2017-02-22)[2024-02-12].<https://doi.org/10.48550/arXiv.1609.02907>.
- [11] YIN Y Q, ZHENG X W, HU B, et al. EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM[J]. Applied soft computing, 2021:106954.
- [12] 高越,傅湘玲,欧阳天雄,等.基于时空自适应图卷积神经网络的脑电信号情绪识别[J].计算机科学,2022,49(4):30-36.
- [13] 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.
- [14] SUN Z, HUANG Z H, DUAN F, et al. A novel multimodal approach for hybrid brain-computer interface[J]. IEEE access, 2020,8:89909-89918.
- [15] DING Y, ROBINSON N, ZENG Q H, et al. LGGNet: learning from local-global-graph representations for brain-computer interface[J].IEEE transactions on neural networks and learning systems, 2024,35(7):9773 - 9786.
- [16] MULLER R, KORNBLITH S, HINTON G. When does label smoothing help?[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. NewYork:ACM,2019:4694-4703.
- [17] WANG Z H, FANG J H, ZHANG J. Rethinking delayed hemodynamic responses for fNIRS classification[J]. IEEE transactions on neural systems and rehabilitation engineering, 2023,31:4528-4538.
- [18] WANG Z H, ZHANG J, ZHANG X C, et al. Transformer model for functional near-Infrared spectroscopy classification[J]. IEEE journal of biomedical and health informatics, 2022,26(2):2559-2569.
- [19] DING Y, ROBINSON N, ZHANG S, et al. TSception: capturing temporal dynamics and spatial asymmetry from EEG for emotion recognition[J].IEEE transactions on affective computing, 2023,14(3):2238 -2250.

【作者简介】

许韵(2000—),女,安徽六安人,硕士研究生,研究方向:类脑计算。

韩伟业(2001—),男,广东汕尾人,硕士研究生,研究方向:类脑计算。

秦伟(1979—),通信作者(email:qinw05080@gmail.com),男,湖南益阳人,博士,研究方向:机器学习、图像识别、多源信息融合等。

(收稿日期:2025-02-27 修回日期:2025-07-02)