

# 基于稀疏注意力机制的多域 SSVEP 信号识别方法

张学军<sup>1,2</sup>, 李夏芸<sup>1</sup>

(1. 南京邮电大学电子与光学工程学院、柔性电子(未来技术)学院, 南京 210023; 2. 南京邮电大学射频集成与微组装技术国家地方联合工程实验室, 南京 210023)

**摘要:** 稳态视觉诱发电位 (Steady-state visual evoked potential, SSVEP) 技术正在脑机接口领域快速发展, 本文面向 4 类 SSVEP 频率识别任务, 针对短时窗下的 SSVEP 易受噪声等非目标信号影响, 特征信息提取不足导致识别率低的问题, 提出一种融合稀疏注意力机制与时频双域特征的卷积神经网络 (Convolutional neural networks integrating sparse attention mechanism with temporal-frequency features, SATF-CNN) 脑电信号分类模型。首先在输入端嵌入动态正弦位置编码模块, 接着构建 Top- $k$  混合稀疏注意力机制的通道提取路径, 结合多阈值稀疏化分数选择筛选关键特征, 最后构建双分支特征提取网络, 通过时域卷积网络分支与快速傅里叶变换频域网络分支分别捕获时域特征与频域特征, 采用 Kolmogorov-Arnold 网络实现特征非线性融合。实验结果表明, 在 1 s 时间窗下的四分类跨受试者实验中, 准确率高达 93.54%, 信息传输速率为 93.13 bit/min。SATF-CNN 模型在稳态视觉诱发电位识别任务中表现出卓越的分类性能, 为临床神经工程和智能交互设备开发提供了新的解决方案。

**关键词:** 脑电信号; 脑机接口; 稳态视觉诱发电位; 注意力机制; 卷积神经网络

中图分类号: TP391 文献标志码: A

## Multi-Domain SSVEP Signal Recognition Based on Sparse Attention Mechanism

ZHANG Xuejun<sup>1,2</sup>, LI Xiayun<sup>1</sup>

(1. College of Electronic and Optical Engineering & College of Flexible Electronics (Future Technology), Nanjing University of Posts and Telecommunications, Nanjing 210023, China; 2. Nation-Local Joint Project Engineering Lab of RF Integration & Micropackage, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

**Abstract:** The steady-state visual evoked potential (SSVEP) technology is rapidly developing in the field of brain-computer interfaces. This paper targets a four-class frequency recognition task to address the low recognition rate caused by insufficient feature extraction and the influence of noise and other non-target signals in short time windows. To this end, a convolutional neural network for EEG signal classification integrating a sparse attention mechanism with time-frequency dual-domain features (SATF-CNN) is proposed. Firstly, a dynamic sinusoidal position encoding module is embedded at the input end. Then, a channel extraction path with a Top- $k$  hybrid sparse attention mechanism is constructed, and key features are selected through multi-threshold sparse fraction selection. Finally, a dual-branch feature extraction network is built, with a time-domain convolutional network branch and a fast Fourier transform frequency-domain network branch capturing time-domain and frequency-domain features respectively. The Kolmogorov-Arnold network is used to achieve nonlinear feature fusion. Experimental results show that in a four-class cross-subject experiment with a 1-second time window, the accuracy rate is as high as 93.54%, and the information transmission rate is 93.13 bit/min. The SATF-CNN model demonstrates outstanding classification performance in SSVEP

**基金项目:** 国家自然科学基金(61977039)

收稿日期: 2025-07-10; 修订日期: 2025-12-17

recognition tasks, providing a new solution for clinical neuroengineering and the development of intelligent interaction devices.

**Key words:** electroencephalography; brain-computer interface; steady-state visual evoked potentials; attention mechanism; convolutional neural network

## 引言

脑机接口 (Brain computer interface, BCI) 技术作为神经科学与信息技术方向交叉融合的重要领域<sup>[1]</sup>, 通过建立生物大脑与机电装置的信息通路, 读取受试者的大脑意图。BCI 核心技术主要包括信号采集、特征提取、指令输出三个部分。根据信号的采集方式, 可分为侵入式、半侵入式和非侵入式三类: 侵入式方法通过手术将电极植入大脑皮层, 这种方式身体极易受到创伤; 半侵入式脑机接口不需要直接穿透皮层组织, 主要应用于癫痫病灶定位等医学领域<sup>[2]</sup>; 非侵入式脑机接口通过头皮表面电极接收脑电信号 (Electroencephalography, EEG), 其中稳态视觉诱发电位因无创性、低实施成本等优势, 成为 BCI 技术应用的主流方向之一。

SSVEP 是指当外界给予视觉系统特定频率闪烁刺激时, 大脑视皮层会产生与刺激频率相关或保持谐波关系的稳态脑电信号, 这种频率响应使原刺激信号能有效地被检测识别。英国神经科学家 David Regan 通过实验首次发现了大脑枕叶皮层对于 SSVEP 信号的响应规律, 利用频率域分析法确定了响应与刺激频率的锁相关系, 为 SSVEP 的生理机制奠定了基础<sup>[3]</sup>。当受试者注视不同频率闪烁的视觉刺激目标时, 系统根据识别出的 SSVEP 信号来执行相应的指令, 如控制机器人的运动方向、对电子设备进行操作等, 这为残障人士提供了一种新的辅助交流和控制手段, 同时也为健康人群拓展了人机交互的新方向。

SSVEP 具有非侵入性及较高的信息传输速率 (Information transfer rate, ITR) 的特点, 在科技和医疗康复领域应用广泛, 但信号在采集过程中易受头皮阻抗、肌电伪迹等噪声干扰, 同时受试者之间存在差异性, 导致信号识别准确率不高。因此, 寻找高效的 SSVEP 识别方法至关重要。传统的 SSVEP 信号识别方法主要分为频域分析和相关成分分析两类。频域分析方法常使用快速傅里叶变换 (Fast Fourier transformation, FFT) 和功率谱密度分析 (Power spectral density analysis, PSDA) 法。Cheng 等通过设计了不同频率闪烁按钮的虚拟电话键盘, 仅使用单通道脑电信号, 利用连续的 FFT 检测刺激频率, 最终实现了 27.15 bits/min 的平均传输速率<sup>[4]</sup>。Materka 等利用 PSDA 方法对枕区 3 个电极采集的 8 种频率编码进行分类, 鲁棒性更强<sup>[5]</sup>。在相关成分分析方法中, 典型相关分析 (Common component analysis, CCA) 法无需训练数据, 通过构造正余弦参考信号, 计算与脑电信号的线性相关性来识别目标频率<sup>[6]</sup>。张莉等利用 CCA 和低通滤波去除了脑电信号中的肌电伪迹<sup>[7]</sup>。Chen 等提出了基于滤波器组的典型相关分析算法 (Filter bank canonical correlation analysis, FBCCA), 在利用基波函数的基础上, 结合谐波子频带, 进一步提高了 SSVEP 的分类准确率, 但未充分的利用脑电信号的时序信息<sup>[8]</sup>。Nakanishi 等提取 SSVEP 信号的时域特征, 提出任务相关成分分析 (Task-related component analysis, TRCA) 信号检测方法, 不仅降低了计算成本, 而且在短视窗下提高了分类准确率和信息传输速率<sup>[9]</sup>。清华大学高小榕等将集成任务相关成分分析 (ensemble task-related component analysis, eTRCA) 与子带卷积神经网络 (sub-band convolutional neural network, sbCNN) 融合, 充分利用 eTRCA 的特征提取能力和 sbCNN 的信号识别能力, 将两模型的分类分数相加, 选择得分最高的频率作为输出, 在两个公开数据集上有较高的识别准确率和信息传输速率<sup>[10]</sup>。

随着卷积神经网络 (Convolutional neural network, CNN)<sup>[11]</sup>和循环神经网络 (Recurrent neural network, RNN)<sup>[12]</sup>为代表的深度学习模型在脑电信号解析中的深入应用, 推动了 SSVEP 识别技术的进一步发展<sup>[13]-[14]</sup>。Kwak 等提出用卷积神经网络对 SSVEP 分类, 仅使用 2 个卷积层和 1 个全连接层, 在 2 s 时间窗下的离线实验中的达到 94.03% 的分类精度<sup>[15]</sup>, 但该

实验主要围绕单一时间窗条件展开,未对不同时间尺度下的模型性能进行系统分析。复卷积神经网络(Complex CNN, CCNN)的应用将频域特征引入深度学习算法,充分利用频谱的复数域信息<sup>[16]</sup>。Lawhern 等提出的 EEGNet 是一种专为脑电信号分类设计的轻量级卷积神经网络,将紧凑型卷积神经网络应用在 P300 诱发电位、运动相关皮层电位(Movement-related cortical potential, MRCP)等不同数据集中,具有较强的泛化能力<sup>[17]</sup>。EEGNet 模型利用时间卷积提取时序特征,并通过深度可分离空间卷积捕捉通道依赖关系,计算参数量小,适用于 SSVEP 等小样本脑电数据分析。Chen 等针对短时窗 SSVEP 分类任务,提出了一种融合注意力机制的 tCNN 结构,从时域滤波器组提取脑电信号,经残差压缩激励模块(Residual squeeze-and-excitation, RSE)增强信号特征,再由 tCNN 进一步处理并分类,成功应用于五自由度仿生手的控制<sup>[18]</sup>。Pan 等将神经卷积网络与长短期记忆(Long short-term memory, LSTM)网络融合,提出了 SSVEPNet 模型,利用频谱归一化技术,显著提升了 SSVEP 系统的频率识别性能<sup>[19]</sup>。Liu 等通过在 SSVEPNet 中引入软饱和非线性模块,改善了短时窗条件下的特征学习效果。然而,该方法主要侧重于时域信号建模,尚未从时频联合建模等多特征角度系统挖掘 SSVEP 信号的判别特性<sup>[20]</sup>。Wan 等提出了一种轻量化的广义零样本学习网络,利用类模板和正弦模板,结合注意力机制提取 SSVEP 信号特征,减少了可训练参数,为降低用户的训练负担提供了一种有效途径<sup>[21]</sup>。

目前,大多数深度学习解码模型在处理 SSVEP 信号序列数据时,通常仅将时间序列或频域特征直接作为网络输入。这种单一域建模方式导致特征提取不充分,从而影响信号解码的准确性。基于传统监督学习的算法通常需要通过复杂的校准过程来提高模型性能,这在很大程度上制约了稳态视觉诱发电位脑机接口系统的实用性和便捷性。

针对以上问题,本文提出一种融合稀疏注意力机制与时频双域特征的脑电信号识别模型(Convolutional neural networks integrating sparse attention mechanism with temporal-frequency features, SATF-CNN)。信号经过动态位置编码,增强时空特征提取性能,解决了传统 CNN 时序建模能力不足的问题;设计多头稀疏注意力机制,通过三级 Top- $k$  稀疏注意力筛选关键特征,与权重自适应融合,最后跨域特征联合特征学习,时域特征采用多尺度卷积建模;频域通过 FFT 转换获取频域信息,为复杂脑电信号解码提供了新的技术途径。

## 1 SATF-CNN 模型

图 1 所示为 SATF-CNN 模型的总体框架图,主要包括位置增强稀疏注意力网络(Position enhanced sparse attention, PESA)和时频特征融合分类(Time-frequency feature extraction, TFE)模块。

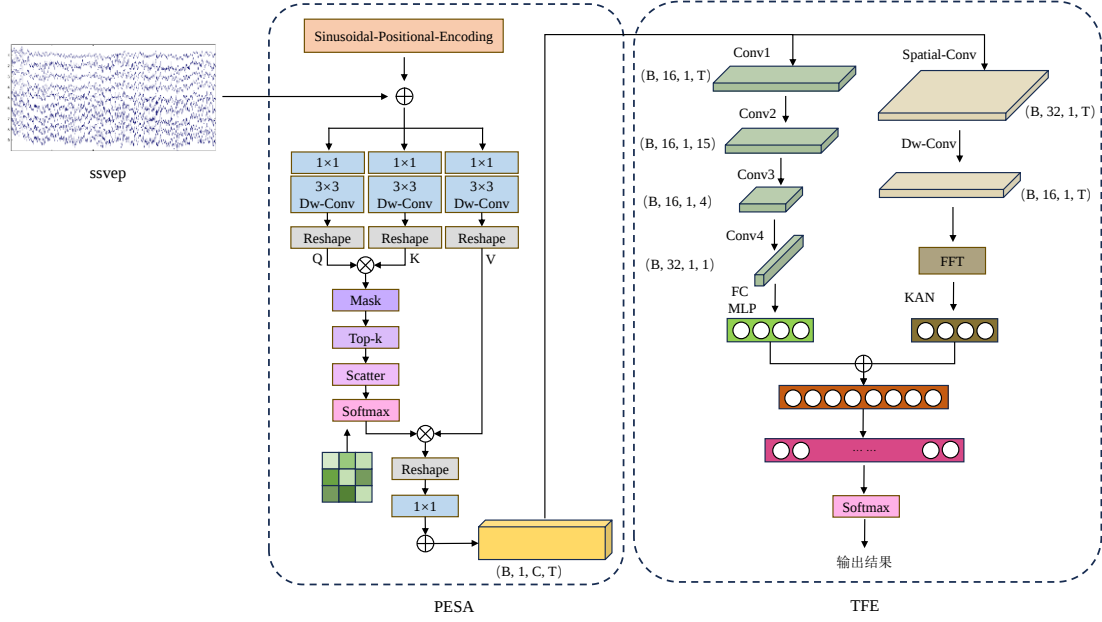


图 1 SATF-CNN 架构图

Fig.1 Architecture diagram of the SATF-CNN

## 1.1 位置增强稀疏注意力网络

为解决传统自注意力计算复杂度高的问题，本文提出位置增强稀疏注意力网络来处理 SSVEP 信号，此模块包含位置编码（Position encoding, PE）和 Top- $k$  稀疏注意力（Top- $k$  sparse attention, TKSA）网络两部分。

### 1.1.1 位置编码

位置编码可以从脑电信号中提取不同时间序列的信息特征。为动态地适应输入特征的通道数和时间步长，将特征维度为  $(B, 1, C, T)$  的原始脑电信号进行正余弦位置编码，其中  $C$  为通道数， $T$  为脑电信号序列长度，信号通过不同频率的正余弦函数组合生成位置编码<sup>[22]</sup>。位置编码计算公式定义为

$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/C}}\right) \quad (1)$$

$$PE(pos, 2i+1) = \cos\left(\frac{pos}{10000^{2i/C}}\right) \quad (2)$$

式中： $pos \in [0, T-1]$  表示时序位置， $i \in [0, C/2-1]$  为通道索引。通过引入指数衰减的频率项  $10000^{-2i/C}$ ，在通道维度上建立了一种频率衰减机制，增强模型对非平稳信号的时序感知能力。位置编码为每个脑电时序信号分配了唯一的编码向量，使后续 Top- $k$  稀疏注意力能够在时间维度对齐下选择关键通道特征，更好地区分与刺激频率相关的时间变化特征与背景脑电波动，尤其在短时间窗条件下能够缓解时序信息不足带来的建模困难，提升模型对时序依赖关系的刻画能力。在本实验中，将位置编码矩阵  $PE \in R^{T \times C}$  与输入张量进行广播相加，引入了时序信息，模型在处理脑电序列时能够识别位置以及随时间变化的规律。

### 1.1.2 Top- $k$ 稀疏注意力

为了提升注意力机制在脑电特征提取过程中的有效性与抗干扰能力，本文引入了一种基于 Top- $k$  分数的稀疏注意力机制。Top- $k$  分数选择最初应用在自然语言处理（Natural language

processing, NLP) 任务中, 通过仅保留对注意力权重贡献较大的特征分量, 将注意力退化为稀疏注意力, 增强了模型的选择能力<sup>[23]</sup>。不同于 NLP 任务在词向量维度进行特征筛选, 本实验使用的是一种基于脑电通道维度的 Top- $k$  选择模块, 这种方法根据查询 ( $Q$ ) 和键 ( $K$ ) 之间的相似性计算注意力得分, 当  $Q$  与  $K$  之间的关联性较弱时, 噪声特征会干扰脑电信号的识别, 因此保留每一位置上得分排名前  $k$  的通道特征, 并将其余部分置零, 构造稀疏注意力权重矩阵, 通过动态筛选关键注意力权重, 由于 SSVEP 信号主要集中分布于枕叶及其邻近区域, Top- $k$  策略的稀疏注意力机制通过选择对分类决策贡献较大的通道特征, 并抑制冗余或噪声通道信息, 有助于模型在多通道 EEG 信号中聚焦于与 SSVEP 相关的关键信息。

位置编码后的脑电信号特征分别通过  $1 \times 1$  卷积和  $3 \times 3$  深度可分离卷积生成  $Q, K, V$ , 并将其沿通道维度划分为多头形式。为了提高注意力机制的数值稳定性, 对  $Q$  与  $K$  进行  $L_2$  归一化处理, 防止内积爆炸。计算初始注意力分数矩阵

$$A = \frac{Q_h K_h^T}{\sqrt{d}}, \quad d = C / N_h \quad (3)$$

式中:  $A \in \mathbf{R}^{B \times N_h \times C \times C}$ ,  $N_h=3$  为注意力头数, 对于每个查询通道, 使用动态多比例阈值融合策略, 生成稀疏注意力掩码, 分别保留  $k_1 = C/2$ 、 $k_2 = 2C/3$ 、 $k_3 = 3C/4$  个最大注意力权重, 其余值概率置为零生成掩码  $M_1$ 、 $M_2$ 、 $M_3$  以实现稀疏化。这种动态选择为

$$\text{SparseAtt}(Q, K, V) = \text{softmax}(T_k(\frac{QK^T}{\sqrt{d}}))V \quad (4)$$

其中  $T_k$  为可学习的参数, 定义为  $[T_k(S)]_{ij} = \begin{cases} S_{ij}, & S_{ij} \in \text{Top}-k(\text{第}j\text{行}) \\ 0, & \text{其他} \end{cases}$ , 经 softmax 激活函数

数输出权重与值矩阵  $V$  相乘, 获得多头输出, 经拼接后通过  $1 \times 1$  卷积投影调整通道维度, 最终恢复为  $x \in \mathbf{R}^{B \times 1 \times C \times T}$  形式的脑电信号特征。

## 1.2 时频特征融合分类

本节提出的时频特征融合分类网络是将脑电特征通过双通道并行架构分别提取时域与频域特征, 两个特征通道共享输入信号, 经过特征融合后, 通过全连接层的处理实现最终分类。

### 1.2.1 时域特征提取网络

时域特征提取网络 (Temporal convolutional network, TCN) 采用四层卷积结构提取信号特征, 为增强时序特征建模过程中的非线性表达能力, 每层卷积后均采用 B-SiLU (Bounded Sigmoid linear unit) 激活函数<sup>[24]</sup>

$$B-SiLU(x) = (x + \alpha) \cdot \sigma(x) - \frac{\alpha}{2}, \alpha = 1.67 \quad (5)$$

式中  $\sigma(x)$  为 Sigmoid 函数。B-SiLU 继承了 SiLU 激活函数的自门控特性, 在正输入区间内保持近似线性增长, 有利于保留具有判别意义的特征幅值; 同时, 引入可调下界对负激活进行有界约束, 可有效抑制短时间窗 SSVEP 信号中由噪声或瞬态干扰引起的不稳定响应, 并避免其在逐层卷积传播过程中被放大, 提升了特征提取稳定性与网络收敛效率。卷积网络具体参数如表 1 所示。

表 1 时域卷积网络参数

Table1 Time-domain convolutional network parameters

| 相关卷积层 | 卷积核尺寸           | 步长 | 输出维度                               |
|-------|-----------------|----|------------------------------------|
| Conv1 | $(9 \times 1)$  | 1  | $(B \times 16 \times 1 \times T)$  |
| Conv2 | $(1 \times T)$  | 3  | $(B \times 16 \times 1 \times 15)$ |
| Conv3 | $(1 \times 12)$ | 1  | $(B \times 16 \times 1 \times 4)$  |
| Conv4 | $(1 \times 4)$  | 1  | $(B \times 32 \times 1 \times 1)$  |

第一个卷积层采用  $9 \times 1$  卷积核进行跨通道特征融合，保留时序信息，通过批归一化层重构信号特征。第二层使用  $1 \times T$  动态卷积核处理输入信号的时序特征，采用步长为 3 填充为 23 的卷积实现时序特征提取，经二次批归一化处理后，通过激活函数提升特征的可分性。第三层使用  $1 \times 12$  大尺度卷积核，通过无填充卷积操作提取局部信号时序关系，经过批归一化层和激活函数确保深层网络训练的稳定性。第四层使用  $1 \times 4$  卷积核对上一层输出的信号进行深度时域特征提取，将脑电信号特征降维，经批归一化与激活函数处理后，特征分类模块将 32 通道特征向量输入全连接层，投影到 4 维目标空间并输出。

### 1.2.2 频域特征提取网络

在频域特征提取网络（Frequency-domain convolutional network, FCN）中，信号先经过大小为  $9 \times 1$  的卷积核进行通道信息特征提取，再利用  $1 \times 1$  逐点卷积调整通道数进行特征降维，将脑电信号特征展平为一维时序信号后，沿着时间轴维度进行快速傅里叶变换，基于共轭对称性保留单边频谱特征。在傅里叶变换前使用卷积网络，拓展了序列，从而提高信号的时间分辨率。最后通过 Kolmogorov-Arnold 网络对频域特征进行非线性投影，生成频域通道输出。

时域特征提取通道与频域特征提取通道输出特征拼接后，通过全连接层进行特征空间变换为 16 维，经过批归一化层和激活函数处理后，再次通过全连接层进行类别投影，将特征维度压缩至 4 维，经 softmax 激活函数对应最终分类任务的目标概率。整个模块融合时频双域信息，通过特征变换与映射，提高了 SSVEP 信号的识别准确率。

## 2 数据集及其处理

### 2.1 数据集

#### 2.1.1 数据集 A

数据集 A 为自采数据集，实验选取了 8 名身体健康视力正常志愿者的 SSVEP 脑电信号记录，其中男性与女性受试者各 4 名，年龄在 22-25 岁之间。实验在光照均匀的安静环境中进行，实验要求受试者坐姿端正，注意力集中，双目与刺激界面保持 70-80 cm 的注视距离。视觉刺激采用黑白块交替闪烁范式，如图 2 所示。通过 E-Prime 2.0 软件分别实现 8 Hz、9 Hz、10 Hz、11 Hz 四种不同频率的刺激。刺激呈现设备选用刷新率稳定在 60 Hz、分辨率为  $2240 \times 1400$  像素的 LED 显示器。

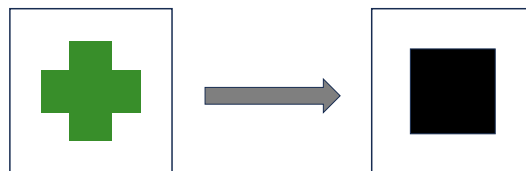


图 2 视觉刺激界面示意图

Fig.2 Schematic diagram of the visual stimulus interface

本实验采用 5 组重复设计，每组包含 4 类频率刺激各一次，每个试次由 3 s 准备期、5 s 刺激注视期和 6 s 休息期组成。SSVEP 信号由 64 导联 Neuroscan 系统采集，采样率为 500 Hz。

实验过程选取覆盖顶叶和枕叶区域的  $P_z$ 、 $PO_3$ 、 $PO_5$ 、 $PO_z$ 、 $PO_4$ 、 $PO_6$ 、 $O_1$ 、 $O_z$ 、 $O_2$  九个通道采集信号，以头顶  $C_z$  通道作为参考，所有电极阻抗均低于  $10\text{ k}\Omega$ ，实验采集现场如图 3 所示。实验开始时，屏幕中央出现 3 s “十” 字视觉提醒，要求受试者集中注意力，随后显示屏中央刺激方块闪烁 5 s，受试者需保持稳定注视，之后界面显示为空白屏 6 s。刺激阶段要求受试者保持注视并避免眨眼，为了避免视觉疲劳，两次重复实验之间受试者有几分钟的休息时间。最终每位受试者共完成 20 试次脑电试验。



图 3 信号采集场景图

Fig.3 Signal acquisition scene diagram

#### 2.1.2 数据集 B

数据集 B 为清华大学脑机接口研究团队提出的 Benchmark 脑电数据集，记录了 35 名健康受试者的 64 脑电通道数据，其中男性 18 人<sup>[25]</sup>。实验刺激界面采用  $5 \times 8$  的矩阵拼写器布局，视觉刺激源采用频率范围 8 Hz 到 15.8 Hz、步长为 0.2 Hz 的 40 种闪烁的字符。每名受试者需完成 6 个区块 (block) 测试，每个区块对应所有的 40 个频率刺激的试验。单次试验切分成 6 s 的脑电数据：前 0.5 s 为视觉定位阶段，屏幕出现红色方块提示受试者集中注意力；中间 5 s 为多频率方块同步闪烁阶段，受试者需迅速注视到目标刺激块；最后 0.5 s 为短暂休息期。最终每位受试者共完成 240 个脑电试验。

### 2.2 数据预处理

对于数据集 A，将信号降采样至 250 Hz，通过 4.8 Hz~90 Hz 通带范围的带通滤波器，去除主要频率信息以外的噪声信号，使用 50 Hz 的带阻滤波器消除工频干扰，运用独立成分分析 (Independent component analysis, ICA) 去除受试者的眼电、肌电信号等伪迹，剔除幅值超过  $100\text{ }\mu\text{V}$  的损坏信号段。考虑到视觉延迟效应的影响，对于 5 s 的刺激注视期，仅提取从 0.2 s 到 5 s 之间的数据片段进行分析。

对于数据集 B，采用已预处理好的标准化数据<sup>[25]</sup>。每位受试者的脑电数据以通道  $\times$  采样点  $\times$  试验  $\times$  区块的四维矩阵形式存储，原始数据降采样至 250 Hz，选取与数据集 A 相同的四个刺激类别与九个电极通道进行实验。

实验采用多尺度时间窗对两个数据集进行分割，时间窗长度从 0.2 s 至 1.0 s，并以 0.1 s 为增量梯度变化。其中，0.2 s 被选为时间窗下限，是在去除刺激起始阶段的瞬态不稳定响应后所保留的最小有效信号长度，确保所截取片段中包含初步形成的频率锁定成分。研究 1.0 s 以内时间窗的脑电数据，是为分析所提出模型在短时条件下的识别性能<sup>[26]</sup>。为优化数据利用率并缓解小样本过拟合问题，采用滑动窗口对连续的脑电信号划分处理，滑动步长设置为当前时间窗长度的一半，样本数量得到扩充后有利于后续的特征提取和信号分类。

### 3 实验与结果

#### 3.1 模型训练

实验在 NVIDIA RTX 4060 GPU、i7-13700H 的硬件设备上进行，采用 Pytorch 框架构建深度学习模型。网络训练阶段选用 Adam 优化器，以交叉熵作为损失函数监督分类任务，每次批量处理 64 个样本数据，共经过 500 轮迭代优化收敛。使用动态学习率调整机制以优化模型训练过程，初始学习率  $\alpha$  设置为  $1 \times 10^{-3}$ ，当连续 20 轮验证集准确率未下降时，学习率降低为原来的一半。采用  $K$  折交叉验证的方式对模型算法进行评估<sup>[27]</sup>。

#### 3.2 评价指标

本文采用两种国际通用的脑机接口系统性能评价指标：分类准确率（Accurate, ACC）和信息传输速率<sup>[28]</sup>，对系统的性能进行综合评估。分类准确率用来衡量模型对 SSVEP 识别的正确率，反映了正确分类的样本数量占总样本数量的比例。信息传输速率本质为单位时间内大脑意图被正确解析的比特量，从信息论角度量化模型的分类效率，与传统通信领域的比特率不同，ITR 综合考虑分类精度和信号时间长度等因素，其计算公式为

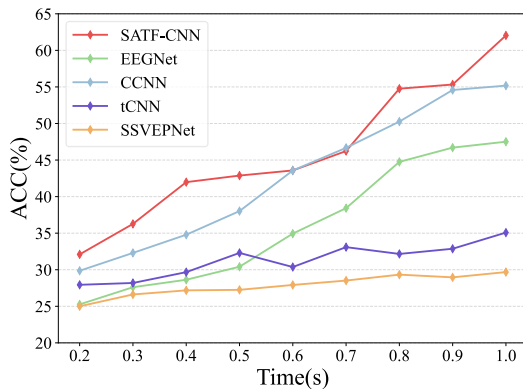
$$ITR = \frac{60}{T} \left[ \log_2 N + P \log_2 P + (1-P) \log_2 \left( \frac{1-P}{N-1} \right) \right] \quad (6)$$

式中： $T$  为切割后的脑电数据时间长度； $P$  是分类准确率； $N$  是分类目标的数量，本实验中  $N=4$ 。

#### 3.3 实验结果分析

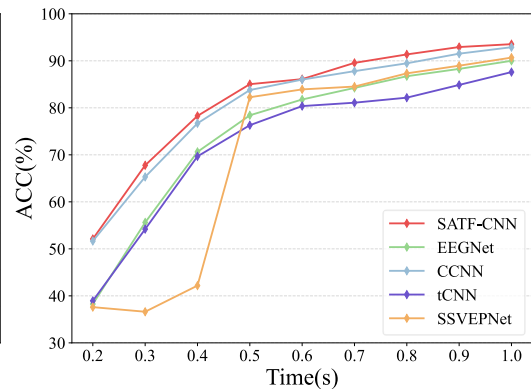
##### 3.3.1 跨受试者实验

跨受试者实验用来评估 SSVEP 识别模型在不同个体间的泛化性。对于低信噪比数据集 A，采用 4 折交叉验证方案，将 8 名受试者划分为 4 组，选择 6 名受试者的数据作为训练集，剩余 2 名受试者的数据作为验证集，在 4 轮实验中，确保每位受试者的数据都作为验证集被测试过。对于高信噪比数据集 B，对受试者数据集采用 5 折交叉验证方案。实验再现了 5 种常用的 SSVEP 解码方法作为基准对比，研究 0.2 s-1.0 s 不同时间窗下的分类准确率和信息传输速率。



(a) 数据集 A

(a) Dataset A



(b) 数据集 B

(b) Dataset B

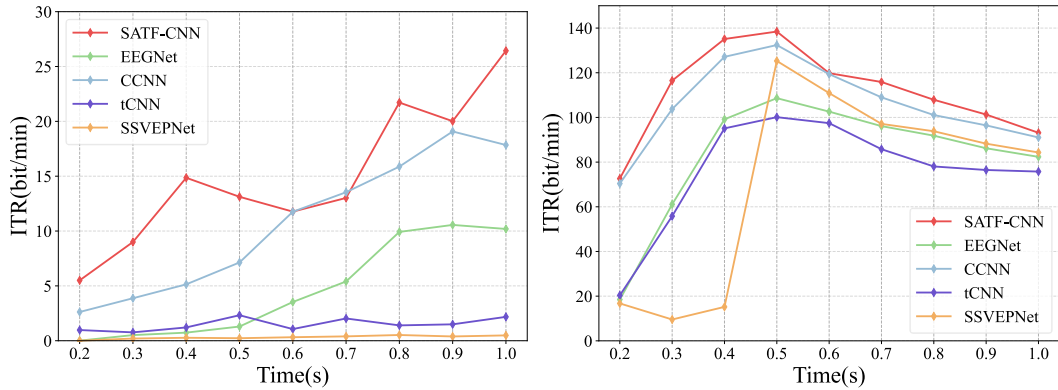
图 4 不同数据长度的准确率

Fig.4 Accuracy of data of different lengths

图 4(a)、(b)展示了 SATF-CNN 模型在识别 SSVEP 信号准确率方面的优势。对于低信噪比数据集 A，在 1.0 s 时间窗下，模型平均分类准确率达到 62.04%，较最优基准方法 CCNN 提升 6.86 个百分点。在 0.7 s 时间窗下，CCNN 模型的准确率略优于 SATF-CNN，局部表现更佳，但在其他 8 种时间窗下，SATF-CNN 综合表现最好。在复杂噪声环境下的脑电信号识别任务中，tCNN 与 SSVEPNet 模型表现出明显的适应性局限，相较之下，SATF-CNN 模型更具有抗噪优势，适用于实际应用中非理想环境下采集的脑电数据。

在数据集 B 中，当信号长度为长时间窗 1.0 s 时，SATF-CNN 模型的准确率高达 93.54%。随着时间窗口缩减至 0.4-0.7 s 的中段区间，当时间窗口为 0.5 s 时，SSVEPNet 出现 40.07% 的异常跃变，而 SATFCNN 准确率平稳增长，这种持续增长的优势证明模型具备优秀的特征融合能力，能够有效整合时空与频谱的多模态信息。当时间窗缩短至 0.2 s 时，SATF-CNN 以 52.11% 的准确率高于 SSVEPNet 模型 14 个百分点以上。

图 5(a)为数据集 A 在不同时间窗下的信息传输速率，SATF-CNN 模型整体呈上升趋势。仅在 0.7 s 时间窗下，SATF-CNN 的 ITR 表现次优，根据公式(6)，该结果反映出在 0.7 s 时间尺度下，SATF-CNN 在分类准确率提升与决策时间增长之间的权衡效果略逊于 CCNN。图 5(b)为数据集 B 在不同时间窗下的信息传输速率变化趋势，SATF-CNN 模型在 0.4 s 时间窗下，ITR 达到 135.11 bits/min 的峰值，此时脑电意图的识别效率达到最优状态，这表明 SATF-CNN 并非仅依赖长时间信息积累，而是能以最高效的方式解码大脑神经活动信息，实现脑电意图识别通量单位时间内最大化。



(a) 数据集 A 信息传输速率

(b) 数据集 B 信息传输速率

(a) ITR of Dataset A

(b) ITR of Dataset B

图 5 不同数据长度的信息传输率

Fig.5 ITR of data of different lengths

### 3.3.2 独立受试者实验

在数据集 A 的中，每位受试者均完成 5 组独立脑电采集实验。为评估模型在独立受试者实验中的性能，采用留一折交叉验证实验策略：依次将每位受试者的单组实验作为测试集，其余 4 组作为训练集，确保所有实验数据均被用于测试。表 2 所示为在 1.0 s 时间窗下，与其他模型性能对比结果。

表 2 数据集 A 的对比试验结果

Table 2 Comparison results on Dataset A

| 方法 | ACC/% | ITR/(bit·min <sup>-1</sup> ) |
|----|-------|------------------------------|
|----|-------|------------------------------|

|                          |              |              |
|--------------------------|--------------|--------------|
| EEGNet <sup>[17]</sup>   | 48.56        | 11.12        |
| CCNN <sup>[16]</sup>     | 56.22        | 19.04        |
| tCNN <sup>[18]</sup>     | 35.34        | 2.29         |
| SSVEPNet <sup>[19]</sup> | 29.37        | 0.43         |
| <b>SATF-CNN</b>          | <b>67.12</b> | <b>33.91</b> |

在严格遵循留一折交叉验证框架的独立受试者实验中，各模型在 1.0 s 时间窗下的性能差异明显。SATF-CNN 模型表现出 67.12% 的准确率和 33.91 bits/min 的最高信息传输速率。CCNN 模型 56.22% 准确率和 19.04 bits/min 传输率表现次优，证明了频域信息在特征提取方面的有效性。适用于不同脑电识别任务的 EEGNet 模型能实现 48.56% 的准确率与 11.12 bits/min 的信息传输率的基础性能。然而，tCNN 与 SSVEPNet 模型的分类准确率较低，缺乏有效的特征选择机制，导致关键 SSVEP 特征被背景噪声掩盖，不能很好的完成信号识别任务。

表 3 所示为在 1.0 s 时间窗下，对于数据集 B 采用留一折交叉验证与其他模型性能对比结果。SATF-CNN 模型准确率高达 92.54%，ITR 为 89.93 bit/min，在两个指标上均取得了最优性能，对 SSVEP 信号有更强的特征提取与判别能力。EEGNet 作为适用于多种脑电任务的轻量化卷积神经网络，其准确率为 90.32%，ITR 为 83.27 bit/min，该模型同样适用于小样本识别任务。CCNN 的性能略优于 EEGNet。而传统的 tCNN 方法表现相对较差，在时序建模与频率分辨方面的局限性。

**表 3 数据集 B 的对比试验结果**  
**Table 3 Comparison results on Dataset B**

| 方法                       | ACC/%        | ITR/(bit·min <sup>-1</sup> ) |
|--------------------------|--------------|------------------------------|
| EEGNet <sup>[17]</sup>   | 90.32        | 83.27                        |
| CCNN <sup>[16]</sup>     | 91.49        | 86.71                        |
| tCNN <sup>[18]</sup>     | 84.68        | 68.36                        |
| SSVEPNet <sup>[19]</sup> | 88.37        | 77.82                        |
| <b>SATF-CNN</b>          | <b>92.54</b> | <b>89.93</b>                 |

### 3.4 消融研究

为验证位置增强稀疏注意力模块（PESA）、时域特征提取模块（TCN）和频域特征提取模块（FCN）对 SATF-CNN 模型算法性能的作用，实验对高信噪比数据集 B 采用消融实验探究，分别使用 3 种消融策略。对于 w/o PESA 模型，移除位置增强稀疏注意力模块，仅考虑 TFE 模块对模型性能的影响；对于 w/o TCN 模型，探究时域消融模型对模型性能的影响；对于 w/o FCN 模型，探究频域消融模型对模型性能的影响。实验结果如图 6 所示。

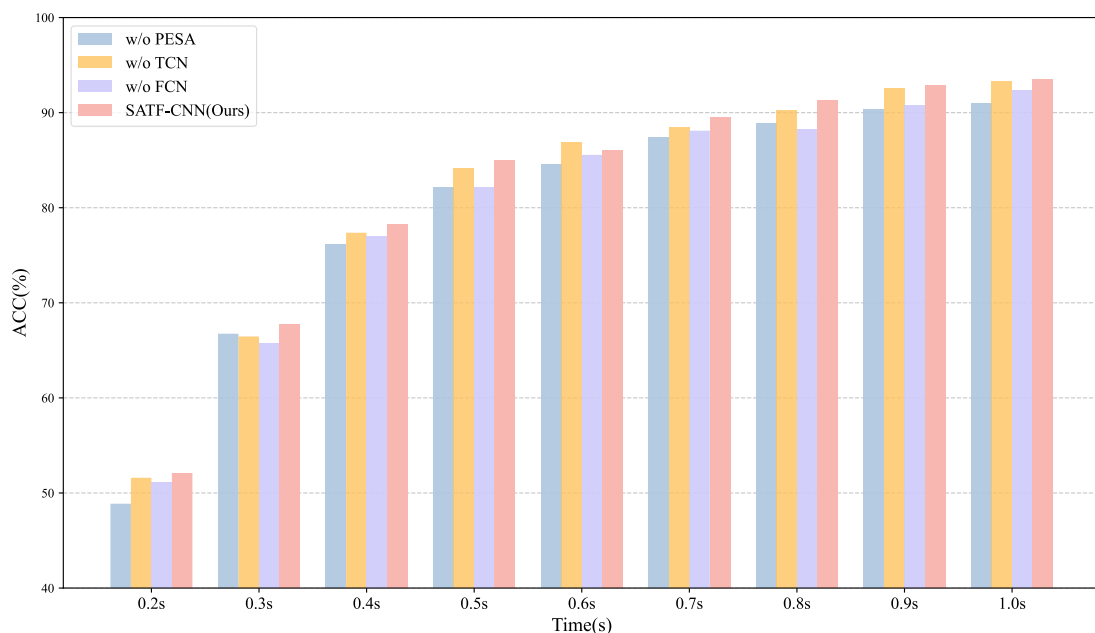
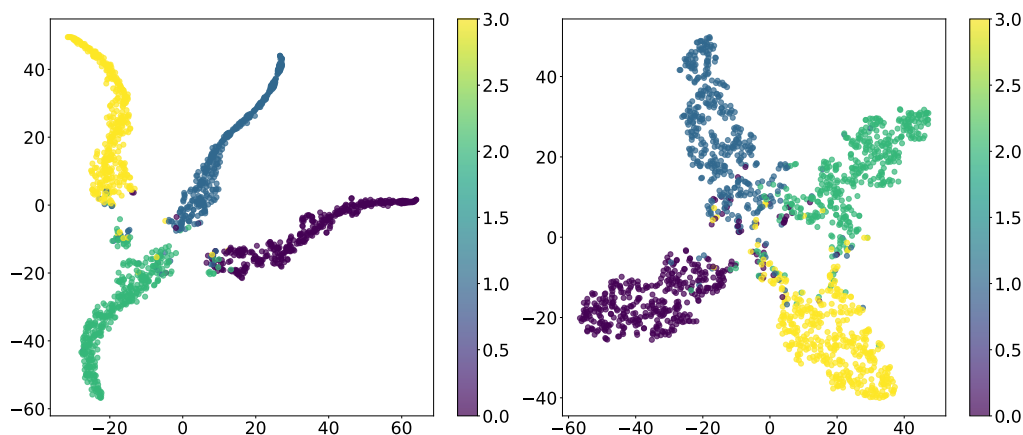


图 6 不同数据长度的消融研究

Fig.6 Ablation study on different data lengths

在 1.0 s 时间窗下，SATF-CNN 模型的平均准确率为 93.54%，经过模块消融后，其他 3 种模型的分类效果均不如原始模型：w/o PESA 模型的准确率为 91.07%，降幅高达 2.47%，w/o TCN 模型的准确率为 93.36%，w/o FCN 的平均准确率为 92.37%。在 9 种不同时间窗长度的分类实验中，w/o PESA 模型在 6 种情况下表现最差，因此当移除稀疏注意力机制时，模型对关键信息的聚焦能力明显减弱，该结果说明 PESA 模块在通道信息增强与有效特征提取过程中发挥着重要作用。当仅保留单一支进行特征提取时，模型对 SSVEP 信号的多尺度表征能力受到限制，导致识别性能下降，因此 TCN 和 FCN 模块分别对于提取时域和频域信息具有不可替代的作用。稀疏注意力机制与时频双分支结构在模型中分别承担着不同的互补功能，联合作用是提升整体识别性能的关键。

为了进一步研究 SATF-CNN 模型提取的特征与消融模块后相比的优越性，本文利用 t-SNE 降维对模型最后一层全连接层的特征分布进行可视化分析<sup>[29]</sup>。SSVEP 信号长度为 1.0 s 时的 t-SNE 可视化结果如图 7。SATF-CNN 模型提取的特征高度紧凑的类内分布，有明显的类间分离，具有较强的特征判别能力。消融模型的特征分布则表现出类内离散度增大且类间边界模糊的现象，尤其在移除位置增强稀疏注意力模块后更加明显。TCN 与 FCN 模块的消融虽影响较小，但仍会导致类间可分性降低。



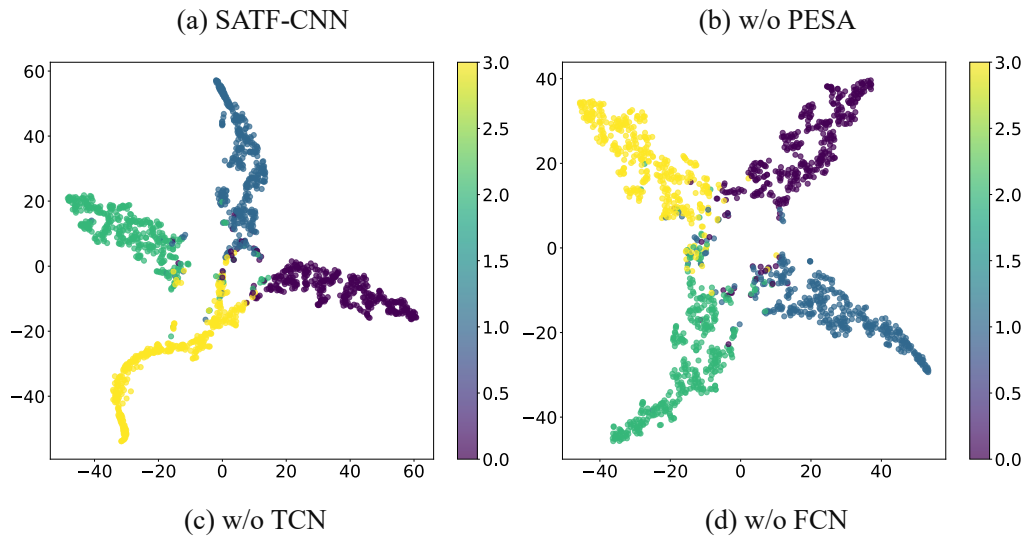


图 7 模块消融 t-SNE 图

Fig.7 t-SNE graphs of module ablation

## 4 结束语

本文提出的 SATF-CNN 模型通过注意力机制与多模态特征融合,有效提升了 SSVEP 识别性能。通过多频正余弦函数编码对脑电信号时序信息进行位置编码,同时设计 Top- $k$  分数稀疏注意力机制,筛选关键特征信息,时频双通道架构提取脑电多域信息,采用 B-SiLU 激活函数通过梯度优化平衡特征保留与噪声抑制。消融实验证明了位置增强稀疏注意力模块以及时频域信息对于特征提取的有效性, SATF-CNN 模型在两个数据集不同时间窗划分下均具有较好的识别性能。

随着脑机接口技术从理论应用于实际生活, SATF-CNN 模型为可穿戴 BCI 设备算法提供新思路。在神经康复领域,可助力运动功能障碍者实现大脑意图操控轮椅。未来的研究中,我们将继续探索更高效的轻量级 SSVEP 信号解码模型,设计更加友好的 SSVEP 刺激范式,以降低受试者视觉疲劳,划分更多长度的时间窗信号,探究不同长短时间窗模型分类性能,帮运动功能障碍者实现行动自由,推动脑机接口医学领域的发展。

参考文献:

- [1] Maiseli B, Abdalla A T, Massawe L V, et al. Brain-computer interface: trend, challenges, and threats[J]. Brain Informatics, 2023, 10(1): 20.
- [2] Sun Y, Chen X, Liu B, et al. Signal acquisition of brain-computer interfaces: A medical-engineering crossover perspective review[J]. Fundamental Research, 2025, 5(1): 3-16.
- [3] Regan D. Some characteristics of average steady-state and transient responses evoked by modulated light[J]. Electroencephalography and Clinical Neurophysiology, 1966, 20(3): 238-248.
- [4] Cheng M, Gao X, Gao S, et al. Design and implementation of a Brain-Computer Interface with high transfer rates[J]. IEEE Transactions on Biomedical Engineering, 2002, 49(10): 1181-1186.
- [5] Materka A, Byczyk M, Poryzala P. A virtual keypad based on alternate half-field stimulated visual evoked potentials[C]//Proceedings of 2007 International Symposium on Information Technology Convergence (ISITC 2007). Jeonju: IEEE, 2007: 296-300.
- [6] Lin Z, Zhang C, Wu W, et al. Frequency recognition based on canonical correlation analysis for SSVEP based BCIs[J]. IEEE Transactions on Biomedical Engineering, 2006, 52(12): 2610 - 2614.

- [7] 张莉, 何传红, 何为, 等. 基于典型相关分析与低通滤波的肌电伪迹去除[J]. 数据采集与处理, 2010, 25(2):255-258.
- [8] Chen X, Wang Y, Gao S, et al. Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain-computer interface[J]. Journal of neural engineering, 2015, 12(4): 046008.
- [9] Nakanishi M, Wang Y, Chen X, et al. Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis[J]. IEEE Transactions on Biomedical Engineering, 2017, 65(1): 104-112.
- [10] Wei Q, Li C, Wang Y et al. Enhancing the performance of SSVEP-based BCIs by combining task-related component analysis and deep neural network[J]. Scientific Reports, 2025, 15(1): 365.
- [11] Fukushima, Kunihiko. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. Biological Cybernetics, 1980, 36(4): 193-202.
- [12] Elman, Jeffrey L. Finding structure in time[J]. Cognitive Science, 1990, 14(2): 179-211.
- [13] Mirowski P, Madhavan D, LeCun Y, et al. Classification of patterns of EEG synchronization for seizure prediction[J]. Clinical Neurophysiology, 2009, 120(11): 1927-1940.
- [14] Bashivan P, Rish I, Yeasin M, et al. Learning representations from EEG with deep recurrent-convolutional neural networks[EB/OL]. (2015-11-19). <https://doi.org/10.48550/arXiv.1511.06448>.
- [15] Kwak N S, Müller K R, Lee S W. A convolutional neural network for steady state visual evoked potential classification under ambulatory environment [J]. PLOS ONE, 2017, 12(2):0172578.
- [16] Ravi A, Beni NH, Manuel J, et al. Comparing user-dependent and user-independent training of CNN for SSVEP BCI[J]. Journal of Neural Engineering, 2020, 17(2): 026028.
- [17] Lawhern V J, Solon A J, Waytowich N R, et al. EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces[J]. Journal of Neural Engineering, 2018, 15(5): 056013.
- [18] Chen J, Sun F, Zhang W, et al. Attention-based multimodal tCNN for classification of steady-state visual evoked potentials and its application to gripper control[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 35(12): 18263-18271.
- [19] Pan Y, Chen J, Zhang Y, et al. An efficient CNN-LSTM network with spectral normalization and label smoothing technologies for SSVEP frequency recognition[J]. Journal of Neural Engineering, 2022, 19(5): 056014.
- [20] Liu B, Gao H, Jiang Y, et al. Research on a soft saturation nonlinear SSVEP signal feature extraction algorithm[J]. Scientific Reports, 2024, 14(1): 17043.
- [21] Wang X, Liu A, Cui H, et al. GZSL-Lite: A lightweight generalized zero-shot learning network for SSVEP-based BCIs[J]. IEEE Transactions on Biomedical Engineering, 2025: 1-11.
- [22] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). 2017: 6000-6010.
- [23] Zhao G, Lin J, Zhang Z, et al. Sparse transformer: Concentrated attention through explicit selection[EB/OL]. (2019-12-25). <https://doi.org/10.48550/arXiv.1912.11637>.
- [24] Horuz CC, Kasenbacher G, Higuchi S, et al. The Resurrection of the ReLU[EB/OL]. (2025-05-28). <https://doi.org/10.48550/arXiv.2505.22074>.
- [25] Wang Y, Chen X, et al. A benchmark dataset for SSVEP-based brain-computer interfaces[J]. IEEE

Transactions on Neural Systems and Rehabilitation Engineering, 2016, 25(10): 1746-1752.

- [26] Chen X, Wang Y, Nakanishi M, et al. High-speed spelling with a noninvasive brain-computer interface[J]. Proceedings of the National Academy of Sciences of the United States of America, 2015, 112(44): E6058-E6067.
- [27] Refaeilzadeh P, Tang L, Liu H. Cross-validation[M]//Encyclopedia of database systems. Boston, MA: Springer, 2009: 532-538.
- [28] Wolpaw J R, Birbaumer N, McFarland D J, et al. Brain-computer interfaces for communication and control[J]. Clinical Neurophysiology, 2002, 113(6): 767-791.
- [29] Van der Maaten L, Hinton G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(86): 2579-2605.

作者简介:



张学军(1969-), 男, 博士, 教授, 硕士生导师, 研究方向: 智能信息处理、脑机接口技术、无线射频识别技术等,  
E-mail: xjzhang@njupt.edu.cn。

李夏芸(2000-), 女, 硕士研究生, 研究方向: 智能信息处理、脑电信号处理,  
E-mail: 1223024601@njupt.edu.cn。

