

基于时空多级交互编码的自适应开放词汇脑电解码

常丽涛, 王永雄, 黄 帅, 王 哲

(上海理工大学光电信息与计算机工程学院, 上海 200093)

摘 要: 在开放词汇场景下进行脑电信号到文本的解码任务中, 从时空关系复杂且具有显著个体差异的脑电信号中提取准确的语义信息是一项关键挑战。针对这一挑战, 本文提出了一种基于时空特征增强、受试者自适应机制和多级交互注意力机制的解码方法。首先, 结合空洞卷积与门控机制, 设计时空特征增强模块, 以提升时序动态建模能力和通道依赖性表征能力。随后, 引入受试者特定令牌, 对不同受试者间的脑电分布差异进行显式建模, 缓解跨受试条件下个体差异对模型解码性能的影响。最后, 提出多级交互注意力机制, 联合通道级、受试者级和全局级注意力, 解析脑电信号中的多层次语义关联。此外, 首次引入基于可学习激活函数的非线性投影网络, 高效表征高维脑电信号, 增强模型对脑电信号的解码能力。实验结果表明, 本文所提出的方法在词汇相似度指标 BLEU 和 ROUGE 以及语义相似度指标 BERTScore 上分别提升 2.23、4.87 和 4.66 个百分点, 验证了该方法在开放词汇脑电解码任务中的有效性与鲁棒性。

关键词: 脑电信号; 脑电图-文本解码; 时空学习; 交互注意力机制

中图分类号: TP391.41; TP183 **文献标志码:** A

引用格式: 常丽涛, 王永雄, 黄帅, 等. 基于时空多级交互编码的自适应开放词汇脑电解码[J]. 数据采集与处理, 2026, 41(4): 1010-1025. CHANG Litao, WANG Yongxiong, HUANG Shuai, et al. A spatio-temporal multi-level adaptive interactive attention network for open-vocabulary EEG decoding[J]. Journal of Data Acquisition and Processing, 2026, 41(4): 1010-1025.

引 言

脑电图 (Electroencephalography, EEG) 信号是指通过在头皮上放置电极采集的大脑电活动信号。因其具备高时间分辨率、非侵入性的特点以及相对较低的成本, 在神经科学研究和临床应用等领域中得到应用^[1-3], 尤其是在脑机接口领域^[4-5]。在脑机接口领域中, 文本解码被认为具有重要意义^[6-7], 它能够帮助失语症患者或运动障碍患者恢复沟通能力, 并为理解大脑语言加工机制提供重要窗口。EEG 信号能够实时反映语言理解与生成过程中多个大脑区域的协同活动, 为基于 EEG 信号的文本解码任务提供了独特的神经生理基础。

最早的 EEG 解码文本模型是通过解码想象打字的动作来间接解码文本。虽然仅是想象而没有实际的打字运动, 个体的大脑仍会激活与实际运动相关的脑区, 在 EEG 信号中反映出特定的模式。基于此, Zhang 等^[8]提出了一种端到端的脑打字系统, 通过结合卷积神经网络与循环神经网络, 从运动想象诱发的 EEG 信号中提取特征, 将其映射为具有语义的短语或句子, 为后续研究提供了结构框架上的经验, 但由于效率低下, 没有获得广泛的应用。

为提高效率,后续研究转向了直接解码阅读文字产生的EEG信号。例如,Pan等^[9]提出了一种基于小波变换和无相关多线性主成分分析的多字符分类方法,并使用分类器对29个字符进行识别,平均分类准确率达到97.59%,在有限词汇集合上取得了较好的性能。然而,相比于有限词汇场景下的研究,开放词汇场景下的文本解码任务具有更高的实际应用价值。在有限词汇场景中,模型本质上进行的是分类任务,仅需学习EEG信号与固定类别标签之间的映射,无需建模语义关联与上下文结构。这类方法虽然能在预定义类别上取得较高准确率,但无法生成训练中未出现的文本内容,难以应对真实语言场景。相比之下,开放词汇场景下的文本解码更接近文本生成任务,模型将原始EEG信号视为语义信息的表征,学习语义信息与上下文依赖关系,因而具备生成同义替换或语义近似表达的能力。这种开放式生成机制更符合失语症患者辅助通信等实际应用场景的需求。开放词汇EEG信号解码的主要挑战在于如何从复杂、多噪声且个体差异显著的EEG信号中准确提取与文本相关的信息^[10-15]。具体而言,首先,EEG信号具有高度复杂的时空特性,这使其对大规模词汇语义表征不稳。其次,词汇之间的语义相似性、多义性以及语境多样性进一步增加了语义判别的难度。此外,EEG信号对语言的抽象表达能力有限,难以有效建模语法结构和长程上下文关系。最后,面向实际应用的实时解码任务对算法计算速度提出了更高的要求。

近期的研究尝试使用预训练的语言模型^[16-18]应对上述挑战。例如Wang等^[10]创新性地提出了先使用Transformer^[19]对EEG信号进行编码,后使用预训练的BART^[13]模型进行解码的框架,首次实现了开放词汇场景下的EEG信号到文本的解码任务,为开放词汇EEG信号到文本的转换建立了基准。利用EEG信号时空关系进行表征学习已被证明有效^[20-25]。Wang等^[10]虽然建立了开放词汇场景下EEG信号到文本的解码任务的框架,但未针对EEG信号复杂时空关系建模,未能充分利用EEG信号中所蕴含的信息。Yang等^[15]提出了一种结合循环神经网络(Recurrent neural network, RNN)与1D卷积神经网络的多尺度网络结构,处理EEG信号的时空关系,然而这种方法不能自适应地加强关键特征,难以同时兼顾局部特征与全局特征,限制了模型的性能。

此外,在使用EEG信号进行解码的相关任务中,另一个难点是受试间的泛化问题^[26-29]。由于每个受试者的大脑活动模式存在显著差异,这使得模型在单一受试上训练和推理所得到的结果难以在其他受试上完美复现。为了解决这一问题,一些方法尝试设置一个独立的处理模块来减少影响。例如,Amrani等^[11]提出使用一个专门的特征矩阵,通过对每个受试者的EEG信号进行个性化编码来应对挑战。但这种方法缺少多层次融合的信息处理,导致在复杂情况下,模型对全局信息捕捉不足,影响整体表现。

综合来看,尽管现有开放词汇下的EEG信号解码文本的方法取得了一定进展,但仍存在若干关键问题。首先,EEG信号具有复杂的时空耦合特性,现有方法对时间动态与空间依赖关系的联合建模能力有限,难以充分挖掘脑电信号中的深层语义信息。其次,多数方法缺乏对跨受试者个体差异的有效建模,在跨受试场景下泛化能力不足。再次,现有研究对不同层级语义信息之间的交互建模不充分,难以同时兼顾局部细粒度特征与全局语义关联。此外,高维EEG信号特征到文本语义空间之间存在复杂的非线性映射关系,而传统线性投影的固定激活函数结构在表达能力方面存在局限。因此,如何在开放词汇场景下实现具有鲁棒性、泛化性与高语义一致性的EEG信号到文本的解码,仍然是当前研究中的重要挑战。

在此背景下,本文提出了一种基于时空多级交互编码的自适应开放词汇脑电解码模型(Spatiotemporal multi-level adaptive interactive attention network for open-vocabulary EEG decoding, ST-MSIAN)。通过多分支卷积结合时空注意力机制,联合建模EEG信号的时间动态与空间依赖关系,提升特征表示的鲁棒性与判别性;同时引入受试者特定令牌并与全局特征融合,实现个体差异与共性语义的协同建

模,增强跨受试场景下的泛化能力。在此基础上,采用多级交互注意力机制,从通道、受试者及全局层面动态融合关键特征,并与带有可学习激活函数的柯尔莫哥洛夫-阿诺德网络(Kolmogorov-Arnold network, KAN)投影头共同构建端到端解码框架,实现从高维EEG信号特征向语言语义空间的精确映射。实验结果表明,该方法在ZuCo(Zurich cognitive language processing corpus)数据集上的词汇相似度指标BLEU、ROUGE和语义相似度指标BERTScore上均显著优于现有方法。

1 本文方法

本文提出了一种多模块融合的端到端解码框架ST-MSIAN。如图1所示,该模型由4个核心模块组成:时空特征增强模块(Spatiotemporal feature enhancement module, SFEM)、多级交互注意力机制(Multi-level interactive attention mechanism, MIAM)、Transformer编码器以及KAN。其中 F 为各模块及其分支输出的特征表示, Q 、 K 、 V 分别为对应注意力模块中查询、键、值矩阵,由MIAM模块的输出经不同的线性变换得到。首先,SFEM从原始EEG信号中提取多尺度的时间与空间特征,并通过门控机制融合为统一表示。随后,引入受试者特定令牌与共享令牌,在MIAM中通过通道级、受试者级以及全局级注意力机制,深度融合个体特征与共性表示。Transformer编码器对融合后的EEG信号特征进行序列建模,增强其上下文依赖与语义表达能力。为提升模型建模高维数据的能力,KAN对编码结果进一步重构完成特征提取。最后由BART完成文本的生成任务。该框架在保持语义一致性的同时,显著提升了模型在开放词汇场景下使用EEG信号进行文本解码任务的精度与泛化能力。

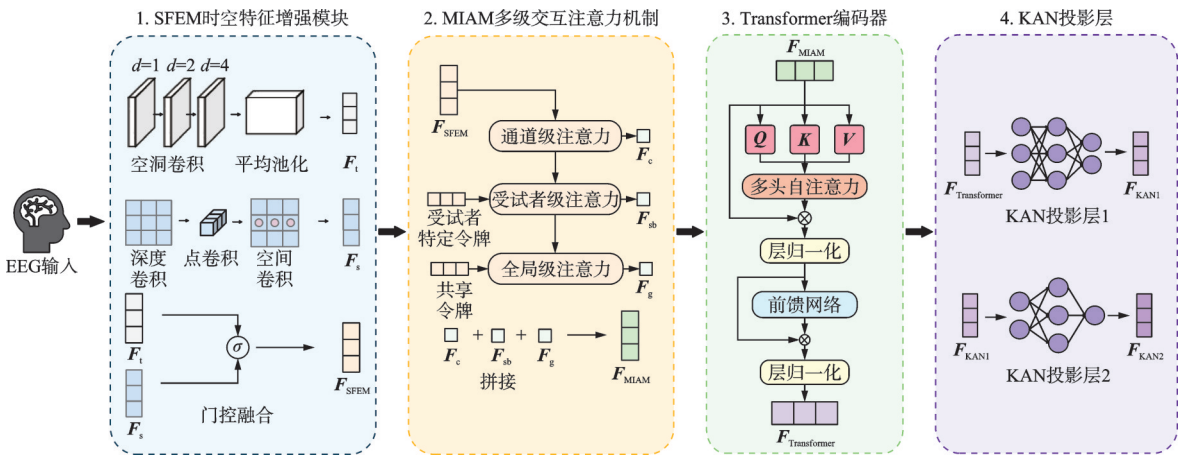


图1 ST-MSIAN模型的主要模块

Fig.1 Main modules of the ST-MSIAN model

1.1 时空特征增强模块

在EEG信号解码任务中,原始EEG信号中同时包含复杂的时间动态和通道间空间依赖。为有效提取和融合这些底层特征,本文在编码器初始阶段引入了SFEM,专注于从单词级EEG信号片段中构建高质量的时空表征,为后续语言生成奠定基础。

SFEM由时间路径与空间路径两条并行子分支构成,如图2所示。其中,时间路径利用多尺度空洞卷积对EEG信号序列进行逐层建模,即

$$F_t^{(i)} = \text{ReLU}(W_t^{(i)} * X + b_t^{(i)}) \quad (1)$$

$$F_t' = F_t^{(1)} + F_t^{(2)} + F_t^{(4)} \quad (2)$$

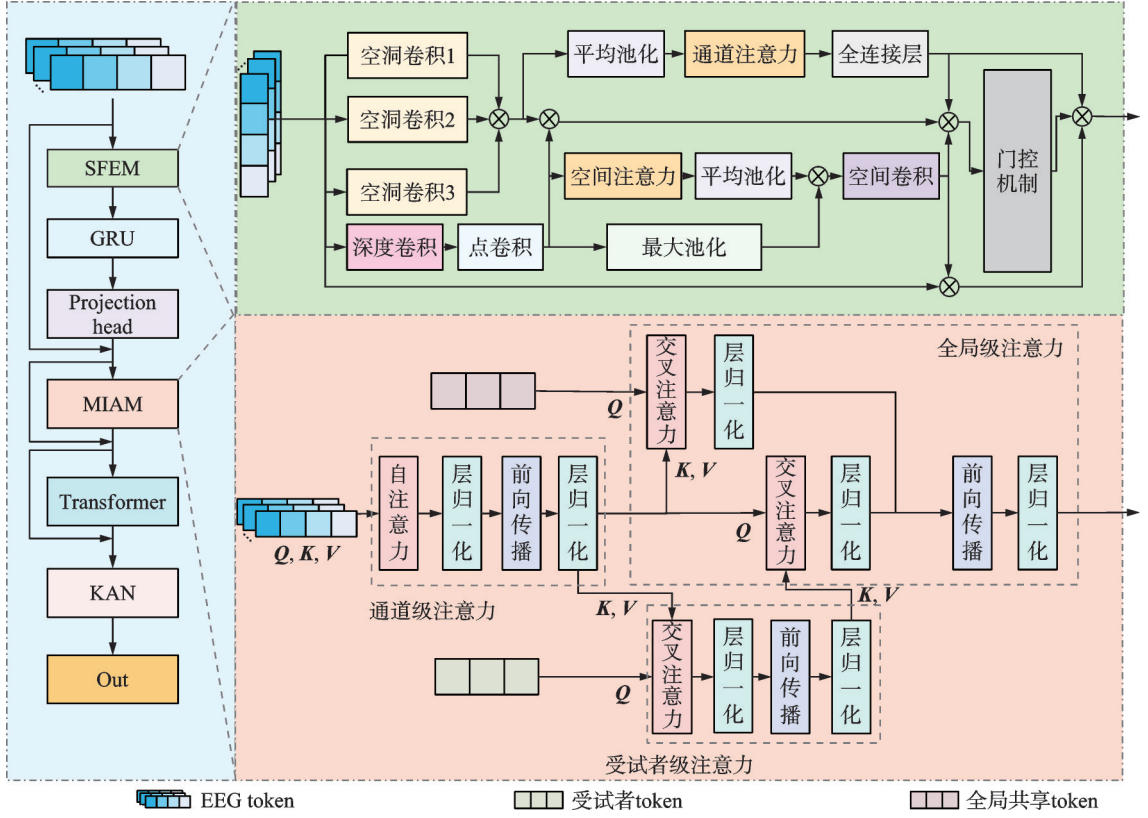


图2 SFEM与MIAM模块实现细节

Fig.2 Implementation details of SFEM and MIAM module

式中: $F_t^{(i)}$ 代表第 i 个膨胀卷积的输出, $W_t^{(i)}$ 和 $b_t^{(i)}$ 分别表示可学习参数, 符号 $*$ 表示膨胀卷积操作, X 为模型输入。为同时捕捉局部短时依赖与跨时长依赖, 本文采用扩张率分别为 1、2 和 4 的膨胀卷积。扩张率 1 对应局部精细特征, 2 和 4 则逐步扩展感受野, 捕获中长时间序动态特征, 实现多尺度时序特征建模。

EEG 信号对应的全局时间表示通过平均池化提取, 即

$$P_t = \text{AvgPool}(F_t') \quad (3)$$

随后, 通道注意力模块以门控方式对时间表示进行重标定, 即

$$S_t = \sigma(W_t^{(2)} \text{ReLU}(W_t^{(1)} P_t)) \quad (4)$$

$$F_t = F_t' \odot S_t \quad (5)$$

式中: σ 代表 Sigmoid 函数, $W_t^{(1)}$ 和 $W_t^{(2)}$ 为可学习参数, “ $\odot$ ” 表示逐元素相乘。

空间路径从深度卷积开始逐通道提取局部特征, 有

$$D_s = \text{DepthwiseConv}(X) \quad (6)$$

采用点卷积, 用于跨通道混合信息, 即

$$P_s = \text{PointwiseConv}(D_s) \quad (7)$$

随后使用通道注意机制关注最关键的空间位置信息, 即

$$A_s = \sigma(f_7[\text{AvgPool}(P_s); \text{MaxPool}(P_s)]) \quad (8)$$

$$F_s = P_s \odot A_s \quad (9)$$

式中 f_7 代表 7×7 的卷积操作。最终, 两个分支的输出通过可学习门控机制融合为统一表征, 有

$$H_g = \text{ReLU}(W_{g1}[F_t; F_s] + b_{g1}) \quad (10)$$

$$G = \sigma(W_{g2}H_g + b_{g2}) \quad (11)$$

$$F_{\text{SFEM}} = G \odot F_t + (1 - G) \odot F_s \quad (12)$$

式中: W_g 和 b_g 为可学习参数, $[\cdot]$ 表示连接操作。该模块通过对多尺度时间依赖与通道空间结构的联合建模, 为解码任务提供了稳健且可泛化的底层表示能力。

SFEM 作为整个模型的信息入口, 在每个词片段的表示计算中被独立调用, 并与后续的门控循环编码器 (Gate recurrent unit, GRU) 编码器共同组成 EEG 信号编码的初始阶段。表 1 展示了 SFEM 具体参数。

表 1 SFEM 参数

Table 1 SFEM parameters

网络层	卷积核/膨胀率	激活函数
空洞卷积 1	$3 \times 3/1$	ReLU
空洞卷积 2	$3 \times 3/2$	ReLU
空洞卷积 3	$3 \times 3/4$	ReLU
深度卷积	1×1	—
点卷积	1×1	—
空间卷积	7×7	Sigmoid

1.2 多级交互注意力机制

MIAM 旨在将 SFEM 生成的时空特征进一步投影至语义表征空间。不同于传统的并行注意力机制, MIAM 采用渐进式协同建模策略, 通过通道层、受试者层和全局层的层级化交互, 逐步从细粒度电极协同模式提炼至跨受试者的全局级统一语义表征。MIAM 的输入来自 3 个层面, 分别是: 经 SFEM 模块所提取的时空特征、用于显式建模个体差异的受试者特定令牌, 以及用于全局信息融合的可学习共享令牌。

(1) 通道级注意力: 作为协同机制的基础层, 通道层注意力关注不同通道之间 EEG 信号的空间协同模式。它利用自注意力处理序列标记, 捕获跨电极的细粒度交互关系, 即

$$A_c = \text{Softmax}\left(\frac{Q_c K_c^T}{\sqrt{d_k}}\right) V_c \quad (13)$$

式中: Q_c 、 K_c 和 V_c 均来自 SFEM 输出的 EEG 信号时空表征, 经不同的线性变换得到; d_k 为键的维度, 取 64。输出 A_c 随后经过层归一化和前馈网络处理, 即

$$F_c' = F_{\text{SFEM}} + A_c \quad (14)$$

$$F_c = F_c' + \text{FFN}(\text{LayerNorm}(F_c')) \quad (15)$$

该层级建立了跨电极通道的基础协同关系, 为后续的个体化建模提供了空间结构化的特征表示。

(2) 受试者级注意力: 基于通道级注意力捕获的空间协同模式, 受试者级引入可学习的受试者特定令牌, 并将其作为查询, 对通道表征进行个体化重加权, 即

$$A_{sb} = \text{Softmax}\left(\frac{Q_{sb} K_{sb}^T}{\sqrt{d_k}}\right) V_{sb} \quad (16)$$

式中: Q_{sb} 为引入可学习的受试者特定令牌, 每位受试者均有一个独立的编码; K_{sb} 和 V_{sb} 来自通道层注意力的输出 F_c , 这使得受试者令牌能够在已建立的通道协同基础上, 进一步学习不同受试者的不同通道组合方式。受试者令牌初始化为零向量, 在训练中通过梯度下降更新, 自适应地捕捉不同受试者的独特模式。输出经过处理得到, 即

$$F_{sb}' = Q_{sb} + A_{sb} \quad (17)$$

$$F_{sb} = F_{sb}' + \text{FFN}(\text{LayerNorm}(F_{sb}')) \quad (18)$$

(3) 全局级注意力: 全局级注意力负责整合来自通道和受试者两个层次的信息, 即

$$A_g = \text{Softmax} \left(\frac{Q_g K_g^T}{\sqrt{d_k}} \right) V_g \quad (19)$$

式中: Q_g 为全局共享语义向量, 通过梯度下降更新, 反应所有受试者共有的语义结构, 而 K_g 和 V_g 为通道层和受试者层注意输出的拼接。最终输出为

$$F_g' = Q_g + A_g \quad (20)$$

$$F_g = F_g' + \text{FFN}(\text{LayerNorm}(F_g')) \quad (21)$$

这些层次之间的交互至关重要。每个层次的输出作为下一层输入, 形成动态且具有上下文感知的特征提取过程。

MIAM 的最终输出可以表示为

$$F_{\text{MIAM}} = \text{Concat}(F_c, F_{\text{sb}}, F_g) \quad (22)$$

通过这种方法, 3个层级的注意力输出经残差融合后形成稳健的脑表征, 不仅使模型在跨受试泛化文本生成任务中具备稳健且可泛化的底层表示能力, 也为下游 Transformer 编码器提供了跨受试者可迁移的高层语义先验。MIAM 模块中所有注意力机制均采用单头注意力。

1.3 Transformer 编码器

在 MIAM 模块输出稳健脑表征后, 模型引入一组多层 Transformer 编码器以完成高层语义建模。具体而言, 系统首先为长度固定的 EEG 信号序列添加可学习的位置嵌入, 显式编码时序关系。随后, 将序列送入 6 层 Transformer 编码器, 每层均包含多头自注意力、前馈网络以及残差和层归一化结构。Transformer 编码器采用标准结构, 每层包括头数为 8, 维度为 832 的多头自注意力和隐藏层维度为 2 048, 丢弃率为 0.1 前馈网络。嵌入维度 832 以匹配 EEG 信号通道数。最终, Transformer 输出的序列特征经归一化后传递给 KAN 层。

1.4 KAN 非线性投影模块

为了增强非线性建模能力, 提升高维度 EEG 信号特征到文本嵌入空间的映射效果, 本文采用了近期提出的 KAN 替代传统 MLP 作为投影头。KAN 通过可学习的激活函数, 将线性变换与基于核的样条插值相结合, 能够自适应地逼近 EEG 信号中复杂的非线性模式且不会显著增加模型参数。具体实现中, 样条阶数为 3, 网格大小为 5, 启用独立的样条缩放, 基础激活函数为 SiLU, 网格范围为 $-1 \sim 1$, 网格间距为 0.02。

在本文框架中, KAN 模块由两层带可学习的激活函数的非线性投影层堆叠组成, 输入维度为 832, 隐藏维度与投影维度均为 1 024, 丢弃率为 0.2, 并通过残差连接投影至预训练 BART 编码器的嵌入空间, 接收 Transformer 编码的 EEG 信号特征。训练过程中, 通过最小化 KAN 输出与 BART 文本嵌入之间的均方误差实现监督学习。虽然 KAN 并非本文提出, 但其作为一个表达能力强、可微分的模块, 有效提升了整个端到端 EEG 到文本解码过程中的泛化能力与语义对齐效果。

1.5 预训练语言解码器

在完成神经语义对齐后, 模型将 KAN 投影输出传递至预训练的 BART 解码器架构中, 进一步完成文本序列生成。该模块作为端到端 EEG 解码流程的语言生成后端, 充分利用了 BART 在语言建模和上下文理解方面的强大能力。训练过程中, BART 解码器以 EEG 信号的编码嵌入为输入, 并通过最小化预测文本与真实文本之间的交叉熵损失实现优化。该模块不仅增强了解码文本的流畅性与语义合理性, 也使得整个模型具备更强的泛化能力与实际应用潜力。

ST-MSIAN 各模块的参数量分布情况见表 2。模型总参数量为 47.1×10^6 , 其中 Transformer 编码器参数量最大, 为 37.8×10^6 , 主要承担全局特征建模任务。相比之下, 本文提出的 SFEM、MIAM 和 KAN 模块参数量相对较小, 分别为 0.51×10^6 、 0.28×10^6 和 0.33×10^6 , 表明所提出模块在整体模型中的参数开销较低。整体来看, ST-MSIAN 在引入时空特征增强、多级交互注意力以及非线性映射机制的同时, 整体模型规模仍保持在较为合理的范围内。

表 2 ST-MSIAN 各模块参数量分布
Table 2 Parameter distribution of different modules in ST-MSIAN

模块	参数量/ 10^6
SFEM	0.51
MIAM	0.28
Transformer	37.8
KAN	0.33
其他	8.09
总计	47.1

2 实 验

2.1 实验设置

在本文的实验中, 所有模型均基于 Pytorch 框架实现, 并在 NVIDIA GeForce RTX 4060Ti GPU 上运行。训练过程中采用交叉熵损失函数作为优化目标, 利用学习率为 $5e-5$ 的 Adam 优化器更新网络参数。批处理大小设置为 4, 并结合动态学习率调度策略调整学习率, 以保证模型的稳定收敛和性能提升。两阶段训练轮数均为 10。

2.2 数据集

本文使用苏黎世认知语言处理语料集 ZuCo 进行实验。该数据集是一个公开的多模态 EEG 信号语言处理数据集, 由瑞士苏黎世大学发布。它主要面向自然语言理解、脑机接口 (Brain computer interface, BCI) 和神经语言建模等方向的研究。所使用的文本材料来自影评语料与维基百科条目。阅读实验包含两类任务: 自然阅读和任务驱动阅读, 同时采集了 EEG 信号与眼动信号, 本实验仅使用其中的 EEG 信号。数据集包括两个版本: ZuCo v1.0^[30] 与 ZuCo v2.0^[31], 分别覆盖 12 名和 18 名受试, 本实验中采用了两个版本中的全部数据。

在数据采集方面, 该数据集使用 128 通道 Geodesic Hydrocel 系统进行高密度 EEG 信号记录, 采样率为 500 Hz, 频率范围设置为 0.1~100 Hz, 并以 Cz 电极作为参考点。原始 EEG 信号数据经过清洗和筛选后, 最终得到共 105 个可用通道。

数据集划分情况如表 3 所示, 数据集按照句子划分, 包含 3 种阅读任务, 对应的语料来源不同: Task 1 来自电影评论中的情感句子; Task 2、Task 3 和 Task 2.2.0 均为维基百科中的普通句子, Task 2 为普通句子, Task 3 要求被试在阅读时聚焦并判断特定语义关系是否出现, Task 2.2.0 要求参与者主动搜索句中的特定语义关系类型。样本被随机分为训练集 (80%)、验证集 (10%) 与测试集 (10%)。测试集中所含句子均未出现在训练或验证过程中, 从而保证模型评估的独立性。

表 3 数据集划分情况

Table 3 Dataset division

任务	句子数量	训练集	验证集	测试集
Task 1	400	3 609	467	456
Task 2	300	2 465	343	350
Task 3	407	4 456	601	522
Task 2.2.0	390	3 372	350	493

2.3 训练策略

传统的机器学习模型往往采用端到端训练策略, 而本文采用了两阶段训练策略, 这是由 EEG 信号的特征所决定。相比端到端联合微调策略, 两阶段训练策略能够有效缓解 EEG 信号噪声大、语义表达

弱以及样本规模有限所带来的训练不稳定问题。端到端训练虽然能够通过全局联合优化实现更强的模态协同能力,但在文本解码任务中,预训练语言模型参数规模远大于EEG编码模块,容易导致梯度失衡与上下文关系的丢失,使模型难以从EEG信号中学习到稳定的语义表示。此外,原始EEG特征与自然语言嵌入空间差异较大,若直接进行文本生成监督,解码器容易受到随机噪声干扰,从而影响整体训练收敛性与生成质量。

在第1阶段的训练中,冻结BART的编码器,并训练ST-MSIAN的输出与之对齐。为了提高模型的收敛速度,采用了循环学习率调度策略。该策略通过在指定的学习率范围内周期性地调整学习率,使得模型能够在不同的学习率下探索不同的参数空间,从而加速收敛。具体而言,设置了一个较小的初始学习率,并采用了Triangular2模式的循环策略,这有助于模型在训练初期以较小的学习率进行精细调整,而在后期逐步增大学习率,以避免陷入局部最优解。在这一阶段,损失函数采用了均方误差损失,该损失函数在多数回归任务中具有较好的表现,适用于本文中的预测任务。经过若干轮训练后,选择验证集上损失最小的模型作为第1阶段的最优模型,并保存该模型用于后续的微调过程。

第2阶段的训练着重于微调BART模型,使其能够更好地适应具体任务的需求。在此阶段,冻结已训练好的ST-MSIAN,微调BART。微调过程使用了较低的学习率,并采用了与第1阶段相同的循环学习率调度策略,为确保模型在微调时能够稳定收敛,本文将最大学习率稍微降低。第2阶段的损失函数采用了交叉熵损失函数,该损失函数在分类任务中表现出色,适合处理多分类问题。通过这种策略,可以有效地调整模型的分类性能,提高预测准确性。

2.4 对比实验

将本文架构与EEG-To-Text^[10]、DeWave^[29]、Amrani^[11]、SEE^[12]和Murad^[18]的当前最先进模型进行比较。选取BLEU、ROUGE与BERTScore^[23]作为综合评价指标,结果如表4所示。BLEU指标中的 $N=1,3$ 指的是 N -gram的阶数,即计算精确匹配时考虑的连续词组长度;ROUGE和BERTScore中的 R 代表召回率,衡量参考文本中的内容有多少被候选文本覆盖到了, P 代表精确率,衡量候选文本中的内容有多少是与参考文本相符的, F 为 R 和 P 的调和平均数,综合平衡两者。

表4 不同模型在ZuCo数据集上的性能
Table 4 Performance of different models on the ZuCo dataset

方 法	BLEU/%		ROUGE/%			BERTScore/%		
	$N=1$	$N=3$	R	P	F	R	P	F
EEG-To-Text ^[10]	40.10	12.50	28.80	31.7	30.18	52.71	48.84	50.70
DeWave ^[29]	41.35	13.92	28.82	33.71	31.07	52.79	48.98	50.81
Amrani ^[11]	41.87	14.53	30.40	35.50	32.75	56.30	54.58	55.42
SEE ^[12]	39.87	12.18	28.23	30.51	29.33	51.32	47.67	49.42
Murad ^[18]	40.13	12.71	26.29	31.36	28.60	53.25	49.36	51.23
本文(ST-MSIAN)	44.10±	16.21±	35.10±	40.54±	37.62±	61.13±	59.06±	60.08±
	0.42	0.41	0.42	0.43	0.32	0.29	0.43	0.38

BLEU指标以 N -gram精确匹配为核心,能够有效衡量候选文本在词汇层面与参考文本的一致程度,是机器翻译等任务中应用最广泛的评价标准。ROUGE指标更强调召回率,特别适合用于摘要类任务,能够从词组覆盖度和句子结构顺序上反映生成文本对参考内容的保留情况。BERTScore基于预训练语言模型的语义表示,能够捕捉同义替换和语义相近的表达,因此在衡量语义相似度层面更符合人

类直觉。三者结合,可以分别从词汇、文本覆盖度和实际语义3个角度对生成结果进行全面评价,避免单一指标可能带来的片面性。

本文方案在词汇相似度指标 BLEU 得分为 44.10%、ROUGE 为 37.62%、语义相似度指标 BERTScore 为 60.08%,在各个方面均优于最先进水平。BLEU 被作为主要评价指标用于统计显著性检验。基于测试样本级 BLEU 分数,与 Amrani 等^[11]的方法进行配对 t 检验,结果表明本文方法在统计意义上具有显著优势(配对 t 检验的显著性水平 $p=0.038<0.05$),其余指标亦表现出一致的性能提升趋势。相比于基线 EEG-To-Text 模型^[10],ST-MSIAN 在每项指标上都表现出更高的性能,证明了其学习时空特征以及采用受试者自适应、多级交互注意力机制和 KAN 的积极影响。图 3 展示了不同模型在 BLEU、ROUGE 和 BERTScore 上的箱型图。通过箱型图能够直观地观察到 ST-MSIAN 模型在文本重构和语义理解方面的优异表现。

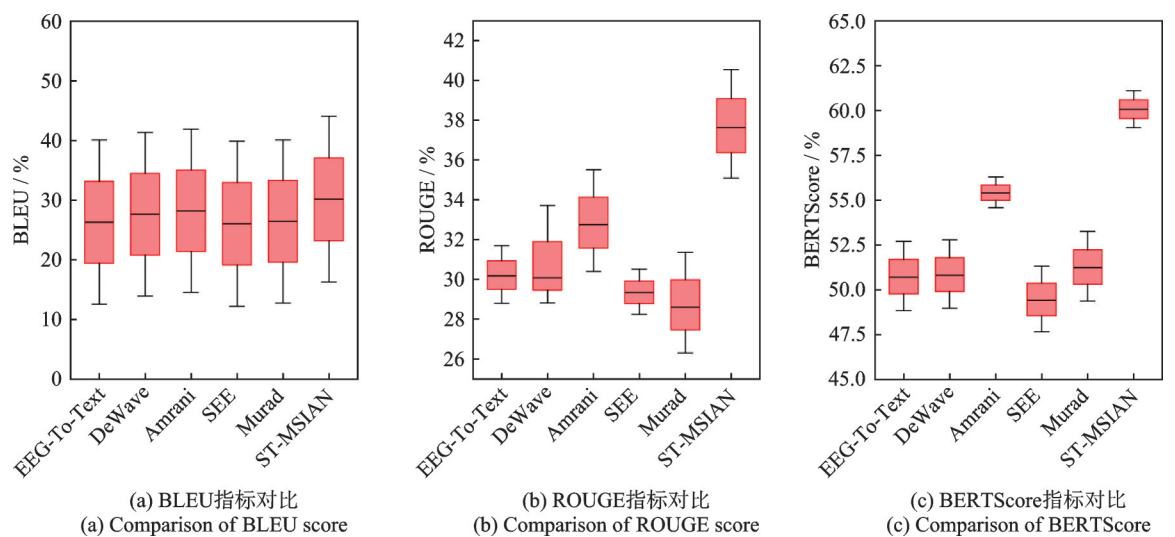


图3 不同模型 BLEU、ROUGE 和 BERTScore 对比

Fig.3 Comparison of BLEU, ROUGE and BERTScore for different models

为评估 ST-MSIAN 的计算效率,表 5 对各方法的参数量、浮点运算次数 FLOPs 及单词级词汇相似度指标 BLEU-1 性能进行了综合对比,所有模型实验条件设置为与本文实验相同。其中,由于 SEE 代码未公开,故无法展示其计算效率。结果表明,在参数量方面,ST-MSIAN 为 47.10×10^6 ,低于 Amrani (85.27×10^6)^[11] 和 DeWave (51.97×10^6)^[29],与 Murad (49.85×10^6)^[18] 和 EEG-To-Text (40.50×10^6)^[10]

表5 不同模型复杂度对比

Table 5 Complexity comparison of different models

方 法	参数量/ 10^6	FLOPs/ 10^9	BLEU-1/%
EEG-To-Text ^[10]	40.50	4.17	40.10
DeWave ^[29]	51.97	5.34	41.35
Amrani ^[11]	85.27	9.13	41.87
SEE ^[12]	—	—	39.87
Murad ^[18]	49.85	5.21	40.13
本文(ST-MSIAN)	47.10	4.89	44.10

处于同一量级。在计算量方面,ST-MSIAN的FLOPs为 4.89×10^9 ,显著低于Amrani(9.13×10^9)^[11],与EEG-To-Text(4.17×10^9)和Murad(5.21×10^9)^[18]相近。然而,在BLEU-1指标上,ST-MSIAN以44.10%的得分显著优于所有对比方法,较第二名Amrani(41.87%)提升了2.23个百分点。上述结果表明,ST-MSIAN在参数量和计算复杂度均处于合理范围的前提下,实现了最优的解码性能,验证了本文模型的有效性与高效性。

表6~8展示本文模型与其他模型的实际预测效果,黑体表示高度相似或部分匹配的词组和短语,体现了模型对关键语义信息的准确捕捉能力。可以看出,Wang等^[10]和Amrani等^[11]的模型虽然在一定程度上可以输出正确的单词,但是无法形成具有实际意义的句子。相比之下本文模型无论是语义还是通顺程度均高于两者。

表6 本文模型预测效果

Table 6 Prediction effectiveness by the proposed model

真实句子	预测句子
(1) He was a member of the Union Railroad, and served in the New York State Legislature.	(1) was a member in the United Railroad, and was served for New York State legislature.
(2) was a United States Senator from Nevada and a member of the Democratic Party.	(2) was the United States Senate from Nevada and a member of the Democratic Party.
(3) Following the 1980 presidential election, Bush and his family moved to Miami-Dade County, Florida.	(3) the 1980 presidential election, Bush and his wife moved to Dade County, Florida.

表7 Wang模型预测效果

Table 7 Prediction effectiveness by Wang's model

真实句子	预测句子
(1) He was a member of the Union Railroad, and served in the New York State Legislature.	(1) was a member of the union rail, and served in the New York.
(2) was a United States Senator from Nevada and a member of the Democratic Party.	(2) was a United States senator from Nevada, the democratic.
(3) Following the 1980 presidential election, Bush and his family moved to Miami-Dade County, Florida.	(3) the 13,41 election, But his family, , ,

表8 Amrani模型预测效果

Table 8 Prediction effectiveness by Amrani's model

真实句子	预测句子
(1) He was a member of the Union Railroad, and served in the New York State Legislature.	(1) was member of union, railroad, New York state.
(2) was a United States Senator from Nevada and a member of the Democratic Party.	(2) was United States senator, Nevada, democratic party.
(3) Following the 1980 presidential election, Bush and his family moved to Miami-Dade County, Florida.	(3) from 1980 election, But his family moved to main Florida.

2.5 消融实验

为了全面评估ST-MSIAN模型中各个组件和设计选择的贡献,本文对模型组件开展了消融实验,

旨在理解不同架构决策对模型性能的影响,所有消融实验均进行5次独立运行,报告平均值与标准差。具体实验结果如表9所示。为验证模型的跨受试泛化能力,本文进行了留一受试者交叉验证实验。在该设置下,每次选择一名受试者的数据作为测试集,其余受试者的数据用于训练,遍历所有受试者后取平均结果。从表9可以看出,ST-MSIAN在跨受试场景下的BLEU-1、ROUGE-*F*和BERTScore-*F*分别达到42.02%、35.06%和56.17%。相比完整模型性能有所下降,这是由于跨受试验证中测试受试者的数据在训练阶段完全未见,对模型泛化能力提出了更严苛的要求。然而,跨受试结果仍显著优于基线方法,充分证明了受试者自适应机制和多级交互注意力机制在处理个体差异方面的有效性。这一结果表明,ST-MSIAN不仅在混合受试训练场景下表现优异,在面对完全未见受试者时也具备良好的泛化能力,为实际应用提供了可能。

表9 消融实验结果对比

Table 9 Result comparison of ablation experiments

模块	BLEU/%		ROUGE/%			BERTScore/%		
	<i>N</i> =1	<i>N</i> =3	<i>R</i>	<i>P</i>	<i>F</i>	<i>R</i>	<i>P</i>	<i>F</i>
w/o SFEM	38.64±	10.72±	28.43±	33.64±	30.82±	55.38±	51.47±	53.35±
	0.81	0.91	0.94	0.87	0.71	0.55	0.76	0.68
w/o 时序分支	39.31±	11.34±	29.83±	35.37±	32.36±	56.64±	52.44±	54.46±
	0.88	0.79	0.82	0.81	0.70	0.53	0.73	0.69
w/o 空间分支	39.63±	11.72±	31.25±	36.23±	33.55±	57.01±	54.30±	55.62±
	0.85	0.74	0.80	0.81	0.69	0.55	0.71	0.70
w/o MIAM	39.70±	12.91±	29.10±	33.74±	31.25±	55.42±	52.09±	53.70±
	0.91	0.81	0.83	0.79	0.70	0.41	0.68	0.63
w/o 通道级注意力	39.60±	11.61±	31.01±	36.13±	33.37±	56.98±	54.12±	55.51±
	0.71	0.51	0.79	0.62	0.77	0.44	0.66	0.61
w/o 受试者级注意力	39.25±	10.29±	30.53±	34.84±	32.54±	56.38±	52.26±	54.24±
	0.67	0.46	0.73	0.61	0.75	0.47	0.69	0.61
w/o 全局级注意力	39.72±	11.62±	31.37±	36.90±	33.91±	57.21±	55.35±	56.26±
	0.72	0.51	0.76	0.66	0.68	0.42	0.61	0.58
w/o KAN	39.57±	11.76±	31.19±	35.27±	33.10±	56.93±	54.57±	55.73±
	0.52	0.41	0.45	0.40	0.48	0.32	0.43	0.39
跨受试	42.02±	14.91±	32.78±	37.68±	35.06±	57.12±	55.25±	56.17±
	0.43	0.41	0.40	0.43	0.33	0.30	0.44	0.40
本文(ST-MSIAN)	44.10±	16.21±	35.10±	40.54±	37.62±	61.13±	59.06±	60.08±
	0.42	0.41	0.42	0.43	0.32	0.29	0.43	0.38

注:w/o表示 without。

在SFEM中,时间和空间两个分支都非常重要。当移除时间分支时,模型性能略有下降,突显了在语言处理任务中捕捉EEG信号时间动态的重要性。这一发现强调了在本文框架中同时考虑EEG数据的空间和时间两个方面非常必要。为验证SFEM时序分支中1、2、4膨胀率配置的合理性,本文还系统评估了时间分支中不同膨胀率组合方案。如表10所示,单一膨胀率配置的BLEU-1得分均低于40.5%,相比本文配置的44.10%显著下降,证明单尺度特征提取的局限性。双尺度组合中,1、4配置获

表 10 SFEM 不同膨胀率对模型性能的影响

Table 10 Impact of different expansion rates of SFEM on model performance

尺度	膨胀率	BLEU-1/%	ROUGE-F/%	BERTScore-F/%
单尺度	$d = 1$	40.23 ± 0.65	33.87 ± 0.54	56.42 ± 0.51
	$d = 2$	39.45 ± 0.71	32.56 ± 0.59	55.78 ± 0.55
	$d = 4$	39.87 ± 0.68	33.21 ± 0.57	56.01 ± 0.53
双尺度	$d = 1, 2$	42.34 ± 0.56	35.78 ± 0.49	58.23 ± 0.47
	$d = 1, 4$	42.67 ± 0.54	36.12 ± 0.47	58.67 ± 0.45
	$d = 2, 4$	41.89 ± 0.59	35.34 ± 0.51	57.89 ± 0.49
三尺度	$d = 1, 2, 4$	44.10 ± 0.42	37.62 ± 0.32	60.08 ± 0.38
	$d = 1, 3, 5$	43.76 ± 0.48	37.68 ± 0.39	59.87 ± 0.42

得 42.67% 的最佳得分,但仍明显低于三尺度配置。本文采用的 1、2、4 配置实现了局部、中等和长距离时序依赖的全面覆盖,在所有测试组合中取得最优性能。相比 1、3、5 的三尺度变体本文配置提升 0.34 个百分点。

关于 MIAM 组件,每个注意力层级都对模型性能有显著贡献。移除任何一个注意力层级都会导致性能明显下降,其中移除受试者级注意力的影响最为显著,与跨受试实验结果一致。这一观察强调了捕捉个体特异性模式在实现准确的 EEG 信号到文本解码过程中的关键作用。MIAM 的协同效果源于 3 个层面的精心设计。首先,在前向信息流层面,受试者层的键值对直接来自通道层输出 F_c 而非原始输入,这确保了个体化建模是在已建立的空间协同模式基础上进行的,避免了受试者令牌直接面对原始高维 EEG 信号,降低了个体化建模的难度。其次,在信息融合层面,采用对 3 个特征进行拼接而非加权求和的并行保留策略,避免了高层抽象覆盖低层细节的问题,使得下游 Transformer 能够同时访问细粒度通道信息 F_c 、个体特异性信息 F_s 和全局共性信息 F_g ,消融实验也验证了多层级交互作用保留的必要性。最后,在参数优化层面,3 层注意力通过反向传播实现联合优化,来自 BART 的语义监督信号首先影响直接对接语义空间的全局层,然后通过梯度反向传播调整受试者层和通道层,这种自上向下的优化使得通道层学会提取有利于个体化的空间模式,受试者层学会识别基于空间模式的个体差异,全局层学会整合两者形成跨受试的统一语义表征。通过这种多维协同设计,MIAM 实现了从局部到整体、从个体到群体的渐进式语义提炼。同时,用多层感知机替代 KAN 导致性能略有下降,但两者参数相差无几,相比 MLP 的固定激活形式,KAN 通过可学习样条函数更适合逼近 EEG 信号中复杂、非平稳的非线性映射关系,尤其在跨受试者场景下表现更稳定。这一结果支持在 ST-MSIAN 中使用 KAN 表示高维的 EEG 数据。

为验证 ST-MSIAN 的跨数据集泛化能力,本文在 ZuCo v1.0 与 v2.0 之间设计了跨版本泛化实验。ZuCo v1.0 与 v2.0 由不同受试者群体参与采集,v1.0 共 12 名受试者,v2.0 共 18 名受试者,两者无重叠,可视为两个独立的数据集。实验分别以单一版本数据进行训练,并在另一版本上进行测试。如表 11 所示,以 ZuCo v1.0 训练、v2.0 测试时,BLEU-1、ROUGE-F 和 BERTScore-F 分别达到 35.39%、28.01% 和 47.28%;以 ZuCo v2.0 训练、v1.0 测试时,3 项指标分别为 36.71%、29.84% 和 48.77%。与主实验合并训练相比,跨版本设置下性能有所下降,这符合训练数据减少且存在分布差异的预期。然而,模型在两种跨版本设置下均保持了较为稳定的性能水平,表明 ST-MSIAN 具备一定的跨受试者群体泛化能力,未过度依赖特定版本的受试者分布。

表 11 跨版本数据集泛化实验

Table 11 Generalization experiments of cross-version dataset

训练集	BLEU-1/%	ROUGE-F/%	BERTScore-F/%
ZuCo v1.0	35.39	28.01	47.28
ZuCo v2.0	36.71	29.84	48.77

总体而言,上述结果展示了ST-MSIAN模型中每个组件的重要性。SFEM、MIAM、受试者自适应机制和KAN都对模型性能起到了关键作用。在确认本文框架有效性的同时,这项全面的消融研究也揭示了未来改进的潜在方向。例如,虽然MIAM中的所有注意力层都发挥了一定的作用,但它们之间的交互可能还有优化空间,以及受限于当前开放词汇EEG解码领域可用数据集稀缺,跨数据集验证暂局限于ZuCo两个版本,未来工作将进一步在更多数据集上验证模型的泛化性。

3 结束语

本文提出了一种创新性的开放词汇场景下使用EEG信号解码文本的框架ST-MSIAN。本文模型中的SFEM、MIAM和受试者自适应机制在开放词汇条件下EEG解码文本任务中展现了卓越的EEG信号处理与语义建模能力,并在KAN投影头的辅助下实现了对预训练语言模型输入空间的高效映射。本文模型不仅显著提升了解码精度,也为大规模开放词汇BCI应用提供了可行的解决方案。实验结果表明,ST-MSIAN在ZuCo数据集上的表现显著优于现有方法,充分证明了所提框架在跨受试者开放词汇EEG信号解码任务中的有效性与鲁棒性。未来工作中,将进一步拓展该框架的应用场景,探索其在更多下游任务中的潜力,并针对EEG信号的长时序依赖与跨模态融合问题进行更深入的研究。ST-MSIAN不仅能够推动开放词汇场景下使用EEG信号解码文本任务的发展,也将在神经康复、辅助交流以及智能脑机接口系统等领域展现出重要应用前景。

参考文献:

[1] DE GANS C J, BURGER P, VAN DEN ENDE E S, et al. Sleep assessment using EEG-based wearables—A systematic review[J]. *Sleep Medicine Reviews*, 2024, 76: 101951.

[2] LIU M, YANG B, MENG L, et al. STA-Net: Spatial-temporal alignment network for hybrid EEG-fNIRS decoding[J]. *Information Fusion*, 2025, 119: 103023.

[3] HUANG S, LUO H, JING H, et al. NEED: Cross-subject and cross-task generalization for video and image reconstruction from EEG signals[C]//*Proceedings of Advances in Neural Information Processing Systems 38 (NeurIPS 2025)*. [S. l.]: NeurIPS, 2025.

[4] 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.

[5] YIN K, LIM E Y, LEE S W. GITGAN: Generative inter-subject transfer for EEG motor imagery analysis[J]. *Pattern Recognition*, 2024, 146: 110015.

[6] ZHOU Y, LI F, LI Y, et al. Enhancing cross-dataset EEG emotion recognition: A novel approach with emotional EEG style transfer network[J]. *IEEE Transactions on Affective Computing*, 2025, 16(3): 2157-2171.

[7] CHANG J, ZHANG Z, QIAN Y, et al. Multi-scale hyperbolic contrastive learning for cross-subject EEG emotion recognition [J]. *IEEE Transactions on Affective Computing*, 2025, 16(3): 1716-1731.

[8] ZHANG X, YAO L, SHENG Q Z, et al. Converting your thoughts to texts: Enabling brain typing via deep feature learning of

- EEG signals[C]//Proceedings of 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom). Athens, Greece: IEEE, 2018: 1-10.
- [9] PAN H, SONG W, LI L, et al. The design and implementation of multi-character classification scheme based on EEG signals of visual imagery[J]. *Cognitive Neurodynamics*, 2024, 18(5): 2299-2309.
- [10] WANG Z, JI H. Open vocabulary electroencephalography-to-text decoding and zero-shot sentiment classification[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2022: 5350-5358.
- [11] AMRANI H, MICUCCI D, NAPOLETANO P. Deep representation learning for open vocabulary electroencephalography-to-text decoding[J]. *IEEE Journal of Biomedical and Health Informatics*, 2026, 30(4): 3593-3600.
- [12] TAO Y, LIANG Y, WANG L, et al. SEE: Semantically aligned EEG-to-text translation[C]//Proceedings of ICASSP 2025—2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Hyderabad, India: IEEE, 2025: 1-5.
- [13] LEWIS M, LIU Y, GOYAL N, et al. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. [S.l.]: Association for Computational Linguistics, 2020: 7871-7880.
- [14] LIU Z M, WANG Y X, VAIDYA S, et al. KAN: Kolmogorov Arnold networks[C]//Proceedings of the International Conference on Learning Representation. [S.l.]: ICLR, 2025.
- [15] YANG J, AWAIS M, HOSSAIN M A, et al. Thoughts of brain EEG signal-to-text conversion using weighted feature fusion-based multiscale dilated adaptive DenseNet with attention mechanism[J]. *Biomedical Signal Processing and Control*, 2023, 86: 105120.
- [16] HUANG S, WANG Y, LUO H. CCSUMSP: A cross-subject Chinese speech decoding framework with unified topology and multi-modal semantic pre-training[J]. *Information Fusion*, 2025, 119: 103022.
- [17] ZHANG R, CHEN Y, XU Z, et al. Recognition of single upper limb motor imagery tasks from EEG using multi-branch fusion convolutional neural network[J]. *Frontiers in Neuroscience*, 2023, 17: 1129049.
- [18] MURAD S A, DAHAL A, RAHIMI N. EEG-to-text translation: A model for deciphering human brain activity[EB/OL]. (2025-05-01). <https://arxiv.org/abs/2505.13936>.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st Annual Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017: 5998-6008.
- [20] GAO C, JIA L, XU J, et al. UMDA-DDSTGN: An unsupervised meta-domain adaptation method using dynamic directed spatial-temporal graph network for EEG-based emotion recognition[J]. *Knowledge-Based Systems*, 2025, 325: 113938.
- [21] WANG H, XU L, YU Y, et al. Global context MambaVision for EEG-based emotion recognition[C]//Proceedings of ICASSP 2025—2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Hyderabad, India: IEEE, 2025: 1-5.
- [22] WANG J, ZHANG C. Cross-modality fusion with EEG and text for enhanced emotion detection in English writing[J]. *Frontiers in Neurorobotics*, 2025, 18: 1529880.
- [23] ZHANG T, KISHORE V, WU F, et al. BERTScore: Evaluating text generation with BERT[C]//Proceedings of the 8th International Conference on Learning Representations. Addis Ababa: ICLR, 2020.
- [24] SHIRODKAR V, REDDY EDLA D, KUMARI A, et al. Event-related desynchronization detection and electroencephalography motor imagery classification using vision transformer[J]. *Intelligent Data Analysis*, 2025, 29(6): 1598-1614.
- [25] OPENAI, ACHIAM J, ADLER S, et al. GPT-4 technical report[EB/OL]. (2023-03-01). <https://arxiv.org/abs/2303.08774>.
- [26] WANG D, WEI Q. SMANet: A model combining SincNet, multi-branch spatial—Temporal CNN, and attention mechanism for motor imagery BCI[J]. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2025, 33: 1497-1508.
- [27] 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.

[28] FORENZO D, ZHU H, SHANAHAN J, et al. Continuous tracking using deep learning-based decoding for noninvasive brain-computer interface[J]. *PNAS Nexus*, 2024, 3(4): 145.

[29] DUAN Y Q, ZHOU J Z, WANG Z, et al. DeWave: Discrete EEG waves encoding for brain dynamics to text translation[C]// *Proceedings of the 37th International Conference on Neural Information Processing Systems*. New Orleans: NeurIPS, 2023: 9907-9918.

[30] HOLLENSTEIN N, ROTSZTEJN J, TROENDLE M, et al. ZuCo, a simultaneous EEG and eye-tracking resource for natural sentence reading[J]. *Scientific Data*, 2018, 5: 180291.

[31] HOLLENSTEIN N, TROENDLE M, ZHANG C, et al. ZuCo 2.0: A dataset of physiological recordings during natural reading and annotation[C]// *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Marseille, France: Association for Computational Linguistics, 2020: 138-146.

作者简介:



常丽涛(2002-),男,硕士研究生,研究方向:多模态脑电信号处理,E-mail: changle714@163.com。



王永雄(1970-),通信作者,男,教授,博士生导师,研究方向:机器学习和机器人智能算法,E-mail: wyxiong@usst.edu.cn。



黄帅(1998-),男,硕士研究生,研究方向:脑电信号处理。



王哲(1994-),男,博士研究生,研究方向:脑电情绪识别。

(编辑:张黄群)

A Spatio-Temporal Multi-level Adaptive Interactive Attention Network for Open-Vocabulary EEG Decoding

CHANG Litao, WANG Yongxiong*, HUANG Shuai, WANG Zhe

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Decoding electroencephalography (EEG) signals into natural language text is a fundamental task in brain-computer interface (BCI) research. Existing approaches often suffer from insufficient modeling of spatiotemporal EEG characteristics, limited capability in capturing hierarchical semantic relationships, and poor generalization across subjects. To address these issues, this study proposes a spatio-temporal multi-level adaptive cross-subject interactive attention network (ST-MSIAN). First, the spatiotemporal feature enhancement module (SFEM) employs parallel temporal and spatial branches to jointly model temporal dynamics and channel dependencies of EEG signals. Multi-scale dilated convolutions with dilation rates of 1, 2, and 4 are adopted to capture both local and long-range temporal dependencies, while spatial convolutions and attention mechanisms enhance inter-channel representations. Subsequently, subject-specific tokens are introduced to explicitly characterize individual neural patterns and reduce the influence of inter-subject variability. Based on these representations, the multi-level interactive attention mechanism (MIAM) progressively extracts hierarchical semantic information through channel-level, subject-level, and global-level attention interactions. The resulting features are further processed by a six-layer Transformer encoder to model contextual dependencies. Finally, a KAN-based projection head with learnable activation functions is employed to establish a more effective nonlinear mapping between high-dimensional EEG features and the language embedding space of a pretrained BART decoder, enabling accurate text generation. Experiments are conducted on the publicly available Zurich cognitive language processing corpus (ZuCo), including both ZuCo v1.0 and ZuCo v2.0 datasets. The proposed ST-MSIAN achieves a BLEU-1 score of 44.10%, a BLEU-3 score of 16.21%, a ROUGE-F score of 37.62%, and a BERTScore-F score of 60.08%. Compared with the strongest competing method, the proposed approach improves BLEU-1, ROUGE-F, and BERTScore-F by 2.23, 4.87, and 4.66 percentage points, respectively. The proposed ST-MSIAN provides an effective and robust solution for cross-subject neural language decoding and offers promising potential for future applications in assistive communication systems and intelligent brain-computer interfaces.

Highlights:

1. A novel ST-MSIAN framework is proposed for open-vocabulary EEG-to-text decoding.
2. A spatiotemporal feature enhancement module jointly models temporal dynamics and spatial channel dependencies.
3. Subject-specific tokens are introduced to explicitly reduce inter-subject variability and improve cross-subject generalization.
4. A multi-level interactive attention mechanism captures channel-level, subject-level, and global semantic representations.

Key words: electroencephalography (EEG) signal; EEG-to-text decoding; spatiotemporal learning; interactive attention mechanism

Foundation item: Natural Science Foundation of Shanghai (No.22ZR1443700).

Received: 2026-01-23; **Revised:** 2026-07-15

***Corresponding author, E-mail:** wyxiong@usst.edu.cn.