iBRAIN

Deep Hyperalignment

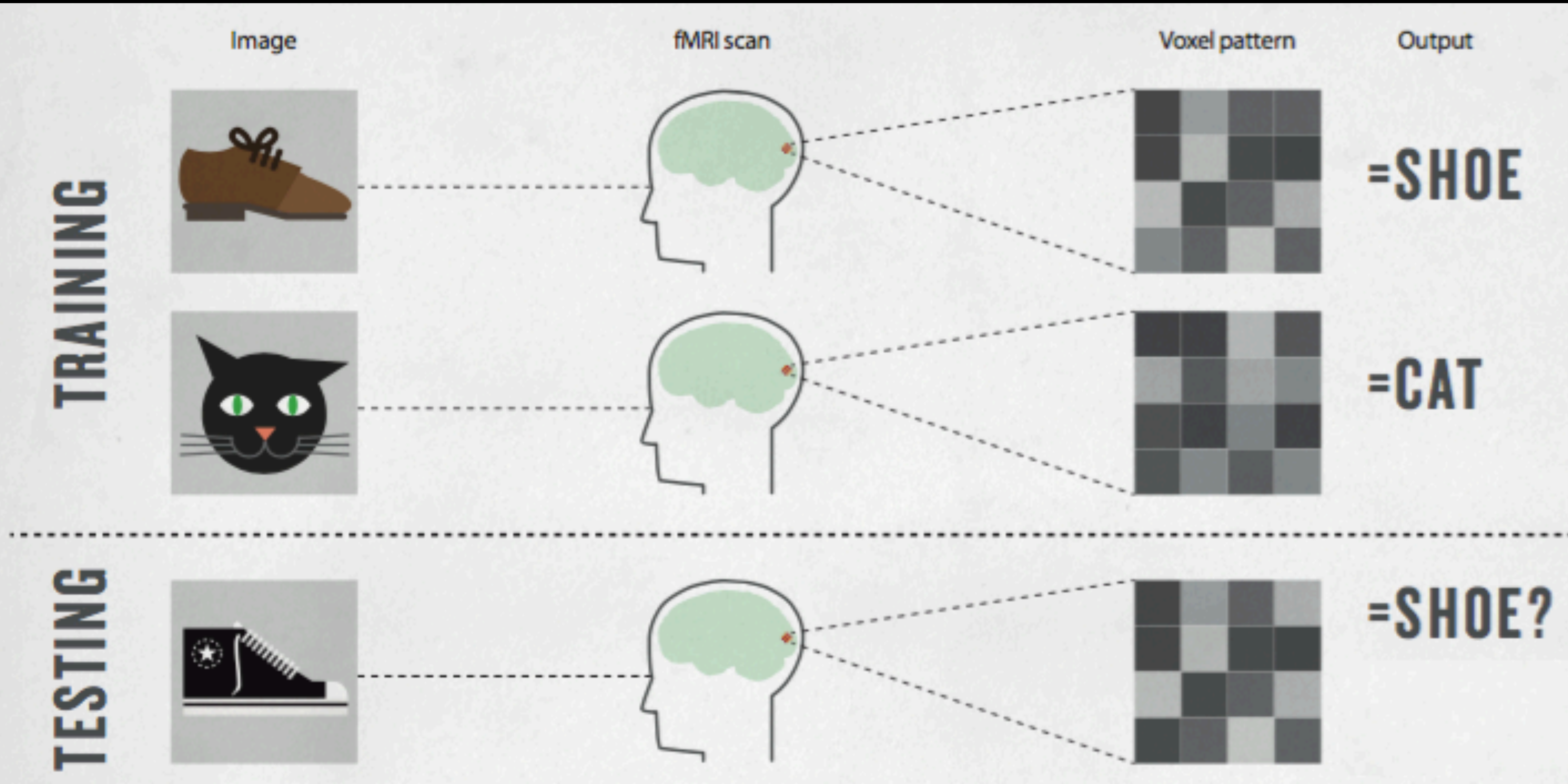
Muhammad Yousefnezhad, Daoqiang Zhang

31st Advances in Neural Information Processing Systems (NIPS-17)

Dec 2017

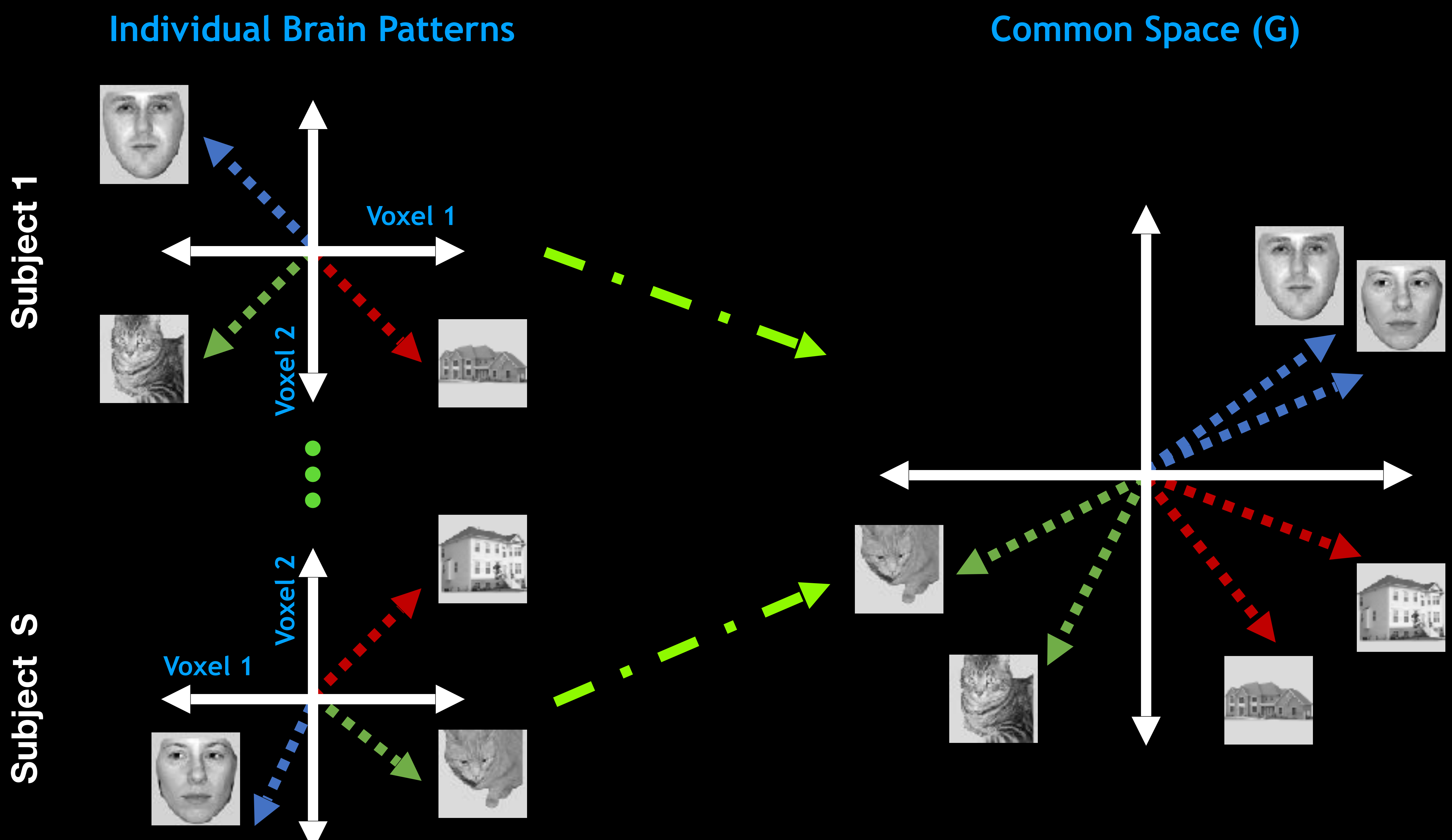
What is
Hyperalignment?

Brain Function Analysis



Smith, Nature, 2013

Hyperalignment



A Generalized Approach

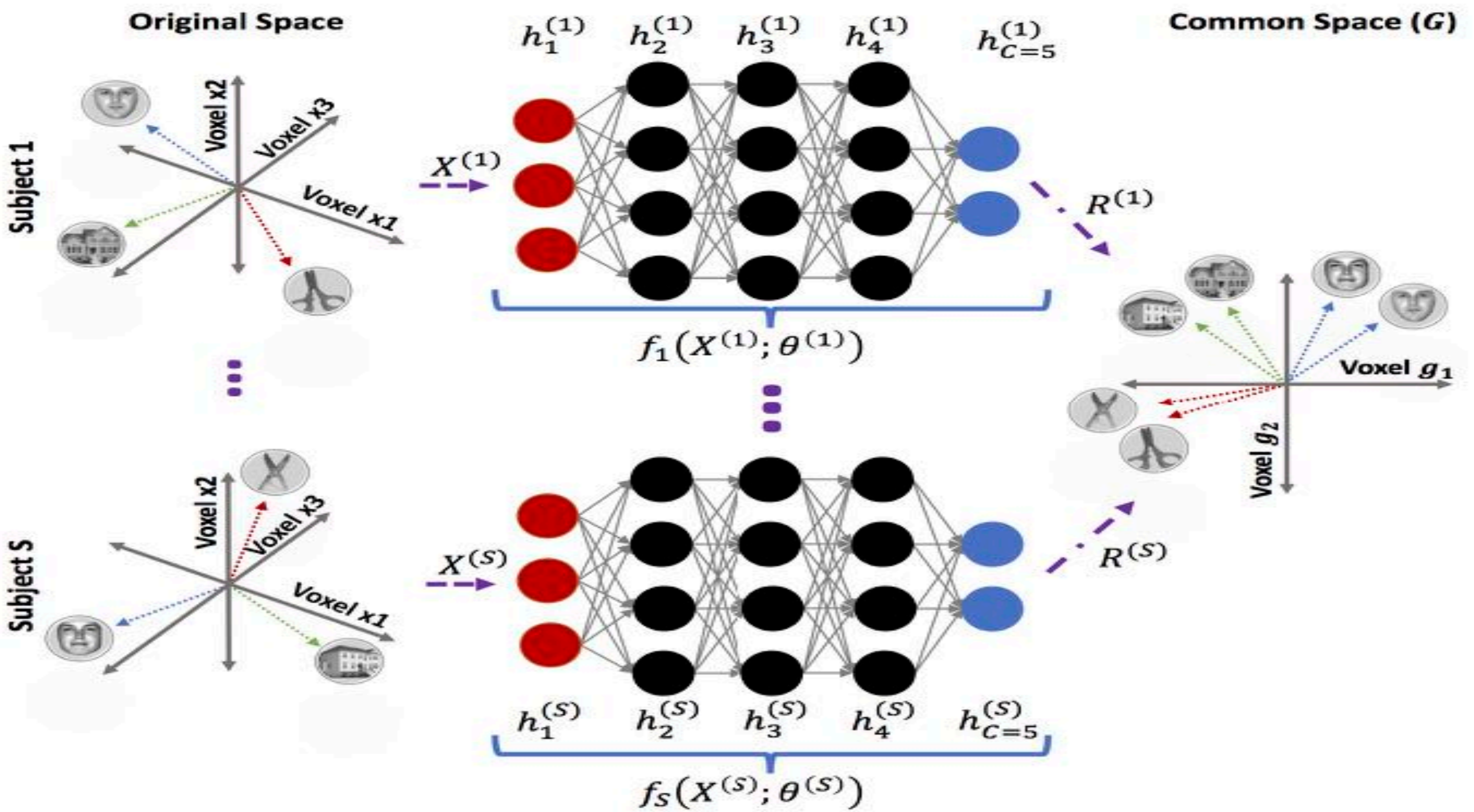
$$\min_{\mathbf{R}^{(i)},\mathbf{R}^{(j)}} \sum_{i=1}^{S} \sum_{j=i+1}^{S} \left\| f\left(\mathbf{X}^{(i)}\right)\mathbf{R}^{(i)} - f\left(\mathbf{X}^{(j)}\right)\mathbf{R}^{(j)} \right\|_F^2$$

s.t. $$\left(\mathbf{R}^{(\ell)}\right)^\top \left(\left(f\left(\mathbf{X}^{(\ell)}\right) \right)^\top f\left(\mathbf{X}^{(\ell)}\right) + \epsilon \mathbf{I} \right) \mathbf{R}^{(\ell)} = \mathbf{I}, \quad \ell = 1{:}S$$

- If $f(\mathbf{x}) = \mathbf{x}$ & $\epsilon = 0$, then we have the original HA
- If $f(\mathbf{x}) = \mathbf{x}$ & $\epsilon \neq 0$, then we have the Regularized HA
- If $f(\mathbf{x})$ is a nonlinear kernel, then we have the Kernel HA

Deep Hyperalignment

Main Idea



Original Space
Common Space (G)
Subject 1
Subject S
Voxel x1
Voxel x2
Voxel x3
$X^{(1)}$
$X^{(S)}$
$h_1^{(1)}$
$h_2^{(1)}$
$h_3^{(1)}$
$h_4^{(1)}$
$h_{C=5}^{(1)}$
$h_1^{(S)}$
$h_2^{(S)}$
$h_3^{(S)}$
$h_4^{(S)}$
$h_{C=5}^{(S)}$
$f_1(X^{(1)};\theta^{(1)})$
$f_S(X^{(S)};\theta^{(S)})$
$R^{(1)}$
$R^{(S)}$
Voxel g_1
Voxel g_2

DHA: Objective Function

- We want to optimize following function:

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

s.t. $$\left(\mathbf{R}^{(\ell)}\right)^\top\left(\left(f_\ell\left(\mathbf{X}^{(\ell)};\theta^{(\ell)}\right)\right)^\top f_\ell\left(\mathbf{X}^{(\ell)};\theta^{(\ell)}\right)+\epsilon\mathbf{I}\right)\mathbf{R}^{(\ell)}=\mathbf{I},\quad \ell=1{:}S$$

where the deep network is defined as follows:

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

$$\mathbf{h}_m^{(\ell)}=\mathbf{g}\left(\mathbf{W}_m^{(\ell)}\mathbf{h}_{m-1}^{(\ell)}+\mathbf{b}_m^{(\ell)}\right),\quad \textbf{where}\quad \mathbf{h}_1^{(\ell)}=\mathbf{vec}\left(\mathbf{X}^{(\ell)}\right)\quad \textbf{and}\quad m=2{:}C$$

Generalized DHA

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

s.t. $\quad \mathbf{G}^\top \mathbf{G} = \mathbf{I}$

where

$$\mathbf{G} = \frac{1}{S} \sum_{j=1}^{S} f_j\left(\mathbf{X}^{(j)};\theta^{(j)}\right)\mathbf{R}^{(j)}$$

DHA: Optimization

- *rank-m* **SVD**

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

- ***Projection Matrix***

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

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

where $\mathbf{D}^{(\ell)}\left(\mathbf{D}^{(\ell)}\right)^\top = \left(\mathbf{\Sigma}^{(\ell)}\right)^\top\left(\mathbf{\Sigma}^{(\ell)}\left(\mathbf{\Sigma}^{(\ell)}\right)^\top + \epsilon\mathbf{I}\right)^{-1}\mathbf{\Sigma}^{(\ell)}.$

- **Sum of *Projection Matrices***

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

Cholesky Decomposition

DHA: Optimization

- ***Objective Function can be reformulated as follows:***

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

- ***So, we have:***

$$\mathbf{AG} = \mathbf{G}\boldsymbol{\Lambda}, \text{ where } \Lambda = \left\{\lambda_1 \ldots \lambda_T\right\}$$

$$\widetilde{\mathbf{A}} = \mathbf{G}\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{\Psi}}^\top \longrightarrow \text{Incremental PCA}$$

- ***DHA mappings can be calculated as follows:***

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

DHA: Optimization

- ***In order to use back-propagation algorithm for seeking an optimized parameters for the deep network, we also have:***

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

where

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

Empirical Studies

Datasets

Table S2: The 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.

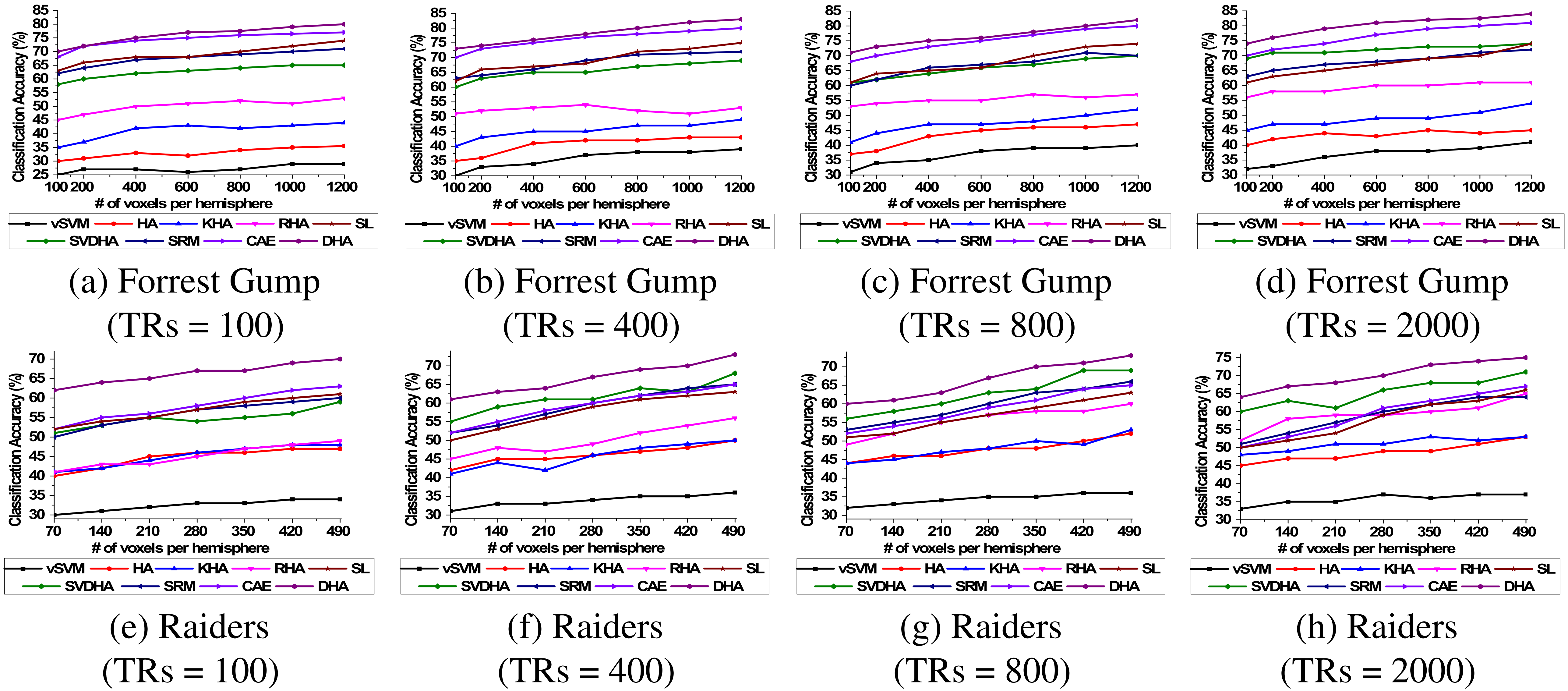
Simple Tasks Analysis

Table 1: Accuracy of HA methods in post-alignment classification by using simple task datasets

↓Algorithms, Datasets→	DS005	DS105	DS107	DS116	DS117
ν-SVM [17]	71.65±0.97	22.89±1.02	38.84±0.82	67.26±1.99	73.32±1.67
HA [1]	81.27±0.59	30.03±0.87	43.01±0.56	74.23±1.40	77.93±0.29
RHA [2]	83.06±0.36	32.62±0.52	46.82±0.37	78.71±0.76	84.22±0.44
KHA [3]	85.29±0.49	37.14±0.91	52.69±0.69	78.03±0.89	83.32±0.41
SVD-HA [4]	90.82±1.23	40.21±0.83	59.54±0.99	81.56±0.54	95.62±0.83
SRM [5]	91.26±0.34	48.77±0.94	64.11±0.37	83.31±0.73	95.01±0.64
SL [9]	90.21±0.61	49.86±0.4	64.07±0.98	82.32±0.28	94.96±0.24
CAE [6]	94.25±0.76	54.52±0.80	72.16±0.43	**91.49±0.67**	95.92±0.67
DHA	**97.92±0.82**	**60.39±0.68**	**73.05±0.63**	90.28±0.71	**97.99±0.94**

Table 2: Area under the ROC curve (AUC) of different HA methods in post-alignment classification by using simple task datasets

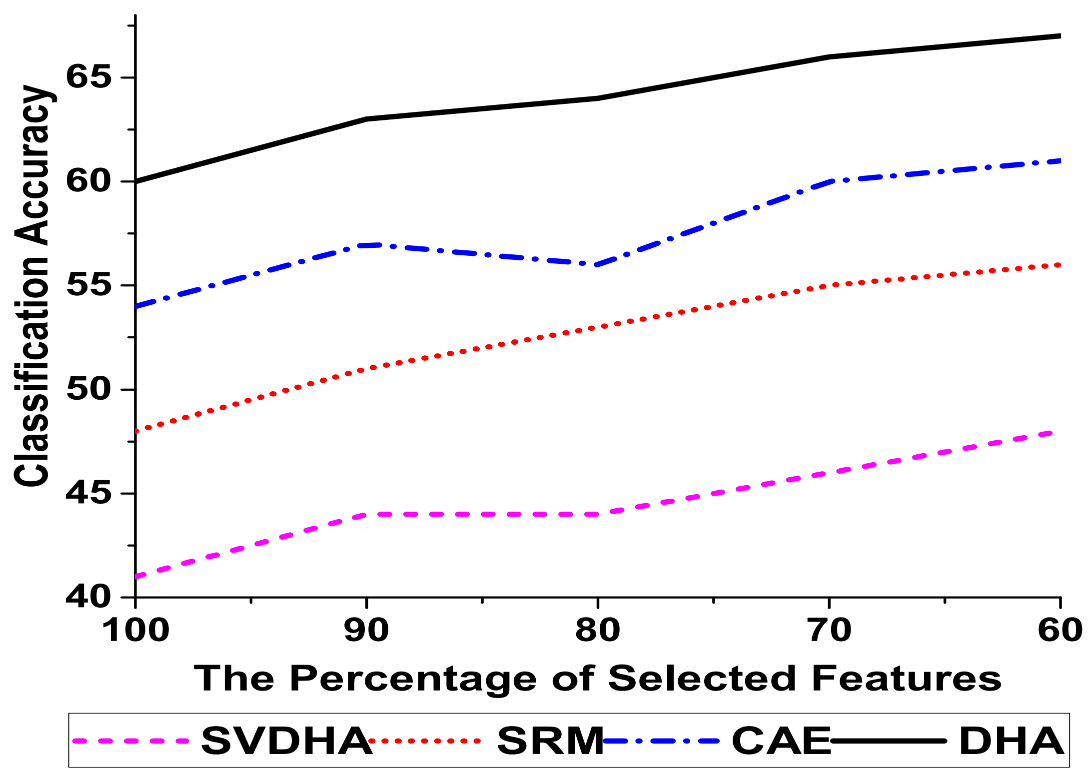
↓Algorithms, Datasets→	DS005	DS105	DS107	DS116	DS117
ν-SVM [17]	68.37±1.01	21.76±0.91	36.84±1.45	62.49±1.34	70.17±0.59
HA [1]	70.32±0.92	28.91±1.03	40.21±0.33	70.67±0.97	76.14±0.49
RHA [2]	82.22±0.42	30.35±0.39	43.63±0.61	76.34±0.45	81.54±0.92
KHA [3]	80.91±0.21	36.23±0.57	50.41±0.92	75.28±0.94	80.92±0.28
SVD-HA [4]	88.54±0.71	37.61±0.62	57.54±0.31	78.66±0.82	92.14±0.42
SRM [5]	90.23±0.74	44.48±0.75	62.41±0.72	79.20±0.98	93.65±0.93
SL [9]	89.79±0.25	47.32±0.92	61.84±0.32	80.63±0.81	93.26±0.72
CAE [6]	91.24±0.61	52.16±0.63	**72.33±0.79**	87.53±0.72	91.49±0.33
DHA	**96.91±0.82**	**59.57±0.32**	70.23±0.92	**89.93±0.24**	**96.13±0.32**



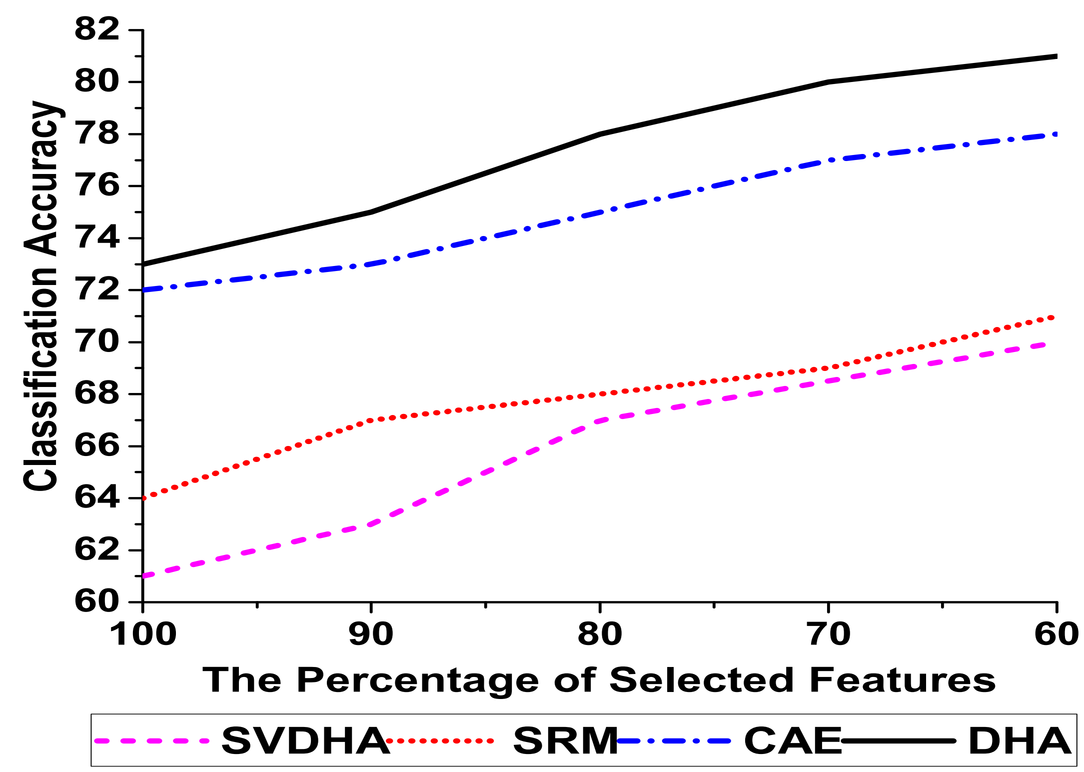
(a) Forrest Gump (TRs = 100)
(b) Forrest Gump (TRs = 400)
(c) Forrest Gump (TRs = 800)
(d) Forrest Gump (TRs = 2000)
(e) Raiders (TRs = 100)
(f) Raiders (TRs = 400)
(g) Raiders (TRs = 800)
(h) Raiders (TRs = 2000)

Figure 1: Comparison of different HA algorithms on complex task datasets by using ranked voxels.

Classification analysis by using feature selection

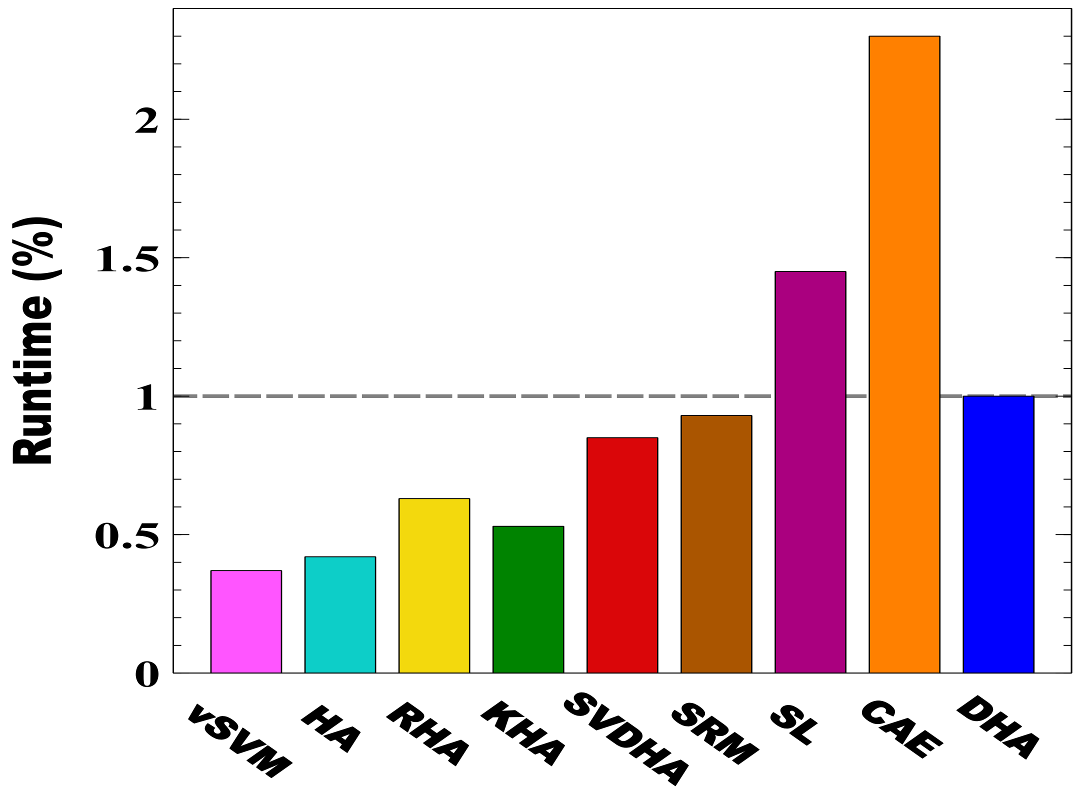


(A) DS105

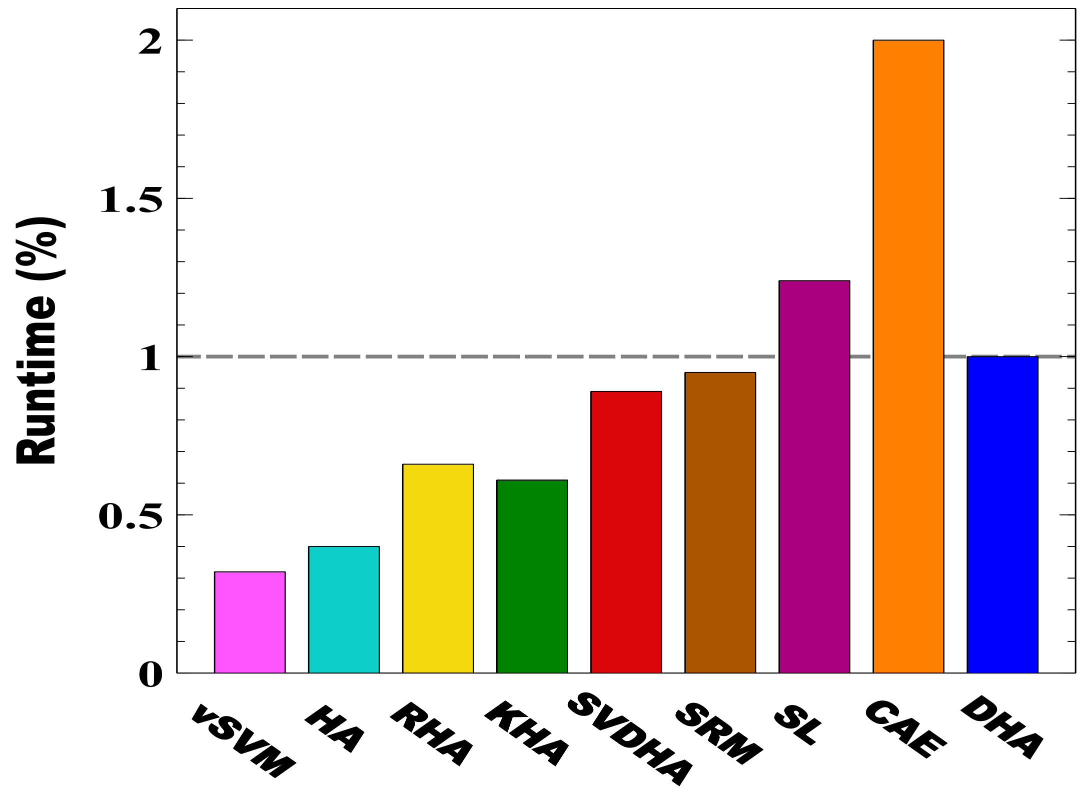


(B) DS107

Runtime Analysis



(A) DS105



(B) DS107

Future
Works

Future Works

- This paper extended a deep approach for hyperalignment methods in order to provide accurate functional alignment in multi-subject fMRI analysis.

- Deep Hyperalignment (DHA) can handle fMRI datasets with nonlinearity, high-dimensionality (broad ROI), and a large number of subjects. Further, its time complexity fairly scales with data size and the training data is not referenced when DHA computes the functional alignment for a new subject.

- In the future, we will plan to employ DHA for improving the performance of other techniques in fMRI analysis, e.g. Representational Similarity Analysis (RSA).

Thank You!

Q & A

For more details, contact:

myousefnezhad@nuaa.edu.cn

myousefnezhad@outlook.com

https://myousefnezhad.github.io/