FCN-LLM：通过图级多任务指令微调赋能大语言模型理解大脑功能连接网络

Xingcan Hu中国科学技术大学
xingcanhu@mail.ustc.edu.cn

Wei Wang
中国科学技术大学
wangweiii@mail.ustc.edu.cn

Li Xiao
中国科学技术大学 xiaoli11@ustc.edu.cn

arXiv:2603.01135v1 [cs.AI] 1 Mar 2026

摘要

大语言模型（LLM）在语言理解和推理方面取得了显著成功，其多模态扩展使其能够理解图像、视频和音频。受此启发，基于静息态功能磁共振成像的脑功能连接网络（FCN）基础模型在临床任务中展现出应用潜力。然而，现有方法未能将功能连接网络与文本模态对齐，限制了大语言模型直接理解功能连接网络的能力。为此，我们提出 FCN-LLM，这是一种通过图级多任务指令微调使大语言模型能够理解功能连接网络的框架。我们采用多尺度功能连接网络编码器，提取脑区、功能子网络和全脑特征，并将其映射到大语言模型的语义空间中。我们设计了多范式指令任务，涵盖人口统计学特征、表型和精神疾病方面的 19 项受试者特异性属性。多阶段学习策略首先使功能连接网络嵌入与大语言模型对齐，然后对整个模型进行联合微调，以捕捉高层语义信息。在大规模多中心功能连接网络数据库上的实验表明，FCN-LLM 在未见数据集上实现了出色的零样本泛化，优于传统监督学习模型和基础模型。本研究提出了一种将脑功能网络与大语言模型相结合的新范式，为神经科学提供了灵活且可解释的框架。

1 引言

近年来，大语言模型（LLM）的快速发展，使其在理解语言（Devlin等人，2019；Chen等人，2024a）、生成连贯对话（Floridi & Chiriatti，2020；Achiam等人，2023）以及进行复杂的文本推理（Guo等人，2025）方面取得了显著突破。然而，依赖纯文本输入使其在感知和推理人类自然处理的丰富多样的信息方面存在根本性局限。为克服这些局限，研究界致力于将图像、视频和语音等其他模态对齐至大语言模型的语义空间；这类模型被称为大型多模态模型（LMM）（Liu等人，2023；Li等人，2023；Lin等人，2023；Boson AI，2025）。通过将其他模态与文本对齐，大型多模态模型能够提取、理解并推理非文本数据中的信息，同时以文本形式生成回复（Han等人，2023；Wang等人，2024c）。大语言模型的此类多模态扩展在众多下游任务中实现了强大的泛化能力和最先进的性能，包括视觉问答（Antol等人，2015年；Marino等人，2019年）、图像描述（Lin等人，2014）、视频推理（Fu等人，2025）、音频转录（Panayotov等人，2015年）和语音基础对话（Wang等人，2024a）。总而言之，大型多模态模型主要以文本作为输入和输出，为人工智能提供了一种强大的新范式，具有卓越的灵活性、直观的可解释性以及与人类自然交互的能力。

与此同时，大语言模型的成功得益于大规模预训练和可扩展的基于Transformer的模型架构（Vaswani等人，2017），这启发神经科学界探索适用于各种神经影像数据模态的类似基础模型（Yang等人，2024b；Hu等人，2025；Wei等人，2025）。其目标是学习稳健且具有泛化能力的表征，并将其迁移至多种下游任务，这对临床转化至关重要。在基于静息态功能磁共振成像（rs-fMRI）得出的脑功能连接网络（FCN）领域，为理解脑功能（Power等人，2011）并支持疾病诊断（Yang等人，2021；Wang等人，2024d）和结局预测（Finn等人，2015年；He等人，2024）等临床应用，研究人员提出了新兴的基础模型；近期工作已超越基于图的有监督深度学习方法（Li等人，2021；Kawahara等人，2017；Kan等人，2022b）。值得注意的是，已有多项研究（Hu等人，2024；Yang等人，2024b；Wei等人，2025）开始汇集大规模多站点功能连接网络数据集，并采用不同的预训练策略构建基础模型，以提升性能。这些基础模型通过从功能连接网络中提取可迁移特征，并在多种下游临床任务中取得优异表现，展现出良好的泛化能力。

尽管已有这些进展，一个关键缺口仍然存在：现有方法未能将功能连接网络（FCN）与文本对齐。这阻碍了多模态大语言模型（MLLM）——它们已牢固确立语义空间作为多模态理解的通用基础——直接解读功能连接网络中丰富的大脑相关信息。一个关键障碍是缺乏面向功能连接网络的系统性语言框架：这与视觉领域形成鲜明对比，在该领域中，图像可自然地与文本配对，用于对比预训练（Radford等人，2021年）或描述生成（Li等人，2022年；Yu等人，2022年），并阻碍了将已验证的多模态对齐策略迁移到功能连接网络——尽管功能连接网络本身包含足够的文本相关表征，足以支持对通用功能连接网络嵌入进行基于语言的多尺度解码。加剧这一局限的是，当前的功能连接网络基础模型仍依赖任务特定微调（使用下游数据），或针对每项任务使用单独的分类器（例如，线性支持向量机（SVM））。因此，它们无法实现真正的零样本泛化，因为它们不能借助多模态大语言模型的语义空间以及功能连接网络内部潜在的文本兼容表征来利用跨模态知识迁移。

为探索功能连接网络在多大程度上能够与文本模态对齐，并回答“大语言模型能否理解功能连接网络”这一引人深思的问题，我们提出了 FCN-LLM，这是一种通过图级多任务指令微调赋能大语言模型理解功能连接网络的框架。具体而言，首先，我们设计了**多尺度功能连接网络编码器**，从功能连接网络的三个互补层级——大脑感兴趣区域（ROI）层级、功能子网络层级（Yeo 等人，2011年）和全脑层级——提取特征，并通过多层感知机（MLP）将其映射到大语言模型的语义空间。其次，鉴于难以用自然语言描述功能连接网络，我们从图级预训练（Hu et al.，2019年）的角度出发，开发了**用于指令微调的多范式任务设计**，要求模型预测与功能连接网络对应受试者的领域特定属性。我们总共整理了涵盖人口统计学特征、表型和精神疾病的19项属性，并围绕预测、判断和比较三种范式合成大量指令-回答对，以指导模型训练。第三，我们提出了**功能连接网络知识的多阶段学习**：第一阶段仅训练多尺度功能连接网络编码器，使功能连接网络嵌入与大语言模型的语义空间对齐；第二阶段则对整个模型进行联合微调，以捕捉功能连接网络的高层抽象语义信息。

我们收集了一个大规模、多中心功能连接网络数据库，用于训练 FCN-LLM，并将其与传统有监督功能连接网络模型以及近期开发的功能连接网络基础模型进行了全面比较。实验结果表明，尽管 FCN-LLM 在域内测试集上的表现略逊于有监督功能连接网络模型，但以零样本方式在未见数据集上测试时，它取得了最先进的泛化性能。这种出色的泛化能力归功于两个关键因素：图级多任务指令微调赋予 FCN 特征出色的泛化属性；大语言模型强大的语言理解与生成能力则带来了内在的灵活性。我们还通过提示条件注意力图提供可解释性，并突出显示子网络级功能连接网络连接，将其作为潜在的脑生物标志物，如附录A.7所示。

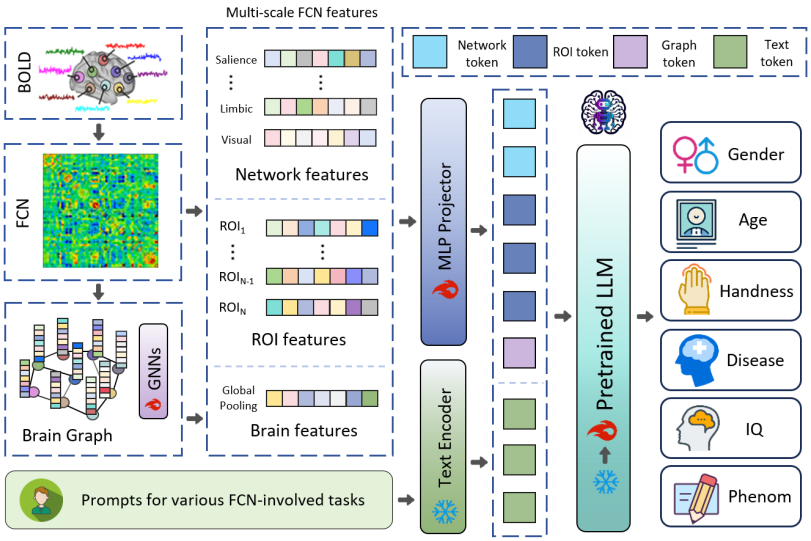


图1：我们提出的FCN-LLM的模型架构，支持与各类涉及功能连接网络的任务相关的查询。它采用多尺度编码器提取功能连接网络特征，并将这些特征映射到文本嵌入空间。

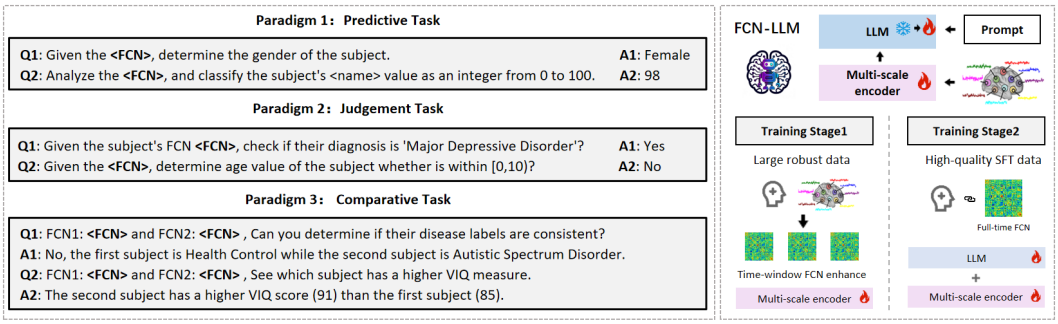
2 相关工作

2.1 面向脑功能连接网络分析的基础模型

通过自监督学习在大规模数据集上预训练的基础模型，在广泛的下游任务中展现出卓越的迁移能力。针对功能连接网络的基础模型，研究者已探索了多种预训练策略。例如，BrainNPT 提出了替换 ROI 预测策略，并在下游应用中采用 [CLS] 标记嵌入（Hu 等，2024年）。BrainMass 采用掩码 ROI 建模和潜在表示对齐来学习稳健的表征，随后通过线性探测进行评估（Yang 等，2024b）。CINP 对功能连接网络与结构磁共振成像进行对比预训练，以构建联合语义空间（Hu 等，2025）。BrainGFM 利用图对比学习和图掩码自编码器，在包含不同脑区划分的多样化脑图谱上进行预训练，从而提升模型在异质的基于功能磁共振成像的脑表征上的泛化能力和适应性（Wei 等，2025）。尽管这些功能连接网络基础模型已在多个下游任务中展现出强大的迁移能力，但仍存在两个关键局限：与大语言模型语义空间缺乏对齐，以及依赖任务特定微调才能有效部署。

2.2 新模态与大语言模型的对齐

近期研究在将新模态（如图像和视频）与大语言模型对齐方面取得了显著进展。一项重要研究方向是引入投影模块，将视觉特征映射到大语言模型的输入空间。例如，BLIP-2 和 MiniGPT-4 使用 Q-Former，通过交叉注意力机制与图像特征交互，再将其投影到大语言模型的输入空间（Li等人，2023；Zhu等人，2023）。LLaVA 则采用轻量且有效的多层感知机，将 CLIP 视觉编码器提取的特征映射为视觉标记（ Liu 等人，2023 ）。另一项研究方向聚焦于多模态指令微调和统一嵌入。例如，ImageBind-LLM（Han 等人，2023）仅使用“绑定网络”和门控层，通过图文对齐进行训练，却能借助 ImageBind 的嵌入无缝泛化至视频、音频和 3D 模态。Video-LLaVA（Lin，2023）进一步推动了这一方向，通过学习图像和视频的统一表征，使大语言模型能够在共享语义空间中处理这两种模态。受基于投影的对齐模块高效性的启发，本文采用多层感知机将功能连接网络特征映射到大语言模型的输入空间，以增强 FCN-LLM 的多模态对齐能力。



Multi-paradigm task design　　Multi-stage learning

图2：所提出的指令微调多范式任务设计（左）以及所提出 FCN-LLM 的 FCN知识学习多阶段学习过程示意图（右）。

2.3 多任务提示微调

多任务提示微调已成为一种将大语言模型适配到众多下游任务的有效方案。它通过在相关任务间迁移有用模式来实现知识共享，从而提升效率并减少冗余（Asai 等人，2022年；Wang等人，2023）。它还能增强泛化能力，因为接触多样化任务有助于形成更鲁棒、更通用的表征，使模型更好地适应此前未见过的任务（Shen 等人，2024年）。此外，多任务提示微调具有灵活性，因为只需组合或扩展轻量级提示，无需重新训练整个模型，便可支持新任务（Wang等人，2023）。最后，多任务提示微调在少样本和零样本设置下尤其有效；与单任务提示微调相比，共享提示能够稳定学习并提升性能（Sanh 等人，2021年；Asai 等人，2022年；Shen 等人，2024年）。鉴于这些优势，本文采用多任务提示微调来构建多样化的指令微调数据，并将功能连接网络知识有效迁移到大语言模型，从而提升泛化能力和鲁棒性。

3 方法

本节组织如下。**（1）我们介绍 FCN-LLM 的架构**（如图1所示），该架构由两个核心组件组成：a）多尺度功能连接网络编码器；b）预训练大语言模型。具体而言，首先通过多尺度功能连接网络编码器将功能连接网络转换为具体的特征嵌入（详见第 3.1 节）；随后，MLP 将这些嵌入投影并对齐到大语言模型的语义空间中，最后由大语言模型从功能连接网络中提取信息并生成文本，以完成下游任务。根据多种大型多模态模型的成功经验（Liu等人，2023；Li等人，2023；Zhu等人，2023），FCN-LLM 中采用的基于投影的对齐方法不仅直观，而且高效。**（2）我们在第 3.2 节提出了一种用于指令微调的多范式任务设计**，如图2左侧所示，该设计包含三种范式：预测、判断和比较。**（3）我们在第 3.3 节构建了一个用于 FCN知识学习的两阶段框架**，如图2右侧所示，该框架包括用于实现稳健的功能连接网络知识对齐的预训练阶段，以及用于高层 FCN 知识精炼的微调阶段。

3.1 多尺度功能连接网络编码器

脑区层级。由静息态功能磁共振成像导出的功能连接网络通常以矩阵表示（以下简称功能连接网络矩阵），其元素表示成对感兴趣区的血氧水平依赖（BOLD）时间序列之间的依赖关系，并以功能连接量化。我们将功能连接网络矩阵的行视为感兴趣区域级别的特征向量，每个向量反映了局部尺度上一个感兴趣区域与其他区域之间的功能连接模式。设 D 为感兴趣区的数量，则感兴趣区域级别的特征向量可表示为$\mathbf{f}_{\text{roi}} \in \mathbb{R}^D$。

功能子网络层级。研究人员通过对大量静息态功能性磁共振成像数据进行聚类分析，将人类大脑皮层系统地划分为七个不同的功能子网络（Yeo等人，2011年）。大脑的子网络划分将功能上密切相关的感兴趣区域整合在一起

。大量研究通过将这些功能子网络与精神障碍、表型和行为相关联，证实了这种功能分区的有效性和必要性（Ereira等人，2024年；Sheffield等人，2015年；Menon，2023）。为充分利用这一先验知识，我们对每个功能子网络内各感兴趣区域的ROI级特征取平均值，以获得子网络级特征向量。

令 $\mathcal{S}_k$ 表示属于第 k 个子网络（$k = 1, 2, \ldots\ldots, N$）的感兴趣区索引集合，其中 N是子网络的数量，$\mathbf{f}_{\text{roi},\, i}$ 表示第 i 个感兴趣区的感兴趣区级特征向量。第 k 个子网络的特征向量计算如下:

$$\mathbf{f}_{\text{sub},k} = \frac{1}{|\mathcal{S}_k|} \sum_{i \in \mathcal{S}_k} \mathbf{f}_{\text{roi},i}. \tag{1}$$

其中，$|\mathcal{S}_k|$ 表示 $\mathcal{S}_k$ 的基数，即第 k 个子网络中感兴趣区的数量。

全脑级。尽管可以在细粒度的感兴趣区和子网络层面解码功能连接网络信息，但同样至关重要的是，利用脑区之间的先验连接，学习能够捕捉全脑组织的表征，因为大尺度网络交互在塑造脑功能方面发挥着核心作用（Bullmore与Sporns，2009）。为此，我们首先对功能连接网络应用绝对值变换和阈值化，以获得邻接矩阵 $\mathbf{A} \in \mathbb{R}^{D \times D}$，将大脑建模为图结构数据，其中节点为感兴趣区，边由A表示，节点特征则是相应感兴趣区的感兴趣区级特征向量，即节点特征矩阵为功能连接网络矩阵。接下来，我们采用两层图卷积网络（GCN）对脑图数据进行平滑处理，然后通过平均池化获得全局尺度上的全脑级特征向量。GCN 的传播规则定义如下:

$$\mathbf{H}^{(l+1)} = \hat{\mathbf{A}}\mathbf{H}^{(l)}\mathbf{W}^{(l)} + \mathbf{b}^{(l)}. \tag{2}$$

其中，$\mathbf{H}^{(l)}$ 表示第 l层的特征矩阵（其中$\mathbf{H}^{(0)}$ 为功能连接网络矩阵），$\hat{\mathbf{A}} = \mathbf{D}^{-1/2}(\mathbf{A} + \mathbf{I})\mathbf{D}^{-1/2}$（其中$\mathbf{I}$ 为单位矩阵，$\mathbf{D}$ 为与$\mathbf{A}$关联的度矩阵），$\mathbf{W}^{(l)}$ 表示第 l层的权重矩阵，$\mathbf{b}^{(l)}$ 为偏置项。

在图卷积网络的第二层中，我们得到经过细化的节点特征矩阵$\mathbf{H}^{(2)} \in \mathbb{R}^{D \times D}$（此处节点特征的输出维度与其原始输入维度一致）。随后，对$\mathbf{H}^{(2)}$中的所有节点特征取平均，计算得到全脑级特征向量，即:

$$\mathbf{f}_{\text{global}} = \frac{1}{D} \sum_{i=1}^{D} \mathbf{h}_i^{(2)}. \tag{3}$$

其中，$\mathbf{h}_i^{(2)}$ 是$\mathbf{H}^{(2)}$的第 i行。

如上所述，我们获得了功能连接网络的层次特征，从而充分打通了功能连接网络信息与大语言模型之间的联系。我们进一步将这些特征拼接为一个统一特征向量，即$\mathbf{f}_{\text{concat}} = [\mathbf{f}_{\text{roi},1}; \mathbf{f}_{\text{roi},2}; \cdots, \mathbf{f}_{\text{roi},D}; \mathbf{f}_{\text{sub},1}; \mathbf{f}_{\text{sub},2}; \cdots, \mathbf{f}_{\text{sub},N}; \mathbf{f}_{\text{global}}] \in \mathbb{R}^{D \times (D+N+1)}$。我们使用多层感知机将$\mathbf{f}_{\text{concat}}$ 映射到大语言模型的输入空间，为后续功能连接网络与文本模态之间的对齐奠定基础。映射过程表示为

$$\mathbf{f}_{\text{aligned}} = \text{MLP}(\mathbf{f}_{\text{concat}}). \tag{4}$$

3.2 指令微调的多范式任务设计

由于多任务提示微调能够促进任务间的知识共享，从而提高学习效率、鲁棒性和泛化性能（Sanh 等人，2021年；Asai 等人，2022年；Shen 等人，2024年），因此为设计多样化的指令微调数据提供了重要动机。为此，我们在图级别构建了三种范式的任务——预测、判断和比较，如图1所示。通过数据合成，这些多范式任务使大语言模型能够高效地获取功能连接网络中蕴含的知识。下文将分别详细介绍这三种范式。

预测范式任务。预测范式要求模型根据输入的功能连接网络及其关联提示，直接预测相应的属性值。本文中，

我们纳入了 19 个属性（详见附录 A.2），并将其分为四类指标：（1）人口统计学指标（性别、年龄、利手）；（2）临床指标（精神疾病状态/诊断）；（3）认知功能指标（FIQ、VIQ、PIQ）；（4）表型指标（人格特质、行为测量和情绪状态）。

判断范式任务。 不同于直接生成答案的预测范式任务，判断范式任务要求模型评估输入功能连接网络与给定候选答案之间的一致性，并据此输出二元决策。具体而言，对于具有有限离散类别集合的属性（即分类属性），模型需要判断给定的候选标签是否与输入功能连接网络相匹配。对于连续属性，模型需要评估真实值是否落在预先定义的区间内（例如，年龄组定义为幼儿期、青春期、成年早期、成年中期和成年晚期；智商范围按每10分为一个区间划分）。该范式可视为对预测任务进行的更粗粒度二分类，其正负样本分布均衡。

比较范式任务。 尽管上述两项任务能够帮助模型学习被试特异性表征，但它们无法捕捉同一任务内的被试间关联。对比学习已被广泛认为是获取功能连接网络稳健且具有判别性表征的有效方法（Wang等人，2022a；2024d），而通过自我博弈将其扩展至大语言模型，进一步展示了其增强对齐与推理能力的潜力（Chen等人，2024b；Wang等人，2024b）。然而，现有方法通常依赖特定于模型的架构设计，并非结构无关。为解决这一局限，我们引入一种将对比学习融入指令微调的比较范式。对于分类属性，模型接收两名被试的功能连接网络，并判断其属性表征是否匹配。对于连续属性，模型比较两名被试的功能连接网络，以确定在预先定义的范围内哪一名被试具有更高的属性值。正负样本对由与预测任务相关的被试组生成，并保持比例均衡。该范式捕捉被试间的群体关系，以数据驱动的方式嵌入对比学习，同时补充上述针对单个被试的预测范式和判断范式。

3.3 功能连接网络知识的多阶段学习

阶段 I：预训练以实现稳健的功能连接网络知识对齐。本阶段利用第 3.2 节中设计的多范式指令微调数据，将功能连接网络的多尺度嵌入标记与大语言模型的文本空间对齐。在训练过程中，我们冻结预训练大语言模型，仅更新多尺度功能连接网络编码器，确保对齐过程专注于连接功能连接网络与文本特征空间，而不会干扰大语言模型固有的语言能力。

值得注意的是，由于大语言模型此前未接触过功能连接网络特征数据，因此需要足够数量的功能连接网络相关任务来促进有效对齐。对于单个受试者的 BOLD时间序列，我们可以通过滑动窗口方法生成多个特定时间窗口的功能连接网络（FCN$_t$），以进行数据增强。具体而言，对于每个感兴趣区域，其长度为 T 的 BOLD时间序列可通过长度为 L 、步长为 P 的滑动窗口划分为 N_{aug} 个片段，其中

$$N_{\text{aug}} = \left\lfloor \frac{T - L}{P} \right\rfloor + 1. \tag{5}$$

其中，$\lfloor \cdot \rfloor$ 表示下取整函数。

第二阶段：面向高层 FCN 知识精炼的微调。 在初始预训练阶段之后，FCN-LLM 获得了识别功能连接网络并捕获基础功能连接网络相关知识的基本能力。然而，仍存在两个关键局限。首先，仅在多尺度功能连接网络编码器上进行训练，限制了模型对功能连接网络的全面理解。其次，大量使用通过滑动窗口生成的增强 FCN 的指令微调数据会引入固有噪声。为克服这些局限，微调阶段采用了两项关键策略。（1）训练在使用原始 FCN 而非增强 FCN 的高质量指令微调数据上进行，同时保留相同任务。（2）大语言模型与多尺度功能连接网络编码器均被解冻并联合训练，使大语言模型能够更深入地整合 FCN 特征，并允许编码器基于高质量数据精炼其表征。这些设计共同增强了模型对功能连接网络知识的高层理解。

表1：本研究中10个数据集的基本信息和使用情况。

用途	Name	Size	可用属性
调优与内部测试	HBN（Alexander等人，2017）	2254	性别、年龄和利手
	HCP（Van Essen等人，2013）	1080	性别、年龄、利手和表型
	QTIM（Strike等人，2023）	1024	性别、年龄和利手
	GSP（Holmes 等人，2015年）	1569	性别、年龄和利手
	ABIDE（Di Martino 等人，2014）	855	性别、年龄、利手、认知和诊断
	注意缺陷多动障碍（联盟，2012年）	872	性别、年龄、利手、认知和诊断
	MDD（Yan 等人，2019年）	2379	性别、年龄、利手和诊断
	SRPBS（Tanaka 等人，2021年）	1397	性别、年龄、利手和诊断
零样本测试	ABIDE II（Di Martino 等人，2014）	1044	性别、年龄、利手、认知和诊断
	CNP（Bilder 等，2020）	261	性别、年龄和诊断

4 实验

4.1 实验设置

数据集。 我们下载了若干公开可用的静息态功能磁共振成像数据集，包括 HBN（Alexander等人，2017）、HCP（Van Essen等人，2013年）、QTIM（Strike等人，2023）、GSP（Holmes 等人，2015年）、ABIDE（Di Martino 等人，2014）、注意缺陷多动障碍（联盟，2012年）、MDD（Yan 等人，2019年）、SRPBS（Tanaka 等人，2021年）、ABIDE II（Di Martino 等人，2014）、CNP（Bilder等人，2020年）。我们在表1中列出了这些数据集的基本信息与用途。简而言之，为构建第 3.2 节所述多样且信息丰富的指令微调任务，以促进大语言模型学习功能连接网络知识，我们将前八个数据集合并，并按 80/20 的比例划分为训练集和测试集。其余两个数据集全部保留用于零样本评估，以模拟模型遇到先前未见数据分布的真实场景。这些数据集共包含 19 个属性，包括年龄、性别、利手、精神疾病诊断、总体智力能力（FIQ）、言语智力能力（VIQ）、操作智力（PIQ），以及十二项特定表型指标（详见附录 A.2），如表1所总结。

数据合成。 针对每种属性，我们在不同范式下生成了 50 条指令提示。从中随机选取一条，并将对应的属性值作为回答，从而构成用于指令微调的问答对。在对齐学习的第一阶段，我们生成了超过 230 万条指令微调数据对；在第二阶段的高层知识学习中，则生成了约 42.6 万条数据对。数据预处理以及各范式下指令微调数据对的确切数量详见附录 A.3 和 A.4。

实现细节。 我们使用 AAL图谱将人脑划分为 $D = 116$个结构感兴趣区。我们遵循（Yeo 等人，2011年）定义了 $N = 7$ 个功能子网络，并根据（Power 等人，2011年）将 AAL模板中的感兴趣区分配到这些子网络中。值得注意的是，并非所有感兴趣区都映射到了预定义的功能子网络。一个两层 GCN 提取全局脑级特征：其邻接矩阵 $\mathbf{A}$ 由功能连接网络中的连接强度取绝对值后得到，并通过 0.5 的阈值滤除弱连接（Wang等人，2024d）；GCN 的隐藏层和输出层维度分别为 256 和 116，并对输出进行平均池化，以获得单个全局特征向量。最后，多尺度编码器将功能连接网络转换为 124 个 FCN令牌（即 $D + N + 1$），并用其替换提示中的特殊占位符⟨FCN⟩，从而构成最终的 FCN-LLM 输入。此外，我们采用 Yang 等人（2024a）的 Qwen2.5-3B 和 7B 作为 FCN-LLM 的预训练大语言模型。有关超参数设置、训练流程和数据集的更多详情，请参阅附录 A.4。

评估指标。 我们在五项任务上评估了该框架：性别分类、疾病分类、年龄预测、认知功能预测和表型预测。具体而言，性别分类是一项二分类任务，用于区分男性和女性受试者。对于疾病分类，由于数据集包含九种疾病（孤独症谱系障碍 (ASD)、注意缺陷多动障碍 (ADHD)、重度抑郁障碍 (MDD)、精神分裂症 (SZ)、强迫症 (OCD) 等），我们根据样本分布将受试者划分为六类：健康对照 (HC)、ASD、ADHD、MDD、SZ，以及由其他精神疾病组成的合并组

表2：三类方法在内部测试集上的性能。

Type	方法	性别分类			疾病分类			年龄预测		认知预测		表型预测	
		准确率	马修斯相关系数	F1	准确率	马修斯相关系数	F1	平均绝对误差	皮尔逊相关系数	平均绝对误差	皮尔逊相关系数	MAE	PCC
监督学习	GCN	65.59	30.14	65.53	66.86	50.26	63.37	6.197	73.13	0.097	79.11	0.146	4.75
	HGCN	63.50	27.97	59.62	66.97	50.73	63.89	5.892	74.91	0.095	78.73	0.144	5.56
	BrainNetCNN	64.28	29.16	62.61	65.83	49.73	64.87	6.185	76.10	0.12	72.72	0.167	-2.16
	BrainGNN	63.86	28.40	63.89	68.11	52.99	66.43	6.062	73.67	0.091	81.42	0.144	3.57
	Transformer	69.60	38.06	69.44	68.23	52.30	65.33	5.041	80.42	0.091	79.50	0.144	0.98
	BNT	71.91	42.80	71.77	71.43	58.48	71.52	4.621	83.79	0.090	80.75	0.143	3.28
基础	BrainNPT	64.71	28.34	63.62	68.10	52.37	64.53	7.570	57.47	0.165	50.23	0.154	1.48
	PTGB	62.29	24.08	61.07	67.09	53.14	65.20	8.285	45.62	0.167	51.64	0.162	4.25
	CINP	63.01	23.78	61.82	67.60	51.90	64.74	6.941	54.24	0.164	53.30	0.148	6.68
	BrainMass	67.28	33.22	67.03	69.26	55.37	63.98	7.084	62.76	0.153	74.27	0.145	5.93
Ours	FCN-LLM (3B)	65.48	29.82	65.39	70.97	57.41	70.78	6.293	75.81	0.089	76.77	0.173	6.95
	FCN-LLM (7B)	66.41	32.21	66.38	69.86	54.44	68.86	6.178	76.38	0.090	77.61	0.138	6.49

表3：三类方法在零样本测试集上的性能。

Type	方法	ABIDEII					CNP		
		疾病分类			认知预测		疾病分类		
		准确率	马修斯相关系数	F1	MAE	PCC	准确率	马修斯相关系数	F1
监督学习	Transformer	22.77	3.59	26.73	0.154	-3.00	36.74	0.16	29.78
	BNT	31.82	6.19	40.16	0.158	10.66	44.44	2.88	30.26
基础	BrainNPT	31.56	2.21	27.59	0.147	4.32	40.71	5.23	31.43
	PTGB	32.14	3.28	29.53	0.145	3.79	38.91	1.24	28.87
	CINP	32.66	1.34	30.67	0.130	5.93	44.44	0.12	29.31
	BrainMass	35.88	5.65	34.04	0.121	4.85	46.36	3.19	29.93
Ours	功能连接网络 token + 支持向量机	32.16	1.65	38.41	0.133	5.37	46.43	2.07	30.02
	FCN-LLM（3B）	53.75	17.69	48.66	0.093	12.47	49.43	9.71	47.62
	FCN-LLM（7B）	54.03	18.23	49.15	0.091	10.26	49.31	9.26	46.72

。年龄预测旨在估计每位受试者的实际年龄，而认知功能预测和表型预测均涉及标准化分数的预测。对于分类任务，我们采用准确率 (ACC)、马修斯相关系数 (MCC) 和 F1分数作为评估指标；对于回归任务，则采用平均绝对误差 (MAE) 和皮尔逊相关系数 (PCC)。由于认知功能预测和表型预测任务均涉及多个属性，我们在相应表格中报告了各属性的平均结果。

4.2 定量结果

FCN-LLM 与基线模型在内部测试集上的性能比较。我们将 FCN-LLM 与六种有监督功能连接网络模型和四种功能连接网络基础模型进行了比较；基线模型的详细描述见附录A.5。如表2所示，有监督的 BNT 在所有任务中均取得了最佳性能，展现了有监督学习在聚焦特定目标时的强大能力。值得注意的是，在所有无监督方法中，我们的 FCN-LLM 取得了最先进的结果，甚至在认知功能预测的平均绝对误差指标和表型预测的皮尔逊相关系数指标上超过了其他方法。这表明，利用大语言模型理解功能连接网络，可以增强对功能连接网络数据整体分布的学习，从而更好地捕捉功能连接网络之间的全局关系模式。与基于 Qwen2.5-3B 的 FCN-LLM 相比，基于 Qwen2.5-7B 的版本在性别分类和回归任务上有所提升，但在疾病分类上的表现略有下降。这一现象可能由两个因素造成：（1）增大大语言模型的规模并不一定会提升其理解功能连接网络的能力；（2）较大的模型可能需要更多指令微调数据，才能有效获取功能连接网络相关知识。

FCN-LLM与基线模型在零样本测试集上的性能比较。我们将FCN-LLM与在内部测试集上性能最佳的两个有监督功能连接网络模型，以及四个功能连接网络基础模型进行了比较。从表3可以看出，在零样本测试集上，FCN-LLM的表现优于功能连接网络基础模型和监督模型，展现出

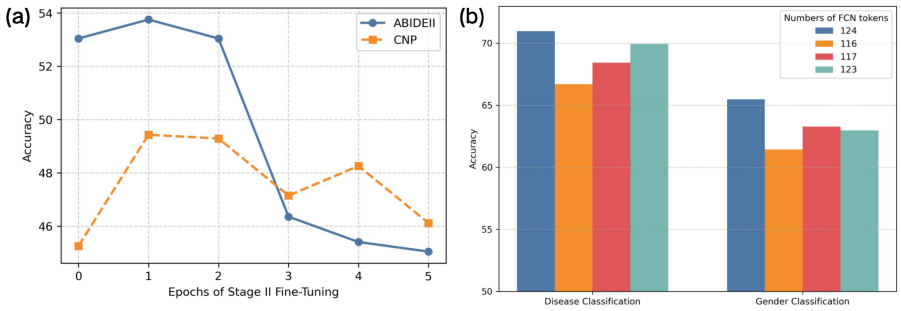


图3：（a）持续训练对阶段2中FCN-LLM泛化性能的影响；（b）对所提出的多尺度功能连接网络编码器进行的消融研究，探讨不同类型FCN令牌的影响——具体包括116个ROI级标记、7个子网络级标记和1个脑级标记的组合。

较强的泛化能力。这一优势主要源于FCN-LLM的灵活性，它支持以任意文本形式进行查询。例如，在仅包含两个类别的ABIDE II数据集中，我们通过将候选标签限制为健康对照和自闭症谱系障碍来调整提示。相比之下，其他模型只能以零样本方式将六分类器迁移到这一二分类场景，导致性能下降。与基于Qwen2.5-3B的FCN-LLM相比，基于Qwen2.5-7B的变体在ABIDE II数据集上表现更好，但在CNP数据集上的性能略有下降，这表明它们在零样本数据集上的泛化能力总体相当。

4.3 消融研究

过度调优对泛化性能的影响。如图3（a）所示，我们展示了在零样本测试集的疾病分类任务上，对多尺度功能连接网络编码器和大语言模型进行第二阶段联合微调的性能。结果表明，尽管第二阶段微调最初能够提升模型性能，但进一步增加微调轮数会导致性能下降。这表明模型可能会对微调数据过拟合，从而降低其在未见数据上的泛化能力。为缓解这一问题，我们将第二阶段的微调轮数设为1轮。这一发现也凸显了未来工作需要丰富指令微调数据的多样性，并引入更大规模FCN数据集。

多尺度 FCN 特征的有效性。为展示多尺度功能连接网络编码器的有效性，我们在内部测试集上评估了不同特征组合在疾病分类和性别分类任务中的表现，如图3(b)所示。结果表明，结合三个尺度的全部特征总体上表现最佳。对于疾病分类，与加入全脑级特征相比，加入功能性子网络级特征带来的性能提升更大，这意味着疾病判别更依赖于局部子网络信息。相反，对于性别分类，加入全局脑水平特征比加入子网络级特征带来的提升更大，表明性别相关差异更多体现在全脑模式中。这些发现凸显了利用功能连接网络的多尺度表征有效捕捉任务特定信息的重要性。

比较范式任务的消融实验。我们针对第 3.2 节提出的比较范式任务开展了更多实验，附录A.6中的详细结果证明了该任务的有效性。

5 结论

我们提出了 FCN-LLM，这是首个将脑功能连接网络与文本模态对齐的框架，使大语言模型能够理解功能连接网络。通过结合多尺度功能连接网络编码器与图级多任务指令微调，FCN-LLM 学习了跨大脑

脑区、子网络和全脑连接的功能知识。在多中心数据集上进行的大量实验表明，FCN-LLM 具有出色的零样本泛化能力，在未见任务上的表现优于现有有监督功能连接网络模型和基础模型。这些结果凸显了将神经科学数据与大语言模型相结合的潜力，为临床预测和脑研究提供了灵活且可解释的平台。未来的工作可探索纳入动态脑连接、扩展至其他神经影像学模态，并利用更丰富的基于语言的知识，进一步提升模型的理解能力和泛化能力。

伦理声明。 作者已仔细阅读并遵守 ICLR 道德准则。本研究不涉及人体受试者、敏感数据或任何引发伦理顾虑的研究实践。我们确认，本研究的所有方面，包括数据使用、实验流程和结果报告，均完全符合伦理标准和会议指南。此外，我们不存在利益冲突，且所有数据处理均依照适用法规确保隐私与安全。

可复现性声明。 我们已竭尽全力确保本研究的可复现性。数据集获取和预处理的相关细节，包括所用数据属性的描述，见附录 A.2 和 A.3。指令微调数据的生成和多范式任务的相关说明见第 3.2 节。预训练和微调阶段所采用的模型架构、训练流程及超参数设置见第 3.1 节、第 3.3 节和表 7。其他实现细节和补充分析见附录 A.4。论文被接收后，我们将公开复现实验所需的代码和脚本。

参考文献

Josh Achiam、Steven Adler、Sandhini Agarwal、Lama Ahmad、Ilge Akkaya、Florencia Leoni Aleman、Diogo Almeida、Janko Altenschmidt、Sam Altman、Shyamal Anadkat 等。Gpt-4 技术报告。*arXiv* 预印本 *arXiv:2303.08774*，2023年。

Lindsay M Alexander、Jasmine Escalera、Lei Ai、Charissa Andreotti、Karina Febre、Alexander Mangone、Natan Vega-Potler、Nicolas Langer、Alexis Alexander、Meagan Kovacs 等。用于儿童心理健康和学习障碍跨诊断研究的开放资源。*Scientific Data*，4(1):1–26，2017。

Stanislaw Antol、Aishwarya Agrawal、Jiasen Lu、Margaret Mitchell、Dhruv Batra、C Lawrence Zitnick 和 Devi Parikh。视觉问答：视觉问答。在 电气电子工程师学会国际计算机视觉会议论文集，第 2425–2433 页，2015年。

Akari Asai、Mohammadreza Salehi、Matthew E Peters 和 Hannaneh Hajishirzi。ATTEMPT：通过注意力机制混合软提示进行参数高效的多任务调优。*arXiv* 预印本 *arXiv:2205.11961*，2022年。

R Bilder、R Poldrack、T Cannon、E London、N Freimer、E Congdon、K Karlsgodt 和 F Sabb。“加州大学洛杉矶分校神经精神表型学联盟 LA5c 研究”，2020年。

Boson AI。Higgs Audio V2：重新定义音频生成的表现力。https://github. com/boson-ai/higgs-audio，2025。GitHub 代码库。发布博客见 https: //www.boson.ai/blog/higgs-audio-v2。

Ed Bullmore 和 Olaf Sporns。复杂脑网络：结构和功能系统的图论分析。*Nature Reviews Neuroscience*，10(3):186–198，2009年。

Jianlv Chen、Shitao Xiao、Peitian Zhang、Kun Luo、Defu Lian 和 Zheng Liu。BGE M3-Embedding：通过自知识蒸馏实现多语言、多功能、多粒度的文本嵌入。*arXiv* 预印本 *arXiv:2402.03216*，2024a。

Zixiang Chen、Yihe Deng、Huizhuo Yuan、Kaixuan Ji 和 Quanquan Gu。自博弈微调将弱语言模型转化为强语言模型。*arXiv* 预印本 *arXiv:2401.01335*，2024b。

ADHD-200联盟。ADHD-200联盟：推动神经影像学在临床神经科学中转化潜力的模型。*Frontiers in Systems Neuroscience*，6:62，2012年。

Hejie Cui、Wei Dai、Yanqiao Zhu、Xuan Kan、Antonio Aodong Chen Gu、Joshua Lukemire、Liang Zhan、Lifang He、Ying Guo 和 Carl Yang。Braingb：基于图神经网络进行脑网络分析的基准测试。*IEEE Transactions on Medical Imaging*，42(2)：493–506，2022年。

Jacob Devlin、Ming-Wei Chang、Kenton Lee 和 Kristina Toutanova。BERT：用于语言理解的深度双向变换器预训练。载于计算语言学协会北美分会 *2019* 年会议论文集：人类语言技术，第 *1* 卷（长篇和短篇论文），第 4171–4186 页，2019年。

Adriana Di Martino、Chao-Gan Yan、Qingyang Li、Erin Denio、Francisco X Castellanos、KaatAlaerts、Jeffrey S Anderson、Michal Assaf、Susan Y Bookheimer、Mirella Dapretto 等。《自闭症脑成像数据交换项目：大规模评估自闭症患者内在脑结构》。《分子精神病学》，19(6)：659–667，2014。

Sam Ereira、Sheena Waters、Adeel Razi 和 Charles R Marshall。《利用默认模式网络有效连接早期检测痴呆》。《自然·心理健康》，2(7)：787–800，2024年。

Oscar Esteban、Christopher J Markiewicz、Ross W Blair、Craig A Moodie、A Ilkay Isik、Asier Erramuzpe、James DKent、Mathias Goncalves、Elizabeth DuPre、Madeleine Snyder 等。fmriprep：一种稳健的功能性磁共振成像预处理流程。《自然·方法》，16(1)：111–116，2019年。

Emily S Finn、Xilin Shen、Dustin Scheinost、Monica D Rosenberg、Jessica Huang、Marvin M Chun、Xenophon Papademetris和RTodd Constable。功能连接组指纹识别：利用脑连接模式识别个体。《自然·神经科学》，18(11):1664–1671，2015年。

Luciano Floridi和Massimo Chiriatti。GPT-3：其本质、范围、局限与影响。《心智与机器》，30(4):681–694，2020年。

Karl J Friston。神经影像学中的功能连接与有效连接：一项综合。《人脑图谱》，2(1-2):56–78，1994年。

Chaoyou Fu、Yuhan Dai、Yongdong Luo、Lei Li、Shuhuai Ren、Renrui Zhang、Zihan Wang、Chenyu Zhou、Yunhang Shen、Mengdan Zhang等。Video-MME：视频分析中多模态大语言模型的首个全面评测基准。载于计算机视觉与模式识别会议论文集，第24108–24118页，2025。

Daya Guo、Dejian Yang、Haowei Zhang、Junxiao Song、Ruoyu Zhang、Runxin Xu、Qihao Zhu、Shirong Ma、Peiyi Wang、Xiao Bi 等。DeepSeek-R1：通过强化学习激发大语言模型的推理能力。*arXiv* 预印本 *arXiv:2501.12948*，2025。

Jiaming Han、Renrui Zhang、Wenqi Shao、Peng Gao、Peng Xu、Han Xiao、Kaipeng Zhang、Chris Liu、Song Wen、Ziyu Guo 等。ImageBind-LLM：多模态指令微调。*arXiv* 预印本 *arXiv:2309.03905*，2023。

Zhibin He、Wuyang Li、Yifan Liu、Xinyu Liu、Junwei Han、Tuo Zhang 和 Yixuan Yuan。FM-App：通过功能性磁共振成像到结构性磁共振成像的知识迁移，实现任意表型预测的基础模型。*IEEETransactions on Medical Imaging*，2024年。

Leanna MHernandez、Jeffrey D Rudie、Shulamite AGreen、Susan Bookheimer 和 MirellaDapretto。自闭症谱系障碍的神经特征：对大脑网络动态的洞见。神经精神药理学，40(1):171–189，2015年。

AJ Holmes、MO Hollinshead、TM O'Keefe、VI Petrov、GR Fariello、LL Wald、B Fischl、BR Rosen、RW Mair、JL Roffman 等。脑基因组超结构项目初始数据发布：包含结构、功能和行为测量。科学数据 2，150031，2015年。

Jinlong Hu、Yangmin Huang、Nan Wang 和 Shoubin Dong。BrainNPT：用于脑网络分类的 Transformer 网络预训练。*IEEE*神经系统与康复工程汇刊，2024年。

Weihua Hu、Bowen Liu、Joseph Gomes、Marinka Zitnik、Percy Liang、Vijay Pande 和 Jure Leskovec。图神经网络预训练策略。*arXiv* 预印本 *arXiv:1905.12265*，2019年。

Xingcan Hu、Wei Wang 和 Li Xiao。利用脑功能连接网络监督学习三维医学影像模型，以进行精神障碍诊断。载于医学图像计算与计算机辅助干预国际会议，第 336–346 页。Springer，2025年。

Xuan Kan、Hejie Cui、Joshua Lukemire、Ying Guo 和 Carl Yang。Fbnetgen：通过功能性脑网络生成进行任务感知的基于 GNN 的功能性磁共振成像分析。载于国际医学影像与深度学习会议，第 618–637 页。PMLR，2022a。

Xuan Kan、Wei Dai、Hejie Cui、Zilong Zhang、Ying Guo 和 Carl Yang。脑网络变换器。神经信息处理系统进展，35：25586–25599，2022b。

Jeremy Kawahara、Colin J Brown、Steven P Miller、Brian G Booth、Vann Chau、Ruth E Grunau、Jill G Zwicker 和 Ghassan Hamarneh。BrainNetCNN：用于脑网络的卷积神经网络；面向神经发育预测。*NeuroImage*，146：1038–1049，2017。

Thomas N Kipf 和 Max Welling。利用图卷积网络进行半监督分类。*arXiv* 预印本 *arXiv:1609.02907*，2016。

Junnan Li、Dongxu Li、Caiming Xiong 和 Steven Hoi。Blip：为统一的视觉-语言理解与生成进行语言-图像预训练引导。发表于国际机器学习会议，第 12888–12900 页。PMLR，2022年。

Junnan Li、Dongxu Li、Silvio Savarese 和 Steven Hoi。Blip-2：利用冻结图像编码器和大语言模型引导语言-图像预训练。在国际机器学习会议上，第 19730–19742 页。PMLR，2023。

Xiaoxiao Li、Yuan Zhou、Nicha Dvornek、Muhan Zhang、Siyuan Gao、Juntang Zhuang、Dustin Scheinost、Lawrence H Staib、Pamela Ventola 和 James S Duncan。Braingnn：用于功能性磁共振成像分析的可解释脑图神经网络。*Medical Image Analysis*，74：102233，2021年。

Bin Lin、Yang Ye、Bin Zhu、Jiaxi Cui、Munan Ning、Peng Jin 和 Li Yuan。Video-llava：通过投影前对齐学习统一的视觉表征。*arXiv* 预印本 *arXiv:2311.10122*，2023。

Tsung-Yi Lin、Michael Maire、Serge Belongie、James Hays、Pietro Perona、Deva Ramanan、Piotr Doll´ar 和 C Lawrence Zitnick。Microsoft coco：上下文中的常见物体。在欧洲计算机视觉会议上，第 740–755 页。Springer，2014。

Haotian Liu、Chunyuan Li、Qingyang Wu 和 Yong Jae Lee。视觉指令微调。神经信息处理系统进展，36:34892–34916，2023。

Weronika Machaj、Przemysław Podg´orski、Julian Maciaszek、Patryk Piotrowski、Dorota Szcze´sniak、Adrian Korbecki、Joanna Rymaszewska 和 Anna Zimny。利用静息态功能磁共振成像评估难治性抑郁症患者主要脑网络内与网络间的连接性。《临床医学杂志》，13(18):5507，2024年。

Kenneth Marino、Mohammad Rastegari、Ali Farhadi 和 Roozbeh Mottaghi。OK-VQA：一项需要外部知识的视觉问答基准测试。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第3195–3204页，2019年。

Vinod Menon。默认模式网络的20年：综述与综合。《神经元》，111(16)：2469–2487，2023。Vassil Panayotov、Guoguo Chen、Daniel Povey和Sanjeev Khudanpur。LibriSpeech：基于公有领域有声读物的自动语音识别语料库。载于*2015*年*IEEE*国际声学、语音与信号处理会议（*ICASSP*），第5206–5210页。电气电子工程师学会，2015年。Jonathan D Power、Alexander L Cohen、Steven M Nelson、Gagan S Wig、Kelly Anne Barnes、Jessica A Church、Alecia C Vogel、Timothy O Laumann、Fran M Miezin、Bradley L Schlaggar等。人脑的功能网络组织。《神经元》，72(4)：665–678，2011年。

Alec Radford、Jong Wook Kim、Chris Hallacy、Aditya Ramesh、Gabriel Goh、Sandhini Agarwal、Girish Sastry、Amanda Askell、Pamela Mishkin、Jack Clark 等人。《通过自然语言监督学习可迁移视觉模型》。载于国际机器学习会议，第 8748–8763 页。PMLR，2021年。

Victor Sanh、Albert Webson、Colin Raffel、Stephen H Bach、Lintang Sutawika、Zaid Alyafeai、Antoine Chaffin、Arnaud Stiegler、Teven Le Scao、Arun Raja 等人。《多任务提示训练实现零样本任务泛化》。*arXiv* 预印本 *arXiv:2110.08207*，2021年。

Julia M Sheffield、Grega Repovs、Michael P Harms、Cameron S Carter、James M Gold、Angus W MacDonald III、J Daniel Ragland、Steven M Silverstein、Douglass Godwin和Deanna M Barch。额顶网络与扣带-岛盖网络完整性以及健康与精神分裂症中的认知。*Neuropsychologia*，73:82–93，2015年。

Sheng Shen、Shijia Yang、Tianjun Zhang、Bohan Zhai、Joseph E Gonzalez、Kurt Keutzer和Trevor Darrell。多任务视觉-语言提示调优。载于*IEEE/CVF*冬季计算机视觉应用会议论文集，第5656–5667页，2024年。

Lachlan T. Strike、Gabriella A.M. Blokland、Narelle K. Hansell、Nicholas G. Martin、Arthur W. Toga、Paul M. Thompson、Greig I. de Zubicaray、Katie L. McMahon和Margaret J. Wright。“昆士兰双生子成像（QTIM）”，2023。

Saori C Tanaka、Ayumu Yamashita、Noriaki Yahata、Takashi Itahashi、Giuseppe Lisi、TakashiYamada、Naho Ichikawa、Masahiro Takamura、Yujiro Yoshihara、Akira Kunimatsu等。多中心、多疾病静息态磁共振影像数据库。*Scientific Data*，8(1):227，2021年。

Nathalie Tzourio-Mazoyer、Brigitte Landeau、Dimitri Papathanassiou、Fabrice Crivello、Octave Etard、Nicolas Delcroix、Bernard Mazoyer和Marc Joliot。使用蒙特利尔神经学研究所MRI单被试大脑的宏观解剖分区，在SPM中对激活进行自动解剖标记。*NeuroImage*，15(1):273–289，2002。

David C Van Essen、Stephen M Smith、Deanna M Barch、Timothy EJ Behrens、Essa Yacoub、Kamil Ugurbil、吴-明尼苏达人类连接组计划联盟等。吴-明尼苏达人类连接组计划：概述。*NeuroImage*，80:62–79，2013。

Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N Gomez、Łukasz Kaiser 和 Illia Polosukhin。《Attention is all you need》。神经信息处理系统进展，第 30 卷，2017。

Bin Wang、Xunlong Zou、Geyu Lin、Shuo Sun、Zhuohan Liu、Wenyu Zhang、Zhengyuan Liu、AiTi Aw 和 Nancy F Chen。《Audiobench：面向音频大语言模型的通用基准测试》。*arXiv* 预印本 *arXiv:2406.16020*，2024a。

Wei Wang、Zhaowei Li、Qi Xu、Yiqing Cai、Hang Song、Qi Qi、Ran Zhou、Zhida Huang、Tao Wang 和 Li Xiao。《Qcrd：面向大语言模型的质量引导的对比式推理蒸馏》。*arXiv* 预印本 *arXiv:2405.13014*，2024b。

Wei Wang、Zhaowei Li、Qi Xu、Linfeng Li、YiQing Cai、Botian Jiang、Hang Song、Xingcan Hu、Pengyu Wang 和 Li Xiao。通过多模态模型中的多尺度对齐推进细粒度视觉理解。*arXiv* 预印本 *arXiv:2411.09691*，2024c。

Wei Wang、Li Xiao、Gang Qu、Vince D Calhoun、Yu-Ping Wang 和 Xiaoyan Sun。面向多中心、多图谱功能性磁共振成像的多视图超边感知超图嵌入学习及功能连接网络分析。*Medical Image Analysis*，94:103144，2024d。

Xuesong Wang、Lina Yao、Islem Rekik 和 Yu Zhang。用于基于群体的功能性磁共振成像分类的对比式功能连接图学习。发表于医学图像计算与计算机辅助干预国际会议，第 221–230 页。Springer，2022a。

Xuezhi Wang、Jason Wei、Dale Schuurmans、Quoc Le、Ed Chi、Sharan Narang、Aakanksha Chowdhery 和 Denny Zhou。《自洽性提升语言模型中的思维链推理》。*arXiv* 预印本 *arXiv:2203.11171*，2022b。

Zhen Wang、Rameswar Panda、Leonid Karlinsky、RogerioFeris、Huan Sun 和 Yoon Kim。多任务提示调优实现参数高效迁移学习。*arXiv* 预印本 *arXiv:2303.02861*，2023。

Xinxu Wei、Kanhao Zhao、Yong Jiao、Lifang He和Yu Zhang。脑图基础模型：面向任意图谱与疾病的预训练和提示调优。*arXiv*预印本 *arXiv:2506.02044*，2025。

Chao-Gan Yan、Xiao Chen、Le Li、Francisco Xavier Castellanos、Tong-Jian Bai、Qi-Jing Bo、Jun Cao、Guan-Mao Chen、Ning-Xuan Chen、Wei Chen等。复发性重度抑郁障碍患者的默认模式网络功能连接降低。美国国家科学院院刊，116(18):9078–9083，2019年。

An Yang、Baosong Yang、Beichen Zhang、Binyuan Hui、Bo Zheng、Bowen Yu、Chengyuan Li、Dayiheng Liu、Fei Huang、Haoran Wei、Huan Lin、Jian Yang、Jianhong Tu、Jianwei Zhang、Jianxin Yang、Jiaxi Yang、Jingren Zhou、Junyang Lin、Kai Dang、Keming Lu、Keqin Bao、Kexin Yang、Le Yu、Mei Li、Mingfeng Xue、Pei Zhang、Qin Zhu、Rui Men、Runji Lin、Tianhao Li、Tingyu Xia、Xingzhang Ren、Xuancheng Ren、Yang Fan、Yang Su、Yichang Zhang、Yu Wan、Yuqiong Liu、Zeyu Cui、Zhenru Zhang 和 Zihan Qiu。Qwen2.5 技术报告。*arXiv* 预印本 *arXiv:2412.15115*，2024a。

Yanwu Yang、Chenfei Ye、Junyan Sun、Li Liang、Haiyan Lv、Linlin Gao、Jiliang Fang、Ting Ma 和 Tao Wu。帕金森病进展过程中脑结构连接的改变：全连接组网络分析。*NeuroImage: Clinical*，31:102715，2021年。

Yanwu Yang、Chenfei Ye、Guinan Su、Ziyao Zhang、Zhikai Chang、Hairui Chen、Piu Chan、Yue Yu 和 Ting Ma。BrainMass：通过大规模自监督学习推进用于诊断的脑网络分析。电气电子工程师学会医学影像汇刊，43(11):4004–4016，2024b。

Yi Yang、Hejie Cui 和 Carl Yang。PTGB：预训练用于脑网络分析的图神经网络。*arXiv* 预印本 *arXiv:2305.14376*，2023。

BT Thomas Yeo、Fenna M Krienen、Jorge Sepulcre、Mert R Sabuncu、Danial Lashkari、Marisa Hollinshead、Joshua L Roffman、Jordan W Smoller、Lilla Z¨ollei、Jonathan R Polimeni 等。通过内在功能连接估计的人类大脑皮层组织。《神经生理学杂志》，2011年。

Jiahui Yu、Zirui Wang、Vijay Vasudevan、Legg Yeung、Mojtaba Seyedhosseini 和 Yonghui Wu。CoCa：对比式图像描述模型是图像-文本基础模型。*arXiv* 预印本 *arXiv:2205.01917*，2022年。

Ru Zhang、StuartB Murray、ChristinaJ Duval、Danny JJWang 和 KayJann。对展现注意缺陷多动障碍行为症状的青春期前青少年进行静息态功能磁共振成像的功能连接与复杂性分析。*Psychiatry Research*，334：115794，2024年。

Deyao Zhu、Jun Chen、Xiaoqian Shen、Xiang Li 和 Mohamed Elhoseiny。MiniGPT-4：利用先进的大语言模型增强视觉-语言理解。*arXiv* 预印本 *arXiv:2304.10592*，2023。

A 附录

A.1 功能连接网络

功能连接网络通常通过计算每对脑区平均 BOLD信号之间的相关性来构建（本研究采用皮尔逊相关）（Friston，1994）。具体而言，假设脑区 a 的 BOLD信号为 x，脑区 b 的 BOLD信号为 y，则皮尔逊相关系数的计算如下：

$$r_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{6}$$

其中，$\bar{x}$ 和 $\bar{y}$ 分别表示 x 和 y的均值，n 表示 BOLD信号中的时间点数量。

功能连接网络刻画了大脑区域间交互的响应，以区域尺度表征人脑。然而，现有研究表明，进一步将脑区划分为子网络（即将功能相似的脑区归为一组），并探索这些子网络之间的交互，是解码疾病相关模式等特征的关键方法。

此外，大量研究表明，利用图神经网络（GNNs）等方法学习全脑尺度的功能连接网络统一表征，在下游任务中展现出巨大潜力（Kipf & Welling，2016）。

A.2 数据集

数据集中的数据分布（原始与增强）如表4所示，包含五个具体部分：性别分布、利手分布、年龄组分布、诊断分布和表型统计。本文采用时间窗功能连接网络增强，滑动窗口长度 L 为100，步长 P 为20。此外，表型统计部分包含 HCP 表型的12项具体指标：1）视觉空间加工测试、正确总数；2）NIH工具箱口语阅读识别测试；3）感知压力；4）愤怒与攻击性；5）NIH工具箱握力测试；6）2分钟步行测试；7）NIH工具箱图片词汇测试；8）NIH 工具箱列表排序工作记忆测验；9）愤怒与敌意；10）孤独感；11）意义与目标；12）NIH 工具箱九孔插板灵巧度测试。有关这些指标的具体定义和来源，请参阅表5。

(a) 性别分布

性别	原始	增强
Male	4237	39002
女性	3516	40548

(b) 利手分布

利手	原始	增强
惯用右手	4417	16987
惯用左手	296	995
左右手混用	240	290

(c) 年龄组分布

年龄区间	原始	增强
$[0, 10)$	1200	2293
$[10, 19)$	1994	7210
$[19, 40)$	3686	63517
$[40, 65)$	761	5622
$[65, 100)$	112	908

(d) 表型统计

表型指标	原始	增强
FIQ	1206	5985
VIQ	1033	6256
PIQ	1046	6249
12 HCP表型	$\sim 860\times12$	$\sim 48986\times12$

（e）诊断分布

诊断	原始	增强	诊断	原始	增强
健康对照	1432	9168	ASD	416	2727
MDD	1183	8178	ADHD	260	1477
精神分裂症	117	901	Pain	19	171
双相情感障碍	33	297	心境恶劣	3	27
其他	18	162		-	

表4：人口统计学特征、诊断分布及表型样本量（原始数据集与时间窗口增强数据集）

A.3 数据预处理

所有静息态功能磁共振成像数据均使用 fMRIPrep（Esteban 等，2019）进行预处理。fMRIPrep 是一种易于使用、先进且能够适应不同扫描采集方案的功能磁共振成像数据预处理流程。对于静息态功能磁共振成像数据，我们进行了切片时间校正、头动校正、颅骨剥离，并将数据空间标准化至 MNI152空间。经过这些预处理步骤后，静息态功能磁共振成像数据的体素大小为 $2mm^3$。为基于自动解剖标记图谱（Tzourio-Mazoyer 等，2002）从静息态功能磁共振成像数据中提取功能连接网络，该图谱将全脑划分为 116 个感兴趣区，我们将成对感兴趣区的血氧水平依赖时间序列之间的皮尔逊相关定义为功能连接。为使扫描时长不同的数据集所表示的功能连接网络图的节点特征维度保持一致，我们采用节点连接特征，即功能连接网络矩阵中与每个节点对应的行，作为节点特征。已证明，所定义的节点特征能够

表5：表型指标的详细信息

变量名称	描述	测量内容	数值较高表示
VSPLOT_TC	视觉空间过程测验，正确总数	视觉空间能力	更强的空间推理能力和更多正确反应
ReadEng_Unadj	NIH 工具箱口语阅读-ing识别测试	阅读 and word 识别能力	更强的阅读和语言解码技能
未调整的感知压力	感知压力	主观压力水平	更高的感知压力
未调整的愤怒攻击性	愤怒与攻击性（来自 PROMIS）	攻击行为愤怒时	更倾向于产生攻击性反应性反应
未调整的力量	NIH 工具箱 Grip 力量测试	身体力量（手部握力）	更强的肌肉力量
耐力 未调整	2分钟步行测试	心血管／身体耐力	更好的耐力和更长的步行距离
PicVocab_未调整	NIH工具箱图片词汇测试	接受性词汇和语言	词汇量更大，且语言理解能力更强
ListSort_未调整	NIH 工具箱列表排序工作记忆 Test	工作记忆能力 ity	更强的工作记忆和认知处理能力
AngHostil_未调整	愤怒－敌意（来自 PROMIS）	敌意态度或感受-ings	更强烈的敌意感
孤独感 未调整的	孤独感（来自 PROMIS）	主观感受 孤独感	更强烈的社会孤立感
MeanPurp_未调整	意义与目标（来自 NIH 工具箱）	生命意义感和目标感	更强的目标感和生活满意度
未调整灵巧度	NIH 工具箱九孔插孔板灵巧度测试	手部和手指运动协调能力	更好的手部灵巧度和精细运动技能

表 6：认知功能指标详情

缩写全称		含义	描述
FIQ	全量表智商	整体智力能 ity	反映一个人的整体智力水平，通常由加权多个分测验得分组合而得
VIQ	言语智商	言语智力能- ity	衡量语言相关认知能力，如词汇、理解-理解力和言语推理
PIQ	操作智商	性能（非言语）智力能力	测量非言语认知能力，例如空间推理、模式识别，以及手眼协调能力

性能优于其他类型的节点特征，例如节点身份、度和基于特征向量的嵌入（Cui等人，2022；Kan等人，2022a）。
对于连续属性，我们通过以下方式统一输出格式，并消除取值范围差异的影响：(1) 应用最小-最大归一化，将所有回归值重新缩放至 [0,1]（He等人，2024）；(2) 将归一化后的值线性映射为 [0,100]范围内的整数。因此，所有回归输出均表示为 [0,100], 中的整数，并作为标准化的FCN-LLM监督信号，从而确保多任务输出格式一致。

A.4 实验设置

我们补充介绍实验配置的详细信息，以便于复现模型。各阶段的超参数汇总于表 7，数据集详情汇总于表8。此外，预测任务的数量与判断任务一致：通过提供候选答案并要求输出“是/否”，即可将每个预测任务转化为判断任务。对于比较任务，我们为每项任务手动设定固定的目标数量（例如50k），使第一训练阶段的样本总数达到700K，与预测任务和判断任务的样本数量大致相当；随后，在第二训练阶段将这一总数减半。

在模型推理过程中，我们根据测试数据集的属性类型和先验知识灵活调整提示。例如，在疾病分类任务中，我们会在提示中提供候选疾病类别，并根据每个测试集中的具体疾病进行调整。此外，我们采用自一致性策略（Wang等人，2022b）汇总多个提示的响应，以期获得最佳性能。我们采用关键词抽取和规则匹配方法，将 FCN-LLM 生成的文本响应解析为可比较标签。无法解析的输出（如乱码文本）则直接计为错误。

Size	阶段 1	阶段 2
批大小	32	32
学习率	1e-3	1e-5
训练轮数	1	1
学习率调度	余弦衰减	余弦衰减
预热比例	0.03	0.03
权重衰减	0	0
BF16	✓	✓
TF32	✓	✓
DeepSpeed 阶段	ZeRO2	ZeRO2
GPUs	8xA100	8xA100
多尺度编码器	训练	训练
LLM	冻结	训练

表 7：模型训练的超参数。

阶段	数据集		样本
阶段1	增强FCN数据	预测任务	806K
		判断任务	806K
		比较任务	700K
第二阶段	高层监督微调数据	预测任务	38K
		判断任务	38K
		比较任务	350K

表8：用于模型训练的数据集详情。

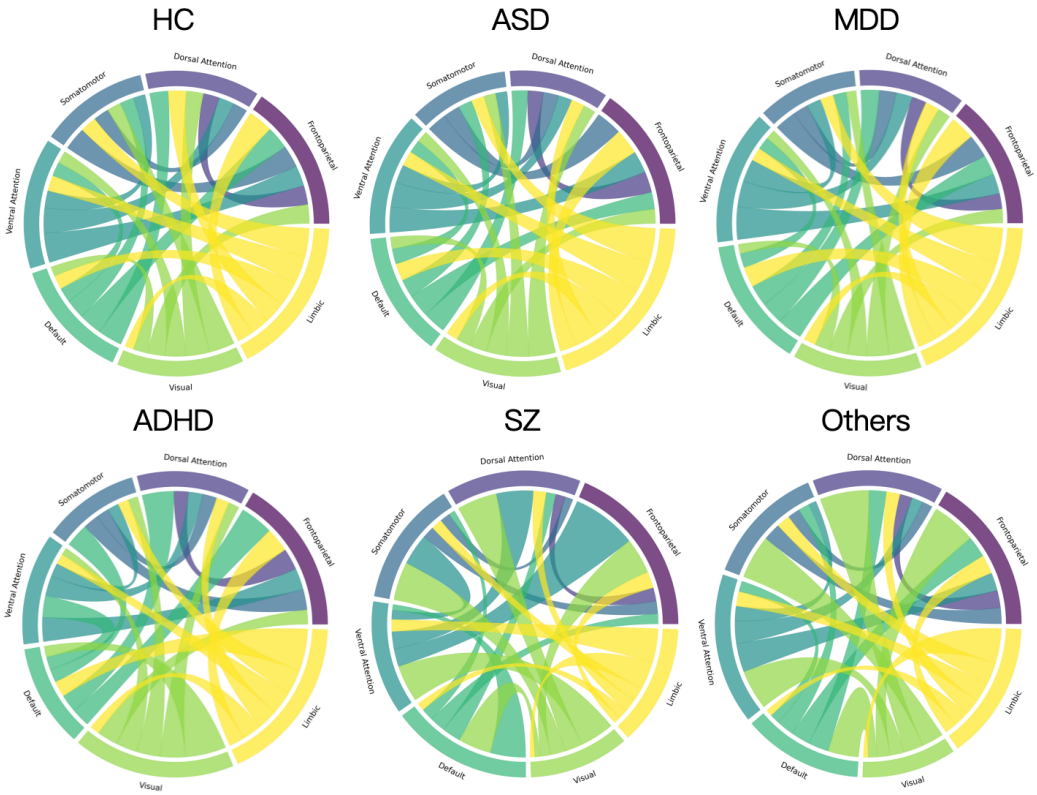


图4：从子网络视角可视化不同疾病组的FCN令牌之间的连接。

A.5 基线

我们主要将 FCN-LLM 与两类用于脑功能连接网络分析的模型进行比较：1）从头开始训练的监督模型，包括 GCN（Kipf & Welling，2016）、HGCN（Wang等人，2024d）、BrainNetCNN（Kawahara等人，2017）、BrainGNN（Li等人，2021）、Transformer（Kan等人，2022b）和 BNT（Kan等人，2022b）；2）用于功能连接网络的基础模型，包括 BrainNPT（Hu等人，2024）、PTGB（Yang等人，2023）、CINP（Hu等人，2025）和 BrainMass（Yang等人，2024b）。

对于有监督功能连接网络模型，我们进一步将初始训练集按90/10的比例划分为训练集和验证子集。通过超参数搜索，我们为每个属性训练了一个独立模型，并选择在验证集上性能最佳的模型。最终性能随后在内部测试集和零样本测试集上进行评估。为使用功能连接网络嵌入进行比较，我们用基础FCN模型提取了训练集和测试集的嵌入，对训练嵌入进行归一化，并为每个属性训练了一个支持向量机分类器。随后将这些分类器应用于内部测试集和零样本测试集，以获得评估指标。

A.6 比较范式任务的消融

如表9所示，我们专门探究了在训练期间加入比较范式任务对所提出的FCN-LLM（3B）模型性能的影响。从结果可以看出，纳入比较范式任务——即模型在训练过程中以数据驱动的方式学习不同样本之间的关系——有助于更有效地进行FCN解码。例如，这使疾病分类任务的性能提升了2.25%。

表9：在内部测试集上，对 FCN-LLM (3B) 的比较范式任务进行消融研究，并从性别分类、疾病分类和年龄预测方面进行评估。

比较任务	性别分类			疾病分类			年龄预测	
	ACC	MCC	F1	ACC	MCC	F1	MAE	PCC
×	64.23	29.03	64.89	68.72	54.69	68.85	6.631	73.28
✓	65.48	29.82	65.39	70.97	57.41	70.78	6.293	75.81

A.7 疾病生物标志物可视化

FCN-LLM基于功能连接网络特征输入和文本提示，对最大似然输出进行建模。通过以参考提示为条件，模型能够识别感兴趣区域特征中对输出影响显著的部分。如图4所示，为检测与疾病相关的脑生物标志物，我们计算了FCN-LLM各层注意力图的平均值，并从子网络视角将FCN令牌之间的连接可视化。

获取 FCN 令牌之间注意力图的过程包括以下几个步骤：（1）在提示模型进行疾病诊断时，选取 FCN 令牌与类别输出（即 HC/MDD/ASD 等）之间的注意力分数；（2）对注意力头维度和类别输出两个维度上的注意力分数分别求和；（3）将归一化注意力分数映射到 FCN 令牌上，以获得各 FCN 令牌之间的交互。

从图中可以得出，在注意缺陷多动障碍组中，静息态功能磁共振成像研究揭示了默认模式网络（DMN）与注意网络之间存在异常的功能交互，这可能会损害注意调节（Zhang等人，2024年）。在 MDD 组中，视觉网络（VN）与显著性网络或背侧注意网络（SN/DAN）之间的功能连接增强，这可能反映了对外部刺激的敏感性提高（Machaj等人，2024年）。在 ASD 组中，DMN 的连接性降低，这可能是社会认知和自我参照加工缺陷的潜在根源。这些发现凸显了疾病特异性的功能网络组织紊乱，并强调了网络交互在认知与行为异常中的作用（Hernandez等人，2015年）。

A.8 注意力分析

请参见图5和图6。

A.9 大语言模型的使用

在本研究中，大语言模型（LLM）主要作为通用辅助工具，用于以下两个方面：

稿件润色。 我们使用大语言模型提升稿件的清晰度、流畅度和学术风格。具体而言，它们协助改写句子、纠正语法错误，并将非正式或口语化的表达转换为正式的学术语言。本稿中提出的所有科学观点、实验设计和解释均由作者构思并撰写。大语言模型未参与研究假设的制定或实验结果的分析。

编程辅助。 我们还使用大语言模型协助完成编程任务，包括编写和重构代码、调试，以及解决环境或运行时问题。具体而言，它们协助编写脚本、提供软件包安装指导，并处理模型实现过程中遇到的错误消息。底层算法、模型设计和实验流程均由作者独立开发和实现，大语言模型仅作为技术助理。

大语言模型生成的内容未被用于替代原始研究贡献，其作用仅限于提供辅助，重点在于改善论文的呈现效果和编程效率。

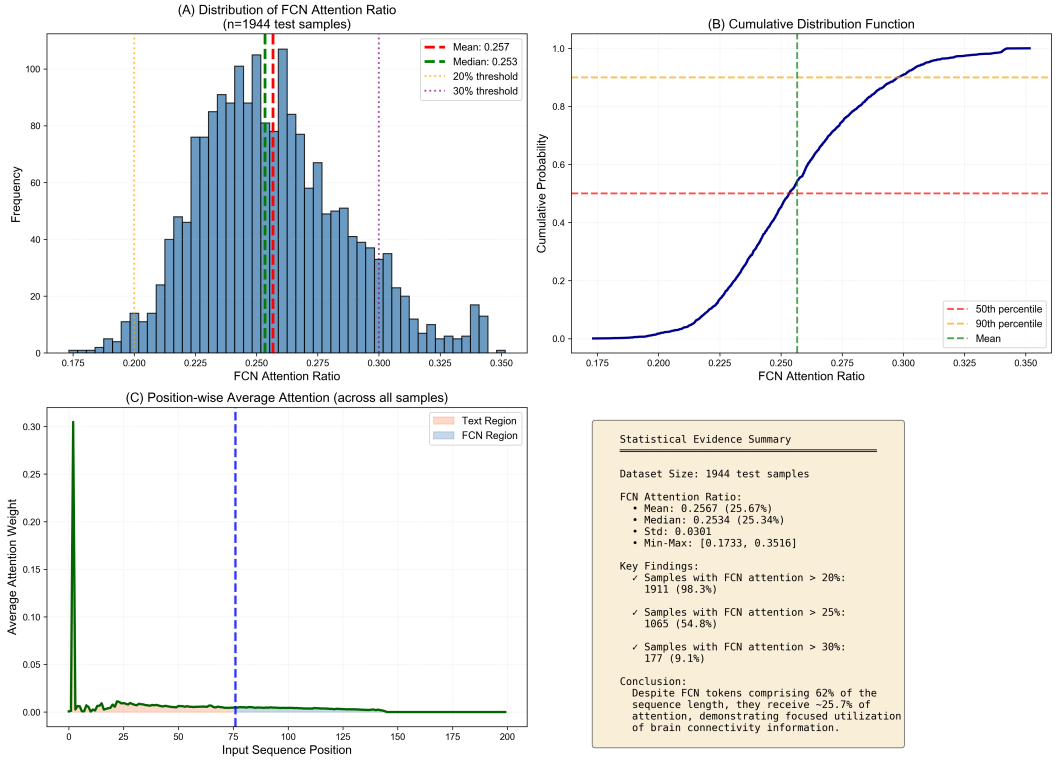


图5：统计证据综合可视化。

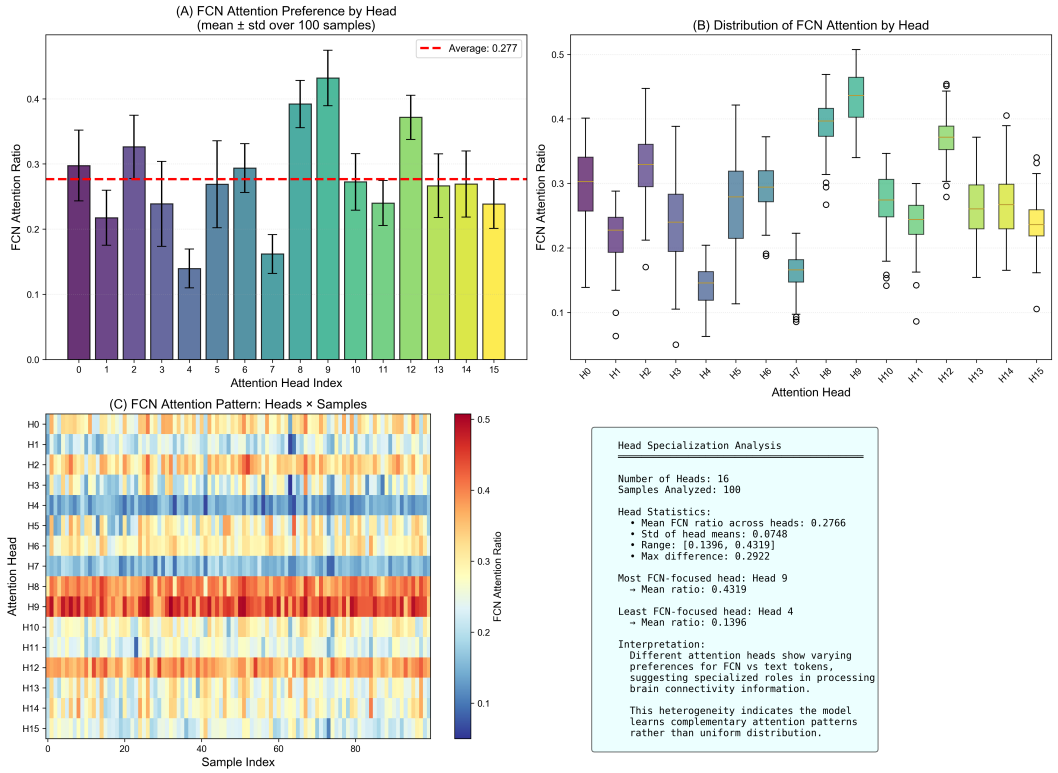


图6：注意力头专门化分析。