Data-Efficient Brain Connectome Analysis via Multi-Task Meta-Learning

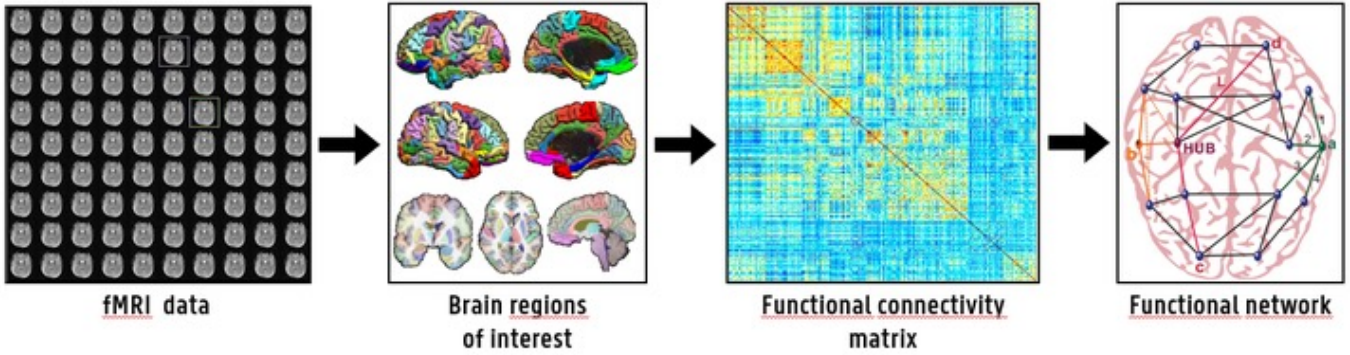
Yi Yang[1*], Yanqiao Zhu[2*], Hejie Cui[1], Xuan Kan[1], Lifang He[3], Ying Guo[1], Carl Yang[1]

[1]Emory University, [2]University of California Los Angeles, [3]Lehigh University
* Equal contribution

Content

Background & Motivation – *Scarcity of Neuroimaging Data*

- Brain network data is usually collected through various techniques such as functional magnetic resonance imaging (fMRI) or diffusion tensor imaging (DTI)
- The data collection process is costly and expensive which leads to severe scarcity of available training resources and data instances
- Modern machine learning techniques on complex topological data requires sufficiently large samples to achieve effective domain knowledge extraction and discriminative power



Joshi et al. 17' What are the steps for preprocessing fMRI data?

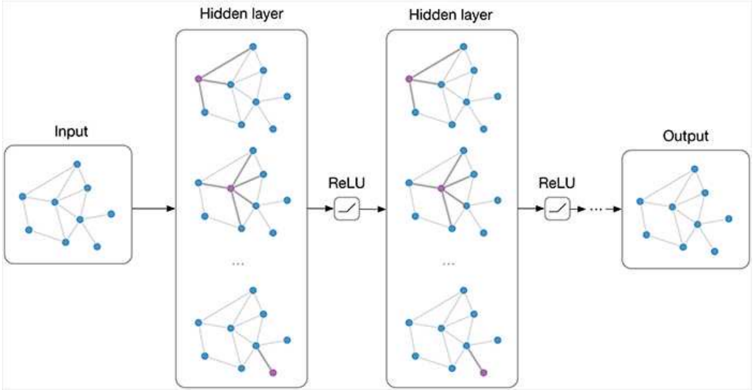
Related Work – GNNs & Meta-Learning

Graph Neural Networks

- ➢ GNN is powerful in learning topological relational information among nodes and edges
- ➢ *BrainGNN* (Li et al. 21') proposed ROI aware graph convolution layer and selective pooling layer
- ➢ With limited training samples, GNN suffers poor performance result and high variances

Meta-Learning on Graphs

- ➢ Mostly surveyed on applicability and feasibility of joint learning on multiple objectives
- ➢ Shared substructure learning may not adapt well on dense brain connectomes

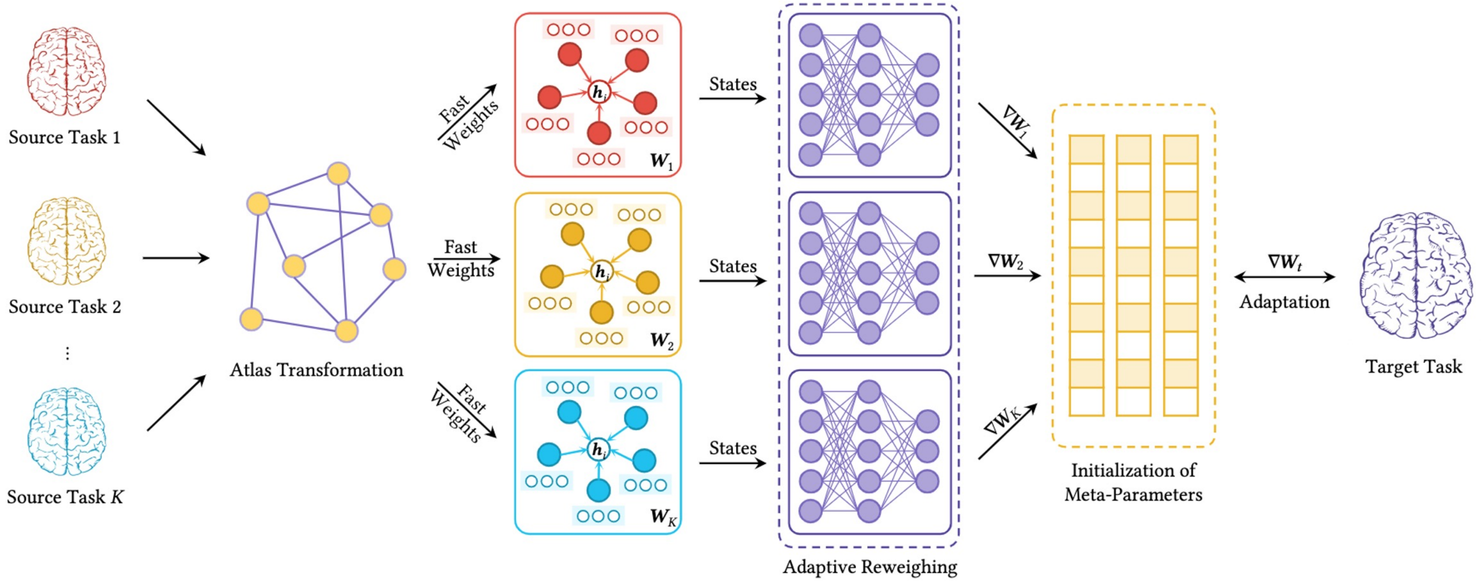


Kipf 16' Graph Convolutional Networks

Problem Formulation

- We consider the brain network data with carefully parcellated ROIs and correlations as edge weighted graph $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i, A_i)$ where the connectivity is represented by a node set $\mathcal{V}$ and an edge set $\mathcal{E} = \mathcal{V} \times \mathcal{V}$
- Our objective is to train an encoder model $f_\theta(\cdot)$ such that θ efficiently converges to optimal θ^* on target dataset $\mathcal{D}^t$ given that θ_0 is initialized via proper pre-training or meta-training on source datasets $\mathcal{D}^s = \{S_1, S_2, \cdots, S_k\}$, where $|\mathcal{D}^s| > |\mathcal{D}^t|$
- In our simplified setting, considering the multiview and multimodality nature of brain network dataset, we regard each view as an independent training objective. That is, given a dataset $\mathcal{D}$ with k modalities, our learning pipeline will extract k tasks into the task distribution τ.

Overall Pipeline



Data Efficient Training – Single Task Transfer Learning

- The first stage involves pre-training the encoder $f_\theta(\cdot)$ on a single source task and its corresponding objective
- We use the binary cross entropy loss objective for graph classification

$$\mathcal{L}_{\mathrm{BCE}} = -\frac{1}{|\mathcal{D}|} \sum_{(\mathcal{G}_i, y_i)\sim\mathcal{D}} y_i \log \sigma(f_\theta(\mathcal{G}_i)) + (1-y_i) \log(1-\sigma(f_\theta(\mathcal{G}_i)))$$

- We then fine-tune the model on target task and evaluate by cross validation
- Sensitive and vulnerable to knowledge gaps between source and target domain

Algorithm 1 Single-task supervised transfer learning (STT)

Input: pre-train task S, fine-tune task T, encoder $f(\theta)$
Require: α: learning rate hyperparameter
Randomly initialize θ
▷ Pre-training phase
while not done **do**
    Evaluate the gradient $\nabla_\theta \mathcal{L}_S f(\theta)$
    Update parameters with SGD: $\theta \leftarrow \theta - \alpha \nabla_\theta \mathcal{L}_S f(\theta)$
end while
▷ Fine-tuning phase
Split T into T_{train} and T_{eval} into K folds
for split **in** K folds **do**
    Get split-specific parameters $\hat{\theta} \leftarrow \theta$
    while not done **do**
        Evaluate the gradient $\nabla_{\hat{\theta}} \mathcal{L}_{T_{\text{train}}} f(\hat{\theta})$
        Update parameters with SGD $\hat{\theta} \leftarrow \hat{\theta} - \alpha \nabla_{\hat{\theta}} \mathcal{L}_{T_{\text{train}}} f(\hat{\theta})$
    end while
    Evaluate ACC, AUC from $f_{\hat{\theta}}(T_{\text{eval}})$
end for

Data Efficient Training – Multi-task Meta-Learning

- **Transitioning to multi-task transfer learning (MTT)**: extending into multi-task setting where $f_\theta(\cdot)$ is pre-trained according to an aggregated (e.g., sum) loss objective

- MTT has limited generalizability power due to the joint training objective

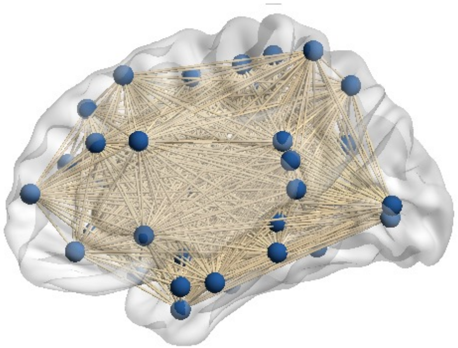
- State-of-the-art meta-learning (Finn et al. 17') architecture demonstrates robustness in generalizing knowledge across domains

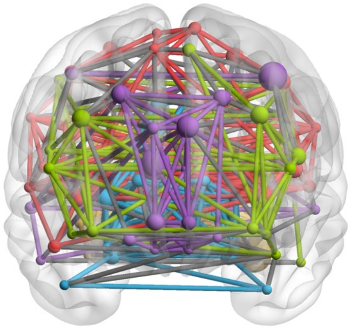
Algorithm 2 Multi-task meta-learning (MML)

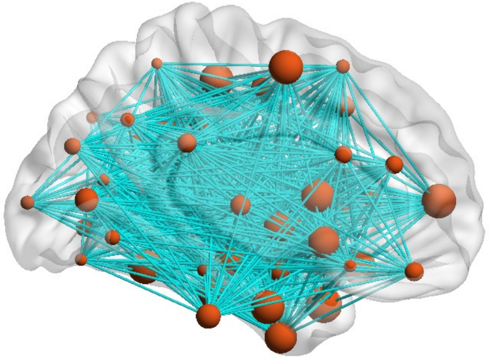
Input: meta-train task pool S_τ, meta-test task T, encoder $f(\theta)$
Require: α, β: learning rate hyperparameters
Randomly initialize θ
▹ Meta-training phase
while not done **do**
    for each task τ_i **in** S_τ **do**
        Sample k datapoints $\mathcal{D}_i$ from τ_i
        Evaluate the gradient $\nabla_\theta \mathcal{L}_{\mathcal{D}_i} f(\theta)$
        Compute the adapted parameters $\theta_i' \leftarrow \theta - \beta \nabla_\theta \mathcal{L}_{\mathcal{D}_i} f(\theta)$
        Sample another set of datapoints $\mathcal{D}_i'$ from τ_i
    end for
    Update parameters $\theta \leftarrow \theta - \alpha \nabla_\theta \sum_{\mathcal{D}_i', \theta_i' \sim S_\tau} \mathcal{L}_{\mathcal{D}_i'} f(\theta_i')$
end while
▹ Meta-test phase
Perform k-fold evaluation on target tasks

Dataset and Experimental Configuration

- **Datasets: [1] BP** (82 ROIs, 97 samples) **[2] HIV** (90 ROIs, 70 samples) **[3] PPMI** (84 ROIs, 718 samples)
- **Backbone encoders**: **[1] BrainNetCNN** (Kawahara et al. 17') **[2] GCN** (Kipf et al. 17') **[3] GAT** (Veličković et al. 18')
- **Training configurations**: We consider PPMI to be source dataset, BP and HIV as target dataset







Experimental Results

Table 1: Performance comparison of our proposed methodologies and baselines in terms of area under the ROC curve (AUC) and accuracy (ACC). The best performing model is highlighted in boldface.

Encoder	Dataset	Modality	DSL		STT		MTT		MML	
			AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
BrainNetCNN	BP	fMRI	0.50±0.13	0.51±0.15	0.55±0.07	0.56±0.08	0.56±0.09	0.56±0.11	0.57±0.10	0.57±0.07
		DTI	0.47±0.16	0.49±0.14	0.53±0.11	0.54±0.12	0.54±0.07	0.54±0.09	0.55±0.13	0.56±0.08
	HIV	fMRI	0.60±0.15	0.59±0.13	0.66±0.14	0.65±0.10	0.66±0.13	0.66±0.11	0.67±0.12	0.67±0.09
		DTI	0.54±0.16	0.53±0.15	0.60±0.09	0.60±0.09	0.60±0.10	0.60±0.12	0.57±0.11	0.61±0.14
GAT	BP	fMRI	0.51±0.13	0.52±0.16	0.57±0.07	0.58±0.05	0.59±0.10	0.59±0.07	0.61±0.07	0.60±0.09
		DTI	0.50±0.09	0.50±0.13	0.53±0.08	0.54±0.10	0.51±0.06	0.55±0.08	0.55±0.08	0.57±0.05
	HIV	fMRI	0.61±0.15	0.61±0.14	0.65±0.07	0.66±0.11	0.66±0.09	0.68±0.06	0.68±0.10	0.69±0.08
		DTI	0.56±0.17	0.55±0.15	0.61±0.07	0.60±0.08	0.62±0.09	0.61±0.10	0.64±0.09	0.62±0.12
GCN	BP	fMRI	0.55±0.11	0.54±0.14	0.59±0.12	0.58±0.13	0.61±0.10	0.60±0.11	**0.62±0.08**	**0.62±0.10**
		DTI	0.51±0.12	0.52±0.11	0.52±0.10	0.54±0.12	0.55±0.09	0.56±0.14	**0.59±0.07**	**0.58±0.11**
	HIV	fMRI	0.63±0.18	0.64±0.12	0.65±0.14	0.68±0.15	0.67±0.12	0.68±0.11	**0.69±0.10**	**0.70±0.09**
		DTI	0.60±0.12	0.58±0.13	0.61±0.11	0.60±0.12	0.63±0.13	0.63±0.15	**0.65±0.12**	**0.64±0.13**

Atlas Transformation

- **Motivation**: Cross dataset brain connectome analysis is challenged by incompatible and non-convertible ROI definitions across datasets
- **Learnable Linear Projections (LP)**: Emphasize on performing projection of the original input feature space by attaching a projection head $\boldsymbol{W} \in \mathbb{R}^{n \times k}$ in front of encoder f_θ
- **Simple Auto-encoding (AE)**: Emphasize on obtaining fixed representation of original feature space. The projection matrix $\boldsymbol{W} \in \mathbb{R}^{n \times k}$ has the objective given as $\arg\min_{\boldsymbol{W}} \| \boldsymbol{X} - \boldsymbol{X}\boldsymbol{W}\boldsymbol{W}^{\boldsymbol{T}} \|^2$

Table 2: Performance with three different atlas transformation techniques.

Dataset	Modality	Zero Pad		LP		AE	
		AUC	ACC	AUC	ACC	AUC	ACC
BP	fMRI	$0.62_{\pm 0.08}$	$0.62_{\pm 0.10}$	$0.62_{\pm 0.13}$	$0.63 \pm_{0.12}$	$\mathbf{0.63}_{\pm 0.09}$	$\mathbf{0.64}_{\pm 0.09}$
	DTI	$0.59_{\pm 0.07}$	$0.58_{\pm 0.11}$	$0.59_{\pm 0.09}$	$0.60_{\pm 0.14}$	$\mathbf{0.60}_{\pm 0.04}$	$\mathbf{0.61}_{\pm 0.10}$
HIV	fMRI	$0.69_{\pm 0.10}$	$0.70_{\pm 0.09}$	$0.71_{\pm 0.13}$	$0.70_{\pm 0.11}$	$\mathbf{0.73}_{\pm 0.10}$	$\mathbf{0.72}_{\pm 0.08}$
	DTI	$0.65_{\pm 0.12}$	$0.64_{\pm 0.13}$	$0.68_{\pm 0.14}$	$0.66_{\pm 0.13}$	$\mathbf{0.69}_{\pm 0.06}$	$\mathbf{0.69}_{\pm 0.08}$

Task Adaptive Reweighing

- **Motivations**: Base meta-learning framework fails to consider learning difficulty of different source tasks which leads to skewed and biased overall generalization.
- **Analysis**: We first investigate the relatedness of source (PPMI) and target (BP, HIV) data by visualizing a correlation derived from computed task embeddings (Achille et al. 19')
- **Observation**: High correlation corroborates with clinical studies on inter-connections among the investigated diseases suggesting that each source task does not contribute equally to the adaptation on target task.
- **Solution**: We leverage a dynamic and optimizable scheme for inner-loop hyperparameter (i.e,. Learning rate, weight decay) selection inspired by (Baik et al. 20'). The task-specific optimization is governed by a non-linear mapping function that the determines rate of convergence.

Algorithm 3 Multi-task meta-learning with adaptive task reweighing (MMAR)

Input: meta-train tasks S_τ, meta-test task T, encoder $f(\theta)$, hyperparameter generator $g(\phi)$
Require: η: outer-loop learning rate
Randomly initialize θ, ϕ

Table 3: Performance with task reweighing techniques.

Dataset	Modality	DSL		MML		MMAR	
		AUC	ACC	AUC	ACC	AUC	ACC
BP	fMRI	$0.55_{\pm 0.11}$	$0.54_{\pm 0.14}$	$0.62_{\pm 0.08}$	$0.62_{\pm 0.10}$	$\mathbf{0.68_{\pm 0.10}}$	$\mathbf{0.66_{\pm 0.08}}$
	DTI	$0.51_{\pm 0.12}$	$0.52_{\pm 0.11}$	$0.59_{\pm 0.07}$	$0.58_{\pm 0.11}$	$\mathbf{0.64_{\pm 0.06}}$	$\mathbf{0.64_{\pm 0.10}}$
HIV	fMRI	$0.63_{\pm 0.18}$	$0.64_{\pm 0.12}$	$0.69_{\pm 0.10}$	$0.70_{\pm 0.09}$	$\mathbf{0.74_{\pm 0.10}}$	$\mathbf{0.76_{\pm 0.08}}$
	DTI	$0.60_{\pm 0.12}$	$0.58_{\pm 0.13}$	$0.65_{\pm 0.12}$	$0.64_{\pm 0.13}$	$\mathbf{0.72_{\pm 0.08}}$	$\mathbf{0.72_{\pm 0.07}}$

Sample another set of datapoints $\mathcal{D}'_i$ from τ_i
end for
Update parameters $\theta \leftarrow \theta - \eta \nabla_\theta \sum_{\mathcal{D}'_i, \theta'_i \sim S_\tau} \mathcal{L}_{\mathcal{D}'_i} f(\theta'_i)$
Update parameters $\phi \leftarrow \phi - \eta \nabla_\phi \sum_{\mathcal{D}'_i, \theta'_i \sim S_\tau} \mathcal{L}_{\mathcal{D}'_i} f(\theta'_i)$
end while
Perform k-fold evaluation on target tasks

Conclusions and Remarks

- **What we did**: We proposed a data-efficient learning framework on brain network dataset. The framework is naturally generic and can be applied to broader spectrum of settings.

- **Current limitations**: **[1]** Brain networks are multimodal and a comprehensive feature extraction requires capturing shared knowledge across modalities. **[2]** Sampling data could be costly which motivates meta-optimization using less training instances.

- **Future directions**: We extend to unsupervised setting for model meta-training and explore on sampling-efficient strategies for brain network learning.

Thank You