Published in final edited form as:
Neuroimage. 2011 February 1; 54(3): 2033–2044. doi:10.1016/j.neuroimage.2010.09.025.

A Reproducible Evaluation of ANTs Similarity Metric Performance in Brain Image Registration

Brian B. Avants, **Nicholas J. Tustison**, **Gang Song**, **Philip A. Cook**, **Arno Klein**[†], and **James C. Gee**

Penn Image Computing and Science Laboratory, University of Pennsylvania, Philadelphia, PA 19104

[†]New York State Psychiatric Institute, Columbia University, NY, NY 10032, USA

1. Introduction

Rapid advancement in biological and medical imaging technologies increases demand for quantitative, computational anatomy tools. The principal tools of this emerging field are deformable mappings between images whether they be driven by similarity metrics which are intensity-based, point-set-based, or both. Several categories of mappings exist in the literature. Of particular recent interest are diffeomorphic transformations which, by definition, preserve topology. Topology preservation is fundamental to making comparisons between objects in the natural world that are thought to differ or change while preserving local neighborhood relations. Cytoarchitectonic brain mapping studies also suggest that the layout of cell types throughout the brain is generally preserved (Schleicher et al., 2009), further motivating diffeomorphic mapping in the context of the brain.

Our limited assessment of published research mirrors the experience of many others who prefer a working paradigm of *reproducible research* (Kovacevic, 2006). Dr. Kovacevic describes “[reproducible research as] the idea that, in ‘computational’ sciences, the ultimate product is not a published paper but, rather, the entire environment used to produce the results in the paper (data, software,etc.).” After an informal survey of 15 published papers, she finds “none had code available” and “in only about half the cases were the parameters [of the algorithm] specified.” The computational sciences research community also voices concerns about reproducibility (Yoo and Metaxas, 2005; Ibanez et al., 2006). In this paper, we discuss our contribution to the open source medical image analysis research community which we call ANTs (Advanced Neuroimaging Tools). Built on an an Insight ToolKit (ITK) framework, this software package comprises a suite of tools for image registration, template building and segmentation based on previously published research. Here, we provide an overview of the package and detail recent technical advances, in the spirit of previous papers published in this journal (Neu et al., 2005; Zhang et al., 2008; Patel et al., 2010) and open source registration tools such as Elastix (Klein et al., 2010b).

The recent outcome from two large-scale comparative image registration algorithm assessments (Klein et al., 2009), http://empire10.isi.uu.nl is perhaps the most persuasive

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

evidence motivating the contributions discussed in this paper. Our Symmetric Normalization (SyN) transformation model (Avants et al., 2008) performs consistently in the top rank across all tests in the Klein et al. (2009) study and finished first overall in the phase one Empire-10 evaluation study of intra-subject thoracic CT registration (http://empire10.isi.uu.nl). Unlike some of the other algorithms in these studies, all of our methods (not just SyN) are open source software.

One difficulty in interpreting the results of these evaluation studies is that each algorithm uses a different combination of transformation model (the geometric constraint on the mapping between brains), similarity metric (the measure that evaluates how similar two images appear), and multi-resolution, optimization, and resampling strategies and parameter settings. Thus, one cannot isolate the effect of transformation model from similarity metric or optimization strategy. Other aspects of implementation may also differ, including whether the authors recommend using whole head or whole brain data. For instance, the DARTEL algorithm (Ashburner, 2007) uses whole head data and segmentation to aid performance while the other methods did not incorporate segmentation. The follow-up evaluation study Klein et al. (2010a) evaluated ART2.0 (Ardekani et al., 2005), SyN, and Freesurfer (Fischl and Dale, 2000) on whole head data and found that both brain extraction and registration via an "optimal" (group-generated) template improve performance. However, Klein et al. (2010a) applied generic parameters for ANTs, including the similarity metric, which might have resulted in suboptimal performance for the whole head component of the study.

Consequently, here we study the effect of the similarity metric on whole head registration-based labeling via an optimal template. We evaluate ANTs affine as well as nonlinear registration performance because accuracy in both stages is critical for successful registration-based brain segmentation/labeling. Furthermore, this problem is faced routinely in brain image processing labs (Ségonne et al., 2004; Sadananthan et al., 2010; Park and Lee, 2009; Lim and Pfefferbaum, 1989; de Boer et al., 2010; Acosta-Cabronero et al., 2008). One advantage of a consistent and modular framework, such as constructed in ANTs, is that it is possible to evaluate a single component of the processing stream while holding all other aspects constant.

The paper organization: Section 2 gives an overview of the transformation models and similarity metrics in ANTs and their use with SyN in population mapping. Section 3 reports results on a series of large-scale experiments using the LPBA manually labeled dataset to evaluate ANTs registration applied to cortical and brain labeling. Finally, we close with a discussion of our findings.

2. Theoretical Overview of ANTs

The following three components provide a common classification schema for registration methods (Brown, 1992; Ibanez et al., 2002):

- the *transformation model*, which includes the regularization kernels,
- the *similarity (or correspondence) measures*, and
- the *optimization strategy*.

In general, image normalization computes the optimal transformation, $\boldsymbol{\phi}$, within a transformation space which maps each $\mathbf{x}$ of image $\mathcal{I}(\mathbf{x})$ to a location in image $\mathcal{J}(\mathbf{z})$ by minimizing a cost function, $\mathcal{C}$ describing the similarity between $\mathcal{I}$ and $\mathcal{J}$ SyN, explained in detail below, symmetrizes this formulation and is available in ANTs. A summary of ANTs transformation models and similarity measures is in Table 1. Details follow in subsequent sections.

2.1. ANTs Transformation Models

A researcher in brain mapping may choose from a variety of transformation models with different degrees of freedom. For deformable transformations, one approach is to optimize within the space of non-topology-preserving, yet physics-based transformations—an approach pioneered by Bajcsy (Bajcsy and Kovacic, 1989). Elastic-type models such as HAMMER (Shen and Davatzikos, 2002), statistical parametric mapping (SPM) (Ashburner and Friston, 2000), free-form deformations (FFD) (Rueckert et al., 1999), and Thirion's Demons (Thirion, 1998) operate in the space of vector fields, which does not preserve topology. In other words, without applying ad hoc constraints, these algorithms may allow the brain topology to change in an uncontrolled way which makes the deformable mappings difficult to interpret in functional or anatomical studies.

Diffeomorphic transformations provide well-behaved solutions with mathematical guarantees about distance in deformation space and regularity. Furthermore, the diffeomorphic space has group structure (Arnold, 1991). Optimizing directly within this space shows remarkable success in various computational anatomy studies involving longitudinal (Avants et al., 2007; Fox et al., 2001), functional (Miller et al., 2005), and population data (Avants et al., 2007). We include three such diffeomorphic algorithms in the ANTs toolbox based on previous research and a new time-parameterized extension to the standard symmetric normalization (SyN) algorithm (Avants et al., 2008).

Regardless of current research trends, however, we recognize that selection of the transformation model is ultimately application-specific, that no single choice is optimal for all scenarios (Wolpert and Macready, 1997), and therefore, the transformation model must be chosen in a principled fashion. Indeed, several non-diffeomorphic algorithms perform quite well in Klein's comparative study of nonrigid registration algorithms (Klein et al., 2009). For this reason, ANTs also includes in its generic framework elastic-type methods among its transformation model options. In this paper, we focus on affine registration and the SyN method due to their proven reliability, speed and flexibility.

2.1.1. Rigid and Affine Linear Transformations—Image registration strategies often begin with a linear transformation for initial global alignment, followed by a deformable transformation with higher degrees of freedom. The linear transformations available within ANTs optimize either a mean-squared difference (MSQ), cross-correlation (CC) or mutual information (MI) similarity metric, each of which are optimized with respect to translation, rotation, and in the case of affine transformations, scaling and shearing. The successive optimization of each component of the linear transformation allows for careful control over increasing degrees of freedom. ANTs also composes the affine transformation with the deformable transformation field before performing any interpolation or downsampling. In this way, ANTs normalization never requires more than a single image interpolation step and is able to refer back to the original full-resolution images. The ANTs implementation of rigid mapping is quaternion-based with additional scaling and shearing terms when affine mapping is desired. The user enables purely rigid mapping by setting the `--do-rigid true` flag.

2.1.2. Vector Field Operators for Regularization—Deformable normalization strategies typically invoke a deformation regularization step which smooths the displacement field, $\mathbf{u}$, or velocity field, $\mathbf{v}$, or both by a linear operator such as the Laplacian or Navier-Stokes operator. One may write this regularization as a variational minimization in terms of its linear operator or in terms of a kernel function operating on the field itself, e.g., $\mathbf{u}_{smooth} = K * \mathbf{u}_{not\ smooth}$, where $K *$ denotes convolution with the Green's kernel, K, for the linear operator, L. ANTs regularization models operate on either the whole mapping ϕ or the gradient of the similarity term or both. The same regularization scheme is available for both diffeomorphic and the recent directly manipulated free-form deformation (DMFFD) (Tustison et al., 2009a) registration. ANTs users may set parameters such that discretized FFD strategies and diffeomorphisms are combined. ANTs enables a variety of choices for *K* including the Gaussian with varying σ and a variety of B-spline functions, both of which induce adequate regularity for normalization models used in ANTs. While additional physical operators will be incorporated in future releases, current B-spline options provide considerable flexibility (Tustison and Gee, 2005) that has yet to be fully explored.

2.1.3. Diffeomorphic Transformations—Diffeomorphisms form a group of differentiable maps with differentiable inverse (Ebin and Marsden, 1970; Mumford, 1998) that is closed under composition. ANTs assumes the diffeomorphism, ϕ, is defined on the image domain, Ω, and maintains an affine transform at the boundary such that $\phi(\partial\Omega) = A(\mathbf{Id})$ where $A(\mathbf{Id})$ is an affine mapping applied to the identity transformation. The map ϕ, over time, parameterizes a family of diffeomorphisms, $\phi(\mathbf{x},t) : \Omega \times t \rightarrow \Omega$, which can be generated by integrating a (potentially) time-dependent, smooth velocity field, $\upsilon : \Omega \times t \rightarrow \mathbb{R}^d$, through the ordinary differential equation (o.d.e.)

$$\frac{d\phi(\mathbf{x},t)}{dt}=\upsilon(\phi(\mathbf{x},t),t),\ \ \phi(\mathbf{x},0)=\mathbf{x}. \tag{1}$$

The existence and uniqueness theorem for o.d.e.'s implies that integrating Equation (1) generates a diffeomorphism. The deformation field yielded by ϕ is $\mathbf{u}(\mathbf{x}) = \phi(\mathbf{x}, 1) - \mathbf{x}$.

One typically encounters somewhat complex intensity transfers between one anatomical instance $\mathscr{J}$ and another instance $\mathscr{I}$. Thus, ANTs enables a variety of similarity metric possibilities beyond the conventional squared difference metric. This leads to the following generalization of the standard Large Deformation Diffeomorphic Metric Matching (LDDMM) equation (Beg et al., 2005):

$$\upsilon^*=\underset{\upsilon}{\text{argmin}}\left\{\int_0^1\|L\upsilon\|^2dt+\lambda\int_\Omega\prod\nolimits_\sim(\mathscr{I},\phi(\mathbf{x},1),\mathscr{J})d\Omega\right\} \tag{2}$$

where $\Pi_\sim$ is a similarity metric depending on the images and the mapping and λ controls the degree of exactness in the matching. We discuss established alternatives for Π in section 2.2.

Exploiting the fact that the diffeomorphism, ϕ, can be decomposed into two components ϕ_1 and ϕ_2, one may construct a *symmetric* alternative to Equation (2). Now define, in $t \in [0,0.5]$, $\upsilon(x,t) = \upsilon_1(x, t)$ and $\upsilon(x, t) = \upsilon_2(x, 1 - t)$ when $t \in [0.5, 1]$. This leads to the symmetric variant of Equation (2),

$$\{\upsilon_1^*,\upsilon_2^*\}=\underset{\upsilon_{1,2}}{\text{argmin}}\left\{\int_0^{0.5}\|L\upsilon_1(x,t)\|^2\,dt+\int_0^{0.5}\|L\upsilon_2(x,t)\|^2\,dt+\lambda\int_\Omega\prod\nolimits_\sim(\mathscr{I}\circ\phi_1(\mathbf{x},0.5),\mathscr{J}\circ\phi_2(\mathbf{x},0.5))d\Omega\right\}. \tag{3}$$

Note that the regularization term, here, is equivalent to that in equation 2. The only change is the splitting of the integral into two time intervals reflecting the underlying optimized components of the velocity field. The corresponding symmetric Euler-Lagrange equations are similar to (Miller et al., 2002). The difference, here, is that in finding υ^*, we minimize the variational energy from either endpoint towards the midpoint of the transformation, as indicated by the data term. This strategy "splits" the optimization dependence equally between both images. Thus, gradient-based iterative convergence deforms $\mathscr{I}$ and $\mathscr{J}$ along the geodesic diffeomorphism, ϕ, to a fixed point midway (intuited by the notion of shape distance) between $\mathscr{I}$ and $\mathscr{J}$ motivating the moniker “Symmetric Normalization” (SyN) for the solution strategy.

Other diffeomorphic algorithms in the research literature include DAR-TEL (Ashburner, 2007) and Diffeomorphic Demons (Vercauteren et al., 2007, 2009), both of which use a constant velocity, exponential model for generating diffeomorphisms. We include—within ANTs options—these four diffeomorphic transformation models for parameterizing $\phi(\cdot)$: Geodesic SyN, Greedy SyN, exponential mapping, and Greedy Exp (based on Diffeomorphic Demons). As shown in Table 1, each of these transformation models can utilize a host of similarity measures both individually and in combination.

Greedy SyN. Although the Geodesic SyN algorithm conforms most closely to the theoretical diffeomorphic foundations culminating in Equation (3), the computational and memory cost is significant due to the dense-in-time gradient calculations and requisite reintegration of the diffeomorphisms after each iterative update. While geodesic SyN is available in ANTs 2.0, the lower-cost, greedy variant called *Greedy SyN* is also available and was the strategy used in the large-scale comparative image registration algorithm assessment of (Klein et al., 2009).

Greedy optimization of Equation (3) calculates the gradient only at the midpoint of the full diffeomorphism, i.e. at $t = 0.5$,

$$\nabla\Pi=\frac{\partial}{\partial\phi_i}\Pi_{\sim}(\mathscr{I}(\phi_1^{-1}(\mathbf{x},0.5)),\ \mathscr{J}(\phi_2^{-1}(\mathbf{x},0.5))) \tag{4}$$

for $i \in \{1,2\}$. $\phi_1(\mathbf{x}, 0.5)$ and $\phi_2(\mathbf{x}, 0.5)$ are then updated from the previous iteration according to

$$\phi_i(\mathbf{x},0.5)=\phi_i(\mathbf{x},0.5)+(\delta K*\nabla\Pi_i(\mathbf{x},0.5))\ \mathrm{o}\phi_i(\mathbf{x},0.5). \tag{5}$$

Choices for the gradient descent parameter, δ, are discussed in section 3.3.3. In this equation, the gradient at the midpoint is mapped back to the origin of each diffeomorphism. We then update the full mapping by explicitly enforcing $\phi_i^{-1}(\phi_i(\mathbf{x},1))=\mathbf{x}$ in the discrete domain, as described in (Avants et al., 2008).

2.2. ANTs Intensity-Based Similarity Metrics

Several intensity-based image metrics appear in the literature with varying performance depending on their application. We include three of the most widely used similarity metrics within ANTs and reviewed in (Hermosillo et al., 2002): mean squared intensity difference (Christensen et al., 1996; Thirion, 1998; Beg et al., 2005; Ashburner, 2007; Vercauteren et al., 2009), cross-correlation (Gee, 1999; Ardekani et al., 2005; Avants et al., 2008), and mutual information (Viola and Wells, 1997; Rueckert et al., 1999; D’Agostino et al., 2003; Crum et al., 2003; Rogelj and Kovacic, 2006; Tao et al., 2009; Loeckx et al., 2010). ANTs implementation of these metrics all follow the same input/output interface and exist within metric-specific classes that inherit base functionality from a generic parent class. Each metric expects only two images as input, along with relevant parameters. The metrics expect the images to exist within the same physical space. We provide specific implementation details for each metric below and note that the code for each implementation is freely available. We restrict discussion to the elements of implementation that are critical to performance. Additionally, we report the derivatives of a pair of images, *I* and *J*, with respect to the identity transform, that is, after they have been mapped to the same space. Mapping to a different domain introduces a Jacobian change of variables as in (Beg et al., 2005) which may be introduced as a product with the derivative terms given here.

2.2.1. ANTs Mean Squared Intensity Difference—

The simplest of the metrics to implement is MSQ. However, a few details are critical. The MSQ derivative equations available in ANTs—via different command line options—are based on Demons algorithm variants (Thirion, 1998). Define g as a gradient vector and $D = I(\mathbf{x}) - J(\mathbf{x})$. Then the MSQ forcing equation may be written

$$\nabla_I \text{MSQ} = \frac{D}{D^2 + g^2} g, \tag{6}$$

where $g = \nabla I(\mathbf{x})$ or $g = g_s = \nabla I(\mathbf{x}) + \nabla J(\mathbf{x})$. These two gradient choices are available as command line options. Additionally, ANTs metrics all implement the complementary force, i.e. the similarity gradient with respect to J. This can be gained for MSQ by setting $D = J(\mathbf{x}) - I(\mathbf{x})$ and $g = \nabla J(\mathbf{x})$. Because SyN uses gradients with respect to both I and J, there is no need to use the “symmetrized gradient” g_s (as in (Thirion, 1998)). However, when using an “asymmetric” ANTs transformation model (e.g. LDDMM, Diffeomorphic Demons or traditional Demons-style elastic matching), using g_s may increase performance by providing additional image forces.

2.2.2. ANTs Cross-Correlation (CC)—

The current version of ANTs bases the correlation derivative on our prior work (Avants et al., 2008), but is much faster due to a sparse, linearized neighborhood updating scheme and a polynomial expansion of the CC terms. This new, accelerated cross-correlation approach is similar to techniques used for efficient low pass, median and texture co-occurrence filtering (Wells, 1986; Clausi and Jernigan, 1998; Huang et al., 1979). One may write the cross-correlation as:

$$\text{CC}(\mathbf{x}) = \frac{\sum_i ((I(\mathbf{x}_i) - \mu_{I(\mathrm{x})})(J(\mathbf{x}_i) - \mu_{J(\mathrm{x})}))^2}{\sum_i (I(\mathbf{x}_i) - \mu_{I(\mathrm{x})})^2 \sum_i (J(\mathbf{x}_i) - \mu_{J(\mathrm{x})})^2}, \tag{7}$$

where $\mathbf{x}$ is at the center *of* $N \times N$ square window (in two dimensions), μ is the mean value within the window centered at $\mathbf{x}$ and $\mathbf{x}_i$ iterates through that window. CC is expensive to compute when done naively but may be sped up by multiplying out the terms and storing local variables for each resulting term. Consider, in general, the polynomial equation, $\Sigma_i (a_i - \mu_a)(b_i — \mu_b)$, which multiplies out to $\Sigma_i(a_i b_i - \mu_b (\Sigma_i a_i) - \mu_a (\Sigma_i b_i) + \Sigma_i \mu_a \mu_b$. Each term in the CC equation above may be represented as this polynomial. Thus, to compute CC within a window, one may keep track of each of these five values: $\Sigma I(\mathbf{x}_i)$, $\Sigma J(\mathbf{x}_i)$, $\Sigma I(\mathbf{x}_i)^2$, $\Sigma J(\mathbf{x}_i)^2$, $\Sigma I(\mathbf{x}_i) J(\mathbf{x}_i)$ along with the number of voxels within the window which is constant except near the edges of an image. With all of these terms, one may compute the derivative of CC as described in Equations (6) and (7) of (Avants et al., 2008). Furthermore, note that—as one iterates through an image—only a few of the voxels that comprise $\Sigma I(\mathbf{x}_i)$, $\Sigma J(\mathbf{x}_i)$, $\Sigma I(\mathbf{x}_i)^2$, $\Sigma J(\mathbf{x}_i)^2$, $\Sigma I(\mathbf{x}_i) J(\mathbf{x}_i)$ change. That is, only the boundaries of the window are updated. In 2D, iterating left to right, the left edge voxels must exit the computation while right edge voxels must enter the computation. ANTs uses this efficient scheme to reduce the total computational expense from $3N^3m+5N^3p$ to $3N^2m+5N^2p$ operations per voxel over a 3D image (with some additional cost for the data structure that comprises the sliding window), where m is the cost of a multiplication and p is the cost of addition. This gives a theoretical speed-up of 5.36 when $N = 9$, $m = 2$, $p = 1$ and 6.65 when $N = 9$, $m = 4$, $p = 1$. In 3D, this results in an empirical speed-up of approximately a factor of 4 for a brain registration with a neighborhood of size 9×9×9, the recommended default for brain mapping with SyN driven by the CC similarity metric. In comparison, the Klein 2009 paper used a 5 × 5 × 5 window. The CPU, compiler and node usage all influence the speed-up factor. The fact that our practical speed-up is near the theoretical limit indicates that an implementation of CC that is not optimized will dominate computation time for deformable registration. Thus, the ANTs optimizations for gradient-based CC are a significant contribution and allows the use of a larger correlation window than before (Klein et al., 2009), which improves performance in whole head image registration. Additionally, our optimizations are distinct from the well-known paper (Lewis, 1995) which optimized non-gradient-based CC assuming a constant neighborhood in one of the two images. We used the default 9 × 9 × 9 window in this evaluation study.

2.2.3. ANTs Mutual Information (MI)—The ANTs implementation of mutual information and its gradient construct an image-based joint histogram and derive marginal distributions from this joint histogram. This implementation relates to work in (Hermosillo et al., 2002; Mattes et al., 2003; Rogelj et al., 2003) which describes the theory well. The basis of the ANTs MI function is the joint histogram of the images *I* and *J* which is constructed by locating a joint intensity value at each position, **x**, and then incrementing the nearest neighbor bin within the joint histogram. We then normalize the joint histogram by its sum to construct the two-dimensional joint probability image $\mathbf{Q}$: $[1, n_h]\times[1, n_h] \rightarrow [0, 1]$, where n_h, default 32, is the number of bins per dimension in the histogram. We also define a sub-voxel mapping from the intensity values in the images *I* and *J* to **Q**. That is, the intensity $i = I(\mathbf{x})$ maps to position *p* within the columns of the joint histogram and intensity $j = J(\mathbf{x})$ maps to *q* in the rows where a linear interpolant is used to find the continuous position. We may then interpolate **Q** at continuous positions with a cubic B-Spline kernel as described in (Mattes et al., 2003). The derivative with respect to $I(\mathbf{x})$ is derived in (Hermosillo et al., 2002):

$$\nabla \mathrm{MI}_I = \left(\frac{d_p \mathbf{Q}(p,q)}{\mathbf{Q}(p,q)} - \frac{d\mathbf{f}(p)}{\mathbf{f}(p)}\right)\nabla I(\mathbf{x}), \tag{8}$$

where *p* is the spatial index to the column of the joint histogram that locates the intensity at $I(\mathbf{x})$ and *q* is the spatial index to the row of the joint histogram that locates the intensity at $J(\mathbf{x})$. The term $d_p\mathbf{Q}(p, q)$ is the spatial gradient of the joint histogram **Q** in the direction of the columns, computed with the B-Spline interpolator. The term $d\mathbf{f}(p)$ is the spatial gradient of the marginal histogram **f** for *I* where the marginal is derived from the joint histogram, as in (Mattes et al., 2003). As with the other metrics, the ANTs MI function also computes the derivative with respect to *J* and uses both in the optimization of the registration.

2.2.4. Feature-Based Metrics—In addition to intensity-based metrics, ANTs contains similarity metrics for registering labeled point sets or label images. These include a landmark matching metric and two point-set metrics (Pluta et al., 2009; Tustison et al., 2009b) which can accommodate point sets of different cardinality. These point-set metrics are applicable alone for strict point-set registration or in parallel with intensity-based metrics for dual intensity/point-set registration. Exact matching and partial (or incompletely labeled) (Pluta et al., 2009) point-set matching are available, though not evaluated here.

2.3. ANTs Template Creation and Labeling

ANTs robustly maps populations to a common space by finding the template and set of transformations that gives the “smallest” parameterization of the dataset. The SyGN (symmetric groupwise normalization, pronounced “sign”) method implements this approach and is fully explained in (Avants et al., 2010b). The size of the parameterization, in the ANTs implementation of SyGN, is given by the metric distance between the average affine transformation and the identity affine transformation as well as the diffeomorphism lengths. No specific guess for the initial template is required. Instead, the template is derived completely from the database of n images, $\{J^i\}$. We denote such a template as image $\bar{I}$. Our previous work (Avants et al., 2010a) updates $\bar{I}$ with respect to both shape and the correlation, but here we use Euclidean distance as a metric for average appearance. In this study, the initial templates are obtained by averaging the data before any transformation is applied.

SyGN optimizes the shape of $\bar{I}$ via a diffeomorphism, ψ (which contains an affine transformation), such that the size and shape of the brain converges to the group mean. This is achieved, in ANTs, by optimizing the following energy iteratively,

$$E_{\bar{I}} = \sum_i E_{SyN,\Pi}(\bar{I}, J^i, \phi^i) \text{ where } \forall i, \phi^i(\mathbf{x}, 0) = \psi(\mathbf{x}), \tag{9}$$

where ψ is a diffeomorphism representing the initial conditions of each $\boldsymbol{\phi}^i$ and SyN gives the solution for each pairwise problem. The algorithm iteratively minimizes the energy $E_{\bar{I}}$ of Equation (9) with respect to the set of $\boldsymbol{\phi}^i$ through distributed computing (instantiated by the ANTs script `buildtemplateparallel.sh`). Additionally, the template appearance and template shape both approach the group mean in the Euclidean space of appearance, the affine space of shape and the diffeomorphic space of shape. This is in contrast to methods such as *congealing* (Learned-Miller, 2006) or (Joshi et al., 2004) in that neither method explicitly optimizes the geometric component of the template. Thus, the ANTs SyGN algorithm yields a robust result across populations, as will be shown in the evaluation section. The method typically converges in well under 10 iterations (usually three to five depending upon the complexity of the deformations in the data). Given a template, and a set of labels, the ANTs program (`ImageSetStatistics`) labels the template by majority voting (Heckemann et al., 2006).

2.4. ANTs Implementation: SVN Revision 603+

ANTs, built upon an ITK foundation, maintains the same coding style as its base. For much of its functionality, ANTs requires version 3.20 of the Insight ToolKit (ITK), necessitating the installation of ITK prior to installing ANTs. All ANTs source code is available via the online source code repository SourceForge.[1] Binaries for Windows, Mac OS X (OSX), 32- and 64-bit LINUX (Linus Torvald's UNIX) are also available from the same online location. For quality assurance and maintenance purposes we established an ANTs test reporting open source "dashboard" [2] on our lab website [3] to monitor compilation and testing of the ANTs program. Such a configuration facilitates reporting of user problems on a multitude of computing platforms. The methods above are all available within ANTs SVN revision 603 and later, compiled against stable ITK version 3.20. A user should download the binaries or compile the source code and run the built-in tests to verify functionality. The ANTs CMakeLists.txt file contains the commands that define the tests and test data used in automated testing (via the CMake ctest command) and allows users to evaluate whether they are getting the expected performance from their own installation. Finally, in Table 2 we give a brief summary of the arguments available for the normalization in the ANTs package. This includes the corresponding variable specification. More information can be found on the ANTs website. This work is based on the 1.9.1 ANTs release at http://sourceforge.net/downloads/advants/ANTS/.

[1] http://sourceforge.net/projects/advants/

[2] http://www.cdash.org

[3] http://www.picsl.upenn.edu/cdash/index.php?project=ANTS

3. Experimental Evaluation

We now apply the above methods using Gaussian regularization of the velocity field, the SyN transformation model, the SyGN template building algorithm and the MSQ, CC and MI metrics to build templates via cross-validation, label the templates by majority voting and apply the templates to the LPBA40 validation dataset.

3.1. 3-D LPBA40 Whole Head Image Normalization Evaluation

The LPBA40 dataset (Shattuck et al., 2008) was collected at the North Shore Long Island Jewish Health System imaging center and is maintained at UCLA. LPBA40 contains 40 images (20 male + 20 female) from normal, healthy ethnically diverse volunteers with average age of 29.2 ± 6.3 years. Each subject underwent 3D SPGR MRI on a 1.5T GE system resulting in $0.86 \times 0.86 \times 1.5 mm^3$ images. Each MRI in the LPBA40 dataset was manually labeled with 56 independent structures at the UCLA Laboratory of Neuro Imaging (LONI). The test-retest reliability of the labeling, across raters, was reported as a minimum Jaccard ratio of 0.697 in the supramarginal gyrus to a maximum of 0.966 in the gyrus rectus. A single labeling of each image is made available to the public and used, here, as silver-standard data for both training and testing in our cross-validation scheme.

3.2. Evaluation Pipeline

The evaluation begins by dividing the dataset of 40 subjects into group A (subjects 1 to 20) and B (subjects 21 to 40). Then, for each (affine, diffeomorphic) metric pair (MSQ, MSQ), (CC, CC), (MI, MI), (MSQ, MI), (CC, MI) we:

1. Construct a group A template via SyGN.
2. Construct a group B template via SyGN.
3. Label each template by majority voting.
4. Map group B to template A and group A to template B.
5. Warp the template labels, with nearest neighbor interpolation, to each individual and evaluate overlap measures with respect to ground truth for both affine and the combined affine and diffeomorphic maps.

Thus, for each evaluation run, we produce two templates (one for group A and one for group B) and mapping of all left-out subjects to the opposite group's template. The scripts that perform this evaluation are available in supplementary material and in the ANTs script base. A visual summary of the pipeline is in Figure 1. Note that the affine registration metrics in ANTs are derived from ITK and explained in ITK documentation. To determine registration quality, we use the Jaccard metric, defined as

$$S(R1, R2) = \frac{\#(R1 \cap R2)}{\#(R1 \cup R2)}, \tag{10}$$

which measures both difference in size and location between two binary segmentations, $R1$ and $R2$. The $\#(R)$ operator counts the number of nonzero pixels in the region, R, which represents a binary object (e.g. a brain or hippocampus labeling).

3.3. Parameter Selection

The theory section characterizes image registration algorithms as a combination of transformation model, similarity and optimization criterion. Here, we detail our experience with the most significant parameters in ANTs and explain default choices and useful parameter ranges. All of the user-controllable parameter choices made in this work are contained within the ANTs scripts `antsIntroduction.sh` which is called by `buildtemplateparallel.sh`) and wrapped by the script `LPBA_Leave_N_Out_ANTS_Evaluation.sh`, located at the Files section of the ANTs sourceforge website[4].

3.3.1. Transformation Models—

The transformation model, itself, is a parameter in ANTs. That is, does one choose SyN, SyN with time (geodesic SyN), an elastic type of model, diffeomorphic demons model? In this work, we selected SyN because it provides a compromise of speed, flexibility and performance. SyN and other diffeomorphic models penalize deformation linearly whereas elastic-style models penalize deformation quadratically. While a discussion of these details is beyond the scope of the paper, linear deformation penalties are fundamental to allowing large deformation and robust brain mappings across many different brain shapes. The only parameter to SyN, directly, is the gradient descent step-size (discussed in section 3.3.3). The second important component of the transformation model is the regularization which is related to the linear operator acting on the velocity and/or deformation field. In ANTs SyN, the default regularization is `Gauss[3,0]`, which indicates that the velocity field is smoothed by a Gaussian filter with variance of 3 × the image spacing. Increasing the value beyond 3 will increase the smoothness of the transformation (and reduce the fineness of detail in the mapping) and decreasing this value (e.g. to zero) will reduce the smoothness. We do not typically change this parameter. One may impose regularization on the deformation field by choosing a non-zero value for the second entry in the regularization option, e.g. `Gauss [3,1]`. The ANTs B-Spline regularization options have yet to be fully explored but show promise in initial experiments.

3.3.2. Similarity Metrics—

In this work, ANTs applies two preprocessing steps that impact the relative appearance of the brain and, thus, the similarity metrics discussed above. ANTs employs a histogram matching algorithm, described in (Avants et al., 2004; Yoo and Metaxas, 2005), as a default within the scripts that may be turned off by excluding the `--Use-Histogram-Matching` option from the command line. This step is suggested in (Noblet et al., 2006) and shown to be valuable in prior (unpublished) ITK evaluations. We also preprocess the data with ANTs bias correction which does not change the appearance significantly unless notable bias is present.

3.3.3. Optimization Strategy—

The ANTs gradient descent and multi-resolution optimization parameters are perhaps the most important to bring to the user's attention particularly if the user is interested in using alternative transformation models (in addition to the need for a good initial rigid/affine mapping before proceeding to deformable registration). We choose the multi-resolution optimization parameters—for both affine and deformable registration—based on the resolution of the input data and the structure within the image relative to this resolution. For typical 1mm^3 T1 MRI, we use three levels in a multiresolution Gaussian pyramid. That is, the registration algorithm begins at the resolution 1mm × 2^n, where n is the number of levels in the pyramid, and proceed through resolutions 1mm × 2^{n-1}, 1mm × 2^{n-2} until the full resolution is reached. In our experience, the 1mm^3 brain's resolution is rarely useful for deformable registration when downsampling proceeds beyond $n = 3$. However, further downsampling is sometimes useful for overcoming weak initialization in affine registration. Thus, when the resolution of input data does not match these expected settings, the user may want to alter the number of resolutions used in the deformable mapping (controlled through the `--number-of-iterations` vector parameter. The gradient descent parameters employed in ANTs are based on prior evaluation studies in affine registration (Song et al.,

[4] script location on web: https://sourceforge.net/downloads/advants/ANTS_Evaluation_Scripts/

2007) and deformable registration (Avants et al., 2008). Due to the linear deformation penalty, this gradient parameter does not typically need to be changed for SyN. Its useful range—for geodesic SyN—is between 0.1 and 1.0 where the optimal value will depend upon the nature of the problem, the regularization choice and the data. For greedy SyN, the useful range is narrower: 0.1 to 0.5 for most problems and for `Gauss[3,0]` regularization. Increasing the deformation field regularization (a non-zero second parameter) may require increasing the gradient step size. While we have found results to be robust to choices for the gradient descent parameter, values that are too large will result in energy oscillation while values that are too small will result in slow convergence.

4. Results

We first establish template stability across population sub-divisions and metrics. We then detail performance differences by comparing evaluation results across metrics.

4.1. Template Stability Across Metrics and Populations

We quantify template stability by choosing the group A (MI,CC) template (arbitrarily) as a reference and mapping all other templates to this reference and comparing the overlap between their labels and the group A (MI,CC) labels. The results are shown in Table 3. The overlap values, gained by affine registration, exceed the overlaps gained by deformable registration for any subject in the dataset. After deformable registration, overlap values exceed the repeatability that is achievable by human raters (Shattuck et al., 2008). The reduction in some of the A group overlap values after deformable registration suggests we are operating near the limit of achievable overlap when using nearest neighbor interpolation. See (Klein et al., 2010a) for examples of this issue. Figure 2 shows the templates derived in this study before and after registration to the CC group A template. The acutance of the MI template is relatively reduced in comparison to the MSQ and CC templates. Recalling that MI outperforms MSQ in terms of Jaccard overlap, one may conclude that the acutance of the template alone is insufficient in terms of determining the anatomical accuracy of a registration strategy.

4.2. Labeling Subjects Outside the Training Set

Five (affine, deformable) metric pairs were chosen for use in the full evaluation pipeline, from template construction to majority voting to labeling the left out subjects. In the first phase, we use the same metric consistently: (MSQ, MSQ), (CC, CC), and (MI, MI). In the second phase, we use MI as the first metric for two more pairs, (MI, MSQ) and (MI, CC), since MI was the best performer for affine registration in the first phase (see Figure 3). On the affine registration level, the mutual information performs best for both brain extraction and labeling of finer structures.

4.2.1. Brain Extraction—The initial affine registration results to the derived template show a clearly superior performance under MI, as verified by pairwise T-tests in Figure 3. At the same time, when all of the deformable metrics are given the same initialization with the MI metric, then they perform similarly, at least on first glance. The concern with using the Jaccard ratio on brain extraction is that small differences in values (even in the thousandths place) may correspond to visually meaningful differences in labeling performance. This is due to the fact that typical errors represent a small component of the binary image. An example of the labeling from one subject is shown in Figure 5. The mean±sd value of the registration-based diffeomorphic brain extractions (for all metric pairings) are: (MSQ, MSQ)= 0.938±0.0197, (CC, CC)= 0.937±0.0210, (MI, MI)= 0.956±0.0056, (MSQ, MI)= 0.955±0.0062, (CC, MI)= 0.958±0.0054. The (MSQ, MSQ) and (CC, CC) results are both significantly lower than the (MI, MSQ) and (MI, CC) results indicating that the affine MI metric boosts performance for CC and MSQ deformable mappings. The top performer on the Segmentation Validation Engine (the SVE http://sve.loni.ucla.edu/) as of May 10, 2010, shows an average Jaccard ratio of 0.9504 as obtained by user "cgaser" using VBM8.0. Thus, these template-based diffeomorphic brain extraction methods are competitive with the state-of-the-art. There is a difference in the brain extractions from the SVE and those distributed with LPBA40. Thus, evaluation numbers are not strictly comparable. The SVE data is kept hidden to prevent overfitting of data to results. As this study intends to use only accessible data, we restrict to evaluation on the public components of LPBA40.

4.2.2. Extraction of Brain Sub-Regions—The trends present in overall brain extraction persist in the evaluation of the sub-region overlaps. The overall error in the sub-regions is shown in Figure 3, while the region-wise are shown in Figure 4. The table reveals the specific regions where performance differs across metrics. Furthermore, unsurprisingly, the trends in the MI-affine column are reflected in the diffeomorphic results (all of which used the MI-affine metric to initialize the diffeomorphic matching). Correlations between the diffeomorphic and MI-affine results are also strong, as in the figure. This further accentuates the importance of affine initialization in determining the deformable outcome. Note that, as shown in (Rohlfing et al., 2004), the Jaccard overlap values of different structures is affected by their surface to volume ratio. One must take this variation into account when interpreting these results.

4.3. Relative Computation Time for Each Metric

We quantify the relative wall-clock computation time of the similarity metrics in terms of the computation time for the mean squares metric (the simplest and fastest of the three). We run 10 iterations of the metric computation at full resolution (that is, without running a registration) on a machine that is nominally idle. The input image was three-dimensional with 256×124×256 voxels, as in LPBA40 data. The results are MSQ=1, MI=14.7, CC=19.1 where MSQ took 22 seconds. Thus, the CC metric is the most time-consuming and the MSQ metric may be the most efficient for performing an initial brain extraction that may be later refined by a post-processing algorithm.

5. Discussion

5.1. Summary

In this paper, we provide an overview of the ANTs toolkit and detail the ANTs implementation of MSQ, CC and MI deformable image registration metrics. We also contribute a new implementation of the CC metric that reduces computation time by a factor of 4–5 with default parameters in 1mm^3 3D brain image registration. We evaluate the impact of these metric choices—and their affine counterparts—on optimal template construction and template-based brain labeling. We use a conservative two-fold cross-validation strategy to show template stability. We establish—quantitatively—that the templates derived from the subsets of the data are more similar to each other than any individual in the dataset. The law of large numbers in anatomical variability, combined with effects of diminishing returns, explain these findings. That is, brains from different individuals sampled in a demographic are coarsely similar and the somewhat random differences tend to average out. In addition, the similarity metric does not have a large impact on the overall template shape. Despite a very high similarity of the templates, there do exist small residual differences after high-dimensional alignment with different metrics and sub-populations. Future work will be required to understand the nature and impact of these differences.

Our results show that mutual information-based affine registration, in ANTs, provides the best initialization for deformable registration. Mutual information, along with normalized mutual information, has advantages as a similarity metric in dealing with scanner variations and pathomorphological changes. It is possible that MI's robustness may prove a requirement, over the long-term, for large-scale clinical studies. Indeed, as shown in our own evaluation, MI performs best of the three metrics for whole head affine registration. It remains to be seen if some variants of MI are also optimal for deformable registration. It is possible that other implementations of CC and MSQ affine registration would perform as well as MI, but we hypothesize that the MI's quality performance is in part due to its well-known robustness to non-matching structure (e.g. features that exist outside the brain and exhibit significant inter-subject variation). The best diffeomorphic results in our study come from initializing with MI-based affine registration, regardless of the deformable metric used.

One surprising result from our study is the relative similarity in brain extraction performance across the deformable metrics, after affine initialization with MI. This suggests that extraction of larger structures is—in this dataset—very sensitive to affine initialization quality and less sensitive to the deformable metric. However, Figure 5 shows that apparently small variation in Jaccard metric may result in visually obvious differences in performance. The supplementary material contains the evaluation values for the Hausdorff distance between brain extractions, which may be a more sensitive measure in this application.

Two-fold cross-validation reduces bias in our results and tests generalization to new data (assuming similar resolution, contrast, etc). While all studies should use such a strategy, some involve sets that are too small to leave out any data (Yushkevich et al., 2009). Other studies simply accept a biased strategy though it is not necessary (Jia et al., 2010). Results are artificially inflated when the same data are used in both testing and training (Vul et al., 2009; Kriegeskorte et al., 2009). This effect makes it challenging to compare results that use cross-validation and those that do not.

The current study also highlights the impact of quality affine initialization in brain labeling performance. We report correlations greater than 0.92 between the initial affine registration result and the final deformable registration result that persists across metrics. Consequently, initialization quality in diffeomorphic image registration is of critical importance.

5.2. Relation to Other Work

In addition to (Klein et al., 2009) and (Klein et al., 2010a), a few other studies compare metrics in affine registration (Studholme et al., 1997; Woods et al., 1998) and deformable registration (Woods et al., 1998; Noblet et al., 2006). Noblet et al (Noblet et al., 2006) use an intensity difference metric (and a few transformations thereof) to show superior performance of a B-Spline algorithm relative to the Demons algorithm. Many aspects of the method were validated, but the method and results were not made public, to our knowledge. Studholme found (Studholme et al., 1997) that mutual information was more robust for rigid registration of PET-MRI head data when compared to other metrics, including cross-correlation. Perhaps the best known evaluation, historically, is that by Hellier et al (Hellier et al., 2003). Relative to Hellier's evaluation, the current work uses a single framework to test different similarity metrics without the confound of different pre-processing and transformation implementations. That is, of the three registration components detailed in the introduction, we hold two constant and evaluate one. Furthermore, in the spirit of open science, our code base, evaluation data, and evaluation software are made fully available.

A comparison of results reported here and those in a recent paper (Heckemann et al., 2010) suggests that multi-template labeling outperforms single-template labeling. As may be seen in Klein 2009, specifically Figure 5, relative overlap performance across algorithms is largely consistent across evaluation datasets. However, absolute performance values have notable variation. Thus, one must take care in directly comparing overlap values from LPBA40 data with those from Hammers 2003/Heckemann 2010, in particular because LPBA40 data is lower resolution. Furthermore, the current study evaluates labeling via *whole-head* normalization, whereas Heckemann 2010 (and Klein 2009) normalize brain-extracted images.

Despite these caveats, multi-template labeling likely yields a performance advantage, in general, and therefore we provide a script `ants_multitemplate_labeling.sh` at https://sourceforge.net/downloads/advants/ANTS_Evaluation_Scripts/ that implements the multi-template strategy with ANTs. In fact, the methods used in this paper to label our group template and those used in standard multi-template labeling are fundamentally similar. Thus, no new developments are needed to implement multi-template labeling with ANTs. We also provide, at the above location, multi-template labeling results derived from the Hammers dataset by applying the `ants_multitemplate_labeling.sh` script to the 19 datasets at http://www.brain-development.org/ (Hammers et al., 2003; Heckemann et al., 2006). Results are competitive with both (Heckemann et al., 2006, 2010) though the latter appears to use a different label set. The closest comparison may be made with (Heckemann et al., 2006), which uses almost the same label set, though with 30 datasets in total. Our results only incorporated the 19 currently available online. However, we currently focus on single-template labeling strategies due to the significant human effort required to generate consistent manual labels across datasets. A common, single template space is used in the large majority of population studies and the main purpose of this paper is to detail an open-source framework to implement and benchmark such studies. Some of these studies are discussed below.

ANTs users employ this technology in a variety of application domains, including but not limited to, fMRI analysis (Yassa et al., 2010), morphometry (Hanson et al., 2010), anatomical labeling of both human and mouse anatomy as well as in computer vision. ANTs has proven successful in large-scale normalization studies in not only healthy subjects, but also diseased subjects with large anatomical variance Avants et al. (2008); Klein et al. (2010a). The SyN method from ANTs recently finished as the top performer in an unbiased registration evaluation using manually landmarked intra-subject pairs of thoracic CT volumes (the EMPIRE-10 challenge for MICCAI 2010, http://empire10.isi.uu.nl). ANTs large-deformation methods easily adapt to processing subjects with epilepsy-induced sclerosis (Avants et al., 2010b), Alzheimer’s disease (Yushkevich et al., 2010), mild cognitive impairment (Yassa et al., 2010), and subjects with autopsy-confirmed frontotemporal dementia, which induces severe ventriculomegaly (Avants et al., 2010a). Finally, a subset of the ANTs toolkit is under development for inclusion in version 4 of the Insight Toolkit which will bring these methods to more users, increase robustness and ensure continued user support.

5.3. Shortcomings of this Study

The goal of this paper was to use the similarity metric as the variable of interest. As such, we did not evaluate the impact of the transformation and regularization models on registration accuracy and leave this to future work. A large number of parameter or algorithm design choices, both subtle and obvious, were also selected by relying upon experience and good engineering principles, but without direct evaluation. For instance, we did not explore the many ranges of possible downsampling strategies that could be employed in our multiresolution framework. As a second example, we did not use partial volume interpolation in our MI implementation as recommended by Maes (Maes et al., 1997). The size of the joint histogram in MI may also impact performance. However, we believe that there is a more fundamental issue with using a global MI measure for intra-modality registration. The more “flexible” MI correspondences (relative to CC or MSQ) may reduce precision. Despite this claim, it is difficult to prove, due to the importance of implementation details, use of normalized or unnormalized MI, or other implementations such as the maximum distance-gradient-magnitude similarity measure (Gan and Chung, 2005). ANTs also provides the ability to incorporate cortical constraints, shown to benefit brain registration (Hellier and Barillot, 2003), but such data is not leveraged in this analysis. Furthermore, we did not evaluate on non-brain or non-MRI data. Thus, we may not be able to generalize these results to other modalities or other organs. Lastly, we note that measuring brain labeling accuracy with respect to expert raters, itself, has limitations. For instance, raters may be systematically incorrect in some structures. Secondly, the biological plausibility of the mappings is not rated, nor is the detection power for subtle group effects on brain structure.

5.4. Final Conclusions

This paper details the primary ANTs normalization strategies and provides overview on other aspects of the toolkit. We focus on the deformable similarity metrics and some of the transformation models available in ANTs, provide the philosophy of implementation and give quantitative justification for default ANTs similarity metrics in both deformable and affine registration. We provide a new fast implementation of the CC metric for deformable registration, quantify the latest ANTs performance on brain labeling the LPBA40 dataset and show that brain extraction performance is competitive with the best available results. However, it is currently challenging to compare our region-wise results on LPBA40 data with other methods. This is in part because there are few reported results on LPBA40 data in the literature, and, secondly, those that exist in the literature use different approaches to or lack of cross-validation. We also highlight the similarity of templates derived from data within a demographic and affirm the importance of affine registration to deformable registration performance. Most importantly, this study provides reference scripts (written in bash with a translation in python) and code that may be reproducibly applied to a common evaluation dataset. We encourage other researchers to compare against these results using a similar two-fold cross-validation design, along with the Jaccard ratio as an evaluation metric. Supplementary material provides Dice overlaps, true/false positive ratios and Hausdorff metrics as well if other researchers prefer these measures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

ANTs is supported by Grant R01EB006266-01 from the National Institute Of Biomedical Imaging and Bioengineering and administered through the UCLA Center for Computational Biology. We would also like to thank Ray Razlighi for checking the ANTs MI code and improving both its clarity and implementation.

References

Acosta-Cabronero J, Williams GB, Pereira JMS, Pengas G, Nestor PJ. The impact of skull-stripping and radio-frequency bias correction on grey-matter segmentation for voxel-based morphometry. Neuroimage. 2008 Feb; 39(4):1654–1665. [PubMed: 18065243]

Ardekani BA, Guckemus S, Bachman A, Hoptman MJ, Wojtaszek M, Nierenberg J. Quantitative comparison of algorithms for inter-subject registration of 3D volumetric brain MRI scans. J Neurosci Methods. 2005 Mar; 142(1):67–76. [PubMed: 15652618]

Arnold, VI. Ordinary Differential Equations. Springer-Verlag; 1991.

Ashburner J. A fast diffeomorphic image registration algorithm. Neuroimage. 2007 Oct; 38(1):95–113. [PubMed: 17761438]

Ashburner J, Friston K. Voxel-based morphometry—The methods. Neuroimage. 2000; 11:805–821. [PubMed: 10860804]

Avants B, Anderson C, Grossman M, Gee JC. Spatiotemporal normalization for longitudinal analysis of gray matter atrophy in frontotemporal dementia. Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv. 2007; 10(Pt 2):303–310.

Avants B, Sundaram T, Duda JT, Gee JC, Ng L. Insight into Images. A K Peters, Ch. Non-Rigid Image Registration. 2004:307–348.

Avants BB, Cook PA, Ungar L, Gee JC, Grossman M. Dementia induces correlated reductions in white matter integrity and cortical thickness: A multivariate neuroimaging study with sparse canonical correlation analysis. Neuroimage. 2010a Apr; 50(3):1004–1016. [PubMed: 20083207]

Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal. 2008 Feb; 12(1):26–41. [PubMed: 17659998]

Avants BB, Tustison NJ, Song G, Gee JC. ANTS: Advanced Open-Source Normalization Tools for Neuroanatomy. Penn Image Computing and Science Laboratory. 2009

Avants BB, Yushkevich P, Pluta J, Minkoff D, Korczykowski M, Detre J, Gee JC. The optimal template effect in hippocampus studies of diseased populations. Neuroimage. 2010b Feb; 49(3):2457–2466. [PubMed: 19818860]

Bajcsy R, Kovacic S. Multiresolution elastic matching. Computer Vision, Graphics, and Image Processing. 1989; 46:1–21.

Beg MF, Miller MI, Trouvé A, Younes L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vision. 2005; 61(2):139–157.

Brown LG. A survey of image registration techniques. ACM Computing Surveys. 1992 December; 24 (4):325–376.

Christensen GE, Rabbitt RD, Miller MI. Deformable templates using large deformation kinematics. IEEE Transactions on Image Processing. 1996; 5(10):1435–1447. [PubMed: 18290061]

Clausi DA, Jernigan ME. A fast method to determine cooccurrence texture features. IEEE Trans. Geo. Rem. Sens. 1998; 36(1):298–300.

Crum WR, Hill DLG, Hawkes DJ. Information theoretic similarity measures in non-rigid registration. Inf Process Med Imaging. 2003 Jul.18:378–387. [PubMed: 15344473]

D'Agostino E, Maes F, Vandermeulen D, Suetens P. A viscous fluid model for multimodal non-rigid image registration using mutual information. Med Image Anal. 2003 Dec; 7(4):565–575. [PubMed: 14561559]

de Boer R, Vrooman HA, Ikram MA, Vernooij MW, Breteler MMB, van der Lugt A, Niessen WJ. Accuracy and reproducibility study of automatic MRI brain tissue segmentation methods. Neuroimage. 2010 Jul; 51(3):1047–1056. [PubMed: 20226258]

Ebin DG, Marsden J. Groups of diffeomorphisms and the motion of an incompressible fluid. Annals of Mathematics. 1970; 92:102–163.

Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci U S A. 2000 Sep; 97(20):11050–11055. [PubMed: 10984517]

Fox N, Crum W, Scahill R, Stevens J, Janssen J, Rossor M. Imaging of onset and progression of Alzheimer's disease with voxel-compression mapping of serial magnetic resonance images. Lancet. 2001; 358:201–205. [PubMed: 11476837]

Gan R, Chung ACS. Multi-dimensional mutual information based robust image registration using maximum distance-gradient-magnitude. Inf Process Med Imaging. 2005; 19:210–221. [PubMed: 17354697]

Gee JC. On matching brain volumes. Pattern Recognition. 1999; 32:99–111.

Hammers A, Allom R, Koepp MJ, Free SL, Myers R, Lemieux L, Mitchell TN, Brooks DJ, Duncan JS. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum Brain Mapp. 2003 Aug; 19(4):224–247. [PubMed: 12874777]

Hanson JL, Chung MK, Avants BB, Shirtcliff EA, Gee JC, Davidson RJ, Pollak SD. Early stress is associated with alterations in the orbitofrontal cortex: a tensor-based morphometry investigation of brain structure and behavioral risk. J Neurosci. 2010 Jun; 30(22):7466–7472. [PubMed: 20519521]

Heckemann RA, Hajnal JV, Aljabar P, Rueckert D, Hammers A. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage. 2006 Oct; 33(1):115–126. [PubMed: 16860573]

Heckemann RA, Keihaninejad S, Aljabar P, Rueckert D, Hajnal JV, Hammers A, Initiative ADN. Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. Neuroimage. 2010 May; 51(1):221–227. [PubMed: 20114079]

Hellier P, Barillot C. Coupling dense and landmark-based approaches for nonrigid registration. IEEE Trans Med Imaging. 2003 Feb; 22(2):217–227. [PubMed: 12715998]

Hellier P, Barillot C, Corouge I, Gibaud B, Goualher GL, Collins D, Evans A, Malandain G, Ayache N, Christensen G, Johnson H. Retrospective evaluation of intersubject brain registration. IEEE Trans Med Imaging. 2003 Sep; 22(9):1120–1130. [PubMed: 12956267]

Hermosillo G, Chef d'Hotel C, Faugeras O. Variational methods for multimodal image matching. International Journal of Computer Vision. 2002 December; 50(3):329–343.

Huang T, Yang G, Tang G. A fast two-dimensional median filtering algorithm. IEEE ASSP. 1979; 27 (1):13–18.

Ibanez, L.; Avila, RS.; Aylward, SR. Open source and open science: how it is changing the medical imaging community; Proc. of the International Symposium on Biomedical Imaging; 2006.

Ibanez L, Ng L, Gee JC, Aylward S. Registration patterns: The generic framework for image registration of the Insight Toolkit. IEEE International Symposium on Biomedical Imaging. 2002 July.:345–348.

Jia H, Wu G, Wang Q, Shen D. Absorb: Atlas building by self-organized registration and bundling. Neuroimage. 2010 Jul; 51(3):1057–1070. [PubMed: 20226255]

Joshi S, Davis B, Jomier M, Gerig G. Unbiased diffeomorphic atlas construction for computational anatomy. Neuroimage. 2004; 23 Suppl. 1:S151–S160. [PubMed: 15501084]

Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang M-C, Christensen GE, Collins LD, Gee J, Hellier P, Song JH, Jenkinson M, Lepage C, Rueckert D, Thompson P, Vercauteren T, Woods RP, Mann JJ, Parsey RV. Evaluation of 14 non-linear deformation algorithms applied to human brain MRI registration. Neuroimage. 2009 Jan.

Klein A, Ghosh SS, Avants B, Yeo BTT, Fischl B, Ardekani B, Gee JC, Mann JJ, Parsey RV. Evaluation of volume-based and surface-based brain image registration methods. Neuroimage. 2010a May; 51 (1):214–220. [PubMed: 20123029]

Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. Elastix: A toolbox for intensity-based medical image registration. IEEE Trans Med Imaging. 2010b Jan; 29(1):196–205. [PubMed: 19923044]

Kovacevic J. From the editor-in-chief. IEEE Transactions on Image Processing. 2006; 15:12. [PubMed: 16435533]

Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci. 2009 May; 12(5):535–540. [PubMed: 19396166]

Learned-Miller EG. Data driven image models through continuous joint alignment. IEEE Trans Pattern Anal Mach Intell. 2006 Feb; 28(2):236–250. [PubMed: 16468620]

Lewis, J. Vision Interface. Canadian Image Processing and Pattern Recognition Society. 1995. Fast normalized cross-correlation; p. 120-123.

Lim KO, Pfefferbaum A. Segmentation of MR brain images into cerebrospinal fluid spaces, white and gray matter. J Comput Assist Tomogr. 1989; 13(4):588–593. [PubMed: 2745775]

Loeckx D, Slagmolen P, Maes F, Vandermeulen D, Suetens P. Nonrigid image registration using conditional mutual information. IEEE Trans Med Imaging. 2010 Jan; 29(1):19–29. [PubMed: 19447700]

Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. Multimodality image registration by maximization of mutual information. IEEE Trans Med Imaging. 1997 Apr; 16(2):187–198. [PubMed: 9101328]

Mattes D, Haynor DR, Vesselle H, Lewellen TK, Eubank W. PET-CT image registration in the chest using free-form deformations. IEEE Trans Med Imaging. 2003 Jan; 22(1):120–128. [PubMed: 12703765]

Miller MI, Trouve A, Younes L. On the metrics and euler-lagrange equations of computational anatomy. Annu Rev Biomed Eng. 2002; 4:375–405. [PubMed: 12117763]

Miller MI, Trouvè A, Younes L. Geodesic shooting for computational anatomy. J. Mathematical Imaging and VisionSubmitted. 2005

Mumford D. Pattern theory and vision. Questions Matheematiques En Traitement Du Signal et de L'Image, Institut Henri Poincare. 1998; 3:7–13.

Neu SC, Valentino DJ, Toga AW. The loni debabeler: a mediator for neuroimaging software. Neuroimage. 2005 Feb; 24(4):1170–1179. [PubMed: 15670695]

Noblet V, Heinrich C, Heitz F, Armspach J-P. Retrospective evaluation of a topology preserving non-rigid registration method. Med Image Anal. 2006 Jun; 10(3):366–384. [PubMed: 16497537]

Park JG, Lee C. Skull stripping based on region growing for magnetic resonance brain images. Neuroimage. 2009 Oct; 47(4):1394–1407. [PubMed: 19389477]

Patel V, Dinov ID, Horn JDV, Thompson PM, Toga AW. Loni mind: metadata in nifti for dwi. Neuroimage. 2010 Jun; 51(2):665–676. [PubMed: 20206274]

Pluta J, Avants BB, Glynn S, Awate S, Gee JC, Detre JA. Appearance and incomplete label matching for diffeomorphic template based hippocampus segmentation. Hippocampus. 2009 Jun; 19(6):565–571. [PubMed: 19437413]

Rogelj P, Kovacevic S, Gee JC. Point similarity measures for non-rigid registration of multi-modal data. Computer Vision and Image Understanding. 2003; 92:112–140.

Rogelj P, Kovacic S. Symmetric image registration. Med Image Anal. 2006 Jun; 10(3):484–493. [PubMed: 15896998]

Rohlfing T, Brandt R, Menzel R, Maurer CR. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. Neuroimage. 2004 Apr; 21(4):1428–1442. [PubMed: 15050568]

Rueckert D, Sonoda L, Hayes C, Hill D, Leach M, Hawkes D. Nonrigid registration using free-form deformations: Application to breast MR images. IEEE Trans Med Imaging. 1999 Aug; 18(8):712–721. [PubMed: 10534053]

Sadananthan SA, Zheng W, Chee MWL, Zagorodnov V. Skull stripping using graph cuts. Neuroimage. 2010 Jan; 49(1):225–239. [PubMed: 19732839]

Schleicher A, Morosan P, Amunts K, Zilles K. Quantitative architectural analysis: A new approach to cortical mapping. J Autism Dev Disord. 2009 Jul.

Ségonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, Fischl B. A hybrid approach to the skull stripping problem in MRI. Neuroimage. 2004 Jul; 22(3):1060–1075. [PubMed: 15219578]

Shattuck DW, Mirza M, Adisetiyo V, Hojatkashani C, Salamon G, Narr KL, Poldrack RA, Bilder RM, Toga AW. Construction of a 3D probabilistic atlas of human cortical structures. Neuroimage. 2008 Feb; 39(3):1064–1080. [PubMed: 18037310]

Shen D, Davatzikos C. Hammer: Hierarchical attribute matching mechanism for elastic registration. IEEE Transactions on Medical Imaging. 2002; 21(11):1421–1439. [PubMed: 12575879]

Song, G.; Avants, B.; Gee, J. Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis. 2007. Multi-start method with prior learning for image registration.

Studholme C, Hill DL, Hawkes DJ. Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. Med Phys. 1997 Jan; 24(1):25–35. [PubMed: 9029539]

Tao G, He R, Datta S, Narayana PA. Symmetric inverse consistent nonlinear registration driven by mutual information. Comput Methods Programs Biomed. 2009 Aug; 95(2):105–115. [PubMed: 19268386]

Thirion JP. Image matching as a diffusion process: an analogy with Maxwell’s demons. Medical Image Analysis. 1998; 2(3):243–260. [PubMed: 9873902]

Tustison NJ, Avants BB, Gee JC. Directly manipulated free-form deformation image registration. IEEE Trans Image Process. 2009a Mar; 18(3):624–635. [PubMed: 19171516]

Tustison NJ, Awate SP, Song G, Cook TS, Gee JC. A new information-theoretic measure to control the robustness-sensitivity trade-off for DMFFD point-set registration. Inf Process Med Imaging. 2009b; 21:215–226. [PubMed: 19694265]

Tustison NJ, Gee JC. N-D C^k B-spline scattered data approximation. Insight Journal. 2005 published online.

Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons using itk’s finite difference solver hierarchy. Insight Journal—2007 MICCAI Open Science Workshop. 2007

Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: Efficient non-parametric image registration. Neuroimage. 2009; 45(1 Suppl):S61–S72. [PubMed: 19041946]

Viola P, Wells WM. Alignment by maximization of mutual information. International Journal of Computer Vision. 1997; 24(2):137–154.

Vul E, Harris C, Winkielman P, Pashler H. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. Perspectives on Psychological Science. 2009; 4(3):274–290.

Wells WM. Efficient synthesis of gaussian filters by cascaded uniform filters. IEEE Trans. Pattern Anal. Machine Intell. 1986; 8(2):234–239.

Wolpert D, Macready W. No free lunch theorems for optimization. IEEE Transactions on Evolutionary Computation. 1997 April; 1(1):67–82.

Woods RP, Grafton ST, Watson JD, Sicotte NL, Mazziotta JC. Automated image registration: Ii. intersubject validation of linear and nonlinear models. J Comput Assist Tomogr. 1998; 22(1):153–165. [PubMed: 9448780]

Yassa MA, Stark SM, Bakker A, Albert MS, Gallagher M, Stark CEL. High-resolution structural and functional MRI of hippocampal CA3 and dentate gyrus in patients with amnestic mild cognitive impairment. Neuroimage. 2010 Jul; 51(3):1242–1252. [PubMed: 20338246]

Yoo TS, Metaxas DN. Open science-combining open data and open source software: Medical image analysis with the Insight Toolkit. Med Image Anal. 2005 Dec; 9(6):503–506. [PubMed: 16169766]

Yushkevich PA, Avants BB, Das SR, Pluta J, Altinay M, Craige C, Initiative ADN. Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: An illustration in ADNI 3T MRI data. Neuroimage. 2010 Apr; 50(2):434–445. [PubMed: 20005963]

Yushkevich PA, Avants BB, Pluta J, Das S, Minkoff D, Mechanic-Hamilton D, Glynn S, Pickup S, Liu W, Gee JC, Grossman M, Detre JA. A high-resolution computational atlas of the human hippocampus from postmortem magnetic resonance imaging at 9.4 t. Neuroimage. 2009 Jan; 44(2):385–398. [PubMed: 18840532]

Zhang J, Liang L, Anderson JR, Gatewood L, Rottenberg DA, Strother SC. Evaluation and comparison of glm- and cva-based fMRI processing pipelines with java-based fMRI processing pipeline evaluation system. Neuroimage. 2008 Jul; 41(4):1242–1252. [PubMed: 18482849]

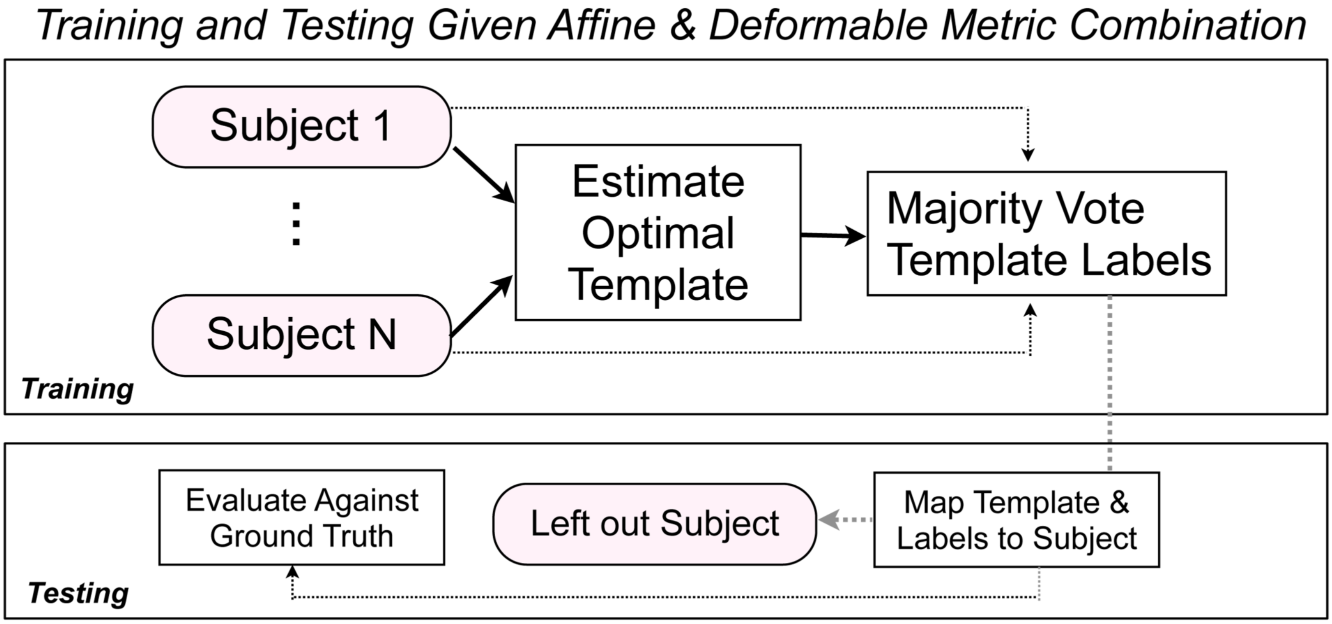

Training and Testing Given Affine & Deformable Metric Combination
Subject 1
Subject N
Estimate Optimal Template
Majority Vote Template Labels
Training
Evaluate Against Ground Truth
Left out Subject
Map Template & Labels to Subject
Testing

Figure 1.
The evaluation pipeline employs two-fold cross-validation and evaluates the following (affine, deformable) metric pairs: (MSQ,MSQ), (CC,CC), (MI,MI), (MI,MSQ), (MI,CC). The LPBA dataset's labels provide the ground truth for the subject registration being evaluated.

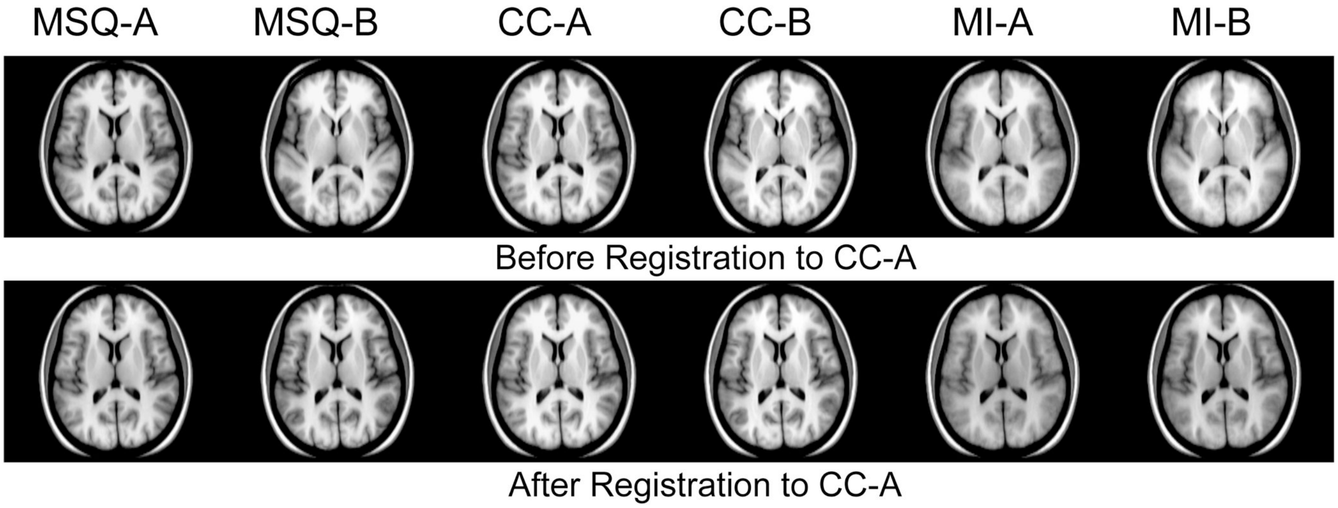
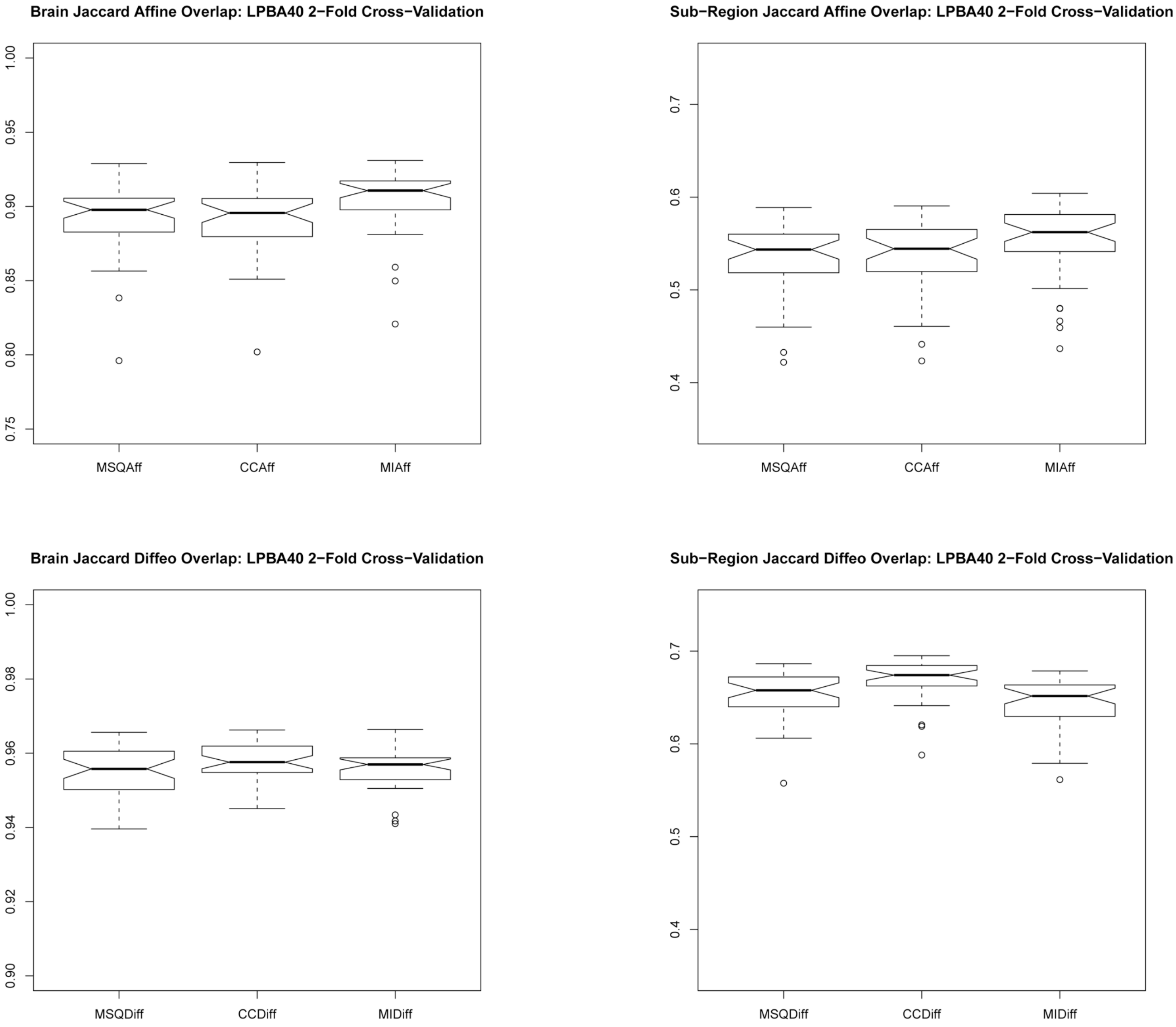
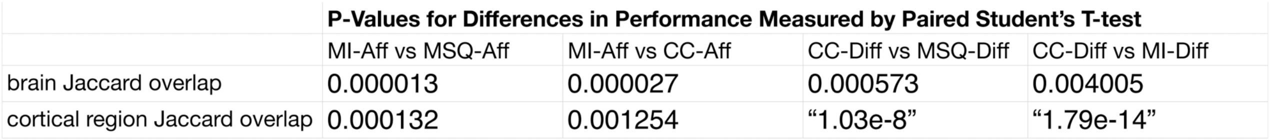


Figure 2.

Two rows of axial slices (neurological convention, i.e. subject left is viewer's left) taken from each of the templates, constructed from subject group A or B, and by (affine, diffeomorphic) registration according to (MI, MSQ), (MI, CC), or (MI, MI). The top row shows the templates before registration to the (MI, CC) group A template and the bottom row shows them after diffeomorphic registration. The high Jaccard overlaps between these templates' label sets quantifies and affirms, from an anatomical perspective, the visual similarity in the appearance of the templates. One may see, in the top row, the relative clarity of the MSQ, CC and MI templates. As template acutance does not strictly increase with our performance evaluation outcome, one may conclude that template clarity, alone, is insufficient to determine the neuroanatomical accuracy of an algorithm.

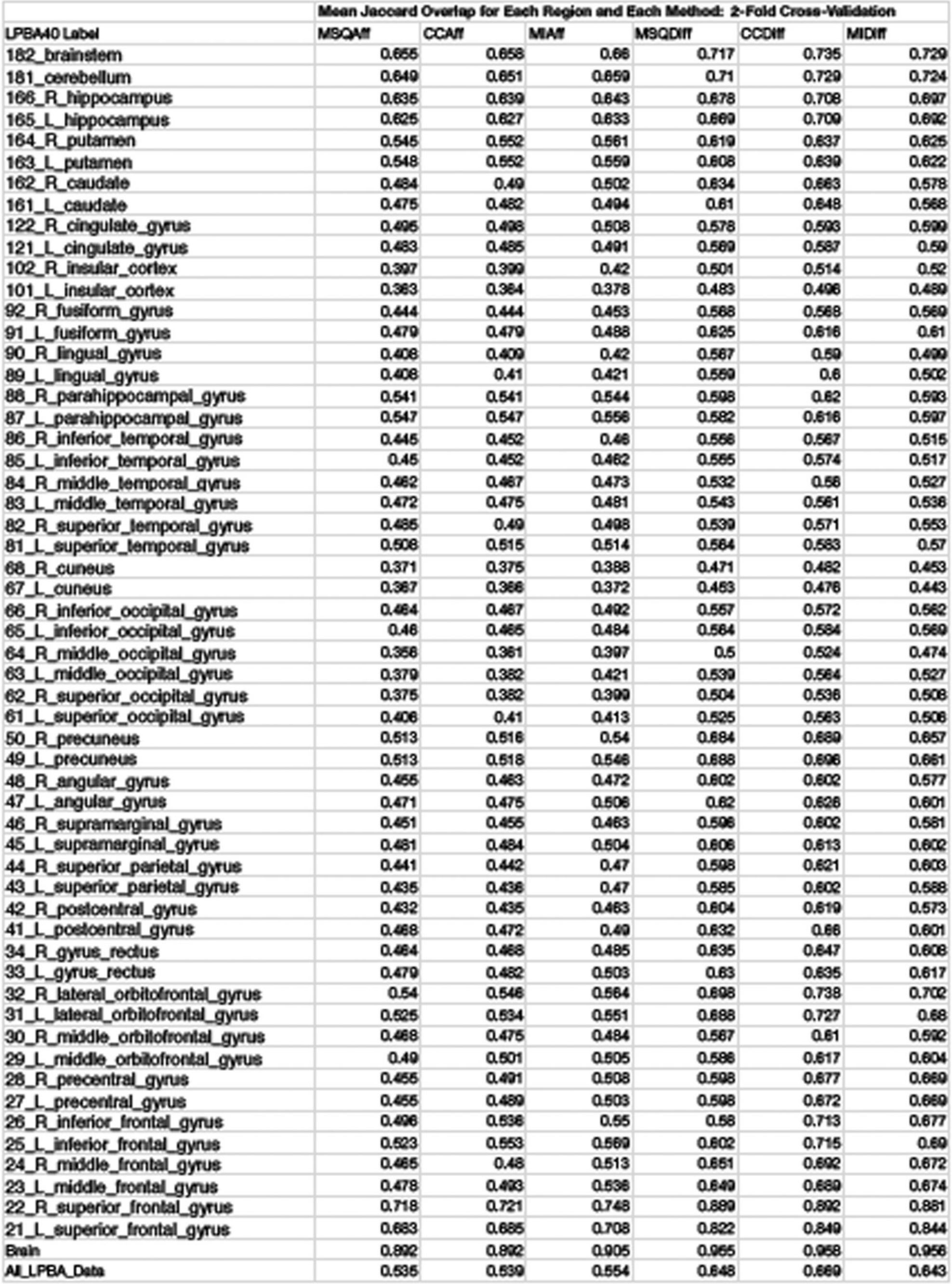




	P-Values for Differences in Performance Measured by Paired Student's T-test			
	MI-Aff vs MSQ-Aff	MI-Aff vs CC-Aff	CC-Diff vs MSQ-Diff	CC-Diff vs MI-Diff
brain Jaccard overlap	0.000013	0.000027	0.000573	0.004005
cortical region Jaccard overlap	0.000132	0.001254	"1.03e-8"	"1.79e-14"

Figure 3.

We use the Jaccard overlap metric (intersection of coregistered labeled regions over their union) to compare performance in this evaluation. The data is visualized with a box and whisker plot, with notches. These plots show the median toward the center of the box. The edges of the box delimit the medians of the data above and below the median. The whiskers and points show the minimum and maximum of the data and any points that are plotted may be considered outliers. We used pairwise Student T-tests to determine whether performance differences are significant. In this figure, the MSQAff, CCAff and MIAff overlap results all report the quality of the affine mapping to the derived template from the (MSQ, MSQ), (CC, CC) and (MI, MI) results. Deformable results for these three runs are not shown. The MI-based affine registration gave the best performance for both brain and cortex labeling thus providing the best initialization for follow-up deformable registration. For this reason, the MSQDiff, CCDiff and MIDiff results all use the MI metric for the affine component and MSQ, CC and MI during deformable registration.



	Mean Jaccard Overlap for Each Region and Each Method: 2-Fold Cross-Validation					
LPBA40 Label	MSQAff	CCAff	MIAff	MSQDiff	CCDiff	MIDiff
182_brainstem	0.655	0.658	0.66	0.717	0.735	0.729
181_cerebellum	0.649	0.651	0.659	0.71	0.729	0.724
166_R_hippocampus	0.635	0.639	0.643	0.678	0.708	0.697
165_L_hippocampus	0.625	0.627	0.633	0.669	0.709	0.692
164_R_putamen	0.545	0.552	0.561	0.619	0.637	0.625
163_L_putamen	0.548	0.552	0.559	0.608	0.639	0.622
162_R_caudate	0.484	0.49	0.502	0.634	0.663	0.578
161_L_caudate	0.475	0.482	0.494	0.61	0.648	0.568
122_R_cingulate_gyrus	0.495	0.498	0.508	0.578	0.593	0.599
121_L_cingulate_gyrus	0.483	0.485	0.491	0.569	0.587	0.59
102_R_insular_cortex	0.397	0.399	0.42	0.501	0.514	0.52
101_L_insular_cortex	0.363	0.364	0.378	0.483	0.496	0.489
92_R_fusiform_gyrus	0.444	0.444	0.453	0.568	0.568	0.569
91_L_fusiform_gyrus	0.479	0.479	0.488	0.625	0.616	0.61
90_R_lingual_gyrus	0.408	0.409	0.42	0.567	0.59	0.499
89_L_lingual_gyrus	0.408	0.41	0.421	0.559	0.6	0.502
88_R_parahippocampal_gyrus	0.541	0.541	0.544	0.598	0.62	0.593
87_L_parahippocampal_gyrus	0.547	0.547	0.556	0.582	0.616	0.597
86_R_inferior_temporal_gyrus	0.445	0.452	0.46	0.556	0.567	0.515
85_L_inferior_temporal_gyrus	0.45	0.452	0.462	0.565	0.574	0.517
84_R_middle_temporal_gyrus	0.462	0.467	0.473	0.532	0.56	0.527
83_L_middle_temporal_gyrus	0.472	0.475	0.481	0.543	0.561	0.536
82_R_superior_temporal_gyrus	0.485	0.49	0.498	0.539	0.571	0.553
81_L_superior_temporal_gyrus	0.508	0.515	0.514	0.564	0.583	0.57
68_R_cuneus	0.371	0.375	0.388	0.471	0.482	0.453
67_L_cuneus	0.367	0.366	0.372	0.463	0.476	0.443
66_R_inferior_occipital_gyrus	0.464	0.467	0.492	0.557	0.572	0.562
65_L_inferior_occipital_gyrus	0.46	0.465	0.484	0.564	0.584	0.569
64_R_middle_occipital_gyrus	0.356	0.361	0.397	0.5	0.524	0.474
63_L_middle_occipital_gyrus	0.379	0.382	0.421	0.539	0.564	0.527
62_R_superior_occipital_gyrus	0.375	0.382	0.399	0.504	0.536	0.508
61_L_superior_occipital_gyrus	0.406	0.41	0.413	0.525	0.563	0.506
50_R_precuneus	0.513	0.516	0.54	0.684	0.689	0.657
49_L_precuneus	0.513	0.518	0.546	0.688	0.696	0.661
48_R_angular_gyrus	0.455	0.463	0.472	0.602	0.602	0.577
47_L_angular_gyrus	0.471	0.475	0.506	0.62	0.626	0.601
46_R_supramarginal_gyrus	0.451	0.455	0.463	0.596	0.602	0.581
45_L_supramarginal_gyrus	0.481	0.484	0.504	0.606	0.613	0.602
44_R_superior_parietal_gyrus	0.441	0.442	0.47	0.598	0.621	0.603
43_L_superior_parietal_gyrus	0.435	0.436	0.47	0.585	0.602	0.588
42_R_postcentral_gyrus	0.432	0.435	0.463	0.604	0.619	0.573
41_L_postcentral_gyrus	0.468	0.472	0.49	0.632	0.66	0.601
34_R_gyrus_rectus	0.464	0.468	0.485	0.635	0.647	0.608
33_L_gyrus_rectus	0.479	0.482	0.503	0.63	0.635	0.617
32_R_lateral_orbitofrontal_gyrus	0.54	0.546	0.564	0.698	0.738	0.702
31_L_lateral_orbitofrontal_gyrus	0.525	0.534	0.551	0.688	0.727	0.68
30_R_middle_orbitofrontal_gyrus	0.468	0.475	0.484	0.567	0.61	0.592
29_L_middle_orbitofrontal_gyrus	0.49	0.501	0.505	0.586	0.617	0.604
28_R_precentral_gyrus	0.455	0.491	0.508	0.598	0.677	0.669
27_L_precentral_gyrus	0.455	0.489	0.503	0.598	0.672	0.669
26_R_inferior_frontal_gyrus	0.496	0.536	0.55	0.58	0.713	0.677
25_L_inferior_frontal_gyrus	0.523	0.553	0.569	0.602	0.715	0.69
24_R_middle_frontal_gyrus	0.465	0.48	0.513	0.651	0.692	0.672
23_L_middle_frontal_gyrus	0.478	0.493	0.536	0.649	0.689	0.674
22_R_superior_frontal_gyrus	0.718	0.721	0.748	0.889	0.892	0.881
21_L_superior_frontal_gyrus	0.683	0.685	0.708	0.822	0.849	0.844
Brain	0.892	0.892	0.905	0.965	0.958	0.956
All_LPBA_Data	0.535	0.539	0.554	0.648	0.669	0.643

Figure 4.

A table representing the data used within the box plots of Figure 3 and also showing region-by-region performance for each method. The correlation of the results in the MIAff column with the (MSQ, CC, MI)Diff columns is (0.921, 0.922, 0.944), suggesting the critical role of the affine initialization.

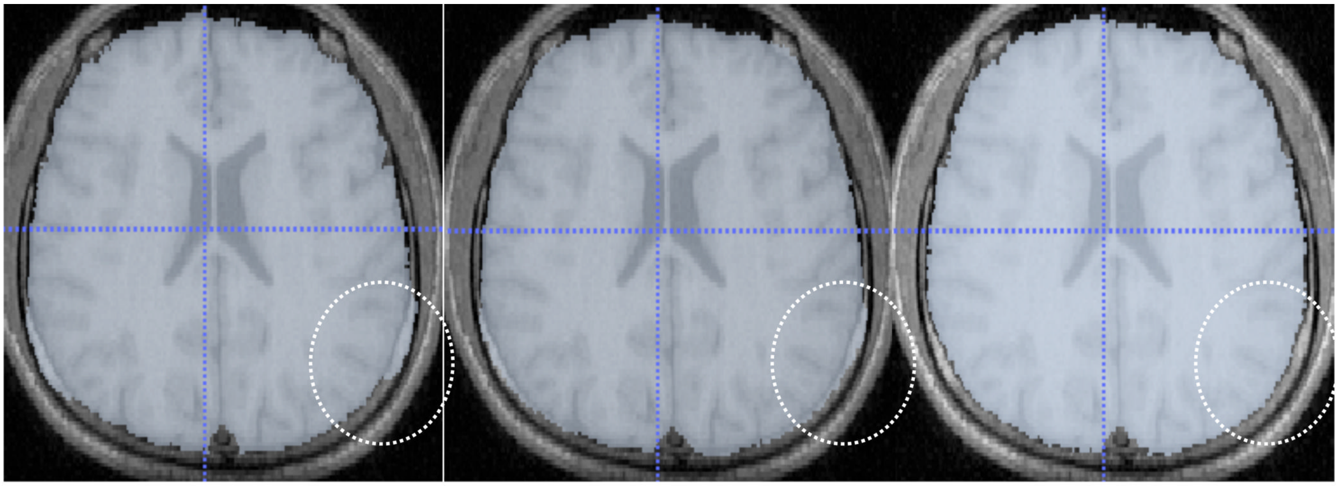


Figure 5.

The Jaccard ratios may be similar but show very different errors, as identified visually. We highlight one region of error in the circled region. The Jaccard values are, from left to right, 0.945078, 0.951205 and 0.952316. The (MI,CC)-based Jaccard mean/sd for diffeomorphic brain extraction, over the full dataset, is: 0.958 ± 0.005. This number is determined from taking the mean and standard deviation of the brain extraction overlaps from mapping the group B to the (MI,CC) group A template and vice versa.

Table 1

Transformations and similarity metrics available in ANTs.

Category	Transformation, ϕ	Similarity Measures	Brief Description
Linear	Rigid†	MSQ, CC, MI	translation and rotation
	Affine†	MSQ, CC, MI	rigid, scaling, and shear
Elastic	Deformable	CC, PR, MI, MSQ, JHCT, PSE	Demons-like algorithm
	DMFFD	CC, PR, MI, MSQ, JHCT, PSE	FFD variant
Diffeomorphic	Exponential	CC,PR, MI, MSQ, JHCT, PSE	minimizes $\upsilon(\mathbf{x})$
	Greedy SyN†	CC, PR, MI, MSQ, JHCT, PSE	minimizes $\upsilon(\mathbf{x}, \mathbf{t})$ locally in time
	Geodesic SyN†	CC, PR, MI, MSQ, JHCT, PSE	minimizes $\upsilon(\mathbf{x}, \mathbf{t})$ over all time

Similarity metric acronyms: MSQ = mean squared difference, CC = cross correlation, PR = CC after subtraction of local mean from the image, MI = mutual information, JHCT = Jensen-Havrda-Charvat-Tsallis divergence, PSE = point-set expectation.

ANTs also provides the inverse of those transformations denoted by the '†' symbol. Only the MSQ, CC and MI metrics are available for both affine and deformable registration and are evaluated here with the Greedy SyN transformation model.

Table 2

The various flags and variables for a variety of image registration possibilities. Additional information can be found on the ANTs website (Avants et al., 2009). The variable Δ represents the gradient step length.

	Argument	Flag	Variables / Sample Parameters
Linear	Iterations	`--number-of-affine-iterations`	N_1xN_2xN_3x…
	Similarity	`--affine-metric-type`	`MI,CC,MSQ`
	Affine or Rigid	`--do-rigid`	`true / false`
Deform.	Image Similarity	`--metric,-m`	`MI,CC,PR,MSQ [  ,param, 1]`
	Point-Set Similarity	`--metric,-m`	`PSE,JHCT [` X, Y`]`
	Iterations/Level	`--iterations,-i`	N_1xN_2xN_3x…
	Regularization	`--regularization,-r`	`Gauss, DMFFD`$[\sigma^2_{gradient}, \sigma^2_{total}]$
	Transformation	`--transformation,-t`	`GreedyExp, Elast, SyN, Exp` [Δ]
	Transformation	`--geodesic`	`SyN` [Δ,# time points,dT]
Misc.	Histogram Match $\mathcal{I}$, $\mathcal{J}$	`--use-histogram-matching`	`1`
	NN Interpolation	`--use-NN`	`0`
	Mask Image	`--mask,-x`	`mask.nii`
	Output Naming	`--output-naming,-o`	`filename.nii`

Table 3

Template stability results across metrics and affine and deformable registration.

Template Stability		
Similarity Metrics	**A → A**	**B → A**
(MI, MSQ) Aff	0.873	0.763
(MI, MSQ) Diff	0.865	0.799
(MI, CC) Aff	1	0.775
(MI, CC) Diff	1	0.815
(MI, MI) Aff	0.880	0.777
(MI, MI) Diff	0.866	0.809