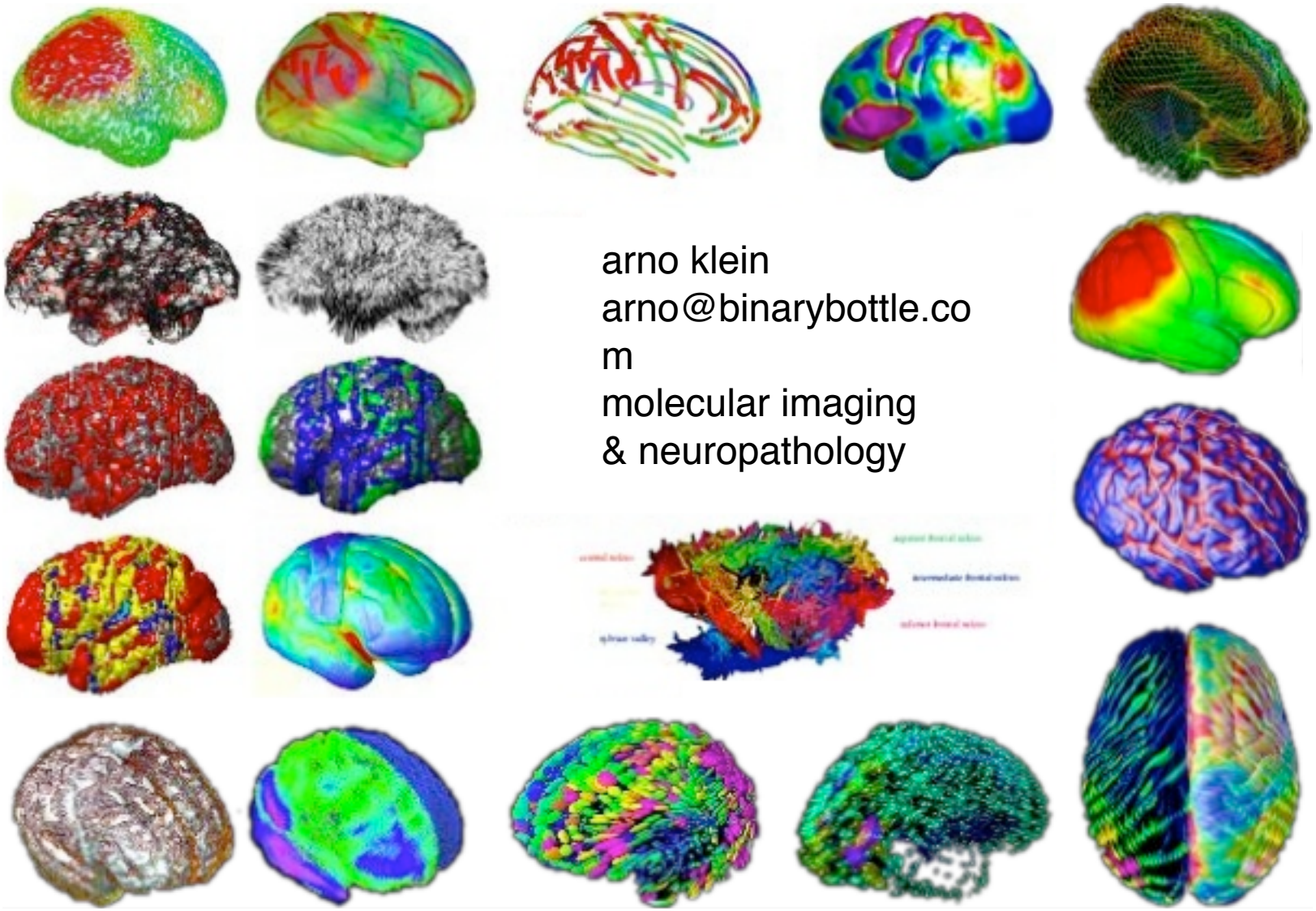
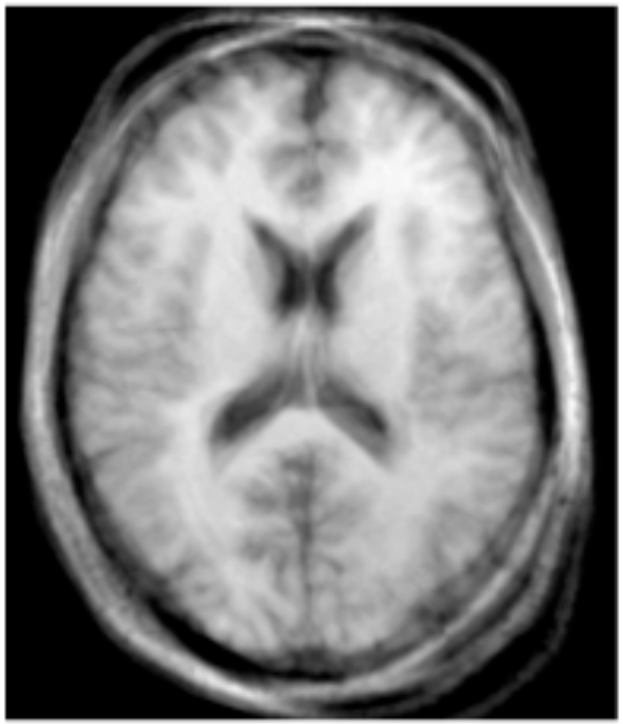
Evaluation of human brain MRI registration algorithms



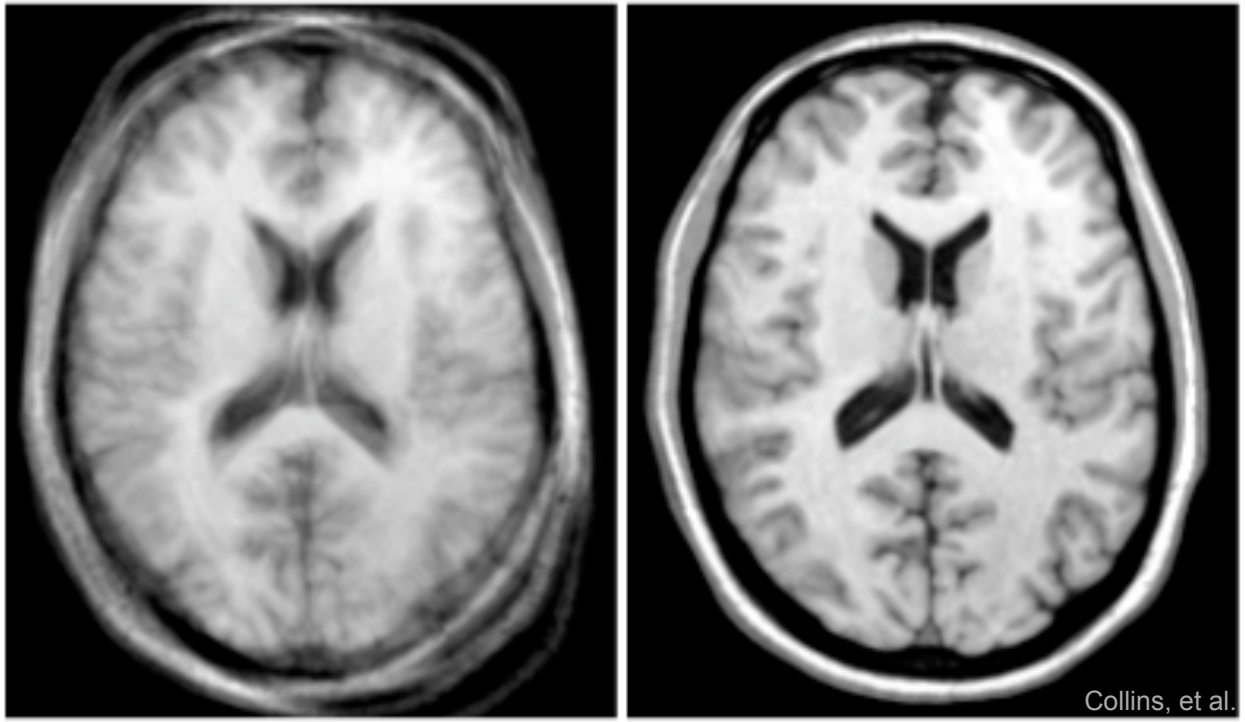
arno klein
arno@binarybottle.com
molecular imaging
& neuropathology

Why register brain images?



Why register brain images?

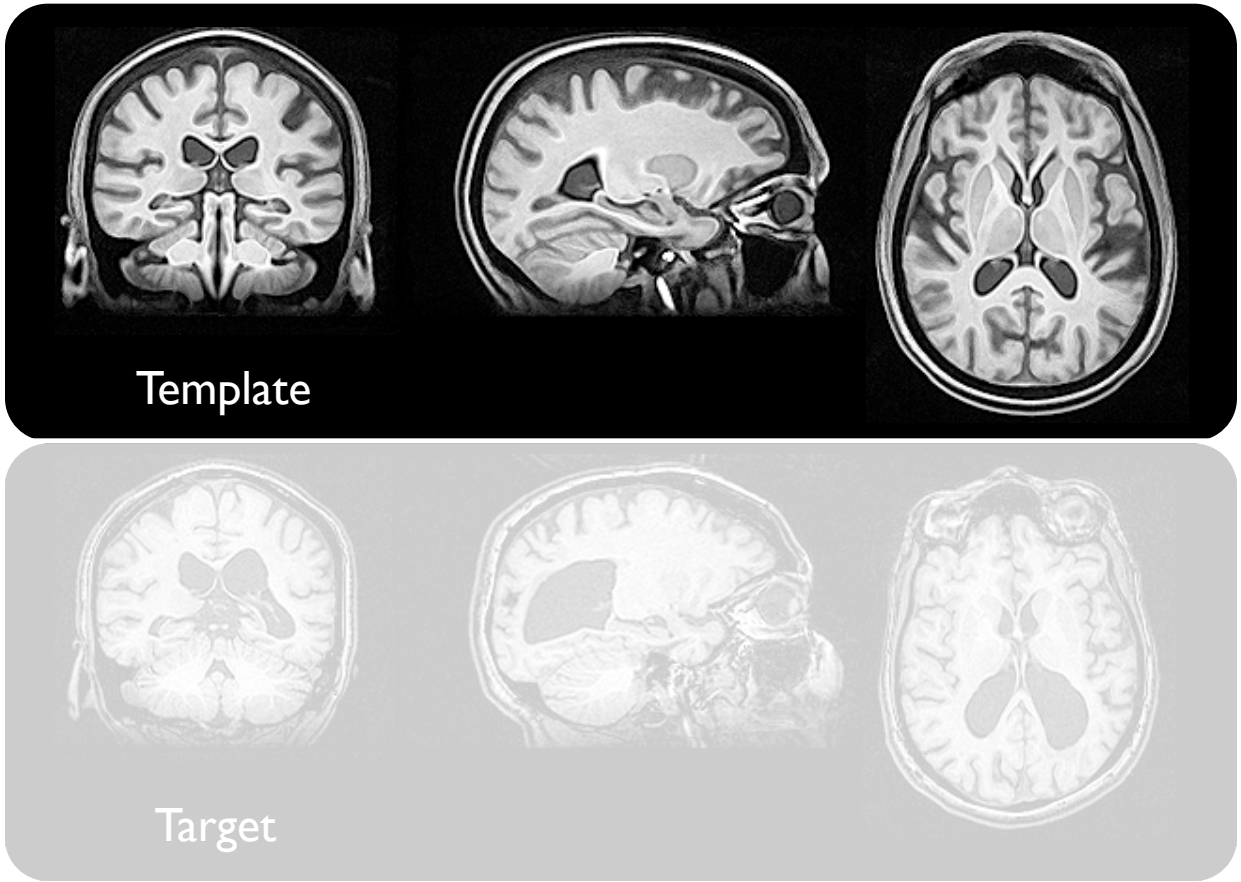
To establish correspondences.



linear nonlinear

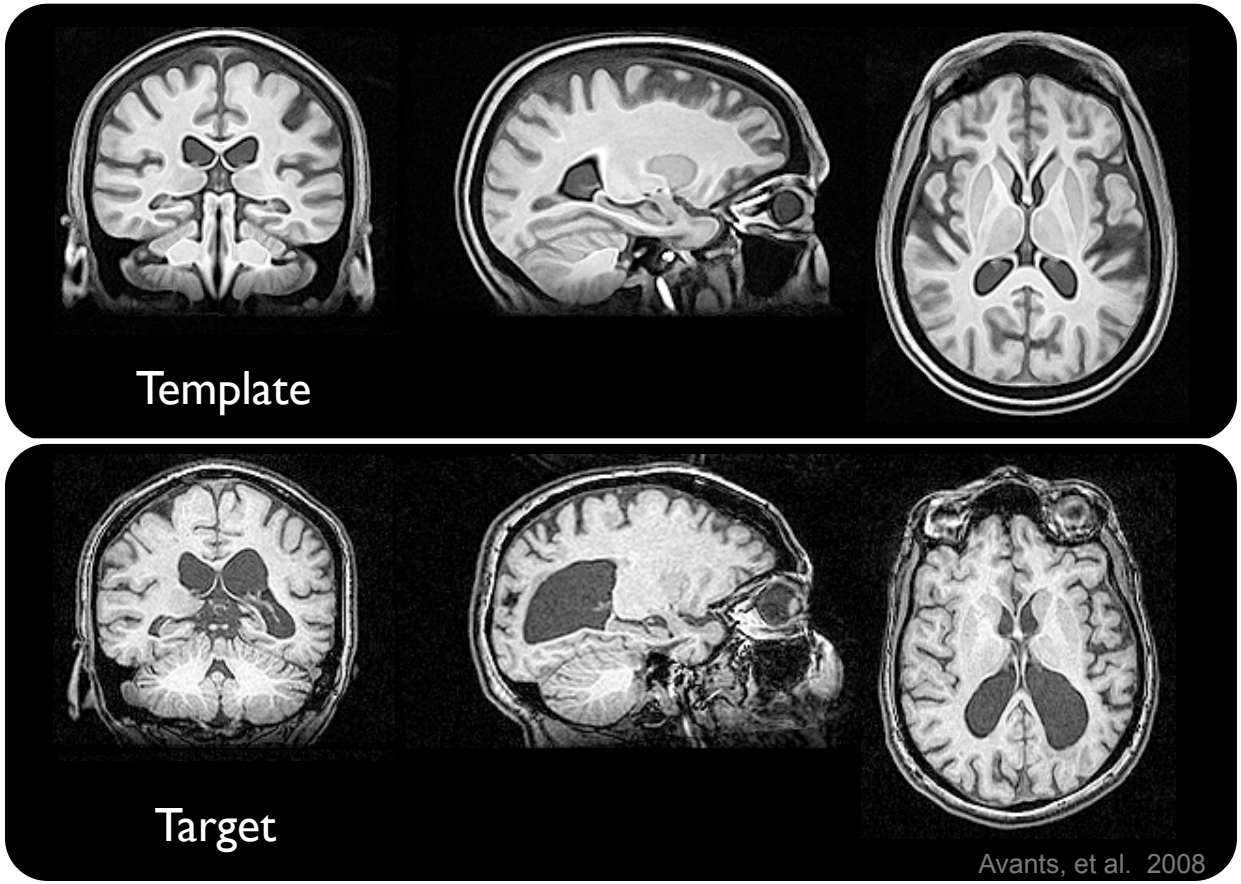
The correspondence problem

template registration



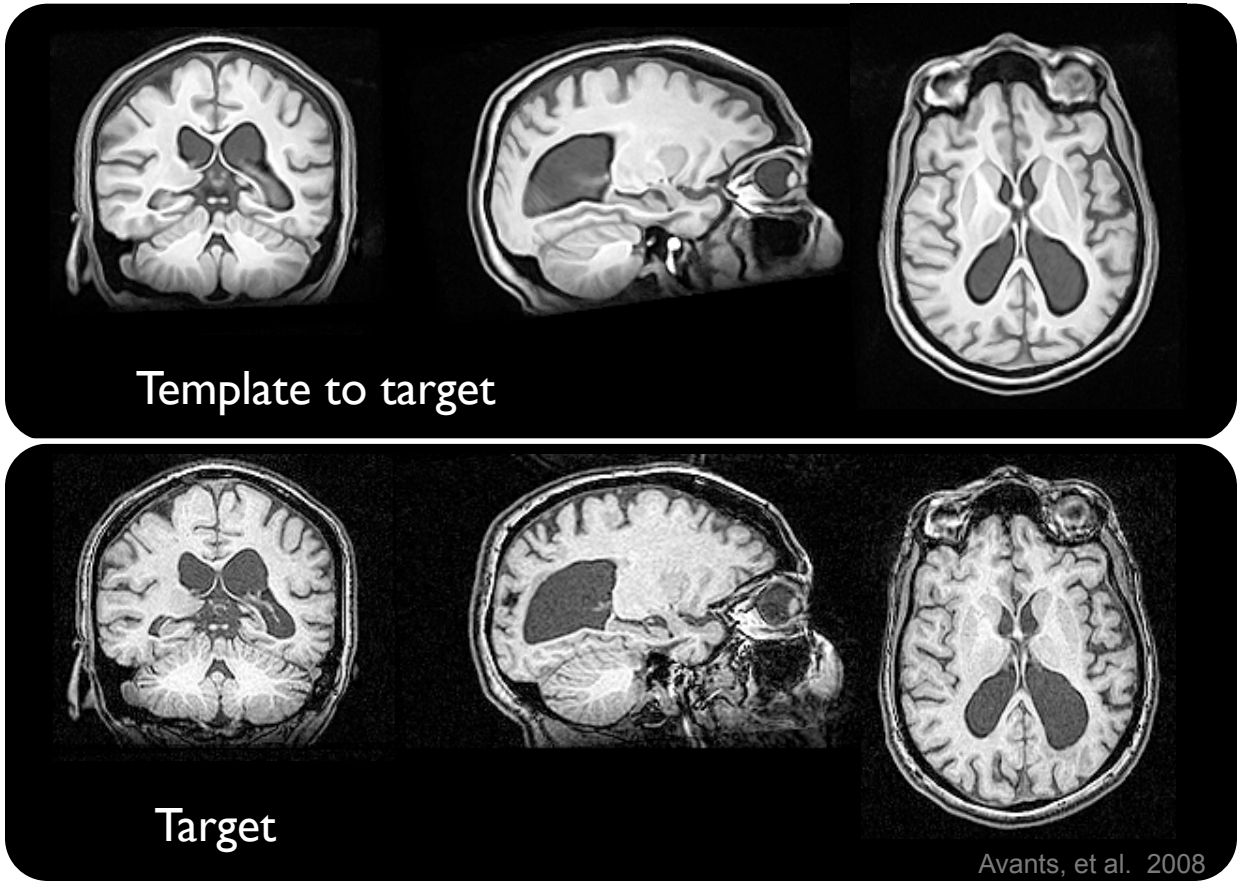
The correspondence problem

template registration



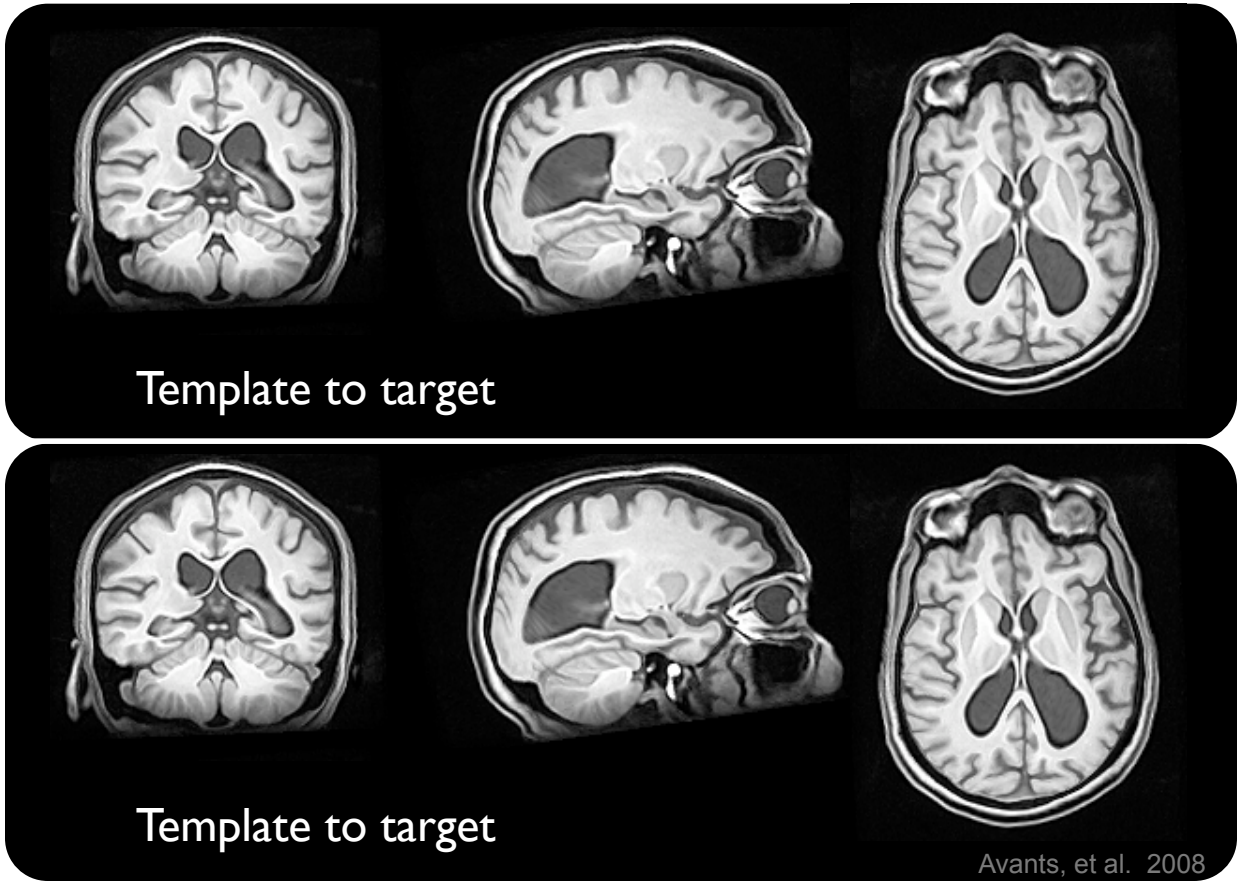
The correspondence problem

precise, but accurate?



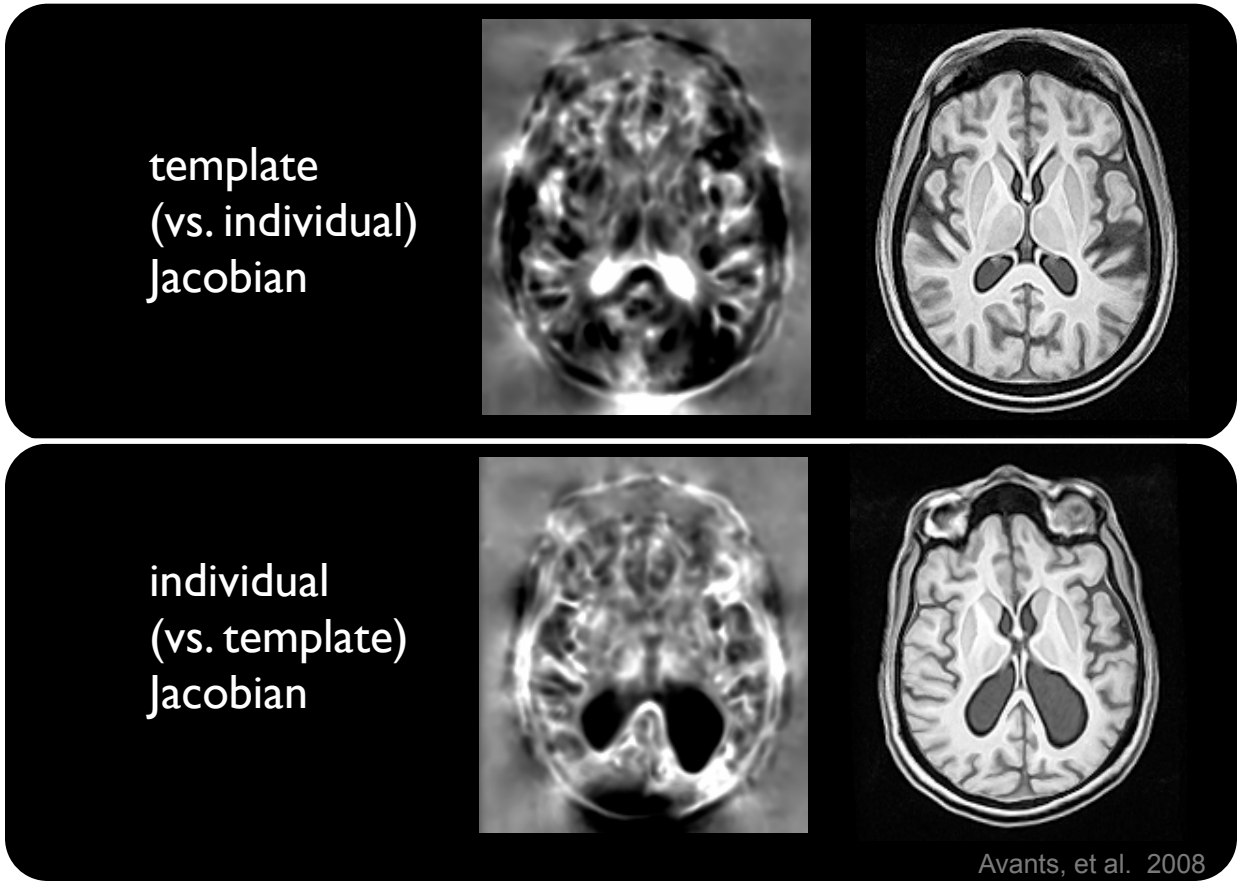
The correspondence problem

precise, but accurate?



Registration application 1:

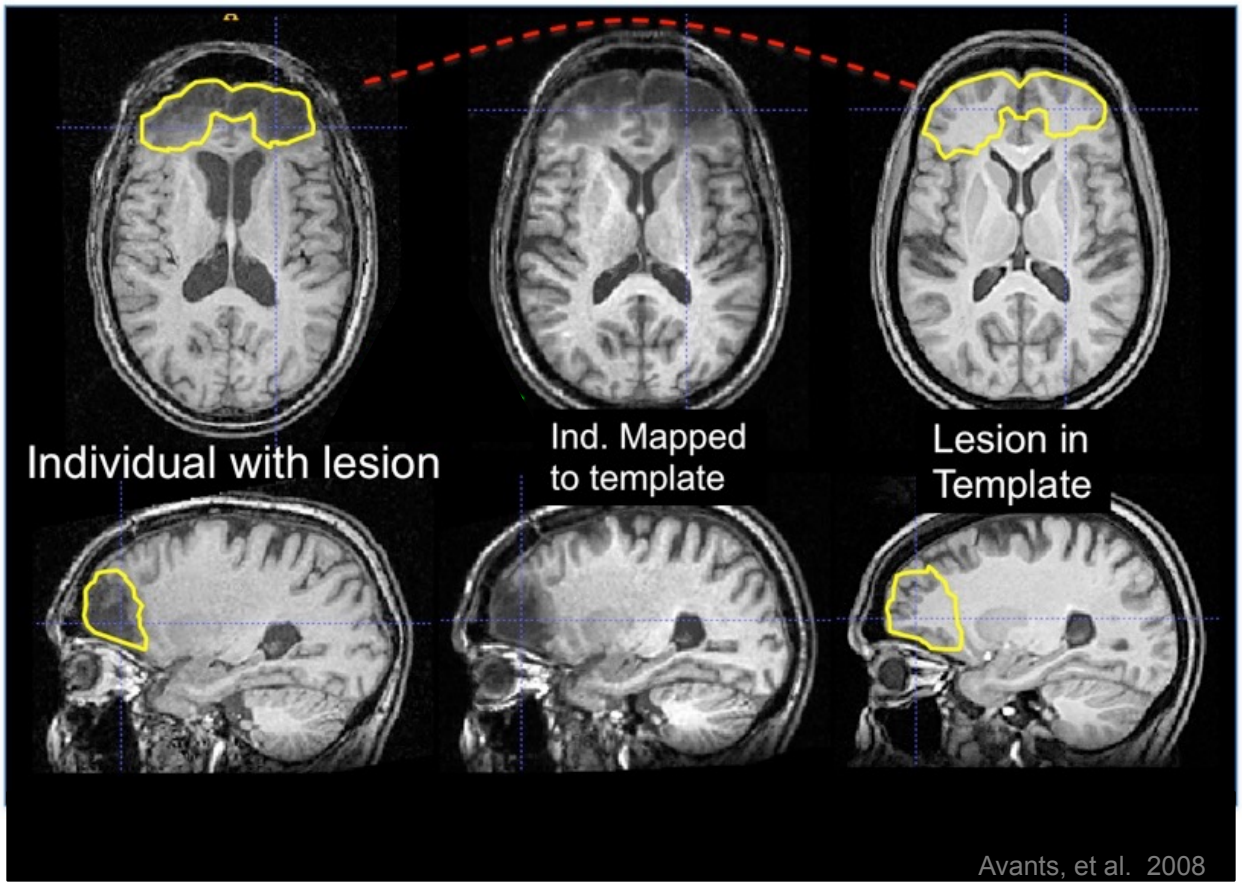
deformation-based morphometry



Jacobian matrix: representation of a differential, an $n \times n$ matrix of first order partial derivatives whose entry in the i-th row and j-th column is $\partial f_i / \partial x_j$. The Jacobian can be thought of as describing the amount of "stretching" that a transformation imposes.

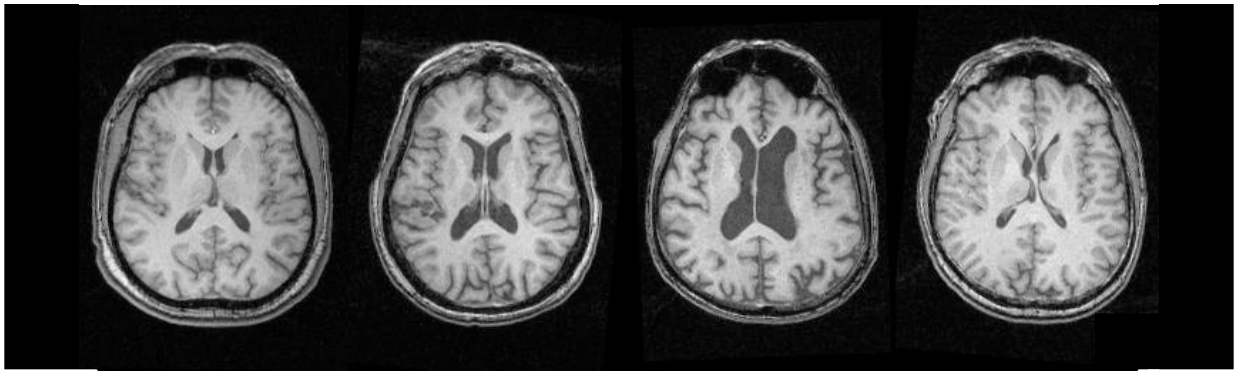
Registration application 2:

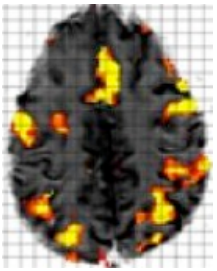
anatomical localization



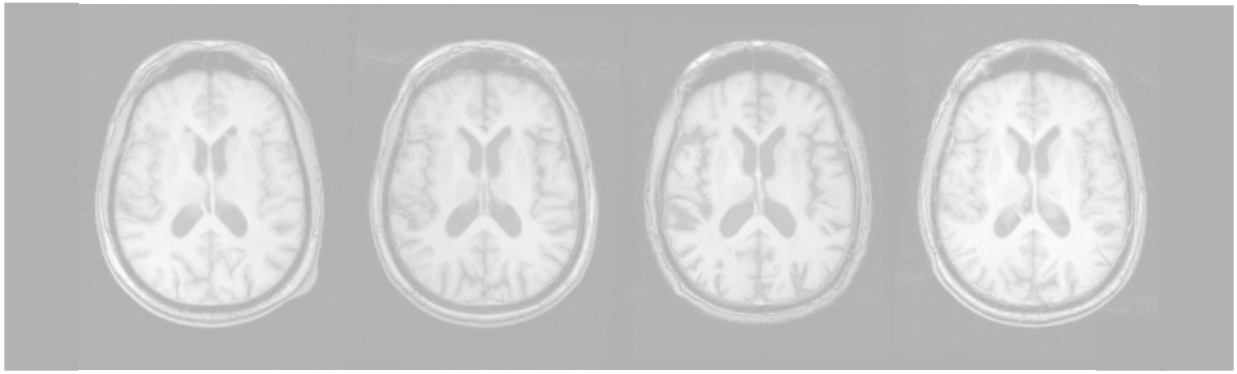
Registration application 3:

group analysis of data





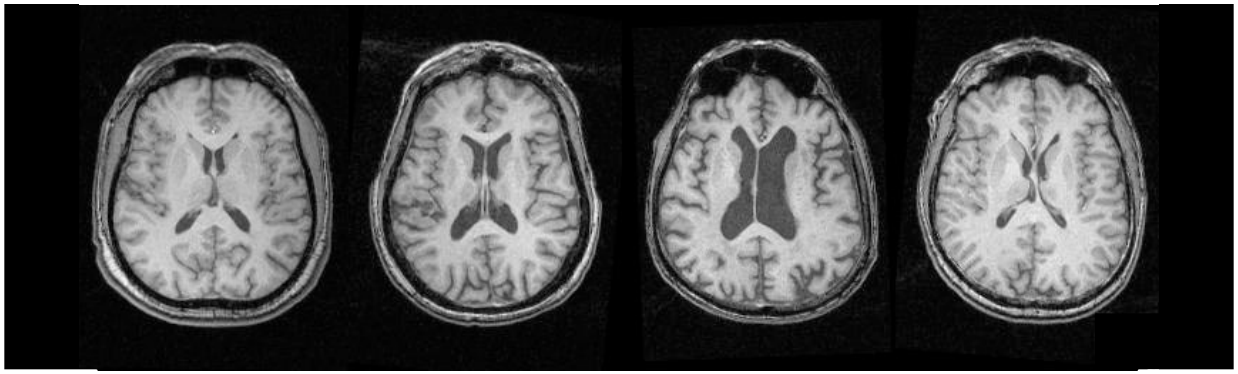
data (structure, function, physiology, landmarks) in individual space

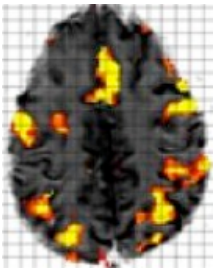


individuals normalized to template space

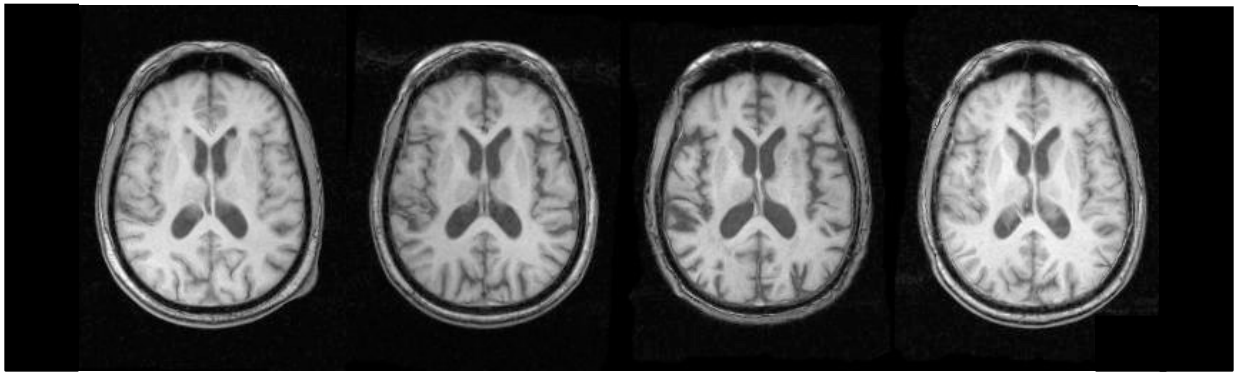
Registration application 3:

group analysis of data





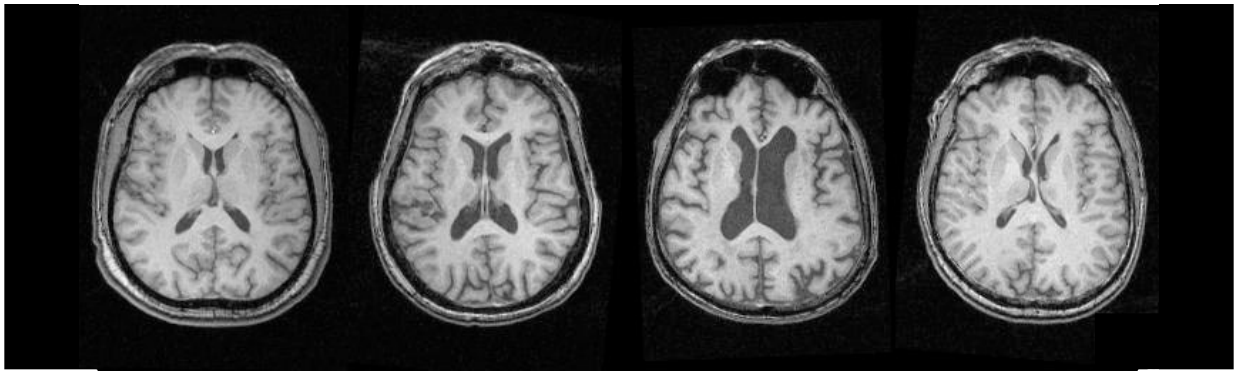
data (structure, function, physiology, landmarks) in individual space

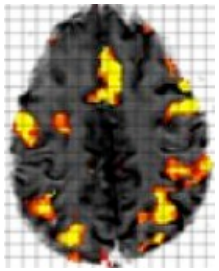


individuals normalized to template space

Registration application 3:

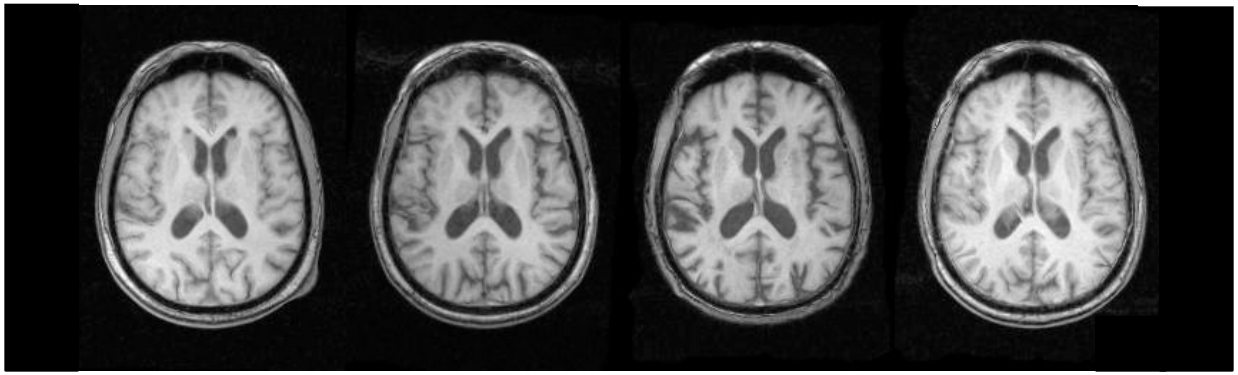
group analysis of data



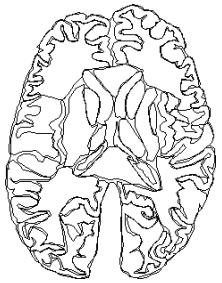


data (structure, function, physiology, landmarks) in individual space

template image



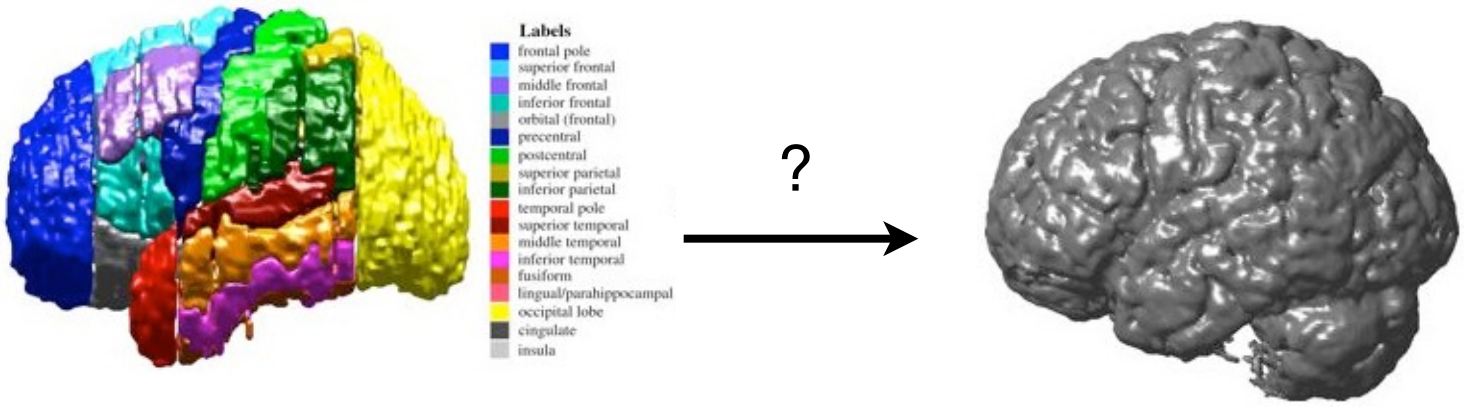
individuals normalized to template space



template labels

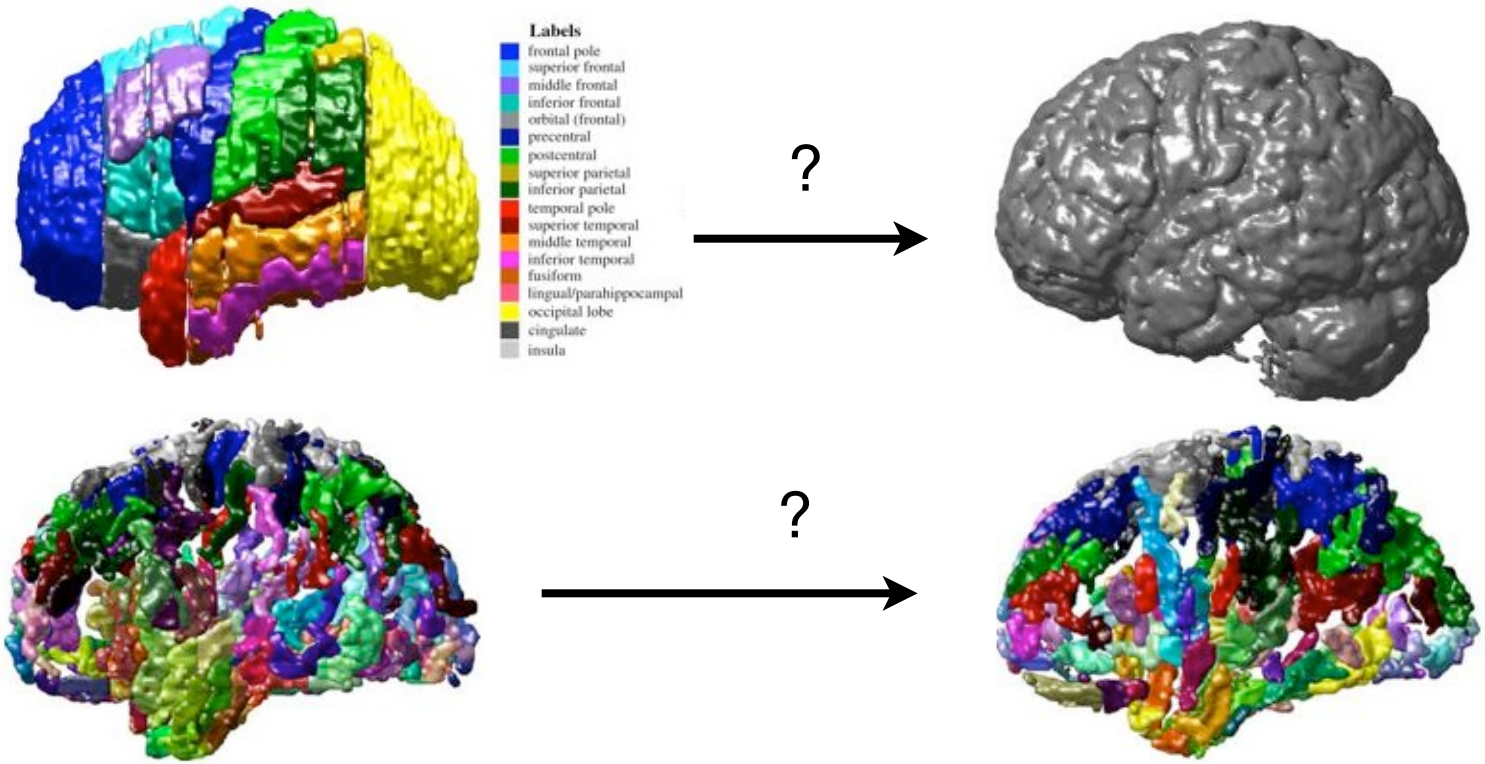
Registration application 4:

atlas-based anatomical labeling



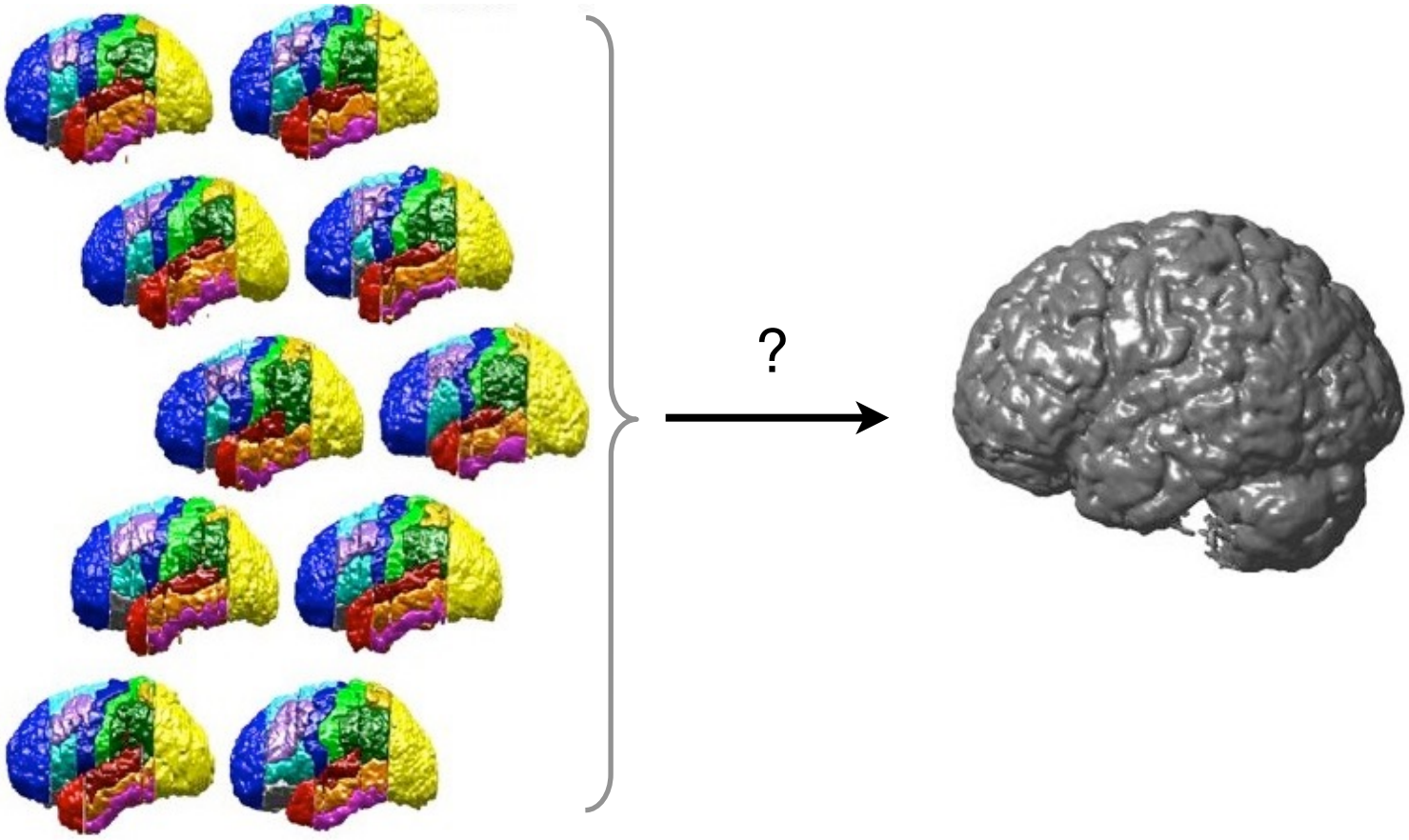
Registration application 4:

atlas-based anatomical labeling

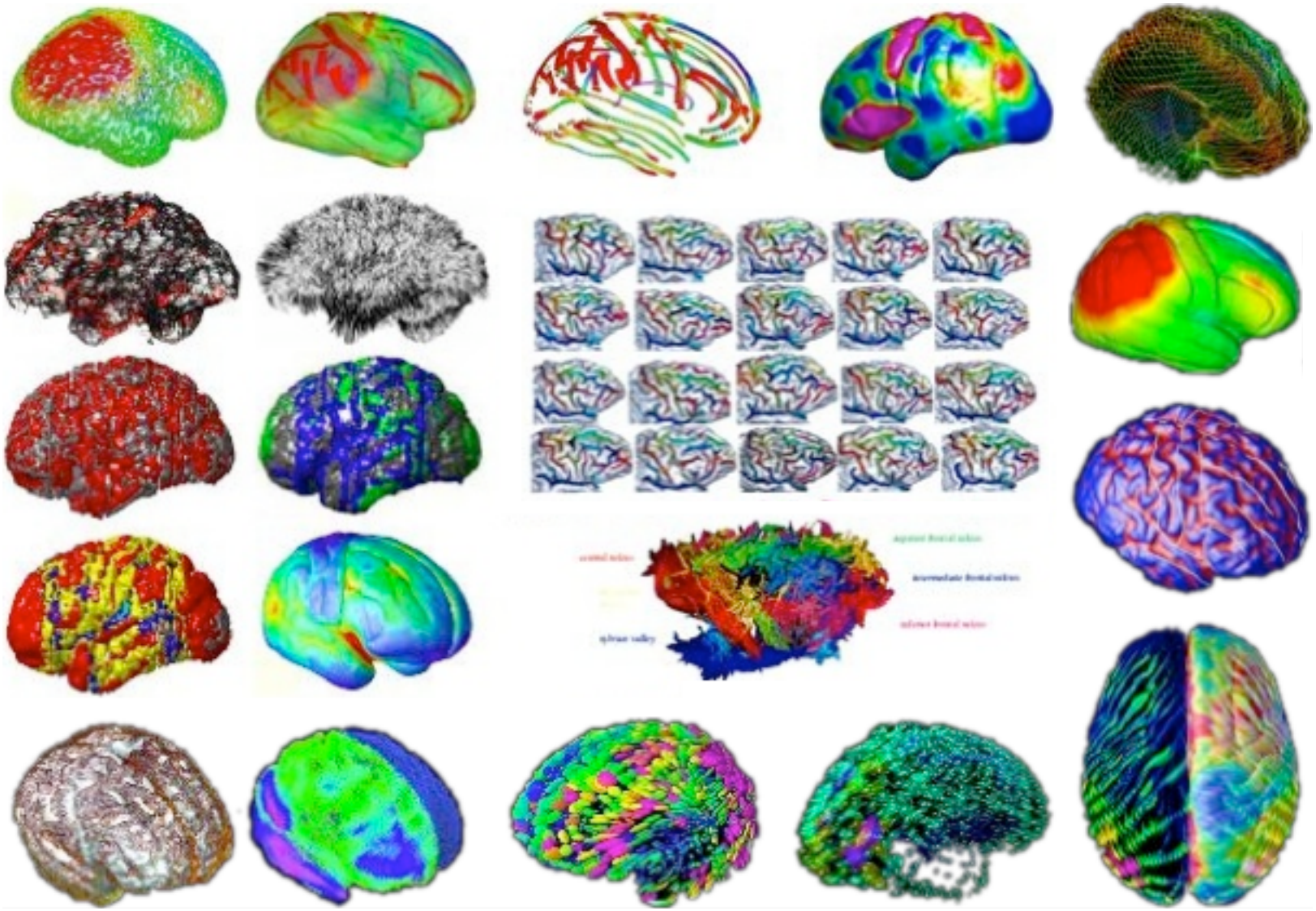


Registration application 4:

atlas-based anatomical labeling



Registration applications



What characterizes a registration algorithm?

Similarity metric: SSD, MSD, (n)CR, (n)CC, MI,...

Transformation: affine, piecewise linear, nonlinear,...

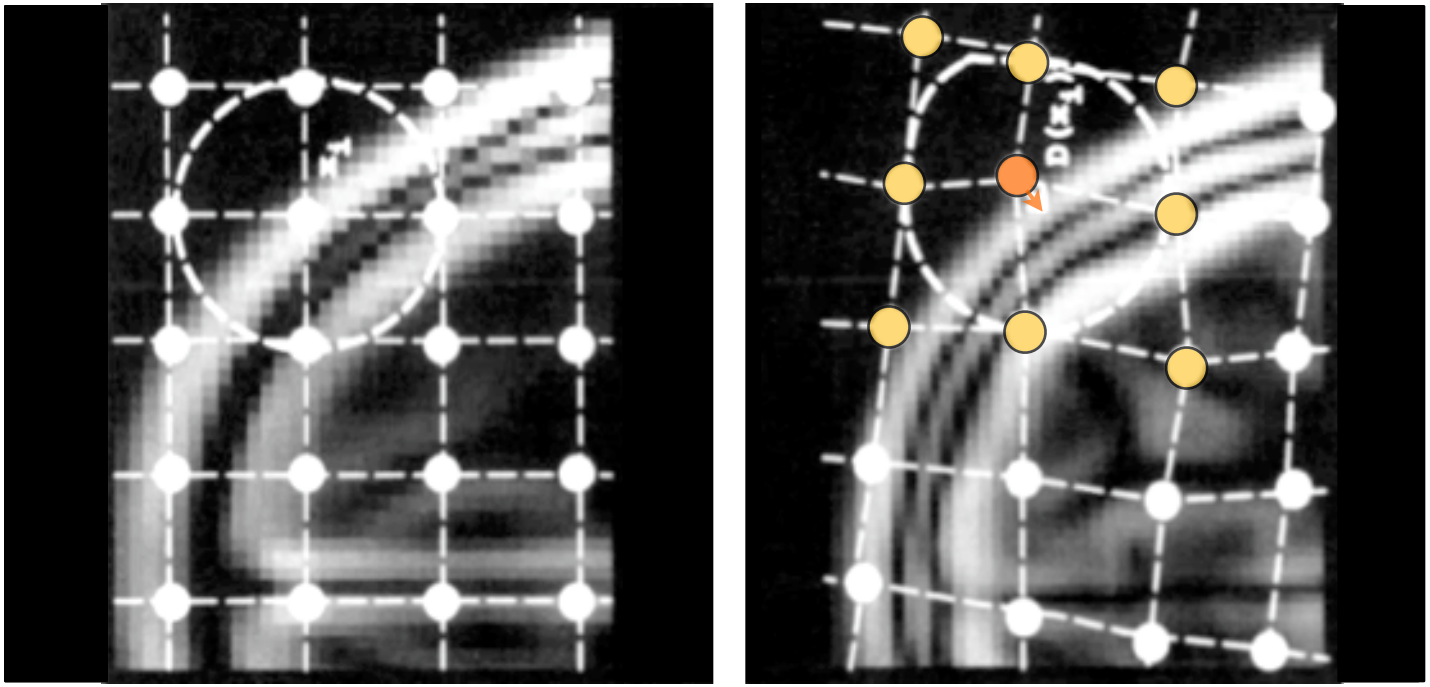
Regularization: multi-resolution/scale, Gaussian blur,...

Optimization: simplex, gradient descent,...

Interpolation: nearest-neighbor, trilinear, cubic, sinc,...

Nonlinear registration example

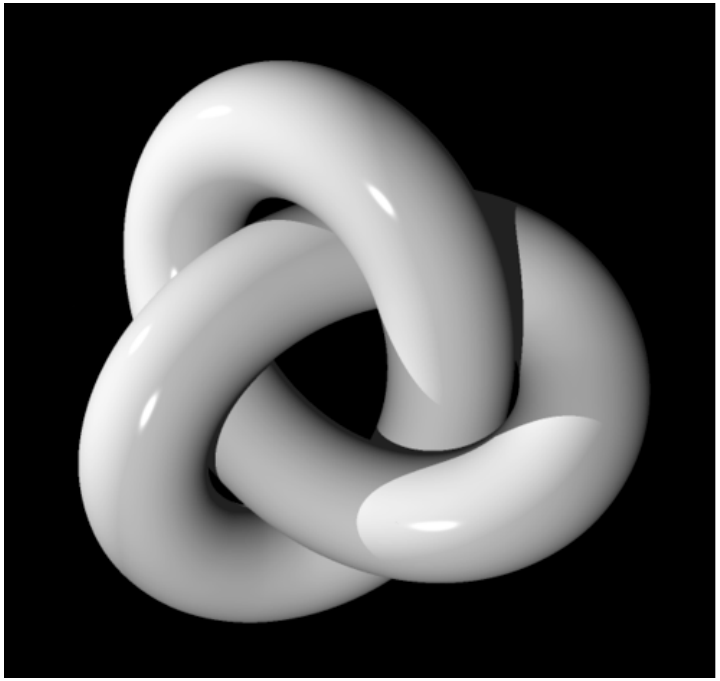
local translations, free-form deformations



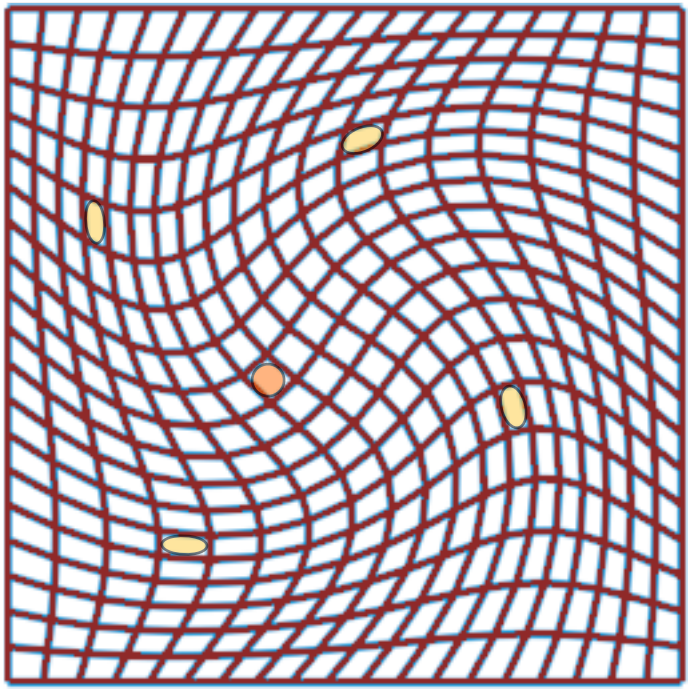
optimization of 3 translational parameters to maximize the neighborhood cross-correlation for each node (ANIMAL)

Nonlinear registration example

topological constraints: diffeomorphisms



homeomorphism: continuous mapping and inverse, but not deformation (trefoil knot, circle)



diffeomorphism: bijective map from manifold M to N and its inverse are differentiable

“Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration”

Participants:	16 from 11 institutions
Algorithms:	>14 software packages
Data:	80 brains manually labeled according to 4 different whole-brain labeling protocols (56 to 128 regions)
Registrations:	Each algorithm applied >2,168 times (>45,000 total)
Evaluation:	8 measures (overlap, distance, volume similarity)
Analysis:	3 independent methods (ranking and statistical tests)

http://www.mindboggle.info/papers/

“Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration”

Participants:	16 from 11 institutions
Algorithms:	>14 software packages
Data:	80 brains manually labeled according to 4 different whole-brain labeling protocols (56 to 128 regions)
Registrations:	Each algorithm applied >2,168 times (>45,000 total)
Evaluation:	8 measures (overlap, distance, volume similarity)
Analysis:	3 independent methods (ranking and statistical tests)

http://www.mindboggle.info/papers/

Participants

Location	Institution	Participants
UK	University of Oxford (FMRIB)	Mark Jenkinson
UK	University of Oxford (FMRIB)	Jesper Andersson
UK	Wellcome Trust Centre for Neuroimaging	John Ashburner
UK	Imperial College	Daniel Rueckert
France	INRIA R. / INSERM	Tom Vercauteren
France	Mauna Kea Technologies / INRIA S.A.	Pierre Hellier
Canada	Montreal Neurological Institute	Louis Collins, Claude Lepage
USA (CA)	UCLA (LONI)	M.-C. Chiang, P. Thompson
USA (CA)	UCLA	Roger P. Woods
USA (IA)	University of Iowa	Gary E. Christensen, J.H. Song
USA (PA)	University of Pennsylvania	Brian Avants, James Gee
USA (NY)	Nathan Kline Institute / NYU	Babak A. Ardekani

“Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration”

Participants:	16 from 11 institutions
Algorithms:	>14 software packages
Data:	80 brains manually labeled according to 4 different whole-brain labeling protocols (56 to 128 regions)
Registrations:	Each algorithm applied >2,168 times (>45,000 total)
Evaluation:	8 measures (overlap, distance, volume similarity)
Analysis:	3 independent methods (ranking and statistical tests)

http://www.mindboggle.info/papers/

Algorithms

Software	Similarity metric	Transformation	DOF
FLIRT (linear)	nCR	linear, rigid-body	9, 6
AIR	MSD	5th-order polynomial warps	168
ANIMAL	CC	local translations	69,000
ART	nCC	FFD based on cubic splines (H,	7,000,000
Diffeomorphic	SSD	np)	21,000,00
Demons	SSD	displacement field (D, np)	0
FNIRT	nMI	cubic B-splines	30,000
IRTK	Jensen-Rényi divergence	cubic B-splines	1,400,000
JRD-fluid	displaced frame difference	viscous fluid; variational calculus	2,000,000
ROMEO	SSD	(D)	2,000,000
SICLE	CC	local affine	8,000
SyN	MSD	3-D Fourier series (D)	28,000,00
SPM5 SPM2-type	MSD	bi-directional diffeomorphism (D)	0
Normalize	generative segmentation	discrete cosine transforms	1,000
Unified	model	discrete cosine transforms	1,000
Segment	multinomial model:	discrete cosine transforms	1,000
DARTEL toolbox	congealing	FDM of viscosity field (Dc)	6,400,000

n = normalized
CC = cross-correlation
CR = correlation ratio
MI = mutual information
MSD = mean of squared
SSD differences
= sum of squared
differences

D = diffeomorphic
Dc = diffeomorphic, constant over
FDM time
FFD = finite difference model
H = free-form deformation
np = homeomorphic
= nonparametric

Algorithms

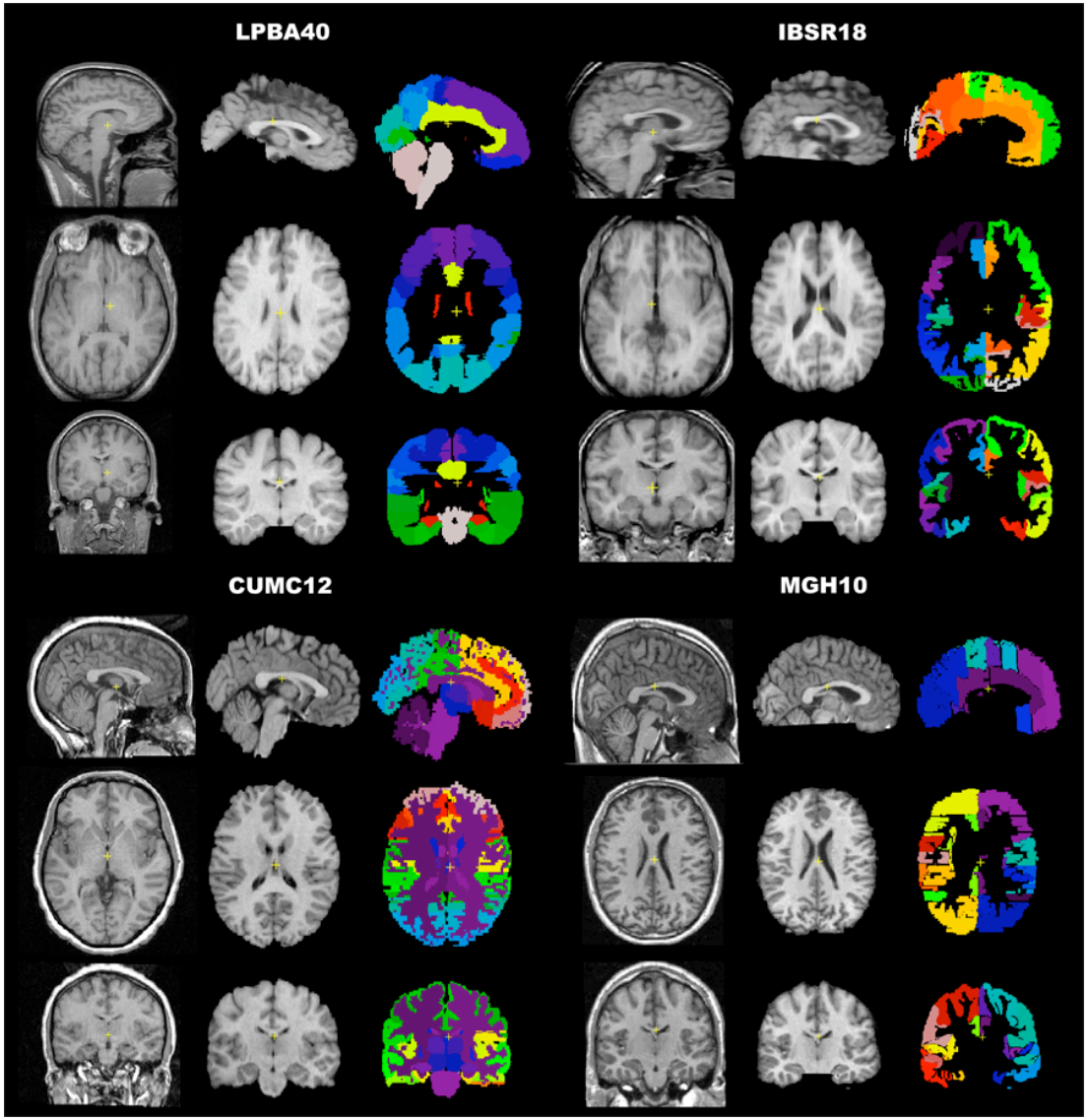
Software	Regularization	Multi-resolution
FLIRT (linear)		
AIR	increase of polynomial order	sparse to fine voxel sampling
ANIMAL	stiffness parameter at each node	local Gaussian
ART		median- & low-pass Gaussian
Diffeomorphic		Gaussian
Demons	membrane energy	down- to up-sampling; # basis
FNIRT	none used in study	components
IRTK	compressible viscous fluid	control mesh
JRD-fluid	explicit method; brightness	multi-resolution
ROMEO	constancy	adaptive multigrid (octree); Gaussian
SICLE	linear elasticity; inverse	# basis components
SyN	consistency	xfm symmetry; Gaussian of velocity field
SPM2-type	transformation symmetry	
Normalize	bending energy; basis cutoff	
Unified	bending energy; basis cutoff	
Segment	bending energy; basis cutoff	full multigrid (recursive)
DARTEL toolbox	linear elasticity	

“Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration”

Participants:	16 from 11 institutions
Algorithms:	>14 software packages
Data:	80 brains manually labeled according to 4 different whole-brain labeling protocols (56 to 128 regions)
Registrations:	Each algorithm applied >2,168 times (>45,000 total)
Evaluation:	8 measures (overlap, distance, volume similarity)
Analysis:	3 independent methods (ranking and statistical tests)

http://www.mindboggle.info/papers/

Data

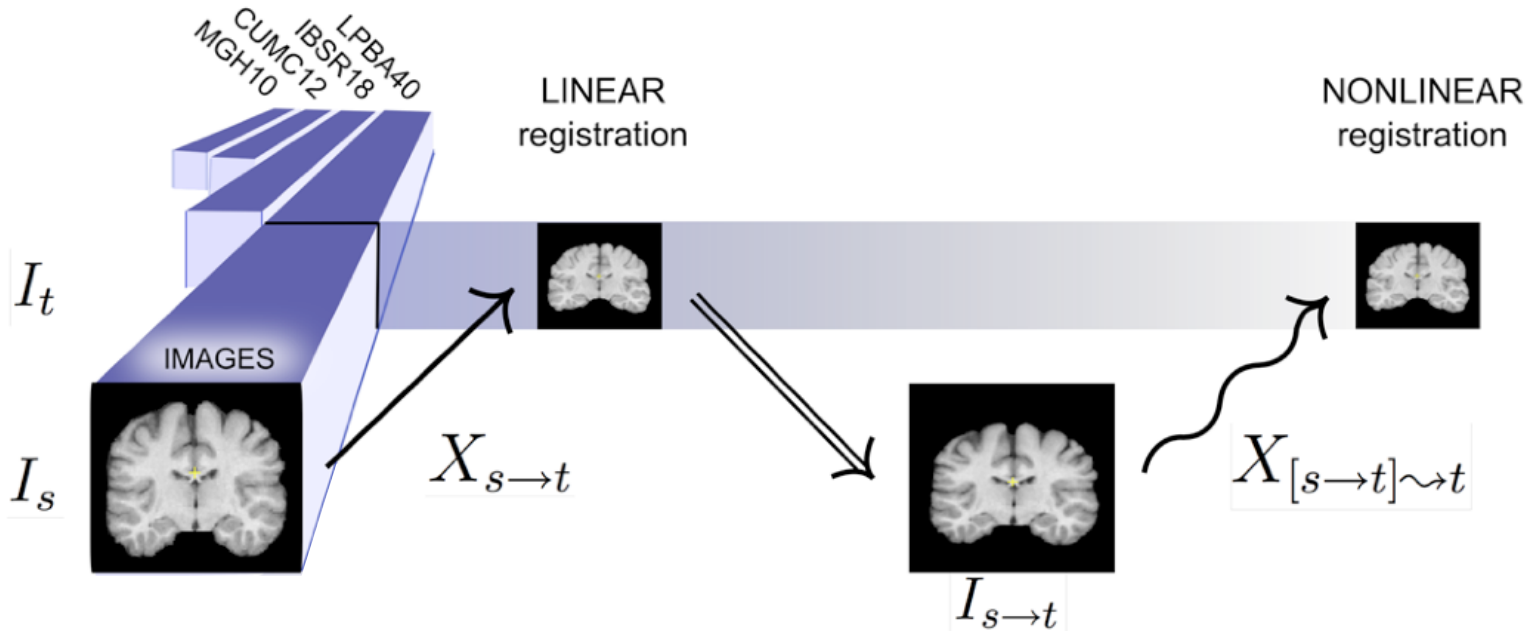


“Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration”

Participants:	16 from 11 institutions
Algorithms:	>14 software packages
Data:	80 brains manually labeled according to 4 different whole-brain labeling protocols (56 to 128 regions)
Registrations:	Each algorithm applied >2,168 times (>45,000 total)
Evaluation:	8 measures (overlap, distance, volume similarity)
Analysis:	3 independent methods (ranking and statistical tests)

http://www.mindboggle.info/papers/

Registrations



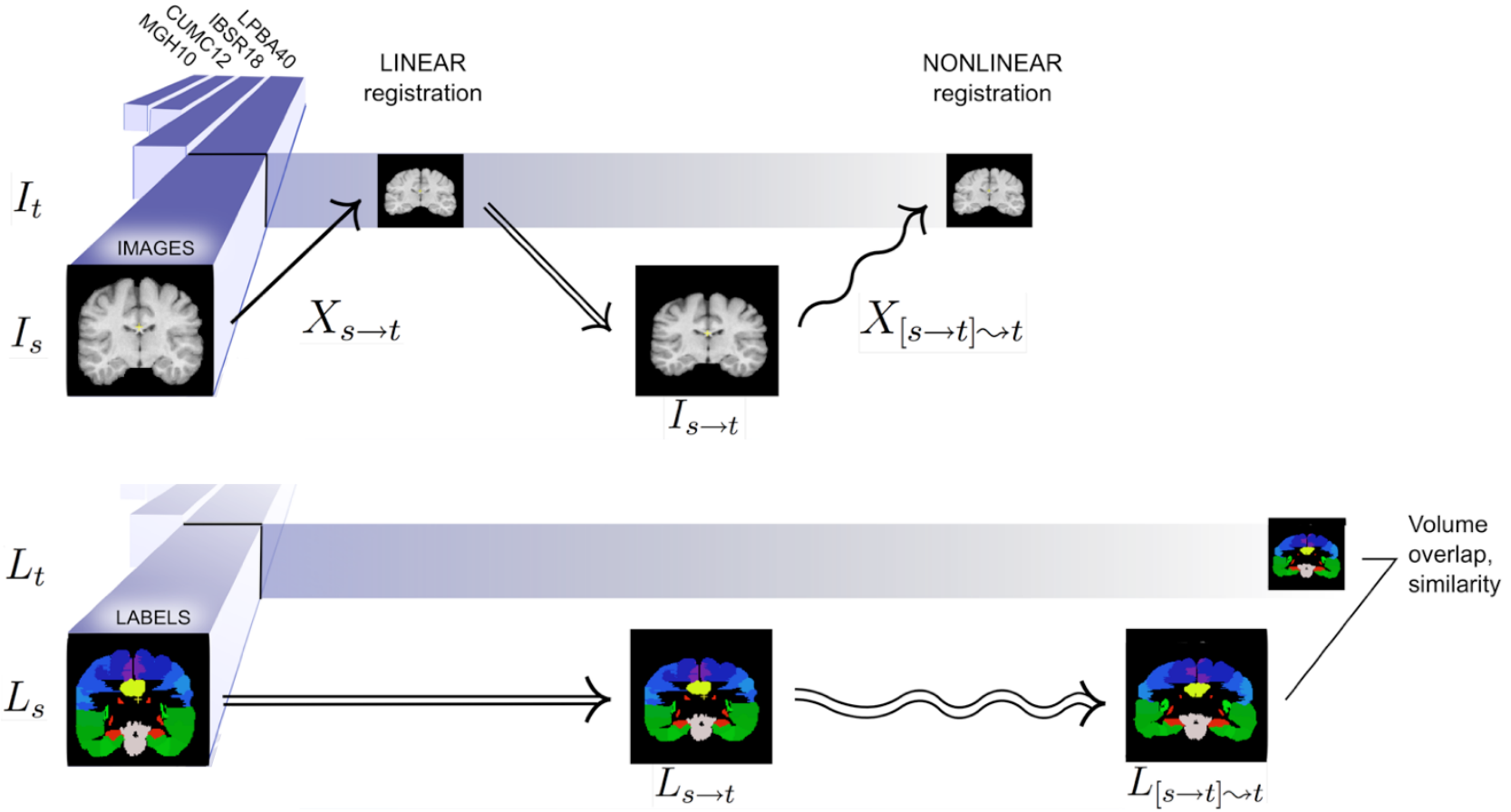
“Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration”

Participants:	16 from 11 institutions
Algorithms:	>14 software packages
Data:	80 brains manually labeled according to 4 different whole-brain labeling protocols (56 to 128 regions)
Registrations:	Each algorithm applied >2,168 times (>45,000 total)
Evaluation:	8 measures (overlap, distance, volume similarity)
Analysis:	3 independent methods (ranking and statistical tests)

http://www.mindboggle.info/papers/

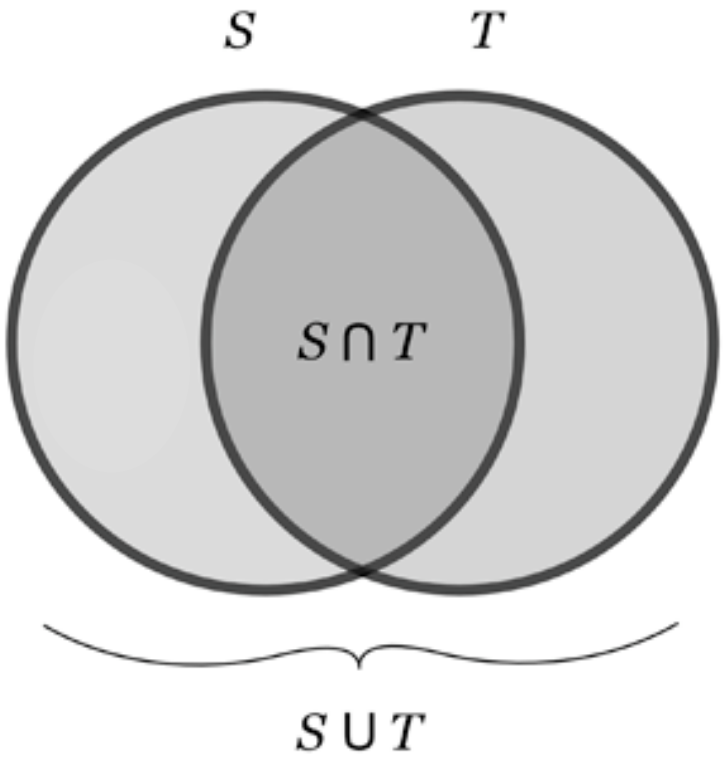
Evaluation

overlap



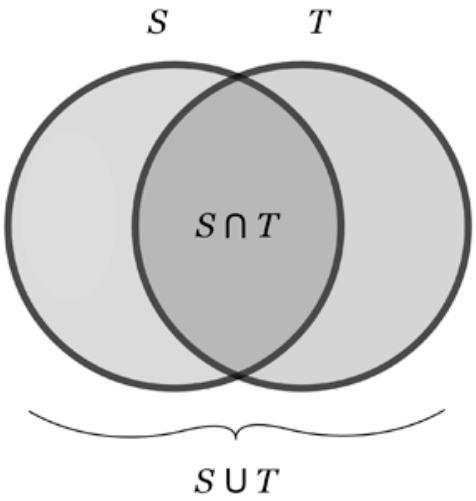
Evaluation

overlap



Evaluation

overlap



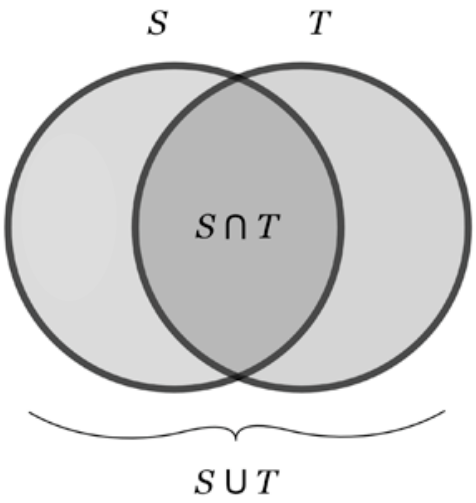
Target overlap: $TO = \frac{\sum_r |S_r \cap T_r|}{\sum_r |T_r|}$

Union overlap: $UO = \frac{\sum_r |S_r \cap T_r|}{\sum_r |S_r \cup T_r|}$

Mean overlap: $MO = 2\frac{\sum_r |S_r \cap T_r|}{\sum_r (|S_r| + |T_r|)}$

Evaluation

overlap and volume similarity

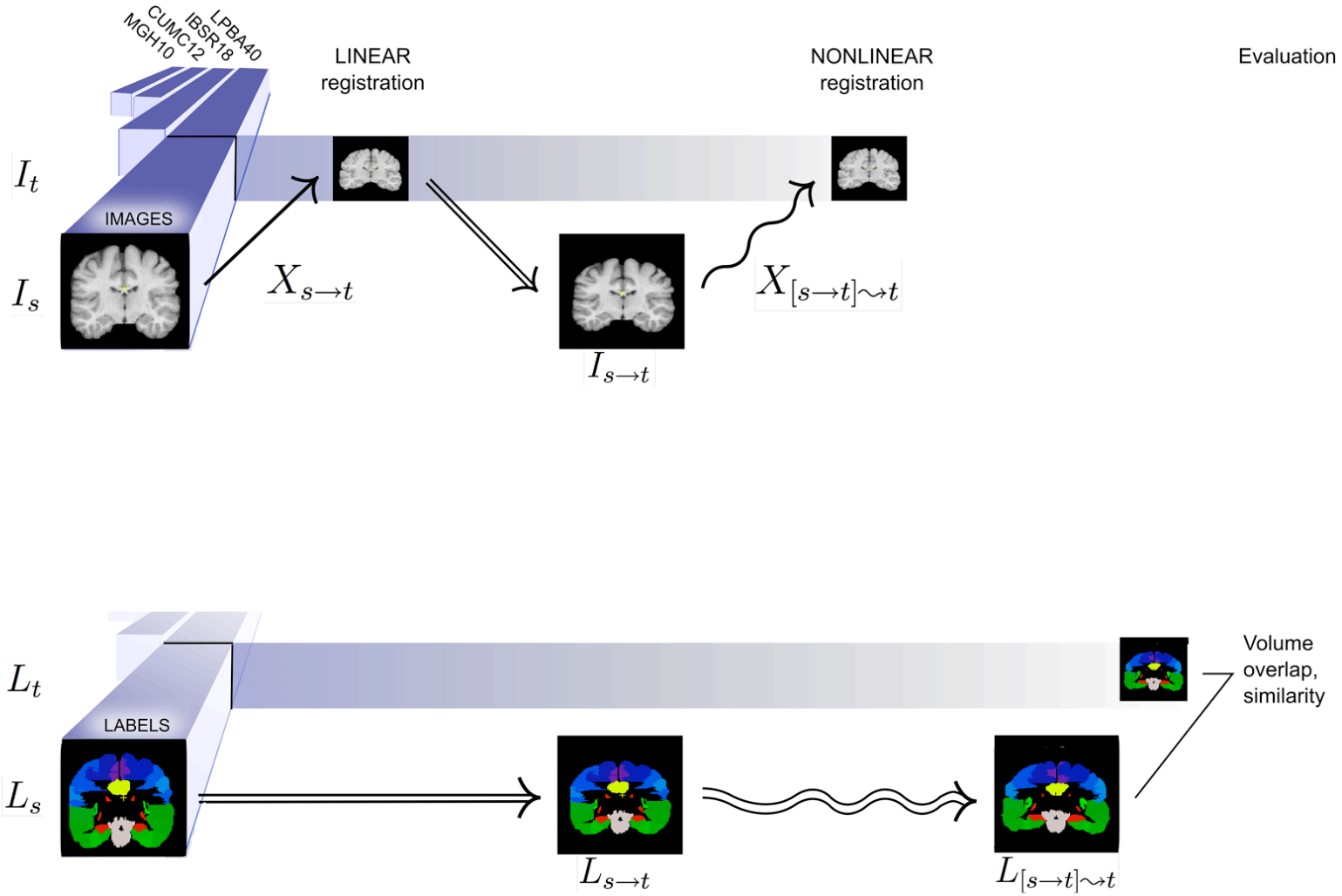


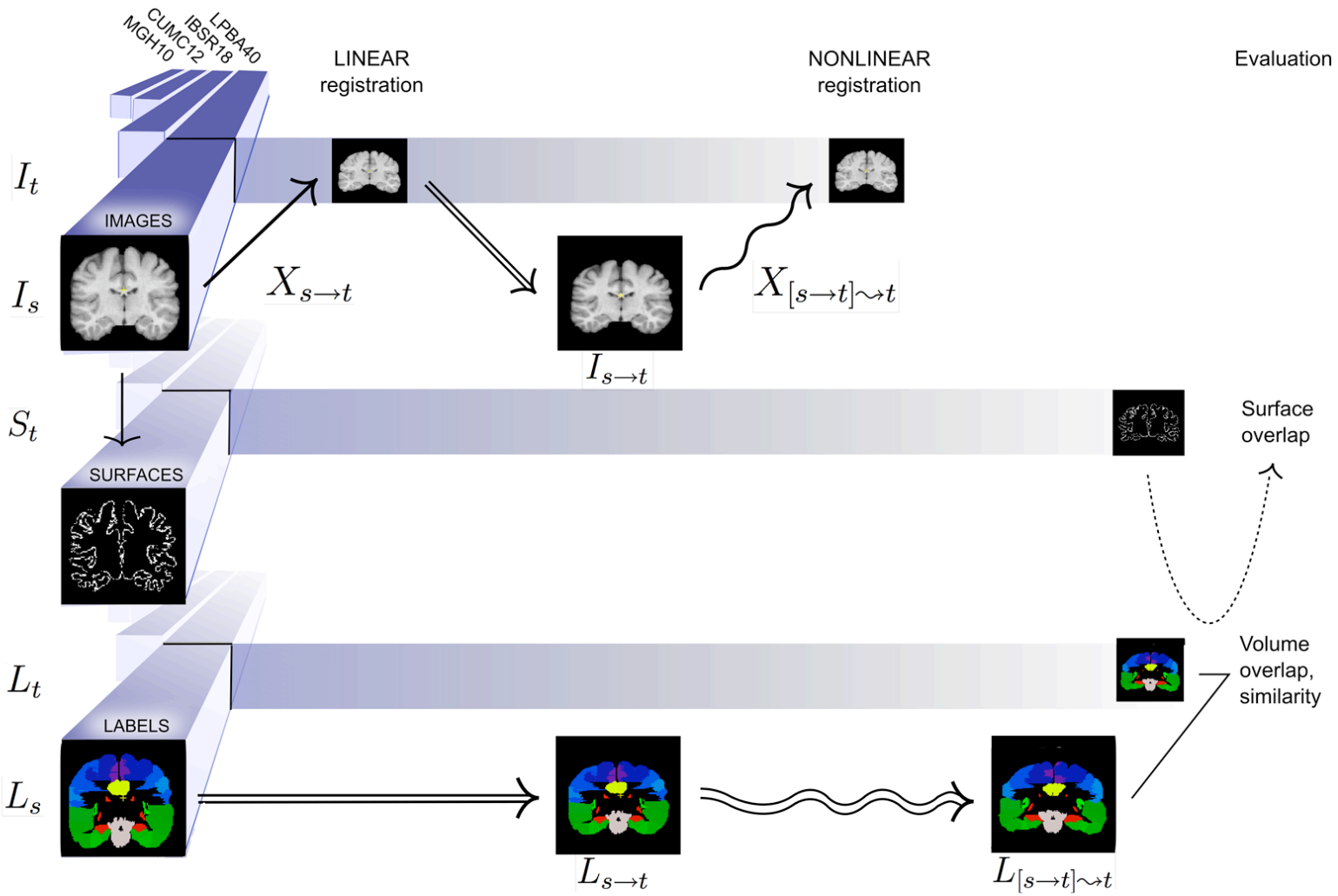
Target overlap: $TO = \frac{\sum_r |S_r \cap T_r|}{\sum_r |T_r|}$

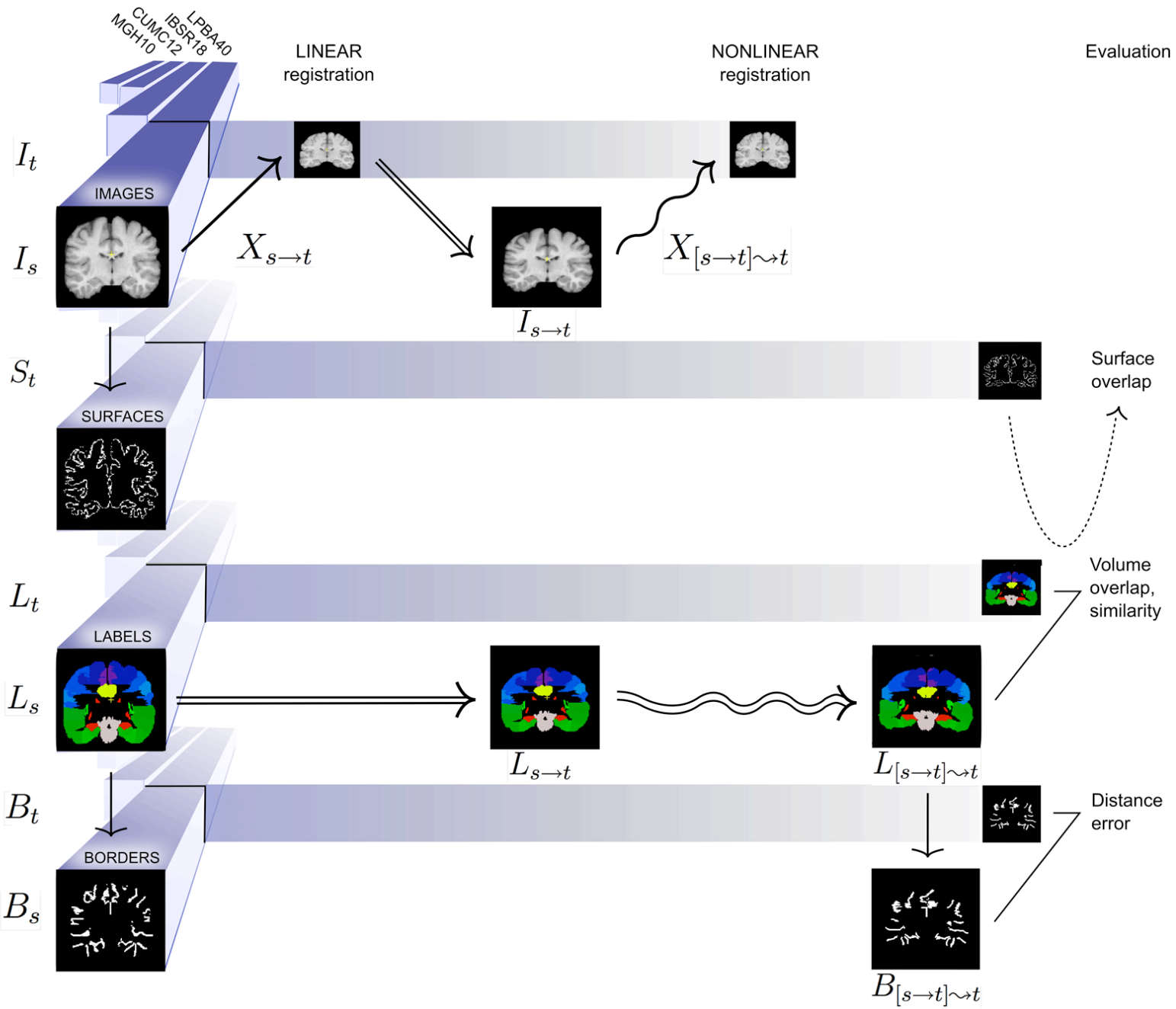
Union overlap: $UO = \frac{\sum_r |S_r \cap T_r|}{\sum_r |S_r \cup T_r|}$

Mean overlap: $MO = 2\frac{\sum_r |S_r \cap T_r|}{\sum_r (|S_r| + |T_r|)}$

Volume similarity: $VS = 2\frac{\sum_r (|S_r| - |T_r|)}{\sum_r (|S_r| + |T_r|)}$


LPBA40
IBSR18
CUMC12
MGH10
LINEAR
registration
NONLINEAR
registration
Evaluation
I_t
IMAGES
I_s
$X_{s\to t}$
$I_{s\to t}$
$X_{[s\to t]\rightsquigarrow t}$
L_t
LABELS
L_s
$L_{s\to t}$
$L_{[s\to t]\rightsquigarrow t}$
Volume
overlap,
similarity


LPBA40
IBSR18
CUMC12
MGH10
LINEAR
registration
NONLINEAR
registration
Evaluation
I_t
IMAGES
I_s
$X_{s\to t}$
$I_{s\to t}$
$X_{[s\to t]\rightsquigarrow t}$
S_t
SURFACES
Surface
overlap
L_t
LABELS
L_s
Volume
overlap,
similarity
$L_{s\to t}$
$L_{[s\to t]\rightsquigarrow t}$


LPBA40
IBSR18
CUMC12
MGH10
LINEAR registration
NONLINEAR registration
Evaluation
I_t
IMAGES
I_s
$X_{s\to t}$
$I_{s\to t}$
$X_{[s\to t]\leadsto t}$
S_t
SURFACES
Surface overlap
L_t
LABELS
L_s
$L_{s\to t}$
$L_{[s\to t]\leadsto t}$
Volume overlap, similarity
B_t
BORDERS
B_s
Distance error
$B_{[s\to t]\leadsto t}$

“Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration”

Participants:	16 from 11 institutions
Algorithms:	>14 software packages
Data:	80 brains manually labeled according to 4 different whole-brain labeling protocols (56 to 128 regions)
Registrations:	Each algorithm applied >2,168 times (>45,000 total)
Evaluation:	8 measures (overlap, distance, volume similarity)
Analysis:	3 independent methods (ranking and statistical tests)

http://www.mindboggle.info/papers/

Analysis

Statistical tests run on small sets of independent samples:

Permutation tests

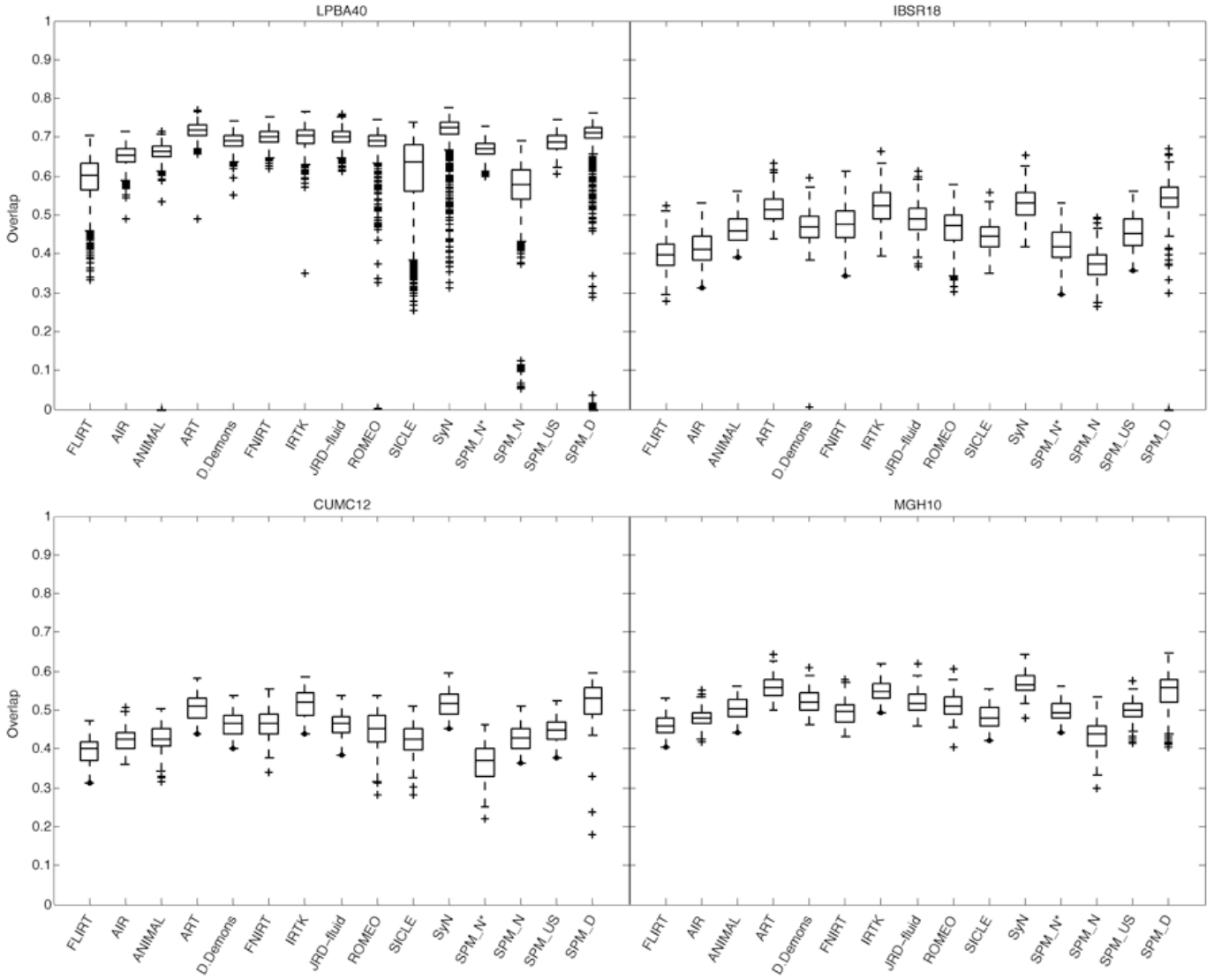
One-way ANOVA tests

Ranking based on practical significance (volume offset):

Indifference-zone ranking

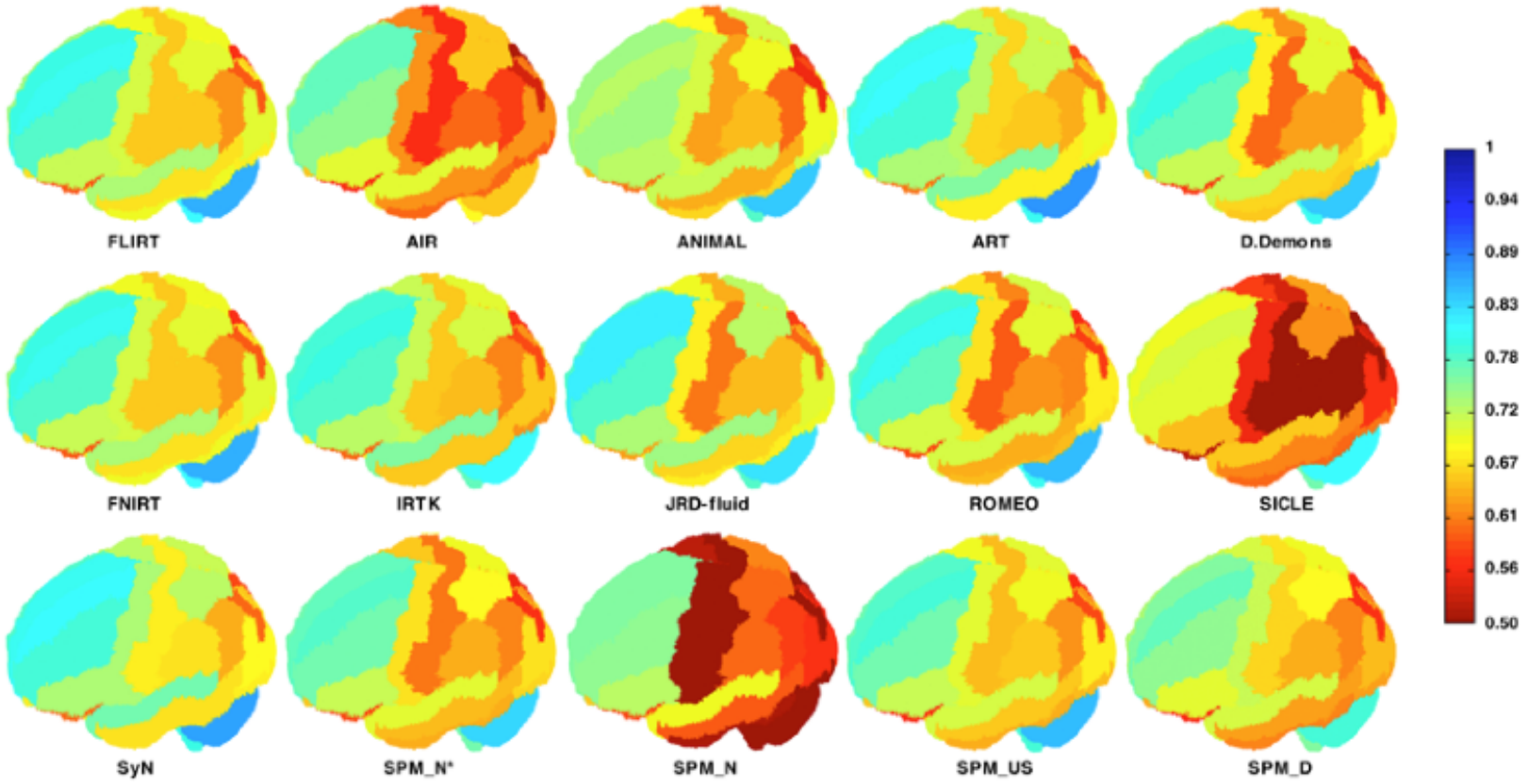
Results

target overlap averaged across all registrations and all regions



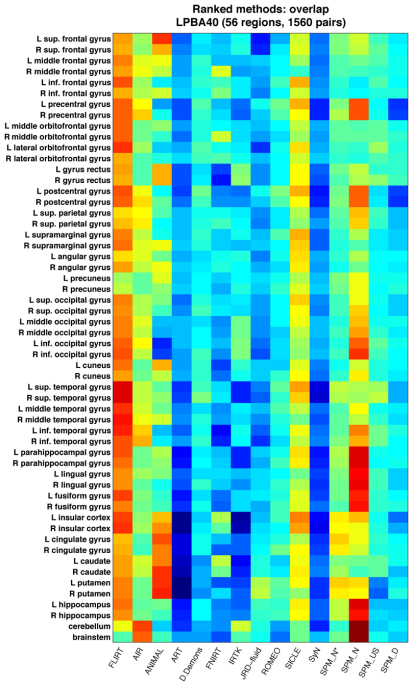
Results

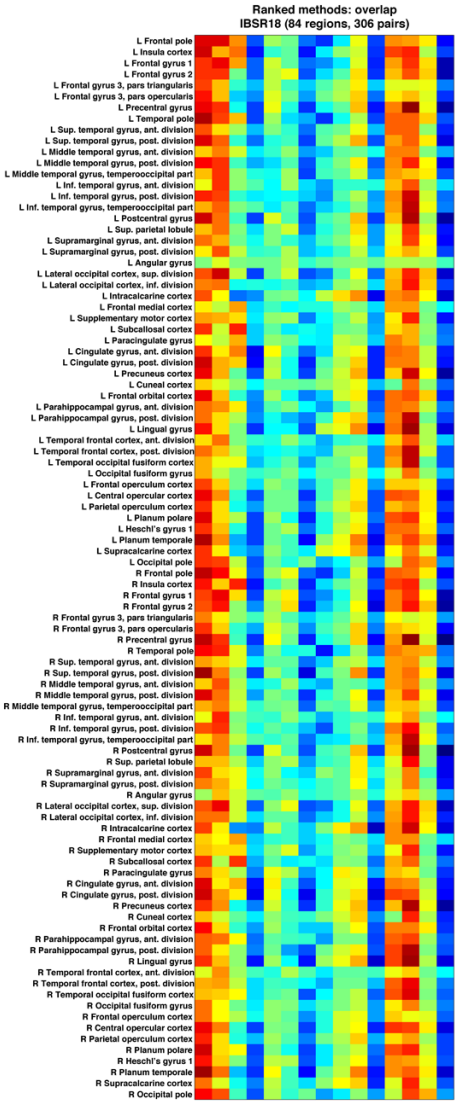
target overlap averaged across all registrations (LPBA40)

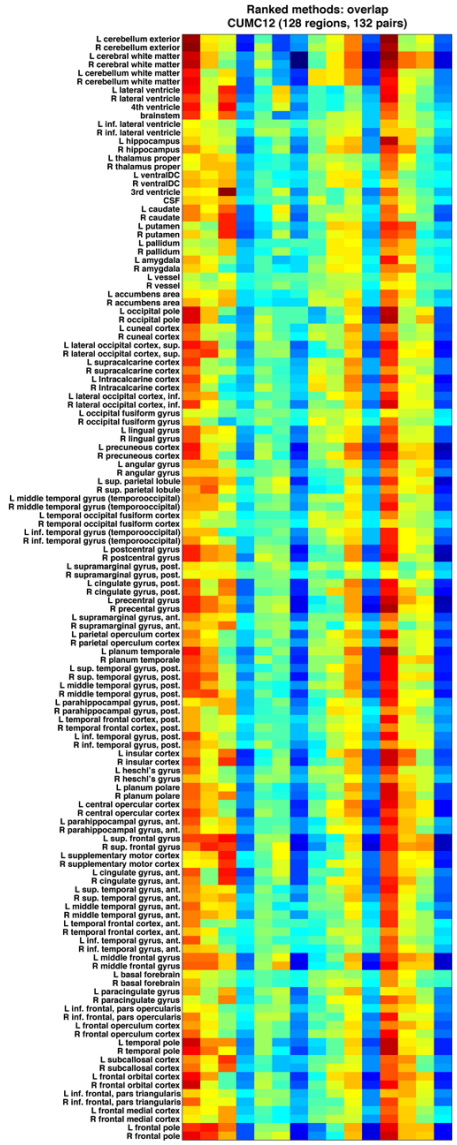


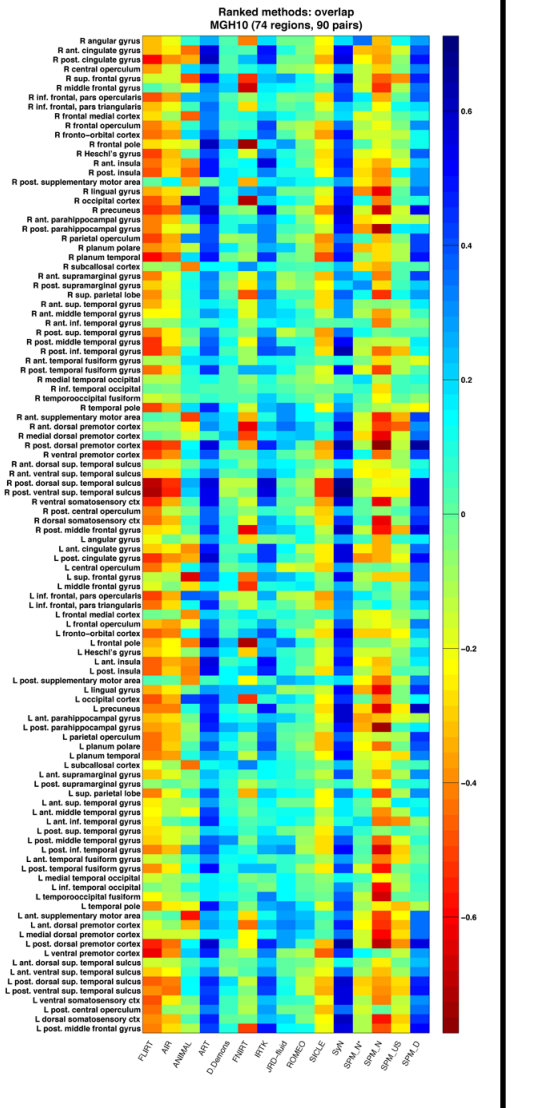
SPM_N*="SPM2-type" Normalize
SPM_N=Normalize
SPM_US=Unified Segmentation
SPM_D=DARTEL pairwise

Results: Indifference-zone ranking







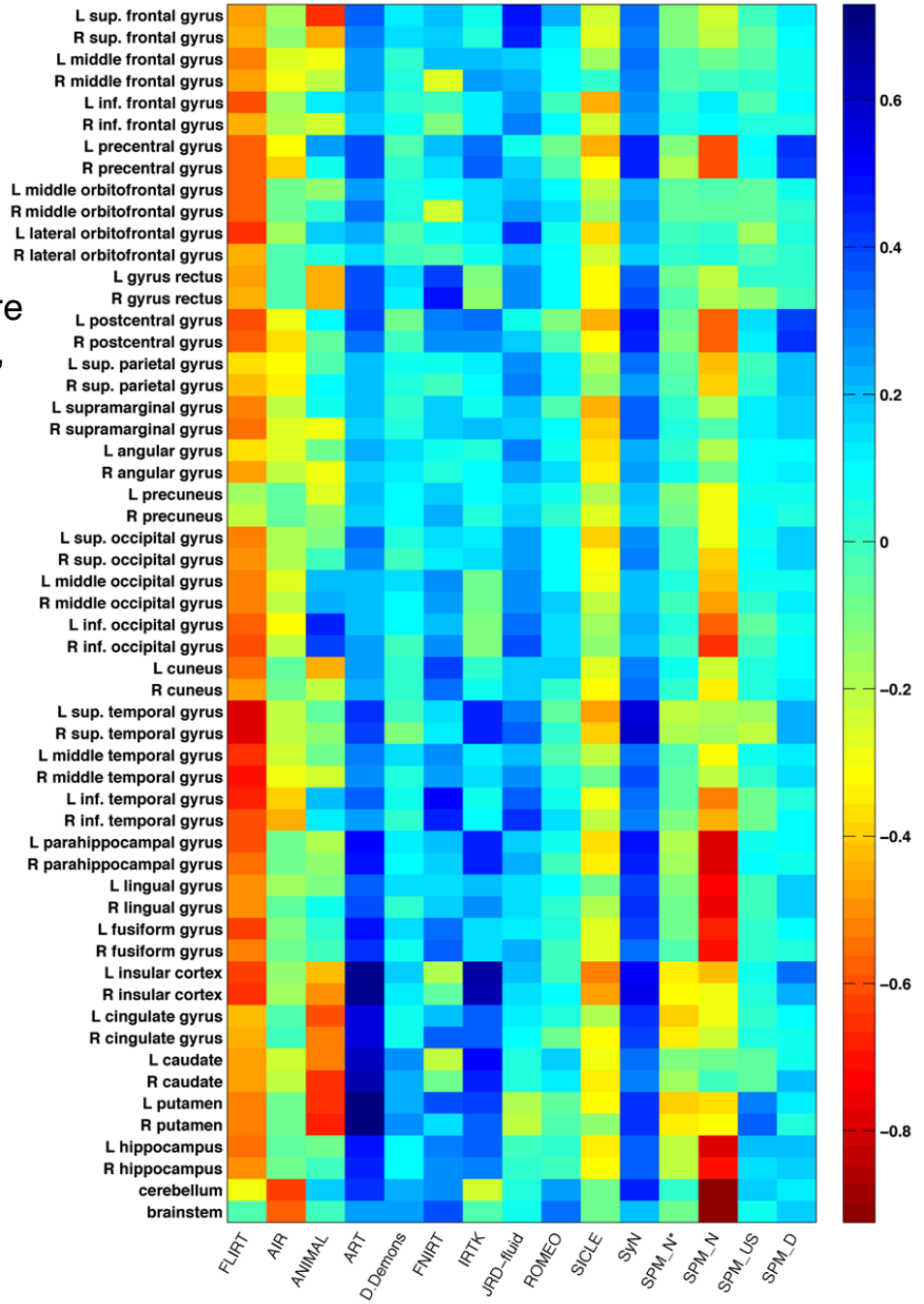


Indifference-zone ranking

Each color cell represents the average score for a given region and for a given algorithm, averaged over 1,560 LPBA40 registrations.

Blue indicates higher accuracy.

The scores are {–1, 0, 1} values indicating the pairwise performance of the algorithm relative to each of the other methods, according to target volume overlap.



SPM_N*=“SPM2-type” Normalize
SPM_N=Normalize
SPM_US=Unified Segmentation
SPM_D=DARTEL pairwise

Indifference-zone ranking

	LPBA40	μ (SD)	IBSR18	μ (SD)	CUMC12	μ (SD)	MGH10	μ (SD)
rank 1	ART	.35 (.07)	SPM_D	.50 (.19)	SPM_D	.47 (.17)	SyN	.39 (.06)
	SyN	.34 (.24)	SyN	.40 (.12)	IRTK	.42 (.07)	ART	.36 (.07)
			IRTK	.35 (.15)	SyN	.41 (.06)		
			ART	.33 (.08)	ART	.35 (.05)		
2			JRD-fluid	.18 (.13)				
3	JRD-fluid	.20 (.08)	FNIRT	.06 (.11)	JRD-fluid	.07 (.07)	IRTK	.26 (.07)
	IRTK	.18 (.15)	D.Demons	.01 (.08)	FNIRT	.07 (.09)	SPM_D	.25 (.28)
	FNIRT	.17 (.08)	ROMEO	.01 (.28)	D.Demons	.05 (.05)		
	SPM_D	.14 (.31)						

Indifference-zone ranking of the registration methods by label set. This table lists the methods that attained the top three ranks after averaging scores across all brain regions then across all registration pairs (μ=mean, SD=standard deviation). The scores reflect a pairwise comparison between methods, according to target overlap Methods within ranks 1, 2, and 3 have positive means lying within one, two, and three standard deviations of the highest mean, respectively. Values are not comparable across label sets (columns). (SPM_D=DARTEL pairwise)

Permutation ranking

	LPBA40	μ (SD)	IBSR18	μ (SD)	CUMC12	μ (SD)	MGH10	μ (SD)
rank 1	ART	.82 (.35)	SPM_D	.83 (.27)	SPM_D	.76 (.24)	SyN	.77 (.37)
	SyN	.60 (.38)	SyN	.72 (.51)	SyN	.74 (.51)	ART	.72 (.45)
	FNIRT	.49 (.66)	IRTK	.67 (.53)	IRTK	.74 (.50)	IRTK	.61 (.51)
	JRD-fluid	.49 (.66)	ART	.60 (.70)	ART	.60 (.70)		
2	IRTK	.43 (.63)	JRD-fluid	.30 (.82)			SPM_D	.27 (.23)
	D.Demons	.13 (.82)					D.Demons	.27 (.69)
	SPM_US	.11 (.83)					JRD-fluid	.24 (.66)
							ROMEO	.06 (.63)
3	ROMEO	.08 (.73)	FNIRT	.16 (.82)	D.Demons	.20 (.84)		
	SPM_D	.07 (.29)	D.Demons	.05 (.84)	FNIRT	.18 (.81)		
					JRD-fluid	.17 (.81)		

Permutation test ranking of the registration methods by label set. This table lists the methods that attained the top three ranks after conducting permutation tests between mean target overlaps (averaged across regions) for each pair of methods, then calculating the percentage of p-values less than or equal to 0.05 (of 100,000 tests for CUMC12 and MGH10 or of 10,000 tests for LPBA40 and IBSR18; μ=mean, SD=standard deviation). Methods within ranks 1, 2, and 3 have positive mean percentages lying within one, two, and three standard deviations of the highest mean, respectively. Values are not comparable across label sets (columns). (SPM_D=DARTEL pairwise)

Average ranks

Algorithm	mean rank	dof	run time: minutes	year
SyN	1.00	28M	77 (15.1)	2008
ART	1.00	7M	20.1 (1.6) [Linux]	2005
IRTK	1.63	1.4M	120.8 (29.3)	1999
SPM5 DARTEL Toolbox	1.88	6.4M	71.8 (6.3)	2007
JRD-fluid	2.50	2M	17.1 (1.0) [Solaris]	2007
Diffeomorphic Demons	3.00	21M	8.7 (1.2)	2007
FNIRT	3.00	30K	29.1 (6.0)	2008
ROMEO	3.50	2M	7.5 (0.5)	2001
ANIMAL		69K	11.2 (0.4)	1994
SICLE		8K	33.5 (6.6)	1999
SPM5 Unified Segmentation		1K	$\simeq$1	2005
"SPM2-type" Normalize		1K	$\simeq$1	1999
SPM5 Normalize		1K	$\simeq$1	1999
AIR		168	6.7 (1.5)	1998

References

To download this lecture:
http://www.mindboggle.info/lectures/

Evaluation study's website:
http://www.mindboggle.info/papers/

Recent review of brain registration algorithms:
"Brain functional localization: a survey of image registration techniques"
IEEE Trans Med Imaging. 2007. 26(4): 427-451. (330 references!)

Registration examples:
http://www.picsl.upenn.edu/ANTS/
ANIMAL: http://dx.doi.org/10.1002/hbm.460030304 and chapter

Mathematics definitions, images:
http://www.wikipedia.org

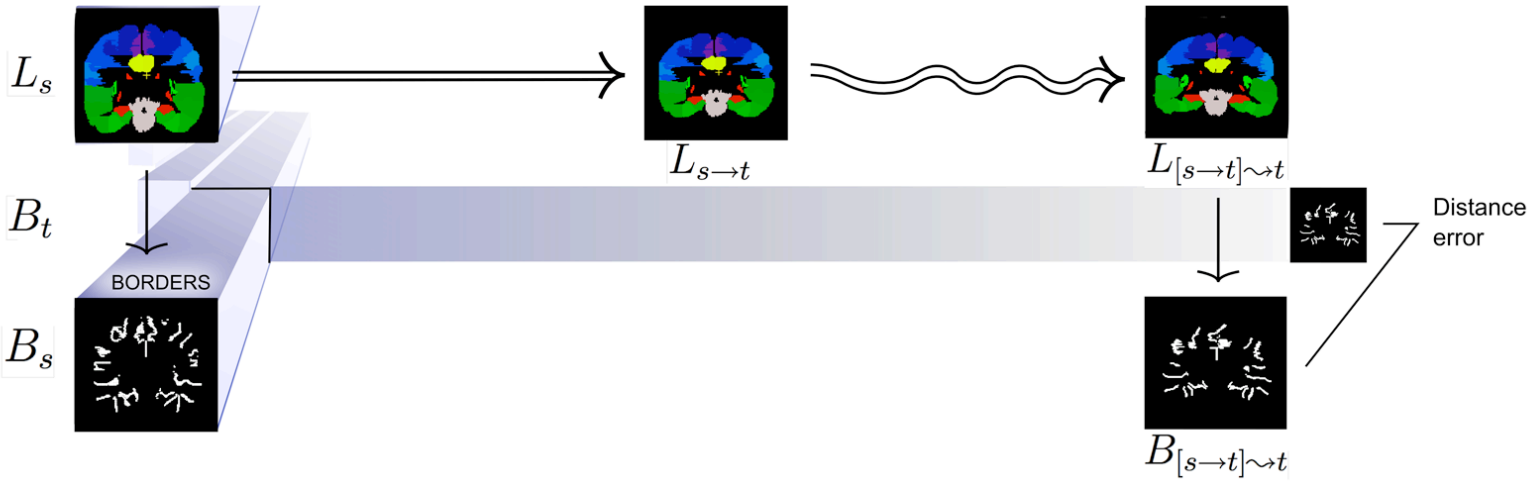
Algorithm	Code	Computer	Input	Setup	Run time: minutes
FLIRT (FSL 4.0)	C++	OSX, Linux, Win,...	Analyze, NiFTI		
AIR 5.25	C	OSX, Unix, Win,... ANSI C compiler	Analyze 8-/16-bit	remove nonbrain structures	6.7 (1.5)
ANIMAL (AutoReg 0.98k)	C, Perl	OSX, Linux, Unix	MINC	intensity correction (option)	11.2 (0.4)
ART	C++	OSX, Linux	Analyze		20.1 (1.6) [Linux]
Diffeomorphic Demons	C++	most (ITK compilable)	Analyze, NifTI, DICOM,... (ITK)		8.7 (1.2)
FNIRT beta	C++	OSX, Linux, Unix	Analyze, NiFTI (writes to Analyze)		29.1 (6.0)
IRTK	C++	OSX, Linux, Win	Analyze, NiFTI, VTK, GIPL	parameter file	120.8 (29.3)
JRD-fluid	C++	Sun	Analyze		17.1 (1.0) [Solaris]
ROMEO	C++	OSX, Linux, Win	Analyze, NiFTI, DICOM,... (ITK)	parameter file	7.5 (0.5)
		900+MB RAM		intensity correction (Hellier, 2003)	
SICLE	C++	OSX, Linux, Solaris, Alpha, Win	Analyze (7.5) 8-bit	dimensions divisible by 16	33.5 (6.6)
		g77/gfortran		intensity correction	
		lapack, f2c		isotropic	
		1+GB RAM		individual parameter files	
SyN beta	C++	most (ITK compilable)	Analyze, NiFTI, DICOM,... (ITK)		77 (15.1)
		1+GB RAM			
SPM5:	Matlab	most (Matlab)			
"SPM2-type" Normalization		Matlab 6.5 onwards	Analyze, NiFTI	smooth targets (Gaussian 8mm FWHM)	<1
Normalization		Matlab 6.5 onwards	Analyze, NiFTI	left-handed orientation	<1
Unified Segmentation		Matlab 6.5 onwards	Analyze, NiFTI	left-handed orientation	≃1
DARTEL Toolbox (pairs)		Matlab 7.0 onwards	Analyze, NiFTI	left-handed orientation	71.8 (6.3)*
				origin near anterior commissure	

Evaluation

distance

Average distance error = the minimum distance from each source region boundary point, S_rB_p, to the set of points making up the target region boundary, T_rB, averaged across P points:

$$DE_r = \frac{1}{P}\sum_{p=1}^{P} mindist(S_rB_p, T_rB)$$



Analysis

permutation tests

1. Select a subset of P independent brain pairs
2. Select a pair of methods (two vectors of P total overlap values)
3. Subtract the two vectors and compute the mean difference D
4. Select a subset of the elements from one of the vectors
5. Swap this subset across the two vectors
6. Subtract the resulting vectors; compute the mean difference D_p
7. Repeat steps #4-6 N times
8. Count the number of times n where[4] $\text{abs}(D_p) \geq \text{abs}(D)$
9. Compute the exact p-value: $\text{p} = \frac{n}{N}$
10. Repeat steps #1-9; compute the fraction of times where $\text{p} \leq 0.05$

Analysis

One-way ANOVA tests

We also performed a standard one-way ANOVA to test if the means of similar subsets of independent average target overlap values obtained by each of the registration methods are the same. We then subjected these results to a multiple comparison test using Bonferroni correction to determine which pairs of means are significantly different (disjoint 95% confidence intervals about the means, based on critical values from the t distribution). We repeated these ANOVA and multiple comparison tests 20 times, each time randomly selecting independent samples from each of the datasets. These tests are not expected to be as accurate as the permutation tests because some of the overlap values have skew distributions and because the p-values are not exact.

Analysis

Indifference-zone ranking

Our third evaluation between methods tested practical significance rather than statistical significance. For example, if a region is registered to another region of equal volume and results in an offset of a single voxel, this is not considered a significant misregistration, but offsets greater than this are considered significant. An evaluation measure of registration accuracy for a given region within a given brain pair is calculated for two different registration methods. If these two values are within delta of one another (referred to as an "indifference zone" when ranking (Bechhofer, 1954)), they are considered equal. The delta must correspond to a practical difference in registration. If we model a region as a cube, then a single-voxel offset along the normal to one of its faces would mean the voxels on that face of the cube reside outside of its target – this is equal to one-sixth of its surface. We therefore set delta to one-sixth of a target region's surface. For the IBSR18, CUMC12, and MGH10 datasets, we assumed the surface to be that of a cube ($6 \times edge^2 - 12 \times edge$, where $edge$ = the edge length of a cube with the volume of the target region, in voxels). For the LPBA40 dataset, we set the surface to the number of voxels bordering adjacent regions.