

Foundation Model-Driven Paradigms in Brain-Computer Interface Encoding and Decoding

WU Xia^{1*}, LI Tongtong^{1,2}, LI Ziyu¹, MA Xiaoqiang², LI Jinke², LI Qing¹, YAO Zhijun²

(1. School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China; 2. School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China)

Abstract: Brain-computer interface (BCI) establishes a mapping relationship between external stimuli and internal neural activity in the brain, providing an effective means to understand brain information processing mechanisms and achieve human-machine intelligent interaction. In recent years, foundational models have achieved breakthrough progress in various computer vision tasks, which has also propelled BCIs from task-specific models toward a general intelligence new paradigm. This paper is the first to review the latest research advances of foundational models in neural encoding and decoding for BCIs. It systematically outlines key studies and research trajectories in natural stimulus encoding-decoding, multimodal brain representation learning, and generalization studies. The analysis identifies current challenges in sample size, data heterogeneity, multimodal fusion, and model interpretability. Finally, it highlights future research directions for general-purpose BCIs. This work aims to provide a systematic reference and research insights for building general BCI models capable of handling complex cognitive scenarios.

Highlights

1. This paper presents the first comprehensive review of foundation model-driven paradigms in brain-computer interface (BCI) encoding and decoding, systematically summarizing recent advances in neural encoding, neural decoding, and unified brain foundation models.
2. It provides an in-depth analysis of key research directions, including natural stimulus-driven brain representation learning, multimodal neural representation modeling, and cross-task generalization mechanisms enabled by foundation models.
3. Based on current progress, this paper identifies major challenges such as data heterogeneity, multimodal fusion, limited data scale, and model interpretability, and outlines promising future directions toward general-purpose BCI systems.

Key words: brain-computer interface (BCI); foundational models; neural encoding; neural decoding; generative artificial intelligence

基础模型驱动的脑机接口编解码新范式

邬霞¹, 李同同^{1,2}, 李子遇¹, 马晓强², 李锦科², 李晴¹, 姚志军²

(1. 北京理工大学计算机学院, 北京 100081; 2. 兰州大学信息科学与工程学院, 兰州 730000)

摘要: 脑机接口 (Brain-computer interface, BCI) 通过建立大脑外部刺激与脑内神经活动之间的映射关系为理解大脑信息处理机制并实现人机智能交互提供了有效手段。近年来, 基础模型在各项计算机视觉任务中取得了突破性进展, 这也推动了 BCI 从依赖任务的专用模型迈向通用智能的新范式。本文首次综述了基础模型在 BCI 神经编码与解码中的最新研究进展, 重点梳理了在自然刺激编解码、多模态脑表征学习及泛化性研究等方面的主要工作和研究脉络, 分析了当前研究在样本规模、数据异质性、多模态融合及模型可解释性等方面所面临的挑战, 最后展望了通用 BCI 的未来研究方向。本文旨在为构建面向复杂认知场景下的通用 BCI 模型提供系统性参考与研究启示。

关键词: 脑机接口; 基础模型; 神经编码; 神经解码; 生成式人工智能

中图分类号: TP391 **文献标志码:** A

引用格式: 邬霞, 李同同, 李子遇, 等. 基础模型驱动的脑机接口编解码新范式[J]. 数据采集与处理, 2026, 41(2): 439-460. WU Xia, LI Tongtong, LI Ziyu, et al. Foundation model-driven paradigms in brain-computer interface encoding and decoding[J]. Journal of Data Acquisition and Processing, 2026, 41(2): 439-460.

引言

脑机接口 (Brain-computer interface, BCI) 是一种革命性人机交互方式, 主要通过采集并解析大脑神经活动信号实现大脑与外部设备之间的信息传递与控制^[1]。近年来, 随着神经信号采集手段与计算技术的持续发展, BCI 已逐步从以实验验证为主的研究阶段走向面向复杂应用场景的实践探索^[2]。2025年7月, 工信部联合多部门发布了《关于推动脑机接口产业创新发展的实施意见》, 这一文件标志着我国 BCI 产业已从“科研探索为主”迈向“国家战略引领的产业化攻坚”的新阶段^[3]。神经编解码是 BCI 系统的核心环节, 神经编码关注外部刺激如何在神经系统中被表示为特定的脑活动模式, 神经解码则从观测到的脑信号中反向推断个体的刺激内容^[4]。神经编解码共同构建了大脑内部活动与外部信息之间的闭环映射关系, 使 BCI 系统实现了大脑表征内容的重构与解释。

近年来, 人工智能 (Artificial intelligence, AI) 方法在 BCI 神经编解码研究中展现出强大的竞争力。然而, 传统基于专用模型 (Task-specific model) 的方法对特定受试者数据依赖性较强, 难以应对跨模态多任务的场景需求。随着大规模预训练技术的成熟, 通用基础模型 (Foundation model, FM) 逐步成为推动 BCI 发展的重要方向。如图 1 所示, 以“Brain-computer interface” (蓝色) 和“Brain-computer interface; Foundation model” (橙色) 为关键词在 Web of Science 数据库中进行检索可以发现: 在过去 10 年中, BCI 相关研究的年发文数量持续增长, 整体呈现出稳定上升趋势。尤其在近 5 年, FM 在表示学习和任务迁移方面的优势逐渐显现, 使其已成为 BCI 神经编解码研究领域的重要组成部分和前沿研究方向之一^[5]。

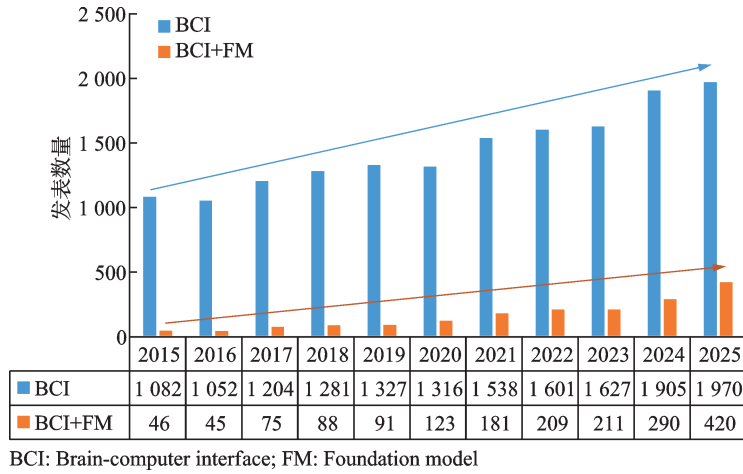


图1 近10年BCI相关研究发文统计

Fig.1 Statistics on BCI-related research publications in the past decade

在上述背景下,本文系统梳理了基础模型赋能BCI编解码的最新研究进展,并在此基础上总结了面向BCI的基础模型整体范式。如图2所示,该范式以自然视觉、语言及语音等刺激为输入,通过基础模型学习刺激表征与神经活动之间的潜在对应关系,在神经编码方向上刻画外部刺激向脑活动模式的映射过程,在解码方向上解析脑信号中所包含的语义信息。这种基于基础模型的编解码闭环映射关系建模为构建通用BCI系统提供了具有可推广性的技术路径。

综上,本文围绕基础模型在神经编解码建模、多模态脑表征学习及跨任务泛化机制等方面的研究进展进行系统梳理,进一步分析其在视觉、语言和听觉等自然刺激场景中的应用潜力。最后,也对该领域当前面临的关键挑战及未来发展方向进行了客观讨论。特别地,本综述首次从基础模型视角回顾神经编解码技术的研究现状,在一定程度上弥补了当前国内缺少可参考文献的局限。这一工作有助于加深对基础模型驱动的BCI编解码研究的理解,并为通用BCI的理论探索与工程实现提供较为完整的参考框架。

1 BCI编解码与基础模型

基于BCI的脑数据处理可归结为神经信息编码与解码问题。近年来,通用基础模型在多个医疗视觉领域展现出鲁棒的特征学习和迁移能力^[6],这为神经编解码提供了新的技术路径。本章将围绕神经编码与解码、基础模型发展脉络及常见大脑信号数据集进行综述,为后续研究提供理论支撑。

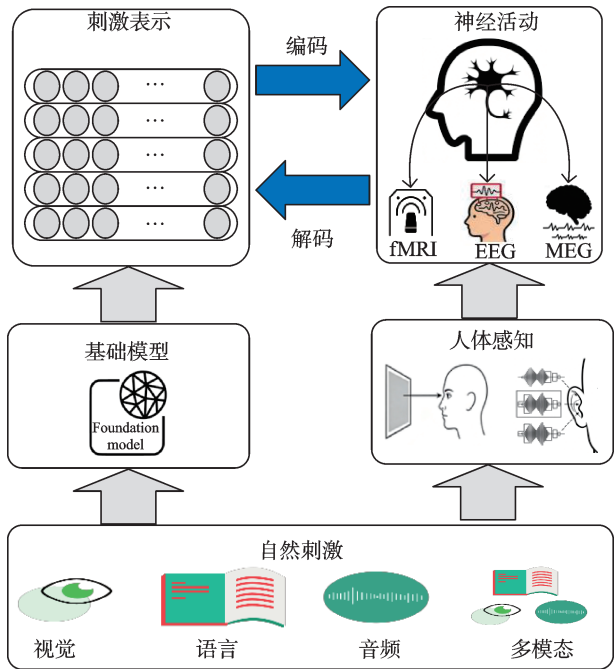


图2 基于基础模型的BCI神经编解码的研究范式

Fig.2 Research paradigm for BCI neural encoding and decoding based on foundation models

1.1 神经编码与解码的理论内涵

刺激是大脑信息处理的起点,它既包括视觉和听觉等外部感觉输入,也可以源于个体自身,如主动产生的运动意图或情绪状态变化等内源信息^[7-8]。这些不同类型的刺激会在大脑中引发相应的神经活动模式,并通过神经元放电、节律振荡或功能连接变化等形式进行表达^[9-10]。在此基础上,神经编码模型主要预测外部刺激条件下的大脑神经反应,而神经解码模型则从观测到的神经活动中重建或识别相应的刺激内容^[11](见图2)。

1.2 基础模型的发展脉络

基础模型主要通过大规模数据驱动的预训练学习具有较强泛化性的特征表示,满足可被灵活迁移至不同任务和应用场景中的统一建模需求^[12]。基础模型的发展历程如图3所示。

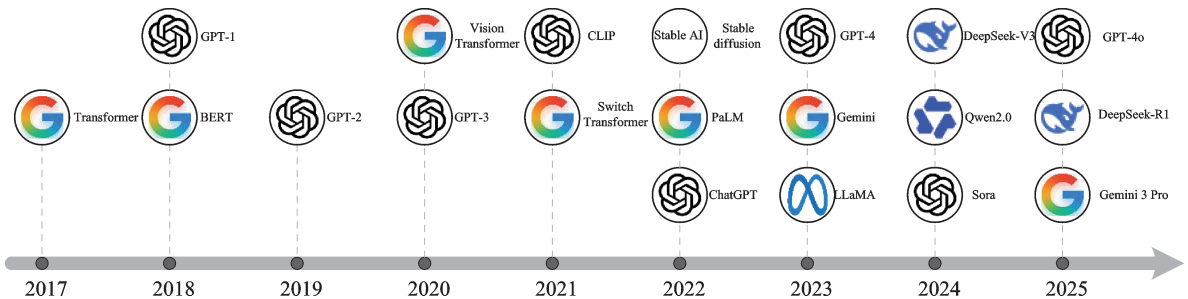


图3 基础模型的发展历程

Fig.3 Evolution of foundation models

2017年,Transformer模型^[13]通过引入自注意力机制(Self-attention mechanism, SAM)有效缓解了传统深度神经网络在建模长距离依赖时存在不足的缺陷,为大规模建模提供了高效计算范式。随后,BERT与GPT-1进一步确立了以“自监督预训练-下游任务微调”的通用建模框架,使模型可以利用大量无标注文本自动学习句法和语义信息,使模型具备上下文理解能力^[14]。2019年以后,随着缩放定律在多项研究中得到验证,基础模型进入了以参数规模和训练数据规模快速增长为特征的发展阶段^[15]。以GPT-3为代表的大规模语言模型在参数量增长后表现出较好的少样本推理能力。与此同时,视觉Transformer(Vision Transformer, ViT)^[16]将Transformer架构成功引入计算机视觉领域,ViT通过将图像划分为序列化的Patch,使视觉建模在结构层面与语言建模实现了统一。此后,跨模态对齐与生成建模成为研究的主流方向,CLIP模型^[17]通过对比学习策略将图像与文本映射至同一表示空间,从而实现了视觉特征与语言语义之间的有效对齐。与此同时,Switch Transformer^[18]的提出标志着模型规模竞赛正式进入“万亿参数”时代,大规模参数下实现计算效率与模型容量平衡成为新的研究热点。

2022年至2023年是生成式基础模型的快速发展期。Stable diffusion模型^[19]的发布使扩散模型(Diffusion model)在图像生成任务中占据主导地位。ChatGPT通过引入人类反馈的强化学习机制,有效提升了模型对用户指令的理解能力。GPT-4与Gemini等模型^[20]则进一步强化了基础模型的多模态理解能力与复杂逻辑推理水平。2024年以后,基础模型的发展方向开始向物理世界建模、高效计算及深度推理发展。Sora模型^[21]在视频生成中对物理规律与时空一致性的鲁棒性表明生成模型开始具备动态场景理解能力。DeepSeek-V3与Qwen2.0等模型在降低训练与推理成本的同时,整体性能甚至超过部分闭源模型^[22-23]。最新的DeepSeek-R1与Gemini 3 Pro则进一步强调了长链条推理能力的学习,标志着基础模型正从以语言处理为核心逐步迈向更具通用性的认知智能阶段^[24-25]。

综上所述,随着基础模型不断向多模态理解和深度推理方向发展,新一代模型在训练阶段对文本、图像、音频与视频等多模态信息进行端到端联合建模。这种高度抽象的特征表示方法为构建真实物理世界与复杂生物信号之间的映射关系提供了有效工具,也为基础模型驱动下的神经编解码奠定了技术基础^[26]。

1.3 大脑数据的独特性与挑战

在BCI研究中,常见的大脑信号数据包括非侵入式与侵入式数据。其中,脑电图(Electroencephalography, EEG)因时间分辨率高、数据采集便捷而被广泛用于运动意图识别、视觉诱发及语言感知等任务;脑磁图(Magnetoencephalography, MEG)通过记录神经电流产生的磁场变化,具备较高的时间分辨率和空间定位能力;功能性磁共振成像(functional magnetic resonance imaging, fMRI)能够以高空间分辨率刻画全脑的功能活动。侵入式信号如皮层脑电(Electrocorticography, ECoG)和神经元放电记录则具有更高的信噪比和更精细的空间分辨率。此外,近红外光谱成像(functional near-infrared spectroscopy, fNIRS)等技术也被用于BCI研究中,用于补充血流动力学信息。常见的大脑信号数据集如表1所示。

表1 常见的大脑信号数据集
Table 1 Common brain signal datasets

| 数据集 | 刺激类型 | 数据模态 | 被试数量 | 下载地址 |
|------------------------------|--------|----------------|------|---------------------------------------------------------------------------------------------------------------------------------|
| NSD ^[27] | 视觉(图像) | fMRI | 8 | http://naturalscenesdataset.org/ |
| GOD ^[28] | 视觉(图像) | fMRI | 5 | https://openneuro.org/datasets/ds001246 |
| Shen等 ^[29] | 视觉(图像) | fMRI | 3 | https://openneuro.org/datasets/ds001506 |
| Vim-1 ^[30] | 视觉(图像) | fMRI | 2 | https://crcns.org/data-sets/vc/vim-1 |
| Narratives ^[31] | 听觉(故事) | fMRI | 345 | https://openneuro.org/datasets/ds002345 |
| Bhattasali等 ^[32] | 听觉(故事) | fMRI | 29 | https://openneuro.org/datasets/ds002322 |
| Gifford等 ^[33] | 视觉(图像) | EEG | 10 | https://osf.io/3jk45/ |
| ZuCo ^[34] | 视觉(阅读) | EEG, 眼动追踪 | 12 | https://osf.io/q3zws/overview |
| OpenMIIR ^[35] | 听觉(音乐) | EEG | 10 | https://github.com/sstober/openmiir |
| Palazzo等 ^[36] | 视觉(图像) | EEG | 6 | https://github.com/perceivelab/eeg_visual_classification |
| SEED ^[37] | 视觉(电影) | EEG, 眼动追踪 | 15 | https://bcmi.sjtu.edu.cn/home/seed/index.html |
| DREAMER ^[38] | 听觉(故事) | EEG, ECG | 23 | https://zenodo.org/records/546113 |
| Features-EEG ^[39] | 视觉(图像) | EEG | 16 | https://openneuro.org/datasets/ds004357/versions/1.0.1 |
| MEG-MASC ^[40] | 听觉(故事) | MEG | 27 | https://osf.io/ag3kj/overview |
| THINGS-data ^[41] | 视觉(图像) | fMRI, MEG | 4 | https://github.com/ViCCo-Group/THINGS-data |
| Wakeman等 ^[42] | 视觉(图像) | fMRI, MEG, EEG | 19 | https://openneuro.org/datasets/ds000117/versions/1.1.0 |

多种大脑信号数据在实验范式和数据特性等方面与传统工程信号数据存在明显差异。从刺激角度看,脑信号数据通常围绕特定类型的刺激进行采集,不同刺激的神经活动在时空结构和信息表达形式上存在较大区别。例如,视觉刺激通过静态图像或动态视频呈现给受试者,这类刺激下的脑信号在空间分布上往往具有一定的一致性。但由于采集成本较高、样本规模受限且个体差异较大等因素,其编解码模型构建与泛化性仍面临较大挑战^[43]。听觉刺激以语音、音乐或环境声音作为输入,侧重于大脑对时间连续信号的感知与理解过程。与视觉刺激相比,听觉刺激诱发的神经活动更强调时间动态特征,信号随刺激内容呈现出明显的时序变化,因此对模型的时间建模能力提出了更高要求。同时,听觉相关脑信号在节律变化和瞬时响应方面往往更加复杂,这也进一步增加了表示学习的难度^[7]。文本刺激通常出现在阅读、语义判断或自然语言理解等任务中,其加工过程涉及词汇、句法和语义等多个层级。相比低层感知刺激,语言刺激诱发的脑活动更具抽象性与上下文依赖性,这对模型的长距离依赖建模与语义对齐能力提出了更高要求^[44]。此外,多模态刺激通过联合呈现视觉、听觉以及文本等多种信息通道,构建更接近真实世界的认知情境,近年来逐渐成为研究热点。在多模态刺激条件下,大脑需

要对不同模态的信息进行对齐与融合,其神经响应不仅包含单一模态特征,还体现出跨模态交互与协同加工的机制。

不同刺激范式和采集手段使脑机接口研究面临更加复杂的建模挑战。就数据特性而言,不同刺激条件下的脑信号普遍具有高噪声、非平稳以及低信噪比等特点,即使在相同实验条件下,不同被试之间的神经响应模式也存在显著差异,而同一被试在不同时间采集到的信号分布也可能发生漂移现象。这种个体差异和时间非稳定性限制了模型在跨被试下的迁移能力。此外,受实验成本、伦理限制及采集条件等影响,现有脑信号数据样本规模普遍受限,标注形式上也受到较大约束,使其难以满足大规模预训练建模的海量数据需求。

综上,由于不同类型的脑信号数据神经活动模式差异明显、噪声水平较高、个体差异显著以及数据规模受限等,所以BCI脑信号建模面临巨大挑战。这些特性决定了通用基础模型在脑信号场景中的应用需要在表示学习目标、训练范式和模型结构层面进行针对性改进,而非简单沿用自然数据领域的成功经验。

2 基础模型驱动的神经营编码

2.1 面向自然刺激的脑表征建模

面向自然刺激的脑表征学习以输入刺激对象,通过人工神经网络等构建刺激特征与人脑神经活动之间的映射关系,来揭示大脑对不同刺激信息的表征方式及其处理机制^[45]。根据实验过程中所采用的外界刺激模态,本节从文本、视觉和语音3种刺激输入综述基础模型驱动下的神经信号表征学习研究。

在文本刺激下的脑表征学习领域,基础模型尤其随着大型语言模型(Large language model, LLM)的引入,彻底重塑了研究者对大脑语言处理机制的理解范式。不同于早期基于静态词向量编码,如Word2Vec^[46],基于Transformer的预训练模型能够结合上下文动态生成词嵌入,这使其在预测大脑对自然连续文本的神经响应时表现出更高的准确率。已有研究表明,LLM的内部表征与人类阅读文本时产生的脑活动存在相关性^[47]。然而,Fodor等^[48]发现,现有基础模型往往依赖于词汇语义的“词袋”特征而非严格的句法规则来构建表征,这一结果表明模型内部的处理方式与大脑在句法层级上的加工机制仍存在本质差异。这一发现使“模型即大脑”的简单类比受到质疑,促使学界开始探索结合了显式句法树结构的混合编码模型,试图在保留LLM强大语义表征能力的同时,增强其对人类大脑句法加工特异性的解释力度^[49-50]。此外,为了突破单一语言研究的局限性,基于多语言预训练模型的跨语言脑编码工作逐渐兴起。De Varda等^[51]利用包含21种语言的fMRI数据集训练多语言编码模型,结果发现经过机器翻译模型对齐后,不同语言在大脑中激活的语义网络高度重叠,这表明大脑存在一个独立于具体表面形式的通用语义表征核心,也即大脑对文本刺激的加工具有跨语言特性^[52]。这一发现为开发通用BCI算法提供了重要依据。当前,该领域研究已不局限于利用现有自然语言处理(Natural language processing, NLP)模型作为分析工具,如Caro等^[53]提出的BrainLM模型通过在大样本fMRI数据上进行自监督掩码训练,学习到了大脑活动的通用时空动力学特征,在零样本条件下即可实现对语言任务脑激活图的精准重建。这种从“用模型理解大脑”到“构建大脑本身的模型”的跨越,标志着神经编码研究进入了一个数据驱动与机制解析并重的新阶段^[54]。

视觉刺激下的脑表征学习主要通过模拟人类视皮层从初级特征提取到高级视觉语义理解的层级化处理过程,旨在建立外部视觉刺激与内部神经活动之间的映射关系。近年来,已有研究发现深度视觉模型在其表征空间中能够与人类视皮层形成较为稳定的线性对齐关系^[55]。基于此,Wu等^[56]构建了基于自监督视觉Transformer的视觉基础模型,揭示视网膜神经元群的时间编码模式,实现了视频脑表征与神经活动的对齐。另有实验表明,通过在多层编码模块中引入EEG信号作为正则化约束,能训练出在物体识别与神经表征上更接近大脑的视觉模型,这为构建具有生物保真度的神经基础模型提供了

新的算法思路^[57-58]。未来,随着可用脑数据和视觉数据规模的持续增长,如何构建能够同时覆盖视觉感知、注意力调控以及记忆检索等多个认知过程的通用视觉脑编码基础模型,将成为该领域亟需突破的重要研究方向。

在自然听觉刺激表征学习研究中,基础模型对语音刺激的建模已不局限于声学层面,还扩展到语义内容和上下文信息等高级语义表示。Rupp等^[59]指出多任务优化的深度模型能够反映听觉皮层不同层次加工的复杂性与时间尺度,这种层次性与人脑对于语音的多层加工过程一致,表明基础模型可以作为计算神经科学中的有效代理进行语音语义层次的建模。针对自然交互场景中连续语音流带来的脑编码难题,研究者开始引入基础语言模型的整体架构来模拟大脑的编码机制。Caucheteux等^[60]构建了基于Transformer的语言模型,通过跨越多个时间尺度对模型激活与脑反应进行映射,发现模型内部表示与人类在听自然语音时的脑活动之间存在稳定的线性对应关系。Borsos等^[61]将语音信号映射为多层离散表示,并通过语言建模范式分别建模局部声学结构与长距离时序依赖,实现了对语音韵律与结构一致性的有效刻画。

未来,面向自然交互场景下复杂自然刺激的脑编码研究正逐步从以单一模态为主的建模方式,转向跨模态的统一表征框架。在此趋势下,融合文本、视觉和语音信息的基础模型有望同时建模语言输入在不同感知通道中的协同加工过程。随着多模态脑机接口技术的持续发展,构建能统一解析多类型自然刺激的通用脑编码基础模型,将为理解人类语言感知的本质提供技术支撑。

2.2 多模态基础模型驱动下的脑表征学习

在认知神经科学领域,人类大脑被视为一个天然的跨模态整合系统,它能将视觉场景、语言描述与听觉信号整合构建统一的世界表征。近年来,随着多模态基础模型的发展,研究者开始利用多模态基础模型提取的多模态对齐特征来探索大脑皮层中的多感知对齐机制。多模态基础模型通过提取不同感知通路的刺激信息,使编码模型突破传统单一模态模型在预测人类大脑活动时的性能限制,学习更加丰富的跨模态神经表征,提高复杂认知任务预测效果^[62]。

已有研究提出基于对比学习(Contrastive learning)的多模态基础架构,通过对齐EEG、MEG及fMRI等多种神经信号在模型的内部表示,实现跨神经信号模态间的特征编码和对齐,并在多个任务中表现出色^[63]。Chen等^[64]构建了大规模EEG-语言对齐数据集,通过将语音、阅读等多模态语言任务映射到一个共同语义空间中,不仅促进了跨语境脑编码模型的语义一致性学习,而且能够在不同语言任务下构建共享特征映射,使得不同受试者之间的脑活动表征共享成为可能,这也在大脑高级认知区响应多源刺激时表现出较高的可解释性^[65]。也有研究借助基于Transformer的通用架构来构建多模态脑编码模型。在最新的Algonauts脑编码竞赛中,融合文本、视觉和听觉特征的多模态Transformer方法,在预测全脑fMRI反应方面明显优于传统单模态模型,同时也提升了对高级联合皮层区域进行功能分类的性能^[66]。这些结果表明,多模态编码架构能有效模拟大脑在自然环境中面对复杂刺激时的集成加工机制,这种跨模态融合的神经表征机制在解释联觉体验与感知互补性方面展现出独特优势^[67]。

此外,多模态表征的引入使得神经编码模型从简单的特征映射进化为复杂的语义逻辑推断。针对动态多模态环境下的脑表征学习,已有研究开始关注时空特征与语义序列在基础模型中的协同演化过程。Chen等^[68]提出的Mind-Video框架利用视觉扩散模型的生成先验知识,同时整合了文本引导的语义嵌入,发现这种多模态融合特征在解释受试者观看电影时的神经活动时,其性能表现远超任何单模态编码器。在其他工作中,通过引入自监督对比学习和大规模无标注数据预训练,表明多模态基础模型所学到的统一表征使得神经编码模型在音频、文本与视觉多源刺激的融合语义层次上具备更强的跨任务适应能力^[69-70]。表2列举了典型的基础模型驱动下的自然刺激脑表征建模研究。

未来,随着大规模脑信号数据集与通用多模态架构的进一步发展,构建能够涵盖人类认知的脑表征基础模型将为通用人工智能与生物大脑的底层接口提供关键的算法路径。

表2 基础模型驱动下的自然刺激脑表征建模

Table 2 Modeling brain representations of natural stimuli driven by foundation models

| 文献 | 发表时间 | 刺激范式 | 数据模态 | 主要贡献 |
|------|------|------|--------------|------------------------------------------|
| [57] | 2024 | 视觉 | EEG | 首次提出利用EEG优化多层模型,使视觉表征更接近人类大脑处理模式 |
| [56] | 2024 | 视觉 | 视网膜神经细胞记录 | 揭示了视网膜神经元对动态自然场景的编码模式,建立了像素与响应间的时空联系 |
| [71] | 2025 | 视觉 | NSD | 实现了跨主体、跨数据集、跨设备的迁移学习,利用“大脑群体智慧”提升性能 |
| [61] | 2023 | 听觉 | 音频 | AudioLM: 将音频生成视为语言建模任务的框架,实现了高质量的语音和音乐续写 |
| [59] | 2025 | 听觉 | EEG | 揭示了人类听觉皮层在处理自然声音时,其处理复杂性和时间尺度呈层级化分布 |
| [51] | 2025 | 语言 | fMRI | 发现多语言模型能捕捉不同语言间共享的意义组件,实现零样本跨语言迁移预测 |
| [48] | 2025 | 语言 | fMRI | 发现人类大脑对句子的表征与大模型存在差异,强调了句子结构在脑中的核心作用 |
| [63] | 2025 | 多模态 | EEG、MEG、fMRI | 建立了视觉、语言与大脑活动的统一嵌入空间,提升了对大脑概念理解的深度 |
| [66] | 2025 | 多模态 | fMRI、视频 | 提出了首个能精确预测全脑对多模态刺激响应的模型,尤其在高级联合皮层表现优异 |

2.3 神经编码基础模型的跨任务泛化机制

传统神经编码模型通常受限于具体的实验范式,例如在视觉任务训练的模型无法直接应用于听觉或记忆任务,而基础模型的引入打破了这种“任务孤岛”现象,它通过学习大脑活动中的通用结构特征使模型具备跨任务迁移的能力。以BrainLM^[53]为代表的大型预训练模型利用掩码自编码器策略在数千小时的静息态与多任务fMRI数据上进行自监督学习,成功提取了独立于具体认知内容的皮层时空依赖规律。这种模型不再依赖于特定的外部刺激标签,而是将大脑活动视为一种遵循特定语法的神经语言,通过预测被掩码的神经信号片段,模型被迫学习大脑各区域间潜在的因果连接与协同模式。

为解决不同受试者及不同扫描设备间的数据异构性问题,构建体素级的通用嵌入空间成为实现跨个体泛化的有效方案。针对跨数据集与跨模态的泛化挑战,最新研究提出了基于图神经网络的BrainGFM框架^[54]将不同脑图谱映射为统一的图拓扑结构,利用图对比学习在超过25 000名受试者的海量数据中学习节点间的拓扑不变性。Cui等^[72]通过引入图提示机制,使基础模型能根据不同的下游任务动态调整其注意力分配策略,从而在完全未见过的临床数据集上实现超越传统迁移学习方法的诊断性能。Beliy等^[71]开发的通用脑编码器则采用了一种以体素为中心的编码架构,通过在“众包大脑”(Crowd of brains)数据上联合训练,模型学会了忽略解剖结构上的个体差异,转而关注功能同构性,这使得利用3T低分辨率数据训练的模型参数可以直接迁移至7T高分辨率数据分析中,极大降低了高精度神经编码的各种数据采集门槛。此外,也有研究^[67]尝试利用EEG与fMRI进行跨模态知识蒸馏,证明了在一种模态上习得的预训练特征可通过适配器网络迁移至另一种模态,为构建通用神经编码器奠定了理论基础。

现有的神经编码基础模型通过自监督学习与图拓扑对齐技术,成功打破了传统专用模型在任务、个体及模态间的难以泛化的限制。相关方法不仅在静息态数据中提取了通用的皮层动力学特征,还通过跨模态知识蒸馏实现了从单一感官刺激向复杂认知任务的零样本迁移,初步验证了大脑功能同构性在计算层面的可操作性。未来,神经编码基础模型的跨任务泛化机制有望进一步向认知层级泛化与闭

环适应方向发展,通过引入因果模型^[73-74]、预测性编码约束^[75]及与大语言模型协同的多尺度对齐^[76]策略,使模型不仅能在不同任务之间迁移已有表征,还能在未知任务与自然交互环境中主动重组神经表示,从而推动通用脑编码器向真正意义上的“认知级基础模型”发展。

3 基础模型驱动的神神经解码

3.1 基于生成式大模型的神神经解码

3.1.1 视觉感知重构:从静态图像到动态视频

视觉感知重构旨在建立大脑神经活动与外部视觉刺激之间的映射关系,是检验基础模型通用表征能力的重要应用场景之一。得益于生成式大模型在语义建模和跨模态对齐能力,该领域的重点已从早期的像素级拟合转向语义一致的高保真生成。受不同神经成像手段物理特性的影响,现有研究主要围绕 fMRI 空间重构与 EEG/MEG 时序解码展开,并向连续视频与三维场景延伸。

fMRI 的高空间分辨率为视觉重构提供了丰富信息。早期 fMRI 解码工作主要依赖卷积神经网络 (Convolutional neural network, CNN) 和生成对抗网络 (Generative adversarial networks, GAN)。例如 BrainDVG^[77]、BigBiGAN^[78]、BrainICG^[79]、SBD^[80] 以及 MindReader^[81], 这些方法在一定程度上验证了解码重构的可行性,但在图像细节与语义准确性上仍存在明显不足。近年来,随着扩散模型与视觉-语言预训练模型的引入, fMRI 重构质量得到了显著提升。具体而言,研究者致力于将 fMRI 信号映射至预训练模型的生成空间进行解码重构,例如 BrainSD^[69]、BrainSD-TGD^[82]、MindDiffuser^[83]、BrainDiffuser^[84]、Brain-Diffuser^[85] 及 MinD-Vis^[86] 等模型通过微调 Stable Diffusion^[19] 或利用 VQ-VAE^[87] 的离散码本,成功实现了从脑活动到高分辨率图像的端到端生成。在此基础上, CAD^[88] 以提升语义一致性为目标,引入注意力机制优化生成质量, LEA^[89] 则专注于潜在嵌入的精细对齐。此外,为提升语义一致性, BrainCLIP^[90]、MindEye^[91]、MindEye2^[92] 等工作利用 CLIP 的多模态嵌入空间作为语义锚点,通过对比学习或多层级特征融合,显著提升了重构图像在语义类别上的准确率; UniBrain^[93]、BrainCaptioning^[94] 及 UMBRAE^[95] 则尝试结合图像与文本的多模态解码。

尽管基于基础模型的视觉重构在单被试范式下发展迅速,但个体间大脑解剖与神经信号分布的差异导致的域偏移 (Domain shift, DS) 已成为迈向多场景应用的核心瓶颈。为此,跨被试解码的研究重心正从传统的单被试单模型转向具有强泛化能力的多被试共享与对齐路线。在 fMRI 跨被试解码工作中,主流方法之一是跨个体功能对齐与特征映射。例如, MindAligner^[96] 通过脑转移矩阵与多级对齐损失,实现了新被试到已知被试的体系级软对齐; STTM^[97] 采用浅层被试适配器结合共享深层解码器的分层架构,在输入端实现跨被试 fMRI 特征的早期对齐并在高层学习共享语义表示,有效缓解了个体差异带来的分布偏移; Neuro-Vision to Language^[98] 则利用 ViT3D 提取统一的三维脑特征并对齐多层视觉嵌入,从而无缝接入 LLM。而对于 EEG/MEG 信号而言,模型不仅需要克服信号分布漂移,还需解决通道布局不一致与低信噪比等严峻挑战。针对这些特性,研究者提出了一系列表征提取与域适应策略以消除个体壁垒。在大规模预训练方面, EEGPT^[99] 构建了通用 Transformer 架构以学习共享表征,间接支持了跨被试与跨数据集的多任务迁移; CL-SSTER^[100] 与 CLISA^[101] 分别在脑电语义理解和情绪解码任务中,通过将不同个体对同一刺激的神经响应构造为正样本对,强制模型学习独立于个体的共享时空表示。面对复杂的流形空间与域偏移, ITSA^[102] 创新性地黎曼与切空间中执行个体重置、分布匹配与监督旋转,有效克服了跨被试与跨电极布局的双重壁垒; Selective-MDA^[103] 引入差异感知的多源域适应框架,通过精准筛选并控制有利的源域被试特征,进一步巩固了高动态脑信号在未见被试上的解码鲁棒性。

此外,得益于 EEG 与 MEG 毫秒级的时间分辨率,这为捕捉视觉加工的瞬时动态提供了可能。基础模型通过强大的先验知识实现“脑补”式重构,有效弥补了其空间信息不足的问题。早期的 EEG 解码研究,如 Brain2Image^[104]、EEG-GAN^[105]、EEG-VGD^[106] 及 NeuroVision^[98] 等模型主要依赖 LSTM 或

GAN进行特征映射,生成的图像较为模糊。近期研究则通过将EEG/MEG信号对齐至CLIP语义空间并结合扩散模型,实现了零样本生成能力,如DreamDiffusion^[107]、NeuroImagen^[86]等模型利用EEG信号生成清晰图像。而ATM^[108]、EEG-CLIP^[109]及SeeEEG^[110]等研究进一步优化了对齐策略,提升了语义检索与生成结果的准确性。CATVis^[111]则关注了上下文感知的视觉重构,而NFM^[62]同样被应用于EEG解码任务中以提取流形特征。作为互补模态,MEG解码呈现出相似的技术路径,MEG-BD^[112]以及UBP^[113]等工作的研究证明,MEG信号在结合大模型后同样能实现高质量的视觉内容还原。

目前,视觉重构研究正由静态图像逐步迈向更具挑战性的动态与立体场景,这对神经编码模型在时间一致性与时空结构建模能力方面提出了更高要求。在fMRI解码领域,Mind-Video^[68]结合了时空掩码自编码器与视频扩散模型,实现了连贯的视频流重构。SSRNM^[114]、f-CVGAN^[115]等工作也从不同角度探索了动态视觉流的编码与解码机制,其中部分工作如NeuroCreat^[116]也涉及视频生成。在EEG领域,EEGMirror^[117]创新地利用孪生网络与对比学习,实现了基于脑电的视频片段检索与重构。此外,MinD-3D++^[118]率先将解码目标拓展至三维空间,利用fMRI信号生成点云或三维网格表示,标志着脑解码技术向感知物理世界的立体化迈出了重要一步。

3.1.2 语言与听觉解码:从语义理解到意图外化

语言与听觉解码旨在建立大脑神经活动与语言符号及声学特征之间的映射,不仅涵盖了对外部语言刺激的语义理解,更致力于实现内部语言意图的直接外化,是构建高性能BCI的关键。当前研究利用LLM与语音预训练模型在海量语料上学习的概率先验,有效解决了神经信号中语言特征稀疏与信噪比低等难题。

在语义理解层面,fMRI被广泛应用于解码长时程的宏观叙事语义。针对NarrativesfMRI等自然语言数据集,PredFT利用Transformer解码器实现了对连续文本流的预测,而MindGPT^[119]与Mind-SA^[120]则通过将脑活动对齐至GPT模型的语义空间,成功解析被试在自然语言感知过程中的高层语义内容。随着多模态大模型的发展,BrainChat^[121]与NeuroCreat^[116]等模型进一步引入了CoCa或LLaMA-3等视觉-语言模型,探索了在复杂情境下的语义解码与创造性生成能力。针对视频或多模态刺激的理解,LLM4Brain^[122]与BP-GPT^[123]结合大语言模型增强了对动态内容的语义解析,From Eye to Mind^[124]借助Vec2Text模型揭示了视觉到语言的语义转化机制。这些工作共同证明,大语言模型提供的上下文预测能力能够有效约束解码过程,从而生成语法连贯且语义准确的文本描述。

在便携式意图解码方面,EEG与MEG成为解码默读文本与想象语音等内源性信号的主要手段。围绕ZuCo^[36]等自然语句阅读理解数据集,EEG2TEXT^[125]与CET-MAE^[126]分别采用了卷积Transformer与多流Transformer架构,从眼动与脑电信号中提取语言特征,实现了对阅读语义的有效建模。DeWave^[127]结合VQ-VAE与BART模型,通过离散编码策略实现了开放词汇的脑电转文本生成。在更具挑战性的想象语音与打字意图解码中,Brain2Qwerty^[128]尝试从EEG与MEG信号中还原拼写意图,而PinVC^[129]则利用CLIP空间的对齐能力探索了基于脑电的文本生成。针对ISD数据集中的语音解码任务,ETS^[130]等方法通过建模神经信号的时序依赖,初步实现了对未发声语言意图的解析。这些尝试为开发非侵入式的静默交流脑机接口提供了重要的技术积累。

除了符号化的语言解码,针对语音波形与频谱图的声学重构也是还原听觉体验的重要维度。在这一领域,不同模态展现出互补的特性。Neural2speech^[131]在fMRI、MEG及sEEG等多模态数据上验证了利用Transformer和wav2vec 2.0等预训练模型直接从脑信号合成语音的可行性。针对精细的声学特征,基于EEG的Dual-DualGAN^[132]则尝试了端到端的语音波形生成。值得注意的是,在侵入式信号领域,SwinTW^[133]模型利用Swin Transformer架构在sEEG与ECoG数据上实现了对语音频谱图的高精度重构。此外,即便在HD-DOT光学成像模态下,相关研究也通过模板匹配等方法探索了听觉刺激信息的解码^[134]。这些工作表明,结合声学预训练模型的生成能力,脑解码技术逐渐具备还原音色、语调等丰富听觉细节的潜力。表3总结了近年来基础模型在不同大脑信号数据上的研究。

表3 基础模型驱动下的神经解码研究

Table 3 Research on neural decoding driven by foundation models

| 数据模态 | 解码模态 | 数据集 | 基础模型 | 方法 | | |
|---------|----------------|--------------------------------|-----------------------------------|---------------------------------|-----------------------------|----------------------------|
| fMRI | 图像 | NSD | CLIP | BrainSD ^[69] | | |
| | | | CLIP | BrainSD-TGD ^[82] | | |
| | | | VQ-VAE | MindDiffuser ^[83] | | |
| | | | CLIP | BrainCLIP ^[90] | | |
| | | | CLIP | BrainDiffuser ^[84] | | |
| | | | CLIP | MindEye ^[91] | | |
| | | | CLIP | MindEye2 ^[92] | | |
| | | | CLIP | MindReader ^[81] | | |
| | | | CLIP | NeuroFusionNet ^[135] | | |
| | | | CLIP | BRAINGUARD ^[136] | | |
| | | | CLIP | MindTuner ^[137] | | |
| | | | CLIP | CSMD ^[138] | | |
| | | | CLIP | FFR ^[139] | | |
| | | | CLIP | NFM ^[62] | | |
| | | | 文本 | NSD | CLIP | CMVDM ^[140] |
| | CLIP | MinD-Vis ^[86] | | | | |
| | DC-MAE | CAD ^[88] | | | | |
| | CNN | SBD ^[80] | | | | |
| | D-VAE | BrainDVG ^[77] | | | | |
| | SwAV ResNet-50 | BigBiGAN ^[78] | | | | |
| | | BrainICG ^[79] | | | | |
| | | Brain-Diffuser ^[85] | | | | |
| | VDVAE | LEA ^[87] | | | | |
| | NSD&-GOD | CLIP | | | NeuroCreat ^[116] | |
| | | CLIP | | | STTM ^[97] | |
| | | Narratives fMRI | | | PredFT ^[75] | |
| | NSD | NSD | | | CoCa | BrainChat ^[121] |
| | | | | | CLIP | MindGPT ^[119] |
| | | | | | CLIP | Mind-SA ^[120] |
| | | | CLIP | NeuroCreat ^[116] | | |
| | | | From Eye to Mind ^[124] | | | |
| CLIP | | | LLM4Brain ^[122] | | | |
| NL fMRI | NL fMRI | GPT-2 | BP-GPT ^[123] | | | |
| | | 图像、文本 | | UniBrain ^[93] | | |
| NSD | NSD | CLIP | BrainCaptioning ^[94] | | | |
| | | CLIP | UMBRAE ^[95] | | | |
| | | CLIP | Mind-Video ^[68] | | | |
| DNV | DNV | MARS | SSRNM ^[114] | | | |
| | | VAE(Encoder) | BrainViVAE ^[141] | | | |

续表

| 数据模态 | 解码模态 | 数据集 | 基础模型 | 方法 | | | | | |
|-----------------------------------------|-----------------------------------------|-----------------------------------------|-----------------------------------------|--------------------------------|-----------------------------|-------------------------------|---------------------------------|-----------------------------------------|------------------------|
| EEG | 三维 语音 | STN | CLIP | f-CVGAN ^[115] | | | | | |
| | | | | Mind2Word ^[142] | | | | | |
| | | NSD | CLIP | Brain2Pix ^[143] | | | | | |
| | | | | NeuroCreat ^[116] | | | | | |
| | | Self-collected dataset ^[118] | CLIP | MinD-3D++ ^[118] | | | | | |
| | | | | neural2speech ^[131] | | | | | |
| | | Self-collected dataset ^[131] | Transformer (Encoder) | CLIP | | | | | |
| | | | | Visual-Guided CNN | | | | | |
| | | 图像 | 图像 | EEG-AVC | CLIP | NeuroImagen ^[86] | | | |
| | | | | | Visual-Guided CNN | EEG-VGD ^[106] | | | |
| | CLIP | | | | EEG-GAN ^[105] | | | | |
| | CLIP | | | | NeuroVision ^[98] | | | | |
| | CLIP | | | | BrainVis ^[36] | | | | |
| | Things-EEG2 | | | Things-EEG2 | CLIP | GWIT ^[125] | | | |
| | | | | | | Perceptogram ^[126] | | | |
| | | | | | | ATM ^[108] | | | |
| | | | | | | UBP ^[144] | | | |
| | | | | | | EEG-CLIP ^[109] | | | |
| | Self-collected dataset ^[104] | MOABB | CLIP | SeeEEG ^[110] | | | | | |
| | | | | NSD | CLIP | Brain2Image ^[104] | | | |
| ZuCo | | | | | | Convolutional Transformer | DreamDiffusion ^[107] | | |
| | | | | | | | Multi-stream Transformer | VQ-VAE | NFM ^[62] |
| | | | | | | | | | Things-EEG2 |
| | ISD | XI Model | CET-MAE ^[130] | | | | | | |
| | | | Self-collected dataset ^[147] | LSTM-RNN | DeWave ^[145] | | | | |
| Self-collected dataset ^[127] | | | | | wav2vec2 | PinVC ^[146] | | | |
| | | | | | | MOUS | wav2vec2 | ETS ^[132] | |
| | | | | | | | | MEG | 图像 |
| | MEG narrative | DINOv2, CLIP | | | | | | | |
| | | | Self-collected dataset ^[150] | CLIP | | | | | |
| MEG | | | | | 图像 | | | | |
| | | | | | | MOUS | wav2vec2 | | |
| | | | | | | | | Self-collected dataset ^[131] | wav2vec2 |
| | sEEG | 文本 | | | | | | | |
| | | | 语音 | FastText | | | | | |
| Intracranial EEG | | | | | 文本 | | | | |
| | | | | | | ECoG | 语音 | | |
| | | | | | | | | HD-DOT | Auditory Stimulus Info |
| | HD-DOT | Template Matching | | | | | | | |
| | | | HD-DOT | Template Matching | | | | | |
| HD-DOT | | | | | Template Matching | | | | |
| | | | | | | HD-DOT | Template Matching | | |
| | | | | | | | | HD-DOT | Template Matching |

3.1.3 跨模态语义对齐与多任务联合生成

尽管前述工作分别展示了视觉与语言解码的进展,但人类的认知过程往往是多感官信息的综合产物。传统解码模型通常局限于单一的输入输出映射,难以捕捉大脑复杂的跨模态语义表征。得益于基础模型提供的共享语义空间,近期研究开始打破单一任务的边界,探索利用同一套神经表征实现跨模态语义对齐与多任务联合生成的可行性。

在fMRI解码研究中,研究者利用大模型的通用特征空间实现了“脑活动到多模态内容”的直接映射。如表3所示,UniBrain^[93]利用预训练大模型的潜在空间作为共享接口,成功实现了从fMRI信号中同时重构高分辨率图像并生成对应的文本描述。这种联合解码策略不仅提升了单一模态的重构质量,还在模型层面保证了视觉结果和语言输出的语义一致性。类似地,BrainCaptioning^[94]通过引入CLIP的图像-文本对比学习机制,使解码器能够理解大脑对视觉场景的语义表征,并将其转化为自然语言描述。此外,NeuroCreat^[116]与LLM4Brain^[122]等工作进一步拓展了这一范式,前者展示了在同一框架下进行图像重构、文本生成乃至创造性编辑的能力,后者则结合Video-LLaMA实现了对复杂视频内容的视觉语义问答。这些工作证明了通过将脑信号对齐至通用的多模态语义空间后,单一解码模型即可具备同时应对多种认知任务的能力。

在EEG解码领域,尽管信号的稀疏性限制了多任务生成的难度,但基于CLIP的语义对齐策略同样展现出跨任务的通用性。如表3所示,PinVC^[129]模型通过将EEG信号映射至CLIP的文本与图像编码空间,使其既能用于视觉图像的检索与重建,又可实现文本内容的解码与生成。这种方法的优势在于摆脱了对特定感知模态的强依赖,可实现跨模态任务的统一建模。总体来看,虽然目前的联合解码研究仍处于起步阶段,但这种基于共享语义空间的多任务生成范式,为未来构建能够同时理解视觉、语言及听觉意图的通用BCI提供了技术路径。

3.2 面向异构脑数据的统一基础模型构建

与前面所提利用视觉或语言大模型辅助特定任务解码的范式不同,构建面向大脑数据的统一基础模型旨在直接从海量、无标注的神经活动中学习通用的时空表征。此类研究借鉴了NLP领域中自监督预训练的成功经验,通过在大规模异构数据集上训练“大脑基础模型(Brain foundation models)”,试图捕捉神经信号中蕴含的内在动力学规律与功能连接模式。根据数据模态的物理属性与融合策略,现有工作主要围绕基于fMRI的大尺度功能网络建模、基于EEG的高通量时序表征学习以及面向异构信号的跨模态统一建模3个方向展开。

功能性磁共振成像提供了全脑范围的体素级活动信息,是构建全脑功能基础模型的首选模态^[153-154]。为了解决fMRI数据维度高且标注稀缺的问题,研究者引入了掩码脑建模(Masked brain modeling)策略。其中,BrainLM^[53]率先采用类似BERT的掩码预测任务,通过在大规模fMRI数据库上进行自监督预训练,学习到了能够泛化至多种心理测量任务的通用脑动力学表征。在此基础上,BrainMass^[155]进一步扩大了数据规模与模型容量,证明了模型规模扩展能够显著提升对脑活动时空模式的表征能力。为了克服生成式预训练在噪声数据上的局限,Brain-JEPA^[156]引入了联合嵌入预测架构,通过在潜在空间而非像素空间进行预测,有效提升了特征的语义抽象层次。近期工作则更加关注时空特征的精细化建模,SLIM-Brain^[157]针对长时程fMRI序列设计了结构化学习机制,而FMMTC^[158]则探索了多任务联合训练对提升模型泛化性的贡献。这些工作共同推动了fMRI分析从单一任务的小模型向通用的大模型范式转变。

相较于fMRI,EEG具有极高的时间分辨率,但面临着信噪比低、电极导联配置不统一以及个体差异显著等挑战。针对这些问题,现有研究致力于构建对导联位置不敏感且具备强抗噪能力的EEG基础模型。Neuro-GPT^[72]与BrainWave^[159]借鉴了GPT的自回归生成范式,通过在大规模EEG语料上进行

预训练,学习到了神经震荡信号中的短时与长时依赖关系。为了解决不同数据集间电极数量与位置不一致的难题,LaBraM^[160]提出基于神经向量量化的波段分割策略,实现了跨数据集、跨设备的导联无关建模,显著提升了模型在下游分类任务中的表现。此外,CBraMod^[161]与REVE^[162]进一步引入了对比学习与多视图编码机制,增强了模型对被试间生理变异的鲁棒性。这些模型证明了在大规模无标注EEG数据上进行预训练,是解决脑机接口领域“数据孤岛”与“泛化困难”问题的有效途径。

构建能够同时处理多种异构神经信号的通用模型是实现全方位脑活动表征的终极目标。然而,不同脑数据模态的巨大鸿沟使得跨模态建模面临严峻挑战。针对电磁信号的融合,BrainOmni^[163]提出了一种面向EEG与MEG的联合预训练框架,通过学习电场与磁场信号间的互补特征,构建统一的电磁神经表征空间。而在跨物理机制的融合方面,EFRM^[164]探索了EEG与fNIRS的联合建模,试图将高时间分辨率的电信号与反映代谢水平的光学信号相结合。这种跨模态的基础模型不仅能够利用不同模态数据的互补优势提升解码性能,更为未来构建涵盖全时空尺度的通用“大脑解释器”奠定了理论与技术基础。

4 现存挑战与未来方向

尽管基础模型为大脑神经编解码任务提供了新范式和新手段,但目前该领域仍面临诸多挑战:

(1) 样本规模受限及模型泛化不足。受限于数据采集设备的高昂成本和实验范式的复杂性,高质量标注的脑信号数据规模远不及传统自然视觉或语言处理领域的样本规模。这导致在预训练阶段基础模型容易陷入过拟合,难以在不同个体、不同脑区甚至不同物种之间进行有效迁移。未来应探索基于自监督对比学习的跨模态预训练技术,利用海量未标注的多模态神经数据提取共性的深度特征,并通过少样本学习或元学习算法,实现模型对新个体神经编码规律的快速适配,构建通用的大脑预训练模型。

(2) 多模态语义融合不足。当前基础模型驱动的大脑编解码研究多集中于单一模态,缺乏对大脑在复杂真实场景下多感官信息集成机制的全面刻画。现有的多模态模型在融合神经特征时往往简单地进行特征整合,无法还原大脑内部复杂的调控逻辑。未来研究可构建具身智能引导的基础模型,将感知信号与运动输出、环境反馈相结合,探究大脑如何在交互过程中构建统一的世界模型,并通过解析多模态基础模型的语义空间,揭示大脑处理抽象概念和逻辑推理的底层编码准则。

(3) 模型结构可解释性不足。尽管基础模型架构在脑表征学习和复杂时序结构方面表现出巨大优势,但其内部表征与真实神经机制之间的对应关系尚不清楚,这限制了其在神经科学领域的可解释性^[165]。此外,现有基础模型往往基于自然语言或视觉任务的架构设计,此类架构与神经信号的时空动态相匹配性不足,并且难以捕捉诸如神经群体协同、空间网络拓扑等重要生理特征^[166]。未来的研究可在模型结构中融入神经生理约束、时序动态机制和空间拓扑实验,以提高大脑编解码模型的生物学可解释性。

(4) 评估标准与可复现性不足。尽管基础模型在BCI领域引起了广泛关注,但当前尚缺乏统一、系统的评估标准与评价体系,导致不同方法之间的性能结果难以直接比较。此外,同时,现有研究多在规模受限的数据集上进行验证,对模型在真实复杂环境中的泛化能力关注不足。未来,亟需建立跨受试者、跨数据集和跨任务的标准评估体系,并明确可解释性和可靠性的量化指标,以推动大脑编解码基础模型向更实际的智能系统迈进。

5 结束语

本文系统综述了基础模型驱动下的BCI编解码研究,揭示了该领域由专用模型向通用基础模型范式演进的发展趋势。具体而言,文中深入探讨了基础模型在神经编码中的脑表征建模和多模态表征学习机制,梳理了基于生成式大模型在神经解码任务中实现的语义重构,分析了通过大规模预训练构建面向异构脑数据的统一基础模型。最后,总结了当前领域在模型可解释性、数据规模及跨模态深度融合等方面面临的挑战。同时,凭借强大的泛化性与推理能力的优势,基础模型展现出重塑BCI研究范式的巨大潜力,并有望在理解大脑信息处理机制与推动应用落地方面发挥更重要的作用。

参考文献:

- [1] WOLPAW J R. Brain-computer interfaces[M]//Handbook of Clinical Neurology: vol 110. [S.l.]: Elsevier, 2013: 67-74.
- [2] 邹霞, 李子遇, 李晴, 等. 脑启智跃:人脑与类脑的协同创新范式[J/OL]. 智能系统学报: 1-15[2025-12-24]. <https://link.cnki.net/urlid/23.1538.tp.20251223.1708.004>.
WU Xia, LI Ziyu, LI Qing, et al. Intelligent leap inspired by brain function mechanisms: A collaborative innovation paradigm between human brain and brain-inspired systems [J/OL]. CAAI Transactions on Intelligent Systems: 1-15[2025-12-24]. <https://link.cnki.net/urlid/23.1538.tp.20251223.1708.004>.
- [3] 杨伊静. 发展脑机接口产业教育革新正当其时——解读《关于推动脑机接口产业创新发展的实施意见》[J]. 中国科技产业, 2025(10): 26-27.
- [4] 邱爽, 杨帮华, 陈小刚, 等. 非侵入式脑-机接口编解码技术研究进展[J]. 中国图象图形学报, 2023, 28(6): 1543-1566.
QIU Shuang, YANG Banghua, CHEN Xiaogang, et al. A survey on encoding and decoding technology of non-invasive brain-computer interface[J]. Journal of Image and Graphics, 2023, 28(6): 1543-1566.
- [5] WU J, REN Z, WANG J, et al. Adabrain-bench: Benchmarking brain foundation models for brain-computer interface applications[J]. arXiv preprint arXiv: 250709882, 2025.
- [6] 钱波, 李富江, 郑常乐, 等. 医疗大模型发展现状与展望[J]. 数据采集与处理, 2025, 40(3): 562-584.
QIAN Bo, LI Fujiang, ZHENG Changle, et al. A review of development and future directions of medical foundation models [J]. Journal of Data Acquisition and Processing, 2025, 40(3): 562-584.
- [7] SANKARAN N, LEONARD M K, THEUNISSEN F, et al. Encoding of melody in the human auditory cortex[J]. Science Advances, 2024, 10(7): eadk0010.
- [8] WANDEL T SK, BJÅNES DA, PEJSA K, et al. Representation of internal speech by single neurons in human supramarginal gyrus[J]. Nature Human Behaviour, 2024, 8(6): 1136-1149.
- [9] PANG J C, AQUINO K M, OLDEHINKEL M, et al. Geometric constraints on human brain function[J]. Nature, 2023, 618(7965): 566-574.
- [10] YAN Y, MURPHY T H. Decoding state-dependent cortical-cerebellar cellular functional connectivity in the mouse brain[J]. Cell Reports, 2024, 43(6): 114348.
- [11] MATHIS M W, ROTONDO A P, CHANG E F, et al. Decoding the brain: From neural representations to mechanistic models[J]. Cell, 2024, 187(21): 5814-5832.
- [12] AWAIS M, NASEER M, KHAN S, et al. Foundation models defining a new era in vision: A survey and outlook [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(4): 2245-2264.
- [13] 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. Red Hook, NY, USA: Curran Associates Inc., 2017: 6000-6010.
- [14] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, volume 1 (long and short papers). Minneapolis, Minnesota, USA: Association for Computational Linguistics, 2019: 4171-4186.
- [15] KAPLAN J, MCCANDLISH S, HENIGHAN T, et al. Scaling laws for neural language models[J]. arXiv preprint arXiv: 200108361, 2020.
- [16] DOSOVITSKIY A. An image is worth 16×16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv: 201011929, 2020.
- [17] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]//Proceedings of International Conference on Machine Learning. [S.l.]: PMLR, 2021: 8748-8763.
- [18] FEDUS W, ZOPH B, SHAZEER N. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity[J]. Journal of Machine Learning Research, 2022, 23(120): 1-39.
- [19] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2022: 10684-10695.

- [20] ACHIAM J, ADLER S, AGARWAL S, et al. Gpt-4 technical report[J]. arXiv preprint arXiv: 230308774, 2023.
- [21] BROOKS T, PEEBLES B, HOLMES C, et al. Video generation models as world simulators[J]. OpenAI Blog, 2024, 1(8): 1.
- [22] LIU A, FENG B, XUE B, et al. Deepseek-v3 technical report[J]. arXiv preprint arXiv: 241219437, 2024.
- [23] TEAM Q. Qwen2 technical report[J]. arXiv preprint arXiv: 240710671, 2024, 2(3).
- [24] TEAM G, GEORGIEV P, LEI V I, et al. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context[J]. arXiv preprint arXiv:240305530, 2024.
- [25] REN J, XIA F. Brain-inspired artificial intelligence: A comprehensive review[J]. arXiv preprint arXiv: 240814811, 2024.
- [26] TUCKUTE G, SATHE A, SRIKANT S, et al. Driving and suppressing the human language network using large language models[J]. Nature Human Behaviour, 2024, 8(3): 544-561.
- [27] ALLEN E J, ST-YVES G, WU Y, et al. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence[J]. Nature Neuroscience, 2022, 25(1): 116-126.
- [28] HORIKAWA T, KAMITANI Y. Generic decoding of seen and imagined objects using hierarchical visual features[J]. Nature Communications, 2017, 8(1): 15037.
- [29] SHEN G, DWIVEDI K, MAJIMA K, et al. End-to-end deep image reconstruction from human brain activity[J]. Frontiers in Computational Neuroscience, 2019, 13: 21.
- [30] KAY K N, NASELARIS T, PRENGER R J, et al. Identifying natural images from human brain activity[J]. Nature, 2008, 452(7185): 352-355.
- [31] NASTASE S, LIU Y, HILLMAN H, et al. The “Narratives” fMRI data for evaluating models of naturalistic language comprehension[J]. Scientific Data, 2021, 8(1): 250.
- [32] BHATTASALI S, BRENNAN J, LUH W M, et al. The Alice datasets: fMRI & EEG observations of natural language comprehension[C]//Proceedings of the 12th Language Resources and Evaluation Conference. Marseille, France: European Language Resources Association, 2020: 120-125.
- [33] GIFFORD AT, DWIVEDI K, ROIG G, et al. A large and rich EEG dataset for modeling human visual object recognition[J]. NeuroImage, 2022, 264: 119754.
- [34] 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(1): 1-13.
- [35] STOBER S. Toward studying music cognition with information retrieval techniques: Lessons learned from the OpenMIIR initiative[J]. Frontiers in Psychology, 2017, 8: 1255.
- [36] PALAZZO S, SPAMPINATO C, KAVASIDIS I, et al. Decoding brain representations by multimodal learning of neural activity and visual features[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(11): 3833-3849.
- [37] ZHENG W L, LU B L. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks[J]. IEEE Transactions on Autonomous Mental Development, 2015, 7(3): 162-175.
- [38] KATSIKIANNIS 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 and Health Informatics, 2017, 22(1): 98-107.
- [39] GROOTSWAGERS T, ROBINSON AK, SHATEK SM, et al. Mapping the dynamics of visual feature coding: Insights into perception and integration[J]. PLOS Computational Biology, 2024, 20(1): e1011760.
- [40] GWILLIAMS L, FLICK G, MARANTZ A, et al. Introducing MEG-MASC a high-quality magneto-encephalography dataset for evaluating natural speech processing[J]. Scientific Data, 2023, 10(1): 862.
- [41] HEBART MN, CONTIER O, TEICHMANN L, et al. THINGS-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior[J]. Elife, 2023, 12: e82580.
- [42] WAKEMAN D G, HENSON R N. A multi-subject, multi-modal human neuroimaging dataset[J]. Scientific Data, 2015, 2(1): 1-10.
- [43] WANG Y, WANG S, ZHANG Y, et al. Representation, alignment, and generation: A comprehensive survey of foundation models for non-invasive brain decoding[J]. bioRxiv, 2025: 2025.2011.2030.691403.
- [44] TANG J, LEBEL A, JAIN S, et al. Semantic reconstruction of continuous language from non-invasive brain recordings[J]. Nature Neuroscience, 2023, 26(5): 858-866.

- [45] OOTA S R, CHEN Z, GUPTA M, et al. Deep neural networks and brain alignment: Brain encoding and decoding (survey)[J]. arXiv preprint arXiv: 230710246, 2023.
- [46] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:13013781, 2013.
- [47] SCHRIMPF M, BLANK I A, TUCKUTE G, et al. The neural architecture of language: Integrative modeling converges on predictive processing[J]. *Proceedings of the National Academy of Sciences*, 2021, 118(45): e2105646118.
- [48] FODOR J, MURAWSKI C, SUZUKI S. When word order matters: Human brains represent sentence meaning differently from large language models[J]. *bioRxiv*, 2025: 2025.2007. 2019.665701.
- [49] RAHIMI M, YAGHOOBZADEH Y, DALIRI M R. Explanations of deep language models explain language representations in the brain[J]. arXiv preprint: arXiv: 2502.14671, 2025.
- [50] BHATTACHARJEE A, ZADA Z, WANG H, et al. Aligning brains into a shared space improves their alignment with large language models[J]. *Nature Computational Science*, 2025: 1-10.
- [51] DE VARDA A G, MALIK-MORALEDA S, TUCKUTE G, et al. Multilingual computational models capture a shared meaning component in brain responses across 21 languages[EB/OL]. (2025-02-01). <https://doi.org/10.1101/2025.02.01.636044>.
- [52] HUTH A G, DE HEER W A, GRIFFITHS T L, et al. Natural speech reveals the semantic maps that tile human cerebral cortex[J]. *Nature*, 2016, 532(7600): 453-458.
- [53] CARO J O, FONSECA A H D O, AVERILL C, et al. BrainLM: A foundation model for brain activity recordings[EB/OL]. (2023-09-12). <https://doi.org/10.1101/2023.09.12.557460>.
- [54] WEI X, ZHAO K, JIAO Y, et al. A brain graph foundation model: Pre-training and prompt-tuning for any atlas and disorder [J]. arXiv preprint arXiv: 250602044, 2025.
- [55] DOERIG A, KIETZMANN T C, ALLEN E, et al. High-level visual representations in the human brain are aligned with large language models[J]. *Nature Machine Intelligence*, 2025, 7(8): 1220-1234.
- [56] WU R, ZHOU F, YIN Z, et al. Aligning neuronal coding of dynamic visual scenes with foundation vision models[C]// *Proceedings of European Conference on Computer Vision*. [S.l.]: Springer, 2024: 238-254.
- [57] LU Z, WANG Y, GOLOMB J D. Achieving more human brain-like vision via human eeg representational alignment[J]. ArXiv preprint: arXiv: 2401.17231 v17232, 2024.
- [58] KNEELAND R, OJEDA J, ST-YVES G, et al. Second sight: Using brain-optimized encoding models to align image distributions with human brain activity[J]. ArXiv preprint: arXiv: 2306.00927 v00921, 2023.
- [59] RUPP K M, HECT J L, HARFORD E E, et al. A hierarchy of processing complexity and timescales for natural sounds in the human auditory cortex[J]. *Proceedings of the National Academy of Sciences*, 2025, 122(18): e2412243122.
- [60] CAUCHETEUX C, GRAMFORT A, KING J R. Evidence of a predictive coding hierarchy in the human brain listening to speech[J]. *Nature Human Behaviour*, 2023, 7(3): 430-441.
- [61] BORSOS Z, MARINIER R, VINCENT D, et al. AudioLM: A language modeling approach to audio generation[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023, 31: 2523-2533.
- [62] FERRANTE M, BOCCATO T, RASHKOV G, et al. Towards neural foundation models for vision: Aligning EEG, MEG, and fMRI representations for decoding, encoding, and modality conversion[J]. *Information Fusion*, 2026, 126: 103650.
- [63] LI D, QIN H, WU M, et al. BrainFLORA: Uncovering brain concept representation via multimodal neural embeddings[C]// *Proceedings of the 33rd ACM International Conference on Multimedia*. [S.l.]: ACM, 2025: 5577-5586.
- [64] CHEN S, LI B, HE C, et al. An EEG dataset for multimodal semantic alignment and neural decoding during reading and listening[J]. *Scientific Data*, 2025, 13(1): 148.
- [65] WATANABE N, MIYOSHI K, JIMURA K, et al. Multimodal deep neural decoding reveals highly resolved spatiotemporal profile of visual object representation in humans[J]. *NeuroImage*, 2023, 275: 120164.
- [66] D'ASCOLI S, RAPIN J, BENCHETRIT Y, et al. TRIBE: TRImodal brain encoder for whole-brain fMRI response prediction[J]. arXiv preprint arXiv: 250722229, 2025.
- [67] TANG J, DU M, VO V, et al. Brain encoding models based on multimodal transformers can transfer across language and vision[J]. *Advances in Neural Information Processing Systems*, 2023, 36: 29654-29666.

- [68] CHEN Z, QING J, ZHOU J H. Cinematic mindscapes: High-quality video reconstruction from brain activity[J]. *Advances in Neural Information Processing Systems*, 2023, 36: 24841-24858.
- [69] TAKAGI Y, NISHIMOTO S. High-resolution image reconstruction with latent diffusion models from human brain activity [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 2023: 14453-14463.
- [70] ZHANG Y, GONG K, ZHANG K, et al. Meta-Transformer: A unified framework for multimodal learning[J]. *arXiv preprint arXiv: 230710802*, 2023.
- [71] BELIY R, WASSERMAN N, ZALCHER A, et al. The wisdom of a crowd of brains: A universal brain encoder[J]. *arXiv preprint arXiv: 240612179*, 2024.
- [72] CUI W, JEONG W, THÖLKE P, et al. Neuro-GPT: Towards a foundation model for EEG [C]//*Proceedings of 2024 IEEE International Symposium on Biomedical Imaging (ISBI)*. [S.l.]: IEEE, 2024: 1-5.
- [73] MA Y, FRAUEN D, JAVUREK E, et al. Foundation models for causal inference via prior-data fitted networks[J]. *arXiv preprint arXiv: 250610914*, 2025.
- [74] FITZ H, HAGOORT P, PETERSSON KM. Neurobiological causal models of language processing[J]. *Neurobiology of Language*, 2024, 5(1): 225-247.
- [75] YIN C, YE Z, LI P. Language reconstruction with brain predictive coding from fMRI data[J]. *arXiv preprint arXiv: 240511597*, 2024.
- [76] GAO C, MA Z, CHEN J, et al. Increasing alignment of large language models with language processing in the human brain[J]. *Nature Computational Science*, 2025, 5(11): 1080-1090.
- [77] REN Z, LI J, XUE X, Et al. Reconstructing seen image from brain activity by visually-guided cognitive representation and adversarial learning[J]. *NeuroImage*, 2021, 228: 117602.
- [78] MOZAFARI M, REDDY L, VANRULLEN R. Reconstructing natural scenes from fMRI patterns using bigbigan[C]//*Proceedings of 2020 International Joint Conference on Neural Networks (IJCNN)*. [S.l.]: IEEE, 2020: 1-8.
- [79] OZCELIK F, CHOKSI B, MOZAFARI M, et al. Reconstruction of perceived images from fmri patterns and semantic brain exploration using instance-conditioned GANs[C]//*Proceedings of 2022 International Joint Conference on Neural networks (IJCNN)*. [S.l.]: IEEE, 2022: 1-8.
- [80] FERRANTE M, BOCCATO T, TOSCHI N. Semantic brain decoding: From fMRI to conceptually similar image reconstruction of visual stimuli[J]. *arXiv preprint arXiv: 221206726*, 2022.
- [81] LIN S, SPRAGUE T, SINGH A K. Mind reader: Reconstructing complex images from brain activities[J]. *Advances in Neural Information Processing Systems*, 2022, 35: 29624-29636.
- [82] TAKAGI Y, NISHIMOTO S. Improving visual image reconstruction from human brain activity using latent diffusion models via multiple decoded inputs[J]. *arXiv preprint arXiv: 230611536*, 2023.
- [83] LU Y, DU C, ZHOU Q, et al. Minddiffuser: Controlled image reconstruction from human brain activity with semantic and structural diffusion[C]//*Proceedings of the 31st ACM International Conference on Multimedia*. [S.l.]: ACM, 2023: 5899-5908.
- [84] OZCELIK F, VANRULLEN R. Natural scene reconstruction from fMRI signals using generative latent diffusion[J]. *Scientific Reports*, 2023, 13(1): 15666.
- [85] MENG L, YANG C. Dual-guided brain diffusion model: Natural image reconstruction from human visual stimulus fMRI[J]. *Bioengineering*, 2023, 10(10): 1117.
- [86] CHEN Z, QING J, XIANG T, et al. Seeing beyond the brain: Conditional diffusion model with sparse masked modeling for vision decoding[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 2023: 22710-22720.
- [87] VAN DEN OORD A, VINYALS O. Neural discrete representation learning[C]//*Proceedings of the 31st International Conference Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2017: 6309-6318.
- [88] SUN J, LI M, CHEN Z, et al. Contrast, attend and diffuse to decode high-resolution images from brain activities[J]. *Advances in Neural Information Processing Systems*, 2023, 36: 12332-12348.
- [89] QIAN X, WANG Y, FU Y, et al. Joint fMRI decoding and encoding with latent embedding alignment[J]. *arXiv preprint arXiv:*

230314730, 2023.

- [90] LIU Y, MA Y, ZHOU W, et al. BrainCLIP: Bridging brain and visual-linguistic representation via CLIP for generic natural visual stimulus decoding from fMRI[J]. arXiv preprint arXiv: 230212971, 2023.
- [91] SCOTTI P, BANERJEE A, GOODE J, et al. Reconstructing the mind's eye: fMRI-to-image with contrastive learning and diffusion priors[J]. *Advances in Neural Information Processing Systems*, 2023, 36: 24705-24728.
- [92] SCOTTI P S, TRIPATHY M, VILLANUEVA C K T, et al. MindEye2: Shared-subject models enable fmri-to-image with 1 hour of data[J]. arXiv preprint arXiv: 240311207, 2024.
- [93] MAI W, ZHANG Z. UniBrain: Unify image reconstruction and captioning all in one diffusion model from human brain activity [J]. arXiv preprint arXiv: 230807428, 2023.
- [94] FERRANTE M, OZCELIK F, BOCCATO T, et al. Brain captioning: Decoding human brain activity into images and text[J]. arXiv preprint arXiv: 230511560, 2023.
- [95] XIA W, DE CHARETTE R, OZTIRELI C, et al. UMBRAE: Unified multimodal brain decoding[C]//*Proceedings of European Conference on Computer Vision*. [S.l.]: Springer, 2024: 242-259.
- [96] DAI Y, YAO Z, SONG C, et al. MindAligner: Explicit brain functional alignment for cross-subject visual decoding from limited fMRI data[J]. arXiv preprint arXiv: 250205034, 2025.
- [97] LIU Y, MA Y, ZHU G, et al. See Through their minds: Learning transferable brain decoding models from cross-subject fMRI [C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. [S.l.]: AAAI, 2025: 5730-5738.
- [98] SHEN G, ZHAO D, HE X, et al. Neuro-vision to language: Enhancing brain recording-based visual reconstruction and language interaction[J]. *Advances in Neural Information Processing Systems*, 2024, 37: 98083-98110.
- [99] WANG G, LIU W, HE Y, et al. EEGPT: Pretrained transformer for universal and reliable representation of EEG signals[J]. *Advances in Neural Information Processing Systems*, 2024, 37: 39249-39280.
- [100] SHEN X, TAO L, CHEN X, et al. Contrastive learning of shared spatiotemporal EEG representations across individuals for naturalistic neuroscience[J]. *NeuroImage*, 2024, 301: 120890.
- [101] SHEN X, LIU X, HU X, et al. Contrastive learning of subject-invariant EEG representations for cross-subject emotion recognition[J]. *IEEE Transactions on Affective Computing*, 2022, 14(3): 2496-2511.
- [102] LAI-TAN N, GU X, PHILIASTIDES M G, et al. Cross-subject and cross-montage EEG transfer learning via individual tangent space alignment and spatial-riemannian feature fusion[J]. arXiv preprint arXiv: 250808216, 2025.
- [103] LEE J, CHOI J W, JO S. Selective multi-source domain adaptation network for cross-subject motor imagery discrimination[J]. *IEEE Transactions on Cognitive and Developmental Systems*, 2023, 16(3): 923-934.
- [104] KAVASIDIS I, PALAZZO S, SPAMPINATO C, et al. Brain2Image: Converting brain signals into images[C]//*Proceedings of the 25th ACM International Conference on Multimedia*. [S.l.]: ACM, 2017: 1809-1817.
- [105] PALAZZO S, SPAMPINATO C, KAVASIDIS I, et al. Generative adversarial networks conditioned by brain signals[C]//*Proceedings of the IEEE International Conference on Computer Vision*. [S.l.]: IEEE, 2017: 3410-3418.
- [106] JIAO Z, YOU H, YANG F, et al. Decoding EEG by visual-guided deep neural networks[C]//*Proceedings of the 28th International Joint Conference on Artificial Intelligence*. Macao, China: AAAI, 2019: 1387-1393.
- [107] BAI Y, WANG X, CAO Y P, et al. DreamDiffusion: High-quality EEG-to-image generation with temporal masked signal modeling and CLIP alignment[C]//*Proceedings of European Conference on Computer Vision*. [S.l.]: Springer, 2024: 472-488.
- [108] LI D, WEI C, LI S, et al. Visual decoding and reconstruction via EEG embeddings with guided diffusion[J]. arXiv preprint arXiv: 240307721, 2024.
- [109] CAO X, GONG P, ZHANG L, et al. EEG-CLIP: A transformer-based framework for EEG-guided image generation[J]. *Neural Networks*, 2026, 194: 108167.
- [110] KIM J M, CHOI W, PARK S J, et al. SeeEEG: Semantic-aware EEG-based multi-modal retrieval-augmented generation for high-fidelity visual brain decoding[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. [S.l.]: IEEE, 2025: 4824-4833.
- [111] MEHMOOD T, AHMAD H, SHAKEEL M H, et al. CATVis: Context-aware thought visualization[C]//*Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*. [S.l.]: Springer, 2025: 98-108.

- [112] BENCHETRIT Y, BANVILLE H, KING J R. Brain decoding: Toward real-time reconstruction of visual perception[J]. arXiv preprint arXiv: 231019812, 2023.
- [113] WU H, LI Q, ZHANG C, et al. Bridging the vision-brain gap with an uncertainty-aware blur prior[C]//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Nashville, TN, USA: IEEE, 2025: 2246-2257.
- [114] KUPERSHMIDT G, BELIY R, GAZIV G, et al. A penny for your (visual) thoughts: Self-supervised reconstruction of natural movies from brain activity[J]. arXiv preprint arXiv: 220603544, 2022.
- [115] WANG C, YAN H, HUANG W, et al. Reconstructing rapid natural vision with fMRI-conditional video generative adversarial network[J]. Cerebral Cortex, 2022, 32(20): 4502-4511.
- [116] JING H, JIANG D, MA Y, et al. Beyond brain decoding: Visual-semantic reconstructions to mental creation extension based on fMRI[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2025: 19258-19268.
- [117] LIU X H, LU B L, ZHENG W L. EEGmirror: Leveraging EEG data in the wild via montage-agnostic self-supervision for EEG to video decoding[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2025: 18273-18283.
- [118] GAO J, FU Y, FU Y, et al. MinD-3D++: Advancing fMRI-based 3D reconstruction with high-quality textured mesh generation and a comprehensive dataset[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(12): 11802-11816.
- [119] CHEN J, QI Y, WANG Y, et al. MindGPT: Interpreting what you see with non-invasive brain recordings[J]. IEEE Transactions on Image Processing, 2025, 34: 3281-3293.
- [120] CHEN J, QI Y, WANG Y, et al. Bridging the gap between brain and machine in interpreting visual semantics: Towards self-adaptive brain-to-text decoding[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2025: 21938-21948.
- [121] HUANG W, MA K, XIE T, et al. BrainChat: Interactive semantic information decoding from fMRI using large-scale vision-language pretrained models[C]//Proceedings of ICASSP 2025—2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). [S.l.]: IEEE, 2025: 1-5.
- [122] ZHENG R, SUN L. LLM4Brain: Training a large language model for brain video understanding[J]. arXiv preprint arXiv: 240917987, 2024.
- [123] CHEN X, DU C, LIU C, et al. Open-vocabulary auditory neural decoding using fMRI-prompted LLM[J]. arXiv preprint arXiv: 240507840, 2024.
- [124] FENG F, NIE J. From eye to mind: Brain2TEXT decoding reveals the neural mechanisms of visual semantic processing[J]. arXiv preprint arXiv: 250322697, 2025.
- [125] LIU H, HAJIALIGOL D, ANTONY B, et al. EEG2text: Open vocabulary EEG-to-text decoding with EEG pre-training and multi-view transformer[J]. arXiv preprint arXiv: 240502165, 2024.
- [126] WANG J, SONG Z, MA Z, et al. Enhancing EEG-to-text decoding through transferable representations from pre-trained contrastive EEG-text masked autoencoder[J]. arXiv preprint arXiv: 240217433, 2024.
- [127] DUAN Y, ZHOU J, WANG Z, et al. DeWave: Discrete EEG waves encoding for brain dynamics to text translation[J]. arXiv preprint arXiv: 230914030, 2023.
- [128] LÉVY J, ZHANG M, PINET S, et al. Brain-to-text decoding: A non-invasive approach via typing[J]. arXiv preprint arXiv: 250217480, 2025.
- [129] JING H, MA Y, YANG P, et al. Pinpointing visual content: Disentangled features in multimodal model for EEG representation learning and decoding[J]. Knowledge-Based Systems, 2025, 329: 114212.
- [130] XIONG W, MA L, LI H. Synthesizing intelligible utterances from EEG of imagined speech[J]. Frontiers in Neuroscience, 2025, 19: 1565848.
- [131] DE ZUAZO X, VERBENI V, KU L C, et al. #neural2speech: Decoding speech and language from the human brain[C]//Proceedings of IberSPEECH 2024. Aveiro, Portugal: International Speech Communication Association (ISCA), 2024: 256-260.
- [132] GUO Y, LIU T, ZHANG X, et al. End-to-end translation of human neural activity to speech with a dual-dual generative

- adversarial network[J]. *Knowledge-Based Systems*, 2023, 277: 110837.
- [133] CHEN J, CHEN X, WANG R, et al. Subject-agnostic transformer-based neural speech decoding from surface and depth electrode signals[EB/OL]. (2024-03-11). <https://doi.org/10.1101/2024.03.11.584533>.
- [134] XIA W, OZTIRELI C. Multigranular evaluation for brain visual decoding[J]. *arXiv preprint arXiv: 250707993*, 2025.
- [135] LANG K, FANG J, SU G. NeuroFusionNet: Cross-modal modeling from brain activity to visual understanding[J]. *Frontiers in Computational Neuroscience*, 2025, 19: 1545971.
- [136] TIAN Z, QUAN R, MA F, et al. BrainGuard: Privacy-preserving multisubject image reconstructions from brain activities [C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. [S.l.]: AAAI, 2025: 14414-14422.
- [137] GONG Z, ZHANG Q, BAO G, et al. MindTuner: Cross-subject visual decoding with visual fingerprint and semantic correction[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. [S.l.]: AAAI, 2025: 14247-14255.
- [138] XU Y, LIU B, SHAO W, et al. Cross-subject mind decoding from inaccurate representations[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. [S.l.]: IEEE, 2025: 15066-15075.
- [139] YU S, CHEN Z, WU W, et al. From flat to round: Redefining brain decoding with surface-based fMRI and cortex structure [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. [S.l.]: IEEE, 2025: 6189-6198.
- [140] ZENG B, LI S, LIU X, et al. Controllable mind visual diffusion model[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. [S.l.]: AAAI, 2024: 6935-6943.
- [141] HAN K, WEN H, SHI J, et al. Variational autoencoder: An unsupervised model for encoding and decoding fMRI activity in visual cortex[J]. *NeuroImage*, 2019, 198: 125-136.
- [142] JIANG S, LI H, MENG Z, et al. Mind2Word: Towards generalized visual neural representations for high-quality video reconstruction[J]. *Expert Systems with Applications*, 2026, 303: 129154.
- [143] LE L, AMBROGIONI L, SEELIGER K, et al. Brain2pix: Fully convolutional naturalistic video frame reconstruction from brain activity[J]. *Frontiers in Neuroscience*, 2022, 16: 940972.
- [144] FERRANTE M, BOCCATO T, PASSAMONTI L, et al. Retrieving and reconstructing conceptually similar images from fMRI with latent diffusion models and a neuro-inspired brain decoding model[J]. *Journal of Neural Engineering*, 2024, 21(4): 046001.
- [145] ZENG H, XIA N, QIAN D, et al. DM-RE2I: A framework based on diffusion model for the reconstruction from EEG to image[J]. *Biomedical Signal Processing and Control*, 2023, 86: 105125.
- [146] LOPEZ E, SIGILLO L, COLONNESE F, et al. Guess what I think: Streamlined EEG-to-image generation with latent diffusion models[C]//*Proceedings of ICASSP 2025—2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.]: IEEE, 2025: 1-5.
- [147] FERRANTE M, BOCCATO T, BARGIONE S, et al. Decoding visual brain representations from electroencephalography through knowledge distillation and latent diffusion models[J]. *Computers in Biology and Medicine*, 2024, 178: 108701.
- [148] GOLDSTEIN A, WANG H, NIEKERKEN L, et al. Deep speech-to-text models capture the neural basis of spontaneous speech in everyday conversations[EB/OL]. (2023-06-26). <https://doi.org/10.1101/2023.06.26.546557>.
- [149] DÉFOSSEZ A, CAUCHETEUX C, RAPIN J, et al. Decoding speech perception from non-invasive brain recordings[J]. *Nature Machine Intelligence*, 2023, 5(10): 1097-1107.
- [150] FU H, WANG H, CHIN J J, et al. BrainVis: Exploring the bridge between brain and visual signals via image reconstruction [C]//*Proceedings of ICASSP 2025—2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.]: IEEE, 2025: 1-5.
- [151] GIDEONI Y, TIMMS R C, JONES O P. Non-invasive neural decoding in source reconstructed brain space[J]. *arXiv preprint arXiv: 241019838*, 2024.
- [152] PESCATORE C R, ZHANG H, HADJINICOLAOU A E, et al. Decoding semantics from natural speech using human intracranial EEG[EB/OL]. (2025-02-10). <https://doi.org/10.1101/2025.02.10.637051>.
- [153] LI Z, ZHU Z, BAI Y, et al. Spatio-temporal pre-trained foundation model for neural decoding with fine-grained optimization [C]//*Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*. [S.l.]: Springer, 2025: 609-618.

- [154] LI Z, ZHU Z, LI Q, et al. Graph pre-trained framework with spatio-temporal importance masking and fine-grained optimizing for neural decoding[J]. Pattern Recognition, 2026, 170: 112006.
- [155] YANG Y, YE C, SU G, et al. BrainMass: Advancing brain network analysis for diagnosis with large-scale self-supervised learning[J]. IEEE Transactions on Medical Imaging, 2024, 43(11): 4004-4016.
- [156] DONG Z, LI R, WU Y, et al. Brain-JEPA: Brain dynamics foundation model with gradient positioning and spatiotemporal masking[J]. Advances in Neural Information Processing Systems, 2024, 37: 86048-86073.
- [157] WANG M, XIA J, YE W, et al. SLIM-Brain: A data-and training-efficient foundation model for fMRI data analysis[J]. arXiv preprint arXiv: 251221881, 2025.
- [158] FAN W, RIZKY LR, ZHANG J, et al. Foundation-model-boosted multimodal learning for fMRI-based neuropathic pain drug response prediction[C]//Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. [S.l.]: Springer, 2025: 238-248.
- [159] YUAN Z, SHEN F, LI M, et al. BrainWave: A brain signal foundation model for clinical applications[J]. arXiv preprint arXiv: 240210251, 2024.
- [160] JIANG W B, ZHAO L M, LU B L. Large brain model for learning generic representations with tremendous EEG data in BCI [J]. arXiv preprint arXiv: 240518765, 2024.
- [161] WANG J, ZHAO S, LUO Z, et al. CBraMod: A criss-cross brain foundation model for EEG decoding[J]. arXiv preprint arXiv: 241207236, 2024.
- [162] OUAHIDI YE, LYS J, THÖLKE P, et al. REVE: A foundation model for EEG—adapting to any setup with large-scale pretraining on 25,000 subjects[J]. arXiv preprint arXiv: 251021585, 2025.
- [163] XIAO Q, CUI Z, ZHANG C, et al. BrainOmni: A brain foundation model for unified EEG and MEG signals[J]. arXiv preprint arXiv: 250518185, 2025.
- [164] JUNG E, AN J. EFRM: A multimodal EEG-fNIRS representation-learning model for few-shot brain-signal classification[J]. Computers in Biology and Medicine, 2025, 199: 111292.
- [165] WANG Y, LIU H, WANG Y, et al. Progress, challenges and future of linguistic neural decoding with deep learning[J]. Communications Biology, 2025, 8(1): 1350.
- [166] KWON J, SHIN Y. Foundation models for neural signal decoding: EEG-centered perspectives toward unified representations [J]. European Journal of Neuroscience, 2026, 63(1): e70376.

作者简介:



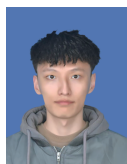
邬霞(1978-),通信作者,女,教授,研究方向:人工智能与脑科学,E-mail: wuxia@bit.edu.cn。



李同同(1996-),男,博士研究生,研究方向:计算机视觉、脑信号分析、医学信息处理,E-mail: ttli2022@lzu.edu.cn。



李子遇(1996-),女,助理研究员,研究方向:脑信号智能分析与类脑智能。



马晓强(1998-),男,博士研究生,研究方向:脑影像诊断与建模研究。



李锦科(2002-),男,硕士研究生,研究方向:脑机智能、大脑编解码与医学图像分析。



李晴(1993-),女,副研究员,研究方向:脑机接口与类脑智能。



姚志军(1977-),男,教授,研究方向:医学影像分析、普适情感计算、数据挖掘、机器学习。

(编辑:王静)