

# 面向认知表现预测的时-空共变混合深度学习模型

李晴<sup>1,2</sup> 徐雪远<sup>1</sup> 邬霞<sup>1</sup>

**摘要** 认知表现预测已经成为当前大脑研究的重要课题. 功能磁共振成像技术由于同时具有较好的时间和空间分辨率, 有潜力为认知表现预测提供数据支持. 为了解决基于功能磁共振成像数据对认知表现进行预测时大脑所具有的时-空共变难刻画问题, 提出了一种新型基于大脑学习机制的时-空共变混合深度学习模型, 即深度稀疏自编码器与循环全连接网络混合模型, 以混合神经网络模型的损失函数误差作为认知表现预测能力的评价标准. 在人类连接组项目数据集上的实验结果表明, 提出的时-空共变混合模型能够有效和稳健地预测认知表现, 并提取到与人脑学习、记忆相关的有意义的脑影像特征, 从而为认知表现预测提供技术支持.

**关键词** 循环自编码器, 时-空共变深度学习模型, 混合深度学习模型, 认知表现预测, 脑启发模型

**引用格式** 李晴, 徐雪远, 邬霞. 面向认知表现预测的时-空共变混合深度学习模型. 自动化学报, 2022, 48(12): 2931-2940

**DOI** 10.16383/j.aas.c220025

## Spatio-temporal Co-variant Hybrid Deep Learning Framework for Cognitive Performance Prediction

LI Qing<sup>1,2</sup> XU Xue-Yuan<sup>1</sup> WU Xia<sup>1</sup>

**Abstract** Cognitive performance prediction has been an important topic for brain research. Functional magnetic resonance imaging is with high resolution in both spatial and temporal dimensions, which has the potential to support cognitive performance prediction. In order to address the problem that it is hard to characterize the spatio-temporal co-variation of brain data when predicting cognitive performance with functional magnetic resonance imaging data, inspired by the brain learning mechanism, a novel spatio-temporal co-variant hybrid deep learning framework has been presented here for evaluation the cognitive performance prediction, named as deep sparse recurrent autoencoder-recurrent fully connected net, to jointly minimize the loss function of the hybrid neural network models. The experimental results on the Human Connectome Project data set have shown that our proposed framework can predict cognitive performance and learn brain studying and memory-related neuroimaging features effectively and robustly, which can support predicting cognitive performance effectively.

**Key words** Recurrent autoencoder, spatio-temporal co-variant deep learning framework, hybrid deep learning framework, cognitive performance prediction, brain inspired model

**Citation** Li Qing, Xu Xue-Yuan, Wu Xia. Spatio-temporal co-variant hybrid deep learning framework for cognitive performance prediction. *Acta Automatica Sinica*, 2022, 48(12): 2931-2940

认知表现预测是指对大脑执行认知任务时学习能力、记忆力、注意力等方面能力的预测, 在教育发展、疾病诊疗等诸多领域具有重要的应用, 一直备受研究人员的关注<sup>[1]</sup>. 认知神经科学证据显示, 神经影像数据所蕴含的大脑信息能够直接反映出大脑认知表现的处理过程<sup>[2]</sup>. 因此, 利用神经影像技术对认知表现进行预测已经成为了人们关注的焦点<sup>[3-4]</sup>. 其

中, 同时具有较好的时间分辨率和空间分辨率的功能磁共振成像 (Functional magnetic resonance imaging, fMRI) 技术已经被广泛应用于构建上述关系.

已有多项基于 fMRI 数据的研究探究了脑功能与认知表现之间的关系, 并取得了不错的结果<sup>[5-6]</sup>. 深度学习作为一种强有力的预测模型, 已经被广泛应用于人脸识别、语言处理、物体检测等多个领域<sup>[7-8]</sup>. 受到已有工作启发, 多种深度学习模型被引入到基于 fMRI 数据的研究中, 如: 深度置信网络<sup>[9]</sup>、卷积神经网络<sup>[10]</sup>、循环神经网络 (Recurrent neural network, RNN)<sup>[11]</sup> 等. 深度学习具有突出的特征提取能力, 使其能够为建立脑功能与认知表现间的关系提供支持, 并进一步实现对认知表现的预测<sup>[12-13]</sup>.

收稿日期 2022-01-10 录用日期 2022-06-16

Manuscript received January 10, 2022; accepted June 16, 2022

北京市自然科学基金 (4212037) 资助

Supported by Natural Science Foundation of Beijing (4212037)

本文责任编辑 何晖光

Recommended by Associate Editor HE Hui-Guang

1. 北京师范大学人工智能学院 北京 100875 2. 北京师范大学  
认知神经科学与学习国家重点实验室 北京 100875

1. School of Artificial Intelligence, Beijing Normal University,  
Beijing 100875 2. State Key Laboratory of Cognitive Neurosci-  
ence and Learning, Beijing Normal University, Beijing 100875

研究表明, 大脑功能网络具有时-空共变的特性, 即不仅在空间上存在脑区协同的脑网络模式, 在时间上存在着动态变化的特性, 且不同空间网络与时间维度序列存在交互作用<sup>[14]</sup>. 受其影响, 大脑的认知表现也在空间维度中受到大脑功能网络的系统性影响, 在时间维度上受到重复性、衰减性等动态性影响<sup>[3]</sup>. 因此, 受大脑功能网络时-空共变特性的启发, 从时-空共变角度来预测认知表现, 将更加完善地对其进行信息处理与分析.

目前, 有研究者提出一种多任务学习模型, 采用最小化奇异值的方式对低秩公共子空间进行表征, 从而联合分析神经影像数据和认知行为数据, 实现了认知表现预测<sup>[15]</sup>. 然而, 以上研究依然未能在基于大脑学习机制的同时, 利用时-空脑数据对认知表现进行预测<sup>[16]</sup>, 从而限制了认知表现预测的能力. 因此, 将大脑学习机制引入到时-空共变深度学习认知表现预测模型中非常有必要. 最近, 一种学习机制启发的深度学习模型被提出, 即深度稀疏循环自编码器算法 (Deep sparse recurrent autoencoder, DSRAE). 该算法被用来对时-空共变 fMRI 特征进行分析, 验证了其时对-空共变脑数据的重建能力及对特征的高效分析能力<sup>[11, 17]</sup>. 在本研究中, 为了更好地预测认知表现能力, 本文结合 DSRAE 模型和循环全连接网络 (Recurrent fully connected

net, RFNet) 构建了一种时-空共变混合深度学习模型 (如图 1 所示), 称为 DSRAE-RFNet. 为了验证本文提出的模型, 本文采用人类连接组项目数据集 (Human connectome project, HCP)<sup>[18]</sup> 中被试执行工作记忆任务的反应时间 (Reaction time, RT) 和准确率 (Accuracy, ACC) 作为认知表现预测的目标.

本研究的目的是构建基于 DSRAE-RFNet 算法的认知表现预测模型. 在保证高预测准确率的同时, 获取有意义的神经影像预测特征作为神经学标记物. 本文的创新点在于:

1) 提出了一种新的时-空共变混合深度学习模型, 采用混合损失函数, 同时利用脑影像数据与认知表现数据对时-空共变混合神经网络模型进行训练. 算法面向认知表现预测需求, 基于 fMRI 数据特点, 利用高维度时-空共变神经影像特征对认知表现进行预测. 相较单一损失函数算法, DSRAE-RFNet 算法以高维度时-空共变脑影像特征为桥梁, 建立了神经影像数据与认知表现之间的关系.

2) 引入了人类大脑学习机制, 能够有效借助大脑变化中的时序依赖关系, 利用神经影像高维度时-空共变特征, 实现认知表现的精准预测. 采用定量分析的方式, 证明了 DSRAE-RFNet 算法作为一种脑启发算法模型, 在认知表现预测方面具有明显

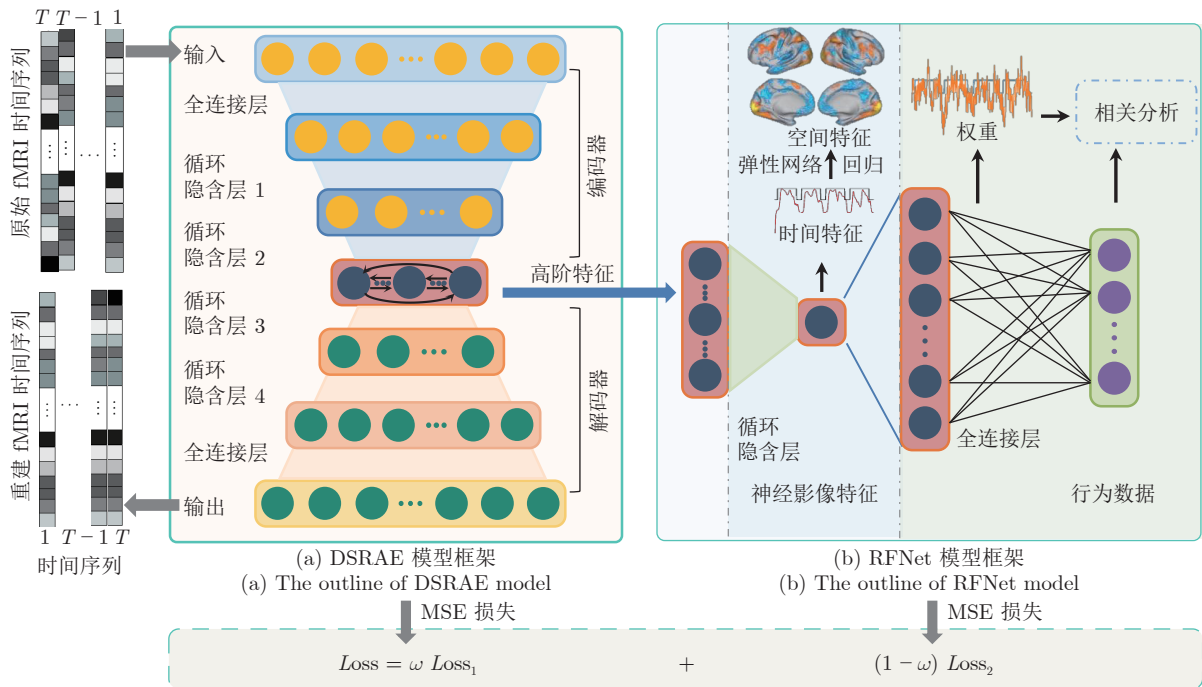


图 1 基于大脑学习机制的时-空共变混合 DSRAE-RFNet 模型框架

Fig.1 The overview of learning mechanism based spatio-temporal co-variant hybrid deep learning framework (DSRAE-RFNet)

优势. 从认知神经科学的角度, 证明了大脑学习机制对认知表现预测的重要性.

3) 在 DSRAE 算法的基础上, 加入了 RFNet 模型, 构建了新的能够有效提取大脑高维度时-空共变特征的认知表现预测模型. 面向 fMRI 数据特点, 利用长短时记忆 (Long short term memory, LSTM) 层获取包含数据有利信息的高维度时-空共变特征, 并利用全连接层对认知表现进行预测. 作为一种受 fMRI 数据本身特点驱动的计算模型, DSRAE-RFNet 模型不仅能够鲁棒地预测认知表现, 而且可以提取到有意义的神经影像预测特征. 该混合学习模型为揭示人类大脑认知机制提供了新思路.

## 1 DSRAE-RFNet 算法模型

### 1.1 模型框架

DSRAE-RFNet 模型框架如图 1 所示, 在进行认知表现预测的同时对相关神经影像特征进行分析. 第 1.2 节描述了 DSRAE 模型, 第 1.3 节描述了 RFNet 模型, 第 1.4 节详细介绍了时-空共变混合深度学习模型. 其中, DSRAE 模型的评价指标为 fMRI 数据的重建错误, RFNet 模型的目的则是将回归预测误差最小化.

### 1.2 DSRAE 模型

为了能够实现时-空共变网络的提取, 进而进行认知表现预测, DSRAE 模型将 LSTM 和自编码器 (Autoencoder, AE) 相结合<sup>[7]</sup>, 有效利用 LSTM 能够捕捉时间序列间时序依赖关系的优势和 AE 能够有效利用空间维度信息的优势, 采用无监督学习的框架对 fMRI 数据进行分析. 图 1(a) 展示了 DSRAE 模型的整体框架, 包含编码器和解码器两部分, 无需任何时序标签即可学习到任务特异性的时-空共变网络特征. 其中, 编码器包含一层输入层、一层全连接层 (可保留大部分空间维度信息) 和两层 LSTM 层; 解码器包含两层 LSTM 层、一层全连接层和一层输出层. 编码器经过全连接层 (128 个节点)、LSTM 层 (64 个节点), 将输入数据的空间维度编码为 32 维特征, 时间序列长度为  $T$ , 在每一层保持不变, 有效提取时-空共变特征. 解码器则与编码器完全相反, 将 32 维特征解码为原始 fMRI 的空间维度. DSRAE 部分的输入和输出都是 fMRI 时间序列, 其隐含层传导特征到 RFNet 模型.

### 1.3 RFNet 模型

图 1(b) 展示了 RFNet 模型, 共包含一层 LSTM

层, 一层全连接层. 从 DSRAE 的隐含层获得的 32 个时序特征经过一层 LSTM 层获得一个唯一的高维特征, 该特征包含了每个时间点的全部空间信息, 作为高阶时-空共变特征. 将神经影像特征与认知行为特征通过全连接层连接, 根据神经影像数据特征的时间维度与认知表现数据维度, 权重矩阵的维度为  $405 \times 10$ . 下面利用相关分析构建基于时-空共变信息的认知表现预测模型. 特别地, 本文采用模型在全连接层的权重与任务设计之间的皮尔逊相关系数 (Pearson correlation coefficient, PCC) 与行为数据进行相关分析, 构建预测模型.

### 1.4 时-空共变混合学习模型

本研究采用如图 1 所示的时-空共变混合学习模型, 基于神经影像数据, 对被试执行不同任务刺激下的认知表现的整体趋势进行预测.

具体来说, DSRAE-RFNet 整体模型在训练阶段的优化公式为:

$$J(\mathbf{W}, \mathbf{b}) = MSE(\mathbf{x}|\hat{\mathbf{x}}) + MSE(\mathbf{x}_f|\mathbf{y}) + \Omega_1 \|\mathbf{O}\|_1 + \Omega_2 \|\mathbf{O}\|_2^2 \quad (1)$$

式中,  $MSE(\mathbf{x}|\hat{\mathbf{x}})$  为 DSRAE 模型的误差函数,  $MSE(\mathbf{x}_f|\mathbf{y})$  为 RFNet 模型的误差函数. 二者均基于均方误差 (Mean square error, MSE) 损失方程构建:

$$L(\boldsymbol{\theta}, g(f(\boldsymbol{\rho}))) \quad (2)$$

式中,  $L$  表示  $g(f(\boldsymbol{\rho}))$  与  $\boldsymbol{\theta}$  之间差异的损失函数. 对于 DSRAE 模型,  $\boldsymbol{\theta}$  和  $\boldsymbol{\rho}$  均为 fMRI 一个时间点上的全脑数据;  $f(\boldsymbol{\rho})$  为编码阶段输出 (编码器第 2 层 LSTM 的节点输出),  $g(f(\boldsymbol{\rho}))$  为解码阶段输出 (DSRAE 解码器最后一层的输出). 对于 RFNet 模型,  $\boldsymbol{\rho}$  为从 DSRAE 模型隐含层获得的时间维度特征,  $\boldsymbol{\theta}$  为认知表现数据;  $f(\boldsymbol{\rho})$  为 RFNet 的 LSTM 层输出,  $g(f(\boldsymbol{\rho}))$  为全连接层输出. 式 (1) 中,  $\mathbf{O}$  表示模型的整体输出.  $\Omega_1$  和  $\Omega_2$  分别为  $L_1$  和  $L_2$  正则惩罚项, 用以避免整个训练过程过拟合.

基于 DSRAE 和 RFNet 模型的均方误差损失加权和, 时-空共变混合学习模型的损失函数定义为:

$$Loss = \omega Loss_1 + (1 - \omega) Loss_2 \quad (3)$$

式中,  $Loss_1$  和  $Loss_2$  分别代表 DSRAE 和 RFNet 模型的均方误差损失,  $\omega$  为 DSRAE 模型的均方误差损失对整个时-空共变混合学习模型的占比权重.

DSRAE-RFNet 算法的主要计算流程如算法 1 所示.

**算法 1.** 基于大脑学习机制的时-空共变混合深度学习模型

**输入.** 数据矩阵  $H \in \mathbf{R}^{s \times t}$  // 输入的 fMRI 数据.

**输出.** 预测矩阵  $B' \in \mathbf{R}$  // 认知表现预测结果、 $S \in \mathbf{R}^{s \times h}$  // 基于弹性网络回归获得的空间网络模式、 $T \in \mathbf{R}^{h \times t}$  // 时间序列.

**循环.**

1) 更新预测矩阵  $B'$ .

2) 根据输入矩阵  $H$ 、标准化认知表现数据  $B$ 、式 (1) 和式 (3) 更新网络模型.

3) 直到收敛或达到规定重复次数.

## 2 算法评估

### 2.1 实验数据

本文使用 750 名神经影像数据 (如表 1 所示) 和行为数据 (如表 2 所示) 均完备的正常成年被试数据作为训练集、验证集和测试集. 所使用的数据来自国际公开数据集 HCP (<https://db.humanconnectome.org>). HCP 数据集采用国际标准的蒙特利尔神经学研究所坐标系统空间对大脑灰质信息进行建模, 该坐标系统空间是根据一系列正常人脑的磁共振图像建立的. 数据预处理步骤包括: 空间维度平滑、时间维度滤波、回归去除脑脊液等干扰信息、以及头动校正. 所有预处理步骤均为 HCP 数据集基于 FreeSurfer 软件<sup>[19]</sup> 完成的.

表 1 工作记忆任务: fMRI 数据信息

Table 1 Working memory task: fMRI data information

参数	任务信息
时间点数量	405 个
扫描持续时间	5 分 01 秒
任务组块数量	8 个
刺激名称	0-back: 被试判断当前呈现内容是否 与预先规定内容一致 2-back: 被试判断当前呈现内容是否 与 2 个位置之前的呈现内容一致 cue: 任务初始阶段或 block 间的间隔

本研究采用工作记忆任务数据作为认知表现预测数据集, 该任务采用  $N$ -back 的方式对被试的工作记忆能力进行评测<sup>[20]</sup>, 任务刺激包括 0-back、2-back 和 cue. 该任务已被证明在不同被试<sup>[20]</sup> 和不同时间<sup>[21]</sup> 均能够激发出较为稳定的大脑活动模式. 任务信息如表 1 所示, 其中, 任务组块表示将相同条件的几个行为实验安排在一起. 扫描信息如下: 矩阵大小为  $90 \times 104$ 、切片数为 72、脉冲重复时间为 0.72 s、回波时间为 33.1 ms、面内视野为  $208 \times 180$  mm、翻转角为  $52^\circ$ 、频带宽度为 2290 Hz/Px、空间分辨率为  $2 \times 2 \times 2$  mm<sup>3</sup>. 被试的反应时间和准确率的 10 项具体认知表现数据信息如表 2 所示.

表 2 工作记忆任务: 认知表现数据信息

Table 2 Working memory task: Cognitive performance data information

刺激	反应时间 (RT)	准确率 (ACC)
0-back	总体反应时间	总体准确率
	人体刺激反应时间	人体刺激准确率
	面孔刺激反应时间	面孔刺激准确率
	地点刺激反应时间	地点刺激准确率
	工具刺激反应时间	工具刺激准确率
2-back	总体反应时间	总体准确率
	人体刺激反应时间	人体刺激准确率
	面孔刺激反应时间	面孔刺激准确率
	地点刺激反应时间	地点刺激准确率
	工具刺激反应时间	工具刺激准确率

### 2.2 算法细节

特别地, 本文利用 Z-值<sup>[22]</sup> 标准化 (如式 (4)) 将 fMRI 数据标准化为均值为 0、标准差为 1 的矩阵:

$$x^* = \frac{x - \bar{x}}{\sigma} \quad (4)$$

式中,  $x$  为原始数据,  $\bar{x}$  为原始数据的均值,  $\sigma$  为原始数据的标准差,  $x^*$  为标准化后的数据. 为了统一认知表现数据的单位, 本文将反应时间数值均/1000, 准确率数值均/100.

本研究采用 HCP 数据集中参与工作记忆任务的 750 名被试数据作为实验数据. 本文将其随机平均分为 3 组, 每组包含 250 名被试. 其中, 每组被试中的 150 名被试作为训练集, 50 名作为验证集, 50 名作为测试集. 第 2.3 节所呈现的结果均基于测试集获得. 反应时间与准确率的认知表现预测结果作为每个组别对应的验证评价指标.

DSRAE-RFNet 采用 Adam 优化器<sup>[23]</sup> 来处理模型中的后向传播过程:

$$\Delta\theta_t = \frac{-\hat{m}_t}{(\sqrt{\hat{n}_t + \epsilon}) \cdot \eta} \quad (5)$$

式中,  $\hat{m}_t$  和  $\hat{n}_t$  分别是  $m_t$  和  $n_t$  的无偏估计,  $\eta$  为学习率, 梯度的一阶矩估计  $m_t$  和二阶矩估计  $n_t$  分别为:

$$m_t = \mu \cdot m_{t-1} + (1 - \mu) \cdot g_t \quad (6)$$

$$n_t = \nu \cdot n_{t-1} + (1 - \nu) \cdot g_t^2 \quad (7)$$

式中,  $\mu$  和  $\nu$  为势能因子,  $g_t$  为梯度.

模型的参数如下: 学习率设置为 0.01, 批大小为 1 (包含 1 次回看步数内的全部信息),  $L_1$  和  $L_2$  正则化参数  $\Omega_1$  和  $\Omega_2$  分别为  $1/10^7$  和  $1/10^4$ <sup>[11]</sup>. 为了比较学习机制对 DSRAE-RFNet 模型的影响, LSTM 的回看步数根据是否采用学习机制设置为 405 (每个被试的重复时间总数量) 和 1 (单个重复时间).

重复次数设置为 10 (若验证集更早收敛, 则采用早停机制直接停止, 如 5~6 个重复次数). 如图 2 所示, 无论是对准确率还是反应时间进行预测, 在 3 组数据集上, 基于 MSE 的验证集损失在 10 个重复次数内均能达到收敛状态. DSRAE-RFNet 模型整体损失中, DSRAE 模型的均方误差损失对整个时-空共变混合学习模型的占比权重  $\omega$  设置为 0.5, 从而有效平衡 DSRAE 模型对神经影像特征的提取精确度和 RFNet 对认知表现的预测精度. 如图 3 所示, 当  $\omega = 0.5$  时 (即神经影像数据和认知表现数据的混合学习比重一致), DSRAE-RFNet 模型对 RT 和 ACC 的预测效果最为稳定.

为了进一步揭示神经影像数据对认知表现预测的认知神经科学依据, 并有效避免噪声对空间网络激活模式生成过程的影响, 本文通过弹性网络回归<sup>[24-26]</sup>, 基于 RFNet 中获得的时序特征生成空间网络:

$$\hat{\beta} = \arg \min_{\beta} \|\mathbf{x}_{f1} - \mathbf{X}\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2 \quad (8)$$

式中,  $\mathbf{X}$  为预测矩阵 (空间网络),  $\mathbf{x}_{f1}$  为 RFNet 的一维高阶特征 (时序特征).  $\lambda_1$  和  $\lambda_2$  分别是弹性网络回归的  $L_1$  和  $L_2$  正则化参数, 能够有效避免噪声、奇异值等的影响, 本文  $L_1$  比例设置为 0.005.

### 2.3 结果与分析

本文将 DSRAE-RFNet 模型在 3 组参与了工作记忆任务的被试上进行了训练、验证和测试, 预测被试的反应时间和准确率. 结果显示, 本文提出

的时-空共变混合深度学习模型能够通过 fMRI 数据有效预测认知表现能力, 并能够同时获得相关的时-空共变脑网络.

#### 2.3.1 学习机制分析

为了对学习机制启发的 DSRAE-RFNet 模型进行验证, 本文在 3 组被试上分别进行了采用学习机制和未采用学习机制的模型测试. 研究表明, 大脑在学习新信息或处理新任务时, 会基于原有时间序列获得时序信息经验, 即便是较长时间间隔之前的信息, 依然会对当前时刻大脑的信息处理能力产生一定程度的影响<sup>[27-29]</sup>. 因此, 本文将回看步数设置为完整 fMRI 时间序列长度, 即 405. 相反地, 回看步数设置为 1 代表模型没有基于人脑学习机制.

表 3 展示了 RT 和 ACC 在 3 组被试测试集上的预测结果. 对于 RT 预测, 使用了学习机制的预测结果在 3 组被试上均比未使用学习机制的模型结果好 (采用双样本  $t$  检验,  $p$  值  $< 0.01$ ).

特别地, 对于第 2 组被试, 使用了学习机制的模型在认知表现预测上的性能具有 0.41 的优势, 甚至超过了未使用学习机制的模型预测结果两倍. 对于 ACC 预测, 采用了学习机制的模型预测结果总体来说高于没有采用学习机制的模型. 对于第 1 组被试, 采用学习机制的模型预测结果没有展现出更多优势, 但在统计学上二者没有差异. 这可能是由于第 1 组被试整体在各个刺激条件下的 ACC 变化不明显 (详见第 2.3.3 节), 导致预测模型敏感性较低.

#### 2.3.2 反应时间表现预测

如表 3 所示, 3 组被试的 RT 预测结果分别是



图 2 DSRAE-RFNet 模型在 3 组被试上对 RT、ACC 分别进行预测时的 MSE 损失图

Fig. 2 The MSE loss when predicting RT and ACC with DSRAE-RFNet model on three groups participants

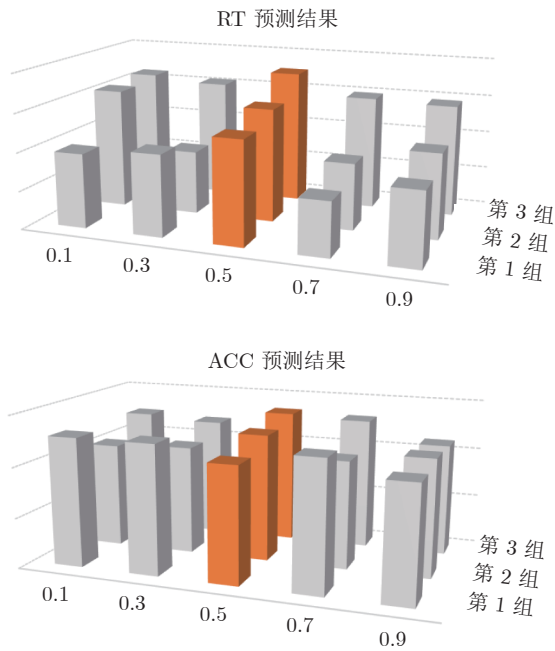


图 3 DSRAE-RFNet 模型在 3 组被试上对 RT、ACC 的预测结果

Fig. 3 The predictive results of RT and ACC with DSRAE-RFNet model on three groups participants

表 3 与未采用学习机制模型比较的认知表现预测结果  
Table 3 Cognitive performance prediction results compared with the model without learning mechanism

认知表现	组别	没有采用学习机制	采用学习机制
RT	第 1 组	0.455	<b>0.700</b>
	第 2 组	0.330	<b>0.740</b>
	第 3 组	0.680	<b>0.776</b>
ACC	第 1 组	<b>0.449</b>	0.429
	第 2 组	0.388	<b>0.477</b>
	第 3 组	0.523	<b>0.536</b>

0.7、0.74 和 0.776。图 4 展示了 RFNet 中全连接层的权重可视化结果及其与 RT 表现之间的关系。对于每一组被试，RFNet 的全连接层建立了该组被试的 RT 数据与神经影像数据之间的潜在关系。如图 4(a) 所示，针对 RT 进行预测时，无论是第几组被测试，RFNet 全连接层的 10 组权重均与任务设计之间表现出明显的正相关性，表明本文的模型所学习到的连接权重具有神经影像学意义。对于 RT 的表现预测包含了 10 个预测条目（如表 2 所示），图 4(b) 展示了上述连接权重与任务设计之间的 PCC 系数与 RT 的认知表现数据间所具有的正相关性，分别为 0.862、0.665 和 0.767。以上结果表明，预测模型学习到的特征同时具有神经科学意义和认知科学意义，也证明了基于大脑学习机制的时-空共变混合

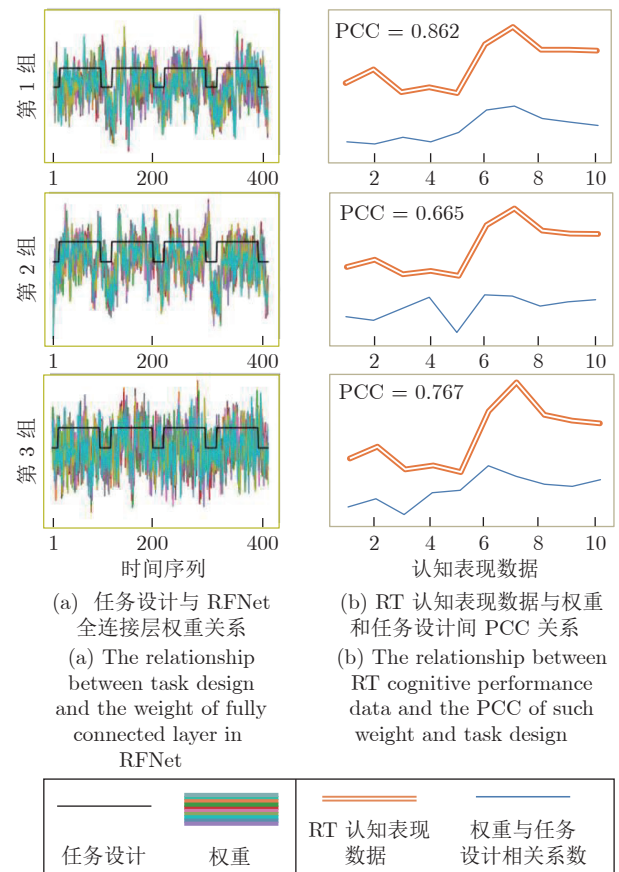


图 4 反应时间表现预测结果

Fig. 4 Reaction time performance predicting results

深度学习模型能够有效建立神经影像脑数据与认知表现之间的关系，为基于 fMRI 数据对认知表现进行预测提供了有力证据。

为了更进一步验证基于大脑学习机制的时-空共变混合深度学习模型在预测 RT 时能够学习到有意义的神经影像学特征，图 5(a) 展示了相关的时间维度和空间维度特征。对于时间维度特征，在对被试 RT 进行预测时，DSRAE-RFNet 学习到的时间序列与任务设计具有高度相关性，在 3 组被试上分别为 0.568、0.624 和 0.325。这表明，提取到的高维时-空共变特征与工作记忆任务刺激波动具有一致性。对于空间维度特征，在对被试执行工作记忆任务的 RT 进行预测时，3 组被试都一致显著地正激活了背侧注意网络和视觉网络。基于广义线性模型 (General linear model, GLM)<sup>[18]</sup>、荟萃分析<sup>[30]</sup>和深度学习<sup>[11, 17, 31]</sup>等研究的结果均表明，两个网络的相关脑区与工作记忆任务具有高度相关性；此外，默认网络显示出了明显的负激活，这与基于 GLM 生成的空间网络非常相似<sup>[18, 32]</sup>，也就是说在执行认知任务过程中，DMN 处于负激活状态。结果证明在 RT 预测过程中所提取到的时-空共变特征是具有

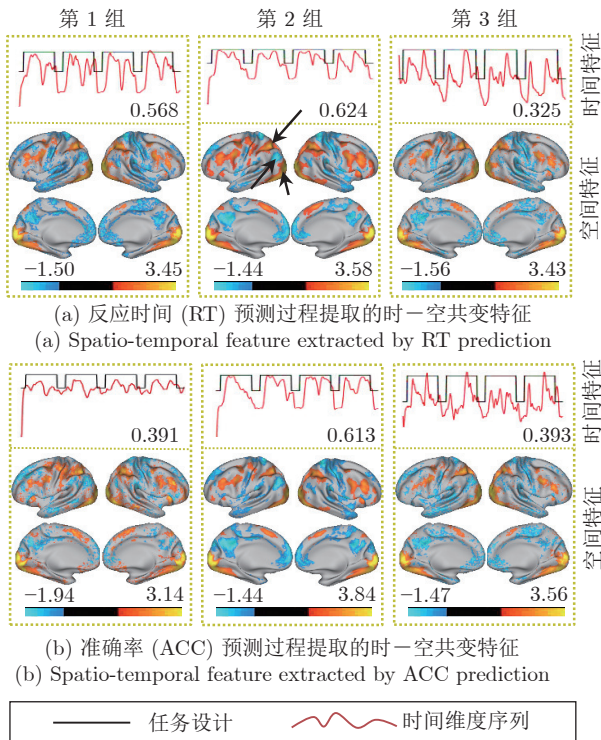


图 5 反应时间与准确率表现预测过程中习得的大脑时-空共变特征

Fig.5 The brain spatio-temporal co-variant features learned from the RT and ACC performance prediction processes

认知神经科学意义的。

### 2.3.3 准确率表现预测

如表 3 所示, 对于 3 组被试的 ACC 表现预测, 基于大脑学习机制的 DSRAE-RFNet 模型的预测准确率分别是 0.429、0.477 和 0.536。在对 ACC 表现进行预测时, 3 组被试的全部 10 个条目 (如表 2 所示) 的全连接层权重均与任务设计具有正相关性 (如图 6(a) 所示), 表明模型学习到的权重具有明显的实际意义。此外, 模型学习到的权重与任务设计 PCC 和 ACC 的认知表现数据具有高度正相关性, 对于 3 组被试分别为 0.499、0.439 和 0.666 (如图 6(b) 所示)。

为了进一步验证预测过程中所学习到的神经影像特征具有实际意义, 本文同样将预测 ACC 表现过程中习得的时-空共变网络特征进行了可视化 (如图 5(b) 所示)。与 RT 预测结果相似, 提取到的时序信息特征与工作记忆的任务设计具有高度相关性。同时, 空间维度网络激活模式与 GLM、荟萃分析及深度学习等算法所获得的空间网络非常相似<sup>[11, 17-18, 30-31]</sup>, 即背侧注意网络和视觉网络正激活, 默认网络负激活。以上结果表明, DSRAE-RFNet

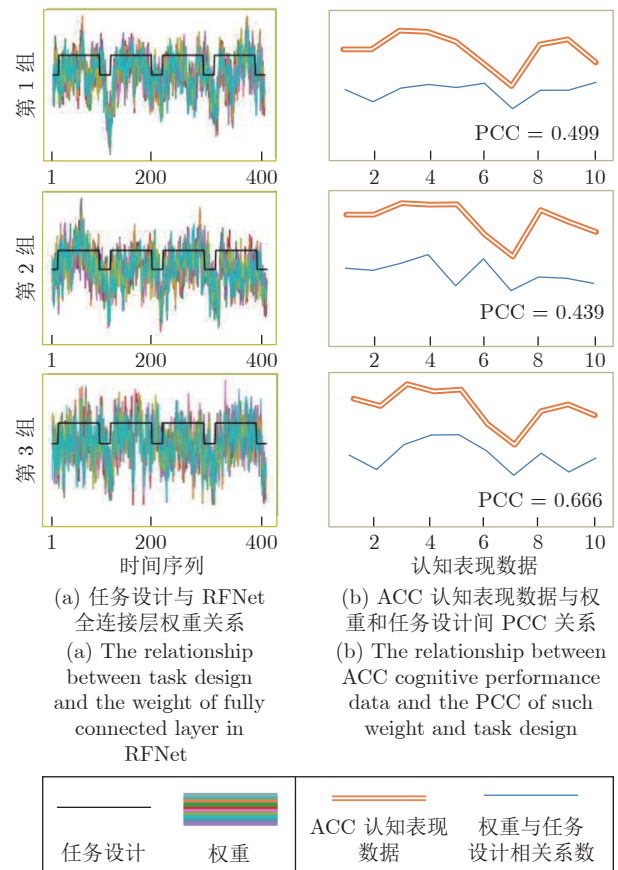


图 6 准确率表现预测结果

Fig.6 Accuracy performance predicting results

模型能够提取到具有认知神经科学意义的时-空脑网络, 从而实现对被试认知表现的预测。

### 2.3.4 算法性能比较

为了证明本文提出的 DSRAE-RFNet 算法的性能, 本文将 DSRAE-RFNet 与经典预测模型 (包括线性和非线性模型) 以及前沿算法进行了比较, 包括: 独立成分分析 (Independent component analysis, ICA)、GLM、RNN、LSTM、AE、深度稀疏循环神经网络 (Deep sparse recurrent neural network, DSRNN)<sup>[33]</sup>、深度变分自编码器 (Deep variational autoencoder, DVAE)<sup>[34]</sup>、时空注意力自编码器 (Spatiotemporal attention autoencoder, STAAE)<sup>[35]</sup>、深度卷积自编码器 (Deep convolutional autoencoder, DCAE)<sup>[10]</sup>。其中, AE 与经典线性模型 ICA、GLM 相似, 提取空间网络模式特征进行认知表现预测; DSRNN<sup>[33]</sup> 与经典非线性模型 RNN、LSTM 相似, 主要对 fMRI 信号中的高维度时序特征进行提取, 从而实现认知表现预测; DVAE<sup>[34]</sup>、STAAE<sup>[35]</sup> 算法分别利用变分生成模块和注意力模块与 AE 相结合, 提取 fMRI 的空间高维度特征;

DCAE<sup>[10]</sup> 算法以 fMRI 数据中的体素为单位, 提取时间维度高维特征作为预测指标. 如表 4 所示, 相较经典和前沿各预测模型, DSRAE-RFNet 在 RT 和 ACC 的预测能力上均表现出较大优势. 对于第 1、3 组被试, 预测能力均达到最优水平, 证明了本文算法在预测能力上的有效性和稳定性.

表 4 与其他预测模型比较的认知表现预测结果

Table 4 Cognitive performance prediction results compared with the other predictive model

组别	RT			ACC		
	第 1 组	第 2 组	第 3 组	第 1 组	第 2 组	第 3 组
ICA	0.677	0.661	0.765	0.090	<b>0.505</b>	0.449
GLM	0.451	0.369	0.292	0.232	0.224	0.343
RNN	0.696	<b>0.747</b>	0.719	0.390	0.453	0.511
LSTM	0.685	0.674	0.731	0.369	0.368	0.345
AE	0.078	0.519	0.663	0.178	0.289	0.202
DSRNN <sup>[33]</sup>	0.692	0.576	0.733	0.413	0.354	0.356
DVAE <sup>[34]</sup>	0.157	0.644	0.512	0.411	0.242	0.216
STAAE <sup>[35]</sup>	0.588	0.431	0.532	0.066	0.122	0.466
DCAE <sup>[10]</sup>	0.384	0.438	0.465	0.225	0.200	0.089
本文算法	<b>0.700</b>	0.740	<b>0.776</b>	<b>0.429</b>	0.477	<b>0.536</b>

为了证明 DSRAE-RFNet 算法在认知表现预测上具有优势的同时, 在预测效率上同样具有优势, 本文在单个图形处理器 (Graphics processing unit, GPU) (RTX 2070) 上比较了不同算法的运行时间 (包括: 在特征提取阶段的单个运行次数内运行时间和在测试阶段的单个被试预测时间). 如图 7 所示, DSRAE-RFNet 算法的运行时间远低于 ICA 方法; 相较其他以 AE 为基础的前沿算法 (DVAE<sup>[34]</sup>、STAAE<sup>[35]</sup>、DCAE<sup>[10]</sup>), 本文所提算法同样具有明显优势; 本文算法与 GLM、RNN、LSTM、AE、DSRNN<sup>[33]</sup> 算法在单个 GPU 上的运行时间基本相似, 但对认知表现的预测效果更具优势 (如表 4 所示). 因此, 基于对各算法运行时间和预测准确率的综合分析, 本文所提 DSRAE-RFNet 算法是一种高效、鲁棒的新型认知表现预测算法.

### 3 结束语

本文提出了一种新型基于大脑学习机制的时-空共变混合深度学习模型 DSRAE-RFNet, 用来预测认知表现, 并同时提取相关的时-空共变神经学特征. 本文采用 HCP 数据集中的工作记忆任务作为本研究的训练、验证及测试集, 获得了非常有意义的结果.

为了证明本文提出的 DSRAE-RFNet 模型的

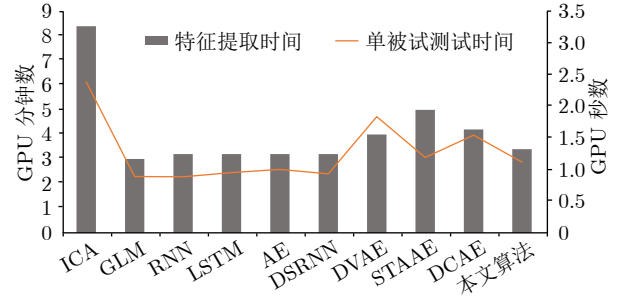


图 7 DSRAE-RFNet 及比较算法在单个 GPU 上单个运行次数的运行时间

Fig. 7 Running time of DSRAE-RFNet and comparable methods on a single GPU during one epoch

鲁棒性, 本文将模型扩展到了 HCP 数据集中提供了 RT 和 ACC 认知表现信息的情感任务、语言任务和关系任务上, 用于对被试的认知能力进行预测. 结果表明, DSRAE-RFNet 模型除了能够对被试执行工作记忆任务时的认知表现进行预测外, 还能够稳定地对多种认知任务的认知表现能力进行预测 (如表 5 所示). 其中, 相较其他三种认知任务, 情感任务的 RT 和 ACC 预测准确率相对偏低, 这可能与被试执行情感任务时对人脸和图形判断的相关认知表现差异并不明显有关<sup>[18]</sup>.

表 5 对不同认知任务的认知表现预测结果

Table 5 Cognitive performance prediction results for different cognitive tasks

认知表现	组别	情感任务	语言任务	关系任务	工作记忆任务
RT	第1组	0.160	0.720	0.667	0.700
	第2组	0.280	0.800	0.990	0.740
	第3组	0.240	0.720	0.760	0.776
ACC	第1组	0.400	0.520	0.693	0.429
	第2组	0.320	0.720	0.820	0.477
	第3组	0.140	0.760	0.840	0.536

总之, 作为一种面向认知表现预测需求的新型时-空共变混合深度学习模型, 本文提出的 DSRAE-RFNet 算法能够基于学习机制, 受 fMRI 数据本身特点驱动, 充分利用认知神经科学先验信息. 算法模型不仅能够基于 fMRI 数据预测认知表现, 同时能够获得认知任务有意义的神经影像学特征. 通过对 3 组被试的验证与测试, 证明了本文提出的 DSRAE-RFNet 模型的有效性和鲁棒性. 与经典线性、非线性以及前沿算法的比较结果同样证明, 本文提出的算法在预测效果和效率上均具有优势. 基于本文研究, 未来将在以下几方面进行研究: 1) 继续尝试其他混合深度学习模型, 如多核学习<sup>[36-38]</sup>方式, 采用自



适应的损失函数优化方式,以期更加有效地利用神经影像特征来预测认知表现; 2) 基于神经网络结构搜索框架,针对不同认知表现数据的预测任务,对 DSRAE-RFNet 中 DSRAE 部分的空间维度特征维度进行优化和搜索<sup>[39]</sup>,以期得到更加具有任务特异性的算法结构; 3) 本文在对大脑时-空共变特征进行分析时,空间维度特征是基于时间维度特征通过弹性网络回归分析得到的,尚未结合和利用深度学习模型. 未来将结合生成模型,对空间特征进行生成; 4) 由于 DSRAE-RFNet 模型并不依赖于认知任务的具体任务刺激,因此有潜力基于静息态 fMRI 数据对认知表现进行预测. 未来将基于静息态 fMRI 数据,利用 DSRAE-RFNet 模型对被试的决策能力等认知行为指标进行预测<sup>[40]</sup>.

### References

- Chhatwal J P, Schultz A P, Dang Y, Ostaszewski B, Liu L, Yang H S, et al. Plasma N-terminal tau fragment levels predict future cognitive decline and neurodegeneration in healthy elderly individuals. *Nature Communications*, 2020, **11**(1): 1–10
- Mayer R E. How can brain research inform academic learning and instruction? *Educational Psychology Review*, 2017, **29**(4): 835–846
- Sigman M, Pena M, Goldin A P, Ribeiro S. Neuroscience and education: Prime time to build the bridge. *Nature Neuroscience*, 2014, **17**(4): 497–502
- Cao P, Liu X, Yang J, Zhao D, Huang M, Zaiane O.  $L_2$ ,  $1-L_1$  regularized nonlinear multi-task representation learning based cognitive performance prediction of Alzheimer's disease. *Pattern Recognition*, 2018, **17**: 195–215
- Jiang X, Zhao L, Liu H, Guo L, Kendrick K M, Liu T. A cortical folding pattern-guided model of intrinsic functional brain networks in emotion processing. *Frontiers in Neuroscience*, 2018, **12**: 575
- Feng C, Yuan J, Geng H, Gu R, Zhou H, Wu X, et al. Individualized prediction of trait narcissism from whole-brain resting-state functional connectivity. *Human Brain Mapping*, 2018, **39**(9): 3701–3712
- Dargan S, Kumar M, Ayyagari M R, Kumar G. A survey of deep learning and its applications: A new paradigm to machine learning. *Archives of Computational Methods in Engineering*, 2019: 1–22
- Liu Xiao-Bo, Liu Peng, Cai Zhi-Hua, Qiao Yu-Lin, Wang Ling, Wang Min. Research progress of optical remote sensing image object detection based on deep learning. *Acta Automatica Sinica*, 2021, **47**(9): 2078–2089  
(刘小波, 刘鹏, 蔡之华, 乔禹霖, 王凌, 汪敏. 基于深度学习的光学遥感图像目标检测研究进展. *自动化学报*, 2021, **47**(9): 2078–2089)
- Hu X, Huang H, Peng B, Han J, Liu N, Lv J, et al. Latent source mining in fMRI via restricted boltzmann machine. *Human Brain Mapping*, 2018, **39**(6): 2368–2380
- Huang H, Hu X, Zhao Y, Makkie M, Dong Q, Zhao S, et al. Modeling task fMRI data via deep convolutional autoencoder. *IEEE Transactions on Medical Imaging*, 2018, **37**(7): 1551–1561
- Li Q, Dong Q, Ge F, Qiang N, Zhao Y, Wang H, et al. Simultaneous spatial-temporal decomposition of connectome-scale brain networks by deep sparse recurrent auto-encoders. In: Proceedings of the 2019 International Conference on Information Processing in Medical Imaging, Hongkong, China: 2019. 579–591
- Viejo G, Khamassi M, Brovelli A, Girard B. Modeling choice and reaction time during arbitrary visuomotor learning through the coordination of adaptive working memory and reinforcement learning. *Frontiers in Behavioral Neuroscience*, 2015, **9**: 225
- Gu Ling-Yun, Lü Wen-Zhi, Yang Yong, Gao Jun-Feng, Guan Jin-An, Zhou Dao. Deception detection study based on PCANet and support vector machine. *Acta Electronica Sinica*, 2016, **44**(8): 1969–1973  
(顾凌云, 吕文志, 杨勇, 高军峰, 官金安, 周到. 基于PCANet和SVM的谎言测试研究. *电子学报*, 2016, **44**(8): 1969–1973)
- Betzell R F, Bassett D S. Multi-scale brain networks. *NeuroImage*, 2017, **160**: 73–83
- Huo Z, Shen D, Huang H. New multi-task learning model to predict Alzheimer's disease cognitive assessment. In: Proceedings of the 2016 Medical Image Computing and Computer Assisted Intervention. Lima, Peru: 2016. 317–25
- Hassabis D, Kumaran D, Summerfield C, Botvinick M. Neuroscience-inspired artificial intelligence. *Neuron*, 2017, **95**(2): 245–258
- Li Q, Dong Q, Ge F, Qiang N, Wu X, Liu T. Simultaneous Spatial-temporal decomposition for connectome-scale brain networks by deep sparse recurrent autoencoder. *Brain Imaging and Behavior*, 2021, **15**: 2646–2660
- Barch D M, Burgess G C, Harms M P, Petersen S E, Schlaggar B L, Corbetta M, et al. Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage*, 2013, **80**: 169–189
- Glasser M F, Sotiropoulos S N, Wilson J A, Coalson T S, Fischl B, Andersson J L, et al. The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 2013, **80**: 105–124
- Drobyshevsky A, Baumann S B, Schneider W. A rapid fMRI task battery for mapping of visual, motor, cognitive, and emotional function. *NeuroImage*, 2006, **31**: 732–744
- Caceres A, Hall D L, Zelaya F O, Williams S C R, Mehta M A. Measuring fMRI reliability with the intra-class correlation coefficient. *NeuroImage*, 2009, **45**(3): 758–768
- Colan S D. The why and how of Z scores. *Journal of the American Society of Echocardiography*, 2013, **26**(1): 38–40
- Kingma D P, Ba J L. ADAM: A method for stochastic optimization. In: Proceedings of the 2015 International Conference on Learning Representations. San Diego, USA. 2015. 1–15
- Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B*, 2005, **67**(2): 301–320
- Paz-Linares D, Vega-Hernández M, Rojas-López P A, Valdés-Hernández P A, Martínez-Montes E, Valdés-Sosa P A. Spatio-temporal EEG source imaging with the hierarchical Bayesian elastic net and elitist lasso models. *Frontiers in Neuroscience*, 2017, **11**: 1–22
- Spisak T, Kincses B, Schlitt F, Zunhammer M, Schmidt-Wileke T, Kincses Z T, et al. Pain-free resting-state functional brain connectivity predicts individual pain sensitivity. *Nature Communications*, 2020, **11**: 187
- Bransford J D, Ann L, Brown A, Cocking R R. How people learn brain, mind, experience, and school committee. *Psychology*, 2004: 116
- Grossberg S. Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks*, 2013, **37**: 1–47
- Vaadia E. Learning how the brain learns. *Nature*, 2000, **405**: 523–525
- Yaple Z A, Stevens W D, Arsalidou M. NeuroImage Meta-analyses of the n-back working memory task? fMRI evidence of age-related changes in prefrontal cortex involvement across the adult

lifespan. *NeuroImage*, 2019, **196**: 16–31

- 31 Qiang N, Dong Q, Zhang W, Ge B, Ge F, Liang H, et al. Modeling task-based fMRI data via deep belief network with neural architecture search. *Computerized Medical Imaging and Graphics*, 2020: 101747
- 32 Wen X, Wang H, Liu Z, Liu C, Li K, Ding M, et al. Dynamic top-down configuration by the core control system during working memory. *Neuroscience*, 2018, **391**: 13–24
- 33 Wang H, Zhao S, Dong Q, Cui Y, Chen Y, Han J, et al. Recognizing brain states using deep sparse recurrent neural network. *IEEE Transactions on Medical Imaging*, 2018, **38**(4): 1058–1068
- 34 Qiang N, Dong Q, Ge F, Liang H, Ge B, Zhang S, et al. Deep variational autoencoder for mapping functional brain networks. *IEEE Transactions on Cognitive and Developmental Systems*, 2021, **13**(4): 841–852
- 35 Dong Q, Qiang N, Lv J, Li X, Liu T, Li Q. Spatiotemporal attention autoencoder (STAAE) for ADHD classification. In: Proceedings of the 2020 Medical Image Computing and Computer-Assisted Intervention. Lima, Peru: 2020. 508–517
- 36 Li Q, Wu X, Xu L, Chen K, Yao L. Classification of Alzheimer's disease, mild cognitive impairment, and cognitively unimpaired individuals using multi-feature kernel discriminant dictionary learning. *Frontiers in Computational Neuroscience*, 2018, **11**: 117
- 37 Wu X, Li Q, Xu L, Chen K, Yao L. Multi-feature kernel discriminant dictionary learning for face recognition. *Pattern Recognition*, 2017, **66**: 404–411
- 38 Zhang Xian-Fa, Hao Kuang-Rong, Chen Lei. Motor imagery EEG classification based on immune multi-domain-feature fusion and multiple kernel learning SVM. *Acta Automatica Sinica*, 2020, **46**(11): 2417–2426  
(张宪法, 郝矿荣, 陈磊. 免疫多域特征融合的多核学习SVM运动想象脑电信号分类. *自动化学报*, 2020, **46**(11): 2417–2426)
- 39 Li Q, Zhang W, Zhao L, Wu X, Liu T. Evolutional neural architecture search for optimization of spatiotemporal brain network decomposition. *IEEE Transactions on Biomedical Engineering*, 2022, **69**(2): 624–634
- 40 Ji Jun-Zhong, Zou Ai-Xiao, Liu Jin-Duo. An overview of identification methods on human brain effective connectivity networks based on functional magnetic resonance imaging. *Acta Automatica Sinica*, 2021, **47**(2): 278–296  
(冀俊忠, 邹爱笑, 刘金铎. 基于功能磁共振成像的人脑效应连接网络识别方法综述. *自动化学报*, 2021, **47**(2): 278–296)



李 晴 北京师范大学认知神经科学与学习国家重点实验室博士后. 2022年获北京师范大学博士学位. 主要研究方向为脑影像智能分析.

E-mail: liqing@bnu.edu.cn

(**LI Qing** Postdoctoral researcher at the State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University. She received her Ph.D. degree from Beijing Normal University in 2022. Her main research interest is brain imaging intelligent analysis.)



徐雪远 2022年获北京师范大学人工智能学院博士学位. 主要研究方向为脑信号智能分析.

E-mail: xuxueyuan@mail.bnu.edu.cn

(**XU Xue-Yuan** He received his Ph.D. degree from Beijing Normal University in 2022. His main research interest is brain signal intelligent analysis.)



邬 霞 北京师范大学人工智能学院教授. 2008年获北京师范大学认知神经科学与学习研究所博士学位. 主要研究方向为医学图像处理. 本文通信作者. E-mail: wuxia@bnu.edu.cn

(**WU Xia** Professor at the School of Artificial Intelligence, Beijing Normal University. She received her Ph.D. degree from Beijing Normal University in 2008. Her main research interest is medical image processing. Corresponding author of this paper.)