Medical Image Analysis 98 (2024) 103292

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

JOSA: Joint surface-based registration and atlas construction of brain geometry and function

Jian Li [a,b,*], Greta Tuckute [c,d], Evelina Fedorenko [c,d,e], Brian L. Edlow [a,b], Adrian V. Dalca [a,f,1], Bruce Fischl [a,f,1]

[a] A. A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, United States of America
[b] Center for Neurotechnology and Neurorecovery, Department of Neurology, Massachusetts General Hospital and Harvard Medical School, United States of America
[c] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, United States of America
[d] McGovern Institute for Brain Research, Massachusetts Institute of Technology, United States of America
[e] Program in Speech Hearing Bioscience and Technology, Harvard University, United States of America
[f] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, United States of America

ARTICLE INFO

Keywords:
Cortical registration
Semi-supervised learning

ABSTRACT

Surface-based cortical registration is an important topic in medical image analysis and facilitates many downstream applications. Current approaches for cortical registration are mainly driven by geometric features, such as sulcal depth and curvature, and often assume that registration of folding patterns leads to alignment of brain function. However, functional variability of anatomically corresponding areas across subjects has been widely reported, particularly in higher-order cognitive areas. In this work, we present JOSA, a novel cortical registration framework that jointly models the mismatch between geometry and function while simultaneously learning an unbiased population-specific atlas. Using a semi-supervised training strategy, JOSA achieves superior registration performance in both geometry and function to the state-of-the-art methods but without requiring functional data at inference. This learning framework can be extended to any auxiliary data to guide spherical registration that is available during training but is difficult or impossible to obtain during inference, such as parcellations, architectonic identity, transcriptomic information, and molecular profiles. By recognizing the mismatch between geometry and function, JOSA provides new insights into the future development of registration methods using joint analysis of brain structure and function.

1. Introduction

Image registration is a key research topic in medical image analysis. Deformable image registration establishes spatial correspondence between a pair of images via a nonlinear spatial transformation. This transformation is often obtained through an optimization procedure that maximizes a similarity measure. Approaches for registration of the brain are either directly based on 3D images from scans of the brain (volume-based), or based on features resampled onto the brain surface (surface-based) (Maintz and Viergever, 1998), or a combination of these (Postelnicu et al., 2009; Joshi et al., 2009). Surface-based registration of the human cerebral cortex, often referred to as cortical registration, extracts a variety of representative information from the brain images and solves the registration problem based on a surface-matching framework.

Cortical registration methods align the complex folding patterns of the cerebral cortex by matching the geometric features of the brain (Davatzikos and Bryan, 1996; Fischl et al., 1999b). Inter-subject cortical registration can improve not only the statistical power of functional group analyses (van Atteveldt et al., 2004; Frost and Goebel, 2012), but also the predictability of cytoarchitecture of the brain (Yeo et al., 2010b; Fischl et al., 2008). However, cortical registration is a challenging task because of the geometric complexity of the cortex and the large variability among individuals. Inter-subject surface alignment is commonly driven by geometric features that describe measures of cortical folding patterns, such as sulcal depth and mean curvature (Fischl et al., 1999b; Yeo et al., 2010a; Conroy et al., 2013). It is often assumed that an accurate registration of folding patterns will also successfully align brain function (Toga and Thompson, 2001). However, the functional variability of anatomically corresponding areas across subjects has

* Corresponding author at: A. A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, United States of America.
E-mail address: jli112@mgh.havard.edu (J. Li).
[1] Co-senior authors with equal contribution.

https://doi.org/10.1016/j.media.2024.103292
Received 27 October 2023; Received in revised form 21 June 2024; Accepted 30 July 2024
Available online 3 August 2024

been widely reported, particularly in regions implicated in higher-order cognition (Fischl et al., 2008; Frost and Goebel, 2012). This variability implies that regions with different functional specializations may not be optimally aligned with a perfect anatomical registration. There may exist many equally good solutions for geometric registration with substantially different performance in registration of function. Another possible cause is the "mismatch" between geometry and function within each subject, i.e., different individuals may use a slightly different region to process the same function.

Traditional model-based deformable registration has been extensively studied (Fischl et al., 1999b; Yeo et al., 2010a; Sabuncu et al., 2010; Guntupalli et al., 2016; Robinson et al., 2014; Vercauteren et al., 2009; Nenning et al., 2017; Avants et al., 2008; Beg et al., 2005; Christensen et al., 1997; Van Essen et al., 1998; Joshi et al., 2007; Lyu et al., 2013, 2019). Typical strategies employ an iterative approach that seeks an optimal deformation field to warp a moving image to a fixed image. Methods usually involve optimization of a similarity measure between two feature maps, e.g., minimizing mean squared error (MSE) or maximizing normalized cross correlation, while regularizing the deformation field to have some desired property, such as smoothness and/or diffeomorphism. Widely used techniques for cortical surface registration map the surface onto the unit sphere and establish correspondence between feature maps in the spherical space (Fischl et al., 1999a). Conventional approaches, such as FreeSurfer (Fischl et al., 1999b), register an individual subject to a probabilistic population atlas by minimizing the geometry MSE weighted by the inverse variance of the atlas convexity, in a maximum *a posteriori* formulation. These anatomical registration methods have been adapted to functional registration by minimizing MSE on functional connectivity computed from functional magnetic resonance imaging (fMRI) data (Sabuncu et al., 2010). Spatial correspondence can also be maximized by finding local orthogonal transforms that linearly combine features around each local neighborhood (Guntupalli et al., 2016, 2018).

Discrete optimization approaches iteratively align local features using spherical meshes from low-resolution to high-resolution (Robinson et al., 2014, 2018). Similar hierarchical approaches use features projected to the spherical harmonics bases (Lyu et al., 2019). Instead of whole-brain geometric features, landmarks, such as sulcal curves, have also been used to drive the registration (Van Essen, 2005; Lyu et al., 2013). To encourage invertibility of the deformation field and preserve anatomical topology, diffeomorphic registration uses an exponentiated Lie algebra, most often assuming a stationary velocity field (SVF) (Ashburner, 2007; Vercauteren et al., 2009). These strategies were extended to the sphere by regularizing the deformation using spherical thin plate spline interpolation (Yeo et al., 2010a). Several methods align functional regions, for example, using Laplacian eigen embeddings computed from fMRI data (Nenning et al., 2017). These methods are successful but solve an optimization problem for each image pair, resulting in a high computational cost.

Increasingly popular learning-based registration methods can be categorized into supervised (Krebs et al., 2017; Sokooti et al., 2017; Yang et al., 2016; Cao et al., 2017) and unsupervised (Dalca et al., 2018a; Balakrishnan et al., 2019; Cheng et al., 2020a; Niethammer et al., 2019; Krebs et al., 2019; de Vos et al., 2019) registration. Supervised registration evaluates the performance of a registration network by comparing the predicted deformation with a "ground truth" deformation. However, due to the lack of the "ground truth", deformation generated by iterative methods or simulation is often used as the target, which fundamentally limits the performance of supervised approaches (Sokooti et al., 2017; Yang et al., 2016; Cao et al., 2017). Further, the iterative registration methods are relatively slow which makes it impractical to generate large numbers of samples for supervised training, often resulting in overfitting. In contrast, unsupervised registration methods predict a deformation field that is used to warp the moving image to the fixed image. The core idea of these methods is to employ a classical loss in the image space, thus forming an end-to-end training pipeline (de Vos et al., 2019; Balakrishnan et al., 2019; Hoffmann et al., 2022; Mok and Chung, 2020). Semi-supervised methods employ additional information, such as segmentation or parcellation maps, to guide registration without requiring them during inference (Balakrishnan et al., 2019; Blendowski et al., 2021). Recent methods extend this strategy to the spherical domain by parameterizing the brain surfaces in a 2D grid that accounts for distortions (Cheng et al., 2020a) or directly on the sphere using spherical kernels (Zhao et al., 2021b,a). These learning-based registration methods substantially improved registration speed by orders of magnitude at inference while achieving a superior or comparable registration accuracy relative to the iterative methods. However, these methods do not directly model structure-function mismatches within a subject.

In this work, we present a diffeomorphic cortical registration framework that

- explicitly uses functional data to drive the geometric registration to optimally align function,
- simultaneously learns multi-modal population-specific atlases during training,
- employs a semi-supervised training strategy that uses task fMRI data to improve functional registration but without requiring task fMRI data during inference.

Specifically, we build on recent unsupervised spherical registration strategies (Balakrishnan et al., 2019; Cheng et al., 2020a) and use a joint deformation field shared by geometry and function to capture the relatively large differences among subjects. We introduce deformation fields that describe relatively small variations between geometry and function within each subject. The learned atlases avoid the biases towards the initial subject used for template construction in iterative approaches. In contrast to the term "unsupervised" commonly used in the literature (Balakrishnan et al., 2019; Cheng et al., 2020a), here we borrow the term "semi-supervised" to highlight the strategy where auxiliary information, such as functional data, can be incorporated during training but are *not* required during inference. This semi-supervised training framework can also be easily extended to any auxiliary data that could be helpful to guide spherical registration but is difficult or impossible to obtain or undesirable to use during inference, such as parcellations, architectonic identity, transcriptomic information, and molecular profiles. In many cases, even functional MRI data are often not available. For instance, many clinical scans only have a T1-weighted image, however, estimation of functional properties of the patients' brain is still highly desired. More importantly, to study the functional properties of the human brain, we need an unbiased way to perform inter-subject registration that does not involve functional data. At inference, JOSA registration is solely based on the geometry of the subject surface, avoiding the potential circularity issue in any subsequent functional analyses, which would preclude unbiased group statistics otherwise. We demonstrate that the proposed framework yields improved registration performance in both anatomical and functional domains.

We presented a preliminary version of this method at the Medical Imaging with Deep Learning 2023 conference (Li et al., 2024). The current paper provides a more detailed description of the method, new and substantially expended experimental results with a comparison to an extensive list of state-of-the-art methods, as well as detailed ablation studies.

2. Methods

We describe **Jo**int **S**pherical registration and **A**tlas building (**JOSA**), a method for surface-based cortical registration with simultaneous atlas construction that explicitly models the cross-subject variation in the relationship between anatomy and functional properties in each subject.

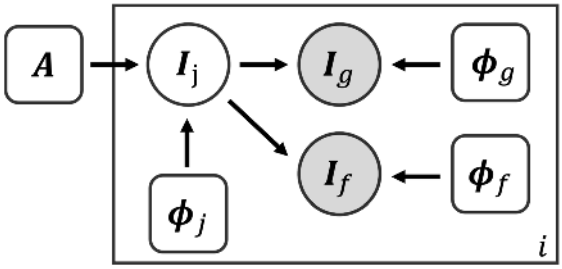

Fig. 1. Graphical representation of the generative model. Circles are random variables. Rounded squares indicate parameters. Shaded quantities are observations. The big plate represents replication. $\boldsymbol{A}$ represents the global atlas, $\boldsymbol{I}$ the input image, $\boldsymbol{\phi}$ deformation field. The subscript j, g, and f stand for joint, geometry, and function, respectively.

2.1. Generative model

Fig. 1 shows the graphical representation for the proposed generative model. Let $\boldsymbol{A}$ be an unknown population atlas with all geometric and functional cortical features of interest. We propose a generative model that describes the formation of the subject geometric $\boldsymbol{I}_g$ and functional $\boldsymbol{I}_f$ features by first warping the atlas $\boldsymbol{A}$ by a subject deformation field $\boldsymbol{\phi}_j$. This model characterizes the differences between subjects and results in a joint multi-feature image $\boldsymbol{I}_j$. Geometric feature $\boldsymbol{I}_g$ is formed given an additional field $\boldsymbol{\phi}_g$ that deforms the geometric features in $\boldsymbol{I}_j$, and similarly for $\boldsymbol{I}_f$ and $\boldsymbol{\phi}_f$. The separation of $\boldsymbol{\phi}_g$ from $\boldsymbol{\phi}_f$ enables us to explicitly model structure-function variability across subjects. This is of critical importance from a neuroscientific perspective because some functional areas are better predicted by folding patterns than others (Fischl et al., 2008).

2.1.1. Deformation priors

Let $\boldsymbol{\phi}_j^i, \boldsymbol{\phi}_g^i, \boldsymbol{\phi}_f^i$ be the joint, geometric, and functional deformation fields for each subject i, respectively. All variables in the model are subject-specific, except for the global atlas $\boldsymbol{A}$, and we omit i for our derivation. We impose the deformation priors

$$\begin{aligned} p(\boldsymbol{\phi}_j) &\sim \exp\{-(\lambda_j \left\|\nabla \boldsymbol{u}_j\right\|^2 + \alpha_j \left\|\bar{\boldsymbol{u}}_j\right\|^2)\} \\ p(\boldsymbol{\phi}_g) &\sim \exp\{-\lambda_g \left\|\nabla \boldsymbol{u}_g\right\|^2\} \\ p(\boldsymbol{\phi}_f) &\sim \exp\{-\lambda_f \left\|\nabla \boldsymbol{u}_f\right\|^2\} \end{aligned} \tag{1}$$

where $\boldsymbol{u}_j$ is the spatial displacement for $\boldsymbol{\phi}_j = Id + \boldsymbol{u}_j$, $\nabla \boldsymbol{u}_j$ is its spatial gradient, and $\bar{\boldsymbol{u}}_j = 1/N \sum_i \boldsymbol{u}_j^i$, N is the number of subjects, and similarly for $\boldsymbol{u}_g$ and $\boldsymbol{u}_f$. The gradient term encourages smooth deformations, while the mean term encourages an unbiased atlas $\boldsymbol{A}$ by penalizing the average deformation over the entire dataset and hence favoring atlases that are "close" to every subject in the training (Dalca et al., 2019).

2.1.2. Data likelihood

We treat the latent joint image $\boldsymbol{I}_j$ as a noisy warped atlas,

$$p(\boldsymbol{I}_j|\boldsymbol{\phi}_j; \boldsymbol{A}) = \mathcal{N}(\boldsymbol{I}_j; \boldsymbol{\phi}_j \circ \boldsymbol{A}, \sigma^2 \mathbb{I}) \tag{2}$$

where $\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, $\circ$ represents spatial transformation, σ represents additive noise, and $\mathbb{I}$ is the identity matrix. The geometric feature image $\boldsymbol{I}_g$ is then a noisy observation of a further-moved joint image $\boldsymbol{I}_j$:

$$p(\boldsymbol{I}_g|\boldsymbol{\phi}_g, \boldsymbol{I}_j) = \mathcal{N}(\boldsymbol{I}_g; \boldsymbol{\phi}_g \circ \boldsymbol{I}_j, \sigma^2 \mathbb{I}). \tag{3}$$

Therefore, the complete **geometric image likelihood** is given by:

$$\begin{aligned} p(\boldsymbol{I}_g|\boldsymbol{\phi}_g, \boldsymbol{\phi}_j; \boldsymbol{A}) &= \int_{\boldsymbol{I}_j} p(\boldsymbol{I}_g|\boldsymbol{\phi}_g, \boldsymbol{I}_j) p(\boldsymbol{I}_j|\boldsymbol{\phi}_j; \boldsymbol{A}) \\ &= \int_{\boldsymbol{I}_j} \mathcal{N}(\boldsymbol{I}_g; \boldsymbol{\phi}_g \circ \boldsymbol{I}_j, \sigma\mathbb{I}) \mathcal{N}(\boldsymbol{I}_j; \boldsymbol{\phi}_j \circ \boldsymbol{A}, \sigma\mathbb{I}) \\ &= \int_{\boldsymbol{I}_j} \mathcal{N}(\boldsymbol{I}_j; \boldsymbol{\phi}_g^{-1} \circ \boldsymbol{I}_g, \sigma\mathbb{I}) \mathcal{N}(\boldsymbol{I}_j; \boldsymbol{\phi}_j \circ \boldsymbol{A}, \sigma\mathbb{I}) \\ &\overset{*}{=} \int_{\boldsymbol{I}_j} \mathcal{N}(\boldsymbol{I}_j; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \mathcal{N}(\boldsymbol{\phi}_g^{-1} \circ \boldsymbol{I}_g; \boldsymbol{\phi}_j \circ \boldsymbol{A}, 2\sigma^2 \mathbb{I}) \\ &= \mathcal{N}(\boldsymbol{\phi}_g^{-1} \circ \boldsymbol{I}_g; \boldsymbol{\phi}_j \circ \boldsymbol{A}, 2\sigma^2 \mathbb{I}) \\ &= \mathcal{N}(\boldsymbol{I}_g; \boldsymbol{\phi}_g \circ \boldsymbol{\phi}_j \circ \boldsymbol{A}, 2\sigma^2 \mathbb{I}) \end{aligned} \tag{4}$$

where in $*$ we used an identity of the product of two Gaussian distributions, and $\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c$ are constants. We use a similar model for the functional image $\boldsymbol{I}_f$.

2.1.3. Learning

Let $\boldsymbol{\Phi} = \{\boldsymbol{\phi}_j, \boldsymbol{\phi}_g, \boldsymbol{\phi}_f\}$ and $\boldsymbol{I} = \{\boldsymbol{I}_g, \boldsymbol{I}_f\}$. We estimate $\boldsymbol{\Phi}$ by minimizing the negative log likelihood

$$\begin{aligned} \mathcal{L}(\boldsymbol{\Phi}|\boldsymbol{I}; \boldsymbol{A}) &= -\log p(\boldsymbol{\Phi}|\boldsymbol{I}; \boldsymbol{A}) = -\log p(\boldsymbol{I}|\boldsymbol{\Phi}; \boldsymbol{A}) - \log p(\boldsymbol{\Phi}) \\ &= -\log \prod_{k\in\{g,f\}} p(\boldsymbol{I}_k|\boldsymbol{\phi}_j, \boldsymbol{\phi}_k, \boldsymbol{A}) - \log \prod_{k\in\{j,g,f\}} p(\boldsymbol{\phi}_k) \\ &= \frac{1}{2\sigma^2}\left(\left\|\boldsymbol{I}_g - \boldsymbol{\phi}_g \circ \boldsymbol{\phi}_j \circ \boldsymbol{A}\right\|^2 + \left\|\boldsymbol{I}_f - \boldsymbol{\phi}_f \circ \boldsymbol{\phi}_j \circ \boldsymbol{A}\right\|^2\right) \\ &\quad + \lambda_j \left\|\nabla \boldsymbol{u}_j\right\|^2 + \lambda_g \left\|\nabla \boldsymbol{u}_g\right\|^2 + \lambda_f \left\|\nabla \boldsymbol{u}_f\right\|^2 + \alpha_j \left\|\bar{\boldsymbol{u}}_j\right\|^2 \\ &\quad + \text{const}. \end{aligned} \tag{5}$$

The first two terms are data fidelity in the subject and atlas space respectively. The remaining four terms regularize the gradient and mean of the spatial displacement, encouraging smooth and unbiased deformation fields.

2.2. Neural network and semi-supervised training strategy

We use a neural network to approximate the function $h_{\theta,\boldsymbol{A}}(\boldsymbol{I}) = \boldsymbol{\Phi}$, where θ and $\boldsymbol{A}$ are network parameters. Fig. 2 shows the proposed network architecture. To work with surface-based data, the cortical surface of each subject is inflated into a sphere and then rigidly registered to an average space using FreeSurfer (Fischl, 2012). Geometric and functional features are parameterized onto a 2D grid using a standard conversion from Cartesian coordinates to spherical coordinates, resulting in a 2D image for each input (Cheng et al., 2020a). We use rigid rotation as one of the preprocessing steps to ensure the zero latitude and longitude of the spheres are reasonably well aligned across subjects.

The network takes such a parameterized geometric image as input and outputs *three* deformation fields in the parameterized coordinates. The joint deformation $\boldsymbol{\phi}_j$, which models the relatively large inter-subject variance, is shared among and composed with individual deformations $\boldsymbol{\phi}_g$ and $\boldsymbol{\phi}_f$.

We learn network parameters and use task fMRI data in a *semi-supervised* manner. As shown in the green block in Fig. 2, the task fMRI data and the corresponding functional atlas are not input into the neural network. Rather they are used only for evaluating the functional terms in the loss function (5). This obviates the need for functional data during inference, as the deformation fields can be inferred using only geometric features. The proposed framework is flexible in the sense that auxiliary data, perhaps from different modalities, can be integrated into the framework for simultaneous multi-modality registration.

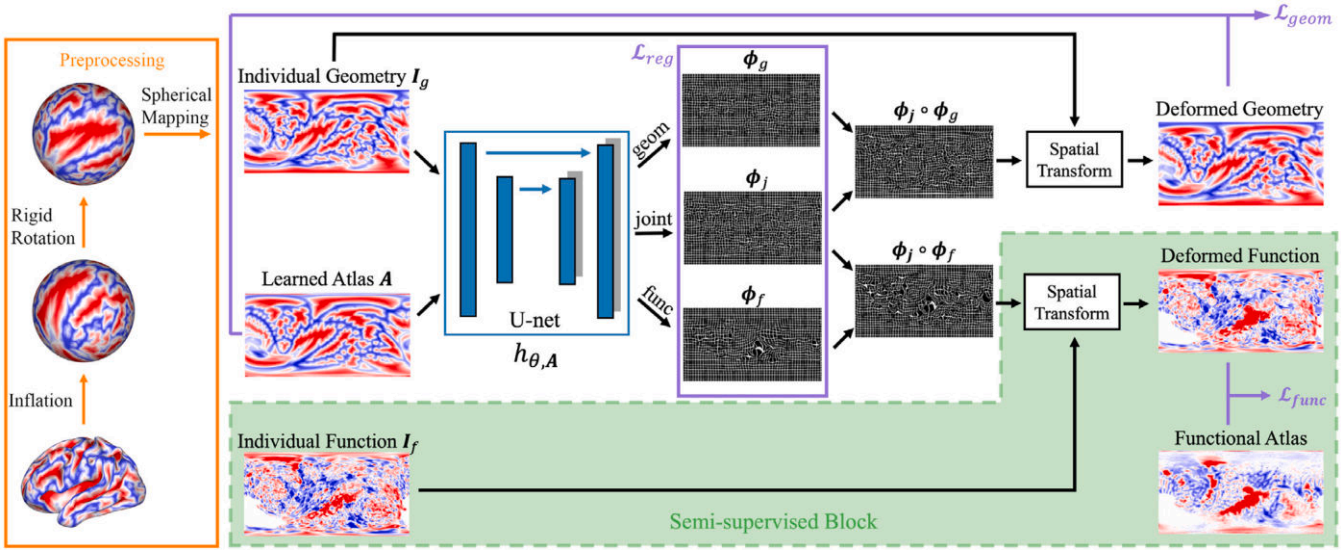

Fig. 2. Network architecture and preprocessing pipeline. The network takes the geometric features from the subject and outputs one joint and two separate deformation fields. The joint and geometric deformation are composted for registration of folding patterns, and likewise for function. Task fMRI data were used for evaluating the functional loss only in a semi-supervised manner, hence, functional task data are not required during inference. The losses $\mathcal{L}_{geom}$ and $\mathcal{L}_{func}$ represent the geometric part and the functional part of the data fidelity terms in Eq. (5). They are evaluated in the atlas and subject space, which also helps avoid atlas drift (Aganj et al., 2017) during atlas construction. $\mathcal{L}_{reg}$ represents the regularization and centrality terms which encourages smooth deformations and an unbiased estimation of the atlas.

2.3. Implementation

We implemented a Unet-like (Ronneberger et al., 2015) network based on the core architecture in VoxelMorph (https://voxelmorph.net) (Balakrishnan et al., 2019; Dalca et al., 2019). We used a 5-layer encoder with [128, 256, 384, 512, 640] filters and a symmetric decoder followed by 2 more convolutional layers with [64, 32] filters. Each layer involves convolution, max-pooling/down-sampling, and LeakyReLU activation. The spherical parameterization leads to denser sampling grids for regions at higher latitudes. We used an L2 loss for geometric features and a correlation loss for functional features. To account for the difference in sampling density, we performed prior and distortion corrections to the losses described in SphereMorph (Cheng et al., 2020a). In short, weights proportional to $\sin(\theta)$, where θ is the elevation, were used to correct the distortion. Recent work also found that varying the locations of the poles in the projections did not meaningfully impact the registration (Cheng et al., 2020a). The parameterized images were standardized identically but separately for structural and functional features, where the median was subtracted for each feature image followed by a division by the standard deviation. To account for the discontinuity across the boundaries of the parameterized map, we implemented a padding strategy, consisting of a reflection padding after a 180-degree circular shift on the top and bottom of the image (north pole and south pole) and a circular padding on the left and right. This enables information to flow continuously across the borders, just as it does on spheres.

During training, we randomly sampled the training data into mini-batches with a batch size of 8. For each batch, we augmented the data by adding Gaussian random deformations with a maximum $\sigma = 8$ and a proper spherical topology and distortion correction at each spatial location. We further augmented the data by adding Gaussian noise with $\sigma = 1$ for geometric features and $\sigma = 6$ for functional features. We used the Adam optimizer (Kingma and Ba, 2014) with an initial learning rate of 10^{-3}. The learning rate was scheduled to decrease linearly to 10^{-4} within the first 500 epochs and then reduced by a factor of 0.9 if the validation loss does not decrease after every 100 epochs. The relative weights between functional loss and geometric loss were set to 0.7:0.3, and the relative weights among different geometric features (`sulc`, `curv`, `inflated.H`) were set to be 1:1:0.1 to emphasize more on the detailed features. We set the regularization hyperparameter λ_j in Eq. (5) for the joint (large) deformation to be 0.05 and λ_g and λ_f for the individual (small) deformations to be 0.1. These hyperparameters were chosen empirically based on the performance on the validation set, which is independent from the test set.

The atlases, as part of the network parameters, were initialized using Gaussian random noise and automatically learned during training. We exploited a three-stage training strategy to further avoid atlas expansion or drift (Aganj et al., 2017): (1) we performed a training run using a single deformation with geometric features only, so that the network learned a geometric atlas; (2) we used and fixed the geometric atlas obtained in step one and trained a JOSA network to learn the functional atlas; and (3) we fixed both the geometric and the functional atlas from the previous two steps and performed a final training for the composite deformation fields.

We used TensorFlow (Abadi et al., 2016) with Keras front-end (Chollet, 2018) and the Neurite package (Dalca et al., 2018b), and all experiments were conducted in a Dell Workstation with dual Intel Xeon Silver 6226R CPUs and an Nvidia RTX6000 GPU. The source code and the trained model of JOSA will be released to the public at https://voxelmorph.net as well as integrated as part of FreeSurfer (Fischl, 2012) package.

3. Experiments

3.1. Language task fMRI data

In this work, we used language task fMRI localizer maps. Rich literature characterizes the language network through individual- and group-level analyses (e.g., Fedorenko et al. (2010), Mahowald and Fedorenko (2016), Lipkin et al. (2022), Lee et al. (2024), Tuckute et al. (2024)). We used data from a large-scale language mapping study (Lipkin et al., 2022) ($N = 800$), where the language network was functionally localized using a task that contrasts reading/listening of sentences versus a perceptually matched control (such as strings

of nonwords or degraded speech), collected using a standard blocked design. Ten different versions of the language localizer task were used (Table 2 in Lipkin et al. (2022)), with the vast majority of subjects (~77%) completing a version that contrasted reading of sentences with nonwords strings. In this localizer version, each stimulus was presented one word/nonword at a time at the rate of 450 ms per word/nonwords (12 words/nonwords per stimulus). Each stimulus was preceded by a 100 ms blank screen and followed by a 400 ms screen showing a picture of a finger pressing a button, and a blank screen for another 100 ms, for a total trial duration of 6 s. Experimental blocks lasted 18 s, and fixation blocks lasted 14 s Each run (consisting of 5 fixation blocks and 16 experimental blocks) lasted 358 s. Subjects completed 2 runs. Subjects were instructed to read attentively and press a button whenever they saw the finger-pressing picture on the screen. Structural and functional data were collected on a 3T Siemens Trio scanner with a 12-channel (G1: $N = 6$, G2: $N = 12$) or a 32-channel (G3: $N = 782$) head coil, at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT. T1-weighted images were collected in 176 sagittal slices with 1 mm isotropic voxels. Functional data (BOLD) were acquired using an EPI sequence with the following parameters: 33 (G1) or 31 (G2, G3) 4 mm thick near-axial slices, 3.0 mm × 3.0 mm (G1) or 2.1 mm × 2.1 mm (G2, G3) in-plane resolution, FoV in the phase encoding (A ≪ P) direction 192 mm (G1) or 200 mm (G2, G3), matrix size 64 × 64 (G1) or 96 × 96 (G2, G3), TR = 2000 ms, and TE = 30 ms.

We preprocessed the data using FreeSurfer v6.0.0 as described in Lipkin et al. (2022). We sampled data onto the FreeSurfer average space (fsaverage), motion corrected, registered using the middle time point of each run, and spatially smoothed with a 4 mm FWHM Gaussian filter. We reconstructed the subjects' surfaces from the T1 images (default *recon-all* parameters). A "sentences vs. control" contrast t-map was originally generated for each subject using first-level GLM analysis based on the blocked design in the fsaverage space (Lipkin et al., 2022). We resampled the t-maps back onto the individual surfaces using FreeSurfer's 'mri_surf2surf' utility as functional features for registration in this work. We randomly split the data into a training set with 600 subjects, a validation set with 100 subjects, and a test set with the remaining 100 subjects.

3.2. Baseline

We compared the registration performance of JOSA to FreeSurfer (Fischl et al., 1999b), Spherical Demons (SD) (Yeo et al., 2010a), Multimodel Surface Matching (MSM) (Robinson et al., 2014), the improved version of MSM with higher-order constraint (MSM-HOC) (Robinson et al., 2018), and SphereMorph (Cheng et al., 2020a) as surface registration baselines. For all baseline methods, we used three geometric features that quantify the folding patterns of the cortex to drive the registration, namely the mean curvature of the inflated surface (`?h.inflated.H`), the sulcal depth map (`?h.sulc`) and the mean curvature (`?h.curv`) of the non-inflated white matter surface, all available as part of the FreeSurfer standard outputs. For FreeSurfer registration, we ran `mris_register` (https://surfer.nmr.mgh.harvard.edu/fswiki/mris_register) for each test subject to register them to the fsaverage. For SD, we ran the pre-compiled MATLAB scripts provided by the authors (https://sites.google.com/view/yeolab/software/sphericaldemonsrelease) to register subjects to fsaverage using their default settings. For MSM and MSM-HOC, we ran the authors' pre-built `msm` executable (https://github.com/ecr05/MSM_HOC) to register each subject to fsaverage with the configurations used in their 2014 and 2018 papers, respectively. For a fair comparison, both versions of MSM were executed in a hierarchical manner such that `?h.inflated.H` was used for initial coarse alignment, then refined by `?h.sulc` and `?h.curv`, consecutively. For SphereMorph (https://voxelmorph.net), we used fsaverage as the registration target and trained the network to predict a single deformation field purely based on the three geometric features. We then used the predicted deformation to warp each subject's functional data to the fsaverage space at the inference time. All baseline methods used fsaverage as their registration target (for MSM-based method it is FS_LR164k, which is a left–right symmetric version of fsaverage). Therefore, the hyperparameters provided by the authors of each method have been optimized for the atlas.

3.3. Evaluation

Qualitatively, we computed the group mean images for the geometric features in the test set for each registration method. For the functional data, we adopted the "top X%" approach commonly used in group task fMRI studies (Lipkin et al., 2022). The functional data for each subject were first thresholded at the top 10% of the t-value map, resulting in a binary map (task-active or inactive) at the individual level. At the group level, for each vertex, we computed the number of task-active subjects over the total number of subjects, yielding a percentage map, which represents the proportion of the 800 subjects for whom that vertex falls in the 10% of vertices across the brain with the highest t-values for sentences > control contrast. This top 10% approach accounts for inter-subject variability in the overall strength of the BOLD signal responses due to trait or state factors, thus eliminating the impact of magnitude of the responses, i.e., a sharper/stronger response will not necessarily yield a higher percentage map. We also computed this percentage map using the test set for each registration method. We visualized the results by superimposing the functional percentage map with the curvature group mean map.

We quantified the geometric registration accuracy using the correlation between the registered individual data and the group mean (Cheng et al., 2020a). Specifically, we computed the Pearson correlation $c_k = corr(\boldsymbol{I}_k, \bar{\boldsymbol{I}})$ between the individual image $\boldsymbol{I}_k$ to the group mean image $\bar{\boldsymbol{I}} = 1/N \sum_k \boldsymbol{I}_k$ for subject k, where N is the number of subjects. We quantified the functional registration performance using the size of the overlapping regions between the suprathreshold individual functional maps at the top 10% of the t-values and suprathreshold percentage map at the group level. We expect a better registration will have a larger overlap with the group map for a reasonably conservative threshold. This also helps avoid a potential evaluation bias when a single-task contrast is used where only a fraction of the surface area is active during tasks. Specifically, we computed $s_k = |\boldsymbol{I}_k^{top10} \bigcap \hat{\boldsymbol{I}}^{th}|$, where $\boldsymbol{I}_k^{top10}$ is the top 10% suprathreshold individual image for subject k, $\hat{\boldsymbol{I}}^{th}$ is the suprathreshold percentage map at the group level at the threshold th, and $|\cdot|$ is the measure of cardinality. We used an empirical threshold $th = 0.75$ for thresholding the group percentage map in this evaluation based on the t-values in the language active regions.

To evaluate the consistency in registration improvement across subjects, we assessed the pair-wise geometric correlation difference $c^k_{Method} - c^k_{Rigid}$ between each of FreeSurfer, SD, MSM, MSM-HOC, SphereMorph, and JOSA and the rigid alignment (i.e., the correlation difference for the *same* subject after registration in comparison to the rigidly aligned version). We similarly assessed the pair-wise functional difference $s^k_{Method} - s^k_{Rigid}$. We statistically tested the median difference between JOSA and the baseline methods using the Wilcoxon Signed Rank test as they are paired measures.

In the quantitative analysis of the functional data, there are two thresholds used for evaluations: one is the top X% and the other is the th for the group thresholding (agreement among subjects). To investigate the impact of the two thresholds to the robustness of the registration performance, we varied one threshold while fixing the other. Specifically, we computed the total number of task-active vertices that agreed among at least 75% of the subjects as a function of the top X% for $X \in \{5, \dots, 15\}$. We also computed the total number of task-active vertices that have a t-value within the top 10% of each individual subject as a function of the agreement threshold $th \in \{0.6, \dots, 0.9\}$.

To demonstrate the impact of the regularization hyperparameters to the registration performance, we ran additional JOSA experiments identical in all aspects except for the regularization weights.

To further illustrate the benefit of JOSA framework, we conducted two additional experiments. We first used the MindBoggle101 dataset (Klein and Tourville, 2012) to demonstrate registration of parcellation. For each subject in the MindBoggle101 dataset, we warped its Desikan-Killiany-Tourville (DKT)-based manual labels to the atlas space using the geometric deformation predicted by the model trained on the language dataset. All subjects from this MindBoggle101 dataset were used for test purpose only (i.e., none of them was included in the training). We evaluated the performance using the Dice score between all pairs of subjects after registration in the atlas space. Second, we tested how well an individual language map from an unseen subject could be "predicted" by the trained registration model. To do so, specifically, for each registration method, we "trained" the model by averaging over the training subjects after registration to the common space (learned atlas for JOSA and fsaverage for all others). We "predicted" individual language response by warping the trained average back to the individual space for each testing subject. We evaluated the performance by computing the overlap between the predicted response and the actual t-value maps for the testing subjects using the same top 10% approach as described above. We statistically tested the difference between JOSA and the baseline methods using the Wilcoxon Signed Rank test.

3.4. Computational efficiency

We compared the computational efficiency of the registration methods by measuring the run time of the registration procedure for each of the test subjects. Iterative methods (FreeSurfer, SD, MSM, and MSM-HOC) were run on a single CPU thread, while the learning-based methods (SphereMorph and JOSA) were evaluated using both a CPU and a single GPU.

3.5. Stability of atlas construction

Atlas initialization can substantially bias the result (Fischl et al., 1999b; Dickie et al., 2017; Cheng et al., 2020b). To investigate the stability and the potential bias of atlas construction using JOSA, we repeated the geometric atlas learning procedure 6 times, each time with randomly selected batches of subjects and their orders through the training process. We qualitatively visualized the difference between each run to the average of all runs. To quantitatively measure the stability of the learned atlases, for each pair of the learned atlases with the 6 trials (hence 15 pairs in total), we computed Pearson correlation between the pairs of atlases.

3.6. Ablation studies

We conducted the following two ablation studies: (1) To illustrate benefit of learning the population-specific and unbiased atlas, we trained two JOSA networks with the only difference being in the registration target, where we used rigid average and FreeSurfer average for comparison to our learned atlas; (2) To assess the effect of modeling the difference between geometry and function, we trained two additional JOSA networks, one with a single (joint) deformation field that was used to align both geometry and function, and the other with two individual deformation fields separately for geometry and function but without the joint deformation. Statistical testing were conducted between JOSA and ablated experiments using the Wilcoxon Signed Rank test, for geometry and function separately.

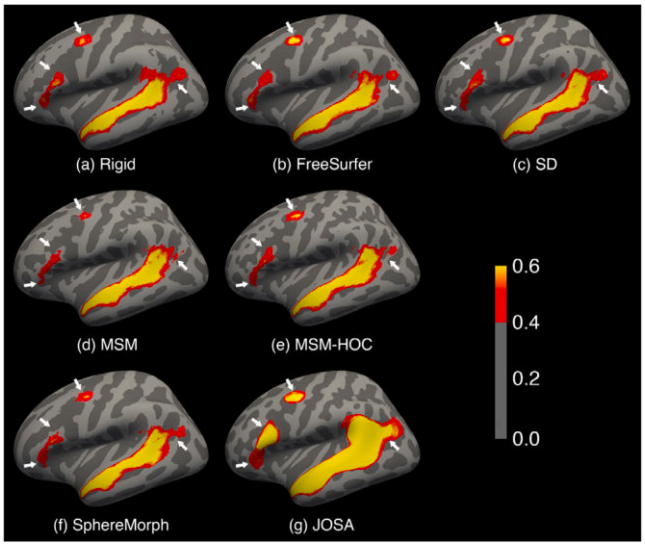

Fig. 3. Group average curvature map of 100 test subjects overlaid with the functional percentage map of language activation. (a) Rigid; (b) SD; (c) FreeSurfer; (d) MSM; (e) MSM-HOC; (f) SphereMorph; (g) JOSA. The language activation map is defined based on the contrast between well-formed sentences and a perceptually matched control stimulus. In the curvature maps, dark gray indicates sulci and light gray indicates gyri. The color bar for functional data (agreement percentage among subjects) is shown on the bottom right. White arrows highlight the key regions in functional alignment where large functional variability across subjects is often observed. JOSA achieved a substantially better alignment in both folding patterns and function with larger functional agreement among subjects.

4. Results

4.1. Qualitative comparison

Fig. 3 shows the group mean curvature maps (one of the geometric features) of the 100 test subjects overlaid with the functional percentage maps of language activation for each of the registration methods. A better registration leads to a larger and higher percentage of agreement across subjects. FreeSurfer, SD, MSM, MSM-HOC, and SphereMorph significantly improved the registration of the folding patterns compared to the rigid alignment, albeit with some noticeable but minor differences in the average curvature patterns. However, all these methods only yield marginal improvement of the functional alignment. In contrast, JOSA achieved a substantially better alignment in both folding patterns and function. In particular, the predominant language region in the superior temporal gyrus shows a substantially higher and larger agreement across subjects. We also find additional/higher agreement of language-responsive regions in the frontal eye field, the area in the inferior frontal gyrus (BA 44/45 near the classic Broca's area), as well as the posterior section of the inferior parietal lobule as indicated by the blue arrows, all consistent with the probabilistic language atlas in earlier work (Lipkin et al., 2022).

4.2. Quantitative evaluation

Quantitatively, Fig. 4(a) shows the pair-wise geometric correlation difference between each of the registration methods to the rigid alignment, and Fig. 4(b) shows the pair-wise functional suprathreshold overlap difference. Values higher than zero indicate consistent registration improvement over rigid alignment. Fig. S.1 in the supplementary material shows the original correlation/overlap values without taking pair-wise differences. Iterative methods such as FreeSurfer, SD, and MSM substantially improved geometric correlations over rigid alignment (by

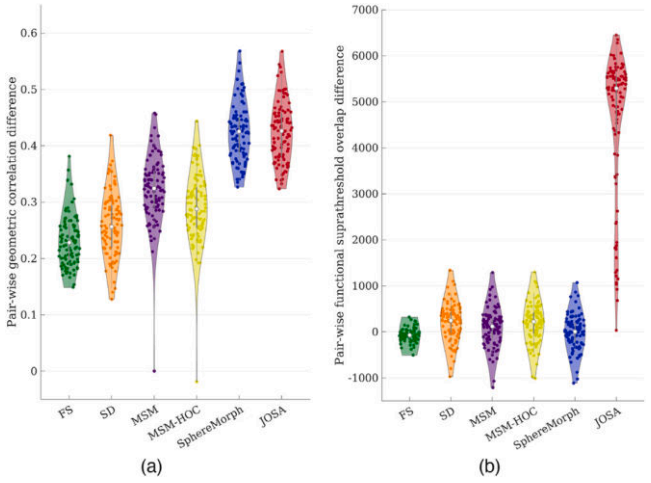

Fig. 4. Violin plot of pairwise difference between each registration method to rigid alignment. (a) Geometric difference measured by correlation; (b) Functional difference measured by the suprathreshold (top 10%, $th = 0.75$) functional overlaps. JOSA yielded substantially higher correlation/larger overlap against rigid alignment compared to the baseline methods and the improvement is consistent across subjects, as shown by the pairwise difference measures. The registration performance of JOSA is significantly better than all baseline methods for both geometry and function with a p-value $<$ 1e−30 except for SphereMorph in geometric alignment.

∼0.3). Learning-based methods achieved a further significant improvement (by ∼0.43) and were comparable between SphereMorph and JOSA. In contrast, only JOSA substantially improved functional registration, while all other methods did not improve or only marginally improved functional registration compared to rigid alignment. The geometric performance difference is statistically significant between JOSA and FS, SD, MSM, MSM-HOC with all p-values $<$ 1e−30 and not significant between JOSA and SphereMorph with a p-value of 0.2. The functional performance difference between JOSA and all baseline methods are statistically significant with p-values $<$ 1e−30.

Fig. 5 shows how the two thresholds used in quantitative evaluation impact the registration performance. Fig. 5(a) shows the number of task-active vertices agreed among at least 75% of the subjects (i.e., th is fixed at 0.75) as a function of the top X%. Fig. 5(b) shows the counterpart where top X% is fixed at 0.1 and th varies from 0.6 to 0.9. All geometry-based registration methods yielded similar functional overlap to that using only rigid alignment. In contrast, JOSA produces substantially higher functional overlap. This significant difference between JOSA and the baseline methods is consistent regardless of the choice of the thresholds.

For geometric features such as sulcal depth, curvature, a sharper average by itself does not necessarily imply a better alignment without considering the regularization of the deformation fields. As the blue curve in Fig. 6 shows, the geometric registration performance improves as the regularization weight is lowered. However, this is *not* the case for functional data. Further reducing regularization of the deformation field will often result in overfitting given the noisy nature of the functional data, hence detrimental to the functional alignment performance. As shown by the red curve in Fig. 6, there a peak point for the functional registration performance over the range of the regularization hyperparameters and the performance in terms of functional alignment decreases as the regularization is further reduced. This observation is also consistent with studies in the literature (Fischl et al., 2008; Robinson et al., 2018). For example, in the MSM-HOC method (Robinson et al., 2018), the performance of functional alignment was improved in comparison to the their original MSM method when a stronger regularization is used (as the method name "higher-order constraint [HOC]" suggested).

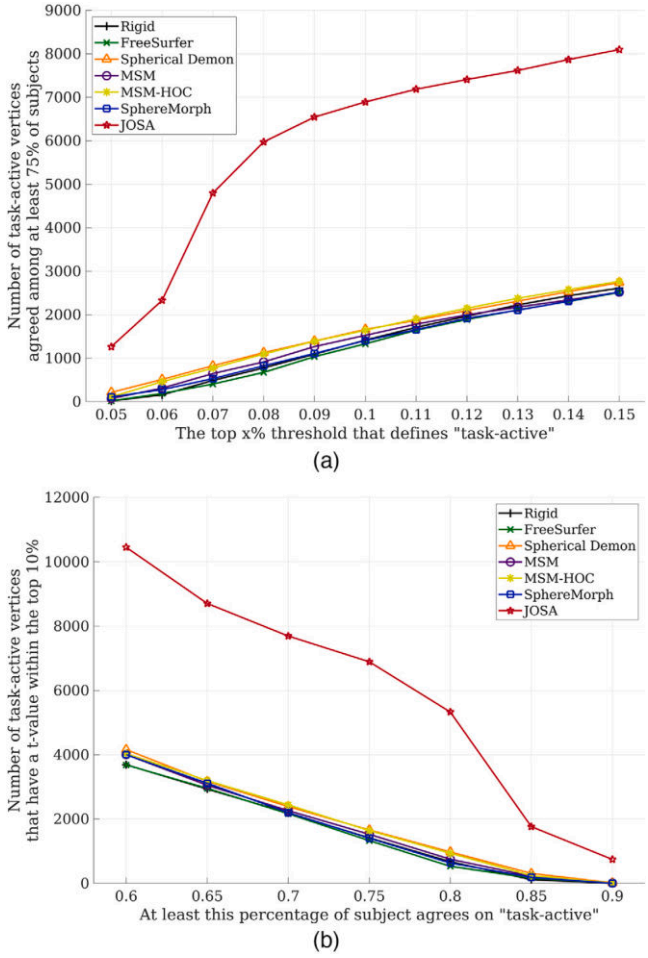

Fig. 5. Impact of the thresholds on functional registration performance. (a) Number of task-active vertices agreed among at least 75% of the subjects as a function of the top X% that defines "task-active". (b) Number of task-active vertices that have a t-value within the top 10% of each subject' task map as a function of the agreement percentage. The functional registration performance of JOSA is substantially higher than that using baseline methods regardless of the choice of the thresholds.

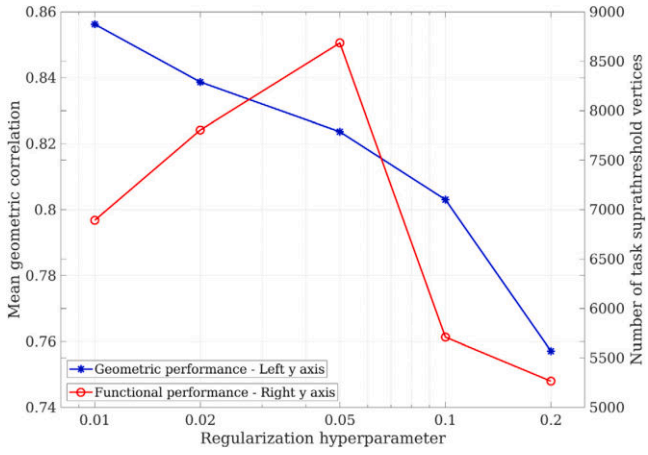

Fig. 6. Impact of the regularization hyperparameters on the registration performance. The blue curve shows the geometric performance measured as correlation (left y-axis) and the red curve shows the functional performance measured as the number of task suprathreshold vertices (right y-axis).

Table 1
Comparison of the run time for individual subject registration. Time is measured in minutes for all methods. For SphereMorph and JOSA, we also show the results in seconds below their corresponding rows for intuitive comparison.

Method	CPU run time total	GPU run time total	GPU run time inference only
FreeSurfer	8.82 ± 2.47	–	–
SD	0.64 ± 0.01	–	–
MSM	21.05 ± 1.87	–	–
MSM-HOC	57.09 ± 8.86	–	–
SphereMorph	0.097 ± 0.008	0.095 ± 0.008	0.003 ± 0.004
SphereMorph (s)	5.81 ± 0.49	5.68 ± 0.51	0.18 ± 0.26
JOSA	$\mathbf{0.097 \pm 0.009}$	$\mathbf{0.094 \pm 0.008}$	$\mathbf{0.003 \pm 0.003}$
JOSA (s)	$\mathbf{5.83 \pm 0.53}$	$\mathbf{5.67 \pm 0.49}$	$\mathbf{0.17 \pm 0.09}$

When using geometric deformation to align parcellation, as shown in the supplementary Fig. S.2(a), there is a clear advantage of the learning-based methods over traditional methods as with the registration of geometric features. This is expected as the DKT parcellation was labeled based on geometric landmarks. JOSA yields a superior performance in alignment of parcellation measured by the Dice score over all baseline methods with p-values ≈ 0, although JOSA is not trained/optimized for any parcellation tasks. Similarly, a higher correlation is observed when predicting individual language responses using JOSA than that using the baseline methods with p-values $<$ 1e−6, Fig. S.2(b), although the effect size is not as large as the group result due to the noisy nature of the fMRI data.

4.3. Run time

Table 1 summarizes the run time for each registration method. For a single subject, all methods started with the spherical cortical surface, and the output is the warped spherical surface. The default FreeSurfer takes approximately 9 min to complete the spherical registration procedure. SD takes approximately 0.6 min to do the same job. MSM and MSM-HOC operated directly on the sphere without parameterization but took significantly longer to execute. In contrast, the learning-based methods take, on average, only 0.097 min ($\approx$ 5.8 s!), providing a speed up by **two orders of magnitude** compared to the iterative methods. The registration speed can be further accelerated by $\sim$ 0.15 s per subject with a GPU implementation. This 0.15 s CPU-to-GPU improvement is significant because, in both SphereMorph and JOSA, the parameterization and resampling back to sphere take most of the time during registration, and the actual inference time through the convolutional neural network is $\sim$ 0.17 s using GPU (last column). That means if the parameterization and resampling can be parallelized and precomputed on a computer cluster, registering 1000 subjects can be done in under 3 min.

4.4. Atlas stability

Fig. 7 shows the geometric atlases learned from the 6 repeated training runs, learned identically in all respects except that the subjects and their orders are randomly selected as the input to the training. Fig. 7(a) shows the mean averaged over the 6 runs and (b) - (g) shows the absolute difference between each run and the mean. These atlases are highly consistent across runs with all major folds nearly identical to each other and subtle differences in the prefrontal, inferior parietal, and middle temporal regions as reflected in the difference images. This strong consistency is quantitatively confirmed by the Pearson correlations 0.94 ± 0.01 among all 15 possible pairs of these atlases. The

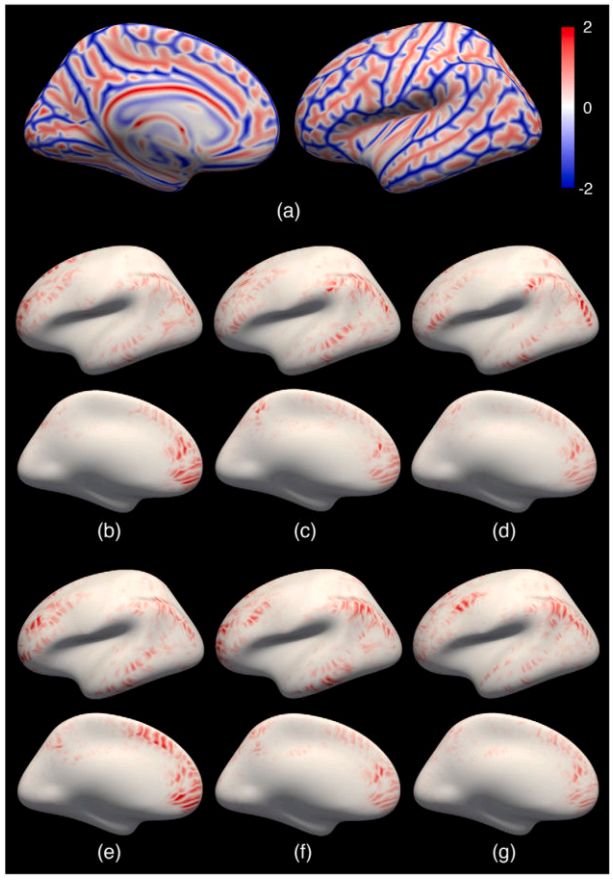

Fig. 7. Learned geometric atlases from 6 repeated training with randomly selected subjects. (a) The mean atlas averaged over the 6 runs. (b) - (g) The absolute difference between each run and the mean. Red indicates sulci and blue indicates gyri. The learned atlases are highly consistent across runs with all major folds nearly identical to each other and subtle differences in prefrontal, inferior parietal, and middle temporal areas.

results support that the atlas construction and its learning procedure in our proposed JOSA method are stable and not biased towards any specific subject.

4.5. Ablation studies

Fig. 8 shows the pair-wise differences from each registration method to the rigid result in the ablation study. The first and the second column shows the result using separate deformations for geometry and function as with JOSA but varying the atlas used as the registration target. The third column shows the result using a single deformation for both geometry and function but the same learned atlas as with JOSA. The fourth column shows the result using two individual deformations separate for geometry and function but without the joint deformation. The last column shows the full JOSA result for easy reference. The substantial improvement in geometry is mainly attributable to a better atlas (p-values $<$ 1e−30 between JOSA and all ablated experiments), whereas the improvement in the registration of function is primarily due to the separate modeling of deformation between geometry and function (p-values $<$ 1e−9). We highlight the importance of this separate modeling of deformations. The third column suggests that the registration performance is sub-optimal for both geometry and function if they share the same deformation, which is commonly assumed (anatomy predicts function).

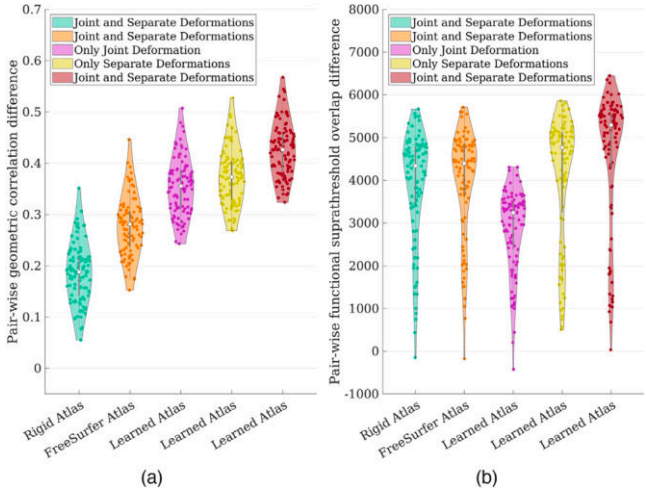

Fig. 8. Violin plot of pair-wise difference between each registration method to rigid alignment for ablation studies. (a) Geometric difference measured by correlation; (b) Functional difference measured by the suprathreshold overlaps. Geometric improvement is mainly attributable to a better atlas (p-values $<$ 1e−30) whereas functional improvement is primarily due to the separate modeling of the deformation fields (p-values $<$ 1e−9).

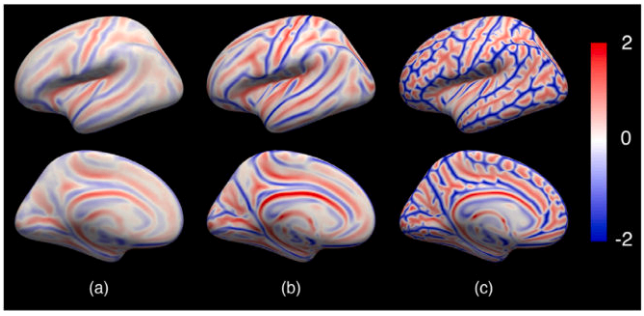

Fig. 9. Comparison of atlases. (a) Rigid atlas; (b) FreeSurfer atlas; (c) JOSA atlas. The learned atlas from JOSA provides more anatomical definition than rigid or FreeSurfer atlas, thus supporting registration with higher resolution and finer details.

5. Conclusion and discussion

In this work, we presented JOSA, a novel registration framework that jointly models the relationship between folding patterns and cortical function. In the early 1900s, the correspondence between the two was discovered in primary motor and sensory regions where a cortical homunculus was drawn to illustrate the distorted mapping of human body parts to sensorimotor functions (Penfield and Boldrey, 1937). Similarly, the functional properties of primary visual cortex have been well studied, where the central and peripheral visual fields are located in the posterior and anterior part of the calcarine sulcus, respectively (Dougherty et al., 2003; Sereno et al., 1995; Tootell et al., 1998; Engel, 1997; Wandell et al., 2007; Engel et al., 1994), and the topographic mapping of the visual field to the visual cortex follows the complex-logarithm rule (Schwartz, 1980; Polimeni et al., 2010). However, the anatomo-functional mapping in higher-order functional areas, e.g., language, emotion, etc., is highly variable among individuals compared with the primary regions (Steinmetz and Seitz, 1991; Fischl et al., 2008; Frost and Goebel, 2012; Robinson et al., 2014). This variability is mainly attributable to the non-unique optimal solutions for geometric registration, where, without the guidance from functional data, their performance can be sub-optimal in registration of function. For example, there are several secondary folds around the active language region near the frontal eye field. Aligning any of them may yield equally good geometric registration, but only one of them may be optimal for functional registration. This is not obvious without access and visualization of the functional data of each individual subject.

By using a semi-supervised training strategy with functional data as the guidance, JOSA is able to choose one appropriate deformation from many valid geometric solutions that produces optimal functional registration. Fig. 4 suggests that both JOSA and SphereMorph (no functional data) achieved comparable performance in geometric alignment, but only JOSA substantially improved functional registration. Moreover, a structure-function mismatch may exist and be different across individuals, i.e., different subjects may use slightly different regions to process similar function. By allowing small deviations of the functional alignment from geometric alignment, JOSA yields superior performance in registration of both folding patterns and functional properties relative to other traditional or learning-based methods. Fig. 8, the 3rd column showed that both the geometric and functional registration are sub-optimal when a single (shared between geometry and function) deformation is used.

In contrast to many other learning-based methods, JOSA also estimates a population-specific atlas as the registration target during training. Fig. 9 shows the learned atlas in (c) compared to atlases constructed by averaging over rigidly registered training subjects (a) and averaged FreeSurfer-registered training subjects in (b). We find that the JOSA atlas provides more anatomical definition, supporting registration with higher resolution and finer details, which also contributes to the improved performance of JOSA (particularly in geometric registration).

Atlas or template construction has been widely studied in classical iterative approaches (Collins et al., 1995; Fischl et al., 1999b; Desikan et al., 2006; Destrieux et al., 2010; Joshi et al., 2022; Lyttelton et al., 2007; Van Essen, 2005). These methods build atlases by repeatedly registering subjects to an estimated atlas, and estimating a new atlas by averaging the registered subjects (Dickie et al., 2017). Their registration is commonly driven by whole-brain geometric features (e.g., sulcal depth, curvature) (Fischl et al., 1999b) or landmarks derived from geometry (Van Essen, 2005). The registration can be also performed in a hierarchical manner from lower resolution to higher resolution (Lyttelton et al., 2007). However, inherited from these iterative construction approaches, a bias towards the initial subject may be introduced to the atlas. JOSA, as one of the very recent learning-based methods, not only facilitates faster atlas construction, but also potentially avoids the bias to any individual subject by random selection of subjects as inputs during training (Cheng et al., 2020b), as illustrated in Fig. 7. In this work, JOSA learned an atlas specific to the population that was used for training (Lipkin et al., 2022). It is, however, straightforward to extend this framework to learn any population-specific atlases, optionally conditioned on clinical attributes if desired (Dalca et al., 2019; Dey et al., 2021; Ding and Niethammer, 2022).

Using a semi-supervised training strategy, the deformation fields can be solely determined from geometric features, which can be derived from a single T1-weighted MR scan. Therefore, JOSA not only lifts the burden of acquiring functional data during inference, but also provides capacity for extension to other comprehensive data modalities that are related to geometry but difficult or even impossible to acquire, such as architectonics, transcriptomics, or proteomics. This geometry-only-based registration model further avoids the circularity issue in any subsequent analyses of functional or any other auxiliary data. In other words, if the functional data is used as one of the inputs to drive the registration, then any evaluation or analysis of the functional data after registration will be potentially biased as the network has observed the data before.

Although JOSA achieved a superior functional registration performance over geometry-focused approaches at the group level, prediction

of language response by registering the atlas back to individuals only marginally outperforms other approaches. This is mainly due to the fact that there is a large variation in language response (or, in general, human brain function) across individuals even in anatomically corresponding areas (Fischl et al., 2008) (see Fedorenko and Blank (2020) for discussion within the domain of language). Moreover, fMRI signal is subject to noise, which may come from multiple sources associated with the recording device or physiology (e.g., system instabilities, head motion). Therefore, prediction (localization) of individual functional responses based on the group result (atlas) remains challenging yet promising direction for the future.

The current approach for better registration of brain function is limited by the use of a single localizer task, the language localizer, used in this study. The partial coverage of the brain by the language-active regions may not provide useful registration guidance for functions that use different parts of the cerebral cortex. We therefore plan to extend JOSA to training with multi-modal datasets, enabling accurate registration of an array of brain functional responses. Although the language dataset used in this study (Lipkin et al., 2022) is one of the largest task fMRI datasets publicly available, another caveat in the training of the registration network is the limited number of samples (subjects). Because the functional data were used only in loss evaluation through the semi-supervised block (i.e., they were not seen by the convolutional layers directly), it poses a significant challenge for the network to learn the anatomo-functional relationships, hence often resulting in an overfitting issue when the sample size is small. We also acknowledge the difficulty in data acquisition. One potential solution in the future is to use a pooled dataset from different task contrasts or even resting-state fMRI data, where subjects may have different functional data available, and the network can be trained dynamically using the corresponding available features, atlases, and loss functions.

Nevertheless, the superior performance of JOSA is mainly attributable to the modeling of the deviation between geometric and functional deformations. We hope this work elucidates the potential mismatch between brain geometry and function that might have been overlooked in the past and provides some new insights into the development of registration methods using joint analysis of brain structure and function in the future.

CRediT authorship contribution statement

Jian Li: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Greta Tuckute:** Writing – review & editing, Validation, Formal analysis, Data curation. **Evelina Fedorenko:** Writing – review & editing, Validation, Supervision, Resources, Data curation. **Brian L. Edlow:** Writing – review & editing, Validation, Supervision, Resources, Funding acquisition. **Adrian V. Dalca:** Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition, Conceptualization. **Bruce Fischl:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: GT was financially supported by the Amazon Fellowship from the Science Hub, administered by the MIT Schwarzman College of Computing, and the International Doctoral Fellowship from the American Association of University Women. EF, BLE, and BF report financial support was provided by National Institutes of Health. EF reports additional financial support from the McGovern Institute for Brain Research, the Brain and Cognitive Sciences Department, the Simons Center for the Social Brain and MIT's Quest for Intelligence. BLE reports additional financial support from the James S. McDonnell Foundation and the Chen Institute MGH Research Scholar Award. BF is an advisor to DeepHealth, a company whose medical pursuits focus on medical imaging and measurement technologies. BF's interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies.

Data availability

Data will be made available on request.

Acknowledgments

Support for this research was provided in part by the BRAIN Initiative Cell Census Network, United States grants U01MH117023 and UM1MH130981, the Brain Initiative Brain Connects consortium, United States (U01NS132181, UM1NS132358), the National Institute for Biomedical Imaging and Bioengineering, United States (R01EB023281, R01EB006758, R21EB018907, R01EB019956, P41EB030006, R01EB033773), the National Institute on Aging, United States (R21AG082082, R01AG064027, R01AG008122, R01AG016495, R01AG070988), the National Institute of Mental Health, United States (UM1MH130981, R01MH123195, R01MH121885, RF1MH123195), the National Institute for Neurological Disorders and Stroke, United States (R01NS052585, R21NS072652, R01NS070963, R01NS083534, U01NS086625, U24NS100591, R01NS105820, R21NS109627, RF1NS115268, R01NS138257, R01NS130119, U01NS121471), the National Institute on Deafness and Other Communication Disorders, United States (R01DC016607, R01DC016950), the NIH Director's Office, United States (DP2HD101400), the James S. McDonnell Foundation, and the Chen Institute MGH Research Scholar Award. This work was also made possible by the resources provided by Shared Instrumentation Grants, United States (S10RR023401, S10RR019307, S10RR023043), the NIH Blueprint for Neuroscience Research, United States (U01MH093765), part of the multi-institutional Human Connectome Project. Additional support was provided by the McGovern Institute for Brain Research, the Brain and Cognitive Sciences Department, the Simons Center for the Social Brain and MIT's Quest for Intelligence, the Amazon Fellowship from the Science Hub, administered by the MIT Schwarzman College of Computing, and the International Doctoral Fellowship from the American Association of University Women (AAUW). Much of the computation resources required for this research was performed on computational hardware generously provided by the Massachusetts Life Sciences Center (https://www.masslifesciences.com). In addition, BF is an advisor to DeepHealth, a company whose medical pursuits focus on medical imaging and measurement technologies. BF's interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies.

Appendix. Supplementary material

See Figs. S.1 and S.2.

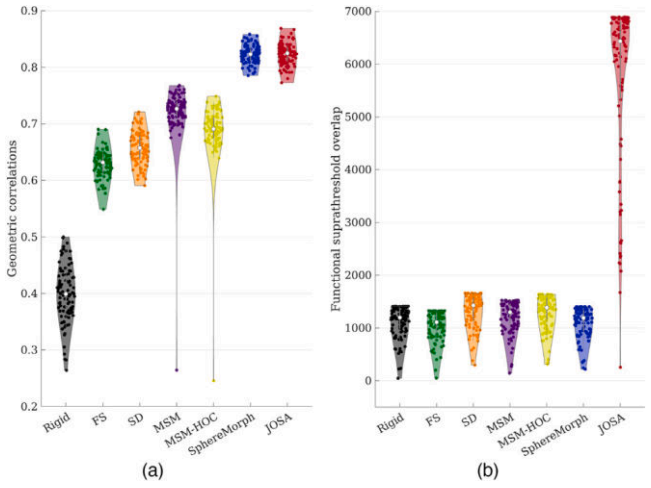

Fig. S.1. Violin plot of the original quantitative measures without taking pair-wise difference. (a) Geometric correlations; (b) Functional overlaps.

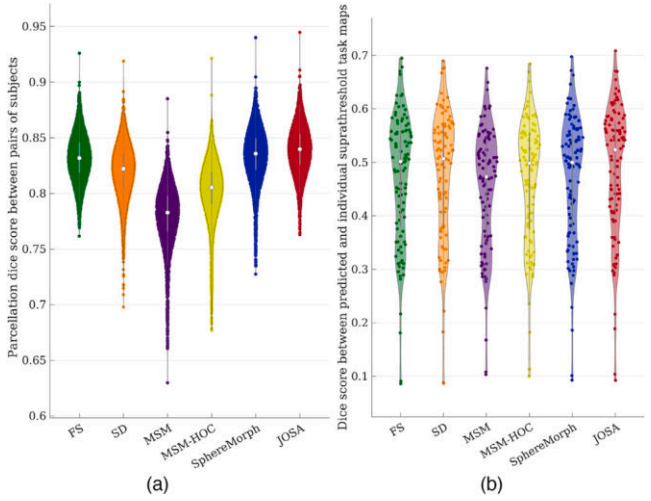

Fig. S.2. Violin plot of pair-wise difference between each registration method to rigid alignment for additional experiments. (a) The parcellation Dice score between all pairs of subjects in the atlas space. The median Dice score is significantly higher in JOSA than baseline methods with p-values ≈ 0. (b) The Dice score between the predicted and the individual suprathreshold task maps in the subject space. The median Dice score is significantly higher in JOSA than the baseline methods with p-values $< 1e{-}6$.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2016. TensorFlow: Large-scale machine learning on heterogeneous distributed systems. http://dx.doi.org/10.48550/arXiv.1603.04467, arXiv:1603.04467.

Aganj, I., Iglesias, J.E., Reuter, M., Sabuncu, M.R., Fischl, B., 2017. Mid-space-independent deformable image registration. NeuroImage 152, 158–170. http://dx.doi.org/10.1016/j.neuroimage.2017.02.055, URL: https://www.sciencedirect.com/science/article/pii/S1053811917301660.

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. NeuroImage 38 (1), 95–113. http://dx.doi.org/10.1016/j.neuroimage.2007.07.007, URL: https://www.sciencedirect.com/science/article/pii/S1053811907005848.

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12 (1), 26–41. http://dx.doi.org/10.1016/j.media.2007.06.004, URL: https://www.sciencedirect.com/science/article/pii/S1361841507000606.

Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V., 2019. VoxelMorph: A learning framework for deformable medical image registration. IEEE Trans. Med. Imaging 38 (8), 1788–1800. http://dx.doi.org/10.1109/TMI.2019.2897538, URL: https://ieeexplore.ieee.org/document/8633930/.

Beg, M.F., Miller, M.I., Trouvé, A., Younes, L., 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vis. 61 (2), 139–157. http://dx.doi.org/10.1023/B:VISI.0000043755.93987.aa.

Blendowski, M., Hansen, L., Heinrich, M.P., 2021. Weakly-supervised learning of multi-modal features for regularised iterative descent in 3D image registration. Med. Image Anal. 67, 101822. http://dx.doi.org/10.1016/j.media.2020.101822, URL: https://linkinghub.elsevier.com/retrieve/pii/S1361841520301869.

Cao, X., Yang, J., Zhang, J., Nie, D., Kim, M., Wang, Q., Shen, D., 2017. Deformable image registration based on similarity-steered CNN regression. In: Medical Image Computing and Computer Assisted Intervention. In: Lecture Notes in Computer Science, Cham, pp. 300–308. http://dx.doi.org/10.1007/978-3-319-66182-7_35, URL: http://link.springer.com/10.1007/978-3-319-66182-7_35.

Cheng, J., Dalca, A.V., Fischl, B., Zöllei, L., 2020a. Cortical surface registration using unsupervised learning. NeuroImage 221, 117161. http://dx.doi.org/10.1016/j.neuroimage.2020.117161, URL: http://www.sciencedirect.com/science/article/pii/S1053811920306479.

Cheng, J., Dalca, A.V., Zollei, L., 2020b. Unbiased atlas construction for neonatal cortical surfaces via unsupervised learning. In: Hu, Y., Licandro, R., Noble, J.A., Hutter, J., Aylward, S., Melbourne, A., Abaci Turk, E., Torrents Barrena, J. (Eds.), In: Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis, vol. 12437, Springer International Publishing, Cham, pp. 334–342. http://dx.doi.org/10.1007/978-3-030-60334-2_33, URL: https://link.springer.com/10.1007/978-3-030-60334-2_33.

Chollet, F., 2018. Keras: the Python deep learning library. Astrophys. Source Code Libr. ascl:1806.022, URL: https://ui.adsabs.harvard.edu/abs/2018ascl.soft06022C.

Christensen, G., Joshi, S., Miller, M., 1997. Volumetric transformation of brain anatomy. IEEE Trans. Med. Imaging 16 (6), 864–877. http://dx.doi.org/10.1109/42.650882, URL: http://ieeexplore.ieee.org/document/650882/.

Collins, D.L., Holmes, C.J., Peters, T.M., Evans, A.C., 1995. Automatic 3D model-based neuroanatomical segmentation. Hum. Brain Mapp. 3 (3), 190–208. http://dx.doi.org/10.1002/hbm.460030304, URL: https://onlinelibrary.wiley.com/doi/10.1002/hbm.460030304.

Conroy, B.R., Singer, B.D., Guntupalli, J.S., Ramadge, P.J., Haxby, J.V., 2013. Inter-subject alignment of human cortical anatomy using functional connectivity. NeuroImage 81, 400–411. http://dx.doi.org/10.1016/j.neuroimage.2013.05.009, URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811913004990.

Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R., 2018a. Unsupervised learning for fast probabilistic diffeomorphic registration. In: Medical Image Computing and Computer Assisted Intervention. In: Lecture Notes in Computer Science, Cham, pp. 729–738. http://dx.doi.org/10.1007/978-3-030-00928-1_82.

Dalca, A.V., Guttag, J., Sabuncu, M.R., 2018b. Anatomical priors in convolutional networks for unsupervised biomedical segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9290–9299, URL: https://openaccess.thecvf.com/content_cvpr_2018/html/Dalca_Anatomical_Priors_in_CVPR_2018_paper.html.

Dalca, A., Rakic, M., Guttag, J., Sabuncu, M., 2019. Learning conditional deformable templates with convolutional networks. In: Advances in Neural Information Processing Systems, vl. 32, Curran Associates, Inc., URL: https://papers.nips.cc/paper/2019/hash/bbcbff5c1f1ded46c25d28119a85c6c2-Abstract.html.

Davatzikos, C., Bryan, N., 1996. Using a deformable surface model to obtain a shape representation of the cortex. IEEE Trans. Med. Imaging 15 (6), 785–795. http://dx.doi.org/10.1109/42.544496.

de Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Isgum, I., 2019. A deep learning framework for unsupervised affine and deformable image registration. Med. Image Anal. 52, 128–143. http://dx.doi.org/10.1016/j.media.2018.11.010, URL: https://www.sciencedirect.com/science/article/pii/S1361841518300495.

Desikan, R.S., Segonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31 (3), 968–980. http://dx.doi.org/10.1016/j.neuroimage.2006.01.021, URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811906000437.

Destrieux, C., Fischl, B., Dale, A., Halgren, E., 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. NeuroImage 53 (1), 1–15. http://dx.doi.org/10.1016/j.neuroimage.2010.06.010, URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811910008542.

Dey, N., Ren, M., Dalca, A.V., Gerig, G., 2021. Generative adversarial registration for improved conditional deformable templates. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3929–3941.

Dickie, D.A., Shenkin, S.D., Anblagan, D., Lee, J., Blesa Cabez, M., Rodriguez, D., Boardman, J.P., Waldman, A., Job, D.E., Wardlaw, J.M., 2017. Whole brain magnetic resonance image atlases: A systematic review of existing atlases and caveats for use in population imaging. Front. Neuroinform. 11, http://dx.doi.org/10.3389/fninf.2017.00001, URL: http://journal.frontiersin.org/article/10.3389/fninf.2017.00001/full.

Ding, Z., Niethammer, M., 2022. Aladdin: Joint atlas building and diffeomorphic registration learning with pairwise alignment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20784–20793.

Dougherty, R.F., Koch, V.M., Brewer, A.A., Fischer, B., Modersitzki, J., Wandell, B.A., 2003. Visual field representations and locations of visual areas V1/2/3 in human visual cortex. J. Vis. 3 (10), 1. http://dx.doi.org/10.1167/3.10.1, URL: http://jov.arvojournals.org/article.aspx?doi=10.1167/3.10.1.

Engel, S., 1997. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cerebral Cortex 7 (2), 181–192. http://dx.doi.org/10.1093/cercor/7.2.181, URL: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/7.2.181.

Engel, S.A., Rumelhart, D.E., Wandell, B.A., Lee, A.T., Glover, G.H., Chichilnisky, E.J., Shadlen, M.N., 1994. fMRI of human visual cortex. Nature 369 (6481), 525. http://dx.doi.org/10.1038/369525a0, URL: https://www.nature.com/articles/369525a0.

Fedorenko, E., Blank, I.A., 2020. Broca's area is not a natural kind. Trends in Cognitive Sciences 24 (4), 270–284. http://dx.doi.org/10.1016/j.tics.2020.01.001, URL: https://linkinghub.elsevier.com/retrieve/pii/S1364661320300036.

Fedorenko, E., Hsieh, P.J., Nieto-Castañón, A., Whitfield-Gabrieli, S., Kanwisher, N., 2010. New method for fMRI investigations of language: defining rois functionally in individual subjects. J. Neurphysiol. 104 (2), 1177–1194. http://dx.doi.org/10.1152/jn.00032.2010, URL: https://www.physiology.org/doi/10.1152/jn.00032.2010.

Fischl, B., 2012. FreeSurfer. NeuroImage 62 (2), 774–781. http://dx.doi.org/10.1016/j.neuroimage.2012.01.021, URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811912000389.

Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B.T.T., Mohlberg, H., Amunts, K., Zilles, K., 2008. Cortical folding patterns and predicting cytoarchitecture. Cerebral Cortex 18 (8), 1973–1980. http://dx.doi.org/10.1093/cercor/bhm225, URL: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhm225.

Fischl, B., Sereno, M.I., Dale, A.M., 1999a. Cortical surface-based Analysis. NeuroImage 9 (2), 195–207. http://dx.doi.org/10.1006/nimg.1998.0396, URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811998903962.

Fischl, B., Sereno, M.I., Tootell, R.B.H., Dale, A.M., 1999b. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8 (4), 272–284. http://dx.doi.org/10.1002/(SICI)1097-0193(1999)8:4<272::AID-HBM10>3.0.CO;2-4, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-0193%281999%298%3A4%3C272%3A%3AID-HBM10%3E3.0.CO%3B2-4.

Frost, M.A., Goebel, R., 2012. Measuring structural–functional correspondence: Spatial variability of specialised brain regions after macro-anatomical alignment. NeuroImage 59 (2), 1369–1381. http://dx.doi.org/10.1016/j.neuroimage.2011.08.035, URL: https://www.sciencedirect.com/science/article/pii/S1053811911009281.

Guntupalli, J.S., Feilong, M., Haxby, J.V., 2018. A computational model of shared fine-scale structure in the human connectome. PLoS Comput. Biol. 14 (4), e1006120. http://dx.doi.org/10.1371/journal.pcbi.1006120, URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006120.

Guntupalli, J.S., Hanke, M., Halchenko, Y.O., Connolly, A.C., Ramadge, P.J., Haxby, J.V., 2016. A model of representational spaces in human cortex. Cerebral Cortex 26 (6), 2919–2934. http://dx.doi.org/10.1093/cercor/bhw068.

Hoffmann, M., Billot, B., Greve, D.N., Iglesias, J.E., Fischl, B., Dalca, A.V., 2022. SynthMorph: Learning contrast-invariant registration without acquired images. IEEE Trans. Med. Imaging 41 (3), 543–558. http://dx.doi.org/10.1109/TMI.2021.3116879.

Joshi, A.A., Choi, S., Liu, Y., Chong, M., Sonkar, G., Gonzalez-Martinez, J., Nair, D., Wisnowski, J.L., Haldar, J.P., Shattuck, D.W., Damasio, H., Leahy, R.M., 2022. A hybrid high-resolution anatomical MRI atlas with sub-parcellation of cortical gyri using resting fMRI. J. Neurosci. Methods 374, 109566. http://dx.doi.org/10.1016/j.jneumeth.2022.109566, URL: https://linkinghub.elsevier.com/retrieve/pii/S0165027022000930.

Joshi, A., Leahy, R., Toga, A.W., Shattuck, D., 2009. A framework for brain registration via simultaneous surface and volume flow. In: Information Processing in Medical Imaging, vol. 5636, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 576–588. http://dx.doi.org/10.1007/978-3-642-02498-6_48, URL: http://link.springer.com/10.1007/978-3-642-02498-6_48.

Joshi, A., Shattuck, D., Thompson, P., Leahy, R., 2007. Surface-constrained volumetric brain registration using harmonic mappings. IEEE Trans. Med. Imaging 26 (12), 1657–1669. http://dx.doi.org/10.1109/TMI.2007.901432, URL: http://ieeexplore.ieee.org/document/4359029/.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. http://dx.doi.org/10.48550/ARXIV.1412.6980, arXiv, URL: https://arxiv.org/abs/1412.6980.

Klein, A., Tourville, J., 2012. 101 labeled brain images and a consistent human cortical labeling protocol. Front. Neurosci. 6, http://dx.doi.org/10.3389/fnins.2012.00171, URL: http://journal.frontiersin.org/article/10.3389/fnins.2012.00171/abstract.

Krebs, J., Delingette, H., Mailhé, B., Ayache, N., Mansi, T., 2019. Learning a probabilistic model for diffeomorphic registration. IEEE Trans. Med. Imaging 38 (9), 2165–2176. http://dx.doi.org/10.1109/TMI.2019.2897112.

Krebs, J., Mansi, T., Delingette, H., Zhang, L., Ghesu, F.C., Miao, S., Maier, A.K., Ayache, N., Liao, R., Kamen, A., 2017. Robust non-rigid registration through agent-based action learning. In: Medical Image Computing and Computer Assisted Intervention. In: Lecture Notes in Computer Science, Cham, pp. 344–352. http://dx.doi.org/10.1007/978-3-319-66182-7_40.

Lee, J.J., Scott, T.L., Perrachione, T.K., 2024. Efficient functional localization of language regions in the brain. NeuroImage 285, 120489. http://dx.doi.org/10.1016/j.neuroimage.2023.120489, URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811923006390.

Li, J., Tuckute, G., Fedorenko, E., Edlow, B.L., Fischl, B., Dalca, A.V., 2024. Joint cortical registration of geometry and function using semi-supervised learning. In: Proceedings of the Sixth Conference on Medical Imaging with Deep Learning. In: Proceedings of Machine Learning Research, 227, PMLR, pp. 862–876, URL: https://proceedings.mlr.press/v227/li24b.html.

Lipkin, B., Tuckute, G., Affourtit, J., Small, H., Mineroff, Z., Kean, H., Jouravlev, O., Rakocevic, L., Pritchett, B., Siegelman, M., Hoeflin, C., Pongos, A., Blank, I.A., Struhl, M.K., Ivanova, A., Shannon, S., Sathe, A., Hoffmann, M., Nieto-Castanon, A., Fedorenko, E., 2022. Probabilistic atlas for the language network based on precision fMRI data from >800 individuals. Sci. Data 9 (1), 529. http://dx.doi.org/10.1038/s41597-022-01645-3, URL: https://www.nature.com/articles/s41597-022-01645-3.

Lyttelton, O., Boucher, M., Robbins, S., Evans, A., 2007. An unbiased iterative group registration template for cortical surface analysis. NeuroImage 34 (4), 1535–1544. http://dx.doi.org/10.1016/j.neuroimage.2006.10.041, URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811906010330.

Lyu, I., Kang, H., Woodward, N.D., Styner, M.A., Landman, B.A., 2019. Hierarchical spherical deformation for cortical surface registration. Med. Image Anal. 57, 72–88. http://dx.doi.org/10.1016/j.media.2019.06.013, URL: https://linkinghub.elsevier.com/retrieve/pii/S1361841518308016.

Lyu, I., Kim, S.H., Seong, J.K., Yoo, S.W., Evans, A.C., Shi, Y., Sanchez, M., Niethammer, M., Styner, M.A., 2013. Group-wise cortical correspondence via sulcal curve-constrained entropy minimization. In: Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M.Y., Weikum, G., Gee, J.C., Joshi, S., Pohl, K.M., Wells, W.M., Zöllei, L. (Eds.), In: Information Processing in Medical Imaging, vol. 7917, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 364–375. http://dx.doi.org/10.1007/978-3-642-38868-2_31, URL: http://link.springer.com/10.1007/978-3-642-38868-2_31.

Mahowald, K., Fedorenko, E., 2016. Reliable individual-level neural markers of high-level language processing: A necessary precursor for relating neural variability to behavioral and genetic variability. NeuroImage 139, 74–93. http://dx.doi.org/10.1016/j.neuroimage.2016.05.073, URL: https://linkinghub.elsevier.com/retrieve/pii/S105381191630194X.

Maintz, J.B.A., Viergever, M.A., 1998. A survey of medical image registration. Med. Image Anal. 2 (1), 1–36. http://dx.doi.org/10.1016/S1361-8415(01)80026-8, URL: https://www.sciencedirect.com/science/article/pii/S1361841501800268.

Mok, T.C., Chung, A.C., 2020. Fast symmetric diffeomorphic image registration with convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR.

Nenning, K.H., Liu, H., Ghosh, S.S., Sabuncu, M.R., Schwartz, E., Langs, G., 2017. Diffeomorphic functional brain surface alignment: Functional demons. NeuroImage 156, 456–465. http://dx.doi.org/10.1016/j.neuroimage.2017.04.028, URL: https://www.sciencedirect.com/science/article/pii/S105381191730321X.

Niethammer, M., Kwitt, R., Vialard, F.X., 2019. Metric learning for image registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8463–8472, URL: https://openaccess.thecvf.com/content_CVPR_2019/html/Niethammer_Metric_Learning_for_Image_Registration_CVPR_2019_paper.html.

Penfield, W., Boldrey, E., 1937. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. Brain 60 (4), 389–443. http://dx.doi.org/10.1093/brain/60.4.389, URL: https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/60.4.389.

Polimeni, J.R., Fischl, B., Greve, D.N., Wald, L.L., 2010. Laminar analysis of 7T BOLD using an imposed spatial activation pattern in human V1. NeuroImage 52 (4), 1334–1346. http://dx.doi.org/10.1016/j.neuroimage.2010.05.005, URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811910007111.

Postelnicu, G., Zollei, L., Fischl, B., 2009. Combined volumetric and surface registration. IEEE Trans. Med. Imaging 28 (4), 508–522. http://dx.doi.org/10.1109/TMI.2008.2004426, URL: http://ieeexplore.ieee.org/document/4601460/.

Robinson, E.C., Garcia, K., Glasser, M.F., Chen, Z., Coalson, T.S., Makropoulos, A., Bozek, J., Wright, R., Schuh, A., Webster, M., Hutter, J., Price, A., Cordero Grande, L., Hughes, E., Tusor, N., Bayly, P.V., Van Essen, D.C., Smith, S.M., Edwards, A.D., Hajnal, J., Jenkinson, M., Glocker, B., Rueckert, D., 2018. Multimodal surface matching with higher-order smoothness constraints. NeuroImage 167, 453–465. http://dx.doi.org/10.1016/j.neuroimage.2017.10.037, URL: https://www.sciencedirect.com/science/article/pii/S1053811917308649.

Robinson, E.C., Jbabdi, S., Glasser, M.F., Andersson, J., Burgess, G.C., Harms, M.P., Smith, S.M., Van Essen, D.C., Jenkinson, M., 2014. MSM: A new flexible framework for Multimodal Surface Matching. NeuroImage 100, 414–426. http://dx.doi.org/10.1016/j.neuroimage.2014.05.069, URL: https://www.sciencedirect.com/science/article/pii/S1053811914004546.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation. http://dx.doi.org/10.48550/ARXIV.1505.04597, arXiv, URL: https://arxiv.org/abs/1505.04597.

Sabuncu, M.R., Singer, B.D., Conroy, B., Bryan, R.E., Ramadge, P.J., Haxby, J.V., 2010. Function-based intersubject alignment of human cortical anatomy. Cerebral Cortex 20 (1), 130–140. http://dx.doi.org/10.1093/cercor/bhp085, URL: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhp085.

Schwartz, E.L., 1980. Computational anatomy and functional architecture of striate cortex: A spatial mapping approach to perceptual coding. Vis. Res. 20 (8), 645–669. http://dx.doi.org/10.1016/0042-6989(80)90090-5, URL: https://linkinghub.elsevier.com/retrieve/pii/0042698980900905.

Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., Tootell, R.B.H., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268 (5212), 889–893. http://dx.doi.org/10.1126/science.7754376, URL: https://www.science.org/doi/10.1126/science.7754376.

Sokooti, H., de Vos, B., Berendsen, F., Lelieveldt, B.P.F., Isgum, I., Staring, M., 2017. Nonrigid image registration using multi-scale 3D convolutional neural networks. In: Medical Image Computing and Computer Assisted Intervention. In: Lecture Notes in Computer Science, Cham, pp. 232–239. http://dx.doi.org/10.1007/978-3-319-66182-7_27.

Steinmetz, H., Seitz, R.J., 1991. Functional anatomy of language processing: Neuroimaging and the problem of individual variability. Neuropsychologia 29 (12), 1149–1161. http://dx.doi.org/10.1016/0028-3932(91)90030-C, URL: https://www.sciencedirect.com/science/article/pii/002839329190030C.

Toga, A., Thompson, P., 2001. The role of image registration in brain mapping. Image Vis. Comput. 19 (1–2), 3–24. http://dx.doi.org/10.1016/S0262-8856(00)00055-X, URL: https://linkinghub.elsevier.com/retrieve/pii/S026288560000055X.

Tootell, R.B.H., Hadjikhani, N.K., Vanduffel, W., Liu, A.K., Mendola, J.D., Sereno, M.I., Dale, A.M., 1998. Functional analysis of primary visual cortex (V1) in humans. Proc. Natl. Acad. Sci. 95 (3), 811–817. http://dx.doi.org/10.1073/pnas.95.3.811, URL: https://pnas.org/doi/full/10.1073/pnas.95.3.811.

Tuckute, G., Lee, E.J., Sathe, A., Fedorenko, E., 2024. A 3.5-minute-long reading-based fmri localizer for the language network. bioRxiv 2024–07.

van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2004. Integration of letters and speech sounds in the human brain. Neuron 43 (2), 271–282. http://dx.doi.org/10.1016/j.neuron.2004.06.025, URL: https://www.sciencedirect.com/science/article/pii/S0896627304003964.

Van Essen, D.C., 2005. A population-average, landmark- and surface-based (PALS) atlas of human cerebral cortex. NeuroImage 28 (3), 635–662. http://dx.doi.org/10.1016/j.neuroimage.2005.06.058, URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811905004945.

Van Essen, D.C., Drury, H.A., Joshi, S., Miller, M.I., 1998. Functional and structural mapping of human cerebral cortex: Solutions are in the surfaces. Proc. Natl. Acad. Sci. 95 (3), 788–795. http://dx.doi.org/10.1073/pnas.95.3.788, URL: https://pnas.org/doi/full/10.1073/pnas.95.3.788.

Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2009. Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45 (1), S61–S72. http://dx.doi.org/10.1016/j.neuroimage.2008.10.040, URL: https://www.sciencedirect.com/science/article/pii/S1053811908011683.

Wandell, B.A., Dumoulin, S.O., Brewer, A.A., 2007. Visual field maps in human cortex. Neuron 56 (2), 366–383. http://dx.doi.org/10.1016/j.neuron.2007.10.012, URL: https://linkinghub.elsevier.com/retrieve/pii/S089662730700774X.

Yang, X., Kwitt, R., Niethammer, M., 2016. Fast predictive image registration. In: Deep Learning and Data Labeling for Medical Applications. In: Lecture Notes in Computer Science, Cham, pp. 48–57. http://dx.doi.org/10.1007/978-3-319-46976-8_6.

Yeo, B., Sabuncu, M., Vercauteren, T., Ayache, N., Fischl, B., Golland, P., 2010a. Spherical demons: Fast diffeomorphic landmark-free surface registration. IEEE Trans. Med. Imaging 29 (3), 650–668. http://dx.doi.org/10.1109/TMI.2009.2030797, URL: http://ieeexplore.ieee.org/document/5223581/.

Yeo, B.T.T., Sabuncu, M.R., Vercauteren, T., Holt, D.J., Amunts, K., Zilles, K., Golland, P., Fischl, B., 2010b. Learning task-optimal registration cost functions for localizing cytoarchitecture and function in the cerebral cortex. IEEE Trans. Med. Imaging 29 (7), 1424–1441. http://dx.doi.org/10.1109/TMI.2010.2049497, URL: http://ieeexplore.ieee.org/document/5482176/.

Zhao, F., Wu, Z., Wang, L., Lin, W., Gilmore, J.H., Xia, S., Shen, D., Li, G., 2021a. Spherical deformable U-net: Application to cortical surface parcellation and development prediction. IEEE Trans. Med. Imaging 40 (4), 1217–1228. http://dx.doi.org/10.1109/TMI.2021.3050072, URL: https://ieeexplore.ieee.org/document/9316936/.

Zhao, F., Wu, Z., Wang, F., Lin, W., Xia, S., Shen, D., Wang, L., Li, G., 2021b. S3Reg: Superfast spherical surface registration based on deep learning. IEEE Trans. Med. Imaging 40 (8), 1964–1976. http://dx.doi.org/10.1109/TMI.2021.3069645.