Analyzing Human Brain Patterns

by using deep approaches

Muhammad (Tony) Yousefnezhad

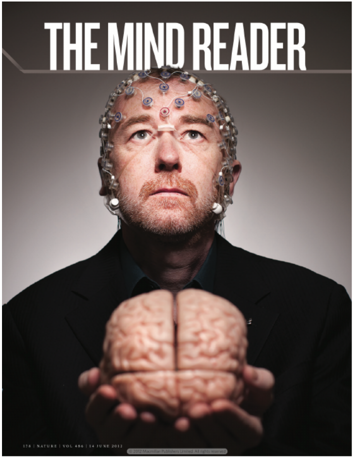
College of Computer Science and Technology
Nanjing University of Aeronautics and Astronautics (NUAA)

Machine Learning, Optimization and Control (MLOC) 2018

Outline

1. Analyzing Brain Patterns
2. Hyperalignment
3. Deep Hyperalignment
4. Deep Hyperalignment: Optimization
5. Experiments
6. Conclusion

The Mind Reader (in theory)



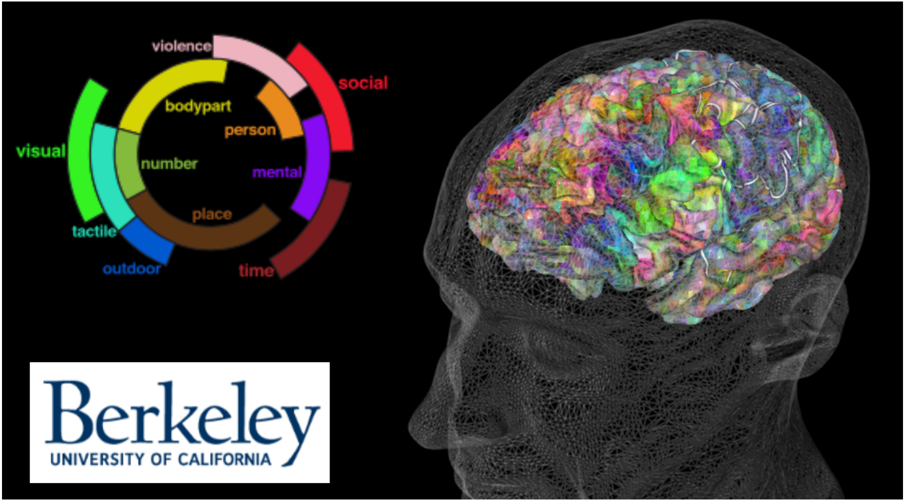
Smith, Nature, 2013

Optogenetics

Recovery Movies from Human Brain

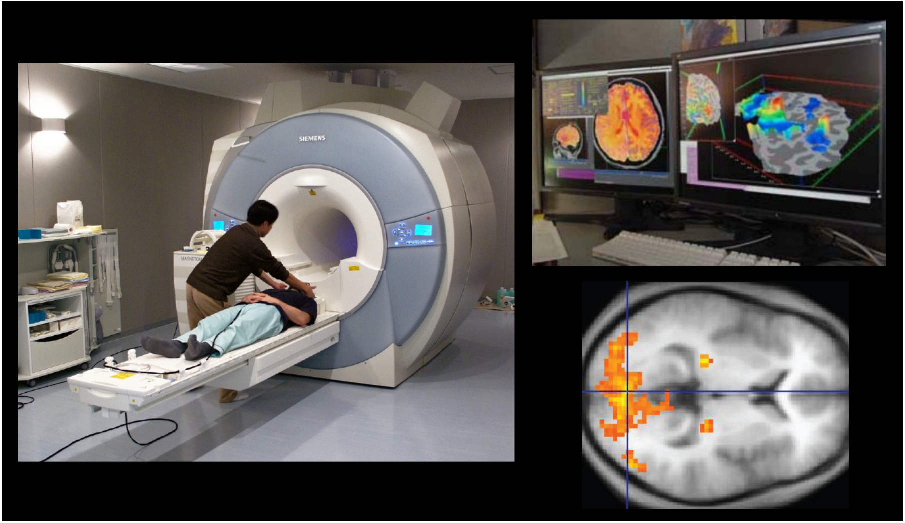
Nishimoto, Current Biology, 2011

Semantic Maps



Huth, Nature, 2016

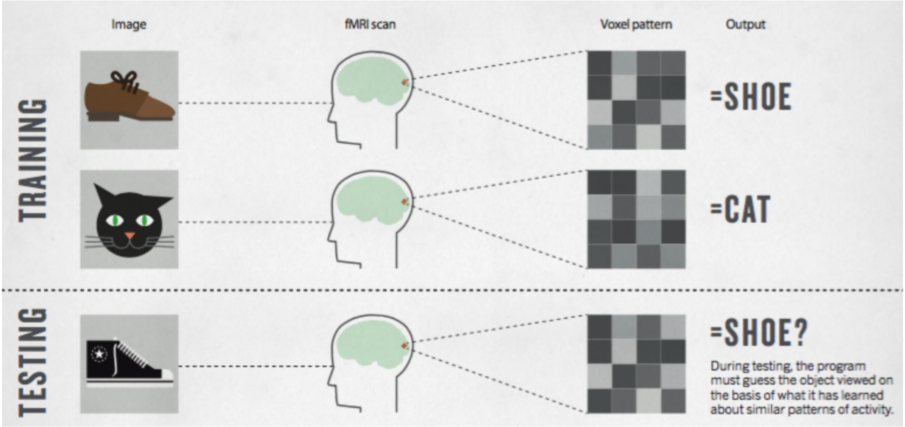
functional Imaging: functional MRI (fMRI)



fMRI vs. Other Modalities

- Prior to the discovery that **within-area patterns** of response in fMRI carried information that **afforded decoding of stimulus distinctions**.
- It was generally believed that the **spatial resolution of fMRI** allowed investigators to ask only which task or stimulus activated a region globally.
- Instead of asking what a regions function is, in terms of a single brain state associated with global activity, fMRI investigators can now ask **what information is represented in a region**, in terms of brain states associated with distinct patterns of activity, and how that information is encoded and organized.
- A wide range of **open source** fMRI datasets.

The Human Brain Decoding: Problem Definition

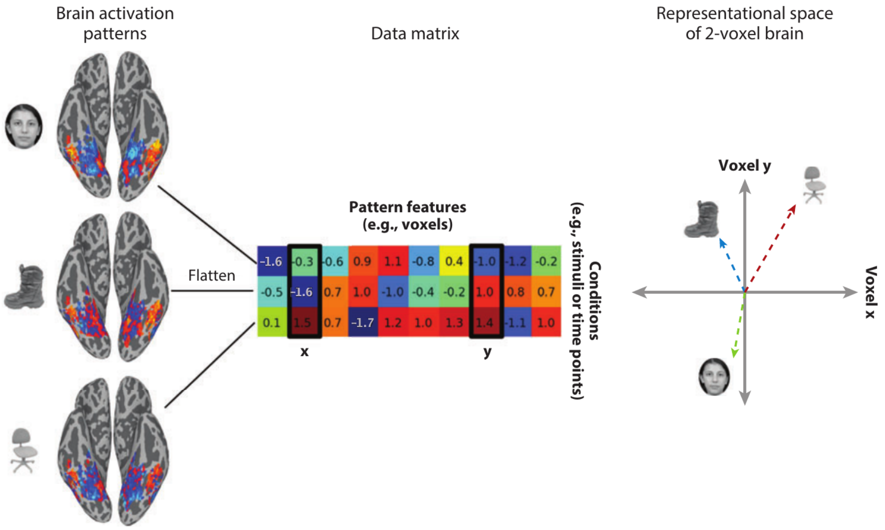


Smith, Nature, 2013

Outline

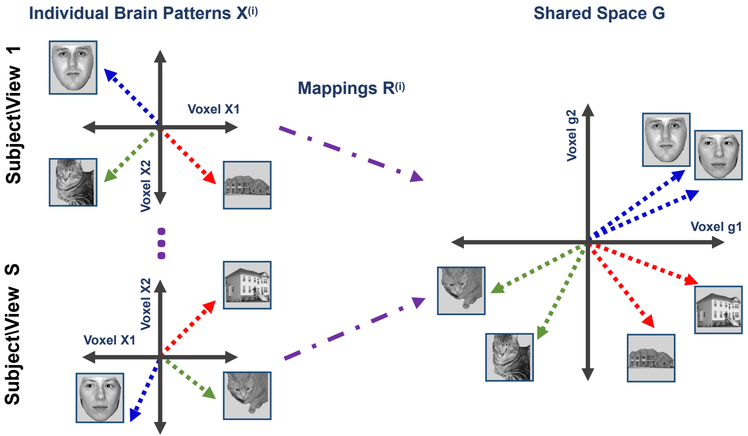
1. Analyzing Brain Patterns
2. Hyperalignment
3. Deep Hyperalignment
4. Deep Hyperalignment: Optimization
5. Experiments
6. Conclusion

Representational Space: Example



Haxby, Annual Review Neuroscience, 2014

Hyperalignment



- The main assumption in Hyperalignment is that the neural actives in different brains are noisy ‘rotations’ of a common space **Haxby, Neuron, 2011**.
- It can be formulated as extracting shared space from multi-view (multi-subject) data.

Classical Hyperalignment

Classical Hyperalignment can be formulated by Generalized Canonical Correlation Analysis (CCA): **Haxby, Neuron, 2011**

$$\min_{\mathbf{R}^{(i)},\mathbf{G}} \sum_{i=1}^{S} \left\| \mathbf{X}^{(i)}\mathbf{R}^{(i)} - \mathbf{G} \right\|_F^2$$

$$\text{subject to } \left(\mathbf{X}^{(\ell)}\mathbf{R}^{(\ell)}\right)^\top \mathbf{X}^{(\ell)}\mathbf{R}^{(\ell)} = \mathbf{I}$$

where the common space can be denoted by:

$$\mathbf{G} \in \mathbb{R}^{T\times V} = \frac{1}{S}\sum_{j=1}^{S} \mathbf{X}^{(j)}\mathbf{R}^{(j)},$$

- $\mathbf{X}^{(\ell)} \in \mathbb{R}^{T\times V}$ denotes the neural activities, and $\mathbf{R}^{(\ell)} \in \mathbb{R}^{V\times V}$ is the mappings.

Regularized Hyperalignment

- RHA's Objective Function can be denoted as follows:

$$\min_{\mathbf{R}^{(i)},\mathbf{G}} \sum_{i=1}^{S} \left\| \mathbf{X}^{(i)}\mathbf{R}^{(i)} - \mathbf{G} \right\|_F^2$$

$$\text{subject to } \left(\mathbf{R}^{(\ell)}\right)^\top \left(\left(\mathbf{X}^{(\ell)}\right)^\top \mathbf{X}^{(\ell)} + \epsilon \mathbf{I}\right)\mathbf{R}^{(\ell)} = \mathbf{I}$$

- The common space: $\mathbf{G} = \frac{1}{S}\sum_{j=1}^{S} \mathbf{X}^{(j)}\mathbf{R}^{(j)}$
- Here, the regularization term ϵ can improve the stability of alignment by providing a better inverse of the covariance matrix for $\mathbf{X}^{(i)}$.

Xu, IEEE SSP, 2012

Kernelized Hyperalignment

- KHA's Objective Function can be denoted as follows:

$$\min_{\mathbf{R}^{(i)},\mathbf{G}} \sum_{i=1}^{S} \left\| \mathbf{\Phi}(\mathbf{X}^{(i)})\mathbf{R}^{(i)} - \mathbf{G} \right\|_F^2$$

$$\text{subject to } \left(\mathbf{\Phi}(\mathbf{X}^{(\ell)})\mathbf{R}^{(\ell)}\right)^\top \mathbf{\Phi}(\mathbf{X}^{(\ell)})\mathbf{R}^{(\ell)} = \mathbf{I}$$

- The common space: $\mathbf{G} = \frac{1}{S}\sum_{j=1}^{S} \mathbf{\Phi}(\mathbf{X}^{(j)})\mathbf{R}^{(j)}$
- Here, $\mathbf{\Phi}(.)$ is a standard kernel function that can handle nonlinear datasets.
- However, classical kernel functions are limited by a restricted fixed representational space.

Lorbert, NIPS, 2012

Outline

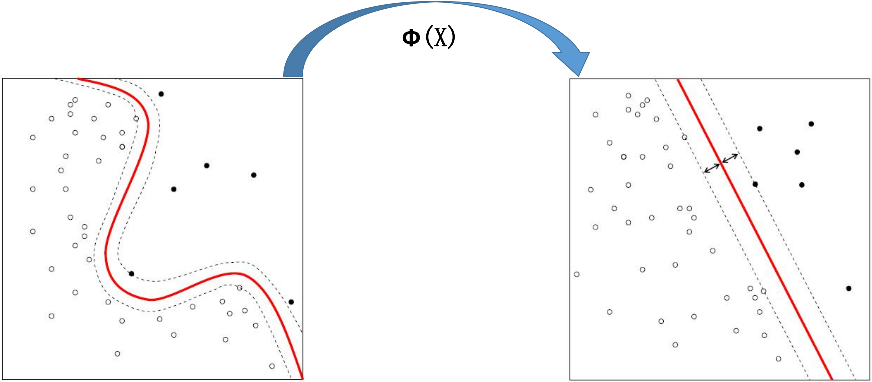
1. Analyzing Brain Patterns
2. Hyperalignment
3. Deep Hyperalignment
4. Deep Hyperalignment: Optimization
5. Experiments
6. Conclusion

Challenges

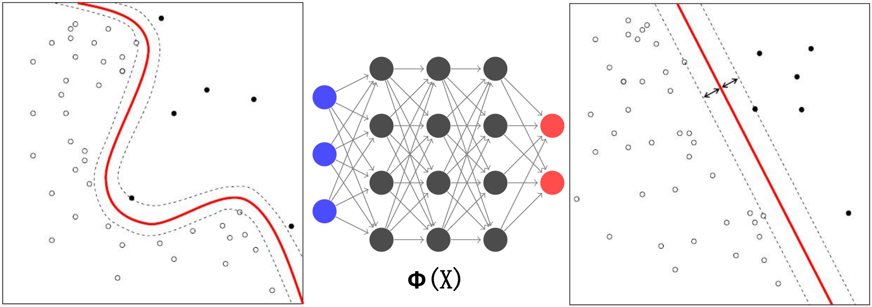
There are some long standing challenges for calculating accurate functional alignments:

- High Dimensionality
- Sparsity
- Nonlinear Features
- Large Number of Subjects

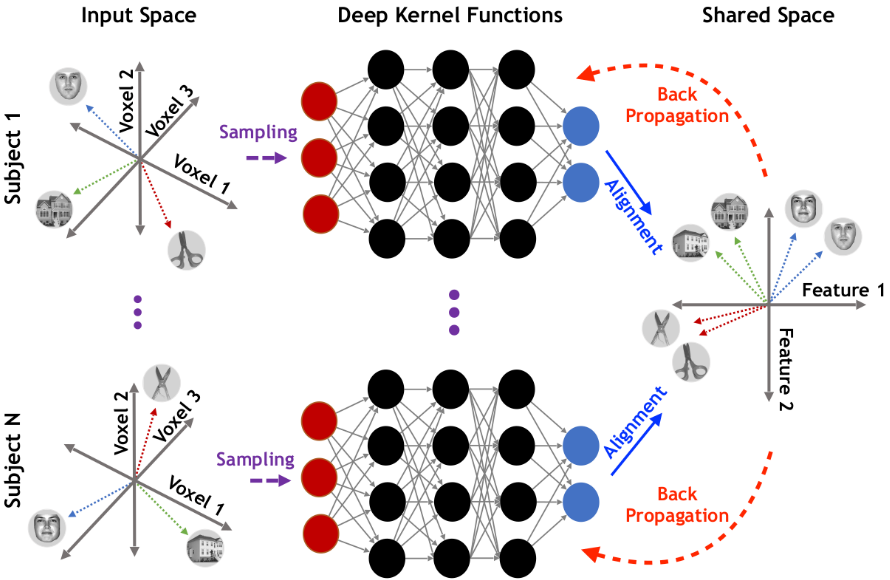
Kernel Function



Deep Kernel Function



Deep Hyperalignment (DHA)



Deep Hyperalignment: Objective Function

- DHA's Objective Function can be denoted as follows:

$$\min_{\mathbf{G},\mathbf{R}^{(i)},\theta^{(i)}} \sum_{i=1}^{S} \left\| \mathbf{G} - f_i\left(\mathbf{X}^{(i)};\theta^{(i)}\right)\mathbf{R}^{(i)} \right\|_F^2$$

$$\text{subject to } \left(\mathbf{R}^{(\ell)}\right)^\top \left(\left(f_\ell\left(\mathbf{X}^{(\ell)};\theta^{(\ell)}\right)\right)^\top f_\ell\left(\mathbf{X}^{(\ell)};\theta^{(\ell)}\right) + \epsilon\mathbf{I} \right)\mathbf{R}^{(\ell)} = \mathbf{I}$$

- The common space: $\mathbf{G} = \frac{1}{S}\sum_{j=1}^{S} f_j\left(\mathbf{X}^{(j)};\theta^{(j)}\right)\mathbf{R}^{(j)}$
- Here, f_ℓ is the deep neural network such as:

$$f_\ell\left(\mathbf{X}^{(\ell)};\theta^{(\ell)}\right) = \text{mat}\left(\mathbf{h}_C^{(\ell)}, T, V_{new}\right),$$

$$\mathbf{h}_m^{(\ell)} = g\left(\mathbf{W}_m^{(\ell)}\mathbf{h}_{m-1}^{(\ell)} + \mathbf{b}_m^{(\ell)}\right)$$

where $\mathbf{h}_1^{(\ell)} = \text{vec}\left(\mathbf{X}^{(\ell)}\right)$ and $m = 2{:}C$.

Outline

1 Analyzing Brain Patterns

2 Hyperalignment

3 Deep Hyperalignment

4 Deep Hyperalignment: Optimization

5 Experiments

6 Conclusion

Deep Hyperalignment: Objective Function

- Firstly, we employ the rank-m SVD as follows:

$$f_\ell\left(\mathbf{X}^{(\ell)};\theta^{(\ell)}\right) \overset{SVD}{=} \mathbf{\Omega}^{(\ell)}\mathbf{\Sigma}^{(\ell)}\left(\mathbf{\Psi}^{(\ell)}\right)^\top, \qquad \ell = 1{:}S$$

- Then, projection matrix can be calculated as follows:

$$\mathbf{P}^{(\ell)} = f_\ell\left(\mathbf{X}^{(\ell)};\theta^{(\ell)}\right)\left(\left(f_\ell\left(\mathbf{X}^{(\ell)};\theta^{(\ell)}\right)\right)^\top f_\ell\left(\mathbf{X}^{(\ell)};\theta^{(\ell)}\right) + \epsilon\mathbf{I}\right)^{-1}\left(f_\ell\left(\mathbf{X}^{(\ell)};\theta^{(\ell)}\right)\right)^\top$$
$$= \mathbf{\Omega}^{(\ell)}\left(\mathbf{\Sigma}^{(\ell)}\right)^\top\left(\mathbf{\Sigma}^{(\ell)}\left(\mathbf{\Sigma}^{(\ell)}\right)^\top + \epsilon\mathbf{I}\right)^{-1}\mathbf{\Sigma}^{(\ell)}\left(\mathbf{\Omega}^{(\ell)}\right)^\top = \mathbf{\Omega}^{(\ell)}\mathbf{D}^{(\ell)}\left(\mathbf{\Omega}^{(\ell)}\mathbf{D}^{(\ell)}\right)^\top$$

- Here, we have a diagonal product $\mathbf{D}^{(\ell)}\left(\mathbf{D}^{(\ell)}\right)^\top = \left(\mathbf{\Sigma}^{(\ell)}\right)^\top\left(\mathbf{\Sigma}^{(\ell)}\left(\mathbf{\Sigma}^{(\ell)}\right)^\top + \epsilon\mathbf{I}\right)^{-1}\mathbf{\Sigma}^{(\ell)}$. Thus, calculating the inverse of matrix is easy!

Yousefnezhad, NIPS, 2017

Deep Hyperalignment: Optimization (Step 1)

Theorem

By considering fixed mapping functions $\mathbf{R}^{(i)}$ and fixed network parameters $\theta^{(i)}$, DHA's Objective Function can be reformulated as follows:

$$\min_{\mathbf{G},\mathbf{R}^{(i)},\theta^{(i)}} \sum_{i=1}^{S} \left\| \mathbf{G} - f_i\left(\mathbf{X}^{(i)};\theta^{(i)}\right)\mathbf{R}^{(i)} \right\| \equiv \max_{\mathbf{G}}\left(tr(\mathbf{G}^\top \mathbf{A}\mathbf{G}) \right)$$

where the sum of projection matrices can be calculated as follows:

$$\mathbf{A} = \sum_{i=1}^{S} \mathbf{P}^{(i)} = \widetilde{\mathbf{A}}\widetilde{\mathbf{A}}^\top, \quad \text{where} \quad \widetilde{\mathbf{A}} \in \mathbb{R}^{T \times mS} = \left[\mathbf{\Omega}^{(1)}\mathbf{D}^{(1)} \dots \mathbf{\Omega}^{(S)}\mathbf{D}^{(S)}\right]$$

Theorem

By using **Incremental SVD**, *the shared space* $\mathbf{G}$ *can be calculated as follows, where* $\mathbf{\Lambda} = \left\{\lambda_1 \dots \lambda_T\right\}$ *is the eigenvalues of* $\mathbf{A}$*:*

$$\mathbf{A}\mathbf{G} = \mathbf{G}\mathbf{\Lambda} \implies \widetilde{\mathbf{A}} = \mathbf{G}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{\Psi}}^\top$$

Deep Hyperalignment: Optimization (Step 2)

Theorem

By considering fixed share space $\mathbf{G}$ *and fixed network parameters* $\theta^{(i)}$*, DHA's mapping functions can be calculated as follows:*

$$\mathbf{R}^{(\ell)} = \left(\left(f_\ell\left(\mathbf{X}^{(\ell)};\theta^{(\ell)}\right)\right)^\top f_\ell\left(\mathbf{X}^{(\ell)};\theta^{(\ell)}\right) + \epsilon\mathbf{I}\right)^{-1}\left(f_\ell\left(\mathbf{X}^{(\ell)};\theta^{(\ell)}\right)\right)^\top \mathbf{G}$$

Deep Hyperalignment: Optimization (Step 3)

Theorem

By considering fixed share space $\mathbf{G}$ *and fixed mapping functions* $\mathbf{R}^{(i)}$*, we use back-propagation algorithm for seeking an optimized parameters for the deep network as follows:*

$$\frac{\partial \mathbf{Z}}{\partial f_\ell\left(\mathbf{X}^{(\ell)};\theta^{(\ell)}\right)} = 2\mathbf{R}^{(\ell)}\mathbf{G}^\top - 2\mathbf{R}^{(\ell)}\left(\mathbf{R}^{(\ell)}\right)^\top\left(f_\ell\left(\mathbf{X}^{(\ell)};\theta^{(\ell)}\right)\right)^\top$$

where $\mathbf{Z}$ *is the sum of the eigenvalues of* $\mathbf{A}$*:*

$$\mathbf{Z} = \sum_{\ell=1}^{T} \lambda_\ell$$

Deep Hyperalignment: Algorithm

Algorithm 1 Deep Hyperalignment (DHA)

Input: Data $\mathbf{X}^{(i)}$, $i = 1{:}S$, Regularized parameter ϵ, Number of layers C, Number of units $U^{(m)}$ for $m = 2{:}C$, HA template $\widehat{\mathbf{G}}$ for testing phase (default $\emptyset$), Learning rate η (default 10^{-4} [13]).
Output: DHA mappings $\mathbf{R}^{(\ell)}$ and parameters $\theta^{(\ell)}$, HA template $\mathbf{G}$ just from training phase
Method:

01. Initialize iteration counter: $m \leftarrow 1$ and $\theta^{(\ell)} \sim \mathcal{N}(0, 1)$ for $\ell = 1{:}S$.
02. Construct $f_\ell\left(\mathbf{X}^{(\ell)};\theta^{(\ell)}\right)$ based on (4) and (5) by using $\theta^{(\ell)}$, C, $U^{(m)}$ for $\ell = 1{:}S$.
03. **IF** ($\widehat{\mathbf{G}} = \emptyset$) **THEN** *% The first step of DHA: fixed $\theta^{(\ell)}$ and calculating $\mathbf{G}$ and $\mathbf{R}^{(\ell)}$ ↓*
04. Generate $\widetilde{\mathbf{A}}$ by using (8) and (10).
05. Calculate $\mathbf{G}$ by applying Incremental SVD [15] to $\widetilde{\mathbf{A}} = \mathbf{G}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{\Psi}}^\top$.
06. **ELSE**
07. $\mathbf{G} = \widehat{\mathbf{G}}$.
08. **END IF**
09. Calculate mappings $\mathbf{R}^{(\ell)}$, $\ell = 1{:}S$ by using (12).
10. Estimate error of iteration $\gamma_m = \sum_{i=1}^{S} \sum_{j=i+1}^{S} \left\| f_i\left(\mathbf{X}^{(i)};\theta^{(i)}\right)\mathbf{R}^{(i)} - f_j\left(\mathbf{X}^{(j)};\theta^{(j)}\right)\mathbf{R}^{(j)} \right\|_F^2$.
11. **IF** $\left((m > 3) \text{ and } (\gamma_m \geq \gamma_{m-1} \geq \gamma_{m-2})\right)$ **THEN** *% This is the finishing condition.*
12. **Return** calculated $\mathbf{G}$, $\mathbf{R}^{(\ell)}$, $\theta^{(\ell)}(\ell = 1{:}S)$ related to $(m\text{-}2)\text{-}th$ iteration.
13. **END IF** *% The second step of DHA: fixed $\mathbf{G}$ and $\mathbf{R}^{(\ell)}$ and updating $\theta^{(\ell)}$ ↓*
14. $\nabla\theta^{(\ell)} \leftarrow \text{backprop}\left({}^{\partial \mathbf{Z}}/_{\partial f_\ell(\mathbf{X}^{(\ell)};\theta^{(\ell)})}, \theta^{(\ell)}\right)$ by using (13) for $\ell = 1{:}S$.
15. Update $\theta^{(\ell)} \leftarrow \theta^{(\ell)} - \eta\nabla\theta^{(\ell)}$ for $\ell = 1{:}S$ and then $m \leftarrow m + 1$
16. **SAVE** all DHA parameters related to this iteration and **GO TO** Line *02*.

Outline

1. Analyzing Brain Patterns
2. Hyperalignment
3. Deep Hyperalignment
4. Deep Hyperalignment: Optimization
5. Experiments
6. Conclusion

Datasets

Title	ID	S	K	T	V	X	Y	Z	Scanner	TR	TE
Mixed-gambles task	DS005	48	2	240	450	53	63	52	S 3T	2	30
Visual Object Recognition	DS105	71	8	121	1963	79	95	79	G 3T	2.5	30
Word and Object Processing	DS107	98	4	164	932	53	63	52	S 3T	2	28
Auditory and Visual Oddball	DS116	102	2	170	2532	53	63	40	P 3T	2	25
Multi-subject, multi-modal	DS117	171	2	210	524	64	61	33	S 3T	2	30
Forrest Gump	DS113	20	10	451	2400	160	160	36	S 7T	2.3	22
Raiders of the Lost Ark	N/A	10	7	924	980	78	78	54	S 3T	3	30

S is the number of subject; K denotes the number of stimulus categories; T is the number of scans in unites of TRs (Time of Repetition); V denotes the number of voxels in ROI; X, Y, Z are the size of 3D images; Scanners include S=Siemens, G = General Electric, and P = Philips in 3 Tesla or 7 Tesla; TR is Time of Repetition in millisecond; TE denotes Echo Time in second; Please see *openfmri.org* for more information.

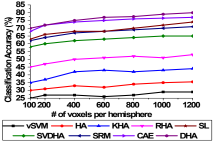
Simple Task Analysis: Accuracy of HA methods

↓Algorithms, Datasets→	DS005	DS105	DS107	DS116	DS117
ν-SVM	71.65±0.97	22.89±1.02	38.84±0.82	67.26±1.99	73.32±1.67
Hyperalignment (HA)	81.27±0.59	30.03±0.87	43.01±0.56	74.23±1.40	77.93±0.29
Regularized HA	83.06±0.36	32.62±0.52	46.82±0.37	78.71±0.76	84.22±0.44
Kernel HA	85.29±0.49	37.14±0.91	52.69±0.69	78.03±0.89	83.32±0.41
SVD-HA	90.82±1.23	40.21±0.83	59.54±0.99	81.56±0.54	95.62±0.83
Shared Response Model	91.26±0.34	48.77±0.94	64.11±0.37	83.31±0.73	95.01±0.64
SearchLight	90.21±0.61	49.86±0.4	64.07±0.98	82.32±0.28	94.96±0.24
Convolutional Autoencoder	94.25±0.76	54.52±0.80	72.16±0.43	**91.49±0.67**	95.92±0.67
Deep HA	**97.92±0.82**	**60.39±0.68**	**73.05±0.63**	90.28±0.71	**97.99±0.94**

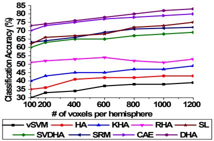
Simple Task Analysis: AUC of HA methods

↓Algorithms, Datasets→	DS005	DS105	DS107	DS116	DS117
ν-SVM [17]	68.37±1.01	21.76±0.91	36.84±1.45	62.49±1.34	70.17±0.59
Hyperalignment (HA)	70.32±0.92	28.91±1.03	40.21±0.33	70.67±0.97	76.14±0.49
Regularized HA	82.22±0.42	30.35±0.39	43.63±0.61	76.34±0.45	81.54±0.92
Kernel HA	80.91±0.21	36.23±0.57	50.41±0.92	75.28±0.94	80.92±0.28
SVD-HA	88.54±0.71	37.61±0.62	57.54±0.31	78.66±0.82	92.14±0.42
Shared Response Model	90.23±0.74	44.48±0.75	62.41±0.72	79.20±0.98	93.65±0.93
SearchLight	89.79±0.25	47.32±0.92	61.84±0.32	80.63±0.81	93.26±0.72
Convolutional Autoencoder	91.24±0.61	52.16±0.63	**72.33±0.79**	87.53±0.72	91.49±0.33
Deep HA	**96.91±0.82**	**59.57±0.32**	70.23±0.92	**89.93±0.24**	**96.13±0.32**

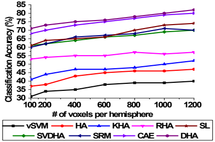
Complex Task Analysis



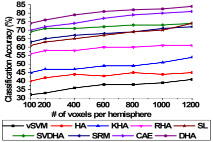
(a) Forrest Gump (TRs = 100)



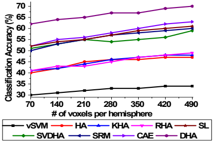
(b) Forrest Gump (TRs = 400)



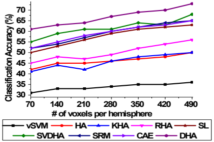
(c) Forrest Gump (TRs = 800)



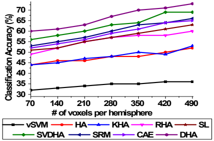
(d) Forrest Gump (TRs = 2000)



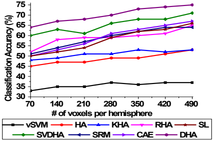
(e) Raiders (TRs = 100)



(f) Raiders (TRs = 400)

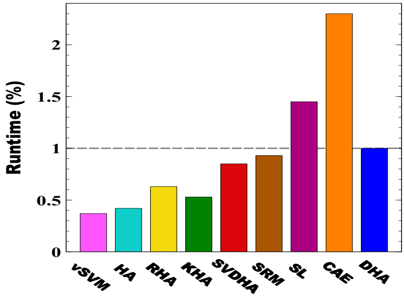


(g) Raiders (TRs = 800)

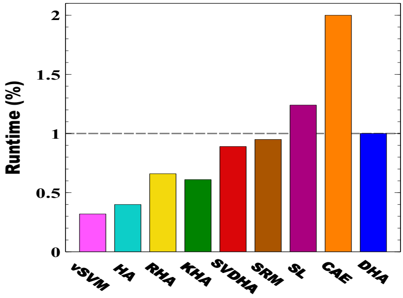


(h) Raiders (TRs = 2000)

Runtime Analysis

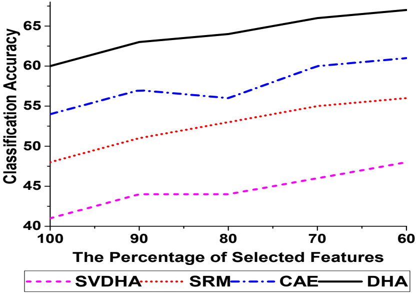


(A) DS105

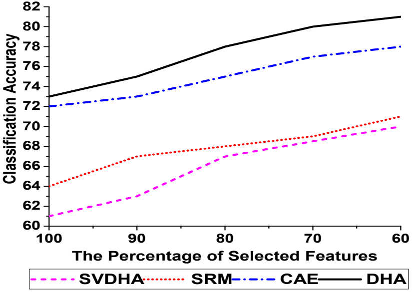


(B) DS107

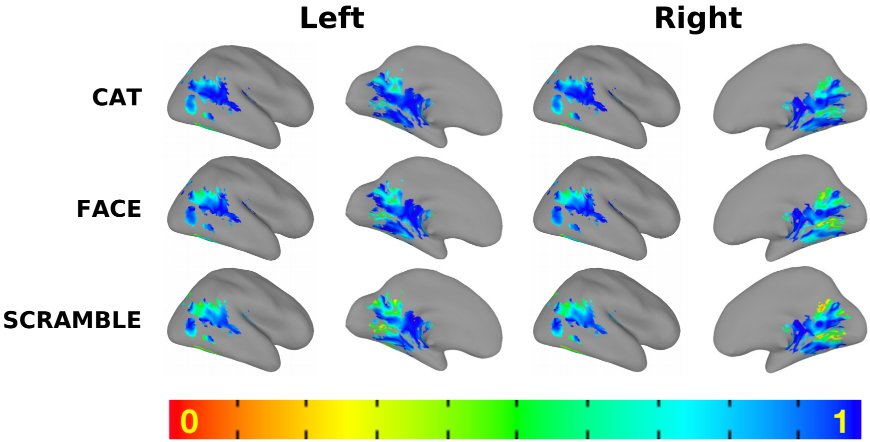
Alignment by selecting features



(A) DS105



(B) DS107


Left
Right
CAT
FACE
SCRAMBLE
0
1

Outline

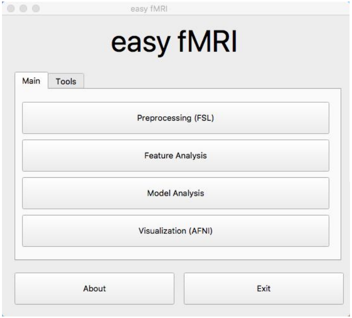
1. Analyzing Brain Patterns
2. Hyperalignment
3. Deep Hyperalignment
4. Deep Hyperalignment: Optimization
5. Experiments
6. Conclusion

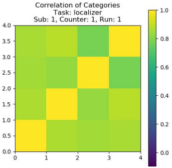
Conclusion

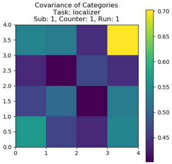
- Our knowledge from human brain is so **limited**.
- In order to understand the human brain, we need to develop new methods in Neuroscience, Psychology, **Mathematics, and Computer Science**.
- Not only can Artificial Intelligence use as a powerful tool for understanding the human brain but also this understanding can be employed reversely to develop AI tools, **e.g. Deep Learning**.

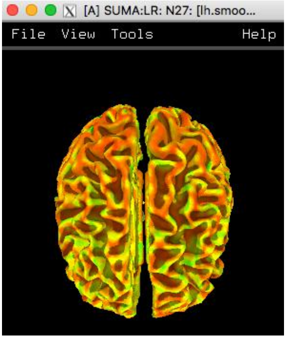
easy fMRI Project

Open Source + Free + Python + SK-Learn + MPI + Tensorflow



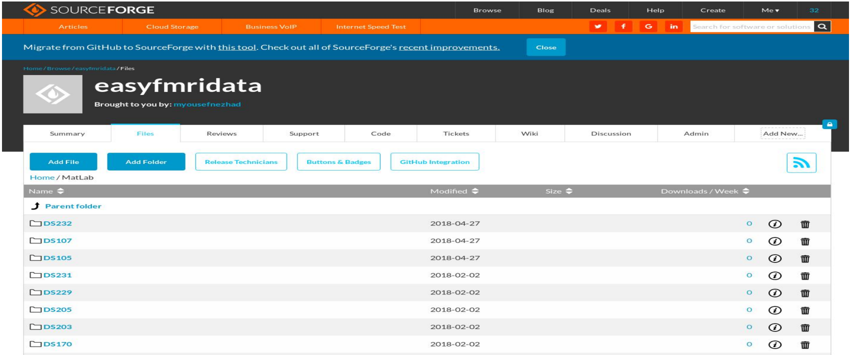






https://easyfmri.gitlab.io/
https://easyfmri.github.io/
https://easyfmri.sourceforge.io/

Matlab + 40 dataset + 200 cognitive tasks + 1000 subjects



https://easydata.gitlab.io/
https://easyfmridata.github.io/
https://easyfmridata.sourceforge.io/

Publications

- **Muhammad Yousefnezhad** and Daoqiang Zhang. ‘Deep Hyperalignment’, NIPS, 2017.
- **Muhammad Yousefnezhad** and Daoqiang Zhang. ‘Local Discriminant Hyperalignment for Multi-Subject fMRI Data Alignment’, AAAI, 2017.
- **Muhammad Yousefnezhad** and Daoqiang Zhang. ‘Multi-Region Neural Representation: A novel model for decoding visual stimuli in human brains’, SIAM SDM, 2017.
- **Muhammad Yousefnezhad** and Daoqiang Zhang. ‘Decoding visual stimuli in human brain by using Anatomical Pattern Analysis on fMRI images’, BICS, China, 2016.
- **Muhammad Yousefnezhad** and Daoqiang Zhang. ‘Weighted spectral cluster ensemble’, ICDM, 2015.
- **Muhammad Yousefnezhad** and Daoqiang Zhang. ‘Anatomical Pattern Analysis for decoding visual stimuli in human brains’. Cognitive Computation, 2017.
- **Muhammad Yousefnezhad**, Sheng-Jun Huang, and Daoqiang Zhang. ‘WoCE: a framework for clustering ensemble by exploiting the wisdom of Crowds theory’, IEEE Transactions on Cybernetics.
- **Muhammad Yousefnezhad**, Ali Reihanian, Daoqiang Zhang, Behrouz Minaei-Bidgoli ‘A new selection strategy for selective cluster ensemble based on Diversity and Independency’, Engineering Applications of Artificial Intelligence.
- Tonglin Xu, **Muhammad Yousefnezhad**, Daoqiang Zhang. ‘Gradient Hyperalignment for multi-subject fMRI data alignment’, PRICAI, China, 2018.

Thank You
Q & A

For more details, contact:

myousefnezhad@outlook.com
https://myousefnezhad.gitlab.io/
https://myousefnezhad.github.io/
https://myousefnezhad.sourceforge.io/
https://ibrain.nuaa.edu.cn