计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
63(1): 189−198, 2026

本期推荐

多层次特征建模与时空依赖挖掘的自监督脑电图分类

李超凡　陈松灿
（南京航空航天大学计算机科学与技术学院　南京　211106）
（模式分析与机器智能工业和信息化部重点实验室（南京航空航天大学）　南京　211106）
（lcffan@nuaa.edu.cn）

Self-Supervised EEG Classification with Multi-Level Feature Modeling and Spatiotemporal Dependence Mining

Li Chaofan and Chen Songcan
（*College of Computer Science and Technology*, *Nanjing University of Aeronautics and Astronautics*, *Nanjing* 211106）
（*Key Laboratory of Pattern Analysis and Machine Intelligence* (*Nanjing University of Aeronautics and Astronautics*), *Ministry of Industry and Information Technology*, *Nanjing* 211106）

Abstract　Multi-channel electroencephalography (EEG), as a non-invasive technique, records brain electrical activity through multiple electrodes placed on the scalp, aiding in understanding individual psychological states and assisting in the diagnosis of various diseases. Given the high cost and technical challenges of annotating large-scale EEG data, self-supervised learning (SSL), as a label-free learning paradigm, has garnered widespread attention in the EEG domain. SSL leverages the intrinsic structure of data to learn representations, thereby effectively enhancing the model's generalization ability. Despite significant advancements in current research, the field still faces the following challenges: First, multi-channel EEG data contain complex spatiotemporal correlations, yet many existing methods are confined to modeling either temporal or spatial dimensions in isolation, rather than integrating both comprehensively, thus limiting the ability to profoundly understand the intrinsic complexities of EEG signals. Second, many current approaches have not effectively integrated segment-level and instance-level information. The former contributes to improving the model's generalization ability, while the latter aids the model in better adapting to downstream (classification) tasks. To address these challenges, we propose and implement a self-supervised pre-training framework that combines contrastive and reconstruction strategies. Specifically, by introducing a channel masking strategy based on temporal mask reconstruction, our framework effectively captures the spatiotemporal relationships of EEG data. By capturing inter-segment relationships in a fine-grained manner, the framework enhances both the model's performance and its generalization ability. Meanwhile, integrating instance-level contrastive learning with masked reconstruction tasks helps the model learn representations with instance discriminability and local perceptual ability, thus facilitating better adaptation to downstream tasks. Additionally, the introduction of the self-paced learning mechanism further enhances the model's generalization ability. Finally, experimental results across multiple EEG tasks demonstrate the effectiveness of the proposed approach.

Key words　electroencephalography classification; self-supervised learning; signal reconstruction; contrastive learning; self-paced learning

收稿日期: 2025-05-26; **修回日期**: 2025-08-18
基金项目: 国家自然科学基金项目(62376126)
This work was supported by the National Natural Science Foundation of China (62376126).
通信作者: 陈松灿(s.chen@nuaa.edu.cn)

摘 要 多通道脑电图（electroencephalography, EEG）作为一种非侵入性技术，通过在头皮上布置多个电极记录大脑电活动，帮助理解个体的心理状态和辅助诊断多种疾病。鉴于标记大量EEG数据的高昂成本和技术挑战，自监督学习（self-supervised learning, SSL）作为一种无需标签的学习范式，通过挖掘数据内在结构进行学习，可有效提升模型的泛化性能，已在EEG领域获得广泛关注。尽管当前研究已取得了显著进展，但仍面临以下挑战：首先，多通道EEG数据蕴含复杂的时空关联，然而众多当前方法仅局限于时间或空间单一维度的建模，未能全面融合两者，限制了对EEG信号内在复杂特征的深入理解。其次，许多现有方法未能有效结合利用片段级与示例级信息，前者有助于提升模型的泛化能力，而后者则能帮助模型更好地适应下游（分类）任务。针对上述挑战，提出并实现一个结合对比和重建的自监督预训练框架。具体而言，在时间维度掩码重建的基础上引入通道掩码策略，有效捕捉EEG数据的时空关系，并细粒度地捕捉片段间的关系，提升了模型的性能和泛化能力。同时，结合示例级对比学习与掩码重建任务，帮助模型学习到具有示例区分性和局部感知能力的表示，从而更好地适应下游任务。此外，引入自步学习机制进一步增强了模型的泛化能力。最后，在多个EEG任务上的实验结果验证了所提方法的有效性。

关键词 脑电图分类；自监督学习；信号重构；对比学习；自步学习

中图法分类号 TP391

DOI: 10.7544/issn1000-1239.202550351 **CSTR:** 32373.14.issn1000-1239.202550351

脑电图（electroencephalography, EEG）是一种非侵入性技术，通过在头皮上布置电极来收集和记录大脑的电信号，以测量其电活动[1]。EEG信号能够反映大脑的内在神经活动，具有高时间分辨率，可以捕捉到毫秒级的大脑动态变化，广泛应用于癫痫诊断[2]、情感分析[3]、睡眠研究[4]以及脑机接口[5]等领域，具有重要的临床和应用价值。

尽管EEG技术在多个领域展现出广泛的应用前景，其仍面临若干挑战。首先，标注样本稀缺，获取大规模标注数据既耗时又昂贵，并且需要专业的医学知识和经验。其次，由于EEG信号的复杂性和多样性，不同专家对同一EEG信号片段的解读往往存在主观差异[6]，这进一步影响了模型的学习效果。此外，在某些情况下，准确了解参与者在认知神经科学实验中的想法或行为可能具有挑战性，因此很难获得可靠的标签。例如，在想象任务中，受试者可能未严格遵循指示，或者所研究的大脑活动过程（如冥想状态、情绪变化等）本身就难以通过客观指标进行准确量化[7]。因此，传统监督学习方法在EEG分析中面临显著瓶颈，难以充分发挥EEG数据的潜力。

自监督学习（self-supervised learning, SSL）作为一种无需标签的学习范式，近年来已受到广泛关注。SSL通过挖掘数据内在结构进行学习，从而有效减少对外部标签的依赖，降低了数据准备的成本和技术门槛。尤其在EEG领域，SSL能够在没有明确标签的情况下，利用数据本身的特性来训练模型，进而提升模型的泛化性能，使其成为解决EEG数据分析难题的一种有前途的方法。尽管SSL在EEG领域取得了显著进展，但仍面临2大挑战：

1）时空依赖性挖掘不足。EEG序列分析的核心在于捕捉和利用观测值中的复杂时间依赖性与通道间的相关性。大脑活动具有动态变化的特性，通过挖掘时间依赖性可更加深入了解大脑在不同状态下的活动规律，而空间依赖性则有助于分析大脑不同区域间的功能连接和网络动态性。例如，通过比较左脑和右脑异常放电的幅度和相位可定位病变；情绪状态的变化也可通过不对称的脑电活动来反映[8-10]。然而，现有方法往往仅局限于时间或空间单一维度的建模，未能充分挖掘EEG数据中的复杂时空关联。

2）未能有效整合来自不同层次的互补信息。低层次信息（即片段级信息）主要聚焦于局部、细粒度的特征变化，这些变化通常反映了大脑活动的即时、微观特性，通过捕捉信号中的瞬时变化和短期趋势，帮助模型理解EEG信号的动态模式，进而获得更好的泛化性能[11]。而高层次信息（即示例级信息）则侧重于整体语义结构和宏观模式，揭示了大脑在不同任务或状态下的全局性响应，对于理解EEG信号的整体行为至关重要。在EEG下游任务中，对未见被试的数据进行分类是一种较为常见的应用场景，多层次信息的整合显得尤为重要。鉴于分类任务的标签通常标记在整个时间序列（即示例）上，预训练阶段学习的示例级信息能够增强模型对这类任务的适应性。此外，考虑到个体间存在的变异性，片段级信息的学习对于提升模型在未知个体上的泛化性能至

关重要。

针对上述挑战，本文提出了一个结合对比学习和掩码重建的自监督学习框架，命名为MSTCR（multi-level spatiotemporal contrastive reconstruction）。为了有效挖掘EEG信号中的时空依赖性，本文针对重建任务设计了一种结合时间掩码和通道掩码的混合掩码策略。在多层次特征建模方面，为了充分利用不同层次信息间的互补优势，本文将数据输入主干编码器后获得低层特征用于重建任务，以捕捉EEG信号的局部依赖关系，接着将低层特征沿时间维度池化后再输入线性层得到高层特征，然后通过对比任务学习高层次的语义信息，由于对比和重建任务在不同的特征空间中执行，因此避免了优化过程中可能出现的任务间冲突。

此外，针对随机掩码导致的样本重建难度分布不均问题，本文引入自步学习（self-paced learning, SPL）构建“由易到难”的训练机制，引导模型从简单掩蔽样本的基础特征入手，逐步适应复杂场景，有效降低局部噪声对重建的干扰，进一步提升模型的泛化能力。

本文的主要贡献包含4个方面：

1）设计了一个适用于多种EEG任务的统一的预训练框架，通过对比和重建任务优化不同层次信息的学习，增强EEG的表征；

2）通过集成时间域和空间域的重建任务，有效学习并捕捉到了EEG数据中复杂的时空依赖性，进一步提升了对EEG信号内在特征的理解；

3）引入自步学习构建“由易到难”的训练机制，增强重建任务对高难度掩蔽样本的鲁棒性；

4）多个EEG任务的实验表明本文方法在绝大多数情况下优于现有方法，尤其在线性探测任务中的AUROC平均提升6.33个百分点。

1 相关工作

现有EEG自监督学习算法根据预置任务（pretext task）的不同大致分为4类[12]，分别是基于预测的方法[7,13]、基于对比的方法[14-17]、基于重建的方法[18-21]和混合的方法[22-23]。基于预测的方法不需要太多的参数和复杂的网络结构，通过创建分类预置任务，预测从无标签数据中生成的离散伪标签，以此学习有效的特征。然而，这类方法通常需要与其他任务相结合才能学习到下游任务的通用表示[12]，本文的框架结合了对比方法和重建方法，以下将主要围绕这2种方法展开讨论。

1.1 基于对比的方法

基于对比的方法通过拉近正样本对和推离负样本对以学习判别性表示，其核心挑战在于正负样本的构建与优化。在正样本构建方面，现有方法主要分为2类：一类是基于时间邻近假设的上下文依赖型策略，例如TNC[14]和TS-TCC[15]将相邻EEG片段视为正样本，但受限于EEG信号的非平稳性，此类方法可能引入采样偏差并难以捕捉全局上下文；另一类则通过数据增广生成多视图来构建正样本，例如SeqCLR[16]采用幅度缩放和时间平移，而TimesURL[24]提出频率-时间混合增广策略，通过保留时序语义关系提升跨任务泛化能力。针对负样本优化，传统方法依赖大规模负样本对以提升判别性，但面临计算成本高昂的问题。为此，“免负样本”设计成为研究热点，例如基于师生架构的BYOL[25]和DINO[26]利用自蒸馏机制防止表示坍缩，而VICReg[27]则通过方差-协方差正则化约束在防止表示坍缩的同时增强表示多样性。然而，现有对比方法通常聚焦于高层语义特征学习，导致低层特征表示趋同，限制了模型对EEG细粒度模式的感知能力。为此，本文提出结合重建任务与对比学习的协同优化框架，通过重建损失引导低层特征学习，以弥补单一对比目标的局限性。

1.2 基于重建的方法

基于重建的方法通过利用未掩码上下文信息恢复被掩蔽的序列片段[28-30]，侧重于捕捉序列的局部关联。此类方法因无需依赖强先验假设且泛化能力突出而备受关注。受BERT[28]模型成功启发，BENDR[18]采用了掩码自编码器方法，通过在特征空间最小化掩蔽区域与重建区域的表示差异实现特征学习。与之不同，MAEEG[19]直接在原始信号空间实施重建，将高层表示映射回原始EEG维度，并计算原始信号与重建信号的均方误差以优化模型。最近，Foumani等人[21]采用联合嵌入预测框架[31]（joint-embedding predictive architecture, JEPA）从表示层面进行重建，结合语义子序列保留（semantic subsequence preserving, SSP）算法生成信息丰富的掩蔽输入来指导模型提取具有丰富语义的表征。

然而受限于架构特性，上述模型虽然能够有效捕获时间维度动态特性，但在捕捉EEG通道间的空间依赖关系方面存在局限。相比之下，MMM[20]专注于脑电数据的空间和拓扑建模，打破了不同脑电拓扑之间的界限，但其时间建模仅依赖微分熵特征[32]静态聚合，难以刻画EEG信号的动态演化规律。鉴于现有EEG自监督方法时空表征割裂的问题，本文

通过设计混合掩码策略,结合通道-时间双维度掩码重建任务,有效捕捉多通道 EEG 数据复杂的时空依赖关系,从而突破现有方法单一维度建模的局限。

2 问题定义

EEG 信号定义为$X=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$,其中$n$表示 EEG 信号的样本个数。EEG 信号通常通过固定在头部的若干电极,以相同采样频率同步采集获得。本文的目标是在没有外部标签的情况下从一个给定的 EEG 数据集X中学习一个非线性嵌入函数$f_\theta:\mathbb{R}^{n\times T\times C}\rightarrow\mathbb{R}^{n\times d}$,将每个 EEG 样本$\boldsymbol{x}_i$映射到一个更加紧凑的表示$\boldsymbol{r}_i$中,然后通过线性探测和全参数微调 2 种方式评估所学表示的质量。1)线性探测:通过在习得的表示上训练一个简单的逻辑回归分类器来衡量表示的可分性;2)全参数微调:在下游分类任务中对整个模型进行微调,通过端到端训练后的性能来评估表示的泛化能力。

3 基于对比和重建的 EEG 自监督框架

3.1 方法介绍

如图 1 所示,本文首先对原始 EEG 序列$\boldsymbol{x}$进行数据增广得到增广后的数据$\boldsymbol{x}'$。接着,对$\boldsymbol{x}$和$\boldsymbol{x}'$分别实施时空掩码操作,生成相应的掩蔽序列$\tilde{\boldsymbol{x}}$和$\tilde{\boldsymbol{x}}'$,其中,$\boldsymbol{x}$和$\boldsymbol{x}'$用于对比任务,旨在挖掘示例级相似信息,而$\tilde{\boldsymbol{x}}$和$\tilde{\boldsymbol{x}}'$则用于重建任务,旨在重建被掩蔽的原始序列。随后,将$\boldsymbol{x}$, $\boldsymbol{x}'$, $\tilde{\boldsymbol{x}}$, $\tilde{\boldsymbol{x}}'$输入主干编码器——时间卷积网络(temporal convolutional networks, TCN),以获取它们在隐空间的表示$\boldsymbol{r}$, $\boldsymbol{r}'$, $\tilde{\boldsymbol{r}}$, $\tilde{\boldsymbol{r}}'$。为了缓解同时优化对比和重建目标可能带来的梯度干扰问题[12],本文将对比和重建任务约束在不同层次进行。对于对比任务,将$\boldsymbol{r}$和$\boldsymbol{r}'$沿时间维度进行最大池化,然后输入一个线性层,从而得到最终的表示$\boldsymbol{h}$和$\boldsymbol{h}'$。这一过程确保了对比任务在高层特征空间进行,这些特征能够反映数据的全局结构和语义信息。而对于重建任务,则将$\tilde{\boldsymbol{r}}$和$\tilde{\boldsymbol{r}}'$输入一个多层感知机(multilayer perceptron, MLP)得到$\boldsymbol{x}_{\mathrm{re}}$和$\boldsymbol{x}'_{\mathrm{re}}$,以重建原始序列中被掩蔽的部分。这一过程在低层特征空间进行,这些特征更侧重于捕捉数据的局部和细节信息。通过在不同的特征空间中执行这些任务,不仅能减少梯度干扰,还能更好地利用每个任务的优势,从而生成更加有效和鲁棒的特征表示。

3.2 对比模块

在 EEG 数据的下游应用中,对即时脑电状态进行分类是一项常见任务,其个体序列/示例通常以整体赋予单个标签。然而,现有基于局部片段的预训练学习范式忽略了序列/示例整体级的全局信息,从而制约了下游任务的适应性,并进而影响了后续微调迁移的性能。为此,本文首先沿时间维度对主干编码器提取的局部(低层)特征进行池化处理,并通过可学的线性投影将其映射到高层语义空间。然后,从数据增广和负样本处理 2 个角度设计示例层面的对比模块,迫使网络捕获具有类别判别性的全局特征。

针对数据增广,现有许多方法直接照搬了计算机视觉与自然语言处理领域的既有策略,忽视了 EEG 序列的独特性,由此引入了不恰当的归纳偏差,导致时间依赖性和语义一致性的破坏。为此,本文借鉴 Liu 等人[24]提出的 FTAug 方法,结合随机裁剪与频率混合策略,利用时域和频域的联合优势来有效减少归纳偏差并保持增广数据的语义一致性。考虑到随机裁剪生成的样本长度变化较大,可能影响

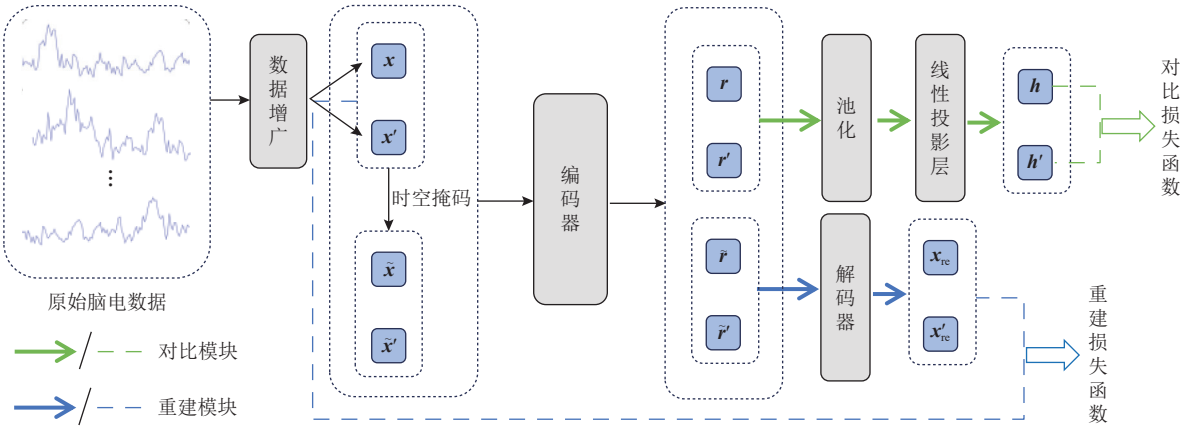


Fig. 1 Framework of MSTCR model

图 1 MSTCR 模型框架

后续重建任务中掩码块大小的设定，为此对裁剪后的样本设置最小长度阈值约束，确保其在合理范围内。对于频率混合操作，为避免丢失关键频域的特征信息，本文保留了幅值最大的k个频率成分，而不对其实施交换操作，从而维持数据在频域上的语义一致性。

针对负样本处理，传统对比方法依赖负样本增强特征判别性并防止表示坍塌，但面临如下挑战：其一，EEG信号易受眼动、肌电等伪迹干扰，导致负样本存在较高的假阴性风险；其二，基于复杂筛选策略的难负例挖掘技术显著增加计算成本。为此，本文采用VICReg[27]方法构建对比损失函数，摆脱对负样本的依赖，通过构建不变性-方差-协方差3重约束实现稳健表征学习，在防止表示坍塌的同时显著增强特征表示的多样性，具体的损失函数$\mathcal{L}_{\mathrm{cl}}$定义为

$$\mathcal{L}_{\mathrm{cl}}(\boldsymbol{h},\boldsymbol{h}')=\lambda s(\boldsymbol{h},\boldsymbol{h}')+\mu[v(\boldsymbol{h})+v(\boldsymbol{h}')]+\gamma[c(\boldsymbol{h})+c(\boldsymbol{h}')]\text{。} \tag{1}$$

其中，$s(\boldsymbol{h},\boldsymbol{h}')$是不变性损失项，旨在约束同一样本不同增广视图间的一致性；$v(\boldsymbol{h})$和$v(\boldsymbol{h}')$是方差损失项，通过维持同批次样本嵌入向量各维度的方差下界，强制同批次不同样本间的特征相异，从而避免模型坍塌；$c(\boldsymbol{h})$和$c(\boldsymbol{h}')$是协方差损失项，通过对嵌入向量协方差矩阵的非对角元素施加惩罚，消除不同维度间的冗余，以增强特征表示的多样性；λ，μ，γ是调节这3项损失的超参数。

3.3 重建模块

掩码重建任务通过构建局部信息缺失的输入样本（遮蔽数据片段），迫使模型基于上下文相关性推理完成缺失部分重建，从而学习数据的内在结构和鲁棒特征表示。针对多通道EEG数据的掩码重建任务，其实现方式依据信号维度可分为时域掩码重建与空域掩码重建2类。前者沿时间轴遮蔽连续时间段，通常利用Transformer模型学习时间依赖性，捕捉EEG信号的短期趋势与长程依赖；后者在通道维度遮蔽部分电极节点，通常采用图神经网络（graph neural networks, GNN）学习脑区间的空间关联。然而由于架构设计差异，Transformer对电极空间拓扑的建模能力较弱，而GNN难以有效捕获跨时间尺度的动态节律特征。为兼顾时序的动态建模与空间的拓扑关联，本文基于TCN[33]主干设计混合掩码策略TSP-RCM（temporal semantic-preserving and random channel masking），如图2所示，实现EEG时空依赖的联合建模。时间维度上，由于EEG时变语义信息需通过连续时间窗口捕捉[34]，本文借鉴SSP[21]方法，通过控制连续可见子片段的最小长度，确保关键神经响应模式的完整性。首先根据给定的总时间步长T、掩码率r和掩码块数N计算保留块的大小L。L的计算为：

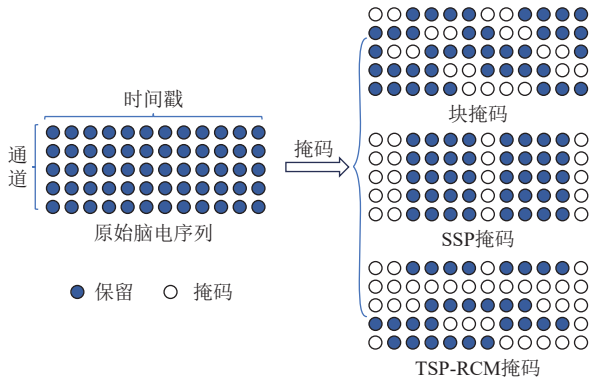


Fig. 2 Illustration of different masking strategies

图2 不同掩码策略例示意图

$$L=\frac{(1-r)\times T}{N}\text{。} \tag{2}$$

接着，随机选择N个起始点，并以各起始点为中心对称扩展得到长度为L的保留区域，然后掩蔽区域外时间点，以此保证掩蔽后序列的时域动态连续性及局部节律完整性。

空间维度上，针对SSP方法采用全通道统一掩码策略（SSP掩码）导致难以建模跨通道空间关联，本文进一步提出通道差异化掩码和随机通道完全掩码策略。具体而言，通道差异化掩码独立生成各通道的掩码位置与长度，当某通道的局部时段被掩蔽时，模型可依据其他通道在同一时间点的未掩蔽信号进行重建，驱使模型结合空间上下文进行重建，此外，通过更加多样化的掩码模式丰富时域特征的学习，避免模型对固定掩码位置的过拟合；随机通道完全掩码策略则通过随机选择若干通道进行完全遮蔽，迫使模型利用剩余未被遮蔽通道的信息来重建全脑功能网络，从而增强模型对全局空间依赖关系的理解。

最后，将掩蔽序列的表示输入解码器，重建被掩蔽部分的原始信号，总的重建损失为：

$$\mathcal{L}_{\mathrm{recon}}=\frac{1}{|\mathcal{B}|}\sum_{i}\left\|m_i\odot\left(\boldsymbol{x}_{i,\mathrm{re}}-\boldsymbol{x}_i\right)\right\|_2^2+\frac{1}{|\mathcal{B}|}\sum_{i}\left\|m_i'\odot\left(\boldsymbol{x}_{i,\mathrm{re}}'-\boldsymbol{x}_i'\right)\right\|_2^2, \tag{3}$$

其中$m_i\in\{0,1\}^{T\times C}$和$m_i'\in\{0,1\}^{T\times C}$是二值掩码矩阵，$m_i=1$和$m_i'=1$表示EEG数据对应位置被掩蔽。

3.4 自步学习机制

传统掩码重建任务中，随机掩蔽策略因缺乏先验知识引导，易生成与任务无关的高噪声掩蔽样本。

具体表现为:时域掩码可能掩蔽任务相关的关键区段,而空域掩码可能选择与其他通道相关性较弱的通道实施掩蔽。此类掩蔽样本会误导模型过度关注局部噪声,而非有效捕获泛化的生理特征。为此,本文引入自步学习框架,通过构建动态难度评估函数与权重分配策略,实现“简单→复杂”的课程化训练机制。具体来说,首先根据样本的重建损失评估样本难度,再依据该损失计算样本权重,计算公式为:

$$v_i = \exp\left(-\frac{\|m_i \odot (\boldsymbol{x}_{i,\mathrm{re}} - \boldsymbol{x}_i)\|_2^2}{\delta}\right), \tag{4}$$

其中年龄参数δ通过控制权重分布锐度调整学习速率,从式(4)可以看出,随着样本重建损失的增加,赋予的权重也会相应的降低。将权重项融入式(4)的重建损失,得到自步重建损失:

$$\mathcal{L}_{\mathrm{sprecon}} = \frac{1}{|\mathcal{B}|}\sum_i v_i \|m_i \odot (\boldsymbol{x}_{i,\mathrm{re}} - \boldsymbol{x}_i)\|_2^2 + \frac{1}{|\mathcal{B}|}\sum_i v_i' \|m_i' \odot (\boldsymbol{x}_{i,\mathrm{re}}' - \boldsymbol{x}_i')\|_2^2 。 \tag{5}$$

结合自步学习之后,模型的整体目标函数为

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\mathrm{cl}} + \lambda_2 \mathcal{L}_{\mathrm{sprecon}}, \tag{6}$$

其中λ_1, λ_2是调节2项损失的超参数。

在实验过程中,首先采用初始重建损失$\mathcal{L}_{\mathrm{recon}}$训练1~5轮,目的是使模型初步习得评估样本难度的能力,之后采用自步重建损失$\mathcal{L}_{\mathrm{sprecon}}$逐步将训练重点从简单样本过渡到复杂样本。

4 实验和结果

4.1 数据集

本文选取EEG领域广泛使用的4个标准数据集进行实验,覆盖多个常见脑电下游任务,分别是Crowdsourced[35], STEW[36], DREAMER[37], TUAB[38],关于数据集的详细信息如表1所示。

Table 1 Detailed Information of Datasets

表1 数据集详细信息

数据集	采样率/Hz	通道数	长度/s	样本数	任务
Crowdsourced	128	14	2	12 296	睁/闭眼检测
STEW	128	14	2	26 136	脑力负荷分类
DREAMER	128	14	2	77 910	情感识别
TUAB	256	16	10	409 455	异常脑电分类

对于数据集的预处理和分割遵循EEG2Rep[21]模型的策略,采用受试者独立(subject-independent)策略将数据集划分为训练集、验证集和测试集,评估模型在跨个体生理差异下的泛化性能。

4.2 评价指标

本文采用准确率(accuracy, ACC)和受试者工作特征曲线下面积(area under the receiver operating characteristic curve, AUROC)作为评价指标,以评估所提方法的性能。其中ACC通过计算正确分类样本占比反映模型全局精度,但对类别不平衡数据敏感;AUROC则以ROC曲线下面积量化模型在不同决策阈值下的综合性能,该曲线通过遍历分类阈值绘制而成,对类别分布不均衡具有鲁棒性。

4.3 基线及实验设置

为验证所提方法的有效性,本文选取6个现有EEG自监督学习方法进行对比:基于对比的TS-TCC[15], TF-C[39], BIOT[40],以及基于重建的BENDR[18], MAEEG[19], EEG2Rep[21]。实验覆盖静息态(Crowdsourced)、认知负荷(STEW)、情感识别(DREAMER)及癫痫检测(TUAB)四类典型脑电任务。其中,前3个数据集通过5次随机种子实验取均值±标准差报告结果,TUAB数据集因计算资源限制,仅进行单次实验,鉴于其数据量丰富且涵盖大量受试者,预计单次实验的结果比较稳定。模型实现基于PyTorch框架,以TCN[33]为主干网络,采用Adam优化器进行20轮预训练,初始学习率为0.001,批次大小设定为128。对比损失超参数λ, μ, γ均设置为1(与原始VICReg[27]保持一致),总损失超参数经网格搜索确定,其中λ_1设置为0.01, λ_2设置为1。

4.4 线性探测实验结果

表2展示了各EEG自监督学习算法在不同数据集上进行线性探测任务的性能评估结果。结果表明,MSTCR在不同被试间变异性(intersubject variability, ISV)场景下均优于其他算法。在STEW和DREAMER数据集中,因认知任务复杂且受试者规模有限,被试间差异较大,MSTCR通过时空掩码重建模块提取跨被试不变性特征,提升了泛化性能。相比之下,Crowdsourced数据集因睁/闭眼检测任务较为简单,被试间差异较小,TUAB数据集因其包含大量患者样本,有效降低被试间差异的影响,在这2个低ISV数据集上,MSTCR借助示例级对比模块增强模型的判别能力,仍保持优越性能。

总体而言,MSTCR结合对比与重建任务的优势,在不同ISV场景下均表现突出,尤其在数据受限时显著提升模型泛化能力。在Crowdsourced, STEW, DREAMER数据集上,AUROC指标相比次优方法分别提高10个百分点、10.69个百分点、5.44个百分点。

Table 2 Performance Comparison of Linear Probing on Different Datasets

表 2 不同数据集上线性探测性能比较 %

方法	Crowdsourced		STEW		DREAMER		TUAB	
	ACC	AUROC	ACC	AUROC	ACC	AUROC	ACC	AUROC
BENDR[18]	70.46±4.14	70.56±4.32	63.03±1.07	63.03±1.07	51.48±1.87	51.68±1.98	72.78	79.87
MAEEG[19]	75.21±2.11	75.01±2.01	67.99±1.86	68.58±1.88	54.24±2.04	53.98±2.68	72.62	79.75
TS-TCC[15]	77.75±2.94	77.83±2.11	64.54±1.62	64.64±1.73	53.60±2.68	53.87±2.28	74.39	81.02
TF-C[39]	64.82±7.23	65.32±6.12	58.84±2.36	58.69±2.43	52.52±1.88	52.26±2.08	69.33	75.75
BIOT[40]	76.23±4.56	76.33±4.12	67.54±2.08	67.69±3.60	53.35±2.41	53.42±1.91	75.11	82.92
EEG2Rep[21]	81.66±2.93	81.67±2.65	69.04±1.04	69.10±1.23	**58.45±1.82**	55.19±1.95	**76.55**	**83.24**
MSTCR（本文）	**82.74±2.07**	**89.84±1.85**	**70.86±0.67**	**76.49±0.82**	58.18±4.19	**58.19±3.92**	75.25	82.55

注：黑体数值表示最优结果，下划线数值表示次优数值；“±”前数值表示平均值，“±”后数值表示标准差。

4.5 全参数微调实验结果

表 3 报告了全参数微调任务的性能评估结果，分析可知，MSTCR 在多数场景下取得最先进性能，但在特定场景表现次优。一方面，对于任务简单且个体间偏移较小的场景（如 Crowdsourced 数据集），预训练泛化效益有限，性能更依赖模型架构与数据的匹配程度；另一方面，对于 DREAMER 这类个体偏移显著的数据集，因个体认知模式差异（强个体特异性）及情感状态主观报告的不确定性（高标签噪声），全参数微调时模型易过拟合个体噪声，过度调整主干网络参数破坏预训练阶段学习到的通用特征，导致性能下降。

Table 3 Performance Comparison of Fine Tuning of Full Parameter on Different Datasets

表 3 不同数据集上全参数微调性能比较 %

方法	Crowdsourced		STEW		DREAMER		TUAB	
	ACC	AUROC	ACC	AUROC	ACC	AUROC	ACC	AUROC
BENDR[18]	83.78±2.35	83.8±2.63	69.74±2.11	69.77±2.03	54.45±2.11	53.02±3.11	76.96	83.97
MAEEG[19]	86.75±3.50	86.21±3.41	72.46±3.67	72.5±3.22	53.63±2.61	52.08±2.36	77.56	86.56
TS-TCC[15]	89.22±1.22	89.22±1.22	71.00±2.98	71.03±3.02	58.16±2.11	55.05±1.79	79.66	87.02
TF-C[39]	82.93±4.02	82.90±4.32	68.65±1.12	68.70±1.75	52.82±1.66	52.86±1.85	72.33	78.48
BIOT[40]	87.95±3.52	87.78±3.09	69.88±2.15	70.11±2.57	53.45±2.01	53.53±2.13	79.21	87.42
EEG2Rep[21]	**94.13±2.11**	**94.13±2.17**	73.60±1.47	74.40±1.50	60.37±1.52	**59.42±1.45**	80.52	**88.43**
MSTCR（本文）	87.02±0.91	93.58±0.85	**74.86±1.62**	**85.35±1.01**	**60.69±1.29**	52.71±0.96	**80.76**	87.71

注：黑体数值表示最优结果，下划线数值表示次优数值；“±”前数值表示平均值，“±”后数值表示标准差。

4.6 计算复杂性比较

本节在 Crowdsourced 数据集上开展计算复杂性对比分析，通过对比模型参数量、总训练时长及推理耗时（原始数据编码为特征表示的时间开销）分析 MSTCR 的计算开销。如表 4 所示，MSTCR 以 TCN 为主干网络，额外引入 MLP 作为解码器，参数量相较于 EEG2Rep 的 CNN+Transformer 架构增加约 10 倍，但 TCN 规避了 Transformer 固有的 2 次计算复杂度，推理时间仅增长 22.48%，训练耗时基本持平。分析表明，MSTCR 的计算开销主要来自主干 TCN，结合对比、重建和自步学习的训练策略未对总训练耗时产生显著影响。值得强调的是，除 TUAB 长时程脑电实验外，所有实验基于单张 RTX4060（8 GB）显卡完成，证明 MSTCR 的实际应用可行性。

Table 4 Computational Complexity Comparison of Different Methods on Crowdsourced Dataset

表 4 Crowdsourced 数据集上不同方法计算复杂性比较

方法	参数量	训练时长/min	推理耗时/s
BIOT[40]	3 318 272	18.56	0.94
EEG2Rep[21]	96 550	19.89	1.29
MSTCR（本文）	1 071 388	20.85	1.58

4.7 参数敏感性分析

本节在 Crowdsourced 数据集上进行超参数敏感

性分析，首先通过改变掩码率r和掩码块数N来探讨它们对下游分类任务的影响。为了排除对比损失的干扰，将对比损失权重λ_1设置为0。如图3所示，当掩码率r设置为0.75时，EEG序列的语义信息被严重破坏，导致性能显著下降。相比之下，当掩码块数N设置为5且掩码率设置为0.05时，EEG序列的语义完整性得以较好保留，模型能获得足够的上下文信息来支持重建，同时，多掩码块的重叠特性导致实际被掩蔽的时间点比例高于预设值，避免模型因信息冗余而退化为简单插值，确保了重建任务能保持适度的复杂度，最终使模型性能达到最优。

此外，本节还对自步学习年龄参数δ进行敏感性分析，展示了其在不同取值下分类性能的变化情况，如图4所示。当$\delta=0.5$时，模型性能出现波动，这是由于此时δ的取值恰好处于对比损失与重建损失的平衡敏感阈值区间。当$\delta<0.5$时，对比损失在优化过程中占据主导地位，导致其在训练前期削弱了对低层次细粒度特征的有效建模，进而引发分类性能的显著下降。而当$\delta>1$时，模型性能趋于稳定，对年龄参数δ的变化不再敏感。

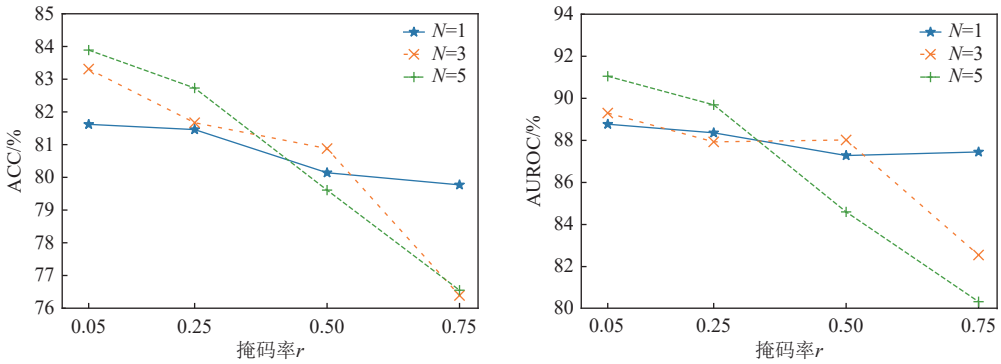


Fig. 3 Sensitivity analysis of mask block number and mask rate

图3 掩码块数及掩码率敏感性分析

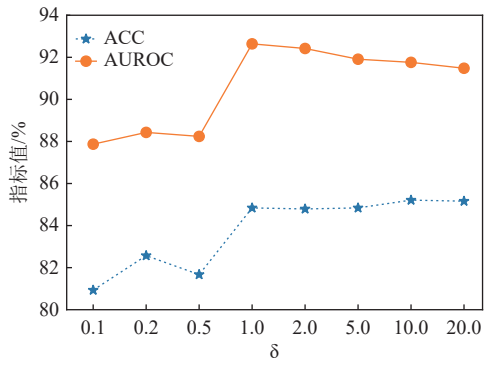


Fig. 4 Sensitivity analysis of age parameter in self-paced learning

图4 自步学习年龄参数敏感性分析

4.8 消融实验

本节在Crowdsourced数据集上通过线性探测任务进行消融实验，首先对损失函数进行消融分析，分别验证对比损失、重建损失及其组成部分的有效性，实验结果如表5所示。由实验结果可知，单独使用对比损失或重建损失均会导致性能显著下降，这说明通过对比和重建任务所习得的知识是互补的，表明同时利用片段级和示例级信息的重要性。对于对比任务，通过添加额外的线性层可以有效减少与重建任务间的冲突，减少优化时的梯度干扰。对于重建任务，在引入自步学习动态赋权机制之后，有效提升性能指标并降低方差，有效缓解因随机掩码导致样本间重建难度不均的问题，进一步提升模型的泛化能力。

Table 5 Ablation Study on Crowdsourced Dataset

表5 Crowdsourced数据集上的消融实验 %

方法	ACC	AUROC
w/o 对比损失	79.47±2.12	85.42±3.33
w/o 对比投影头	80.57±1.84	87.55±2.94
w/o 重建损失	74.57±1.90	81.32±2.03
w/o 自步学习	81.10±2.35	87.89±2.47
MSTCR（本文）	**82.74±2.07**	**89.84±1.85**

注：w/o表示去掉对应的组成成分，黑体数值表示最优结果，“±”前数值表示平均值，“±”后数值表示标准差。

此外，为评估本文所提掩码策略的有效性，将TSP-RCM与通道差异块掩码、SSP掩码、通道差异SSP掩码3种策略进行对比实验，结果如表6所示。其中，通道差异块掩码指的是在原始序列上随机选择块进行掩码；SSP掩码则相反，通过选择块进行保留以最大程度维持掩码后序列的语义完整性；“通道差异”表示不同通道的掩码位置不同。对比通道差

异块掩码与通道差异 SSP 掩码的结果可见，对序列数据掩蔽时，保证其语义信息不被破坏的重要性；SSP 掩码与通道差异 SSP 掩码的结果对比表明，通道间采取不同掩码位置可获得更优性能。TSP-RCM 在通道差异 SSP 掩码的基础上，进一步随机选择通道进行完全掩蔽以学习全局空间依赖关系，最终实现了最佳性能。

Table 6 Performance Comparison of Masking Strategies

表 6 掩码策略的性能比较 %

掩码策略	ACC	AUROC
通道差异块掩码	78.89±1.05	85.83±0.40
SSP 掩码	76.24±1.47	83.01±2.00
通道差异 SSP 掩码	81.14±2.96	87.75±3.33
TSP-RCM（本文）	**82.74±2.07**	**89.84±1.85**

注：黑体数值表示最优结果，“±”前数值表示平均值，“±”后数值表示标准差。

5 结 论

本文针对现有 EEG 自监督学习方法在时空建模方面存在不足，且未能充分利用脑电序列不同层次信息的问题，提出了一种结合对比学习和掩码重建的自监督预训练框架。该框架通过引入时空混合掩码策略，有效捕捉 EEG 数据中的时空关联；同时，通过精心设计的架构引导对比和重建任务在不同层次上进行，确保模型充分发挥各任务优势，有效学习多层次信息，从而增强模型在多种应用场景下的适应性。在多个典型 EEG 任务上进行实验评估，验证了所提方法的有效性。

本文的核心贡献在于构建了适用于多场景的自监督学习脑电表征框架，为脑电信号跨个体迁移提供了通用且鲁棒的技术基础。为进一步提升预训练模型在不同数据场景下的适配能力，未来工作将结合目标任务的生理机制与数据分布特征（如信号噪声水平）设计针对数据特性的微调迁移策略，以更好地满足实际应用需求。

作者贡献声明：李超凡提出算法模型，制定实验方案并完成实验，撰写和修改论文；陈松灿提出研究方向和指导意见，并修改论文。

参 考 文 献

[1] Teplan M. Fundamentals of EEG measurement[J]. Measurement Science Review, 2002, 2(2): 1−11

[2] Gotman J. Automatic recognition of epileptic seizures in the EEG[J]. Electroencephalography and Clinical Neurophysiology, 1982, 54(5): 530−540

[3] Bos D O. EEG-based emotion recognition[J]. The Influence of Visual and Auditory Stimuli, 2006, 56(3): 1−17

[4] Rodenbeck A, Binder R, Geisler P, et al. A review of sleep EEG patterns. Part I: A compilation of amended rules for their visual recognition according to Rechtschaffen and Kales[J]. Somnologie, 2006, 10(4): 159−175

[5] Jia Ziyu, Lin Youfang, Liu Tianhang, et al. Motor imagery classification based on multiscale feature extraction and squeeze-excitation model[J]. Journal of Computer Research and Development, 2020, 57(12): 2481−2489 (in Chinese)
(贾子钰, 林友芳, 刘天航, 等. 基于多尺度特征提取与挤压激励模型的运动想象分类方法[J]. 计算机研究与发展, 2020, 57(12): 2481−2489)

[6] Younes M, Raneri J, Hanly P. Staging sleep in polysomnograms: Analysis of inter-scorer variability[J]. Journal of Clinical Sleep Medicine, 2016, 12(6): 885−894

[7] Banville H, Chehab O, Hyvärinen A, et al. Uncovering the structure of clinical EEG signals with self-supervised learning[J]. Journal of Neural Engineering, 2021, 18(4): 046020

[8] Davidson R J, Fox N A. Asymmetrical brain activity discriminates between positive and negative affective stimuli in human infants[J]. Science, 1982, 218(4578): 1235−1237

[9] Lin Yuanpin, Yang Yihsuan, Jung T P. Fusion of electroencephalographic dynamics and musical contents for estimating emotional responses in music listening[J]. Frontiers in Neuroscience, 2014, 8: 94

[10] Zheng Weilong, Lu Baoliang. 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

[11] Atienza A, Bardram J, Puthusserypady S. Contrastive learning is not optimal for quasiperiodic time series[C]//Proc of the 33rd Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2024: 3661–3668

[12] Weng Weining, Gu Yang, Guo Shuai, et al. Self-supervised learning for electroencephalogram: A systematic survey[J]. arXiv preprint, arXiv: 2401.05446, 2024

[13] Cai Donghong, Chen Junru, Yang Yang, et al. MBrain: A multi-channel self-supervised learning framework for brain signals[C]//Proc of the 29th ACM SIGKDD Conf on Knowledge Discovery and Data Mining. New York: ACM, 2023: 130−141

[14] Tonekaboni S, Eytan D, Goldenberg A. Unsupervised representation learning for time series with temporal neighborhood coding[J]. arXiv preprint, arXiv: 2106.00750, 2021

[15] Eldele E, Ragab M, Chen Zhenghua, et al. Time-series representation learning via temporal and contextual contrasting[C]//Proc of the 30th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2021: 2352−2359

[16] Mohsenvand M N, Izadi M R, Maes P. Contrastive representation learning for electroencephalogram classification[C]//Proc of the 6th

Machine Learning for Health Workshop. Cambridge, MA: MIT, 2020: 238−253

[17] Kan Haoning, Yu Jiale, Huang Jiajin, et al. Self-supervised group meiosis contrastive learning for eeg-based emotion recognition[J]. Applied Intelligence, 2023, 53(22): 27207−27225

[18] Kostas D, Aroca-Ouellette S, Rudzicz F. BENDR: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data[J]. Frontiers in Human Neuroscience, 2021, 15: 653659

[19] Chien H Y S, Goh H, Sandino C M, et al. MAEEG: Masked auto-encoder for EEG representation learning[J]. arXiv preprint, arXiv: 2211.02625, 2022

[20] Yi Ke, Wang Yansen, Ren Kan, et al. Learning topology-agnostic EEG representations with geometry-aware modeling[C]//Proc of the 37th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2023: 53875−53891

[21] Foumani N M, Mackellar G, Ghane S, et al. EEG2Rep: Enhancing self-supervised EEG representation through informative masked inputs[C]//Proc of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM, 2024: 5544−5555

[22] Ho T K K, Armanfard N. Self-supervised learning for anomalous channel detection in EEG graphs: Application to seizure analysis[C]//Proc of the 37th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2023: 7866−7874

[23] Li Yamei, Luo Shengqiong, Zhang Haibo, et al. MtCLSS: Multi-task contrastive learning for semi-supervised pediatric sleep staging[J]. IEEE Journal of Biomedical Health Informatics, 2022, 27(6): 2647−2655

[24] Liu Jiexi, Chen Songcan. TimesURL: Self-supervised contrastive learning for universal time series representation learning[C]//Proc of the 38th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2024: 13918−13926

[25] Grill J-B, Strub F, Altché F, et al. Bootstrap your own latent: A new approach to self-supervised learning[C]//Proc of the 34th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2020: 21271−21284

[26] Caron M, Touvron H, Misra I, et al. Emerging properties in self-supervised vision transformers[C]//Proc of the 18th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2021: 9650−9660

[27] Bardes A, Ponce J, LeCun Y. VICReg: Variance-invariance-covariance regularization for self-supervised learning[J]. arXiv preprint, arXiv: 2105.04906, 2021

[28] Devlin J. BERT: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint, arXiv: 1810.04805, 2018

[29] He Kaiming, Chen Xinlei, Xie Saining, et al. Masked autoencoders are scalable vision learners[C]//Proc of the 40th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 16000−16009

[30] Baevski A, Hsu W N, Xu Qiantong, et al. Data2vec: A general framework for self-supervised learning in speech, vision and language[C]//Proc of the 39th Int Conf on Machine Learning. New York: ACM, 2022: 1298−1312

[31] Assran M, Duval Q, Misra I, et al. Self-supervised learning from images with a joint-embedding predictive architecture[C]//Proc of the 41st IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 15619−15629

[32] Shi Lichen, Jiao Yingying, Lu Baoliang. Differential entropy feature for EEG-based vigilance estimation[C]//Proc of the 35th Annual Int Conf of the IEEE Engineering in Medicine and Biology Society. Piscataway, NJ: IEEE, 2013: 6627−6630

[33] Bai S, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. arXiv preprint, arXiv: 1803.01271, 2018

[34] Wu Haixu, Hu Tengge, Liu Yong, et al. TimesNet: Temporal 2D-variation modeling for general time series analysis[C/OL]//Proc of the 11st Int Conf on Learning Representations. 2023[2025-03-02]. https://openreview.net/forum?id=ju_Uqw384Oq

[35] Williams N S, King W, Mackellar G, et al. Crowdsourced EEG experiments: A proof of concept for remote EEG acquisition using EmotivPRO Builder and EmotivLABS[J]. Heliyon, 2023, 9(8): e18433

[36] Lim W L, Sourina O, Wang Lipo. STEW: Simultaneous task EEG workload data set[J]. IEEE Transactions on Neural Systems Rehabilitation Engineering, 2018, 26(11): 2106−2114

[37] Katsigiannis S, Ramzan N. DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices[J]. IEEE Journal of Biomedical Health Informatics, 2017, 22(1): 98−107

[38] Lopez S, Suarez G, Jungreis D, et al. Automated identification of abnormal adult EEGs[C]//Proc of the 2015 IEEE Signal Processing in Medicine and Biology Symp (SPMB'2015). Piscataway, NJ: IEEE, 2015: 1−5

[39] Zhang Xiang, Zhao Ziyuan, Tsiligkaridis T, et al. Self-supervised contrastive pre-training for time series via time-frequency consistency[C]//Proc of the 36th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2022: 3988−4003

[40] Yang Chaoqi, Westover M, Sun Jimeng. BIOT: Biosignal transformer for cross-data learning in the wild[C]//Proc of the 37th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2023: 78240−78260



Li Chaofan, born in 2001. Master candidate. Student member of CCF. His main research interests include machine learning and pattern recognition.

李超凡, 2001 年生。硕士研究生。CCF 学生会员。主要研究方向为机器学习、模式识别。



Chen Songcan, born in 1962. PhD, professor, PhD supervisor. Senior member of CCF. His main research interests include pattern recognition, machine learning, and neural computing.

陈松灿, 1962 年生。博士, 教授, 博士生导师。CCF 高级会员。主要研究方向为模式识别、机器学习、神经计算。