Deep Hyperalignment

iBRAIN

Muhammad Yousefnezhad, Daoqiang Zhang

School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China

MOTIVATION

★ A generalized approach for classical Hyperalignment (HA):

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

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

If $f(\mathbf{x}) = \mathbf{x}$ & $\epsilon = 0$, then we have the original HA

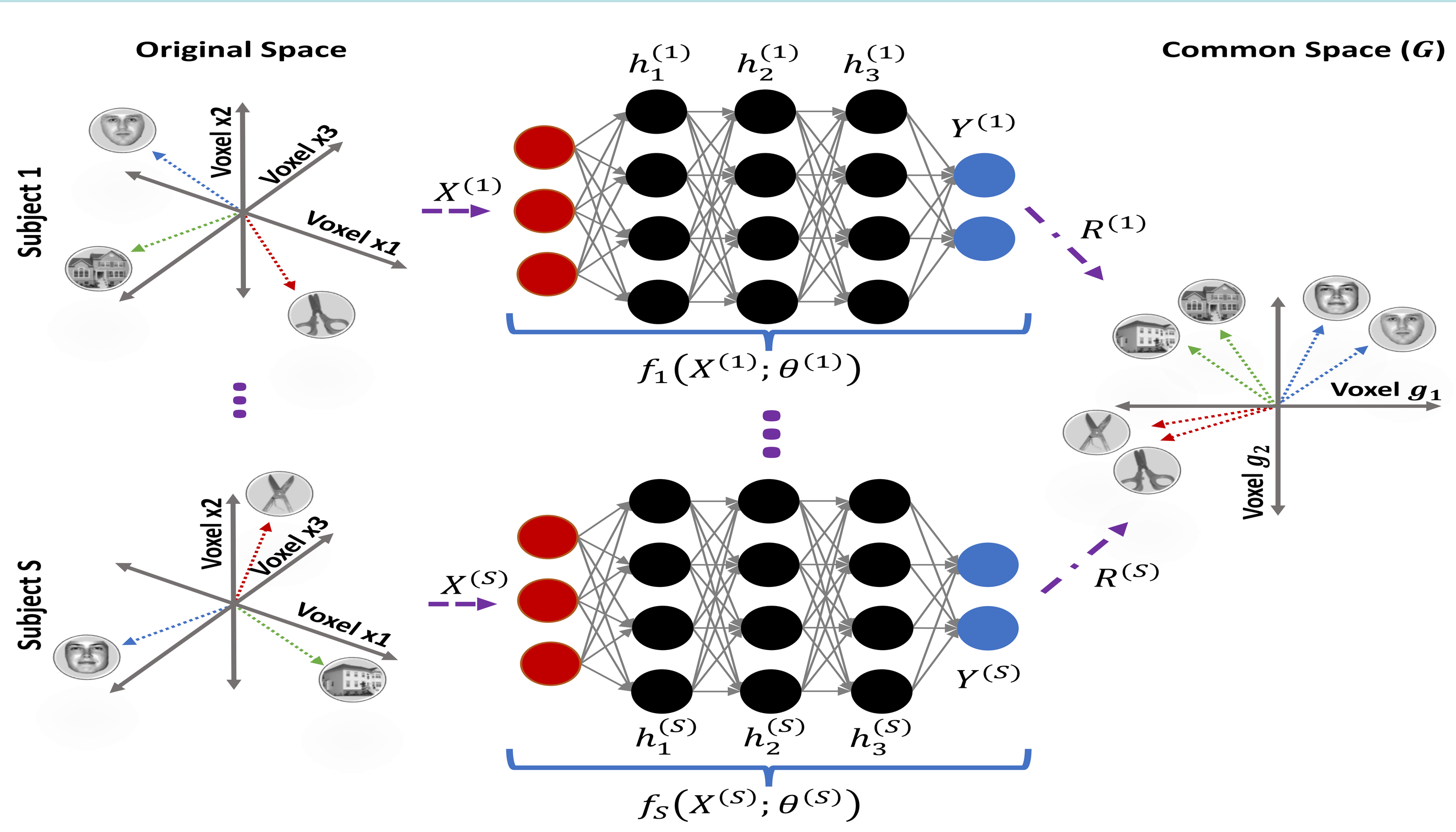
If $f(\mathbf{x}) = \mathbf{x}$ & $\epsilon \neq 0$, then we have the Regularized HA

If $f(\mathbf{x})$ is a nonlinear kernel, then we have the Kernel HA

$\mathbf{X}^{(\ell)}$: denotes brain activities
$\mathbf{R}^{(\ell)}$: is DHA mapping

➡ **These methods are limited by a restricted fixed kernel function.**

METHOD



Graphical abstract of Deep Hyperalignment (DHA)

★ Objective Function:

$$\min_{\mathbf{G},\mathbf{R}^{(i)},\theta^{(i)}} \sum_{i=1}^{S} \left\| \mathbf{G} - f_i(\mathbf{X}^{(i)};\theta^{(i)})\mathbf{R}^{(i)} \right\|_F^2 \qquad \textbf{s.t.}\quad \mathbf{G}^\top\mathbf{G} = \mathbf{I}$$

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

$\theta^{(\ell)}$: Network Parameters
g(): Nonlinear Function
Vnew: # of voxels after mapping
C: # of hidden layers
T: Timepoints

where Deep network is defined as follows:

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

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

★ Optimization:

STEP 1: DHA shared space can be calculated as follows:

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

where A is sum of projection matrices:

$$\mathbf{P}^{(\ell)} = f_\ell(\mathbf{X}^{(\ell)};\theta^{(\ell)}) \left(\left(f_\ell(\mathbf{X}^{(\ell)};\theta^{(\ell)}) \right)^\top f_\ell(\mathbf{X}^{(\ell)};\theta^{(\ell)}) + \epsilon\mathbf{I} \right)^{-1} \left(f_\ell(\mathbf{X}^{(\ell)};\theta^{(\ell)}) \right)^\top, \quad \mathbf{A} = \sum_{i=1}^{S} \mathbf{P}^{(i)}$$

STEP 2: DHA mappings can be calculated as follows:

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

STEP 3: Deep Network parameters can be iteratively calculated as follows by using SGD:

$$\theta^{(\ell)} \leftarrow \theta^{(\ell)} - \eta\nabla\theta^{(\ell)}, \quad \textbf{where}\quad \mathbf{Z} = \sum_{\ell=1}^{T}\lambda_\ell \quad \textbf{and}$$

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

DATASETS

★ This paper utilizes 7 datasets for running empirical studies:

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.

EXPERIMENTAL RESULTS

★ Simple Tasks Analysis:

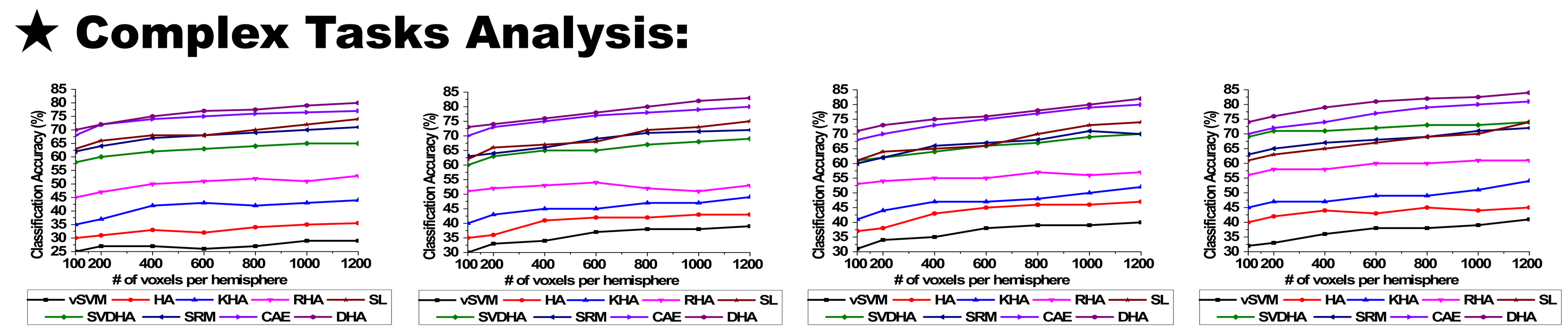
Table 1: Accuracy of HA methods in post-alignment classification by using simple task datasets

↓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**

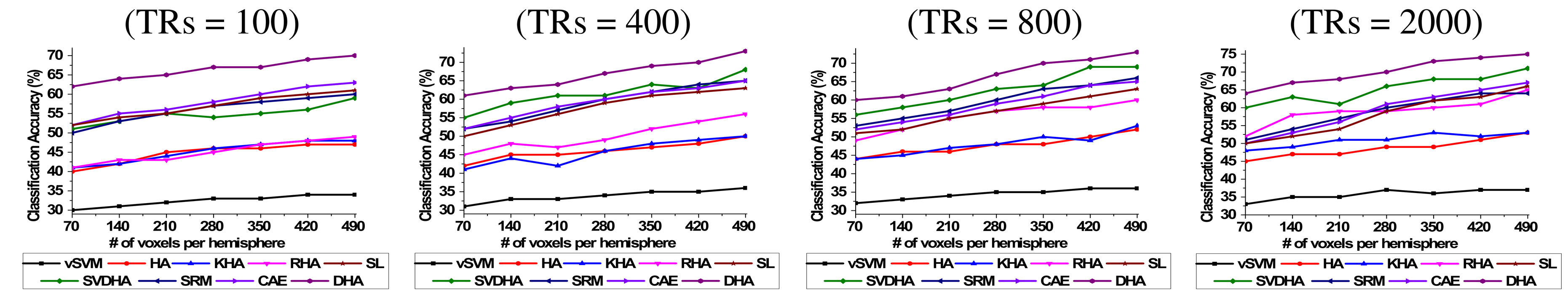
Table 2: Area under the ROC curve (AUC) of different HA methods in post-alignment classification by using simple task datasets

↓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**

★ Complex Tasks Analysis:

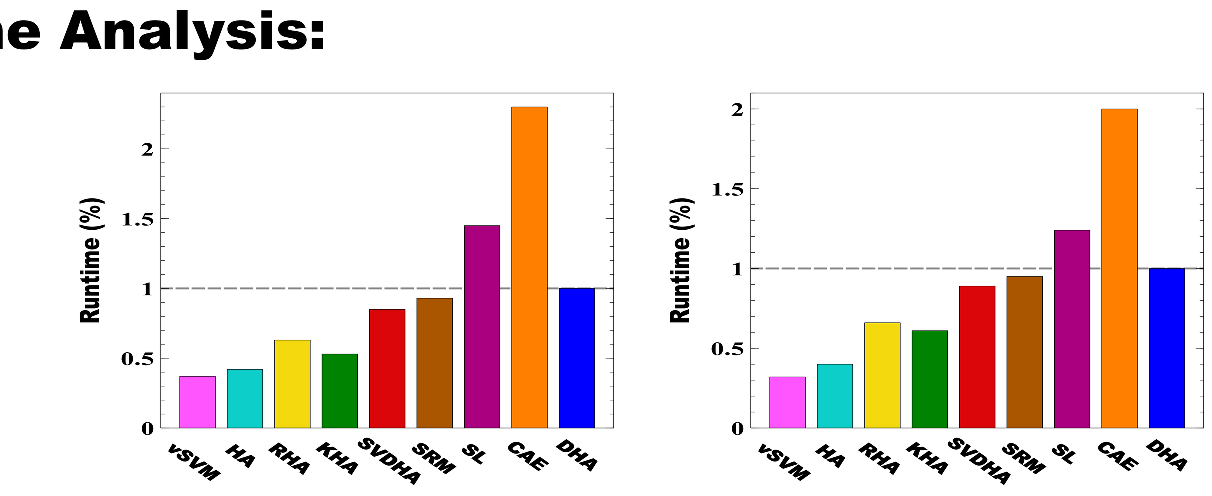


(a) Forrest Gump (TRs = 100) (b) Forrest Gump (TRs = 400) (c) Forrest Gump (TRs = 800) (d) Forrest Gump (TRs = 2000)



(e) Raiders (TRs = 100) (f) Raiders (TRs = 400) (g) Raiders (TRs = 800) (h) Raiders (TRs = 2000)

★ Runtime Analysis:



(A) DS105 (B) DS107

CONCLUSION

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. DHA is parametric and uses rank-m SVD and stochastic gradient descent for optimization. Experimental studies on multi-subject fMRI datasets confirm that the DHA method achieves superior performance to other state-of-the-art HA algorithms. 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), multi-modality, and hub detection.