

SYSTEM BOUNDARY

大脑教给 AI 的三件事

坐标、记忆与闭环

从系统工程理解 NeuroAI 与脑机接口

2026.10.05—10.11 研究观察 · Zach 授权合成声音

AI 概念插画

脑机接口的信号链

环节	我们想知道的变量	可能发生的变化
任务 / 意图	现在想做什么	真正的认知变化
神经活动	群体活动与内部状态	学习、疲劳、上下文
组织 / 设备	观测如何被混合	位置、阻抗、采集条件
解码 / 输出	状态怎样变成动作	任务标定、模型失配

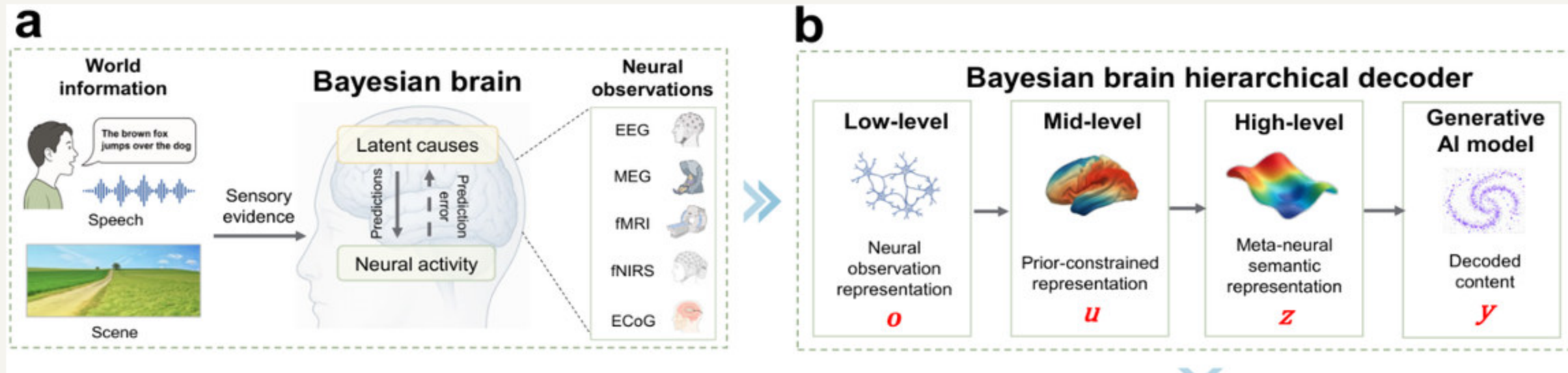
读数改变，可能来自被测对象，也可能来自测量链路

不同模态，打开不同窗口

测量窗口	主要观测	选择时先问
EEG / MEG	头皮电位 / 微弱磁场	快速变化能否定位、能否跨日稳定？
ECoG	靠近皮层的电活动	侵入式局部覆盖是否符合任务需要？
fMRI	血氧相关的间接响应	空间覆盖与血流动力学如何建模？

先定要区分的机制，再定模态组合；本页为概念对照

解码是一个带先验的逆问题



观测 $\approx$ 测量模型 (隐状态) + 噪声

已知结构缩小可行解空间；先验错误也会带来偏差

稳定坐标，动态观测

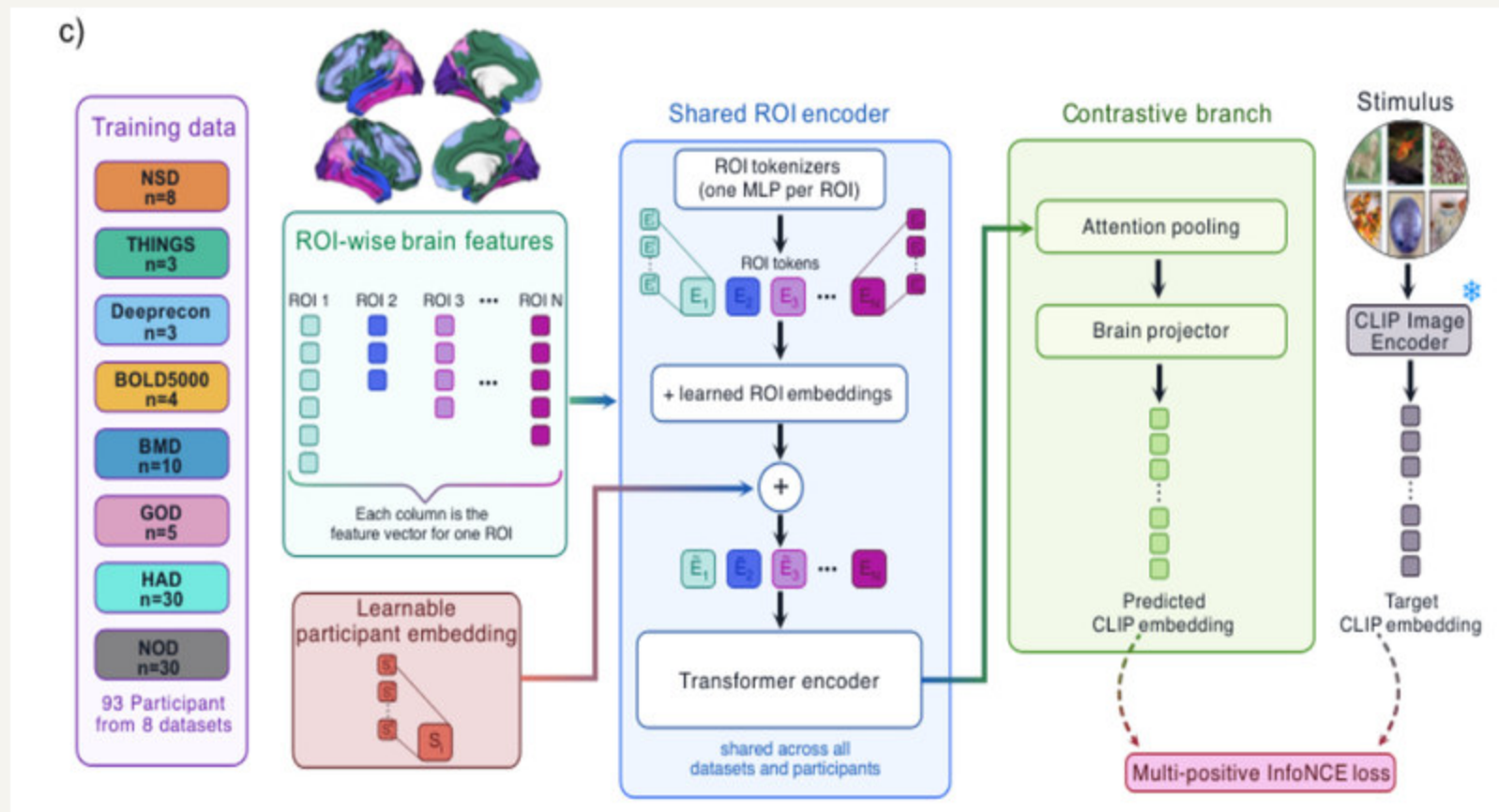
皮层几何本征模态：由皮层形状定义的空间基函数

跨 session 表征相似度



表示更一致，不等于所有在线 BCI 任务都同幅改善

共享的可能是关系几何

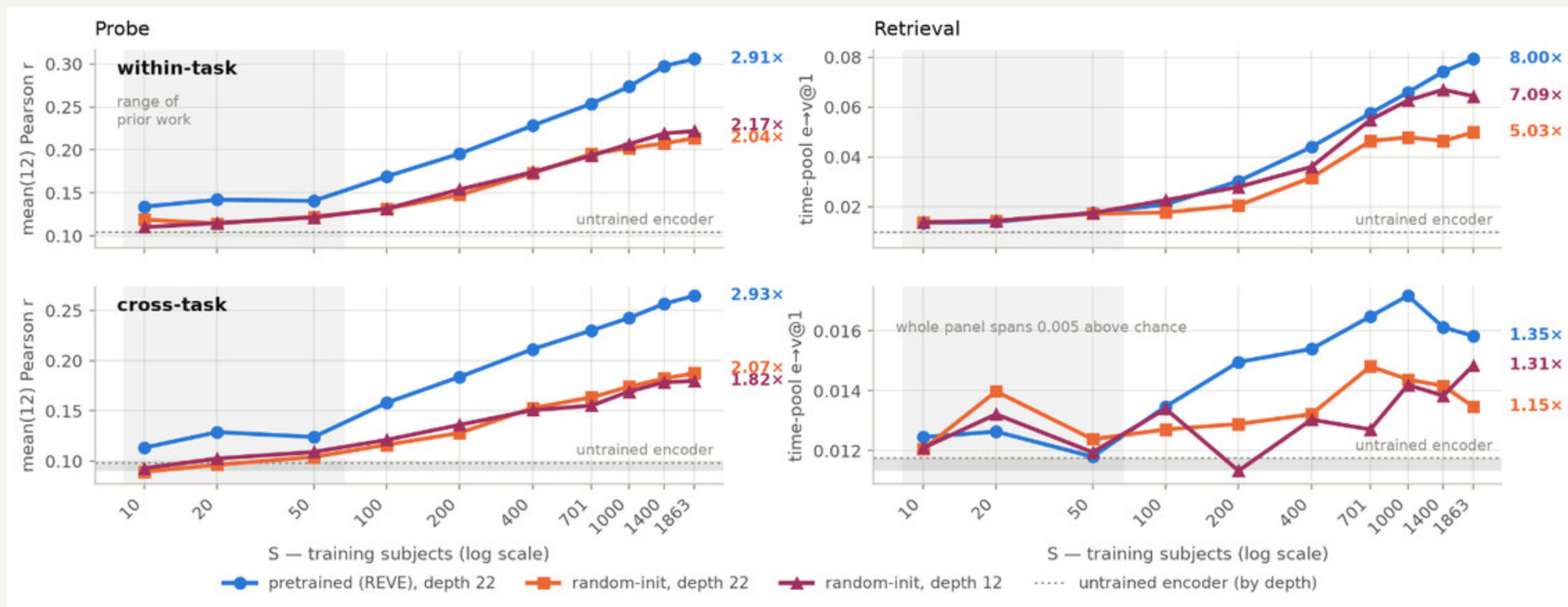


8 个数据集
77 名独立个体
93 条参与记录

脑区 token
+ 个体条件
→ 共享语义空间

CLIP 是监督目标；共享几何仍需跨目标验证

人数规模有用，但存在起点



约 50 人后出现稳定规模收益；蓝线为预训练 EEG encoder

换任务后，两种迁移结果不同

验收方式	允许重新学习什么	未见电影上的结果
冻结后直接检索	不使用新任务标签重拟合	仍接近机会水平
重新拟合线性 probe	只学习新的轻量读出	表示可用，随人数改善

编码器能力

≠

固定标定关系的适用范围

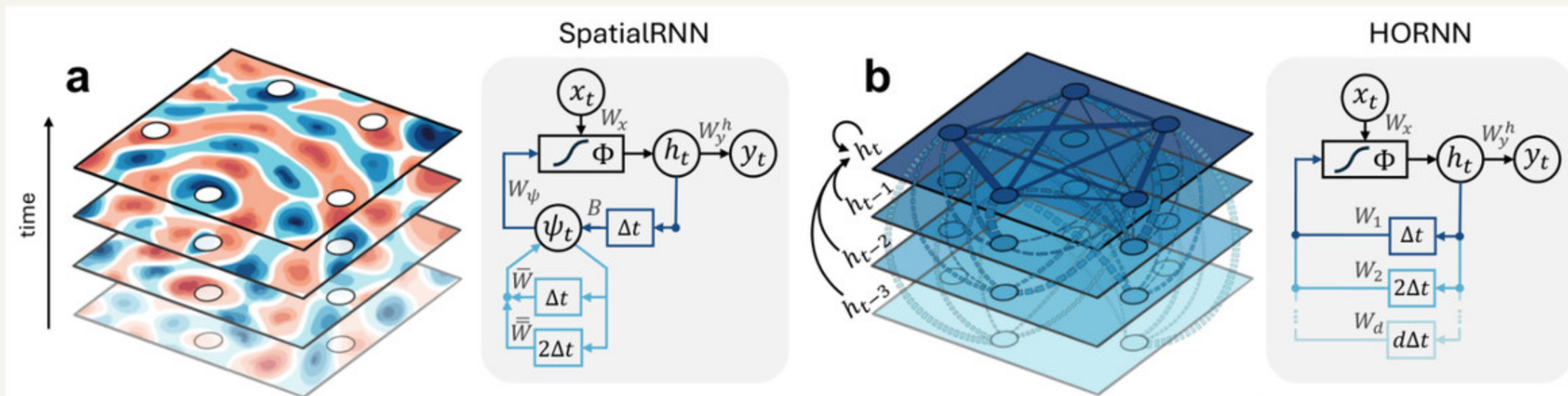
“能迁移”必须说明：允许改哪些参数、用了哪些标签

抗漂移，也要保留真实变化

实验操纵	理想响应	失败意味着什么
同任务，换日期 / 传感器	相关表示保持稳定	测量漂移没有被隔离
同设备，换任务 / 干预	相关表示随状态改变	真实变化被过度抹平

稳定性与敏感性要同时验收

空间场也能承载历史



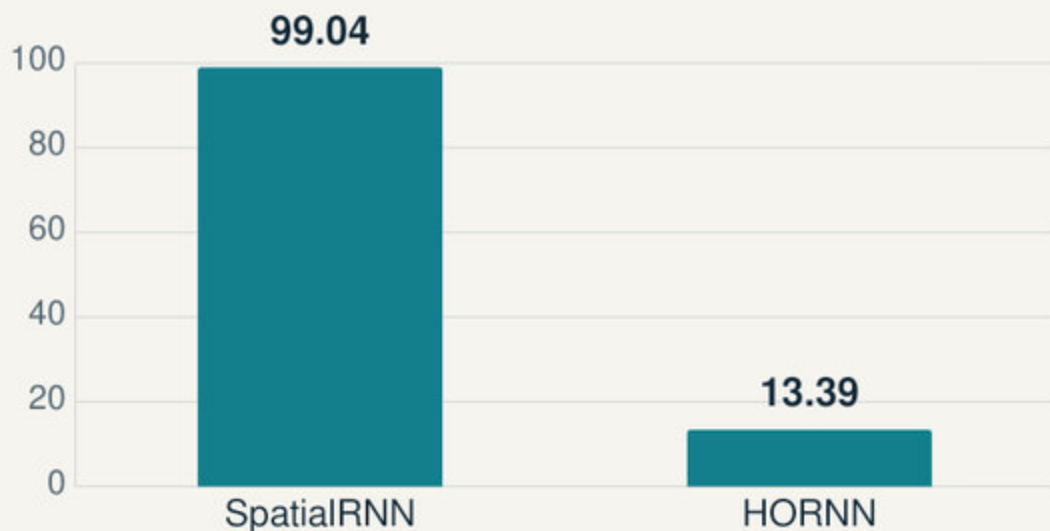
过去输入的影响，可以保留在場的振幅、相位与传播模式中

原图为模型示意；波纹叠层不是实验录像

无限历史依赖 $\neq$ 无限存储容量

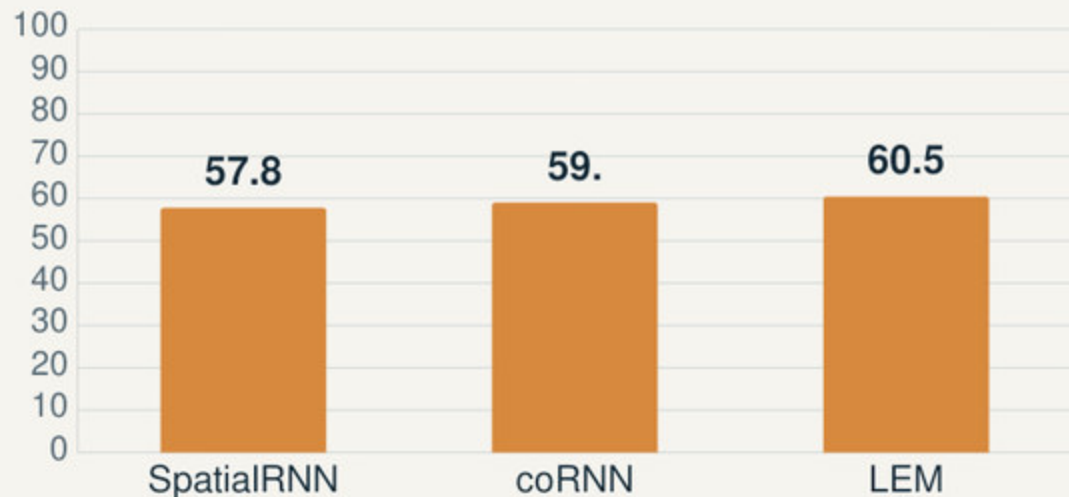
6,815 参数 vs 17,744 参数

复制任务准确率 % · T=1,024



第二项任务没有全面胜出

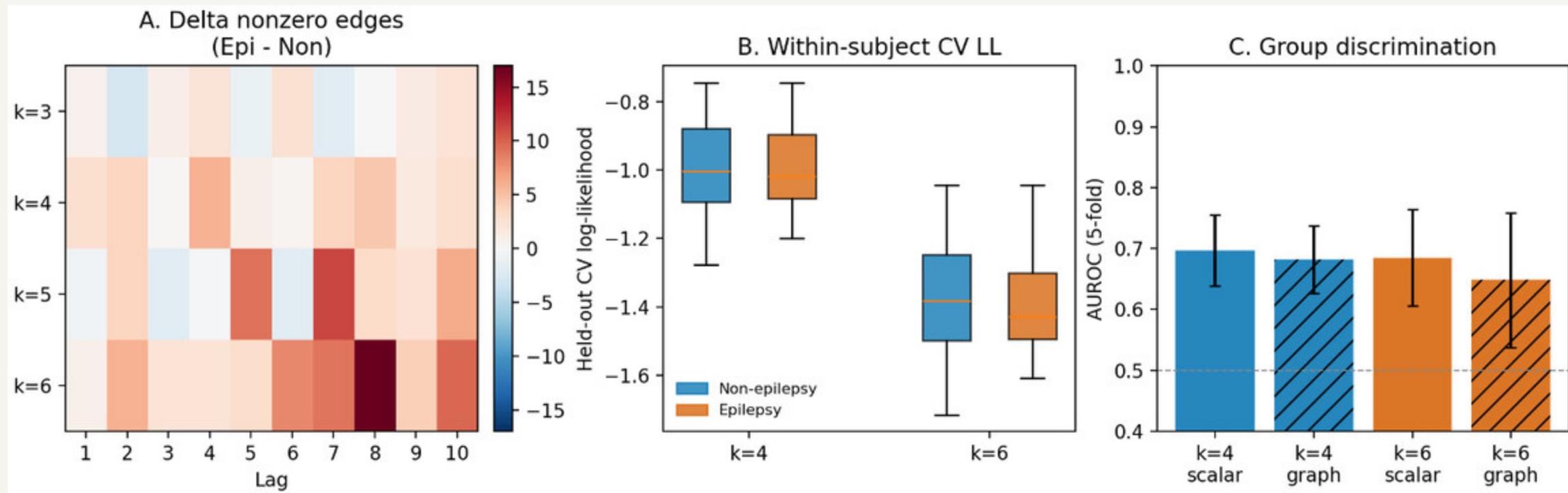
图像任务准确率 % · T=1,000



下一步：与现代 SSM 做同参数、同训练预算比较

相似预测精度，不同内部动力学

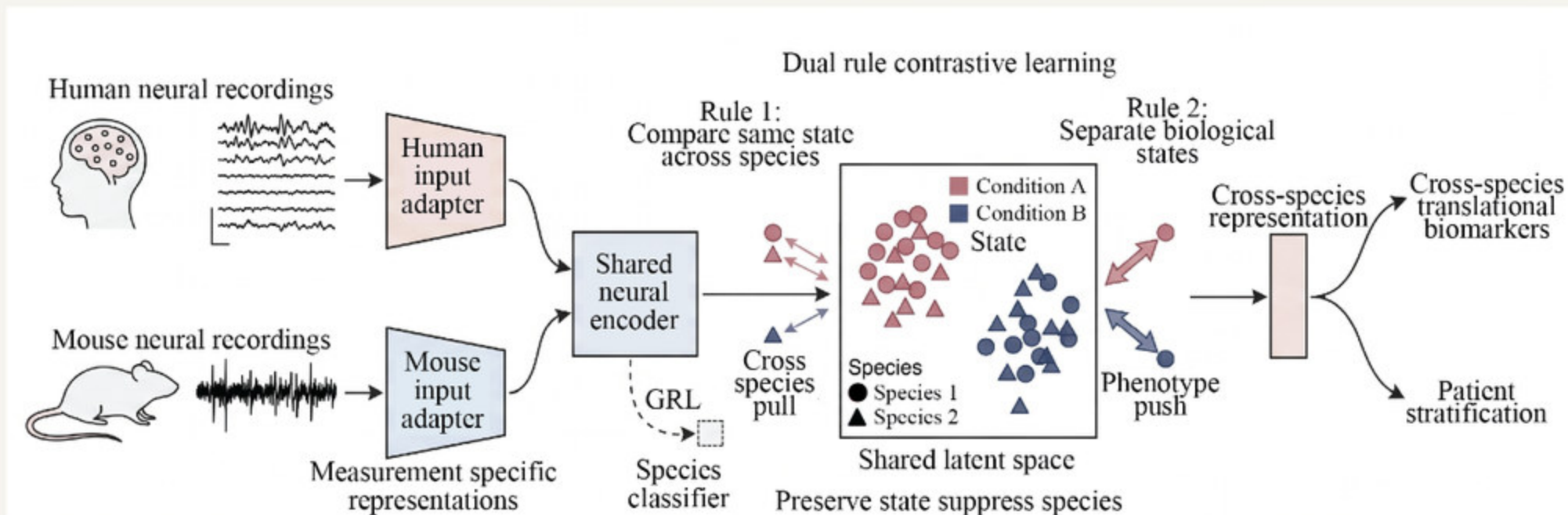
198 人 · 相似的 next-state fit · 不同的转移依赖结构



还要看状态转移、停留时间与干预后的恢复过程

跨物种对齐，关注干预方向

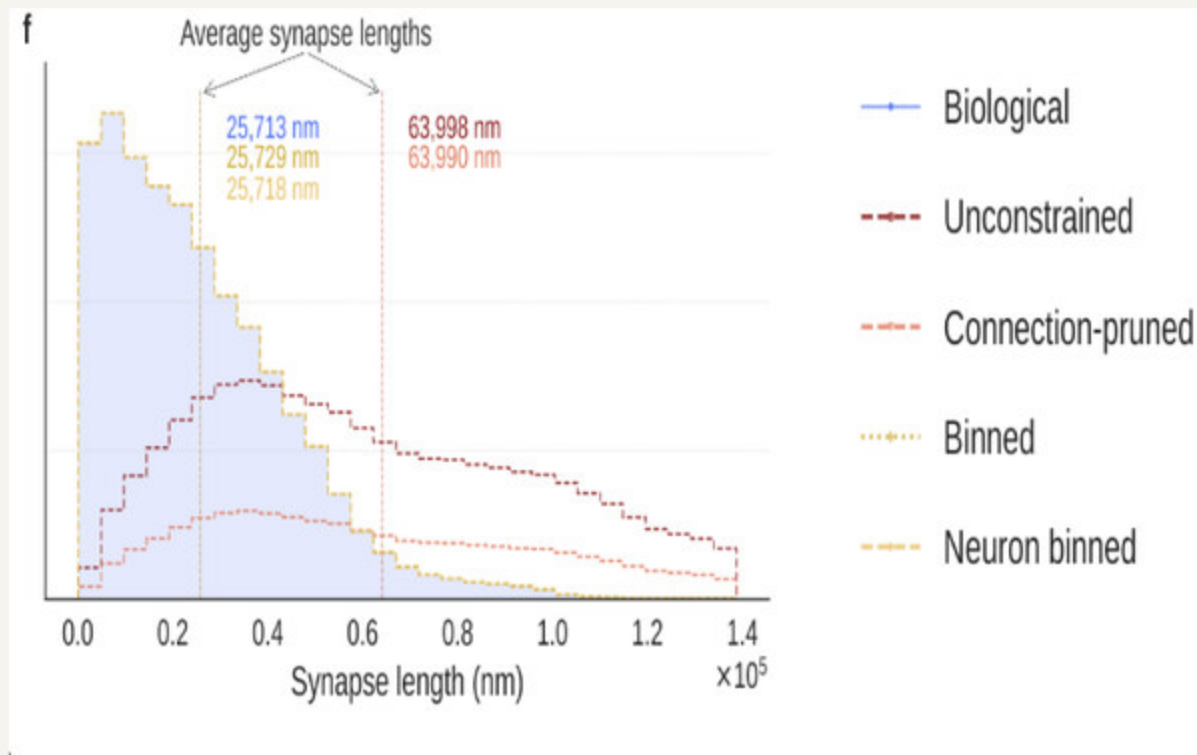
同状态跨物种靠近；不同状态仍要分开



药效排序 $\rho=0.87$ ，只有 10 个药物—模型组合

下一步：冻结模型后，前瞻、盲法预测未见药物

连接组优势取决于布线约束



形状任务

生物图 $\approx 64\%$
无约束随机图 $\approx 70\%$

无约束随机图
也使用了更多长程布线

结构优势必须放在准确率、通信距离和成本的共同预算中

稀疏模型仍然可能映射失败

连接概率	实际映射开始丢突触	10,000 神经元时
$p = 0.01$	约 6,000 神经元	突触损失 >10%
$p = 0.05$	约 4,000 神经元	突触损失 >25%

$p=0.01$ 的解析估计：约 13,216 个神经元才出现明显损失

存得下不等于接得上：映射器、开关与路由都是架构变量

训练稳定器可以不进入部署路径

假设动作价值	向左	向右	停止
原来的 Q 值	5	3	1
共同加上偏移 +10	15	13	11

动作排序没变，TD 学习目标却可能被污染

训练：辅助 ANN 修正共同偏移 / 部署：移除辅助路径

表中数值为解释用假设示例，不是论文实测

一个可检验的候选架构

层	承担什么	先定位哪类错误
测量与编码	模态、几何、时间戳	传感器或设备漂移
稳定坐标	结构与关系约束	跨日 / 跨人对齐失败
动态状态与历史	上下文、循环 / 场状态	长历史丢失
读出与控制	任务标定、动作、反馈	换任务后失效
训练支持	教师、校正、回滚条件	在线适应学偏

工程综合假说：模块先保持可替换，收益验证后再考虑硬化

闭环性能要计算整条时间链

时间链环节	要记录的量	常见混淆
采集窗口	历史跨度、步长、样本到齐时刻	快更新 $\neq$ 短观测窗口
传输 / 预处理	时间戳、缓冲、滤波延迟	只报告模型推理耗时
输出 / 反馈	执行、感知、刺激事件时间	离线准确率 $\neq$ 在线闭环性能

2 秒窗口 + 几毫秒推理 $\neq$ 几毫秒端到端响应

先做三个能失败的小实验

实验	改变什么	看什么结果
漂移实验	日期 / 传感器 与 任务分开操纵	稳定且保留真实变化
长历史实验	模型、空间相位、拓扑	同预算下有效记忆
迁移实验	整个人、整天、整设备留出	零校准与少量校准的差距

先得到可重复的软件收益，再计算路由、能耗和时延

哪一层，值得保持稳定？

传感器编码器
内部状态空间
最后的动作读出

答案决定：采什么数据、怎样校准、如何验证升级

SYSTEM BOUNDARY · 图解讲义与原始论文见节目页面

