Image Segmentation Algorithms

Dzung L. Pham, Ph.D.

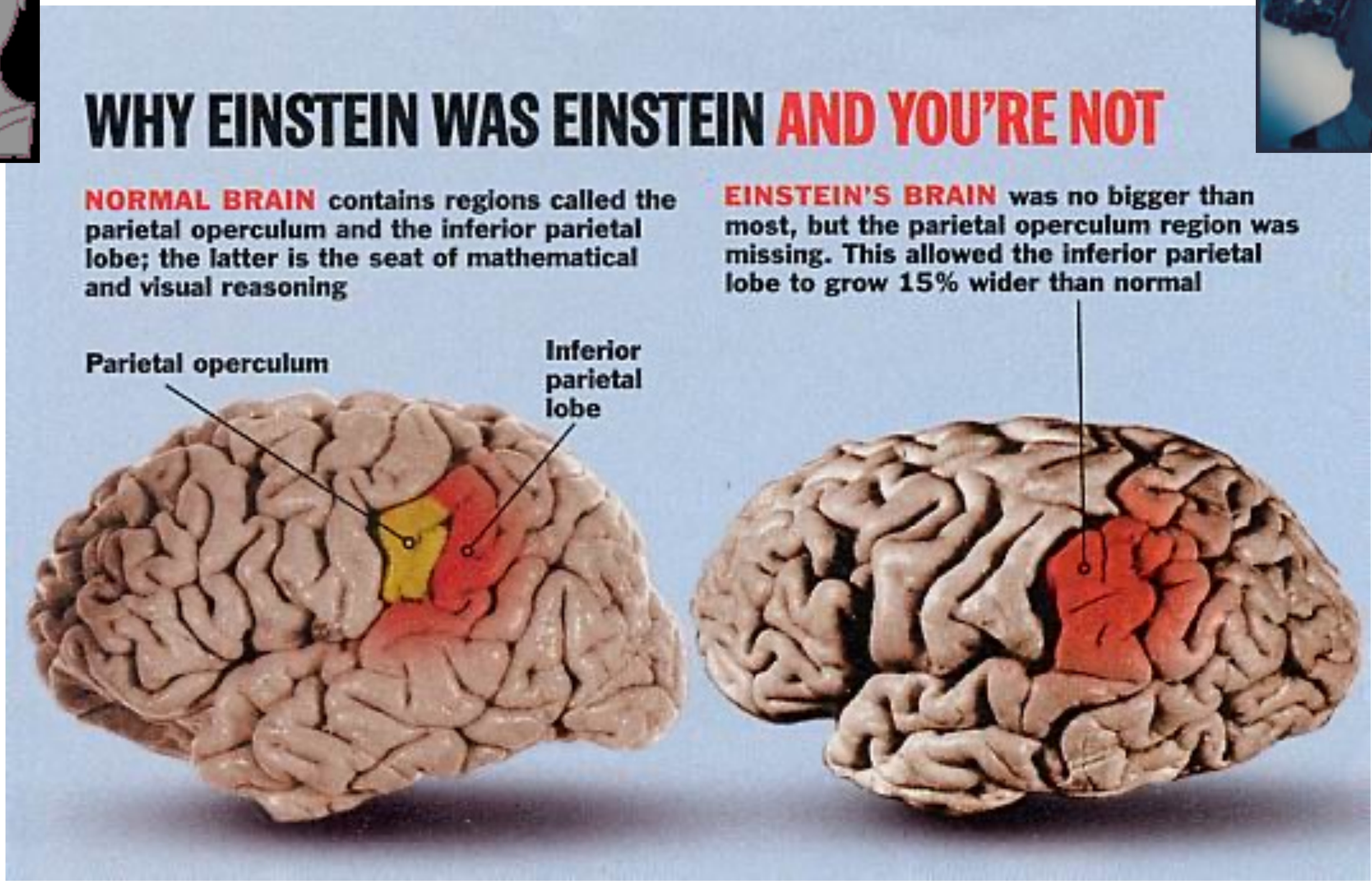
Director, CNRM Image Processing Core
Senior Scientist, Henry Jackson Foundation
Associate Professor, Uniformed Services University

CNRM

http://www.usuhs.mil/cnrm

July 17, 2017

Phrenology Strikes Back



Time Magazine 1999

- The brain is functionally compartmentalized
- Differences or changes in the size and shape of brain structures are associated with aging and disease processes

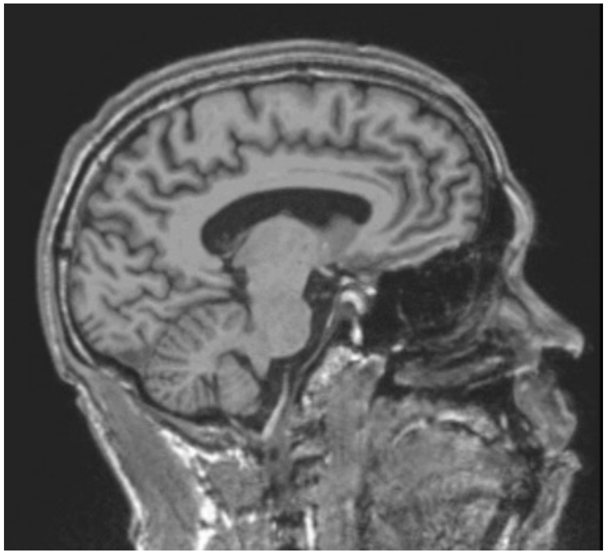
http://www.museumofquackery.com/

Overview

- **Segmentation algorithms** are algorithms that delineate anatomical structures and/or other regions of interest

- **Objective**: provide an overview of MR brain segmentation algorithms to better understand how they may be used in your research

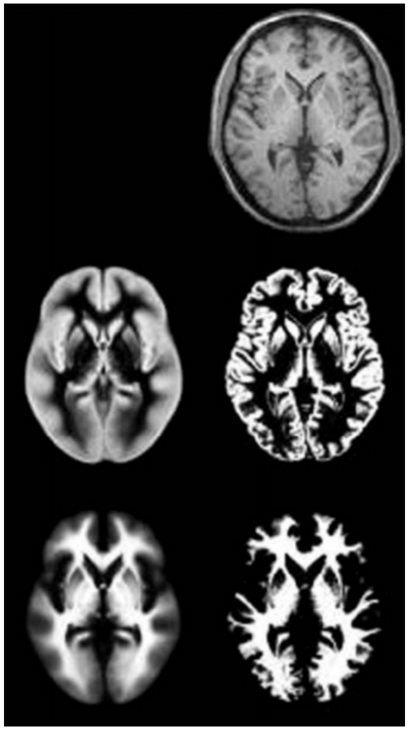
  1. Why am I here?
  2. Challenges of segmentation
  3. Approaches
     a) semi-automatic
     b) automatic
  4. Practical examples
  5. Current directions



fMRI Applications

- Important for volume quantification, morphometrics, localization and monitoring of pathology, computer-aided diagnosis, treatment planning, computer-integrated surgery, and others

- Why is segmentation useful in fMRI data analysis?

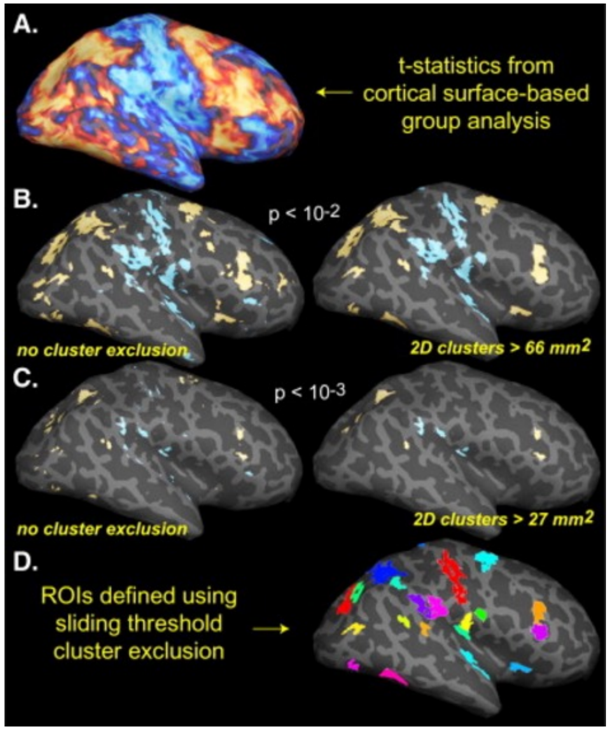


SPM

Ashburner et al 2005

AAL

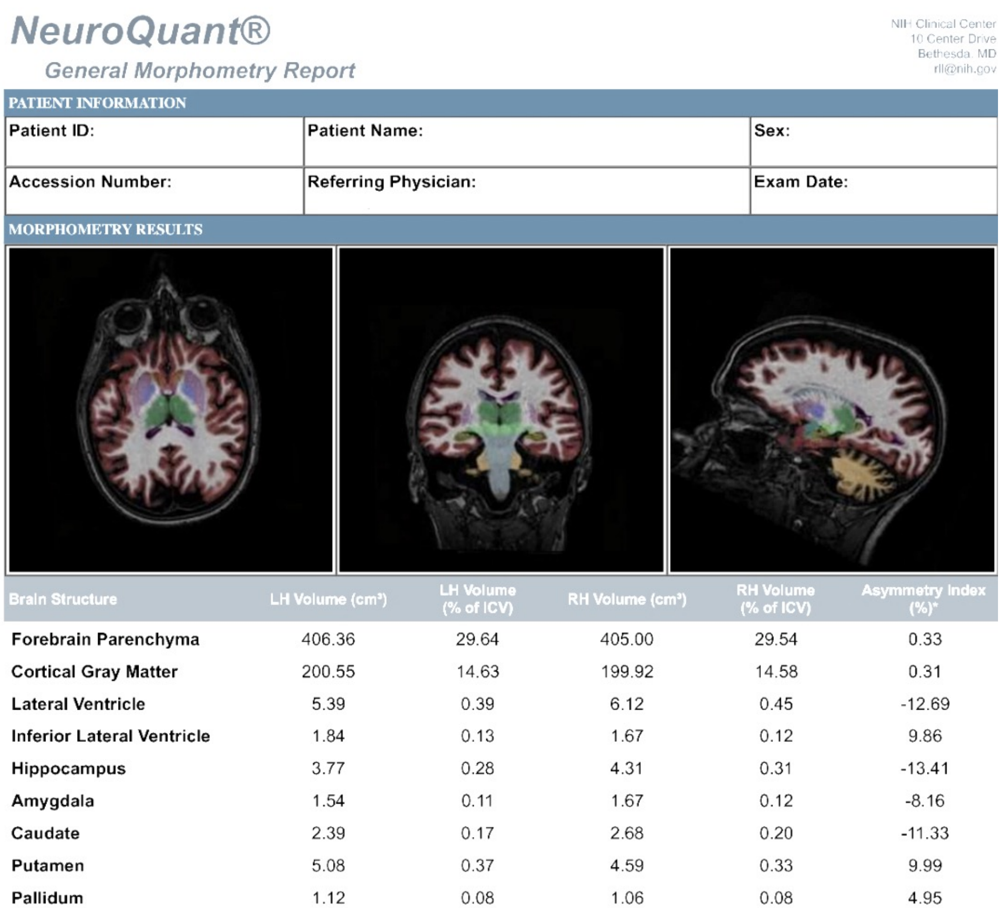
Tsourio-Mazoyer et al 2002



FreeSurfer

Hagler et al 2006

Other Applications



NeuroQuant®

General Morphometry Report

NIH Clinical Center
10 Center Drive
Bethesda, MD
rll@nih.gov

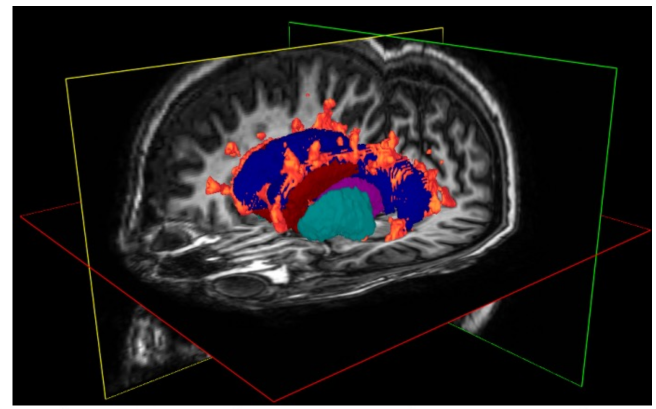
PATIENT INFORMATION

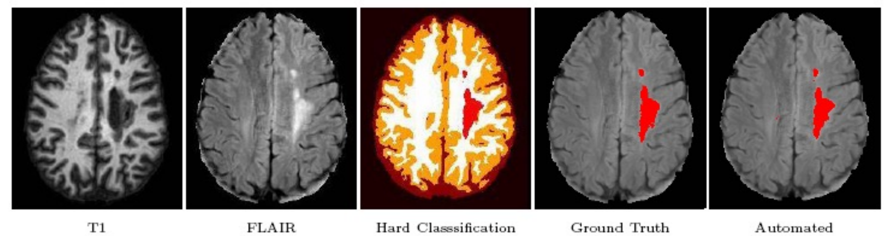
Patient ID:	Patient Name:	Sex:
Accession Number:	Referring Physician:	Exam Date:

MORPHOMETRY RESULTS

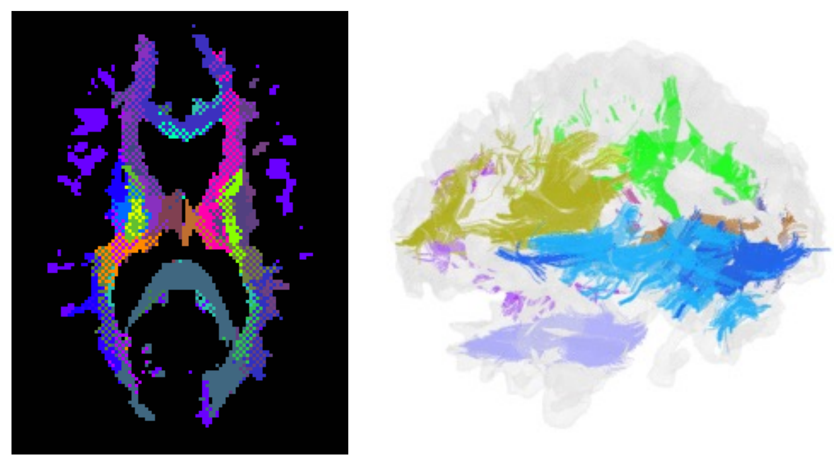
Brain Structure	LH Volume (cm³)	LH Volume (% of ICV)	RH Volume (cm³)	RH Volume (% of ICV)	Asymmetry Index (%)*
Forebrain Parenchyma	406.36	29.64	405.00	29.54	0.33
Cortical Gray Matter	200.55	14.63	199.92	14.58	0.31
Lateral Ventricle	5.39	0.39	6.12	0.45	-12.69
Inferior Lateral Ventricle	1.84	0.13	1.67	0.12	9.86
Hippocampus	3.77	0.28	4.31	0.31	-13.41
Amygdala	1.54	0.11	1.67	0.12	-8.16
Caudate	2.39	0.17	2.68	0.20	-11.33
Putamen	5.08	0.37	4.59	0.33	9.99
Pallidum	1.12	0.08	1.06	0.08	4.95

NeuroQuant output from Cortechs Labs
FDA Approved Brain Segmentation



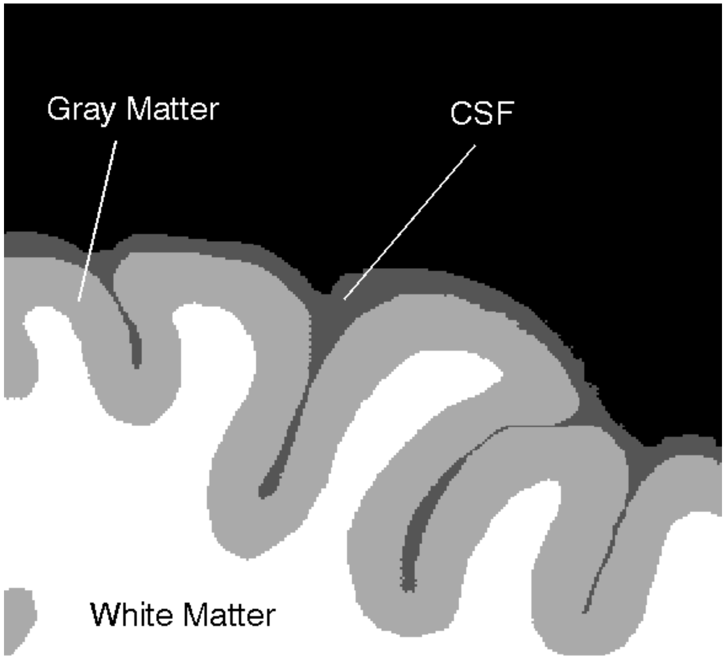


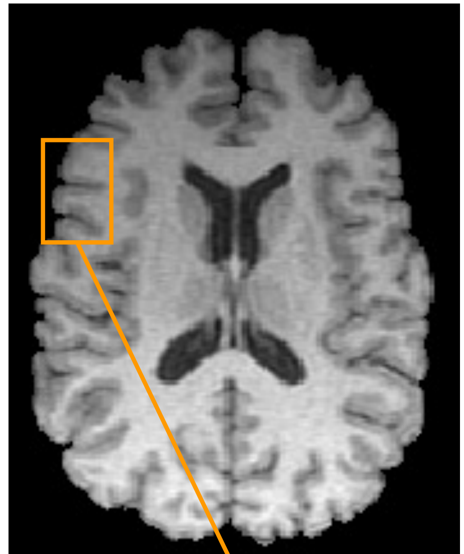
MS Lesion Segmentation

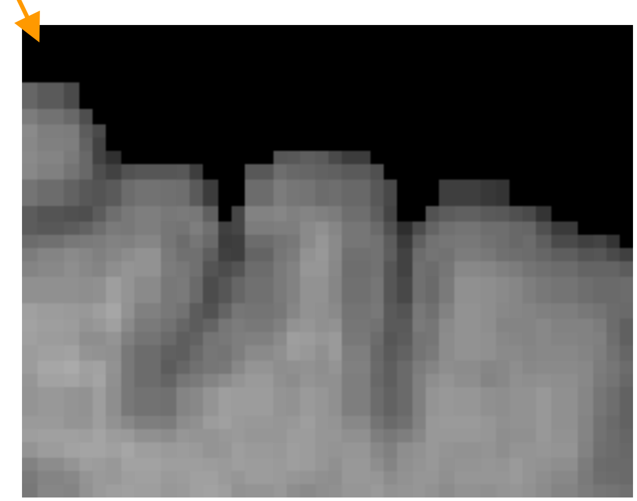


DTI White Matter Tract Segmentation

Challenges







- Manual delineation is not that fun, we want automation
- Structural MR images suffer from artifacts such as noise, RF inhomogeneities, partial volume effects, ghosting, others
- Validation and training data - what is the truth?

Segmentation Approaches

- We will review some of the more commonly used semi-automatic and automatic approaches to MR brain segmentation (not exhaustive)
- There are advantages and disadvantages to these different approaches. In practice, multiple algorithms are typically combined
- Good recent review papers: Gonzallez-Villa et al, Art Intel Med 2016
  Despotovic et al, Comput Math Methods Med 2015

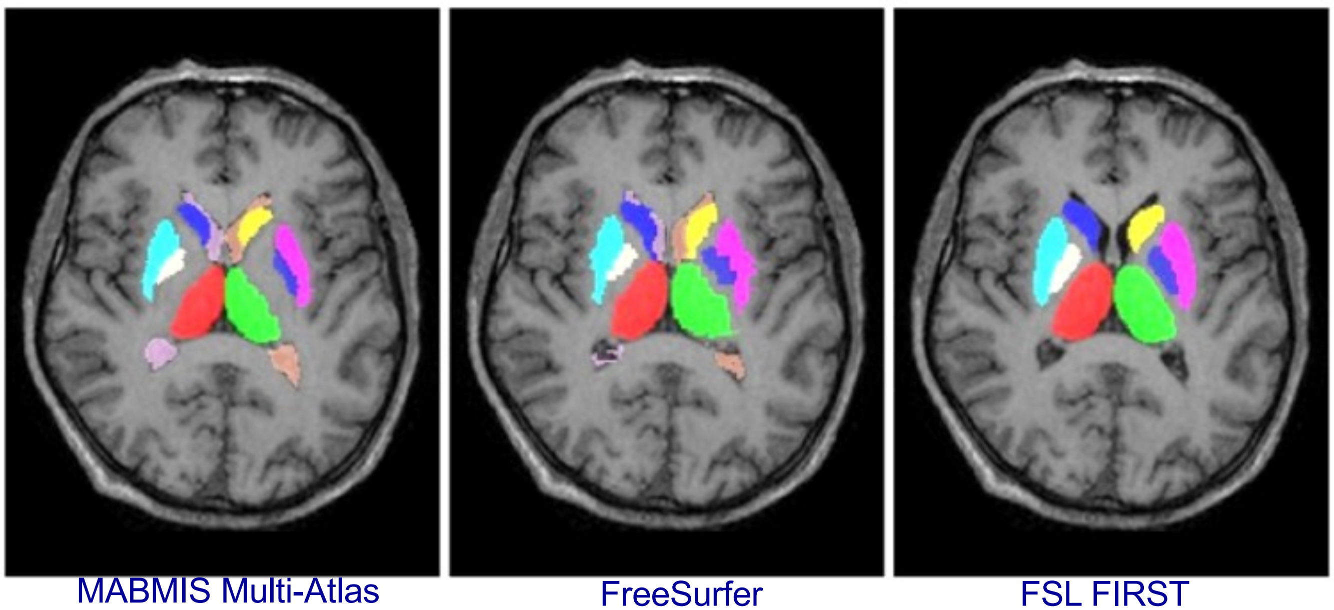
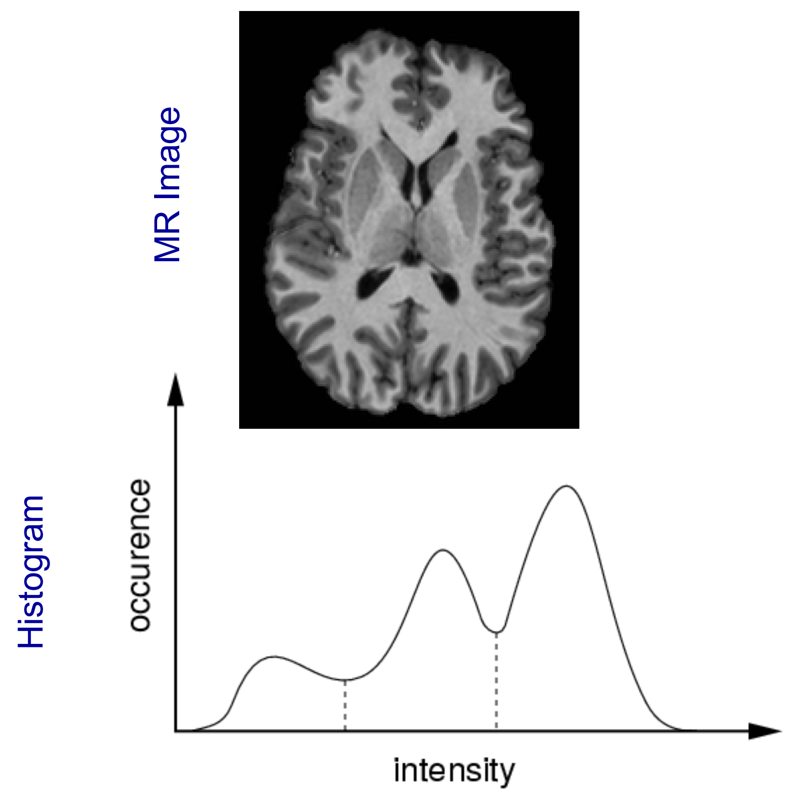
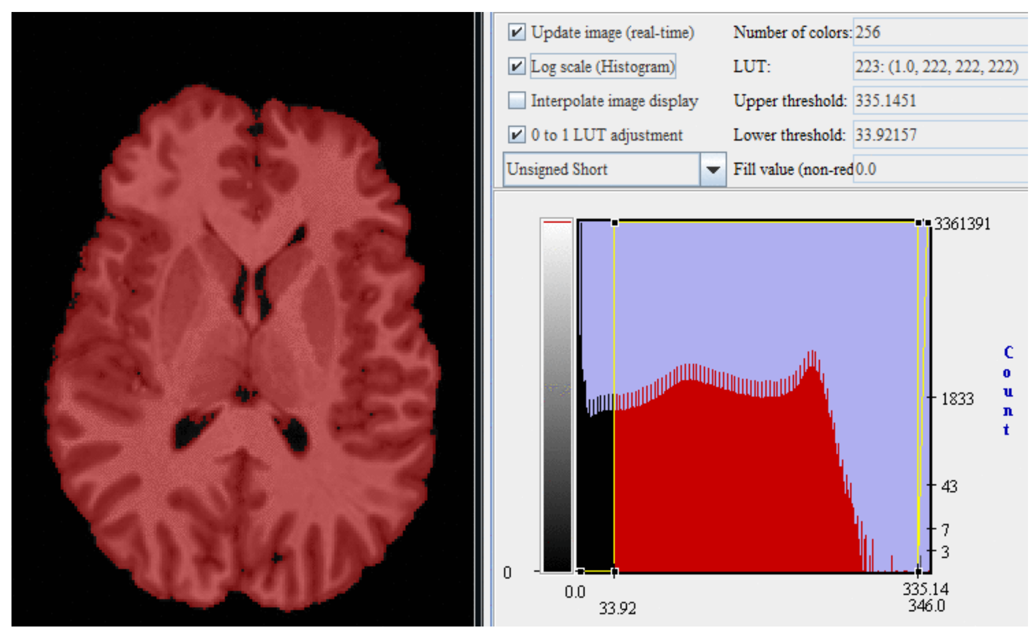


Figure from Gonzalez-Villa et al 2016

Methods: Thresholding

- **Thresholding** segments scalar images by creating a binary partitioning of image intensities
- Simple and effective but can be sensitive to imaging artifacts
- Automated versions exist (eg. Otsu's method)



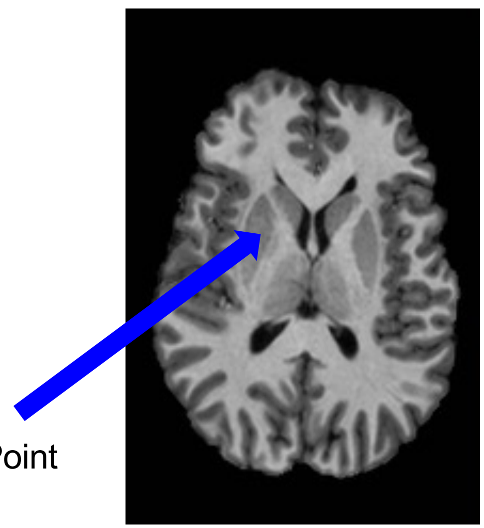


Example of thresholding with MIPAV

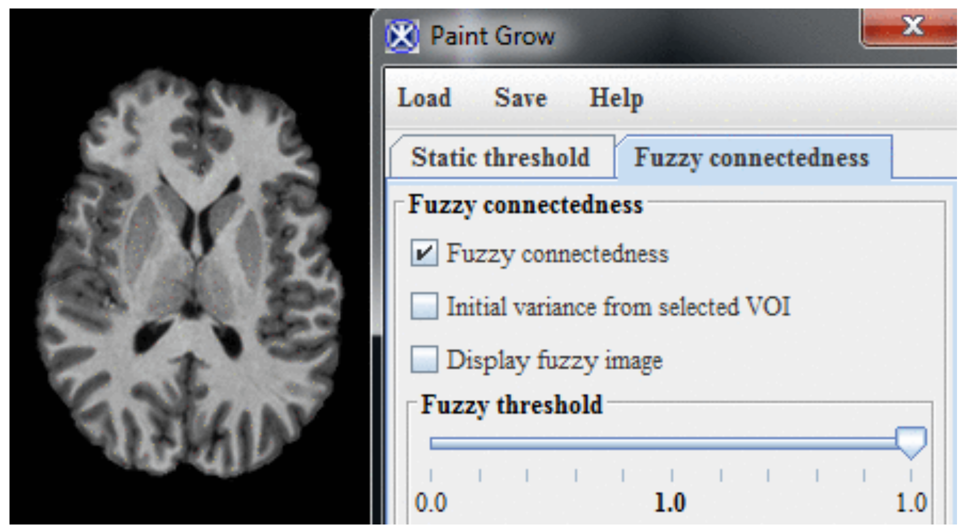
http://mipav.cit.nih.gov
AFNI's 3dcalc command

Methods: Region growing

- **Region growing** is a technique for extracting an image region that is connected based on some predefined criteria
- Often used to aid delineation of small structures like lesions
- Requires manual interaction to specify a seed point, although automated variations have been proposed
- Growing is typically governed by the local intensities or edge features



Seed Point



Example of 3-D region growing with MIPAV

http://mipav.cit.nih.gov
http://www.slicer.org

Methods: Deformable models

- **Deformable models** are physically motivated techniques for delineating region boundaries by using closed curves or surfaces that deform under the influence of internal and external forces
- Formulated as a force-balance equation, where external forces drive the model and internal forces maintain the local smoothness
- Can be semi-automatic or full automated

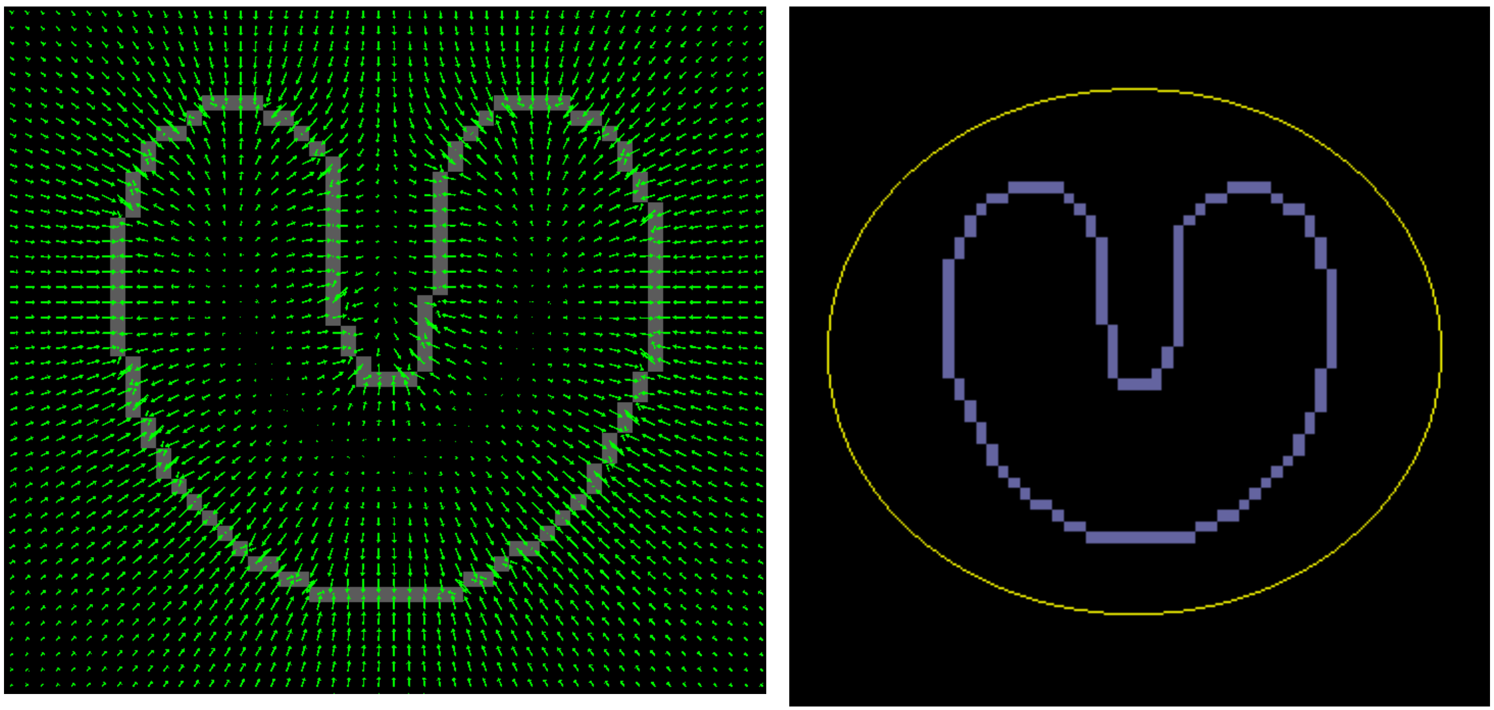
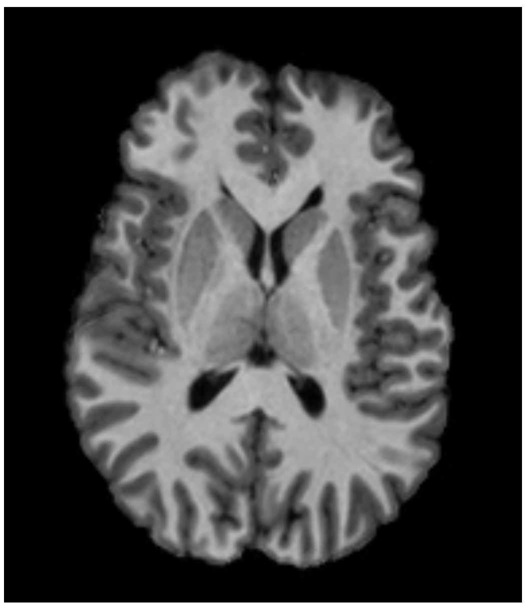


Figure from http://iacl.ece.jhu.edu

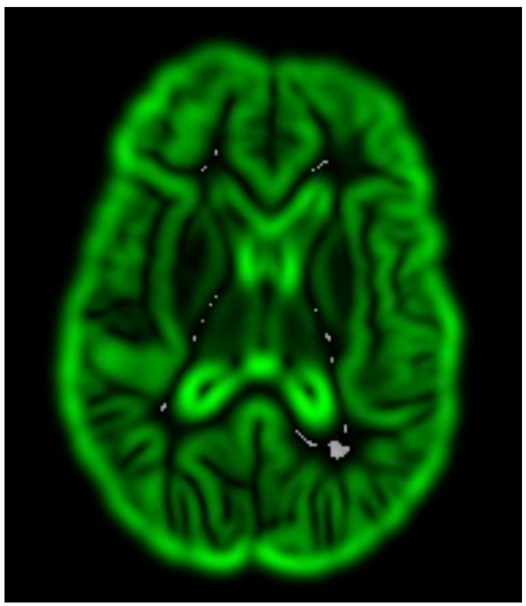
http://www.iacl.ece.jhu.edu/static/gvf/
http://www.itksnap.org

Methods: Deformable models

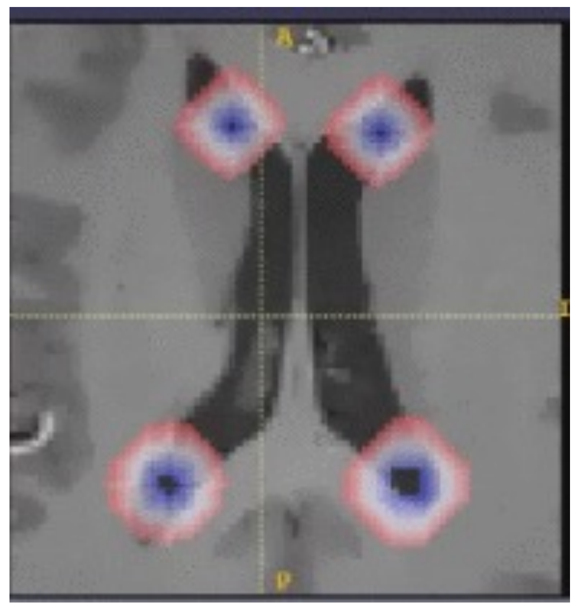
- Deformable models typically are driven by edges, although region based forces are also possible
- Parametric DMs typically have fixed topology, but can have continuous resolution
- Geometric or level-set DMs can adapt to topology, but are defined by on a voxel grid



MR Image



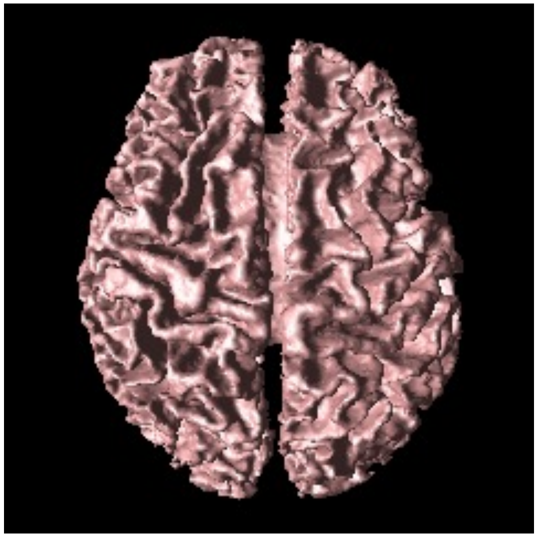
Edge Map computed by Spatial Derivatives



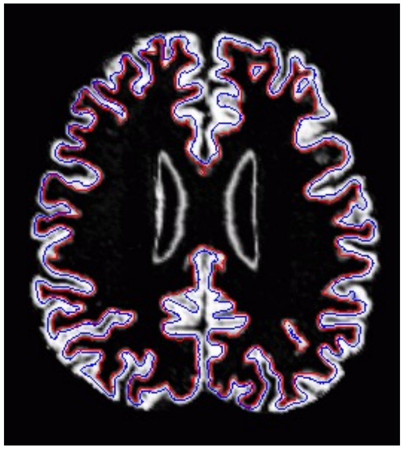
Video of level set evolution from http://www.itksnap.org

http://www.iacl.ece.jhu.edu/static/gvf/
http://www.itksnap.org

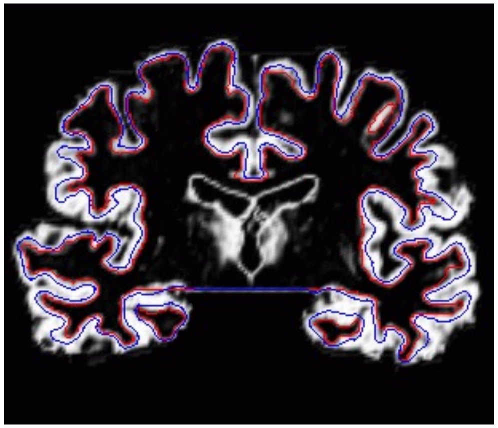
Methods: Deformable Models for Reconstructing the Cortex



Deforming cortex



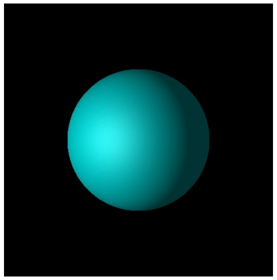
Axial view of inner and central cortical surface on gray matter

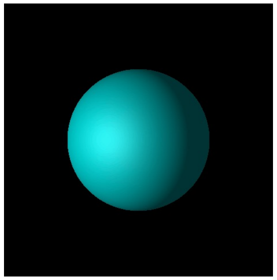


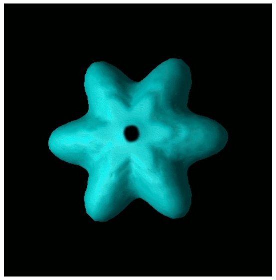
Coronal view of inner and central cortical surface on gray matter

- Models surface as an elastic sheet with forces designed to push the surface towards the gray matter
- Topology *preservation* is an important property for reconstructing the cortex

Demonstration of topology changes







http://www.nitrc.org/projects/toads-cruise
https://surfer.nmr.mgh.harvard.edu/

Methods: Classifiers

- **Classifiers** partition a feature space derived from the image by using data with known labels
- The feature space typically consists of the intensity values but can also include derived features or neighboring intensities
- Classifiers may be unsupervised/supervised and generative/discriminative

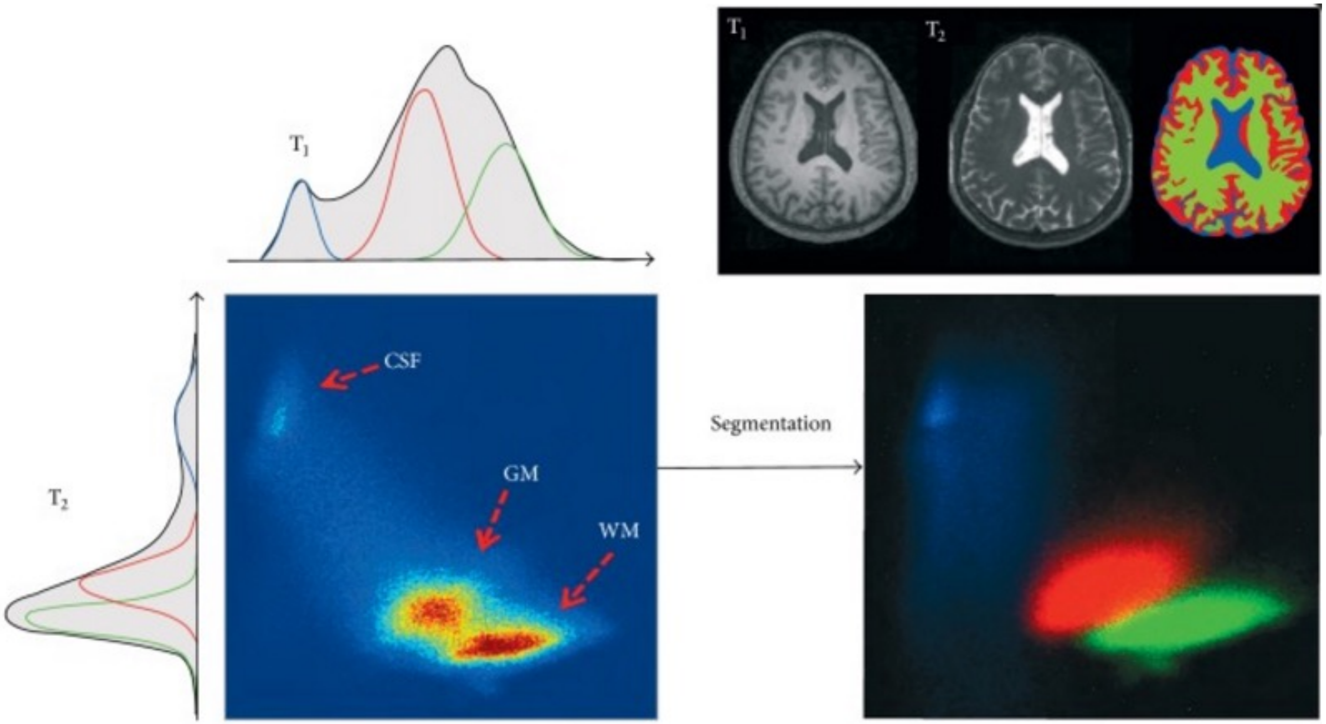


Figure from Despotovic et al. 2015

Methods: Generative Models

Image model:

$$y_j = \sum_{k=1}^{K} z_{jk} v_k + \eta_j, \; j \in \Omega$$

where

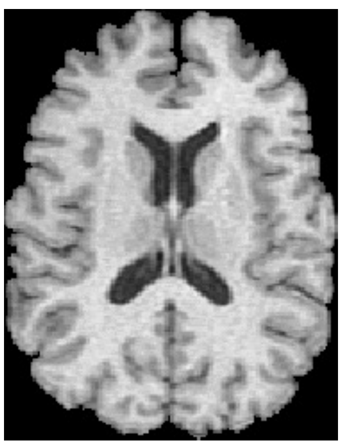
$$z_{jk} = \begin{cases} 1 & \text{if pixel } j \text{ is in tissue class } k \\ 0 & \text{otherwise} \end{cases}$$

- Assume the true image is K-valued with unknown intensities v_k (K assumed known)
- Assume noise is white and Gaussian with variance σ^2
- Segmentation is obtained by estimating z_{jk} and v_k
- This is a generative model because we have an explicit expression for synthesizing observable data
- Can be solved using a k-means algorithm or expectation-maximization

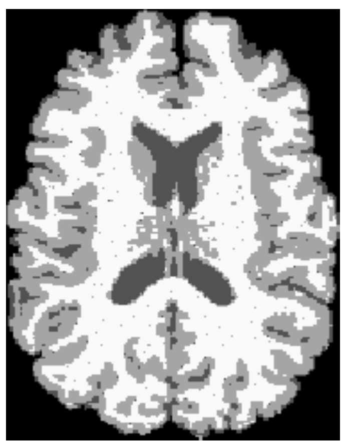
Methods: Unsupervised Classifiers

- **Unsupervised classifiers (aka clustering algorithms)** essentially perform the same function as classifiers without the use of training data (unsupervised)
- Common clustering algorithms: K-means, EM algorithm, fuzzy c-means

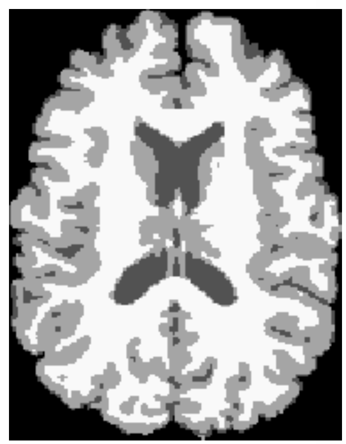
$$\min_{z,v} \sum_{j}^{N} \sum_{k=1}^{K} z_{jk} \left\| y_j - v_k \right\|^2$$



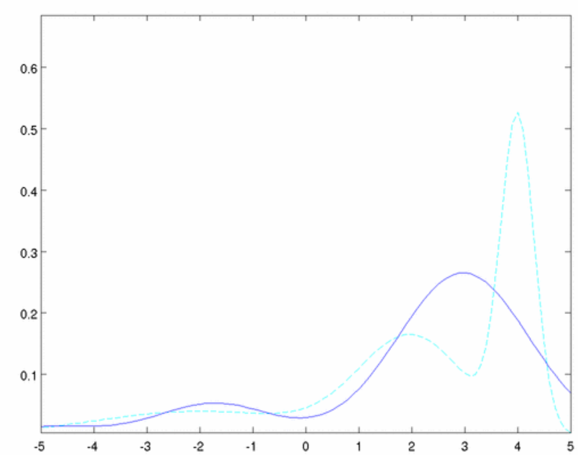
MR Image



K-means segmentation



Regularized K-means segmentation



Fitting the histogram with a Gaussian mixture

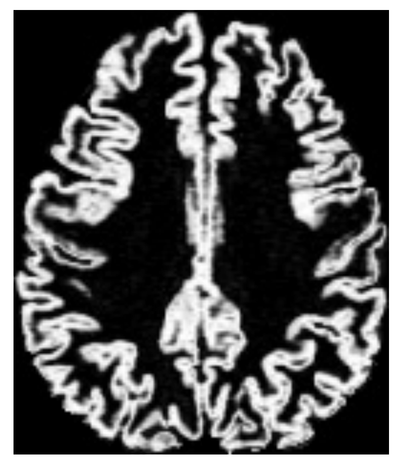
https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FAST
http://mipav.cit.nih.gov
AFNI 3dkmeans

Methods: Soft Segmentation

Class 1 - ■ Class 2 - □

	Hard Membership	Soft Membership	Hard Membership	Soft Membership
	1	1	0	0
	1	0.8	0	0.2
	?	0.5	?	0.5
	0	0.3	1	0.7
	0	0	1	1

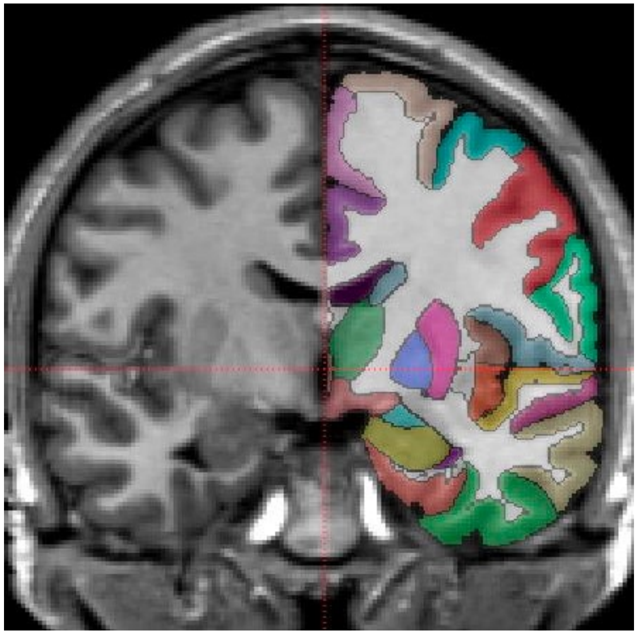
Hard



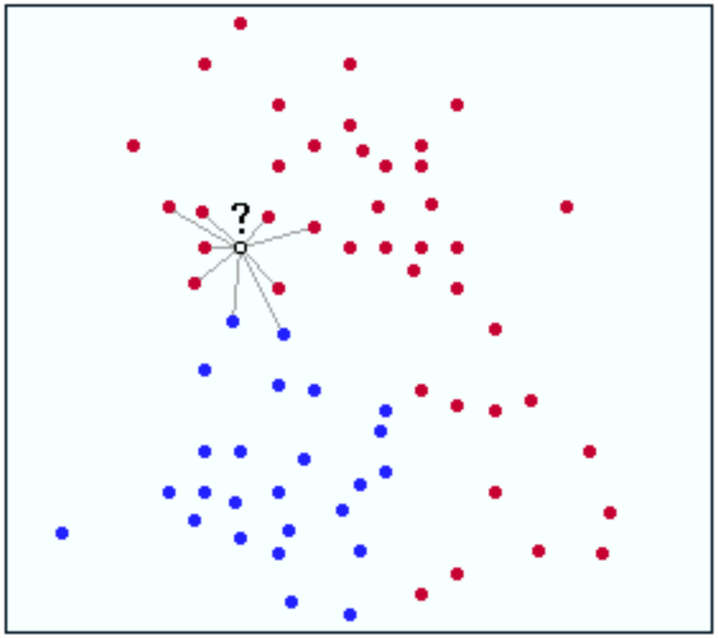
Soft

- Soft segmentations/classifications preserve more information from the image

Methods: Supervised Classifiers



Manually segmented data from
http://www.neuromorphometrics.com



Red and blue circles represent training data from two classes. How should the white circle be classified?

- Supervised classifiers attempt to partition the feature space based on training data
- Training data are data with known labels. For segmentations, this is typically manually segmented data.
- Supervised classifiers generally outperform unsupervised classifiers in terms of accuracy.

Methods: Supervised Classifiers

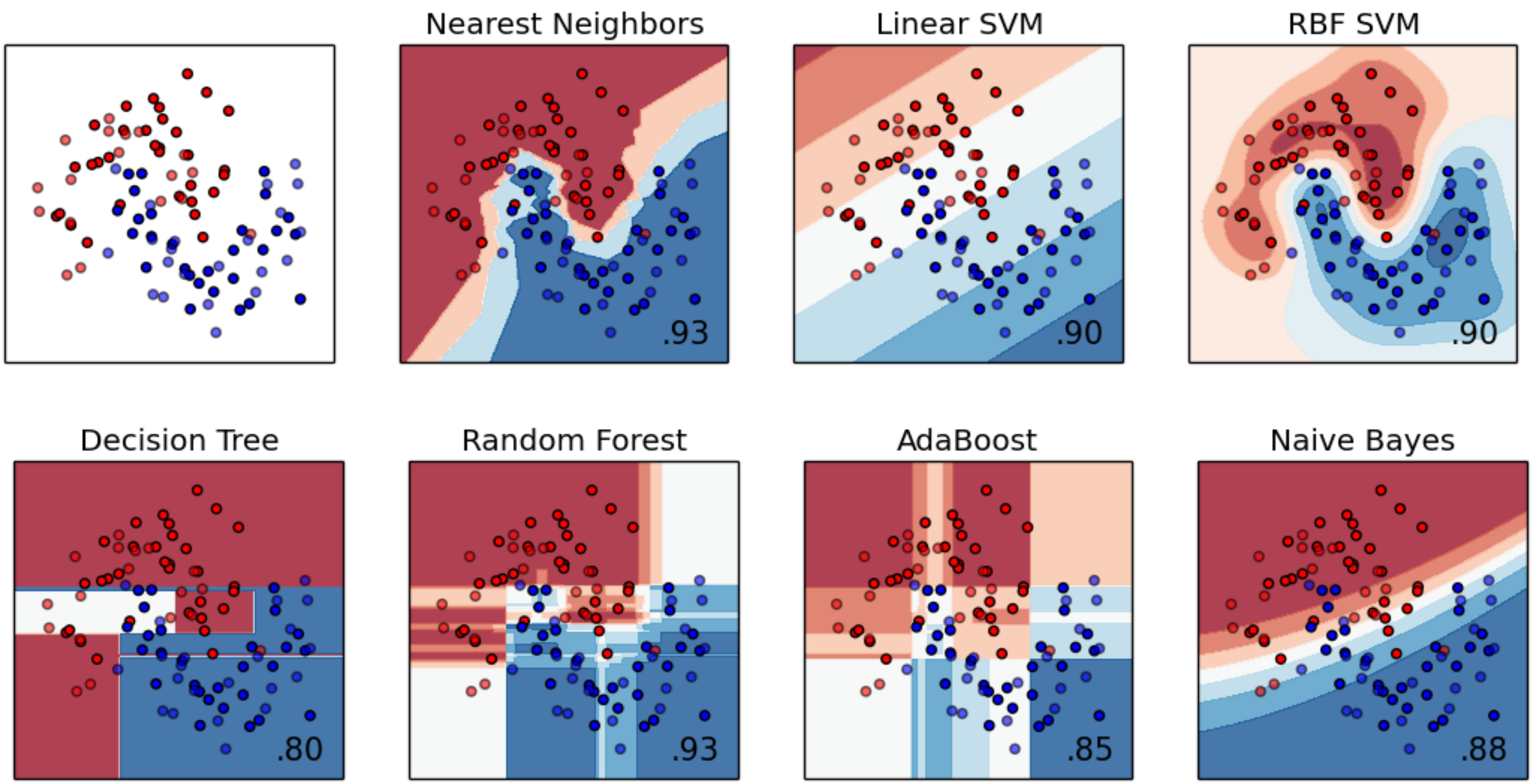


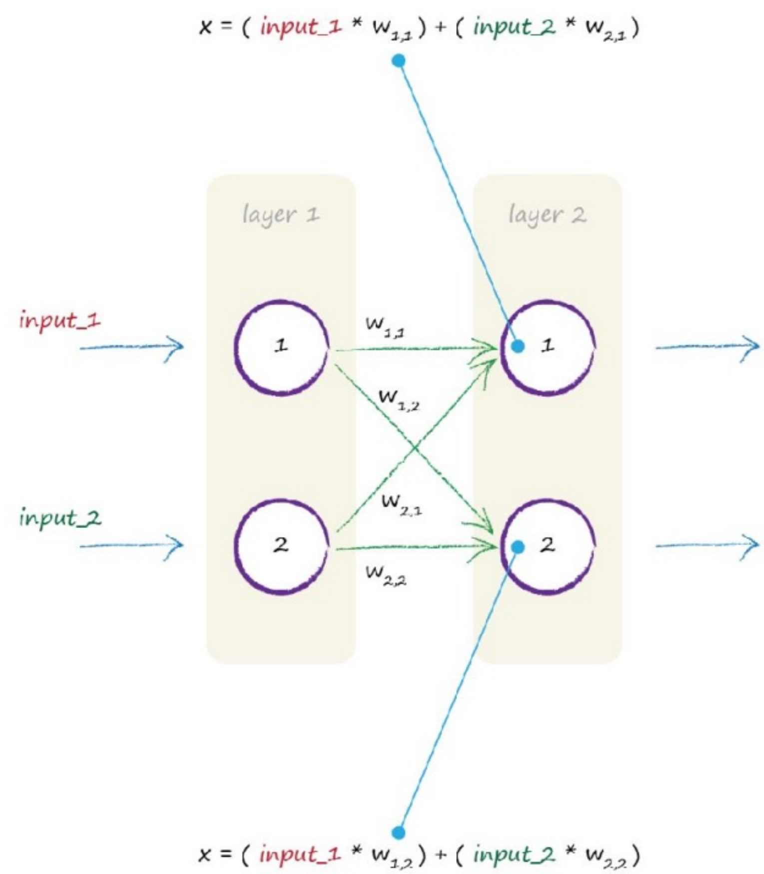
Figure from http://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html

- There are a number of supervised classifiers with different advantages and disadvantages with respect to ease of implementation and amount of training data required
- Performance will depend on the properties of the data

Many Python, R, Matlab libraries are available
https://www.nitrc.org/projects/s3dl

Methods: Artificial neural networks

- **Artificial neural networks** are parallel networks of processing elements or nodes that simulate biological learning
- Used as classifiers, they can be highly effective, but mathematically vague
- Training data are used to determine the weights
- Because of their interconnected structure, important textural features can be automatically determined
- Neural networks can be efficiently implemented on parallel computers or graphical processing units



Example of a simple neural network
from Rashid 2016

Methods: Deep learning networks

- Deep learning networks are simply big neural networks
- Have shown to be highly effective in a number of learning applications. In neuroimaging, they are increasingly being used for segmentation.
- Main disadvantage are that large training data sets are typically needed for optimum performance.

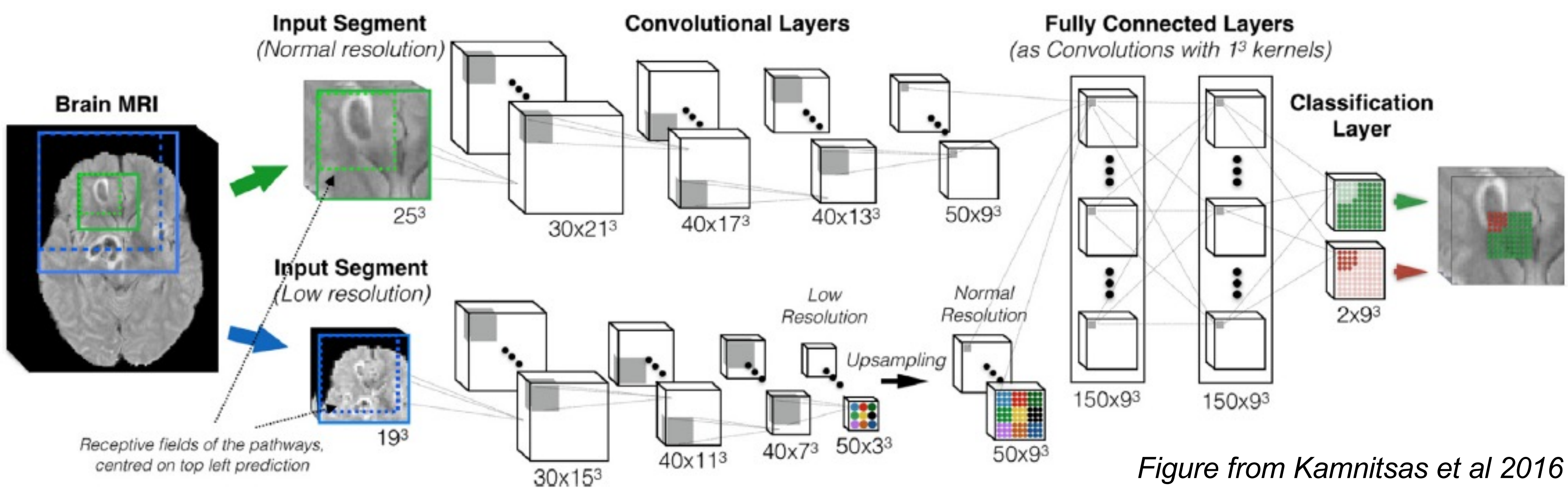
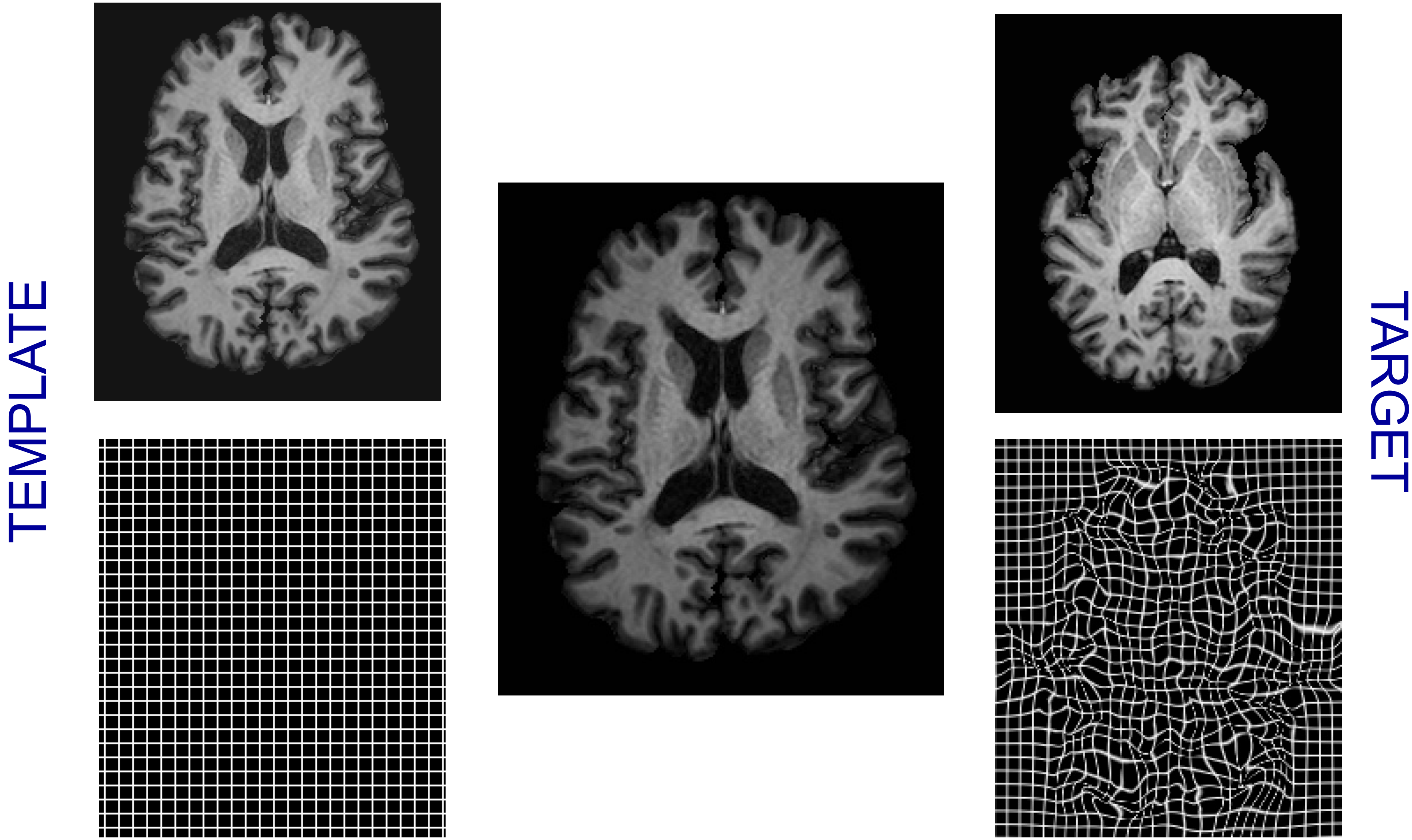


Figure from Kamnitsas et al 2016

https://github.com/Kamnitsask/deepmedic
https://github.com/TJKlein/DeepNAT
https://list.nih.gov/cgi-bin/wa.exe?SUBED1=DLMIB&A=1

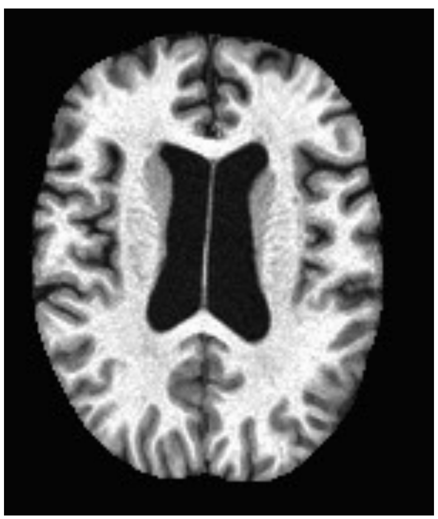
Methods: Deformable Registration



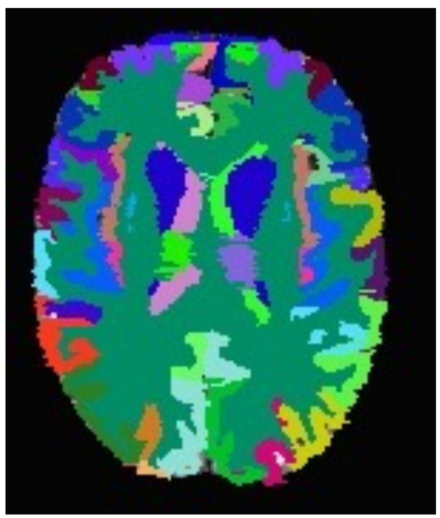
http://stnava.github.io/ANTs/
http://www.mpheinrich.de/software.html
http://elastix.isi.uu.nl/

Methods: Multi-Atlas Label Fusion

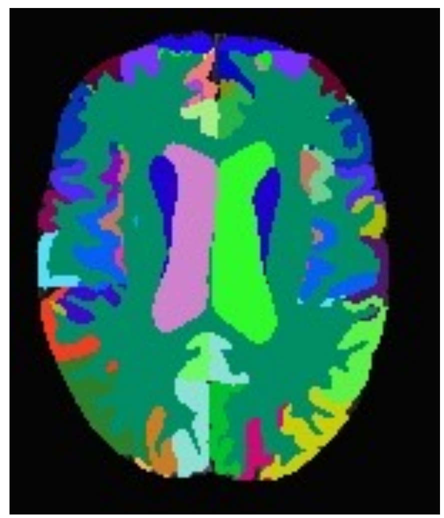
- **Atlas-guided** approaches register the image to a pre-segmented template or atlas
- A single atlas is often insufficient because large differences in anatomy can not be accommodated by the registration
- **Label fusion** techniques use multiple atlases and select a label based on each deformed results
- Majority voting is an example of a simple label fusion approach
- Can find boundaries even when there are no distinguishing features



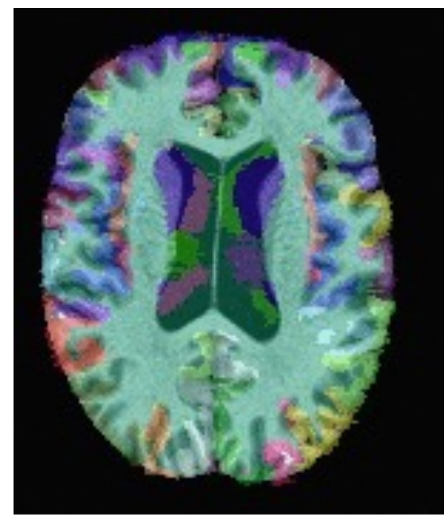
Target image



Single atlas result



30 atlas label fusion

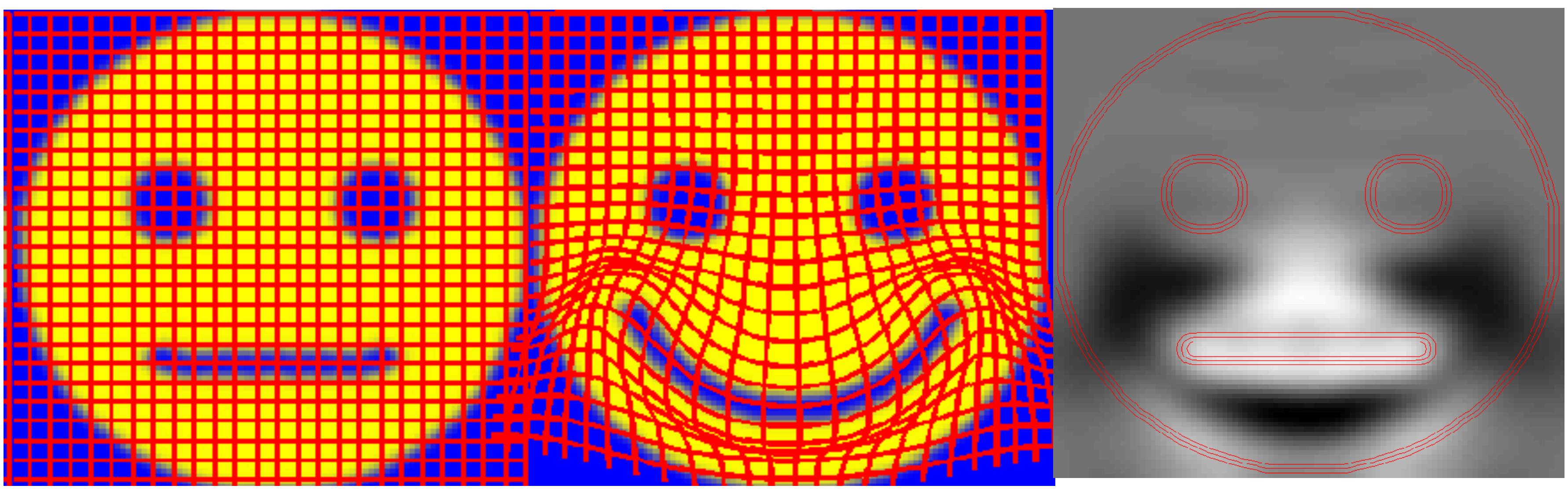


Varying from 1 to 30 atlases

https://www.nitrc.org/projects/picsl_malf/
https://github.com/ledigchr/MALPEM
https://www.nitrc.org/projects/masi-fusion

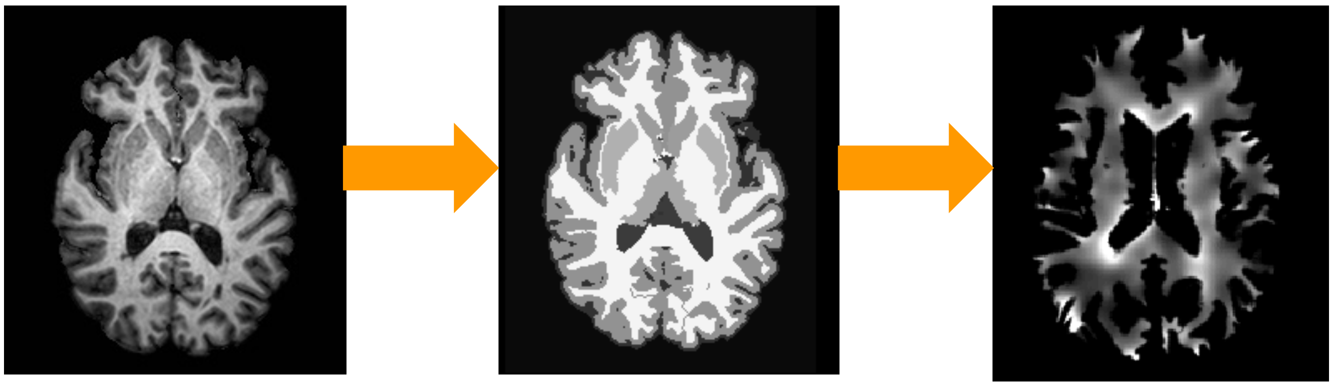
Methods: Voxel-based Morphometry

- Approach popularized by SPM package for localizing regional volumetric group differences (not really a segmentation method)
- Obtains a tissue classification within a normalized space for two groups in a data set (*e.g. disease vs. control*), and then determines areas of statistically significant differences
- Tissue classes are modulated by the jacobian of the deformation
- Bright regions indicate volume expansion, dark indicate compression

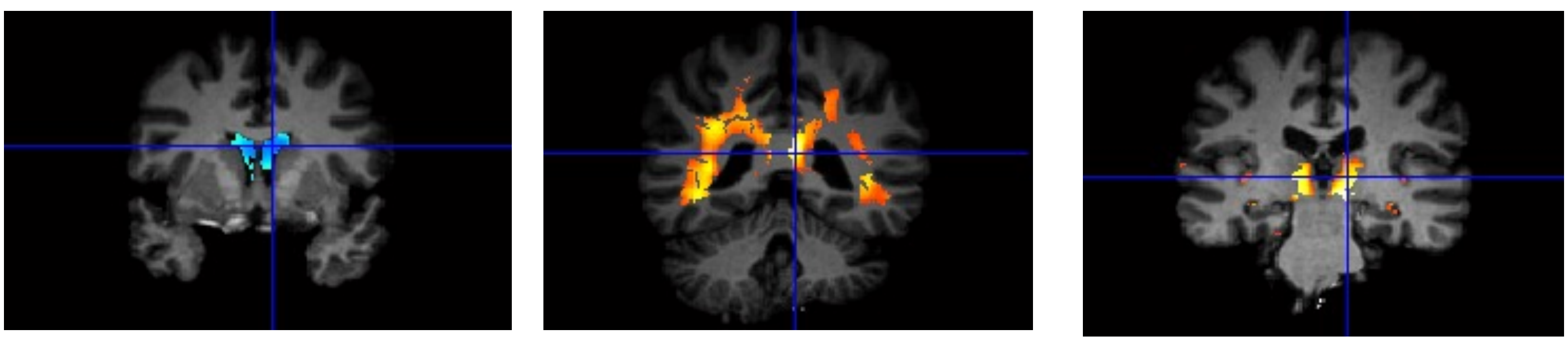


Example image | Deformed image | Jacobian

Methods: VBM Analysis



Pre-processed structural T1 | Tissue Segmentation | Modulated Images

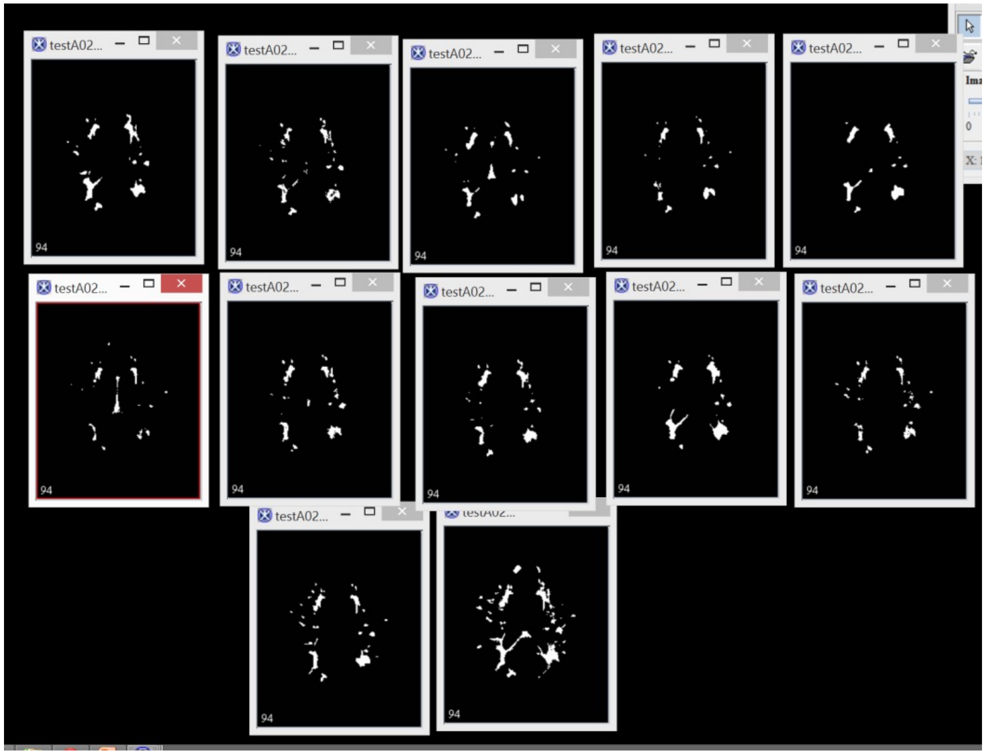


Statistically significant regions of volumetric expansion/contraction of tissue in MS vs. Healthy Controls

http://www.fil.ion.ucl.ac.uk/spm/
https://www.cbica.upenn.edu/sbia/software/dramms/index.html

Validation

- Validation of automated methods is typically compared against manual delineation
- Computational or physical phantoms may also be used for validation
- Figures of merit can be based on region information, such as the number of misclassified pixels, or boundary information, such as distance to the true boundary
- Reproducibility is often not a good metric since most algorithms are deterministic

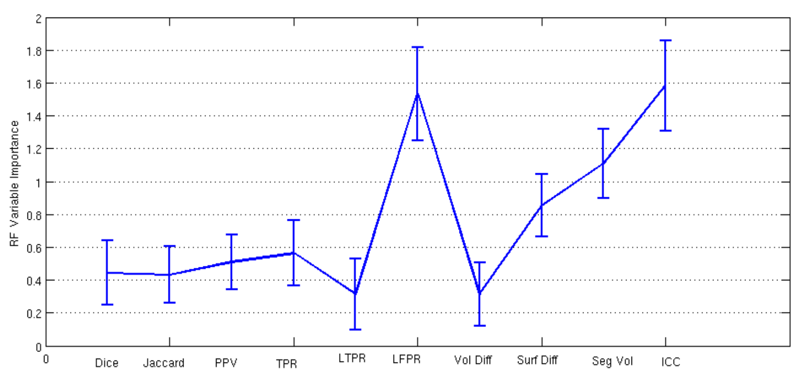


Results of different algorithms from the 2015 Longitudinal MS Lesion Segmentation Challenge

https://smart-stats-tools.org/node/26

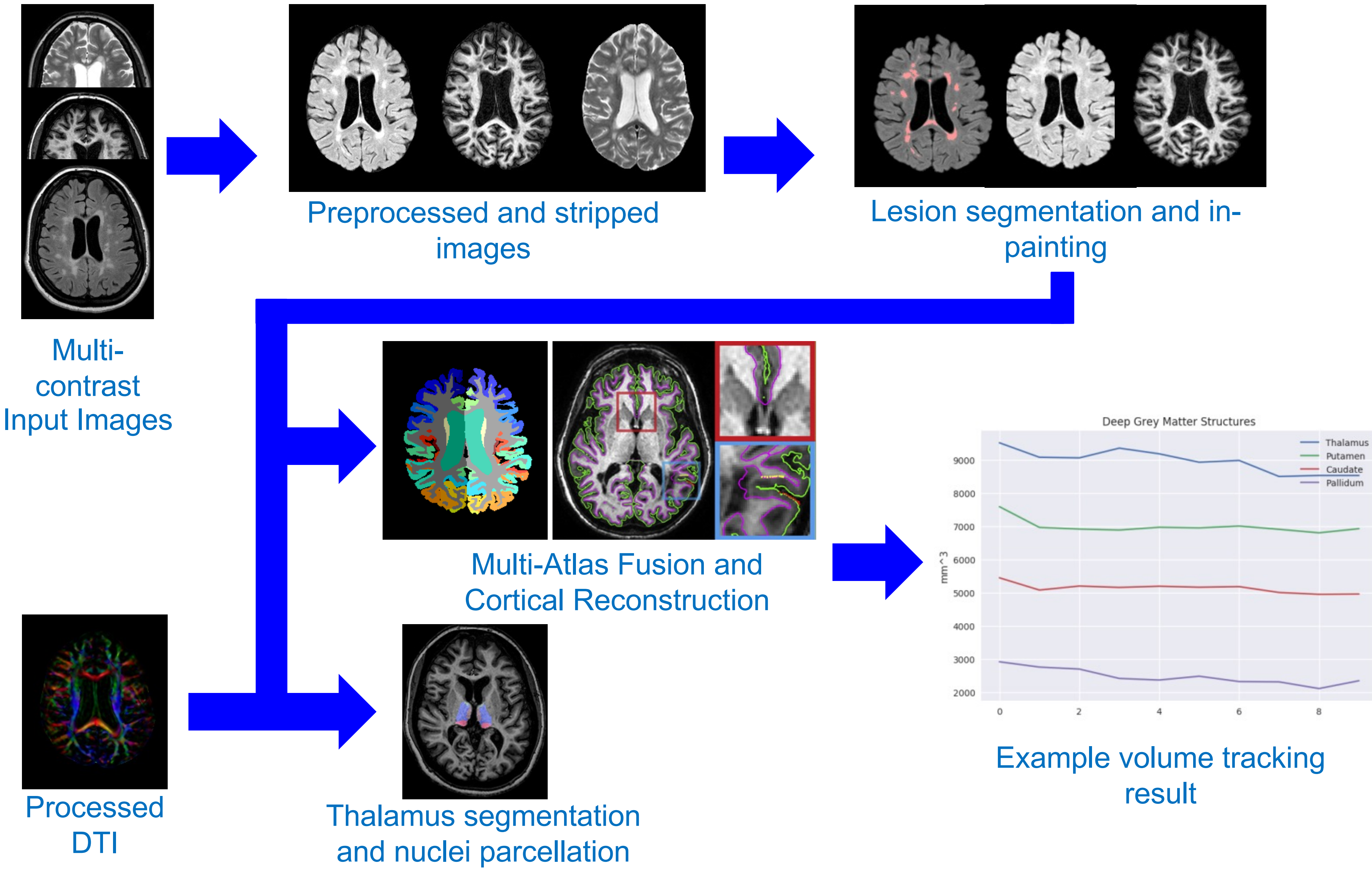
Validation Metrics

- Dice Overlap – the ratio of twice the number overlapping voxels to the total number of voxels in each mask
- Jaccard Overlap – the ratio of the number of overlapping voxels to the number of voxels in the union of each mask
- PPV (positive predictive value) – the ratio of voxel-wise true positives to the sum of true and false positives
- TPR (sensitivity, voxel based) – the ratio of voxel-wise true positives to the sum of true positives and false negatives
- LTPR (lesion TPR based on lesion count) – the ratio of lesion-wise true positives to the sum of true positives and false negatives
- LFPR (lesion FPR based on lesion count) – the ratio of lesion-wise false positives to the sum of false positives and true negatives
- Volume Difference – absolute difference in volumes divided by the true volume
- Surface Difference – average symmetric surface distance
- Segmentation Volume – total volume of segmentation mask for reference purposes
- Manual Volume – total volume of reference mask for reference purposes
- Volume Change Correlation – average linear correlation of changes in lesion volumes between successive time-points
- New lesion detection TPR – ratio of number of new lesions detected to number of true new lesions
- New lesion detection FPR – ratio of new lesions falsely detected to number of true new lesions
- Volume correlation – Pearson's correlation coefficient of all volumes
- Longitudinal volume correlation – Pearson's correlation coefficient of volumes within a subject

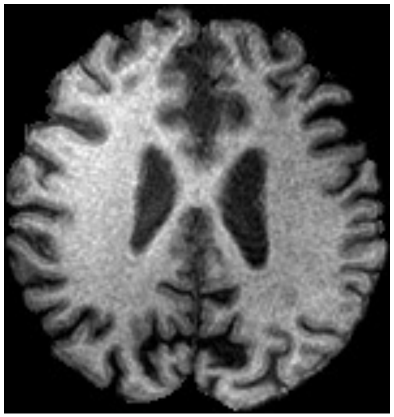


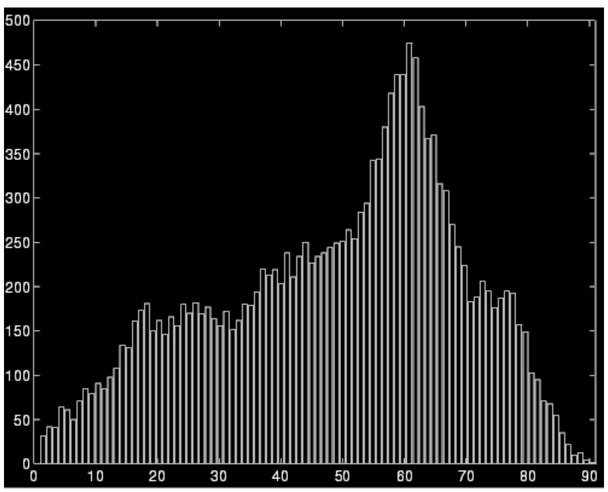
How quantitative metrics relate to a neuroradiologists' subjective rating of performance

Example Segmentation Pipeline

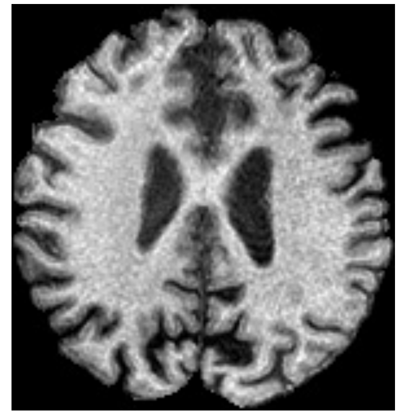


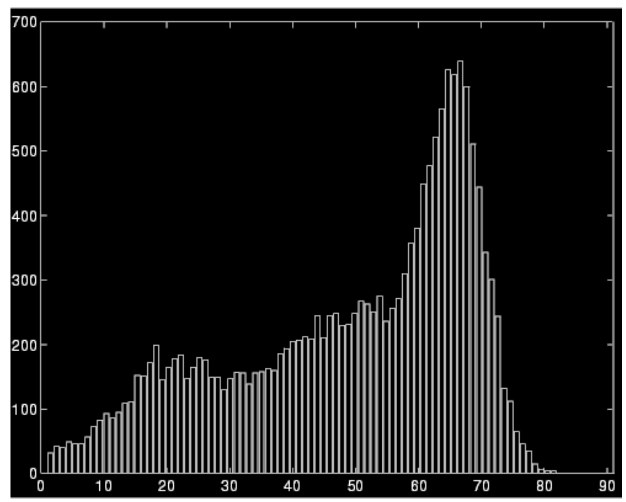
Preprocessing: Inhomogeneity Correction





Before correction



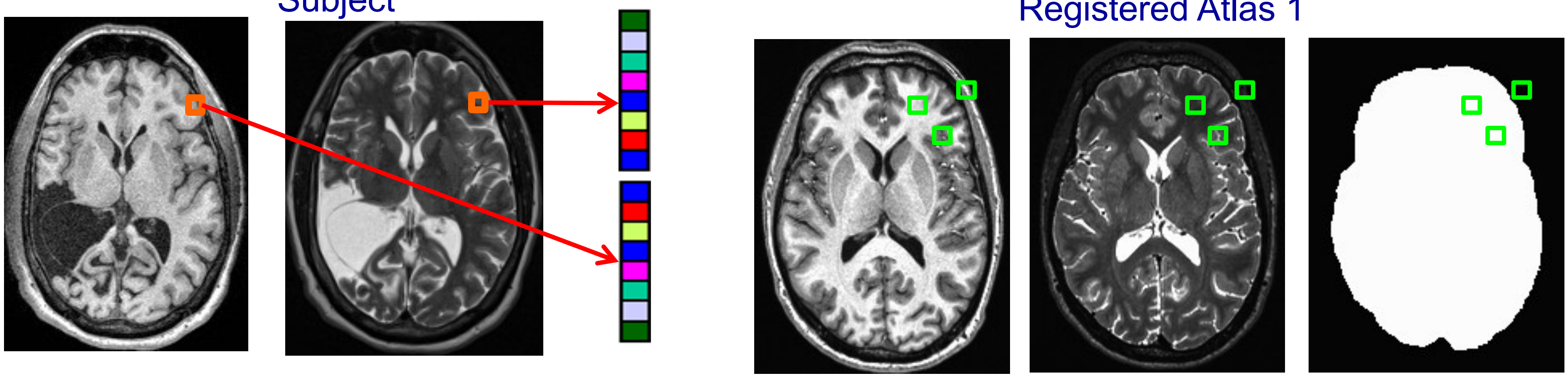


After correction

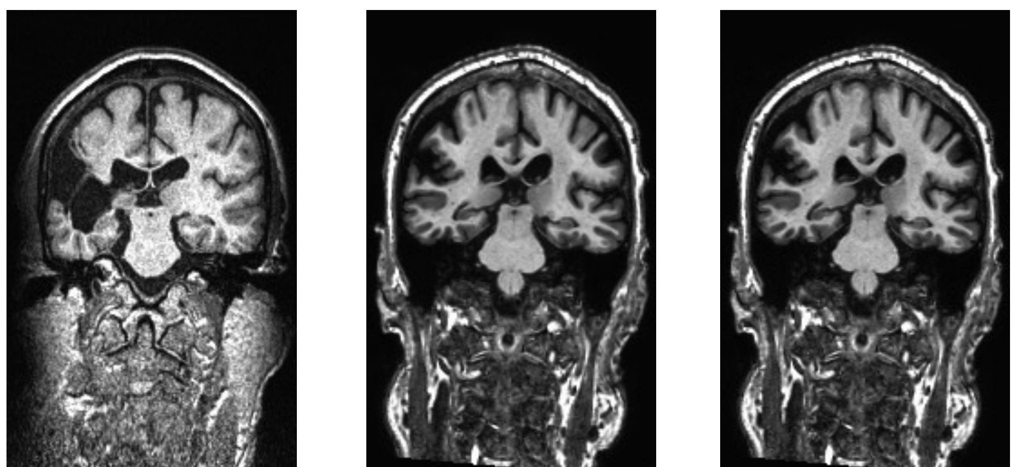
N4ITK http://stnava.github.io/ANTs/

Example Brain Extraction:MONSTR

- Three Primary Innovations: 1) Use multiple MRI contrasts, 2) Apply only a coarse deformable registration, 2) Perform patch matching to allow for greater flexibility in finding correspondences between the training atlases, and the subject data



Deformable Registrations can perform poorly

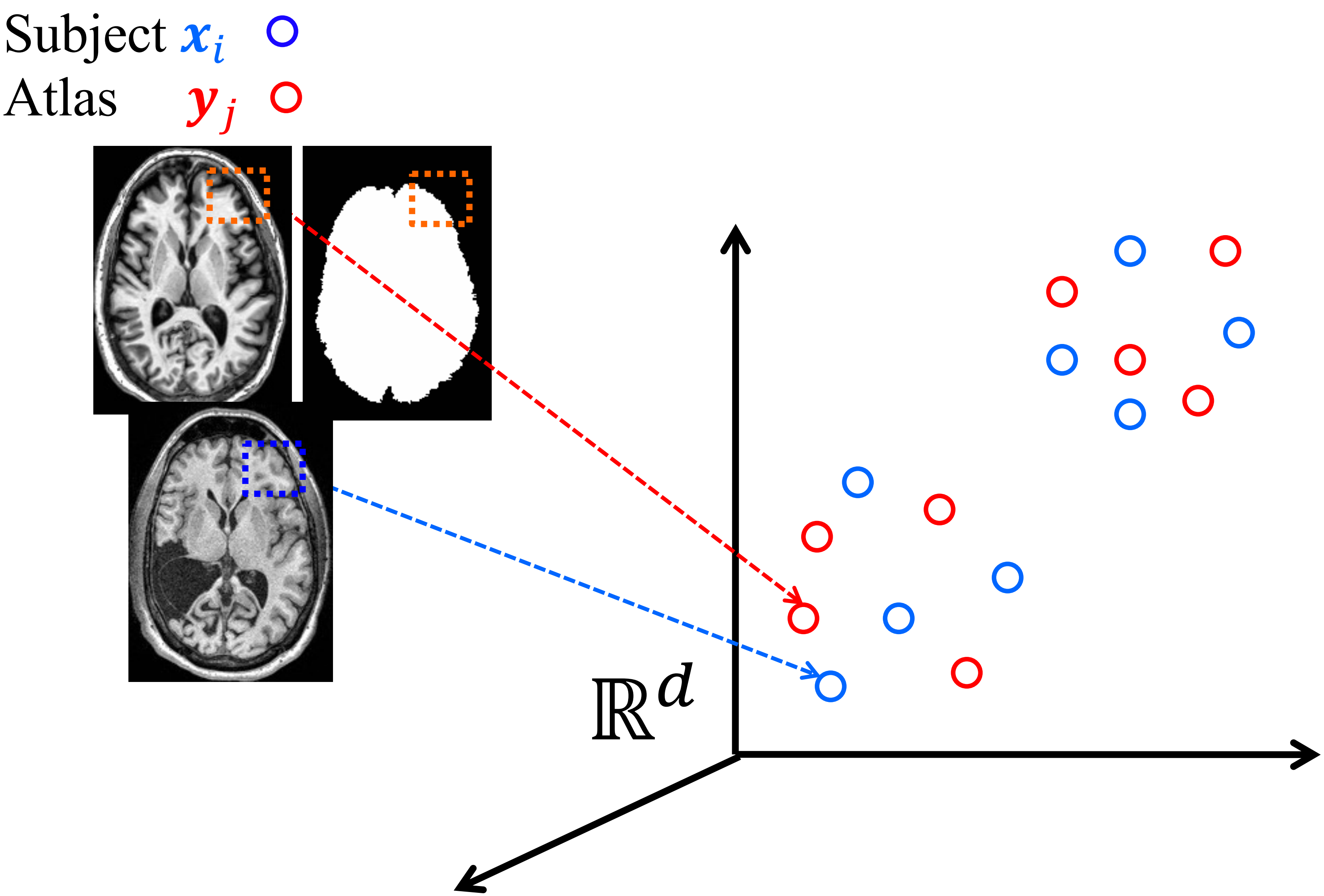


TBI   Atlas   Deformation

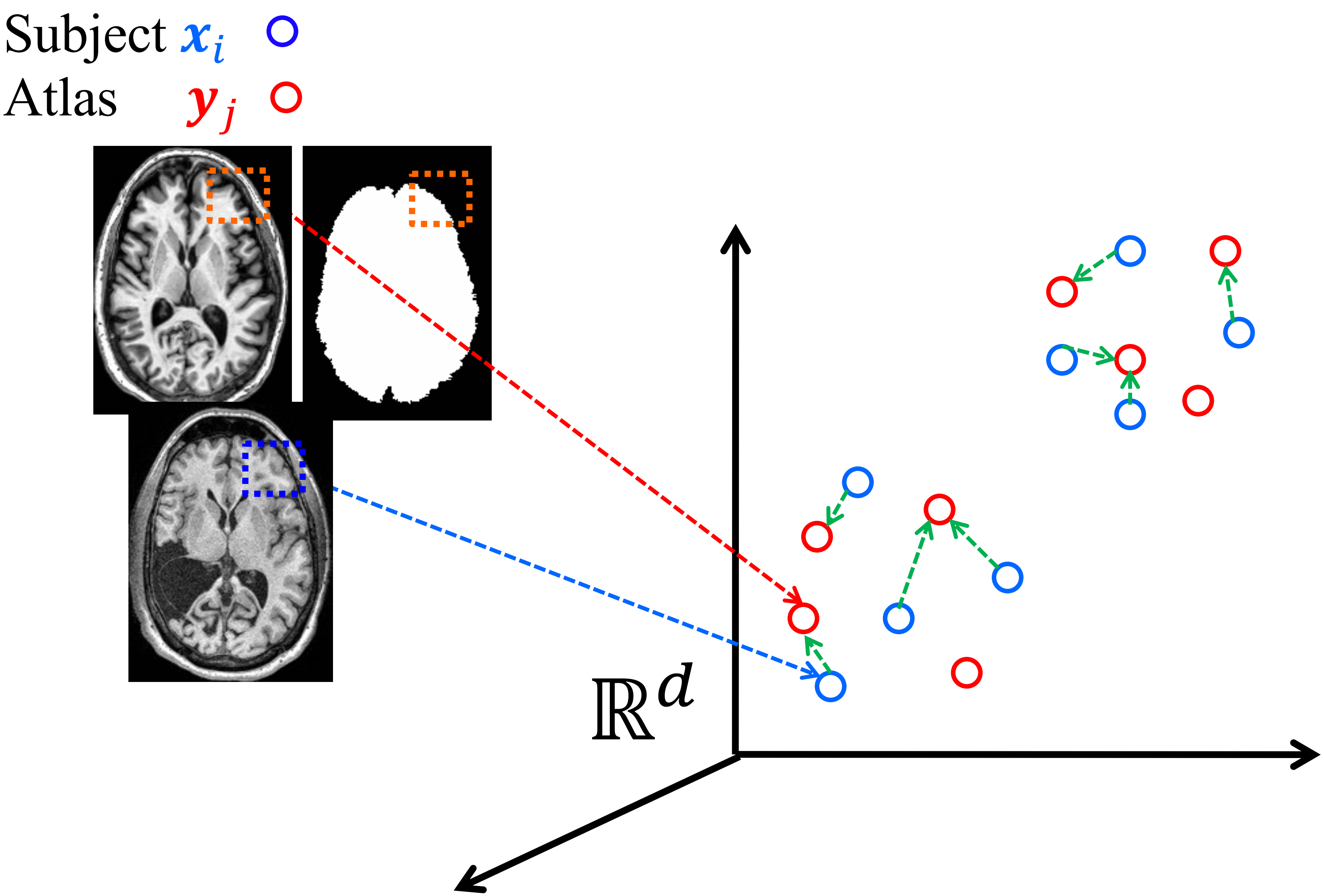
http://www.nitrc.org/projects/monstr

from Roy et al 2017

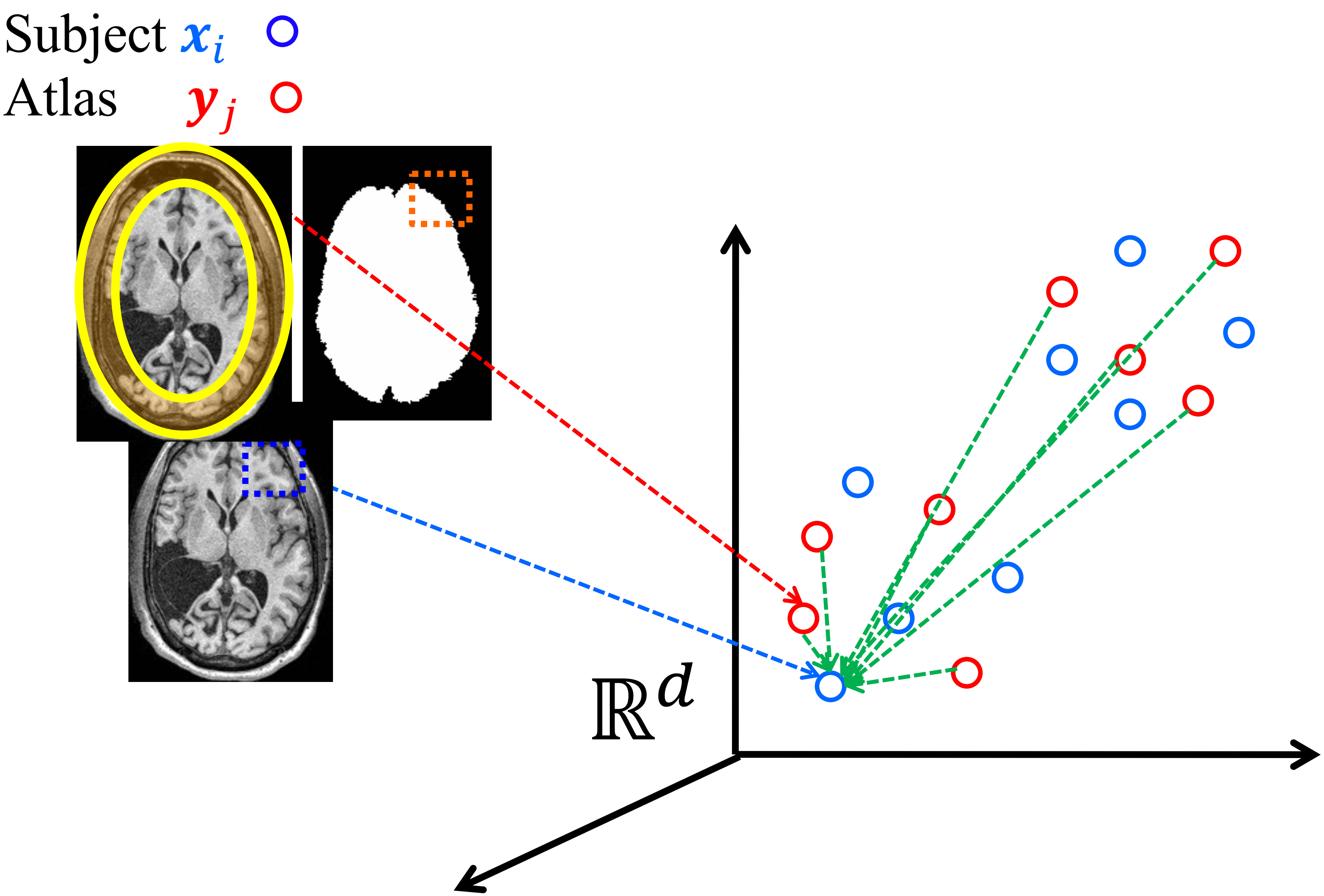
Patch Matching



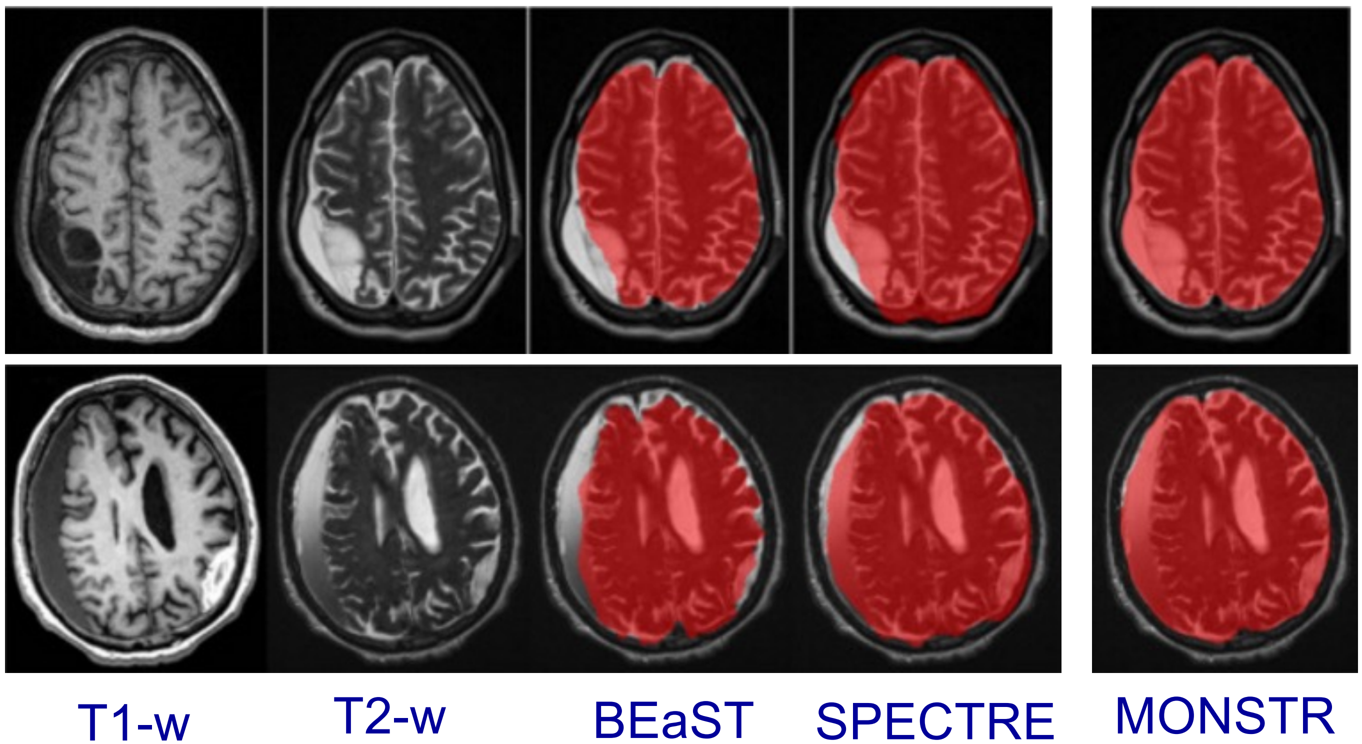
Nearest Neighbor



Weighted Patches

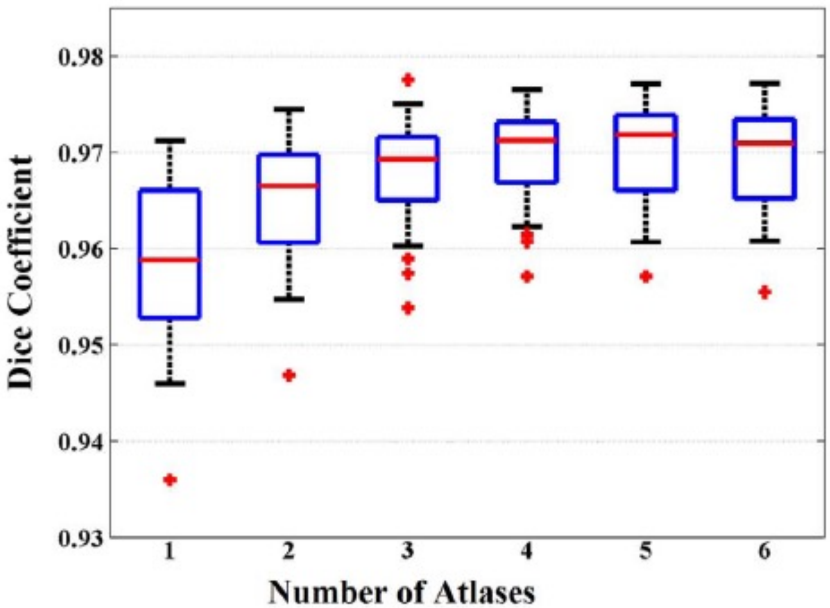


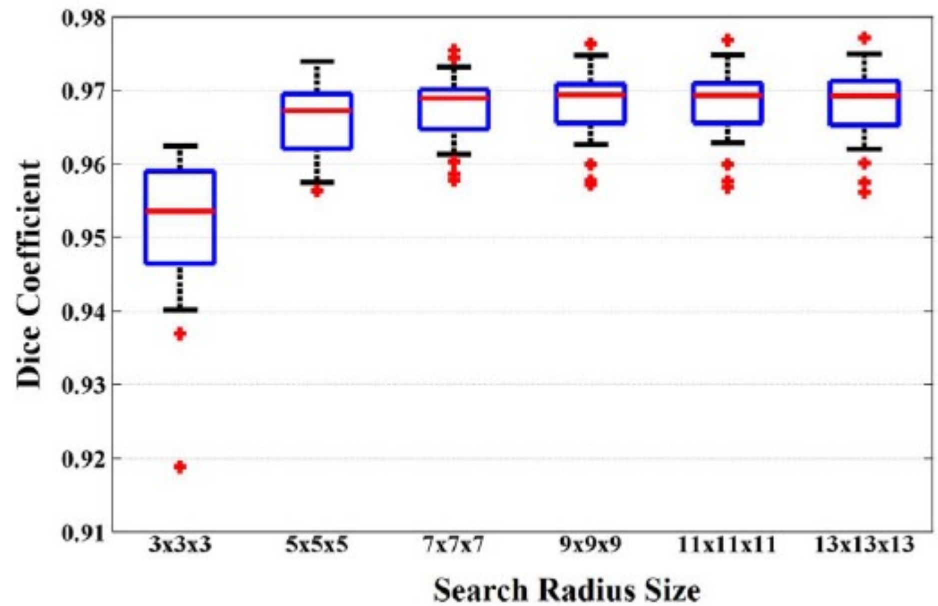
MONSTR: Multi-cONtrast brain STRipping



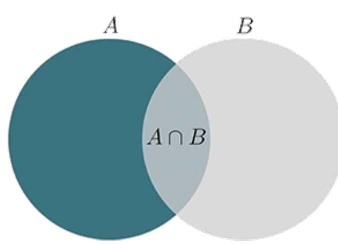
- Validated on multiple data sets
- Compared against other methods on both healthy and diseased brain images
- Improvements were statistically significant

MONSTR Tuning





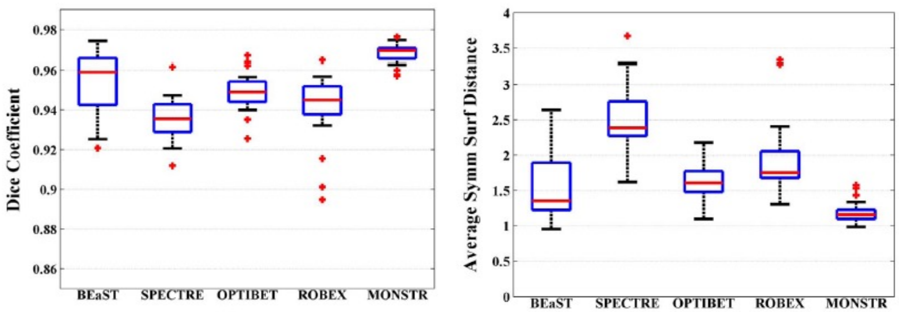
- "Tuning parameters", how did we determine how to select these
- Performed using validation (manually delineated) data



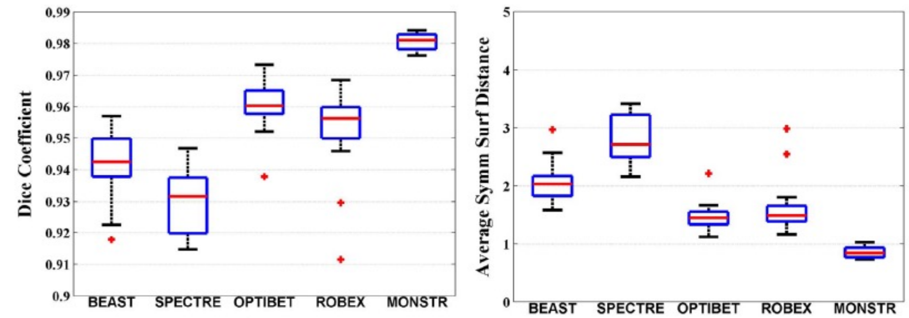
$$DC = \frac{2|A \cap B|}{|A| + |B|}$$

Computation of Dice Coefficient

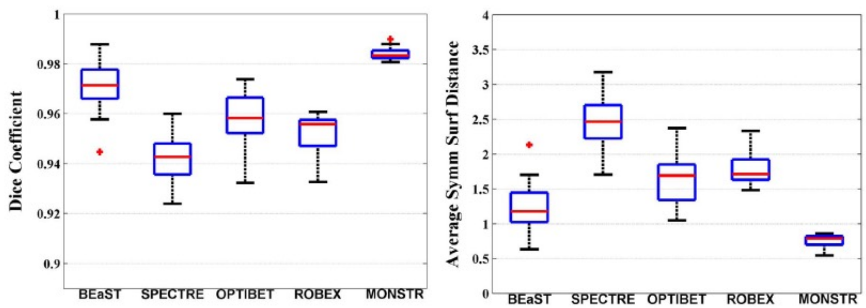
MONSTR Evaluation



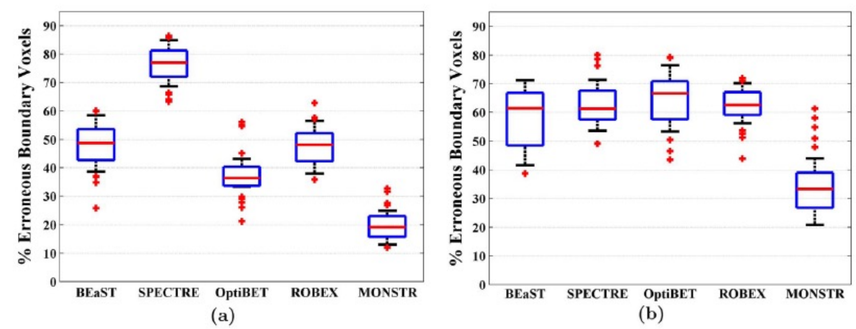
ADNI-29



TBI-19



NAMIC-20



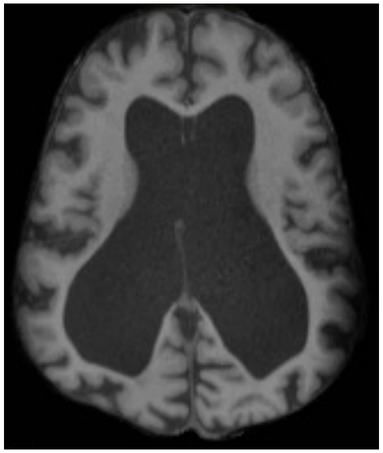
MOV-20 and TUMOR-36

- The last chart demonstrates validation without a manual delineation- a complementary imaging modality (CT) was used instead

http://adni.loni.usc.edu/
https://www.na-mic.org/wiki/Downloads

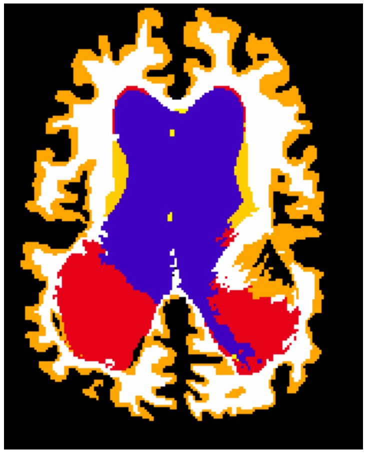
Future Directions: Pathology

Lesions can have a heterogeneous appearance

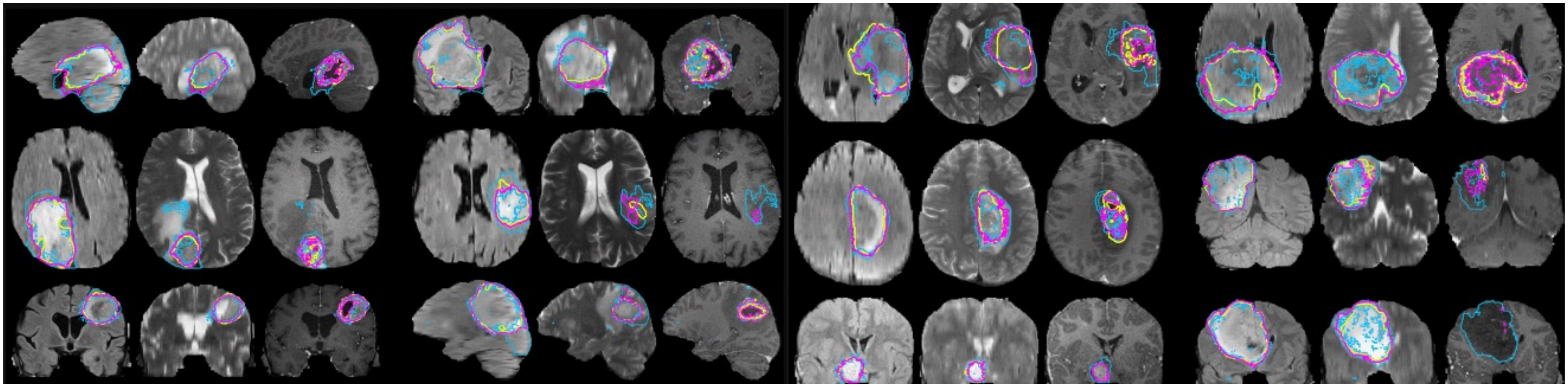


Subject with Normal Pressure Hydrocephalus

Standard FreeSurfer



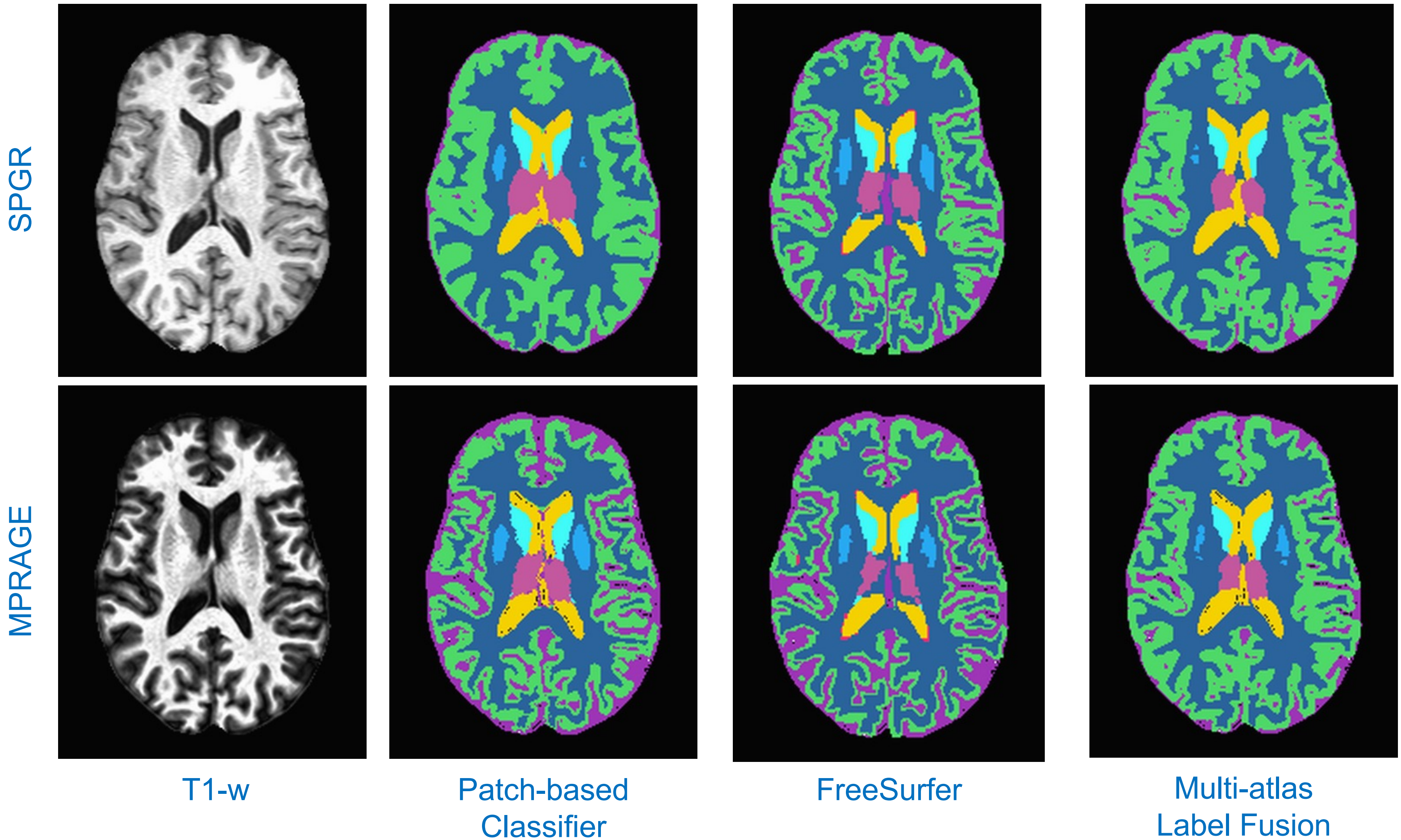
Tweaked FreeSurfer result



BRATS tumor challenge data (*Menze et al 2015*)

http://braintumorsegmentation.org/

Future Directions: Contrast Variations



- Contrast differences affect the segmentation result, even when the underlying anatomy is the same

Conclusions

- sMRI and fMRI should be friends
- There are a large number of approaches and publicly available tools available for brain image segmentation
- Achieving good performance for specific tasks often requires tailored pipelines involving multiple processing steps and multiple segmentation algorithms
- With the shift towards supervised algorithms, the quantity and quality of training data has become extremely important