

协同聚合优先的频域约束端到端脑电情绪识别方法

闭应洲^{1,2}, 袁月^{1,2}, 刘善锐³, 周淑姮^{1,2}, 霍雷刚^{1,2}, 甘秋静^{1,2}

(1. 南宁师范大学人工智能学院, 南宁 530199; 2. 南宁师范大学广西人机交互与智能决策重点实验室, 南宁 530199; 3. 广西职业师范学院计算机与信息工程学院, 南宁 530007)

摘要: 脑电 (Electroencephalogram, EEG) 情绪识别是新一代人机交互与脑机接口 (Brain-computer interface, BCI) 系统中的关键技术, 在国家脑科学战略与智能医疗等领域具有重要应用前景。然而, EEG 信号固有的非平稳性、低信噪比以及医学样本稀缺性, 使得现有深度学习方法在真实应用场景下面临显著挑战。现有主流方法通常采用在原始单通道信号上直接进行频域特征提取的策略, 该过程不仅容易将高频噪声误判为有效特征, 还忽视了大脑不同功能区之间的空间协同机制, 导致模型泛化能力受限。针对上述问题, 本文提出了一种协同聚合优先的端到端脑电情绪识别框架, 构建了基于聚合卷积的协同频域特征提取网络 (Synergy-frequency aggregation network, SFAN)。该方法突破传统“先频域提取、后空间融合”的处理范式, 首先通过协同聚合卷积模块对多通道 EEG 信号进行自适应加权融合, 模拟脑区间的动态协同关系, 生成高鲁棒性的协同表征; 随后在此基础上进行频域特征提取与深度时空建模。该策略在频域分析前有效抑制了通道级噪声, 并通过多样化的聚合模式显著丰富特征空间, 从而缓解了小样本条件下的过拟合问题。在 SEED-IV 数据集上取得了 97.42% 的分类准确率, 在 SEED-V 数据集上取得了 94.91% 的分类准确率, 实验结果表明, 所提出的方法在跨被试与跨会话情绪识别任务中均取得了优于现有方法的分类性能, 验证了在频域分析前引入脑区协同建模对于提升 EEG 情绪识别鲁棒性的有效性。

关键词: 脑电; 情绪识别; 深度学习; 协同聚合; 频域建模

中图分类号: TP391; R318

文献标志码: A

引用格式: 闭应洲, 袁月, 刘善锐, 等. 协同聚合优先的频域约束端到端脑电情绪识别方法[J]. 数据采集与处理, 2026, 41(5): 1535-1547. BI Yingzhou, YUAN Yue, LIU Shanrui, et al. A Synergy-aggregation-first frequency-domain constrained end-to-end EEG emotion recognition method[J]. Journal of Data Acquisition and Processing, 2026, 41(5): 1535-1547.

引言

随着人工智能与神经科学的深度融合, 脑机接口技术正逐步从实验研究走向实际应用。作为一种能够非侵入式反映大脑功能状态的信号形式, 脑电 (Electroencephalogram, EEG) 在情绪识别、智能医疗康复、高负荷作业人员状态监测以及新一代人机交互系统中展现出重要价值^[1-7]。在中国“十四五”规划及展望“十五五”规划的战略背景下, 脑科学与类脑智能被确立为重点前沿研究方向, 这不仅为 EEG 情绪识别研究提供了持续的政策与资源支持, 也对相关方法在复杂真实应用场景下的稳定性、鲁棒性与泛化能力提出了更高要求。

EEG 是一种典型的多通道非平稳时序信号 (图 1(a)), 其频谱振荡特性与人类情绪和认知状态密切

基金项目: 国家自然科学基金 (62067007); 广西学位与研究生教改课题 (JGY2023236); 广西科技创新平台计划项目 (桂科 LT2600640001)。

收稿日期: 2026-01-21; **修订日期:** 2026-07-06

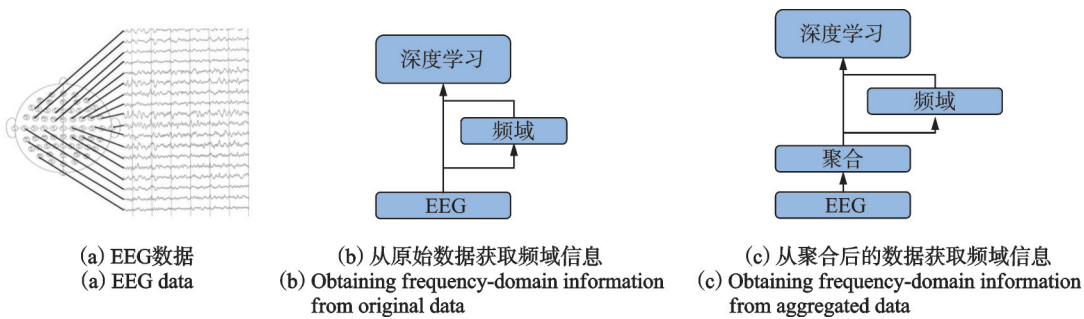


图1 EEG情绪识别中频域信息的获取方法

Fig.1 Methods for obtaining frequency-domain information in EEG emotion recognition

相关^[8-9]。大量神经生理学研究表明,不同频段的神经振荡在情绪加工过程中具有明确的功能意义,如表1所示, θ 波(4~8 Hz)与认知控制和记忆活动相关, α 波(8~13 Hz)反映放松与注意力调节状态,而 β 波与 γ 波则更多参与高级认知与感知整合过程。因此,频域特征被广泛视为EEG情绪识别中最重要的生理先验之一。基于这一认识,现有研究(图1(b))通常在原始EEG信号上直接进行快速傅里叶变换或小波变换的方法,提取功率谱密度、差分熵等频域特征,并结合传统机器学习或深度神经网络完成情绪分类任务。例如,张学军等^[10]提出了融合频域特征的三维卷积神经网络3D-CTM,将其用于脑电情感识别;Chen等^[11]构建了多频率-空间特征融合的EEG情绪识别方法,并采用多输入CNN(MPICNN)进行分类;Xu等^[12]提出了基于Transformer的频谱-空间-时序联合建模框架;Zhong等^[13]则综合利用时域、频域和空间特征,并与传统机器学习方法进行系统对比,进一步验证了多域特征融合在情绪识别任务中的有效性。

尽管上述“直接频域提取”的方法在一定程度上提升了模型性能和可解释性,但在医学级EEG数据场景下,其固有缺陷日益凸显。一方面,EEG信号幅值微弱且极易受到工频干扰、眼电和肌电伪迹等噪声的影响^[14],在原始单通道信号上直接进行频域变换往往会同步放大高频噪声成分,使得提取的频域特征中混杂大量冗余信息,在样本规模有限的条件下容易导致模型学习到不稳定甚至错误的判别模式,从而引发严重的过拟合问题。另一方面,更为关键的是,经科学研究表明,情绪的产生并非由单一脑区的局部活动所决定,而是额叶、颞叶和顶叶等多个功能区在不同频段上的动态协同结果^[15-16],这种逐通道独立的频域分析方式在方法论上割裂了大脑的空间协同机制。然而,现有流程在频域分析之前便将物理通道视为相互独立的信号源,从而限制了模型对稳定的、具有生理一致性的跨脑区协同特征的建模能力。

基于直接频域提取方法在噪声抑制与空间协同建模上具有双重局限,目前研究逐步聚焦特征协同聚合的建模策略,试图通过空间协同信息的引入优化频域特征质量,但现有研究仍存在显著的方法论短板。其一,现有方法普遍遵循在频域之后对多通道特征进行协同融合,但这种范式未能从根本上解决原始EEG信号信噪比低、噪声干扰大的问题,导致频域特征的鲁棒性受限。例如,MCMFuse^[17]虽然进行了频域协同选择,但其频域特征提取仍建立在原始电极信号基础上;ACCNet^[18]虽实现了自适应频带分解,但处理对象依然是原始多通道信号。其二,部分研究尝试在时域引入空间约束后再进行频域

表1 EEG不同频段的功能

Table 1 Functions of different EEG frequency bands

频段	频率范围/ Hz	典型生理功能
Delta(δ)	0.5~4	深度睡眠、基础脑电活动
Theta(θ)	4~8	认知活动、记忆相关
Alpha(α)	8~13	放松、闭眼状态、抑制外部干扰
Beta(β)	13~30	警觉、注意力、认知负荷
Gamma(γ)	>30	高级认知、感知整合

分析,但其空间建模能力受限于局部感受野。如 Borra 等^[19]提出采用可学习的带通滤波器组进行频带分解,然后对各频带进行空间滤波,但其空间滤波器的感受野局限于卷积核尺寸和电极的物理邻域,难以覆盖额叶-颞叶等空间距离远但功能协同强的脑区对。

针对上述问题,本文提出了一种协同聚合优先的 EEG 情绪识别新范式,通过重构时域、空间与频域特征的建模顺序,从源头提升频域特征的鲁棒性。具体而言,设计了一种端到端的协同聚合卷积模块,如图 1(c)所示,在频域分析之前对多通道 EEG 信号进行自适应加权聚合,模拟大脑不同功能区之间的动态连接关系,将离散且噪声较大的电极信号重构为若干组高信噪比的协同表征。在此基础上,对聚合后的协同表征进行频域特征提取,实现对情绪相关周期性结构的有效刻画。同时,通过引入多组聚合核构建多样化的协同模式,显著丰富特征空间,从而缓解小样本条件下的特征稀疏与过拟合问题。

本文的主要贡献如下:

- (1) 提出协同聚合优先的 EGG 情绪识别范式。突破“先频域提取、后空间融合”的传统处理范式,率先在频域分析前引入脑区协同建模,从源头抑制通道级噪声干扰。
- (2) 融合全局频域与局部卷积特征建模机制。设计协同聚合卷积模块,通过时间卷积与空间卷积的级联结构实现多通道信号的自适应加权融合,构建频域映射与频率融合模块,实现跨协同模式的频谱自适应融合。
- (3) 跨数据集适应性强。模型在 SEED-IV 和 SEED-V 数据集上均表现优异,验证了结构设计在通道适配与泛化能力方面的优势。

1 方 法

1.1 方法概述

图 2 给出了本文模型总体框架结构。如图 2(a)所示,本文提出的基于聚合卷积的协同频域特征提

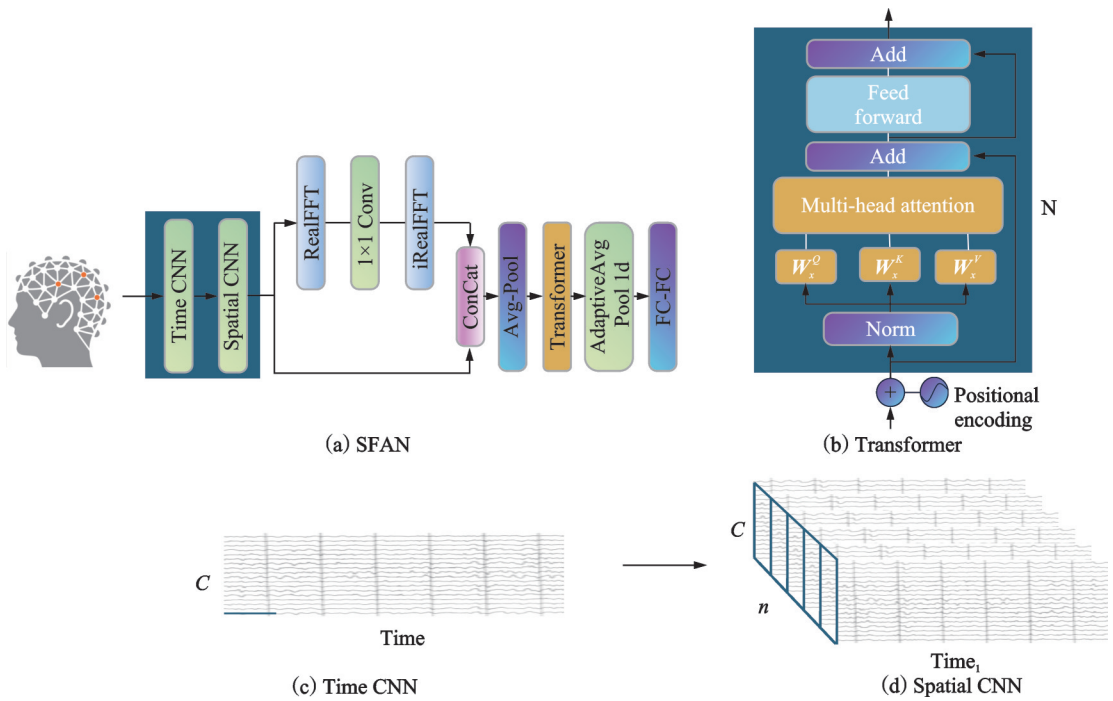


图 2 本文模型总体框架结构图
Fig.2 Overall framework structure diagram of the proposed model

取网络(Synergy-frequency aggregation network, SFAN)由多个关键模块组成,整体为端到端结构。模型的输入为预处理后的多通道EEG信号,其数据形状为 $(1, C, \text{Time})$,其中 C 表示电极通道数,Time表示时间长度。针对传统方法在原始单通道信号上直接进行频域特征提取而忽视脑区协同关系的问题, SFAN在整体设计上采用协同聚合优先的建模思路,在频域特征提取之前对多通道EEG信号进行时空协同建模。

模型首先通过由时间卷积(Time CNN,如图2(c)所示)和空间卷积(Spatial CNN,如图2(d)所示)共同构成的协同聚合卷积模块,对EEG信号进行时空特征建模。时间卷积沿时间维度提取局部时序特征并过滤噪音,而空间卷积则重点建模不同电极通道之间的空间关系,综合考虑不同脑区之间的协同作用以及电极通道排列对空间结构的影响。通过在空间卷积中引入多个卷积核,模型能够对不同通道组合方式进行自适应聚合,从而生成多种潜在的协同模式,对应多个特征图。经过Spatial CNN后,每个特征图在空间维度上被压缩为1,其输出数据形状为 $(n, 1, T_2)$,其中 n 表示卷积核的数量,即生成的协同表征数量。该过程本质上是利用与电极通道数相同大小的卷积核对多通道信号进行加权融合,从而实现通道信息的协同聚合。

在获得聚合后的协同表征后,模型采用一维实数快速傅里叶变换(RealFFT)将信号从时间域映射至频率域,以提取情绪相关的频域特征。随后,在频域中引入 1×1 卷积操作,在通道维度上实现跨通道的线性频率融合,使模型能够自动强调关键协同通道的频率响应并抑制冗余信息。之后,通过逆快速傅里叶变换(iRealFFT)将融合后的频域特征还原至时间域,并与协同聚合卷积模块输出的时域特征进行拼接(Concat),形成融合时域结构与频域约束的联合特征表示。

最后,将联合特征输入基于Transformer结构的特征提取模块(图2(b)),对序列进行高级特征建模,以捕捉EEG信号中的长程时间依赖关系,并进一步提升特征的判别能力。模型最终通过分类模块输出情绪类别的预测概率 $\hat{y}$ 。以下将对SFAN模型中各组成模块的具体结构与实现方式进行详细说明。

1.2 协同聚合卷积模块

协同聚合卷积模块旨在在频域分析之前,对多通道EEG信号进行有效的时空协同建模,从而生成具有较高信噪比和生理一致性的协同表征。该模块由Time CNN与Spatial CNN顺序构成,其中前者用于在保持通道结构的前提下提取多尺度时间特征,而后者则进一步对多通道、多特征图信息进行协同聚合建模。

设输入的预处理EEG信号为 $X \in \mathbf{R}^{1 \times C \times \text{Time}}$ 。时间卷积模块沿时间维度对输入信号进行卷积运算,通过引入 n 个时间卷积核,在每个电极通道上并行提取多种局部时序特征,从而生成 n 个时间特征图。对于第 n 个时间卷积核,其输出可表示为

$$X_{\text{tem}}^{(n)}(c, t) = \sum_{\tau=0}^{k-1} X(c, t + \tau) w_{\text{tem}}^{(n)}(\tau) \quad (1)$$

式中: $w_{\text{tem}}^{(n)}(\tau)$ 为第 n 个时间卷积核在时间偏移 τ 处的权重, k 表示时间感受野大小。经过时间卷积后,可得到时间特征表示 $X_{\text{tem}} \in \mathbf{R}^{n \times C \times \text{Time}_1}$,其中 Time_1 为时间卷积后的时间长度。该过程使模型能够从不同时间尺度对EEG信号进行建模,为后续空间协同聚合提供丰富且稳定的时序特征基础。

在时间卷积输出的多特征图基础上,空间卷积模块进一步沿通道维度对特征进行协同聚合建模。考虑到EEG信号中不同电极通道对应不同脑区,且情绪活动通常由多个脑区的协同作用所决定, Spatial CNN通过对各通道特征进行加权融合,显式刻画脑区之间的空间协同关系。具体而言,对于时间卷积输出的第 n 个时间特征图,其对应的空间聚合结果可表示为

$$S^{(n)}(t) = \sum_{c=1}^C X_{\text{tem}}^{(n)}(c, t) w_{\text{spa}}^{(n)}(c) \quad (2)$$

式中 $w_{\text{spa}}^{(n)}(c)$ 表示第 n 个空间卷积核在第 c 个通道上的权重。需要说明的是,空间卷积阶段同样生成 n 个特征图,其数量与时间卷积阶段保持一致,并在空间卷积运算后依次引入批归一化(Batch normalization)和 ELU(Exponential linear unit)激活函数,以提升特征分布的稳定性并增强非线性表达能力,但两者在语义上对应的是不同层级的特征表示:时间卷积刻画的是多尺度时间特征,空间卷积刻画的是在该时间尺度下的脑区协同聚合结果。

经过空间卷积后,每个特征图在通道维度上被压缩为 1,其整体输出可表示为 $S \in \mathbf{R}^{n \times 1 \times \text{Time}_2}$,其中 Time_2 为对应的时间长度。该过程本质上是一种数据驱动的空间滤波与通道聚合操作,能够在抑制单通道噪声干扰的同时,突出对情绪识别更为关键的跨脑区协同模式。

综上,协同聚合卷积模块通过“多时间特征图提取-对应空间协同聚合”的建模方式,在频域分析之前完成了对多通道 EEG 信号的时空协同表示学习,为后续频域特征提取与融合提供了更加鲁棒且具有判别性的输入基础。

1.3 频域映射与频率融合模块

在协同聚合卷积模块中,模型已将多通道 EEG 信号映射为一组具有明确脑区协同语义的时域表征。为进一步刻画这些协同模式在频率层面的差异性与稳定性,SFAN 在此基础上引入频域映射与频率融合模块,对聚合后的协同信号施加显式的频域约束。

设协同聚合卷积模块的输出表示为 $S \in \mathbf{R}^{n \times 1 \times \text{Time}_2}$,其中 n 表示协同表征的数量,每一条时间序列对应一种潜在的脑区协同模式。为了保证频域变换过程的数值稳定性,模型首先对 S 在时间维度上进行归一化处理,随后对每个协同表征独立施加一维实数快速傅里叶变换(1D RealFFT),将其从时间域映射至频率域,其过程可表示为

$$\bar{F} = \text{RealFFT}(\text{LayerNorm}(S)) \in \mathbf{R}^{n \times 1 \times \text{Time}_2/2} \quad (3)$$

由于实数 FFT 的输出为复数形式,为便于后续采用标准卷积算子进行建模,本文将频谱特征拆分为实部与虚部,并在通道维度进行拼接,从而构建等价的实值频域表示。该操作在不损失频谱信息的前提下,使频域特征能够无缝融入深度学习框架。

在获得实值频域特征后,SFAN 进一步在频率域中引入 1×1 卷积操作,以实现不同协同表征之间的频率融合与重加权,其过程可形式化表示为

$$\hat{F} = \text{Conv}_{1 \times 1}(\bar{F}) \in \mathbf{R}^{n \times 1 \times \text{Time}_2/2} \quad (4)$$

从信号建模角度来看,该 1×1 卷积在协同表征维度上对频率响应进行线性组合,使每一个输出协同特征的频谱分量均由多个输入协同模式的频率信息加权构成,从而能够自适应地强调与情绪状态高度相关的频段,并抑制冗余或不稳定的频率成分。

在完成频域融合后,模型通过 iRealFFT 将融合后的频域特征重新映射回时间域,其结果表示为

$$\tilde{S} = \text{iRealFFT}(\hat{F}) \in \mathbf{R}^{n \times 1 \times \text{Time}_2} \quad (5)$$

由此得到的时间序列不仅保留了原始协同结构的时序信息,同时受到频域特性的显式约束,使模型更加关注具有稳定周期性和生理意义的情绪相关模式,而非瞬态噪声波动。

最后,模型将频域约束后的时间特征 $\tilde{S}$ 与协同聚合卷积模块输出的时域特征 S 在特征维度上进行 Concat 拼接,形成联合特征表示,在该阶段未额外引入归一化或尺度调整操作,其原因在于,EEG 信号本身幅值较小且具有较强的个体差异性,过度的归一化处理可能会削弱不同样本之间的相对幅值差异,从而影响情感相关判别信息的保留。

1.4 Transformer 特征建模与分类模块

在频域映射与频率融合模块获得联合特征 $Z \in \mathbf{R}^{n \times 1 \times \text{Time}_2}$ 之后,首先沿时间维度进行平均池化以降

低计算复杂度并平滑特征序列。经过池化后的特征在时间维度上被压缩,同时保持每个协同表征的判别信息。随后,将特征重新排列为两维矩阵(Time', n),以满足 Transformer 的输入要求,其中 Time' 表示降采样后的时间长度。

重排后的特征通过一层全连接映射,生成 Transformer 的输入嵌入序列。Transformer 特征建模与分类模块利用多头自注意力机制与前馈网络对序列进行全局建模,从而捕捉协同表征在不同时间步的长程依赖关系,并进一步增强特征的判别能力。Transformer 的输出随后通过一维自适应平均池化(Adaptive average pooling 1D)在时间维度上进一步压缩,使输出特征维度固定,便于后续分类层处理。

池化后的特征进入两层全连接网络完成分类决策。第一层全连接结合 ELU 激活函数进行非线性映射,即

$$H_1 = \text{ELU}(W_1 f_{\text{pool}} + b_1)$$

(6)

式中: f_{pool} 为自适应池化后的特征, W_1 与 b_1 分别为权重和偏置。第二层全连接结合 Softmax 激活函数输出情绪类别的预测概率,即

$$P = \text{softmax}(W_2 H_1 + b_2)$$

(7)

式中: W_2 与 b_2 分别为第二层全连接层的权重和偏置, P 表示模型对各类别的预测概率分布。通过 Softmax 归一化,保证输出概率和为 1,便于解释和多类别分类任务使用。

综上,Transformer 特征建模与分类模块通过时间池化、嵌入映射和全局注意力机制有效捕捉协同表征的长程时序依赖,并结合自适应池化与两层全连接实现最终分类。这一设计在保证模型性能的同时,降低了计算复杂度,并提升了在跨被试和跨会话 EEG 情绪分类任务中的稳定性与鲁棒性。

1.5 模型超参数分析

为更加直观地展示模型在各阶段的特征表示变化,本文对主要模块的输入输出尺寸进行统计,如表 2 所示。该统计结果有助于分析模型在时间域、空间域及高阶表示阶段的特征演化过程。由表 2 可以看出,输入信号尺寸为 $1 \times 62 \times 400$ 。经过时间卷积后,时间维度由 400 压缩至 376,同时通道数扩展至 40;随后通过空间卷积实现跨通道信息整合,将空间维度由 62 压缩至 1,实现空间特征聚合。在频域约束与特征融合之后,通过平均池化操作将时间维度由 376 进一步降至 21,从而有效降低后续 Transformer 编码阶段的序列长度,提高建模效率。Transformer Encoder 堆叠 N 层,其输入输出维度保持为 21×256 ,保证特征表示在高维嵌入空间中的稳定性。最终通过自适应池化与全连接层完成特征压缩与类别映射。整体来看,模型在前端卷积阶段逐步完成时间与空间信息压缩,在进入 Transformer 之前

表 2 本文模型各主要模块输入输出尺寸

Table 2 Input-output dimensions of main modules of the proposed model

模块	输入尺寸	输出尺寸	参数量
Time CNN	$1 \times 62 \times 400$	$40 \times 62 \times 376$	1 040
Spatial CNN	$40 \times 62 \times 376$	$40 \times 1 \times 376$	99 240
BatchNorm + ELU	$40 \times 1 \times 376$	$40 \times 1 \times 376$	80
FFT	40×376	40×376	0
AvgPool2d	$80 \times 1 \times 376$	$80 \times 1 \times 21$	0
Transformer Encoder $\times N$	21×256	21×256	$\approx 0.6 \times 10^6 \times N$
AdaptiveAvgPool1d	21×256	21×40	0
全连接层 FC(840→32)	840	32	26 912
全连接层 FC(32→class)	32	5	$33 \times \text{class}$

显著降低序列长度,使得自注意力计算更加可控。特征维度变化过程清晰、递进合理,有助于保证模型表达能力与训练稳定性。

2 实 验

为了验证本文所提的SFAN在EEG情绪识别任务中的有效性,在两个公开EEG数据集上进行了系统实验。将SFAN与当前主流方法进行对比,并评估其在不同实验设置下的性能表现,从而验证模型的泛化能力与稳定性。

2.1 数据集

为了评估SFAN的性能,本文在SEED-IV和SEED-V两个公开EEG数据集上进行实验。两个数据集均来源于上海交通大学BCMI实验室,具有不同的实验设计和情绪类别,能够全面检验模型的适应性、鲁棒性及跨会话、跨受试者的泛化能力。

SEED-IV^[20]是SEED数据集的扩展版本,将原有的3类情绪扩展为4类:快乐(Happy)、悲伤(Sad)、恐惧(Fear)和中性(Neutral)。该数据集共招募15名受试者(6男9女),每位受试者参与3轮实验,时间间隔至少间隔一天以减少疲劳效应和实验顺序偏差。在每轮实验中,受试者依次观看24段情绪诱导视频(每类6段),每段视频时长约2 min,视频内容涵盖电影片段、日常生活短片和实验室定制片段,以诱发目标情绪。EEG信号通过62个标准电极通道采集,采样率为200 Hz,并采用4~47 Hz带通滤波,去除低频漂移与高频干扰。为适应深度学习模型训练,连续EEG信号按2 s不重叠窗口划分为独立样本,每个样本保持原始通道布局和时间顺序,用于后续特征提取和模型训练。

SEED-V^[21]在SEED-IV的基础上作了进一步扩展,将情绪类别增加至5类:快乐(Happy)、悲伤(Sad)、恐惧(Fear)、厌恶(Disgust)和中性(Neutral)。该数据集共招募16名受试者,每位受试者参与多次实验会话,每次会话包含不同顺序的情绪视频播放。实验视频长度与SEED-IV类似,每段视频约2 min,视频总数根据情绪类别均衡分配,以保证各类情绪样本数量相近。EEG采集通道与采样率与SEED-IV一致,预处理流程包括带通滤波、伪迹去除及基线校正。为了保持与SEED-IV实验条件的一致性,原始EEG信号同样按2 s不重叠窗口划分为独立样本,确保模型输入的一致性。

两个数据集均按照4:1的比例划分为训练集和测试集。各数据集的输入格式及样本数量统计信息如表3所示。

表3 SEED-IV和SEED-V数据集样本分布
Table 3 Sample distribution of the SEED-IV and SEED-V datasets

数据集	数据形状	总样本数	训练样本数	测试样本数
SEED-IV	62×400	75 645	60 516	15 129
SEED-V	62×400	22 008	17 606	4 402

2.2 模型训练与参数设置

模型训练采用Adam优化器,初始学习率设为0.000 1,并结合学习率衰减策略对训练过程进行动态调整,以保证模型稳定收敛。训练过程中,使用交叉熵损失函数优化模型参数。整个实验在配备NVIDIA RTX 3090 GPU(24 GB显存)的高性能计算环境下运行,使用CUDA12.1与Python3.11实现模型构建与训练,从而提升训练效率并加速收敛。

2.3 基线比较

为了评估SFAN在多类别情绪识别任务中的性能,本文将其与当前主流EEG情绪识别方法进行了对比实验。实验分别在SEED-IV和SEED-V数据集上进行,覆盖四分类和五分类任务,能够全面验证模型在不同情绪类别数量下的表现,结果如表4、5所示。

从表4可知,在SEED-IV数据集上进行的四分类(快乐、悲伤、恐惧和中性)实验中,SFAN取得了

97.42% 的分类准确率,明显高于所有对比方法。在对比方法中,EESCN采用脉冲神经网络结合卷积结构进行特征提取,取得 79.65% 的准确率;3D-CTM通过卷积 Transformer联合多域特征建模时空依赖,达到 87.44%;BF-GCN利用图卷积建模 EEG 通道间连接关系,准确率为 82.03%;SS-EMERGE 为 81.51%;MSA 模型通过自注意力和记忆遗忘机制增强特征判别力,取得 95.90% 的准确率。与这些方法相比,SFAN能够充分挖掘 EEG 信号中跨通道的协同情绪频域特征,表现出更强的多类情绪区分能力。

如表 5 所示,在 SEED-V 数据集上进行的五分类实验(快乐、悲伤、恐惧、厌恶和中性)中,SFAN同样表现出显著优势,达到 94.91% 的分类准确率。在对比方法中,Conformer与 ASTDF-Net 分别基于卷积 Transformer 和时空双流注意力结构,准确率分别为 77.32% 和 83.63%;PGCN与 SOGNN 采用图结构建模跨被试 EEG 情绪特征,准确率分别为 73.45% 和略低于此范围;AP-CapsNet 与 CLIER 通过胶囊网络和隐式情绪调节机制增强模型表达能力,准确率分别为 80.79% 和 82.46%;MSA 模型在该数据集上达到 85.92%。相比之下,SFAN 在复杂五分类任务中能够更有效地捕捉情绪间微小差异和动态变化,显示出更强的表达能力和稳健性。

2.4 消融实验

为了系统评估频域信息在 EEG 情绪识别中的作用,以及频域特征提取时机对模型性能的影响,本文在 SEED-V 数据集上设计并开展了消融实验。具体而言,实验对比了 3 种不同的处理策略:完全不引入频域信息,仅基于时空特征进行建模;在原始多通道 EEG 信号上直接进行频域特征提取;在协同聚合卷积模块之后再进行频域特征提取。3 种策略均在相同实验设置下进行训练与测试,并以分类准确率作为主要评价指标。实验结果如表 6 所示,对应的训练过程分别如图 3、4 和 5 所示。

从表 6 以及图 3 可以观察到,在不引入频域信息的情况下,模型在训练集上的性能持续提升,而测试集性能提升有限,呈现出一定程度的过拟合现象。这表明在样本规模受限的医学 EEG 数据场景下,仅依赖时空特征难以充分刻画情绪相关的判别信息。进一步地,如表 6 和图 4 所示,当在原始 EEG 信号上直接提取频域特征时,模型性能反而显著下降,分类准确率降至 67.47%,其主要原因在于原始单通道 EEG 信号中普遍存在工频干扰及眼电、肌电等高频噪声,直接进行频域变换会同步放大这些噪声成分,使得频域特征中混杂大量冗余甚至误导性信息,在小样本条件下进一步加剧模型的过拟合问题。

表 4 在 SEED-IV 数据集上的对比实验结果

Table 4 Comparative experimental results on the SEED-IV dataset

年份	方法	准确率/%
2024	EESCN ^[22]	79.65
2024	3D-CTM ^[10]	87.44
2024	BF-GCN ^[23]	82.03
2025	SS-EMERGE ^[24]	81.51
2026	MSA ^[25]	95.90
2026	SFAN	97.42

表 5 在 SEED-V 数据集上的对比实验结果

Table 5 Comparative experimental results on the SEED-V dataset

年份	方法	准确率/%
2023	Conformer ^[26]	77.32
2023	PGCN ^[27]	73.45
2023	AP-CapsNet ^[28]	80.79
2023	ASTDF-Net ^[29]	83.63
2024	CNN ^[30]	82.06
2025	CLIER ^[31]	82.46
2026	MSA ^[25]	85.92
2026	SFAN	94.91

表 6 在 SEED-V 数据集上不同处理策略下的实验结果

Table 6 Experimental results of different processing strategies on the SEED-V dataset

频域信息处理策略	分类准确率/%
不提取频域信息	85.05
在原始数据上提取频域信息	67.47
在聚合后提取频域信息	94.91

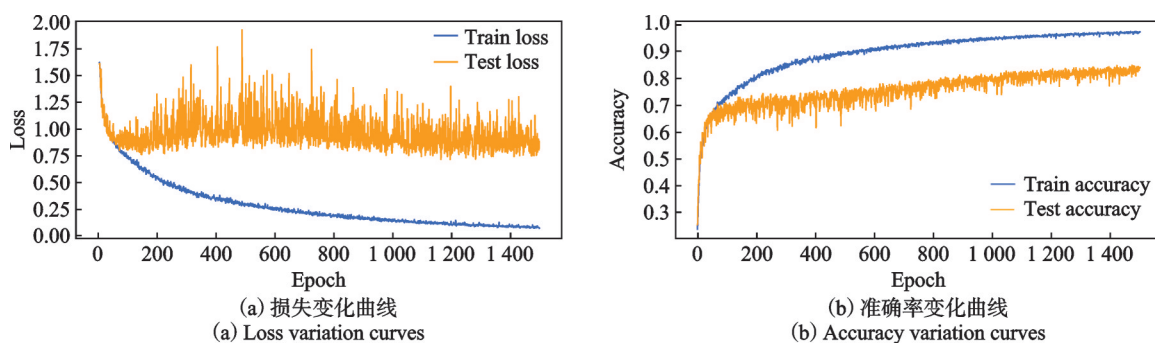


图3 不提取频域信息的训练过程图

Fig.3 Training process without frequency-domain feature extraction

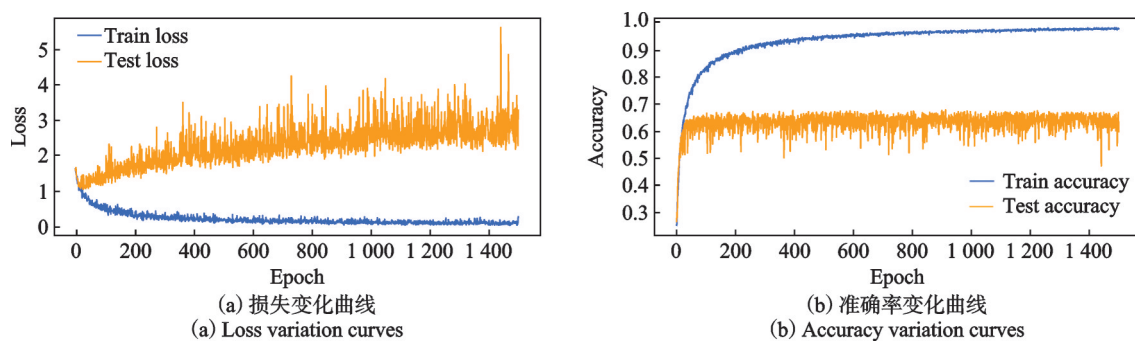


图4 在原始EEG数据上提取频域信息的训练过程图

Fig.4 Training process with frequency-domain feature extraction on raw EEG signals

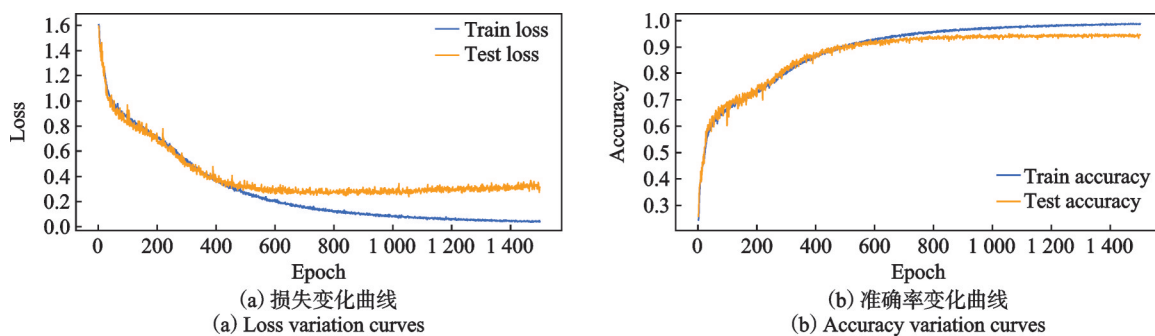


图5 在聚合后提取频域信息的训练过程图

Fig.5 Training process with frequency-domain feature extraction after aggregation

相比之下,在协同聚合卷积之后再行频域特征提取的策略表现出明显优势。如图5所示,模型训练过程更加稳定,测试精度显著提升,并最终在SEED-V数据集上取得94.91%的分类准确率。该结果表明,协同聚合卷积模块在频域分析前对多通道EEG信号进行协同建模,有效提升了信号的信噪比和结构一致性,使得后续提取的频域特征更加稳定且具有判别力。上述消融实验结果充分验证了本文提出的“协同聚合优先、频域分析随后”设计理念的有效性,进一步说明在频域建模之前引入脑区协同机制对于提升EEG情绪识别性能和鲁棒性具有关键作用。

为了研究 Transformer 模块的作用以及其块数 N (图 2(b) 右侧) 对模型性能的影响, 本文在 SEED-V 数据集上进行了消融实验, 比较了从 1 层到 9 层不同块数下的分类准确率, 结果如图 6 所示。从图 6 可以看出, 随着块数的增加, 模型的性能总体呈上升趋势, 在 4 层时达到最高准确率 94.91%, 然后随着块数进一步增加, 准确率略有波动但整体保持在高水平。这表明适当增加 Transformer 块数能够增强模型对时序依赖的建模能力, 但过多的块数可能导致过拟合或增加训练难度, 对性能提升作用有限。

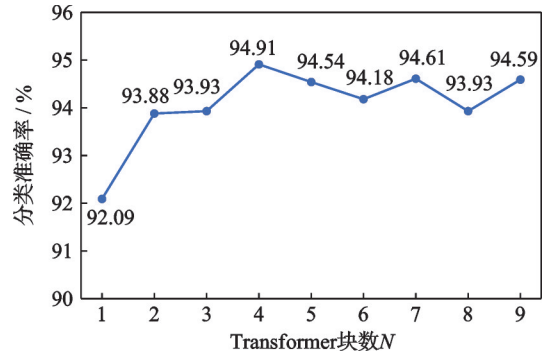


图 6 Transformer 模块中块数 N 对模型性能的影响

Fig.6 Effect of the number of Transformer blocks N on model performance

3 时间卷积层的特征可视化与抑噪能力分析

为了验证时间卷积层在特征提取过程中对噪声的抑制作用, 本文基于 SEED-V 数据集测试集, 对原始 EEG 信号与卷积层输出特征进行对比分析。考虑到 EEG 信号在采集过程中易受到肌电、眼电及环境干扰的影响, 这类噪声通常表现为时域上的高频突变, 本文引入信号一阶差分 (First-order difference) 作为刻画指标, 其定义为

$$\Delta x_i = x_{i+1} - x_i \quad (8)$$

一阶差分能够有效反映信号的局部变化程度: 当信号平滑时差分值趋近于零, 而高频噪声会导致差分分布变宽。因此, 通过分析差分分布可以间接评估信号中的噪声强度。如图 7 所示, 原始信号的一阶差分分布呈现出较宽且平坦的形态, 表明其包含较多高频波动成分; 而经过时间卷积层处理后, 差分分布明显向零点集中, 峰度增加且分布范围显著收缩, 说明模型在提取时序特征的同时有效抑制了高频扰动。从定量结果来看, 原始信号的一阶差分标准差为 3.20×10^{-6} , 经时间卷积层处理后下降至 1.95×10^{-7} , 对应的平滑度提升比例达到 93.91%。该结果表明, 大部分由高频突变引起的干扰在特征提取阶段已被削弱。结合模型整体性能的提升, 可以认为时间卷积层在保留判别性信息的同时, 对非平稳噪声具有显著的抑制作用, 从而增强了特征表示的鲁棒性。

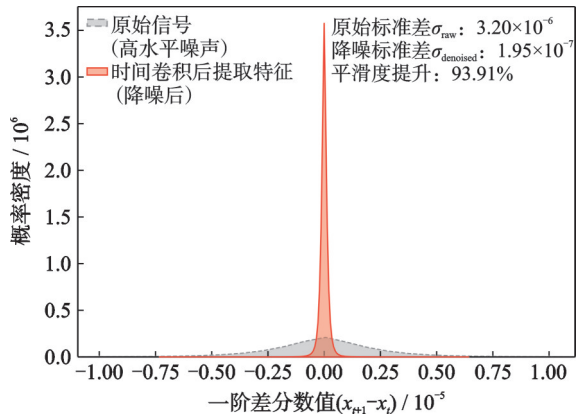


图 7 时间卷积层处理前后的特征分布情况

Fig.7 Feature distribution before and after the time CNN layer treatment

4 结束语

针对传统 EEG 情绪识别方法在单通道频域特征提取中忽视脑区协同作用、易受噪声干扰以及小样本过拟合的问题, 提出了一种基于协同聚合卷积与频域约束的端到端脑电情绪分类框架 SFAN。以协同聚合卷积模块建模多通道时空协同特征, 结合频域映射与 Transformer 时序注意力, 提升特征鲁棒性并精准捕捉情绪相关动态模式。实验结果表明, SFAN 在 SEED-IV 和 SEED-V 数据集上均取得了显著优于现有方法的分类性能, 验证了模型在复杂多类别情绪识别任务中的判别能力和稳健性。

尽管取得了良好效果, 本文方法仍存在一些局限。首先, 在便携式 EEG 设备上的适应性尚未验证。

其次,跨实验室、跨采集设备的泛化能力需要进一步评估,本文模型仅利用 EEG 信号进行情绪分类,未充分利用面部表情、眼动、生理信号等多模态信息。考虑到多模态信号能够提供互补的情绪线索,未来可探索在 SFAN 框架中融合多模态特征,进一步增强情绪识别的准确性与稳健性。此外,模型使用了 Transformer,计算复杂度相对较高,针对本文模型计算复杂度较高、难以满足实时应用需求的局限,未来可探索轻量化卷积与注意力剪枝策略,在保持识别精度的前提下降低模型参数量与推理延迟,使其能够部署于便携式 EEG 设备或移动端平台,实时应用仍有提升空间。未来工作可在以下几个方向展开:探索轻量化卷积与注意力剪枝策略以优化计算效率,研究跨设备迁移学习提高泛化能力,以及结合多模态信号(如面部表情或生理信号)进一步增强情绪识别性能。

参考文献:

- [1] 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.
- [2] ZHANG Tong, ZHENG Wenming, CUI Zhen, et al. A deep neural network-driven feature learning method for multi view facial expression recognition[J]. *IEEE Transactions on Multimedia*, 2016, 18(12): 2528-2536.
- [3] LIU Zhentao, XIE Qiao, WU Min, et al. Speech emotion recognition based on an improved brain emotion learning model[J]. *NeuroComputing*, 2018, 309: 145-156.
- [4] BRITTON J C, PHAN K L, TAYLOR S F, et al. Neural correlates of social and nonsocial emotions: An fMRI study[J]. *NeuroImage*, 2006, 31(1): 397-409.
- [5] 倪哲文, 全海燕. 基于双残差 LSTM 和 DCGAN 的脑电信号驱动视觉图像重建模型[J]. *数据采集与处理*, 2026, 41(1): 244-258.
NI Zhewen, QUAN Haiyan. EEG signal-driven visual image reconstruction model based on double residual LSTM and DCGAN [J]. *Journal of Data Acquisition and Processing*, 2026, 41(1): 244-258.
- [6] 刘佳, 张祥瑞, 陈大鹏, 等. 结合双流注意力与对抗互重建的双模态情绪识别方法[J]. *电子与信息学报*, 2026, 48(1): 277-286.
LIU Jia, ZHANG Yangrui, CHEN Dapeng, et al. Bimodal emotion recognition method based on dual-stream attention and adversarial mutual reconstruction[J]. *Journal of Electronics & Information Technology*, 2026, 48(1): 277-286.
- [7] 闭应洲, 刘善锐, 霍雷刚, 等. 基于参数共享的多特征图内外交互脑电图分类模型[J]. *数据采集与处理*, 2025, 40(4): 950-961.
BI Yingzhou, LIU Shanrui, HUO Leigang, et al. A parameter-sharing multi-feature map interaction model for EEG classification[J]. *Journal of Data Acquisition and Processing*, 2025, 40(4): 950-961.
- [8] 朱子骄, 金子敏, 王广华, 等. 基于脑电监测技术的情绪识别研究进展[J]. *传感技术学报*, 2025, 38(11): 1921-1930.
ZHU Zijiao, JIN Zimin, WANG Guanghua, et al. Emotion recognition using EEG monitoring technology[J]. *Chinese Journal of Sensors and Actuators*, 2025, 38(11): 1921-1930.
- [9] 陈旭东, 钟恒, 皇甫洁, 等. 脑电信号情绪识别综述[J]. *计算机应用*, 2023, 43(S1): 323-332.
CHEN Xudong, ZHONG Heng, HUANGFU Jie, et al. Review of emotion recognition of EEG signals[J]. *Journal of Computer Applications*, 2023, 43(S1): 323-332.
- [10] 张学军, 王天晨, 王泽田. 基于多域信息融合的卷积 Transformer 脑电情感识别模型[J]. *数据采集与处理*, 2024, 39(6): 1543-1552.
ZHANG Xuejun, WANG Tianchen, WANG Zetian. Convolutional Transformer EEG emotion recognition model based on multi-domain information fusion[J]. *Journal of Data Acquisition and Processing*, 2024, 39(6): 1543-1552.
- [11] CHEN Q, MAO X, SONG Y, et al. An EEG-based emotion recognition method by fusing multi-frequency-spatial features under multi-frequency bands[J]. *Journal of Neuroscience Methods*, 2025, 415: 110360.
- [12] XU Y, DU Y, LI L, et al. AMD-ET: Attention based multiple dimensions EEG Transformer for emotion recognition[J]. *IEEE Transactions on Affective Computing*, 2023, 15(3): 1067-1077.
- [13] ZHONG Q, ZHU Y, CAI D, et al. Electroencephalogram access for emotion recognition based on a deep hybrid network[J]. *Frontiers in Human Neuroscience*, 2020, 14: 589001.
- [14] PILLALAMARRI R, SHANMUGAM U. A review on EEG-based multimodal learning for emotion recognition[J]. *Artificial Intelligence Review*, 2025, 58(5): 131.
- [15] KHAZAEI E, MOHAMMADZADE H. Temporal analysis of functional brain connectivity for EEG-based emotion recognition [J]. *arXiv preprint arXiv: 2112.12380*, 2021.
- [16] 刘柯, 张孝, 李沛洋, 等. 基于脑功能网络和共空间模式分析的脑电情绪识别[J]. *计算机应用研究*, 2021, 38(5): 1344-1349.

- LIU Ke, ZHANG Xiao, LI Peiyang, et al. EEG emotion recognition based on common spatial pattern of brain functional network[J]. *Application Research of Computers*, 2021, 38(5): 1344-1349.
- [17] YANG Y, LI Y, ZENG T, et al. MCMFuse: A multi-level co-selection model in the frequency domain to enhance EEG signal analysis for emotion recognition[C]//*Proceedings of 2024 39th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. Dalian, China: IEEE, 2024: 1885-1889.
- [18] TIAN D, WANG Y, GONG P, et al. ACCNet: Adaptive cross-frequency coupling graph attention for EEG emotion recognition[J]. *Neural Networks*, 2010, 191(000): 12.
- [19] BORRA D, MAGOSSO E. Unveiling multi-domain signatures of EEG oscillations using a fully-interpretable convolutional neural network[J]. *Computer Methods and Programs in Biomedicine*, 2025, 271: 109008.
- [20] ZHENG W L, LIU W, LU Y, et al. EmotionMeter: A multimodal framework for recognizing human emotions[J]. *IEEE Transactions on Cybernetics*, 2018, 49(3): 1110-1122.
- [21] LIU W, QIU J L, ZHENG W L, et al. Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition[J]. *IEEE Transactions on Cognitive and Developmental Systems*, 2021, 14(2): 715-729.
- [22] XU F F, PAN D, ZHENG H, et al. EESCN: A novel spiking neural network method for EEG-based emotion recognition[J]. *Computer Methods and Programs in Biomedicine*, 2024, 243: 107927.
- [23] LI C, TANG T, PAN Y, et al. An efficient graph learning system for emotion recognition inspired by the cognitive prior graph of EEG brain network[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 36(4): 7130-7144.
- [24] AHUJA C, SETHIA D. SS-EMERGE-self-supervised enhancement for multidimension emotion recognition using GNNs for EEG[J]. *Scientific Reports*, 2025, 15(1): 14254.
- [25] 刘善锐, 闭应洲, 霍雷刚, 等. 融合记忆力和自注意力机制的脑电图情绪识别模型[J]. *电子与信息学报*, 2026, 48(3): 1017-1026.
- LIU Shanrui, BI Yingzhou, HUO Leigang, et al. An EEG emotion recognition model integrating memory and self-attention mechanisms[J]. *Journal of Electronics & Information Technology*, 2026, 48(3): 1017-1026.
- [26] SONG Y, ZHENG Q, LIU B, et al. EEG conformer: Convolutional Transformer for EEG decoding and visualization[J]. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2023, 31: 710-719.
- [27] ZHOU Y, LI F, LI Y, et al. Progressive graph convolution network for EEG emotion recognition[J]. *NeuroComputing*, 2023, 544: 126262.
- [28] LIU S, WANG Z, AN Y, et al. EEG emotion recognition based on the attention mechanism and pre-trained convolution capsule network[J]. *Knowledge-Based Systems*, 2023, 265: 110372.
- [29] GONG P, JIA Z, WANG P, et al. ASTDF-Net: Attention-based spatial-temporal dual-stream fusion network for EEG-based emotion recognition[C]//*Proceedings of the 31st ACM International Conference on Multimedia*. [S.l.]: ACM, 2023: 883-892.
- [30] ASLAN M, BAYKARA M, ALAKUŞ T B. Analysis of brain areas in emotion recognition from EEG signals with deep learning methods[J]. *Multimedia Tools and Applications*, 2024, 83(11): 32423-32452.
- [31] LI D, JIN Z, SHEN Y, et al. EEG emotion recognition based on an implicit emotion regulatory mechanism[J]. *IEEE Transactions on Artificial Intelligence*, 2025, 6(11): 3005-3017.

作者简介:



闭应洲(1967-),男,博士,教授,硕士生导师,研究方向:智能计算、智能信息处理, E-mail:byzhou@163.com。



袁月(2002-),女,硕士研究生,研究方向:人工智能、脑电图。



刘善锐(1998-),通信作者,男,硕士,研究方向:人工智能、脑电图, E-mail: 511332602@qq.com。



周淑桓(2003-),女,硕士研究生,研究方向:人工智能。



霍雷刚(1986-),男,博士,副教授,研究方向:图像处理、人工智能。



甘秋静(2000-),女,硕士研究生,研究方向:人工智能。

(编辑:王静)

A Synergy-Aggregation-First Frequency-Domain Constrained End-to-End EEG Emotion Recognition Method

BI Yingzhou^{1,2}, YUAN Yue^{1,2}, LIU Shanrui^{3*}, ZHOU Shuheng^{1,2}, HUO Leigang^{1,2}, GAN Qiuqing^{1,2}

(1. School of Artificial Intelligence, Nanning Normal University, Nanning 530199, China; 2. Guangxi Key Lab of Human-Machine Interaction and Intelligent Decision, Nanning Normal University, Nanning 530199, China; 3. College of Computer and Information Engineering, Guangxi Vocational Normal University, Nanning 530007, China)

Abstract: The electroencephalogram (EEG)-based emotion recognition is a key technology in next-generation human-computer interaction and brain-computer interface (BCI) systems, with significant applications in national brain science initiatives and intelligent healthcare. However, the inherent non-stationarity, low signal-to-noise ratio, and scarcity of medical EEG samples present substantial challenges for existing deep learning methods in real-world scenarios. Most current approaches directly extract frequency-domain features from single-channel raw signals. This strategy not only risks misinterpreting high-frequency noise as informative features but also neglects the spatial synergy among different brain regions, limiting the model's generalization ability. To address these issues, this paper proposes a synergy-first end-to-end EEG emotion recognition framework, introducing the synergy-frequency aggregation network (SFAN) for collaborative frequency-domain feature extraction based on aggregation convolutions. Departing from the conventional "frequency extraction followed by spatial fusion" paradigm, the SFAN first employs a channel aggregation convolution module to adaptively fuse multi-channel EEG signals, simulating dynamic inter-regional brain synergies to generate highly robust collaborative representations. Frequency-domain feature extraction and deep spatiotemporal modeling are then performed on these representations. This strategy effectively suppresses channel-level noise prior to frequency analysis and enriches the feature space through diverse aggregation patterns, alleviating overfitting under small-sample conditions. The SFAN achieves a classification accuracy of 97.42% on the SEED-IV dataset and 94.91% on the SEED-V dataset. Experimental results demonstrate that the proposed method consistently outperforms existing approaches in cross-subject and cross-session emotion recognition tasks, validating the effectiveness of incorporating inter-regional brain collaboration prior to frequency-domain analysis for enhancing EEG emotion recognition robustness.

Highlights:

1. This paper breaks through the traditional processing paradigm of "frequency domain extraction first, spatial fusion later", and introduces brain region collaborative modeling before frequency domain analysis to suppress channel level noise interference from the source.
2. Design a collaborative aggregation convolution module to achieve adaptive weighted fusion of multi-channel signals, construct frequency domain mapping and frequency fusion modules, and achieve spectrum adaptive fusion across collaborative modes.
3. The model performs well on both the SEED-IV and SEED-V datasets, verifying the advantages of structural design in channel adaptation and generalization ability.

Key words: electroencephalogram (EEG); emotion recognition; deep learning; synergy aggregation; frequency-domain modeling

Foundation items: National Natural Science Foundation of China (No.62067007); The Innovation Project of Guangxi Graduate Education (No.JGY2023236); Guangxi Science and Technology Innovation Platform Program Project (No.Guikue LT2600640001).

Received: 2026-01-21; **Revised:** 2026-07-06

***Corresponding author, E-mail:** 511332602@qq.com.