Towards Learning-Based Holistic Brain Image Segmentation

Zhuowen Tu

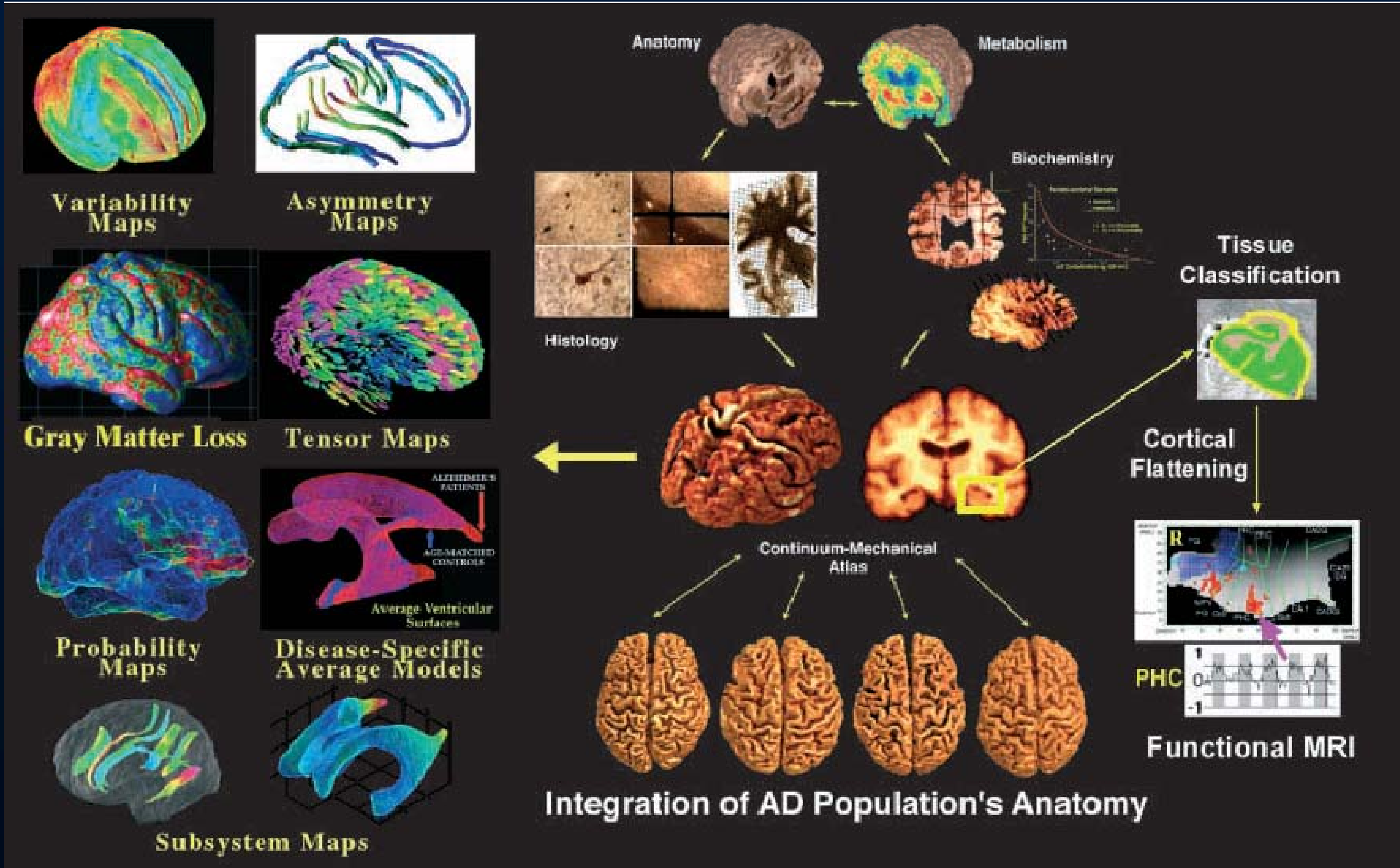
Lab of Neuro Imaging

University of California, Los Angeles

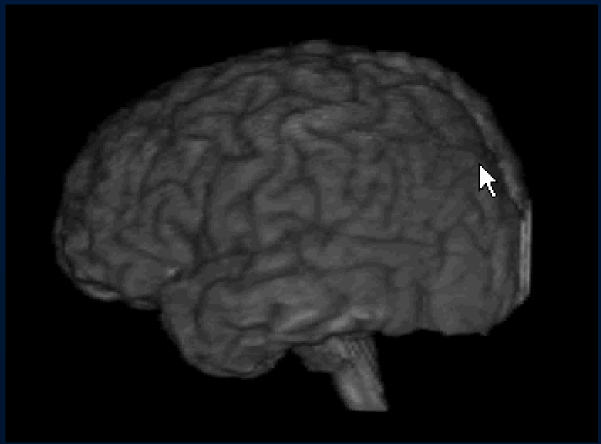
in collaboration with (A. Toga, P. Thompson et al.)

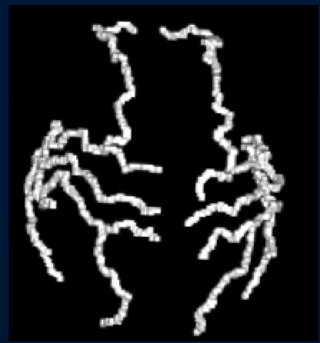
Supported by (NIH CCB Grant U54 RR021813)

Building A Brain Atlas
(Thompson and Toga 2004)

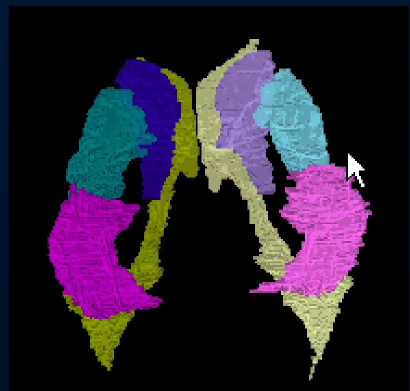


Brain Anatomical Structure Parsing

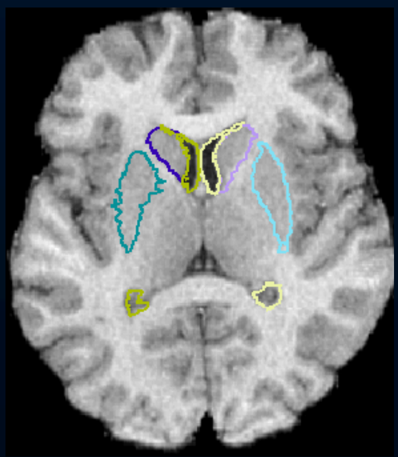


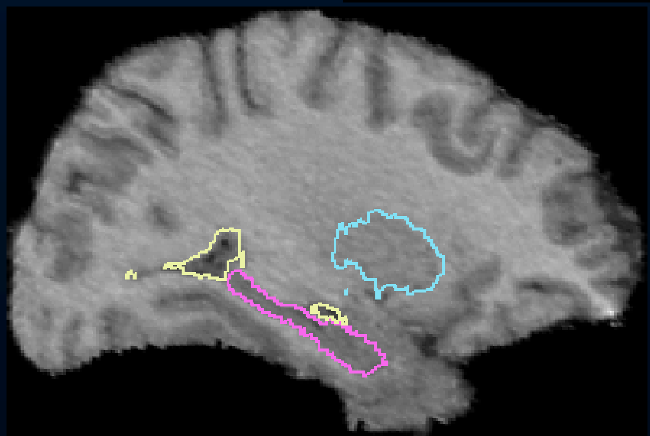


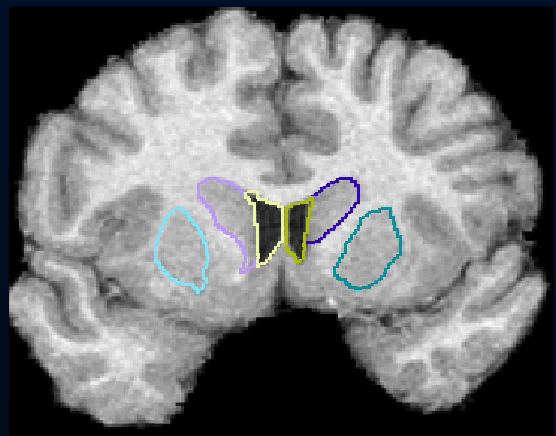
cortical structures: major sulci curves



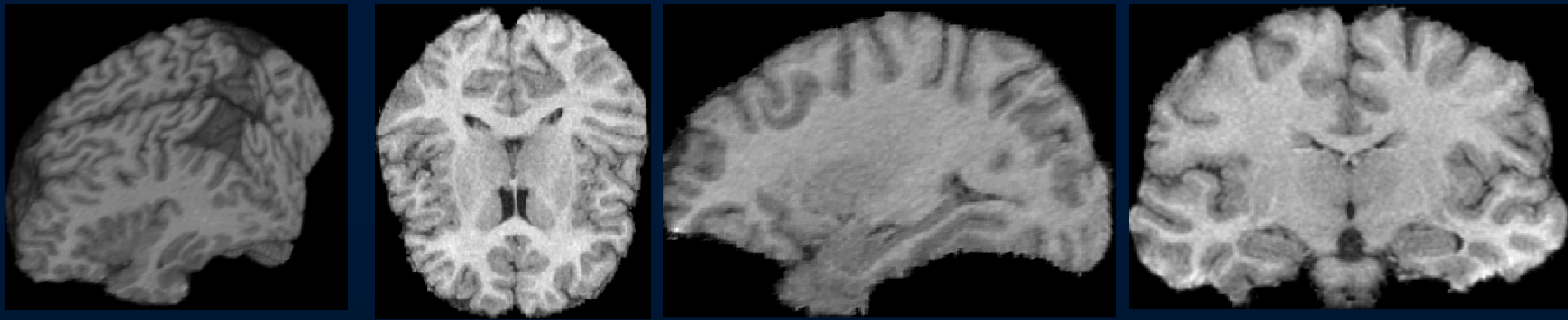
sub-cortical structures







Challenges for Automatic Segmentation



1. Large volume size for high resolution 3d MRI.
2. Very weak intensity patterns. (large inter-class similarity and intra-class variation)
3. Hard to capture 3D shape info due to the high dimension space and limited number of training data.
4. Hard to capture the high-level knowledge and adapt to different protocols.

Existing Work

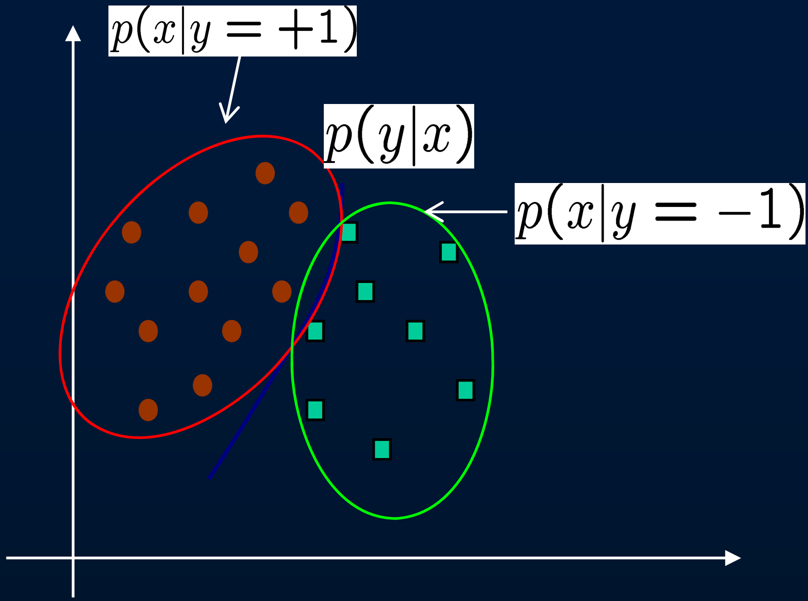
Generative model (shape) driven:

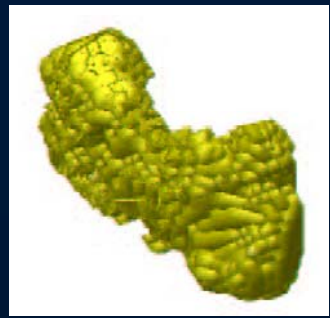
- •Markov random fields (Fischl et al. 2002)
- •Active shape model (Cootes et al. 2001)
- •M-rep (Pizer et al.)
- •Joint PCA shape constraints (Yang et al. 2004)
- •Atlas-based (Li et al. 1993)

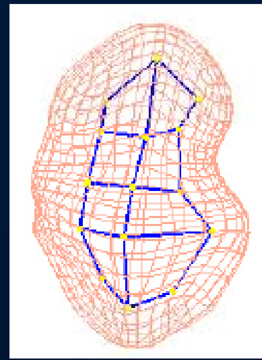
Classification-based (discriminative):

- •Knowledge-based (Li et al. 1993)
- •Feature classification (Liu et al. 2004)
- • SVM voxel classification (Lao et al. 2004)

Comparisons



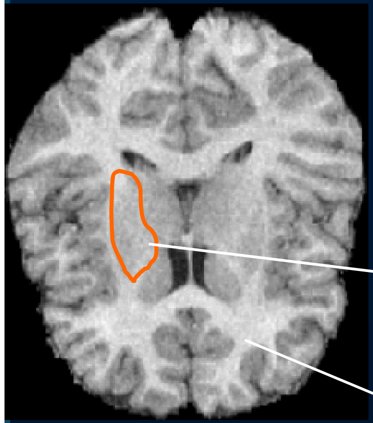

Yang et al. 2004


Pizer et al. 2003

Algorithms	Appearance Model	Shape Model	Inference
Fischl et al. [4]	generative: i.i.d. Gaussians	generative: local constraints	expectation maximization
Yang et al. [5]	generative: i.i.d. Gaussians	generative: PCA on shape	variational method
Pohl et al. [6]	generative: i.i.d. Gaussians	generative: PCA on shape	expectation maximization
Pizer et al. [7]	generative: i.i.d. Gaussians	generative: M-rep on shape	multi-scale gradient descent
Woolrich and Behrens [8]	generative: i.i.d. Gaussians	generative: local constraints	Markov Chain Monte Carlo
Li et al. [9]	discriminative: rule-based	None	rule-based classification
Rohlfing et al. [10]	discriminative: atlas based	somewhat	voxel classification
Descombes et al. [11]	discriminative: extracted features	generative: geometric properties	Markov Chain Monte Carlo
Lao et al. [12]	discriminative: SVM	None	voxel classification
Lee et al. [13]	discriminative: SVM	generative: local constraints	iterated conditional modes

A Bayesian Framework

Input: $\mathbf{V}$ Solution: $W = (R_1, R_2, ..., R_n)$

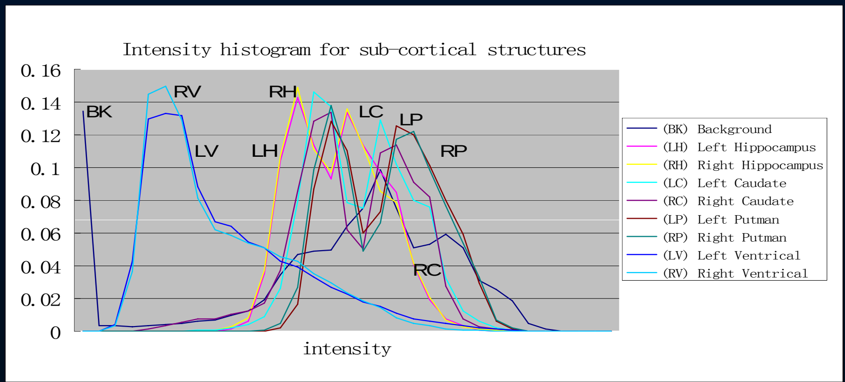


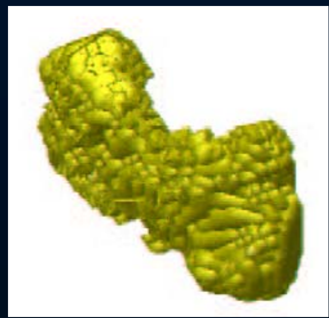
$$p(W|\mathbf{V}) \propto p(\mathbf{V}|W)p(W)$$

$$\propto \prod_{k=1}^{n} p(\mathbf{V}(R_k)|R_k)p(R_k)$$

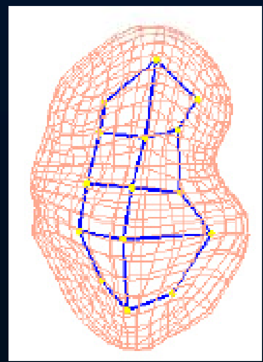
Appearance model

3D shape model





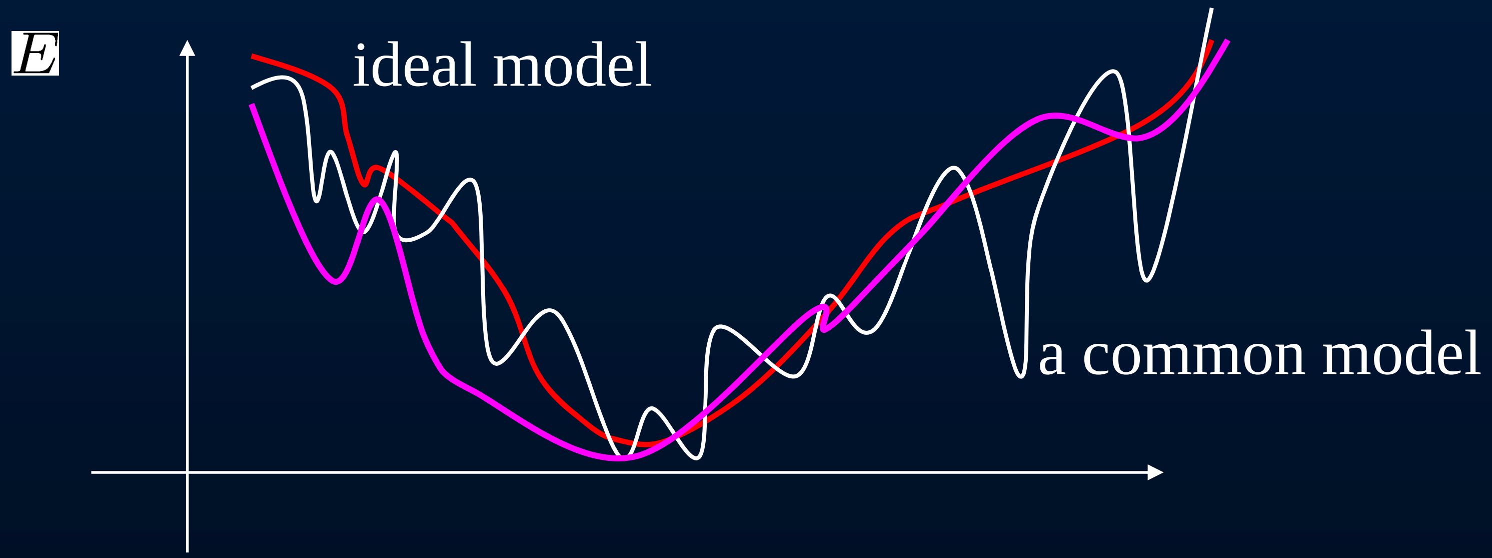
Yang et al. 2004



Pizer et al. 2003

Intensity histograms of different structures

Learning Energy



Are We Getting the Right Model?

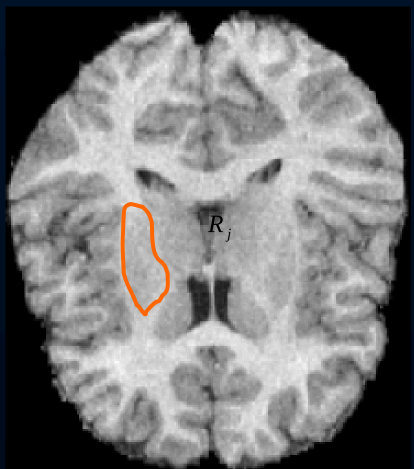
An ideal model: $$p(W|\mathbf{V}) \propto p(\mathbf{V}|W)p(W)$$

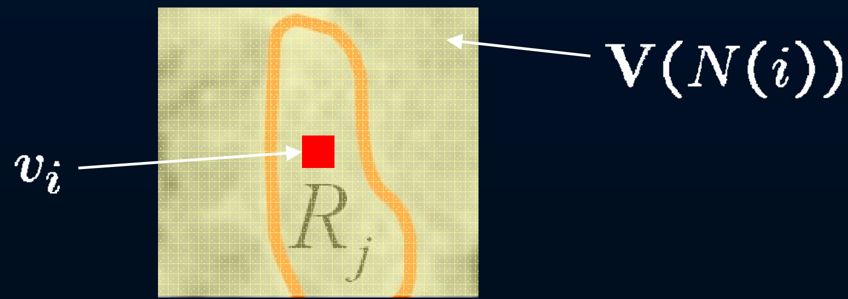
$$E(W,\mathbf{V}) = -\log p(\mathbf{V}|W) - \log p(W)$$

The full generative appearance model p(**V**|W) is very hard to obtain!

$$E_H(W,\mathbf{V}) = -\log \prod_i p(v_i, y_i = j|\mathbf{V}(N(i)/i)) - \log p(W)$$

$$\rightarrow -\log \prod_i p(y_i = j|\mathbf{V}(N(i)) - \log p(W)$$





A discriminative model (classification)!

The Algorithm

Training (given a set of annotated volumes):

(1) Learn multi-class classification model using PBT.

(2) Learn PCA shape model for each structure.

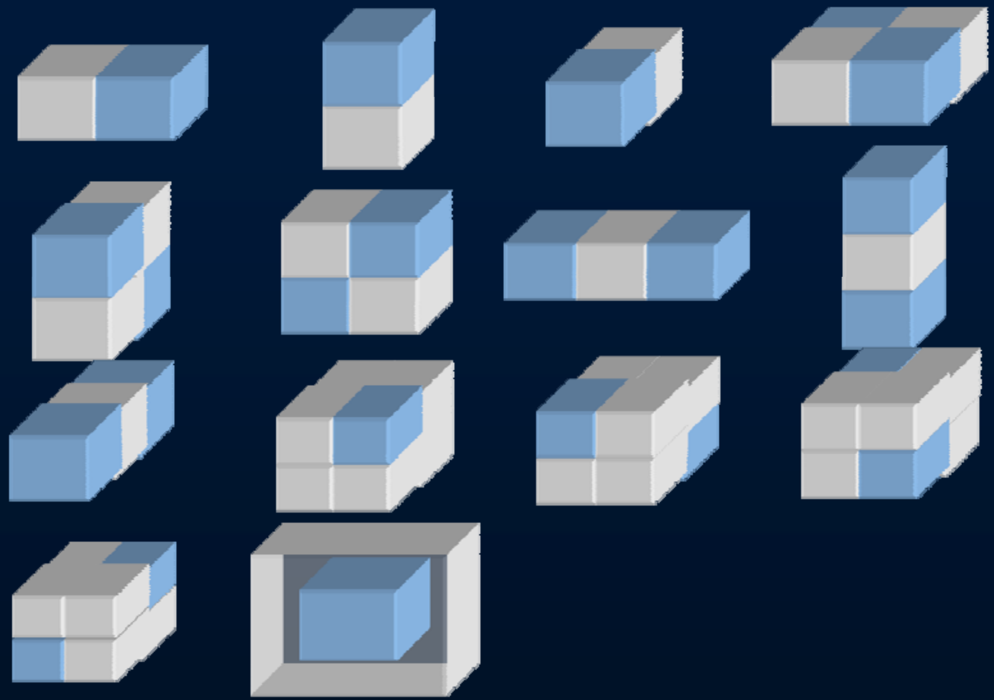
$$E = \alpha_1 \sum_{i=1}^{n} \sum_{s \in R_i} -\log p(l_s = i \,|\, V(N(s)) + \alpha_2 \sum_{i=2} -\log p(S_i) + \alpha_3 \sum_{i=1} -\Lambda(S_i)$$

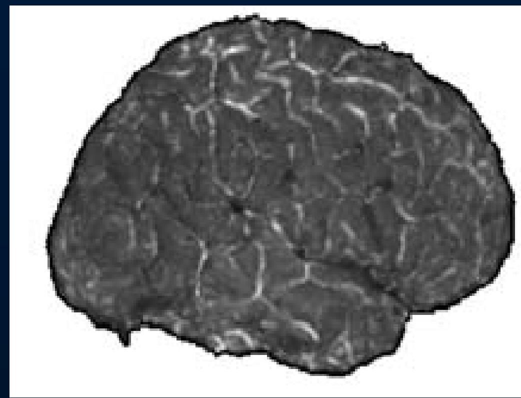
Testing (given a volume)

1. Compute classification using learned PBT.
2. Obtain the initial segmentation.
3. Perform region competition based on the proposed 3D representation.

Features

Around 10,000 features in the candidate pool: Gradients, Curvatures, Haars



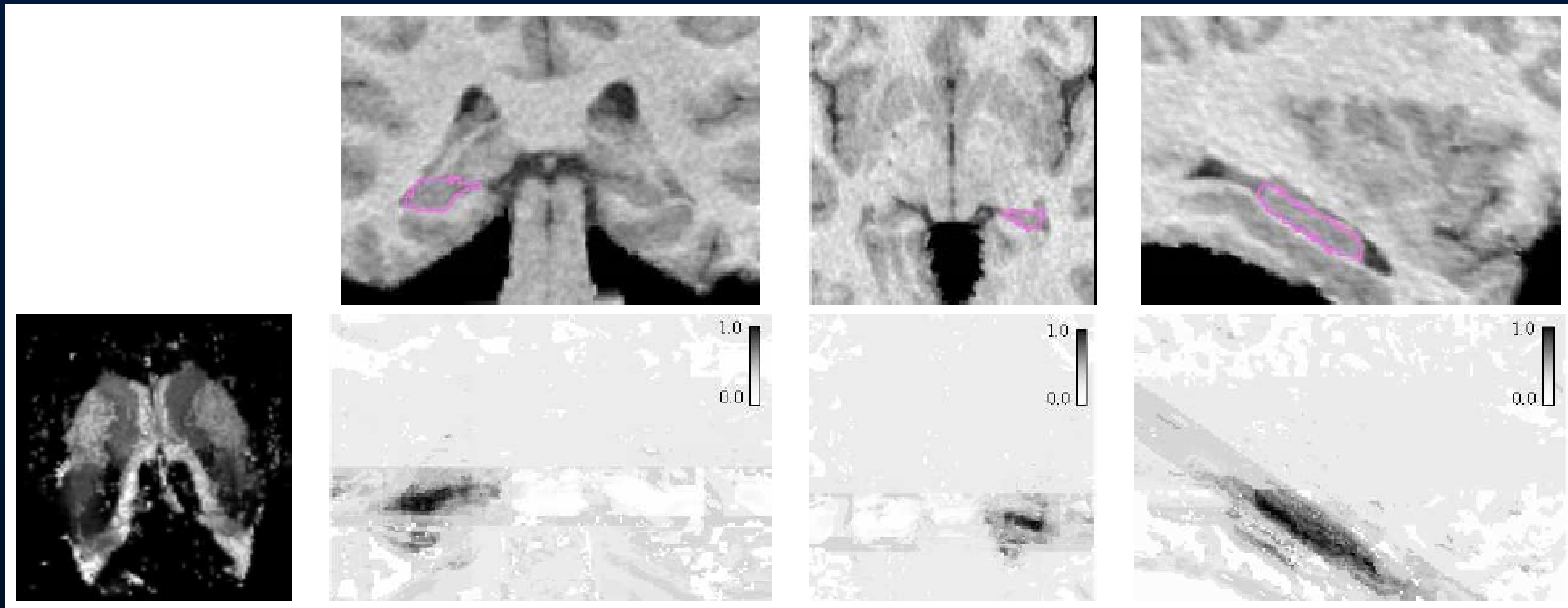


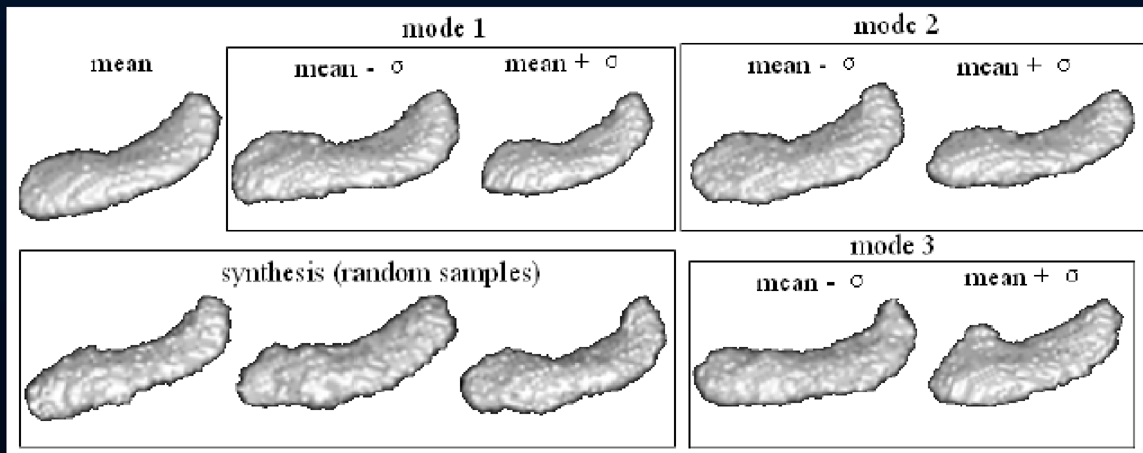
curvatures

(1) Very fast to compute using integral volume.

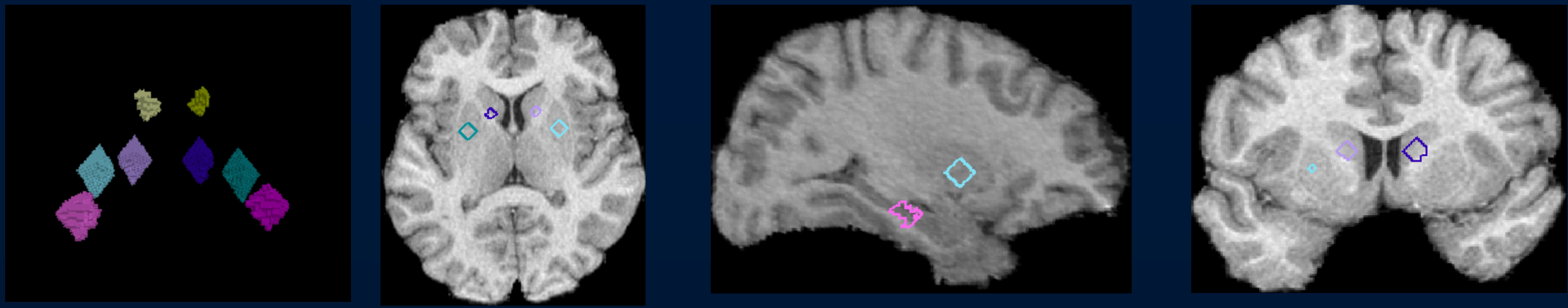
(2) Combine information at different scales.

Discriminative and Generative Models Learned

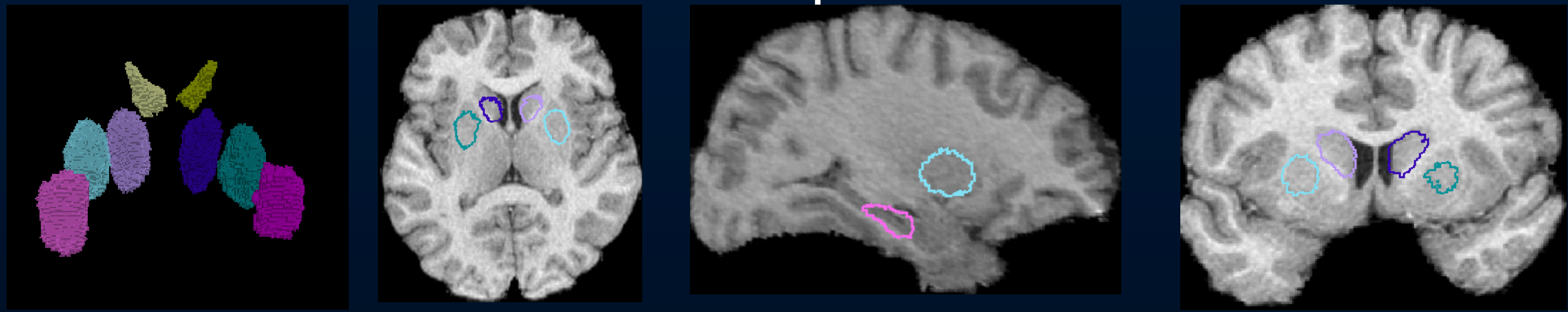




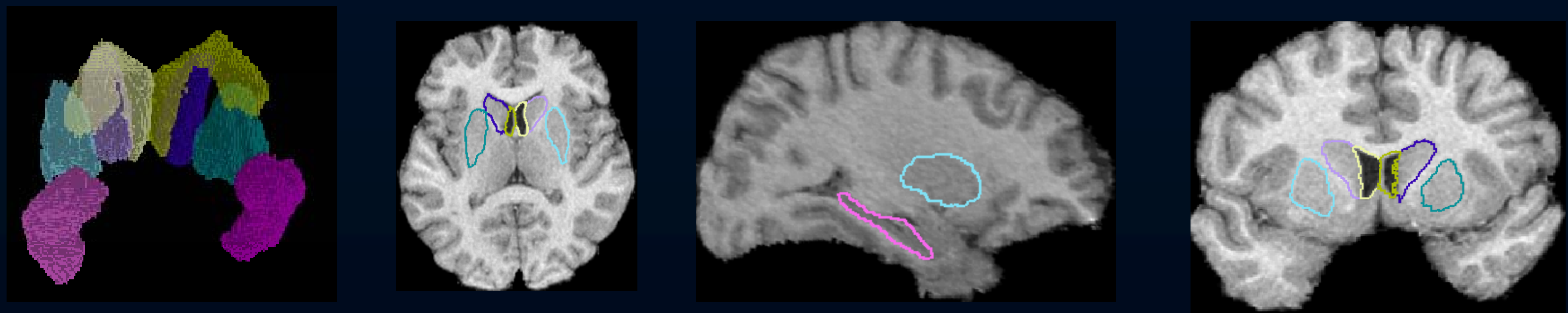
Results



Step=1

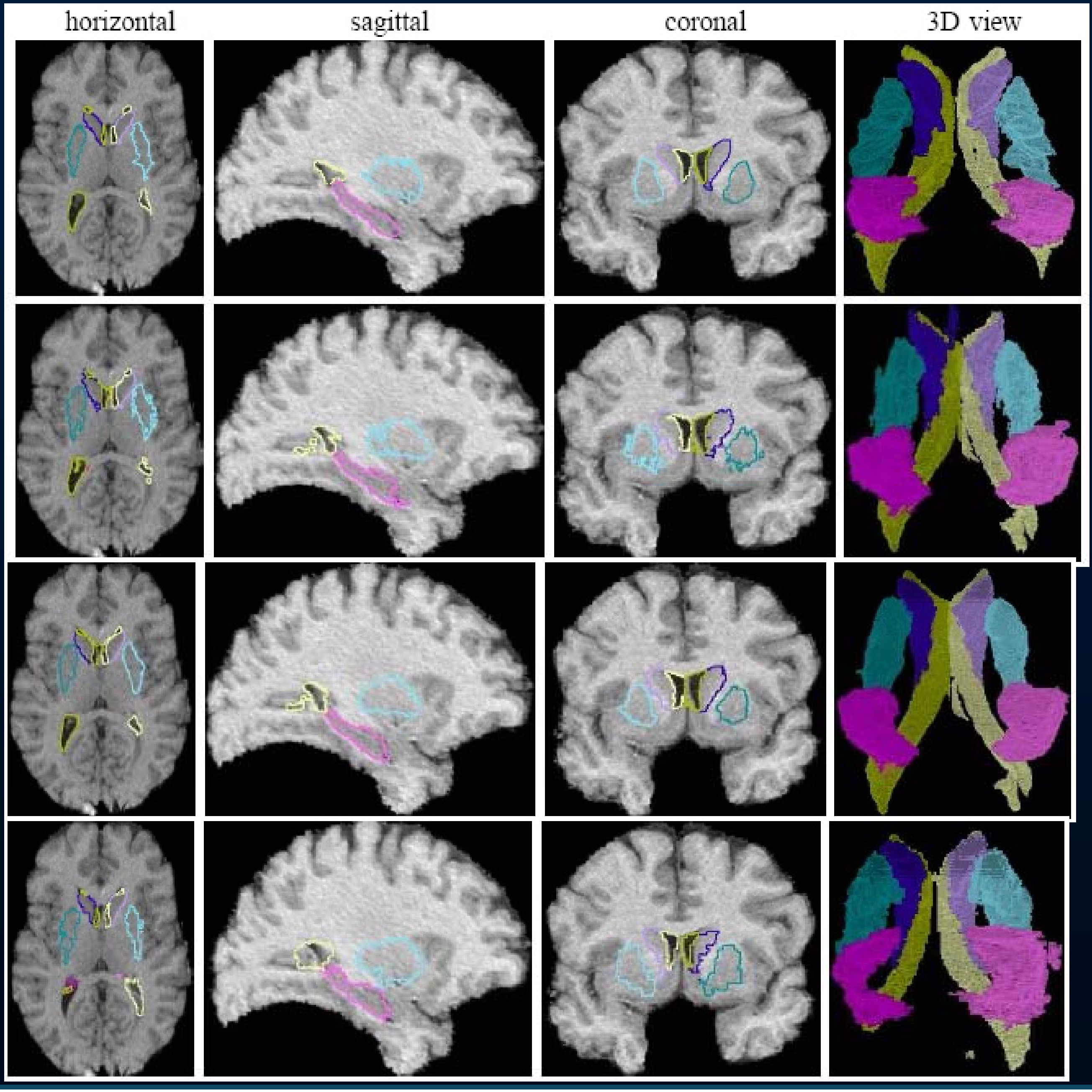


Step=2



Step=3

Results on The Testing Data



manual

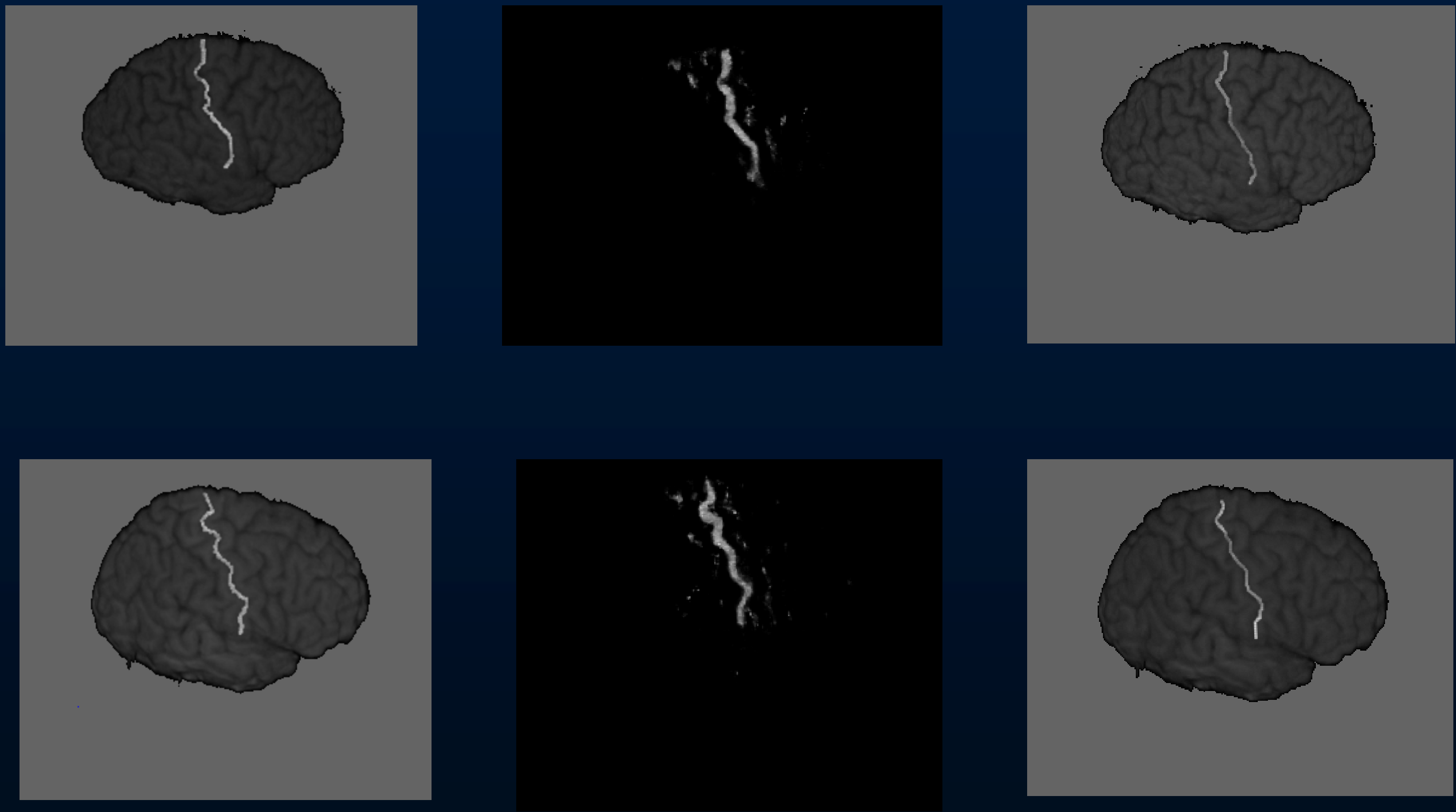
discriminative
model only

hybrid model

FreeSurfer
Fishl. et al.

Sulci-Detection

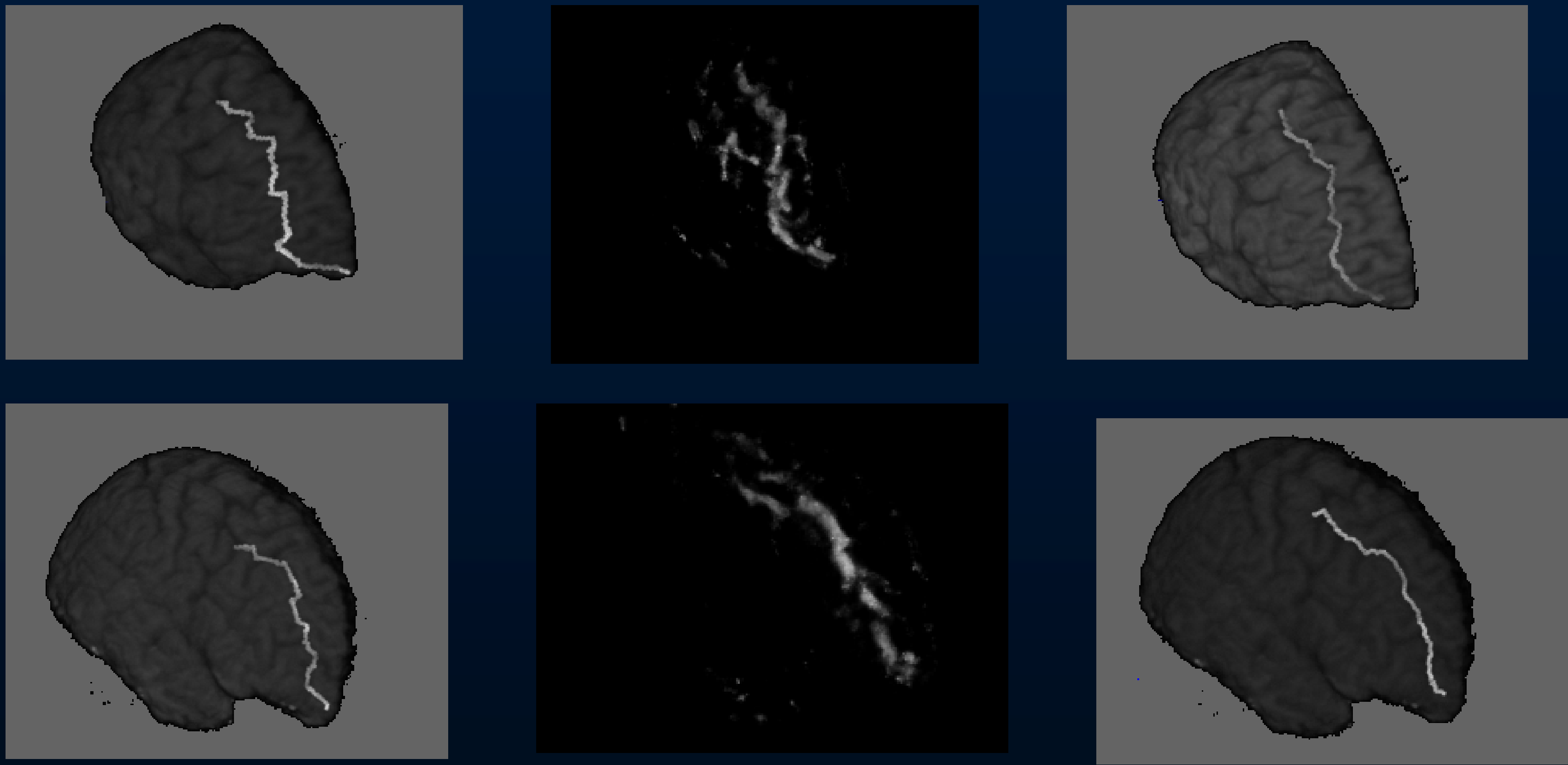
True | Prob | Result



Results on Training set: Central sulcus

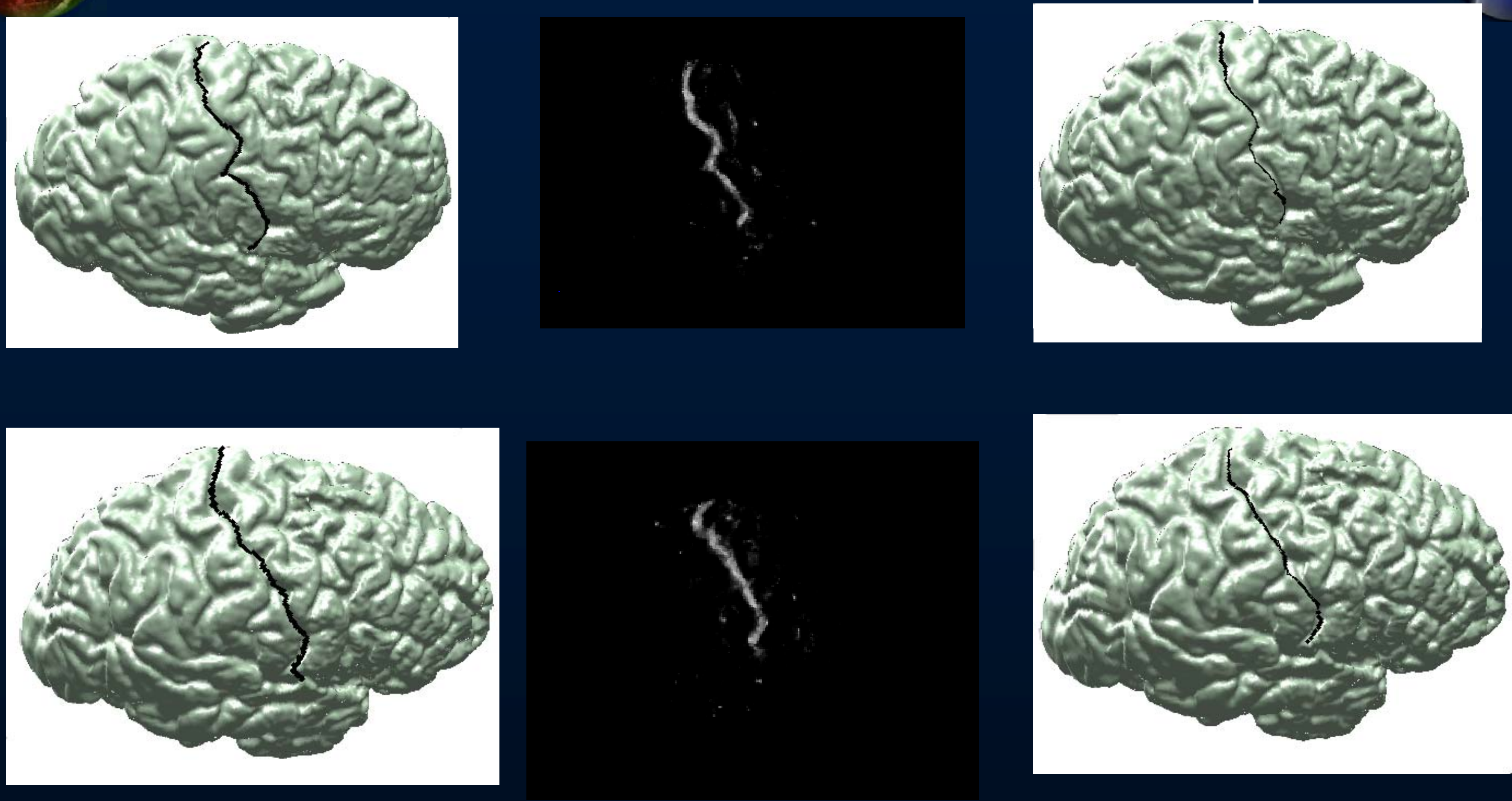
Results

True Prob Result



Results on Testing set: Superior Frontal sulcus

True | Prob | Result



Results on Training set: central sulci on surface

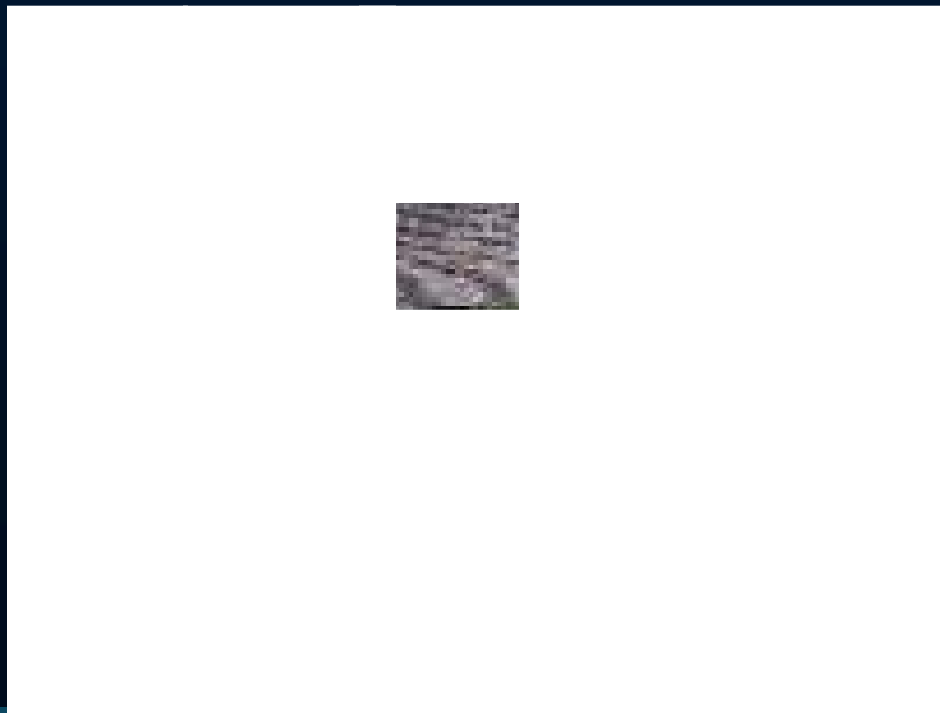
Disadvantages:

- The models only capture the appearance variation in terms of local image “patch” and the joint statistics of different structures are not captured.
- The global shape model is not play the significant role.

Seamless combine context and shape with complex appearance

Context

For object recognition, contexts come in from both **within-object** (parts) and **between-objects** (configurations).



Challenges

$$p(Y|X) \propto p(X|Y)p(Y)$$

Modeling:

It is often very hard to learn p(X|Y) and p(Y) for complex patterns.

Computing:

Computing for the optimal solution that maximizes the posterior is not an easy task. A desired algorithm should bb both efficient and effective.

We are looking for the joint statistics of p(Y|X), “context”.

Problems with MRFs, BP, and CRFs

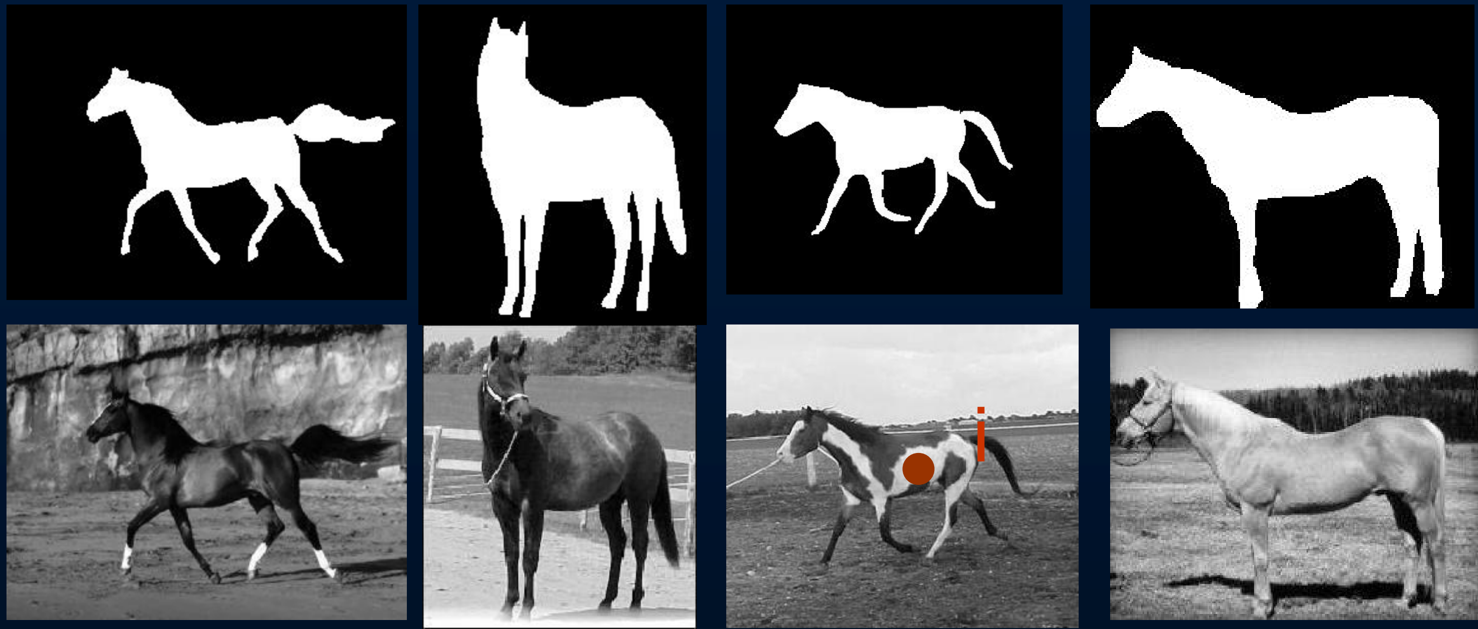
- Use fixed topology on limited number of neighborhood connections (context).
- Usually slow and it takes many steps for the message to propagate.
- Not guaranteed to find the global optimal solution.
- Modeling and computing processes are separate (maybe an advantage in some situations).

Auto-Context

Target $p(Y|X)$ directly

Y_j

X_j

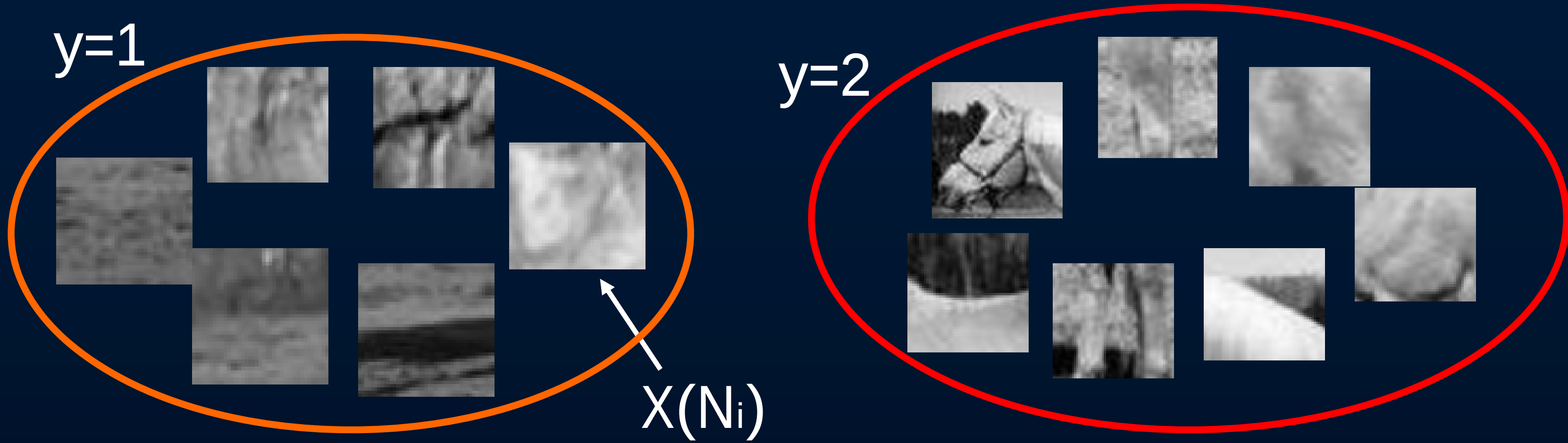


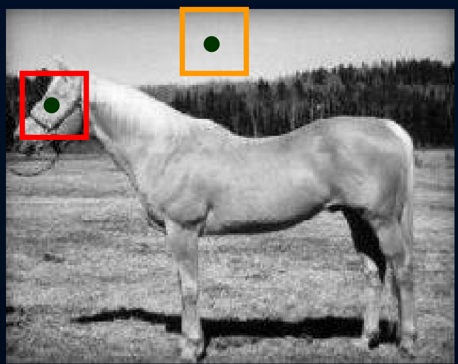
To learn $p(y_i|X)$

$$p(y_i|X) = \int p(y_i, y_{-i}|X) dy_{-i}$$

A Classification Approach

Training Set: $S = \{(y_i, X(N_i)), i = 1..n\}$

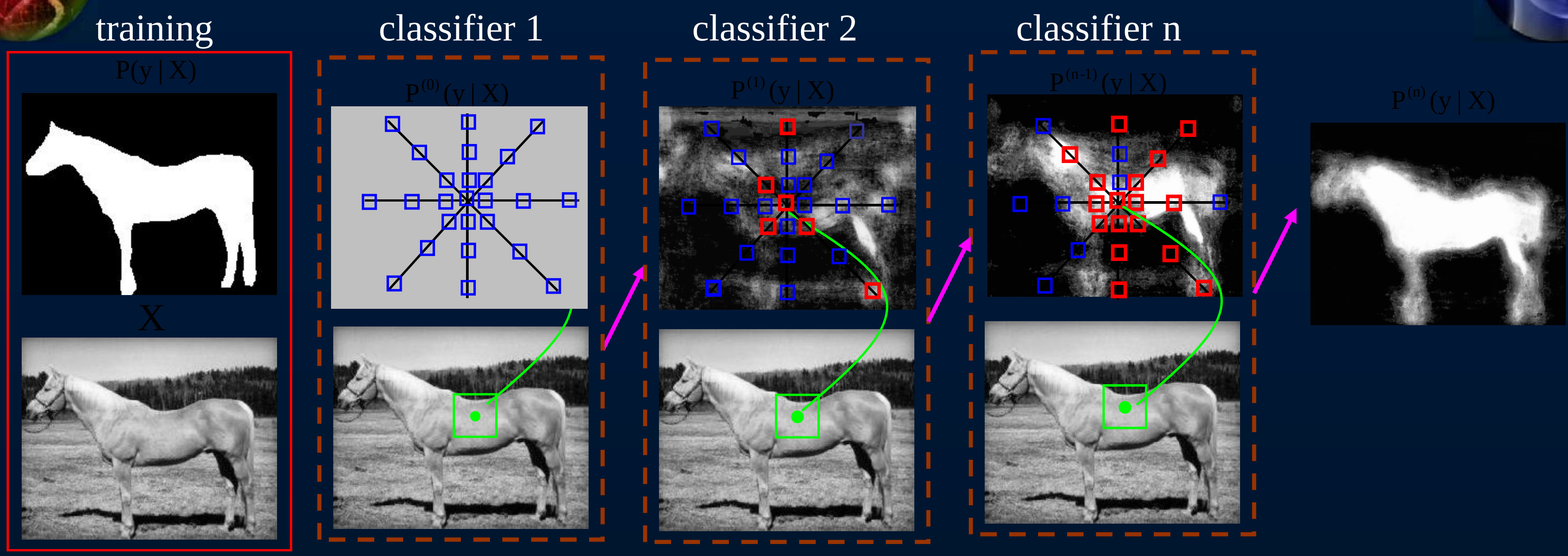




$$p(y = k|X(N)) = \frac{e^{F_k(X(N))}}{\sum_{k=1}^{K} e^{F_k(X(N))}}$$

$$\sum_{k=1}^{K} F_k(X(N)) = 0$$

Auto-Context



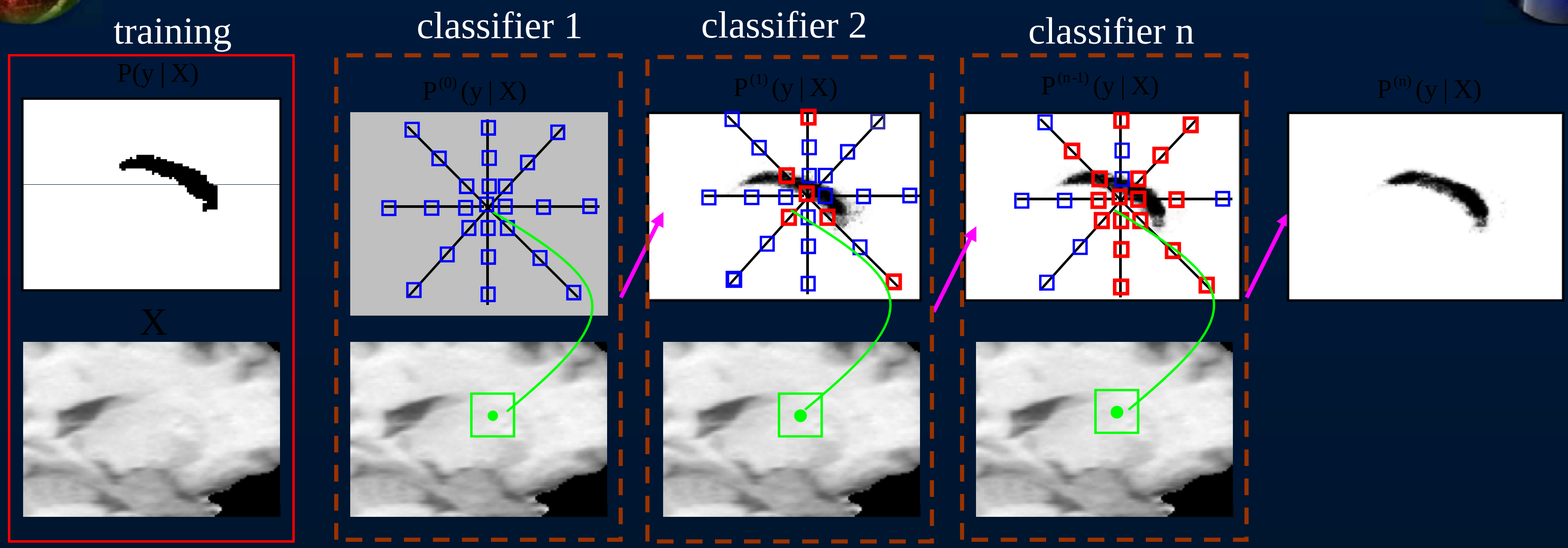
$$p^{(n)}(y_i | X(N_i), \mathbf{P}^{(n-1)}(i)) \rightarrow p(y_i, X) = \int p(y_i, y_{-i}, X) dy_{-i}$$

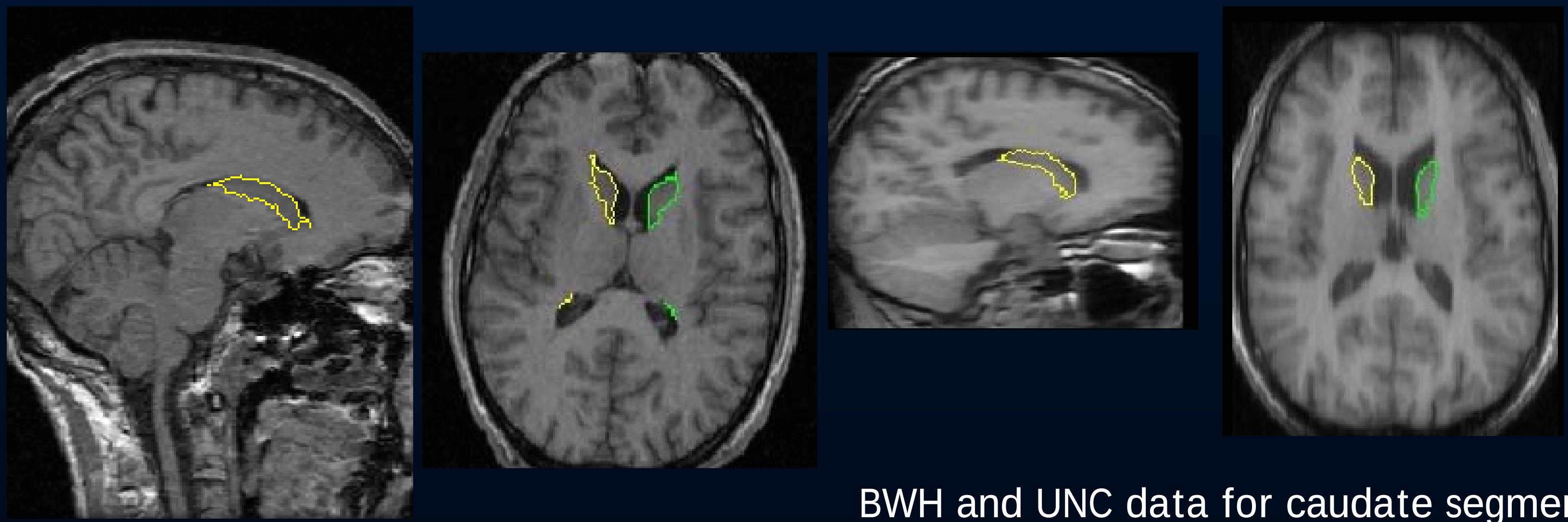
Features:

(1) appearances on X(N), 20,000 Gradients, Gabor, Haar at different scales

(2) context (shape) on P, 10,000 on a fairly large neighborhood

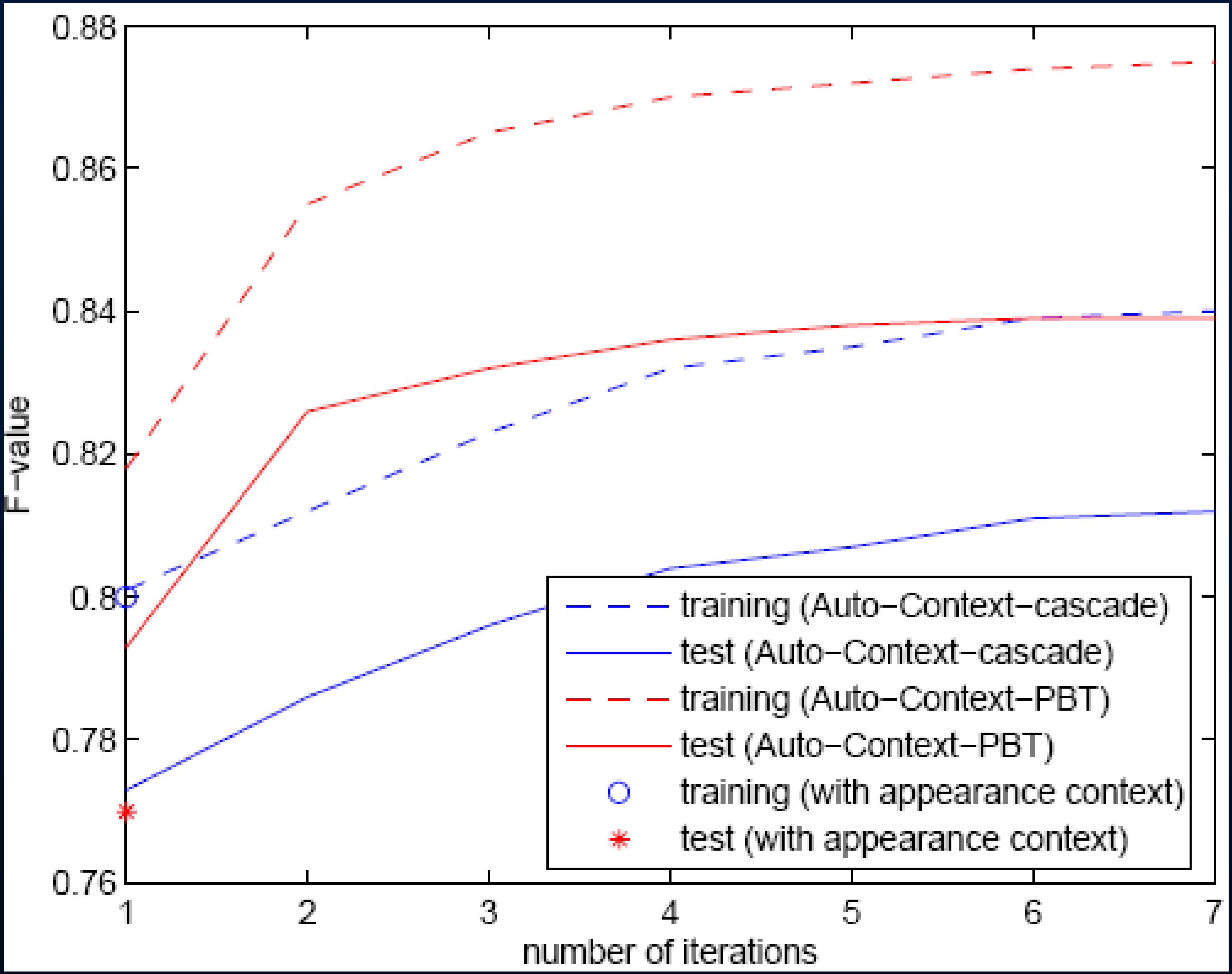
Segmenting Caudate





BWH and UNC data for caudate segmentation

Comparisons with Segmentation Methods



MICCAI

Workshop on 3D Segmentation in the Clinic:

- A Grand Challenge -

MICCAI 2007 BRISBANE AUSTRALIA

For more Information: www.MICCAI.org	Welcome!	Date: 29. October '07 Location: Brisbane, Australia
Home	10th International Conference on Medical Image Computing an	

- News
- Dates
- Program
- Data
- Evaluation
- Organizers
- Download
- Proceedings

Compare your algorithm to others:

Tune it with supplied training data

Evaluate it on specific test images

Get objective results

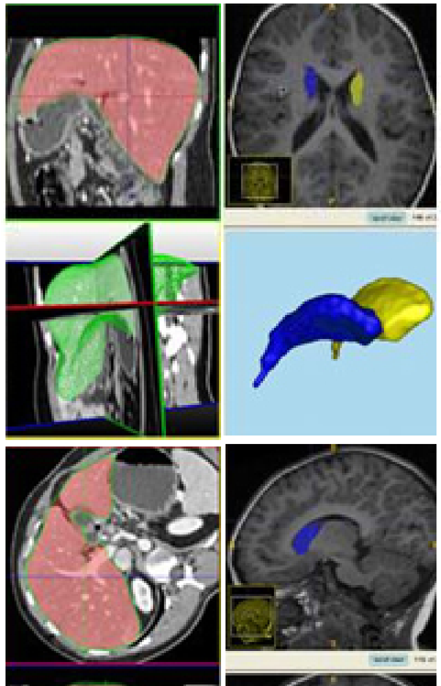
Win prizes, fame and glory

[Read workshop proceedings]()

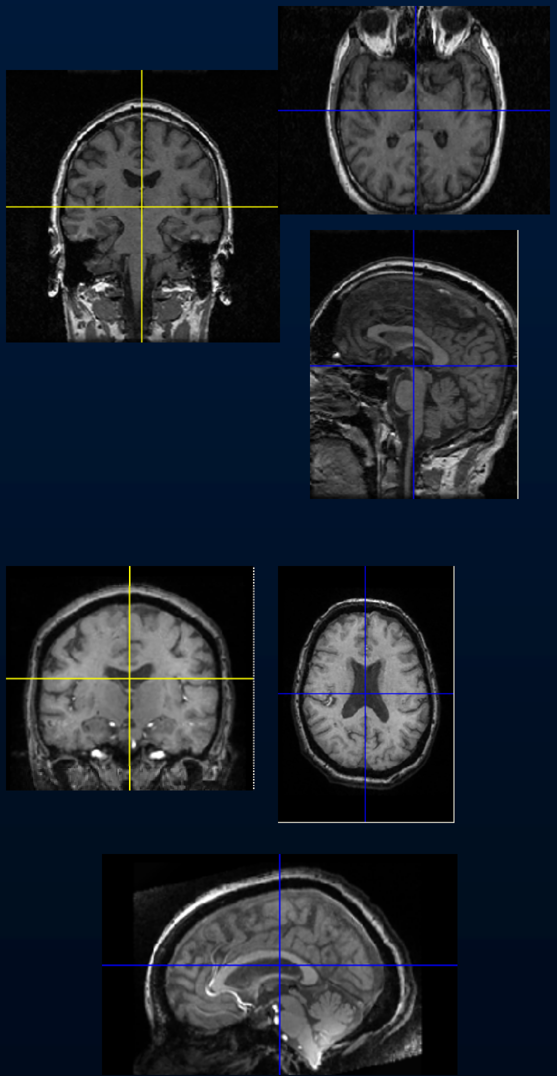
The workshop is over but the Challenge continues:

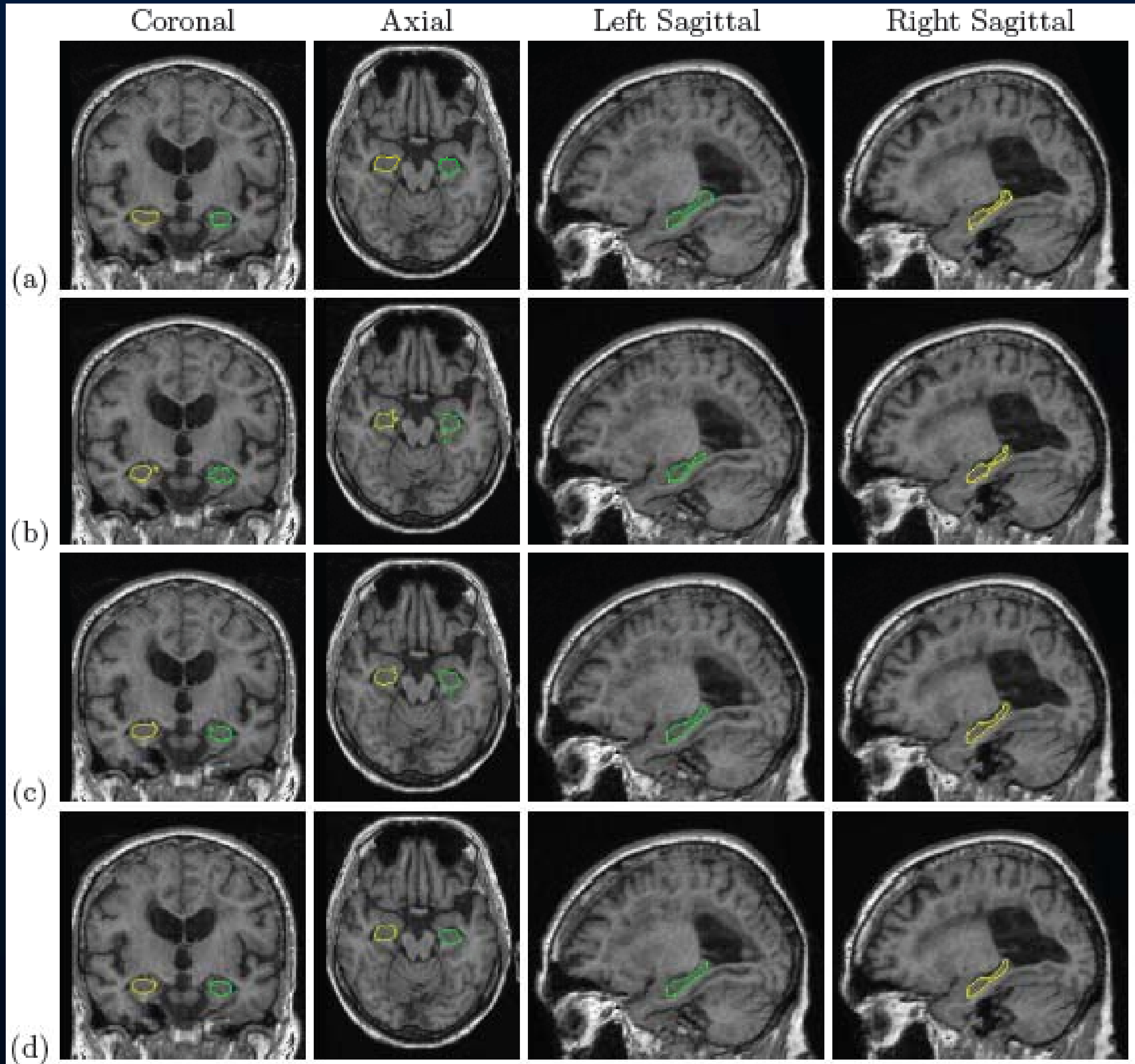
[Online Caudate Segmentation]()

[Online Liver Segmentation]()



Grand Challenge Competition





Grand Challenge Competition

Case	OE	Score	VD	Score	AD	Score	RMSD	Score	MD	Score	Total
UNC Ped	40.35	74.62	-23.21	59.46	0.86	68.25	1.21	78.38	5.64	83.41	72.82
UNC Eld	38.75	75.63	-17.23	69.77	0.75	72.15	1.14	79.64	6.79	80.02	75.44
BWH PNL	41.76	73.73	-26.62	53.78	1.51	49.10	3.50	42.05	25.27	28.41	49.42
Average All	40.84	74.31	-23.93	58.30	1.22	57.89	2.53	57.45	17.33	50.62	59.71

Hybrid Model

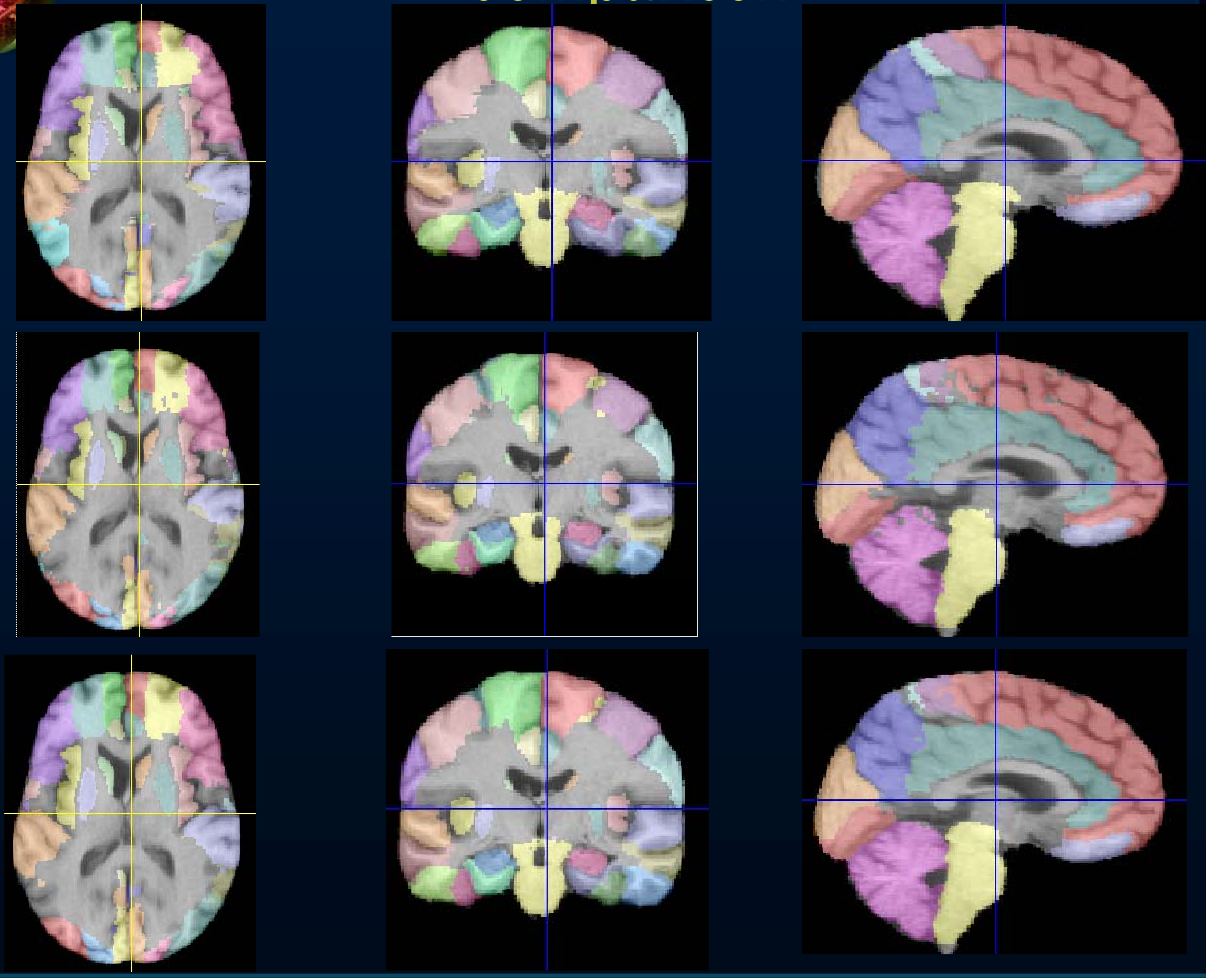
UNC Ped	33.42	78.98	-12.05	76.50	0.68	74.76	1.09	80.47	12.09	64.44	75.03
UNC Eld	36.79	76.86	-0.69	80.04	0.72	73.37	1.31	76.53	17.61	48.21	71.00
BWH PNL	32.07	78.50	-13.62	74.42	1.17	76.55	1.75	76.45	12.83	62.26	73.64
Average All	33.34	78.26	-10.60	76.03	0.97	75.51	1.52	77.31	13.67	59.78	73.38

Auto-Context

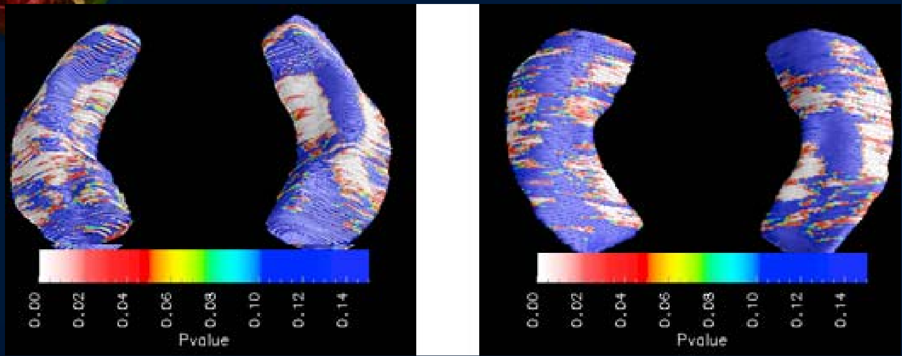
Test/Re-Test	UNC 03 [mm^3]	UNC 04 [mm^3]	UNC 09 [mm^3]	UNC 11 [mm^3]	UNC 17 [mm^3]	UNC 18 [mm^3]	UNC 21 [mm^3]	UNC 22 [mm^3]	UNC 24 [mm^3]	UNC 25 [mm^3]	Mean [mm^3]	Stdev [mm^3]	COV [%]
Left	3008	2896	2965	2862	2849	2917	2844	2836	3078	3041	2930	88	3.0
Right	2908	2980	2948	2806	2825	2971	2845	2889	3125	3059	2936	103	3.5
Total	-										-	-	3.3

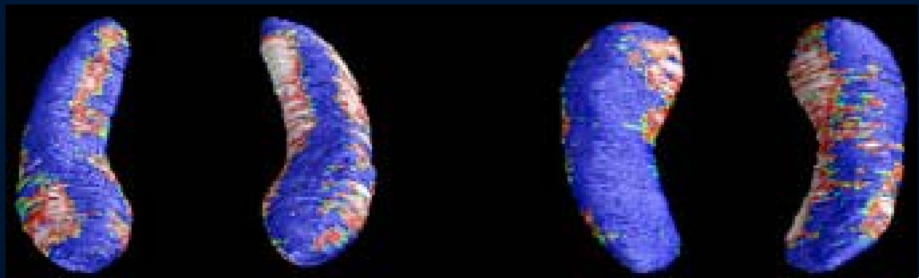
Table 2. The volumetric measurements of the 10 data sets of the same young adult acquired on 5 different scanners within 60 days. The coefficient of variation (COV = standard deviation / average, last column) indicates the stability of the algorithm in a test/re-test situation including scanner variability.

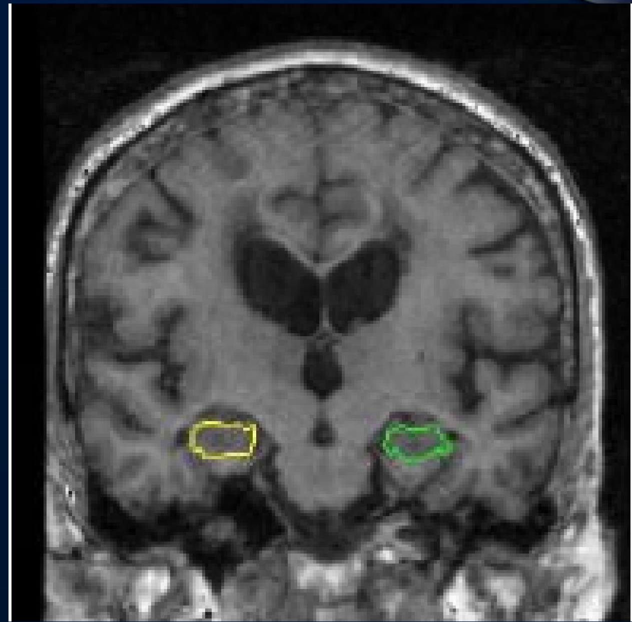
Comparison

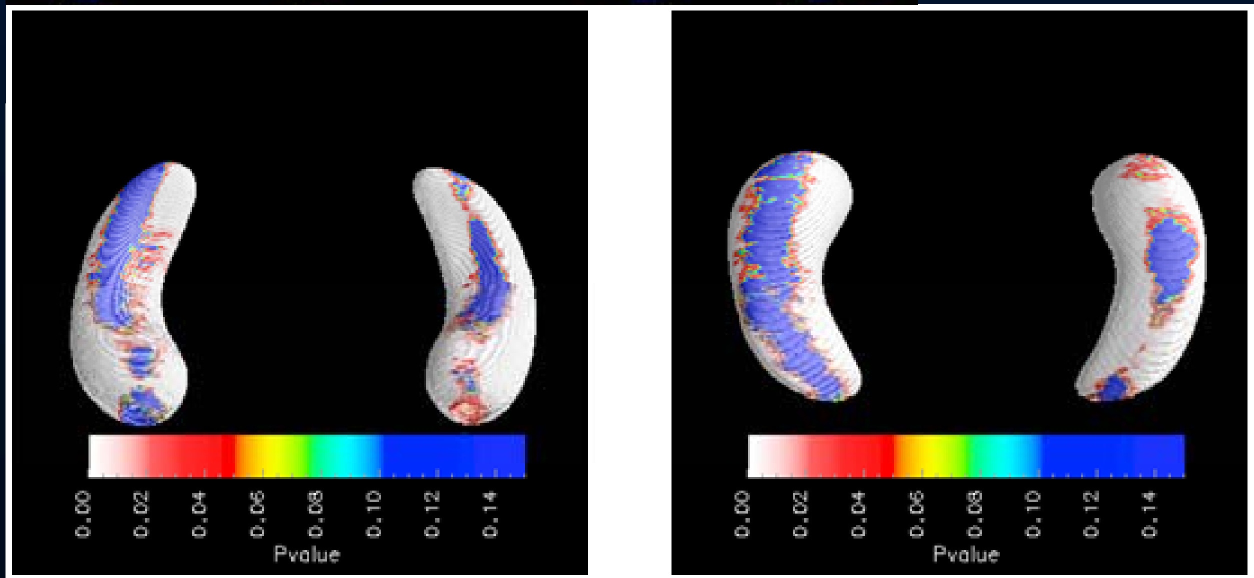


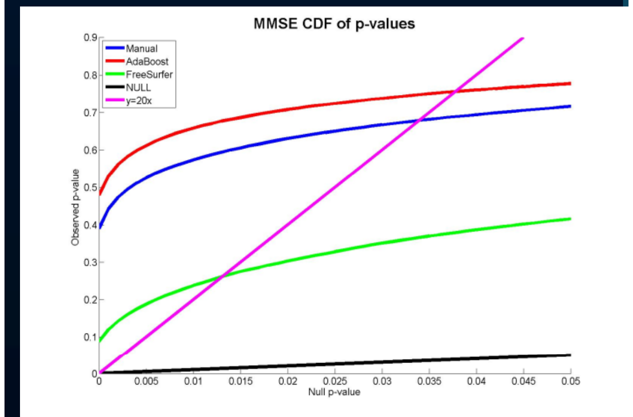
Hippocampus Image Segmentation (Morra l.)







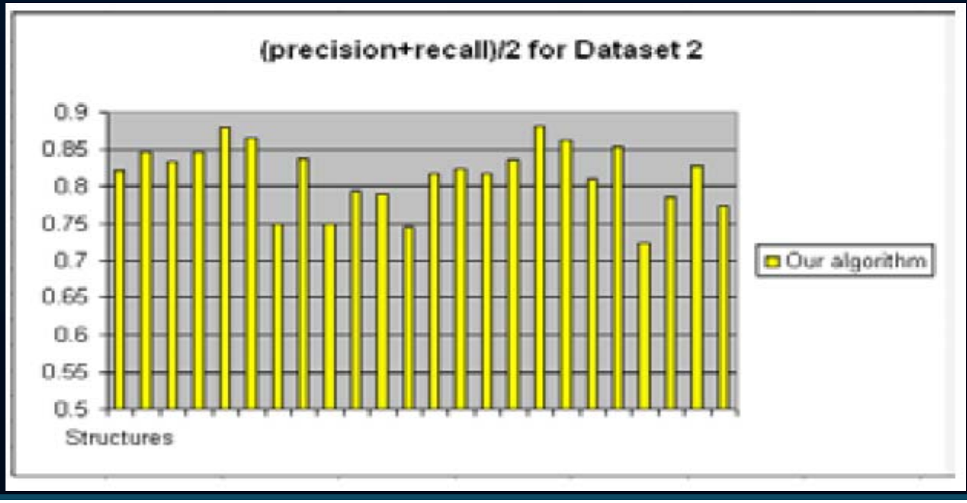




More Results

Manual Delineation

Automatic Segmentation



Convergence of Auto-Context

Theorem: The turbo context algorithm monotonically decreases the training error.

$$\epsilon_t = -\sum_i \log p^{(t)}(y_i | X(i), \mathbf{P}^{(t-1)}(i))$$

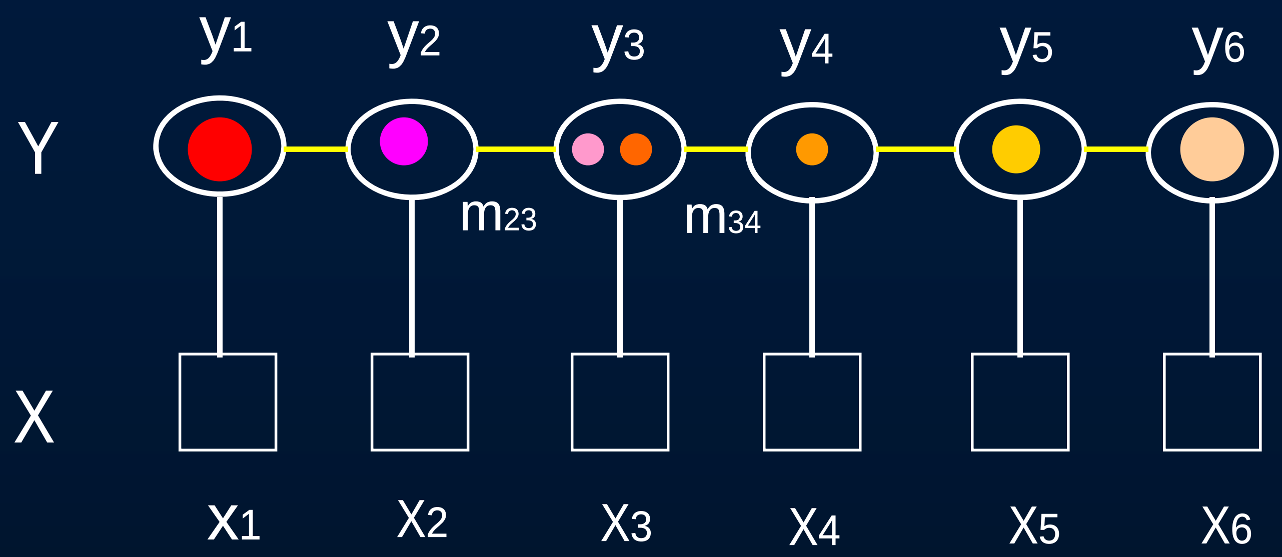
$$p^{(t)}(y_i | X(i), \mathbf{P}^{(t-1)}(i)) = \frac{e^{F_k^{(t)}(X(i), \mathbf{P}^{(t-1)}(i))}}{\sum_{k=1}^{K} e^{F_k^{(t)}(X(i), \mathbf{P}^{(t-1)}(i))}}$$

$$\epsilon_{t-1} = -\sum_i \log \mathbf{P}^{(t-1)}(i)(y_i)$$

And:

$$p^{(t)}(y_i | X(i), \mathbf{P}^{(t-1)}(i)) = \mathbf{P}^{(t-1)}(i)(y_i)$$

Belief Propagation on (MRFs, CRFs)

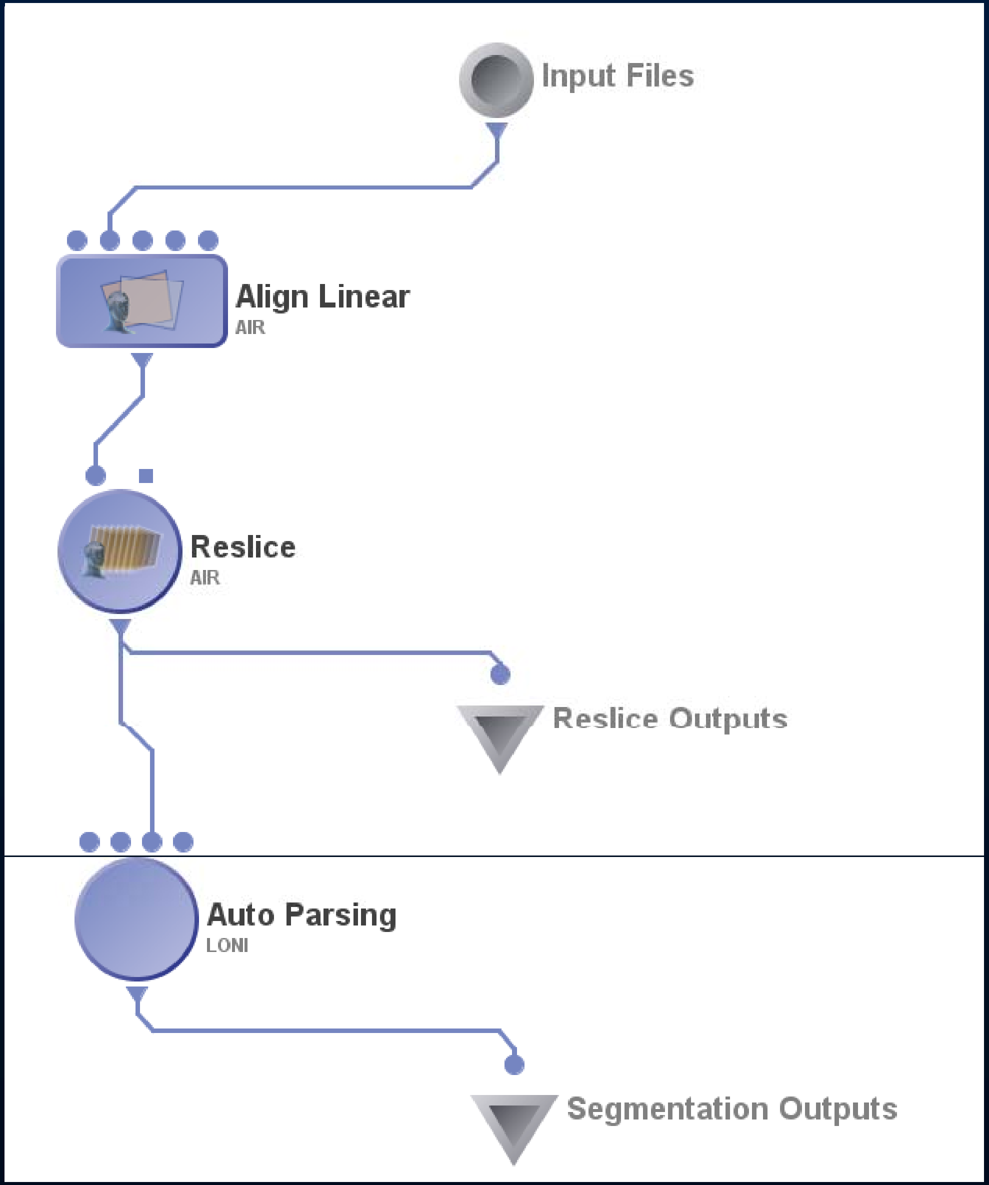


$$p(Y|X) = \frac{1}{Z} \prod_i \phi_i(x_i|y_i) \prod_{(i,j)} \psi(y_i, y_j)$$

$$p_i(y_i) = \frac{1}{Z}\phi_i(y_i) \prod_{j \in N(i)} m_{ji}(y_i)$$

$$m_{ij}(y_j) \leftarrow \sum_{y_i} \phi_i(y_i)\psi_{i,j}(y_i, y_j) \prod_{k \in N(i)\backslash j} m_{ki}(y_i)$$

Brain Parser Pipeline



Conclusions for Auto-Context

Advantages:

- Learns low-level and context model in an integrated framework.
- Very easy to implement.
- Significantly faster than MCMC and BP (30~50 seconds) on MRFs or CRFs.
- General and avoid heavy algorithm design.
- Learning and computing use the same procedures.
- Can be applied in other domains.

Conclusions for Auto-Context

Disadvantages:

- Require training for different problems.
- Explicit high-level information is not included.
- Training time might be long. (half day to a week)
- Require all labeled data (fully supervised).

Thank you!

Questions?