Local Discriminant Hyperalignment for multi-subject fMRI data alignment

Muhammad Yousefnezhad, Daoqiang Zhang

Nanjing University of Aeronautics and Astronautics

Presented by: Daoqiang Zhang

Motivation

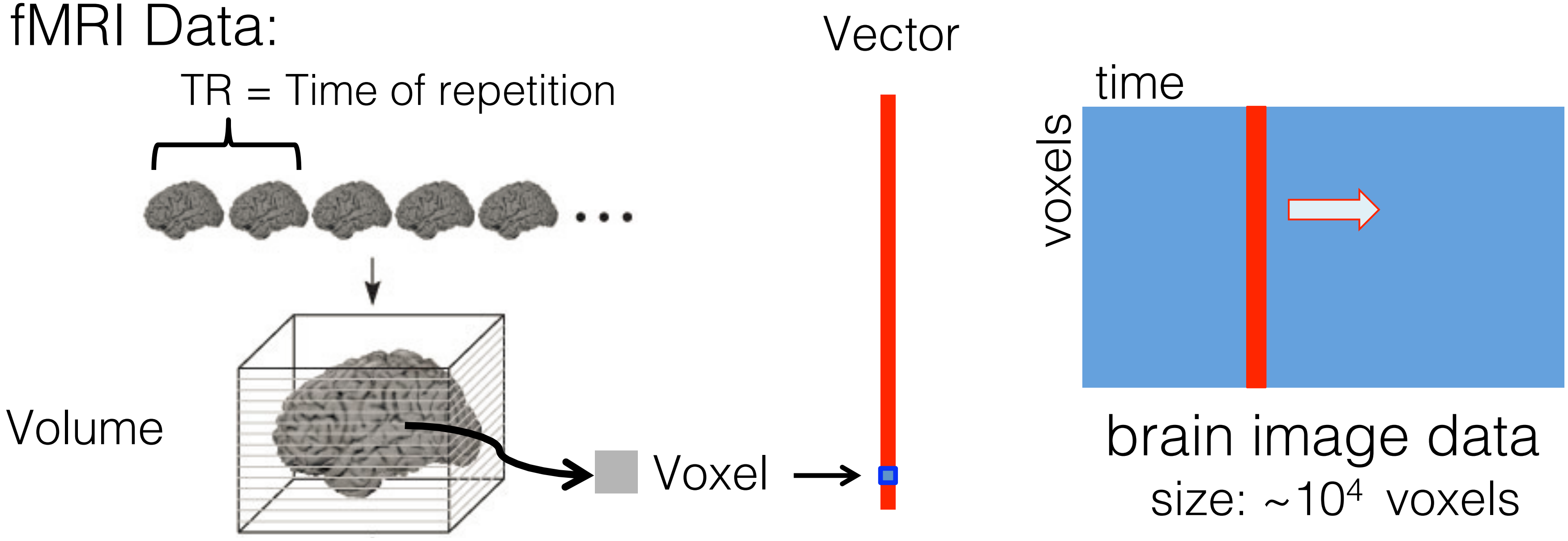
- Modern fMRI studies of human cognition use data from multiple subjects.

- Employing the supervised information in MVP methods for functional aligning the multi-subject fMRI data.

Outline

fMRI Data: Vectorization

fMRI Data:



Main disciplines in the human brain decoding

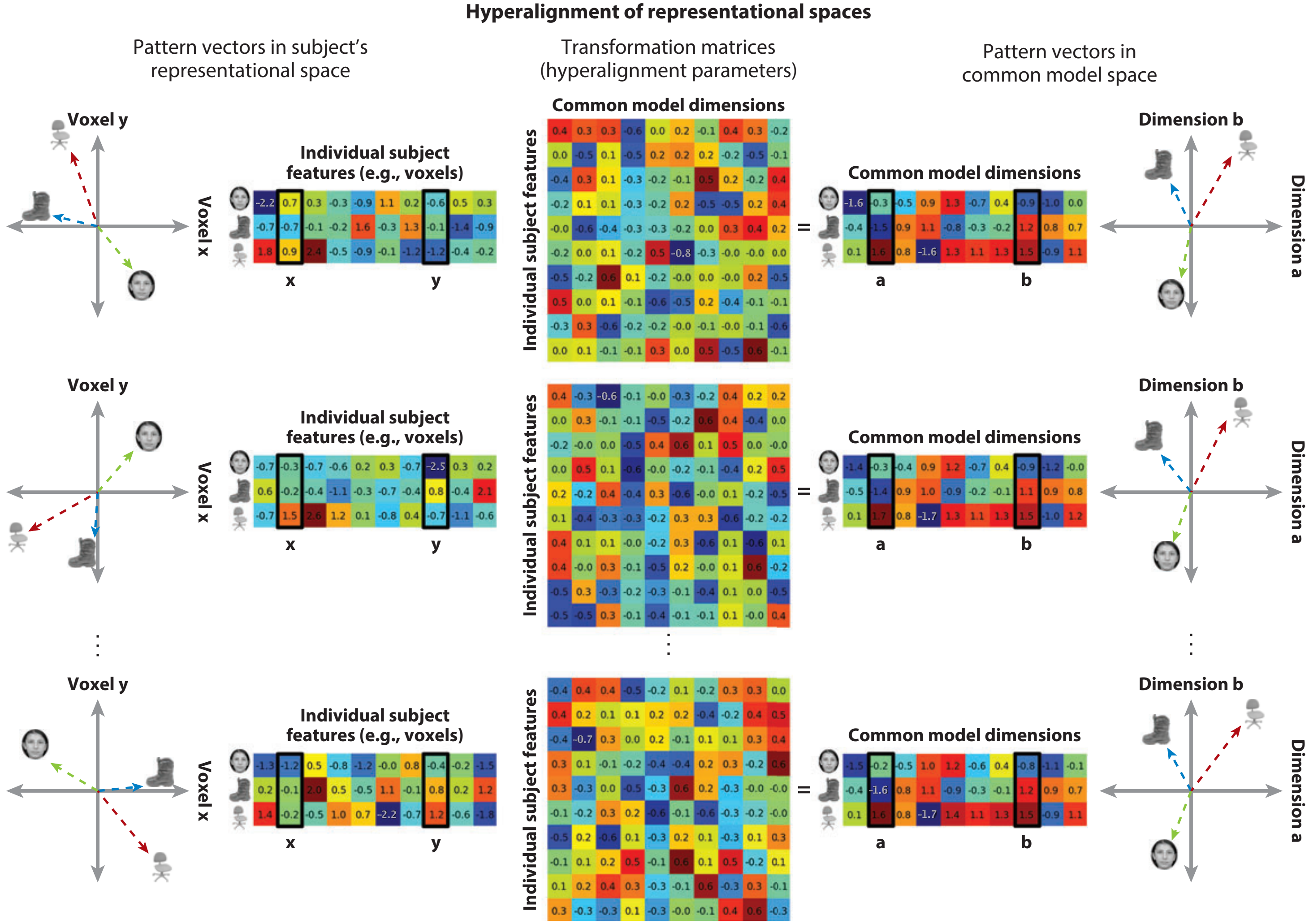
- **Multivariate Pattern Analysis (MVP)**
  - **Creating a *classification model* for new stimuli**
- **Representational Similarity Analysis (RSA)**
  - **Understanding new patterns by using *clustering***
- **Hyperalignment**
  - **Matching generated patterns in *multi-subject* problems**
- **Stimulus-model-based encoding & decoding**
  - **Matching generated models for *new category of stimuli***

Main research areas in the human brain decoding

- **Multivariate Pattern Analysis (MVP)**
  - **Creating a classification m… for n… timuli**
- **Rep…**
  - **U…**
- **…**
  - **…blems**
- **Sti…coding**
  - **Matchi… new stimuli**

This study seeks a *Supervised Hyperalignment* utilized for MVP classification problems

Hyperalignment: Representational Space



[Haxby et al. 2014]

Inter-Subject Correlation (ISC)

$$\text{ISC}(\mathbf{X}^{(i)},\mathbf{X}^{(j)}) = ({}^{1}/{}_{V})\text{tr}((\mathbf{X}^{(i)})^{\top}\mathbf{X}^{(j)}) =$$
$$\frac{1}{V}\sum_{n=1}^{V}\left(\mathbf{x}_{.n}^{(i)}\right)^{\top}\mathbf{x}_{.n}^{(j)} = \frac{1}{V}\sum_{m=1}^{V}\sum_{n=1}^{V}\mathbf{x}_{mn}^{(i)}\mathbf{x}_{mn}^{(j)}$$

- **For $i-th$ subject: $X^{(i)} = \left\{x_{mn}^{(i)}\right\} \in \mathbb{R}^{T\times V}$, where T denotes the number of time point in units of TRs, V is number of voxels.**

- **The column representation of functional activities in $n-th$ voxel:**

$$\mathbf{x}_{.n}^{(i)} \in \mathbb{R}^{T} = \left\{\mathbf{x}_{mn}^{(i)} | \mathbf{x}_{mn}^{(i)} \in \mathbf{X}^{(i)} \text{ and } m = 1{:}T\right\}$$

Hyperalignment based on ISC function

$$\rho = \underset{i,j=1:S}{\arg\max} \sum_{i<j} \mathrm{ISC}(\mathbf{X}^{(i)}\mathbf{R}^{(i)}, \mathbf{X}^{(j)}\mathbf{R}^{(j)})$$

$$= \underset{i,j=1:S}{\arg\max} \sum_{i<j} \sum_{m=1}^{V} \sum_{n=1}^{V} \mathbf{x}_{mn}^{(i)} \mathbf{r}_{nm}^{(i)} \mathbf{x}_{mn}^{(j)} \mathbf{r}_{nm}^{(j)}$$

- **where $R^{(i)} = \left\{r_{mn}^{(i)}\right\} \in \mathbb{R}^{V\times V}$ is the HA solution for $i-th$ subject.**
- **If the functional activities are column-wise standardized $X^{(i)} \sim \mathcal{N}(0,1)$, the ISC lies in $[-1,+1]$, where the large values represent better alignment.**
- **The general assumption in the basic hyperalignment is that the $R^{(i)}$ are noisy 'rotation' of a common template.**

Hyperalignment: Formulation

$$\rho = \arg\min_{i,j=1:S} \sum_{i<j} \|\mathbf{X}^{(i)}\mathbf{R}^{(i)} - \mathbf{X}^{(j)}\mathbf{R}^{(j)}\|_F^2$$

$$\text{subject to} \quad (\mathbf{R}^{(\ell)})^\top \mathbf{A}^{(\ell)} \mathbf{R}^{(\ell)} = \mathbb{I}, \qquad \ell = 1{:}S$$

- **$A^{(\ell)}, \ell = 1{:}S$ are symmetric and positive definite.**
- **$A^{(\ell)} = \mathbb{I}$: we have hyperalignment or a multi-set orthogonal Procrustes problem, which is commonly used in share analysis**
- **$A^{(\ell)} = \left(X^{(\ell)}\right)^\top X^{(\ell)}$: we have a form of multi-set Canonical Correlation Analysis (CCA).**

Hyperalignment: Formulation (cont.)

Lemma 1. *The equation* (4) *is equivalent to:*

$$\rho = \arg\min \sum_{i=1}^{S} \|\mathbf{X}^{(i)}\mathbf{R}^{(i)} - \mathbf{G}\|_F^2$$

$$\text{subject to} \qquad (\mathbf{R}^{(\ell)})^\top \mathbf{A}^{(\ell)} \mathbf{R}^{(\ell)} = \mathbb{I}, \qquad \ell = 1{:}S$$

where $\mathbf{G} \in \mathbb{R}^{T \times V}$ *is the HA template:*

$$\mathbf{G} = \frac{1}{S} \sum_{j=1}^{S} \mathbf{X}^{(j)}\mathbf{R}^{(j)}$$

- **HA template (G) can be used for functional alignment in the test stage before MVP analysis.**
- **Most of previous studies used CCA for finding this template.**

Hyperalignment: Formulation (cont.)

Lemma 2. *Canonical Correlation Analysis (CCA) finds an optimum solution for solving* (4) *by exploiting the objective function* $\max\limits_{i,j=1:S}\Big((\mathbf{R}^{(i)})^\top\mathbf{C}^{(i,j)}\mathbf{R}^{(j)}\Big)$*, and then* $\mathbf{G}$ *also can be calculated based on* (6)*. Briefly, the CCA solution can be formulated as follows:*

$$\rho = \arg\max_{i,j=1:S}\left(\frac{(\mathbf{R}^{(i)})^\top\mathbf{C}^{(i,j)}\mathbf{R}^{(j)}}{\sqrt{((\mathbf{R}^{(i)})^\top\mathbf{C}^{(i)}\mathbf{R}^{(i)})((\mathbf{R}^{(j)})^\top\mathbf{C}^{(j)}\mathbf{R}^{(j)})}}\right) \tag{7}$$

where $\mathbf{C}^{(i)} \in \mathbb{R}^{V\times V} = \mathbb{E}\Big[(\mathbf{X}^{(i)})^\top\mathbf{X}^{(i)}\Big] = (\mathbf{X}^{(i)})^\top\mathbf{X}^{(i)}$*,* $\mathbf{C}^{(j)} \in \mathbb{R}^{V\times V} = \mathbb{E}\Big[(\mathbf{X}^{(j)})^\top\mathbf{X}^{(j)}\Big] = (\mathbf{X}^{(j)})^\top\mathbf{X}^{(j)}$*, and* $\mathbf{C}^{(i,j)} \in \mathbb{R}^{V\times V} = \mathbb{E}\Big[(\mathbf{X}^{(i)})^\top\mathbf{X}^{(j)}\Big] = (\mathbf{X}^{(i)})^\top\mathbf{X}^{(j)}$*. The solution of CCA can be obtained by computing a generalized eigenvalue decomposition problem*

Outline

Remark

- Consider fMRI time series included visual stimuli, where **two subjects** watch two photos of cats as well as two photos of human faces:

Stimuli sequence: $[\text{cat1}, \text{face1}, \text{cat2}, \text{face2}]$

- The unsupervised solution finds two mappings to maximize the correlation in the voxel-level, where **the voxels for each subject** are only compared with the **voxels for other subjects with the same locations**.

- **Unsupervised HA solution** is shown by:

$$(\mathbf{S1{:}cat1} \uparrow \mathbf{S2{:}cat1})\ ;\ (\mathbf{S1{:}face1} \uparrow \mathbf{S2{:}face1});$$
$$(\mathbf{S1{:}cat2} \uparrow \mathbf{S2{:}cat2})\ ;\ (\mathbf{S1{:}face2} \uparrow \mathbf{S2{:}face2})$$

Remark (cont.)

- The CCA solution here just **maximized the correlation for the stimuli in the same locations**, while they must also **maximize the correlation between all stimuli in the same category** and **minimize the correlation between different categories of stimuli**.

- Our approach for solving mentioned issues can be shown by:

$$(\mathbf{S1}{:}\mathbf{cat1,2} \uparrow \mathbf{S2}{:}\mathbf{cat1,2});\ (\mathbf{S1}{:}\mathbf{face1,2} \uparrow \mathbf{S2:face1,2});$$
$$(\mathbf{S1}{:}\mathbf{cat1,2} \downarrow \mathbf{S2}{:}\mathbf{face1,2});\ (\mathbf{S1}{:}\mathbf{face1,2} \downarrow \mathbf{S2}{:}\mathbf{cat1,2})$$

Remark (cont.)

- **The CCA solution here just maximized the correlation for the stimuli in the same locations, while they must also maximize the correlation between all stimuli in the same category and minimize the correlation between different categories of stimuli.**

- **Our approach for solving mentioned issues can be shown by:**

within-class terms

$$(\mathbf{S1}{:}\mathbf{cat1,2} \uparrow \mathbf{S2}{:}\mathbf{cat1,2});\ (\mathbf{S1}{:}\mathbf{face1,2} \uparrow \mathbf{S2} : \mathbf{face1,2});$$

$$(\mathbf{S1}{:}\mathbf{cat1,2} \downarrow \mathbf{S2}{:}\mathbf{face1,2});\ (\mathbf{S1}{:}\mathbf{face1,2} \downarrow \mathbf{S2}{:}\mathbf{cat1,2})$$

Remark (cont.)

- **The CCA solution here just maximized the correlation for the stimuli in the same locations, while they must also maximize the correlation between all stimuli in the same category and minimize the correlation between different categories of stimuli.**

- **Our approach for solving mentioned issues can be shown by:**

$$(\mathbf{S1}{:}\mathbf{cat1,2}\uparrow\mathbf{S2}{:}\mathbf{cat1,2});\ (\mathbf{S1}{:}\mathbf{face1,2}\uparrow\mathbf{S2}:\mathbf{face1,2});$$

$$(\mathbf{S1}{:}\mathbf{cat1,2}\downarrow\mathbf{S2}{:}\mathbf{face1,2});\ (\mathbf{S1}{:}\mathbf{face1,2}\downarrow\mathbf{S2}{:}\mathbf{cat1,2})$$

between-classes terms

Local Discriminant Hyperalignment (LDHA)

- **This paper proposes Local Discriminant Hyperalignment (LDHA), which combines the idea of locality into CCA.**

- **Since unaligned (before applying the HA method) functional activities in different subjects cannot be directly compared with each other, the neighborhoods matrix α is defined as follows:**

$$\alpha_{nm} = \alpha_{mn} = \begin{cases} 0 & \mathbf{y}_m \neq \mathbf{y}_n \\ 1 & \mathbf{y}_m = \mathbf{y}_n \end{cases}, \quad m, n = 1{:}T, m < n$$

where $\boldsymbol{Y} = \{\boldsymbol{y_m}\} \in \mathbb{R}^{\boldsymbol{T}}$ is class labels in the train-set.

LDHA (cont.)

- **Within-class neighborhoods** $W^{(i,j)} = \left\{w_{mn}^{(i,j)}\right\} \in \mathbb{R}^{V \times V}$**:**

$$\mathbf{w}_{mn}^{(i,j)} = \sum_{\ell=1}^{T}\sum_{k=1}^{T} \alpha_{\ell k}\mathbf{x}_{\ell m}^{(i)}\mathbf{x}_{kn}^{(j)} + \alpha_{\ell k}\mathbf{x}_{\ell n}^{(i)}\mathbf{x}_{km}^{(j)}$$

- **Between-classes neighborhoods** $B^{(i,j)} = \left\{b_{mn}^{(i,j)}\right\} \in \mathbb{R}^{V \times V}$**:**

$$\mathbf{b}_{mn}^{(i,j)} = \sum_{\ell=1}^{T}\sum_{k=1}^{T} (1-\alpha_{\ell k})\mathbf{x}_{\ell m}^{(i)}\mathbf{x}_{kn}^{(j)} + (1-\alpha_{\ell k})\mathbf{x}_{\ell n}^{(i)}\mathbf{x}_{km}^{(j)}$$

LDHA (cont.)

- **Supervised Covariance matrix:**

$$\widetilde{\boldsymbol{C}}^{(\boldsymbol{i},\boldsymbol{j})} = \boldsymbol{W}^{(\boldsymbol{i},\boldsymbol{j})} - \left(\frac{\boldsymbol{\eta}}{\boldsymbol{T}^2}\right)\boldsymbol{B}^{(\boldsymbol{i},\boldsymbol{j})}$$

- $\boldsymbol{\eta}$ **is the number of non-zero cells in the matrix** $\boldsymbol{\alpha}$**, and** $\boldsymbol{T}$ **is the number of time points in unites of TRs.**

- **LDHA objective function is denoted as follows:**

$$\rho = \underset{i,j=1:S, i<j}{\arg\max} \frac{(\mathbf{R}^{(i)})^\top \widetilde{\mathbf{C}}^{(i,j)} \mathbf{R}^{(j)}}{\sqrt{((\mathbf{R}^{(i)})^\top \mathbf{C}^{(i)} \mathbf{R}^{(i)})((\mathbf{R}^{(j)})^\top \mathbf{C}^{(j)} \mathbf{R}^{(j)})}}$$

$$\text{subject to} \quad (\mathbf{R}^{(\ell)})^\top \mathbf{C}^{(\ell)} \mathbf{R}^{(\ell)} = \mathbb{I}, \qquad \ell = 1{:}S$$

LDHA Algorithm

Algorithm 1 Local Discriminate Hyperalignment (LDHA)

Input: Data points $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(j)}$, class labels $\mathbf{Y}$:
Output: Hyperalignment parameters $\mathbf{R}^{(i)}$ and $\mathbf{R}^{(j)}$:
Method:

1. Generate α by (9).
2. Calculate $\mathbf{W}^{(i,j)}$, $\mathbf{B}^{(i,j)}$ by using (10) and (11).
3. Calculate $\widetilde{\mathbf{C}}^{(i,j)}$.
4. Compute $\mathbf{H}^{(i,j)} = \left(\mathbf{C}^{(i)}\right)^{-1/2} \widetilde{\mathbf{C}}^{(i,j)} \left(\mathbf{C}^{(j)}\right)^{-1/2}$.
5. Perform SVD: $\mathbf{H}^{(i,j)} = \mathbf{P}^{(i,j)} \mathbf{\Lambda}^{(i,j)} \left(\mathbf{Q}^{(i,j)}\right)^{\top}$.
6. Return $\mathbf{R}^{(i)} = \left(\mathbf{C}^{(i)}\right)^{-1/2} \mathbf{P}^{(i,j)}$

   and $\mathbf{R}^{(j)} = \left(\mathbf{C}^{(j)}\right)^{-1/2} \mathbf{Q}^{(i,j)}$.

A MVP template based on LHDA

Algorithm 2 A general template for MVP analysis by using Local Discriminate Hyperalignment (LDHA)

Input: Train Set $\mathbf{X}^{(i)}, i = 1{:}S$, Test Set $\widehat{\mathbf{X}}^{(j)}, j = 1{:}\hat{S}$:
Output: Classification Performance (ACC, AUC):
Method:

01. Initiate $\mathbf{R}^{(i)}, i = 1{:}S$.
02. **Do**
03. **Foreach** subject $\mathbf{X}^{(i)}, i = 1{:}S$:
04. Update $\mathbf{R}^{(i)}$ by Alg. 1 and $\mathbf{X}^{(\ell)}, \ell = i{+}1{:}S$.
05. **End Foreach**
06. **Until** $\mathbf{X}^{(i)}\mathbf{R}^{(i)}, i = 1{:}S$ do not change in this step.
07. Train a classifier by $\mathbf{X}^{(i)}\mathbf{R}^{(i)}, i = 1{:}S$
08. Initiate $\widehat{\mathbf{R}}^{(j)}, j = 1{:}\hat{S}$.
09. Generate $\mathbf{G}$ based on (6) by using $\mathbf{R}^{(i)}, i = 1{:}S$
10. **Foreach** subject $\widehat{\mathbf{X}}^{(j)}, j = 1{:}\hat{S}$:
11. Compute $\widehat{\mathbf{R}}^{(j)}$ by *classical* HA (Eq. 5,7) and $\mathbf{G}$.
12. **End Foreach**
13. Evaluate the classifier by using $\widehat{\mathbf{X}}^{(j)}\widehat{\mathbf{R}}^{(j)}, j = 1{:}\hat{S}$.

Outline

Simple Tasks Analysis

Table 1: Accuracy of Classification Methods

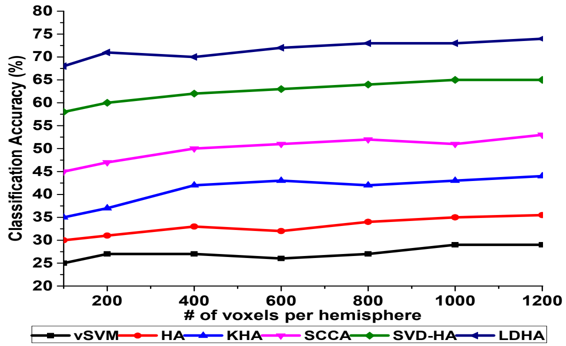
Data Sets	ν-SVM	HA	KHA	SCCA	SVD-HA	LDHA
DS005 (2 classes)	71.65±0.97	81.27±0.59	83.06±0.36	85.29±0.49	90.82±1.23	**94.32±0.16**
DS105 (8 classes)	22.89±1.02	30.03±0.87	32.62±0.52	37.14±0.91	40.21±0.83	**54.04±0.09**
DS107 (4 classes)	38.84±0.82	43.01±0.56	46.82±0.37	52.69±0.69	59.54±0.99	**74.73±0.19**
DS117 (2 classes)	73.32±1.67	77.93±0.29	84.22±0.44	83.32±0.41	**95.62±0.83**	95.07±0.27

Table 2: Area Under the ROC Curve (AUC) of Classification Methods

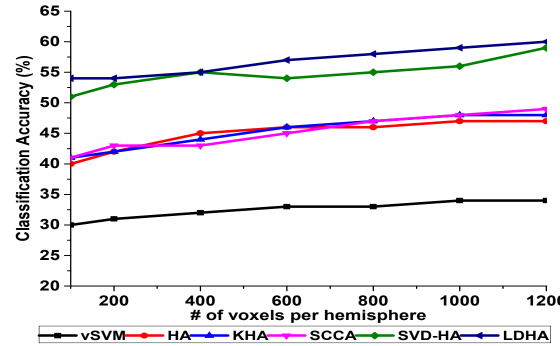
Data Sets	ν-SVM	HA	KHA	SCCA	SVD-HA	LDHA
DS005 (2 classes)	68.37±1.01	70.32±0.92	82.22±0.42	80.91±0.21	88.54±0.71	**93.25±0.92**
DS105 (8 classes)	21.76±0.91	28.91±1.03	30.35±0.39	36.23±0.57	37.61±0.62	**53.86±0.17**
DS107 (4 classes)	36.84±1.45	40.21±0.33	43.63±0.61	50.41±0.92	57.54±0.31	**72.03±0.37**
DS117 (2 classes)	70.17±0.59	76.14±0.49	81.54±0.92	80.92±0.28	92.14±0.42	**94.23±0.94**

DS005: Mixed-gambles task
DS105: Visual Object Recognition
DS107: Word and Object Processing
DS117: Multi-subject, multi-modal human neuroimaging dataset

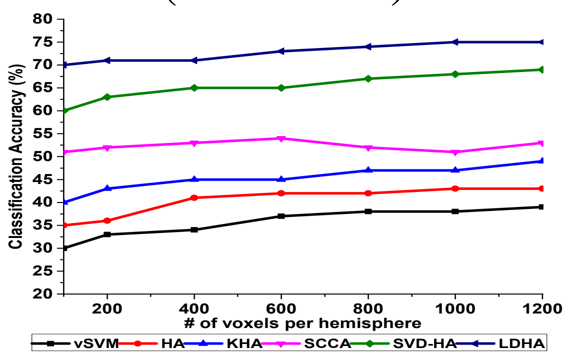
Complex Tasks Analysis



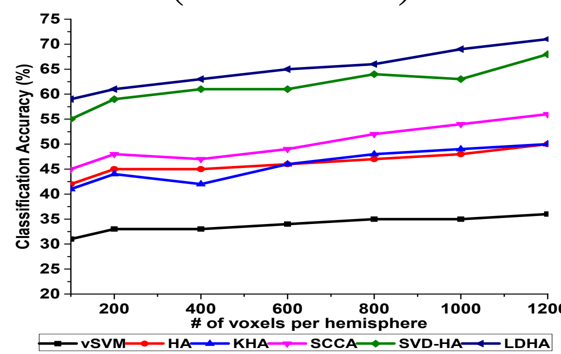
(a) Forrest Gump (TRs = 100)



(b) Raiders of the Lost Ark (TRs = 100)

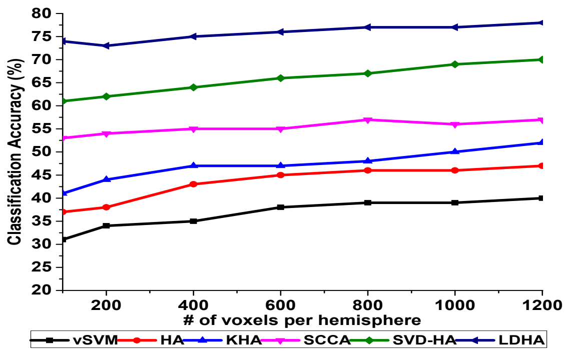


(c) Forrest Gump (TRs = 200)

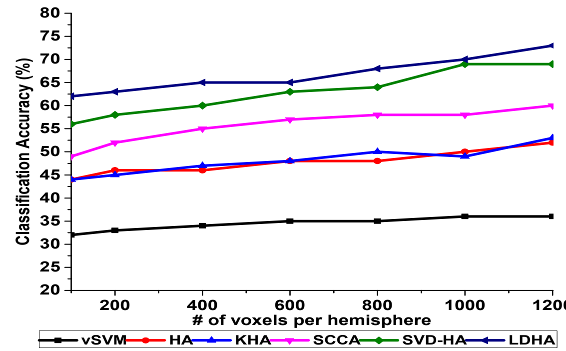


(d) Raiders of the Lost Ark (TRs = 200)

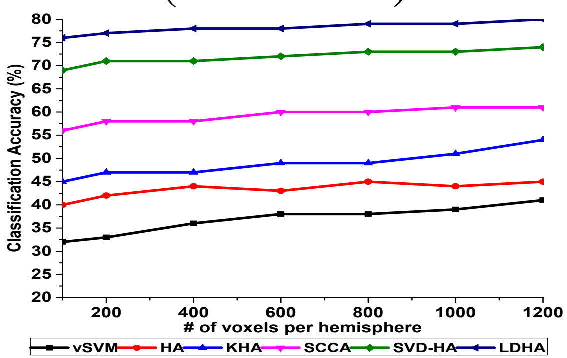
Complex Tasks Analysis (cont.)



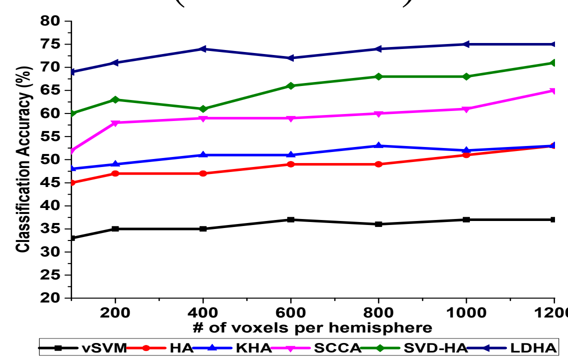
(e) Forrest Gump
(TRs = 400)



(f) Raiders of the Lost Ark
(TRs = 400)



(g) Forrest Gump
(TRs = 2000)



(h) Raiders of the Lost Ark
(TRs = 2000)

Outline

Conclusion

- **We propose LDHA method for MVP classification by combining the idea of locality into CCA.**

- **Experimental studies on multi-subject MVP analysis demonstrate that the LDHA method achieves superior performance to other state-of-the-art HA algorithms.**

- **We will plan to develop:**
  - ✓ **A kernel-based version of LDHA.**
  - ✓ **Whole-brain hyperalignment approach based on LDHA.**

Thanks for your attention!

For more details, contact:

dqzhang@nuaa.edu.cn

http://parnec.nuaa.edu.cn/zhangdq/