Joint cortical registration of geometry and function via semi-supervised learning

Author name(s) withheld EMAIL(S) WITHHELD
Address withheld

Editors: Under Review for MIDL 2023

Abstract

Surface-based image registration is an important component of brain image analysis and aims at establishing spatial correspondence between the cortical surfaces of different subjects. Current classical and learning-based approaches focus on accurate registration of folding patterns of the cerebral cortex, with the assumption that geometry predicts function, thus hoping functional areas can also be well aligned. However, structure/functional variability of anatomically corresponding areas across subjects has been widely reported. In this work, we introduce a novel learning-based cortical registration framework, JOSA, which jointly aligns folding patterns and functional maps with optimal, learned atlases. We demonstrate that JOSA can substantially improve registration performance in both anatomical and functional domains. Further, using a semi-supervised training strategy, this framework obviates the need for functional data at test time, which may greatly benefit its future translation to clinical and other neuroscientific applications in which training data may not be observable at inference time.

Keywords: Cortical registration, Semi-supervised Learning

1. Introduction

Image registration is a key research topic in medical image analysis. Deformable image registration aims at establishing spatial correspondence between a pair of images by seeking a dense spatial transformation that maximizes some similarity measures between the two images. Approaches that address the problem of registering two brain scans typically belong to one of the two categories: volume-based or surface-based methods (Maintz and Viergever, 1998). Surface-based registration, also referred to as cortical registration, extracts geometric features representing the anatomical structure of the brain and solves the registration problem based on a surface-matching framework.

Cortical registration is able to achieve high accuracy in aligning highly complex folding patterns of the human cerebral cortex (Davatzikos and Bryan, 1996; Fischl et al., 1999), and it has been shown that inter-subject cortical registration can improve the statistical power of group functional analysis of the brain (van Atteveldt et al., 2004; Frost and Goebel, 2012). Due to the geometric complexity of the cortex and the large variability between individuals, cortical surface registration remains challenging. Inter-subject surface alignment is commonly driven by geometric features that describe measures of cortical folding patterns, such as sulcal depth, curvature (Fischl et al., 1999; Yeo et al., 2010; Conroy et al., 2013). However, functional variability of anatomically corresponding areas within subjects has also been widely reported (Steinmetz and Seitz, 1991; Fischl et al., 2008; Frost and Goebel, 2012). Regions with different functional specializations may not be accurately aligned even when a perfect anatomical registration is achieved.

© 2023 Anonymous.

Here, we propose a novel diffeomorphic cortical registration framework that jointly describes the relationship between geometry and function. Based on the unsupervised spherical registration framework "SphereMorph" (Balakrishnan et al., 2019; Cheng et al., 2020), we use a "big" deformation field shared by geometry and function to model the relatively large inter-subject difference, and two "small" deformation fields to allow relatively small variations between geometry and function within subjects. We demonstrate the utility of this method via a semi-supervised training strategy using task functional magnetic resonance imaging (fMRI) data. The experimental results show that our framework yields better registration accuracy in aligning both folding patterns and functional maps in comparison to the state-of-the-art approaches. This work presents two technical contributions:

1. We propose a learning-based cortical registration framework that jointly models the relationship between geometry and function, which yields better registration performance in both domains.

2. We develop a semi-supervised training strategy that uses task fMRI data to improve functional registration but without a need for task fMRI data at test time. This semi-supervised training framework can also be extended to any auxiliary data that could be helpful to guide registration but difficult to obtain at test time, e.g., parcellations, architectonic identity, transcriptomic information, molecular profiles, etc., depending on the specific purpose.

2. Related work

2.1. Model-based spherical registration

Model-based deformable registration has been extensively studied in the past (Fischl et al., 1999; Yeo et al., 2010; 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). Typical strategies often employ an iterative approach that seeks an optimal deformation field to warp a moving image to a fixed image. A widely used technique for cortical surface registration is to map the surface onto a unit sphere and establish correspondence between feature maps in the spherical space. Spherical registration methods usually involve optimization of some similarity measure between two feature images, e.g., minimizing mean squared error (MSE) or maximizing normalized cross correlation (NCC), while regularizing the deformation field to have some desired property, such as smoothness and/or diffeomorphism. FreeSurfer (Fischl et al., 1999) registers an individual subject to a probabilistic population atlas by minimizing the MSE in convexity weighted by the inverse variance of the atlas convexity in a maximum *a posteriori* formulation. Functional Connectivity Alignment (Sabuncu et al., 2010) adapted the FreeSurfer method to functional registration by using MSE on functional connectivity computed from fMRI data. Hyperalignment (Guntupalli et al., 2016) and Connectivity Hyperalignment (Guntupalli et al., 2018) maximize the spatial correspondence by finding local orthogonal transforms that linearly combine features around the neighborhood at each spatial location. Multimodal Surface Matching (MSM) (Robinson et al., 2014, 2018) uses a discrete optimization approach that iteratively aligns local features using spherical meshes from low-resolution to high-resolution.

To encourage invertibility and preserve the topology after registration, diffeomorphism has been introduced into registration. A fast diffeomorphic registration method (Ashburner,

2007) was developed using an exponentiated Lie algebra with an assumption of a stationary velocity field (SVF). Spherical Demons (Yeo et al., 2010) extended the Diffeomorphic Demons method (Vercauteren et al., 2009) to cortical surfaces by regularizing the deformation field using spherical thin plate spline interpolation. Functional Demons (Nenning et al., 2017) has further extended the Spherical Demons method to align functional regions using Laplacian eigen embeddings computed from fMRI data. These methods can achieve state-of-the-art performance but need to solve an optimization problem for each image pair, resulting in a high computational cost.

2.2. Learning-based spherical registration

With the increase in popularity of deep learning techniques, many learning-based registration methods have been proposed. They can be categorized into supervised registration (Krebs et al., 2017; Sokooti et al., 2017; Yang et al., 2016; Cao et al., 2017) and unsupervised registration (Dalca et al., 2018a; Balakrishnan et al., 2019; Cheng et al., 2020; Niethammer et al., 2019; Krebs et al., 2019; de Vos et al., 2019).

Supervised registration aims at predicting the deformation fields directly and evaluates the performance by comparing the predicted deformation with a "ground truth" deformation. However, due to the lack of the "ground truth", deformation generated by some existing model-based methods is often used as the target, which fundamentally limits the performance of supervised approaches. Further, the dependency on relatively slow model-based registration methods also makes it impractical to generate a large number of samples for training, which is the reason that most of the supervised registration approaches are based on small patches (Sokooti et al., 2017; Yang et al., 2016; Cao et al., 2017).

In contrast, unsupervised registration outputs a deformation field given a pair of images and uses differentiable Spatial Transformer Networks (Jaderberg et al., 2015) to warp the moving image to the fixed image based on the generated deformation. The loss is evaluated in the image space, thus forming an end-to-end training pipeline. Deep Learning Image Registration (de Vos et al., 2019) was developed for hierarchical multi-resolution affine or deformable image registration. VoxelMorph (Balakrishnan et al., 2019) models an SVF as latent variables that generate topology-preserving diffeomorphic transforms by using a scaling and squaring layer (Arsigny et al., 2006). SphereMorph (Cheng et al., 2020) further extends VoxelMorph to the spherical domain by parameterization of the brain surfaces in a 2D grid that accounts for distortions that occurred during parameterization by using a space-variant metric tensor. Instead of working on the parameterized Euclidean space, Spherical U-net (Zhao et al., 2021) performed convolutions directly on the sphere using spherical kernels.

These learning-based registration methods substantially improved the registration speed by orders of magnitude (at test time) while achieving superior or comparable registration accuracy relative to classical methods. However, no existing method allows for intra-subject variability between geometry and function when performing inter-subject registration.

3. Method

3.1. Joint SphereMorph with Atlas Building

We propose a **Jo**int **S**phereMorph with **A**tlas building (JOSA) method for inter-subject registration with modeling of intra-subject variation between each individual's geometry

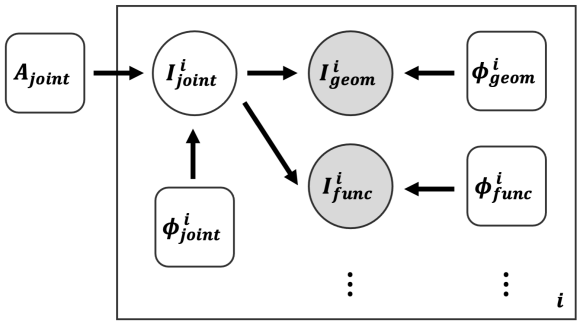


Figure 1: A graphical representation of the generative model. Circles are random variables. Rounded squares indicate parameters. Shaded quantities are observations. The big plate represents replication.

and function. Fig. 1 shows the graphical representation of the generative model for JOSA. Let $\boldsymbol{A}_{joint}$ be the joint atlas with all features combined, $\boldsymbol{I}_{joint}$ and $\boldsymbol{\phi}_{joint} : \mathbb{R}^2 \to \mathbb{R}^2$ be the corresponding individual image and deformation field. $\boldsymbol{I}_{geom}$, $\boldsymbol{I}_{func}$, $\boldsymbol{\phi}_{geom}$, $\boldsymbol{\phi}_{func}$ represent images and deformation fields for geometric features and functional features, respectively.

We propose a generative model that describes the formation of $\boldsymbol{I}_{geom}$ and $\boldsymbol{I}_{func}$ by first warping the atlas $\boldsymbol{A}_{joint}$ by $\boldsymbol{\phi}_{joint}$ to account for (large) inter-subject differences. Then $\boldsymbol{I}_{geom}$ is separated from $\boldsymbol{I}_{joint}$ and followed by an additional warping $\boldsymbol{\phi}_{geom}$, and similarly for $\boldsymbol{I}_{func}$. Built upon the core structures used in VoxelMorph (Balakrishnan et al., 2019) and SphereMorph (Cheng et al., 2020), we leverage a convolutional neural network with SVF integration layers and spatial transform layers to achieve diffeomorphic registration.

We learn network parameters using a semi-supervised training strategy, which obviates the need for functional data at test time. We also note that the framework shown in Fig. 1 is very flexible in the sense that augmentation of data modality can be easily integrated into the framework for simultaneous multi-modality registration, e.g. diffusion or parcellation maps in parallel to geometric and functional features.

3.2. Generative Model

For ease of mathematical derivation, we use $\boldsymbol{I}_k, k = [0, 1, 2, \ldots, N]$ to represent images. $k = 0$ indicates the joint image, i.e., $\boldsymbol{I}_0 = \boldsymbol{I}_{joint}$ and $k = [1, 2, \ldots, N]$ indicate individual images with total number of N modalities, i.e., $\boldsymbol{I}_1 = \boldsymbol{I}_{geom}$, $\boldsymbol{I}_2 = \boldsymbol{I}_{func}$, and so on. Likewise, we use $\boldsymbol{\phi}_k, k = [0, 1, 2, \ldots, N]$ to represent the deformation fields.

Deformation prior. We model the deformation prior $\boldsymbol{\phi}_k$ to have two desired properties: 1) we assume deformation fields are smooth in order to be anatomical plausible for brain images; and 2) we expect an unbiased atlas for the population (Dalca et al., 2019).

$$p(\boldsymbol{\phi}_k) \sim \exp\{-(\lambda_k \|\nabla \boldsymbol{u}_k\|^2 + \alpha_k \|\bar{\boldsymbol{u}}_k\|^2)\} \tag{1}$$

where $\boldsymbol{u}_k$ is the spatial displacement for $\boldsymbol{\phi} = Id + \boldsymbol{u}_k$, $\nabla \boldsymbol{u}_k$ is its spatial gradient, and $\bar{\boldsymbol{u}}_k = 1/n \sum_i \boldsymbol{u}_k^i$. The first term encourages small gradients (smoothness) of the deformation term encourages a small deformation on average over the entire dataset (unbiasedness).

Data likelihood. We use Gaussian to model the data likelihood. For the joint image $\boldsymbol{I}_0$,

$$p(\boldsymbol{I}_0|\boldsymbol{\phi}_0;\boldsymbol{A}) = \mathcal{N}(\boldsymbol{I}_0;\boldsymbol{\phi}_0\circ\boldsymbol{A},\sigma^2\mathbb{I}) \tag{2}$$

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 identity matrix. For individual feature images $\boldsymbol{I}_k$,

$$p(\boldsymbol{I}_k|\boldsymbol{\phi}_k,\boldsymbol{I}_0) = \mathcal{N}(\boldsymbol{I}_k;\boldsymbol{\phi}_k\circ\boldsymbol{I}_0,\sigma^2\mathbb{I}). \tag{3}$$

3.3. Learning

Let $\boldsymbol{\Phi} = \{\boldsymbol{\phi}_k\}_{k=0,1,2,\ldots}$ be the parameter set containing all deformation fields. Let $\boldsymbol{I} = \{\boldsymbol{I}_k\}_{k=1,2,\ldots}$ be the image sets. We estimate $\boldsymbol{\Phi}$ by maximization of its log likelihood,

$$\hat{\boldsymbol{\Phi}} = \underset{\boldsymbol{\Phi}}{\operatorname{argmax}}\ \log p(\boldsymbol{\Phi}|\boldsymbol{I};\boldsymbol{A}) = \underset{\boldsymbol{\Phi}}{\operatorname{argmax}}\ \log p(\boldsymbol{I}|\boldsymbol{\Phi};\boldsymbol{A}) + \log p(\boldsymbol{\Phi}) \tag{4}$$

Equivalently, we can minimize the negative log likelihood $\mathcal{L}(\boldsymbol{\Phi}|\boldsymbol{I},\boldsymbol{A})$

$$\mathcal{L} = -\log\left(p(\boldsymbol{I}_0|\boldsymbol{\phi}_0;\boldsymbol{A})\prod_{k=1}^{N}p(\boldsymbol{I}_k|\boldsymbol{\phi}_k,\boldsymbol{I}_0)\right) - \log\left(\prod_{k=1}^{N}p(\boldsymbol{\phi}_k)\right). \tag{5}$$

Using Eq. (1)–(3), $\mathcal{L}$ can be written as

$$\begin{aligned}\mathcal{L} = \ & \frac{1}{2\sigma^2}\left\|\boldsymbol{I}_0-\boldsymbol{\phi}_0\circ\boldsymbol{A}\right\|^2 + \frac{1}{2\sigma^2}\sum_{k=1}^{N}\left\|\boldsymbol{I}_k-\boldsymbol{\phi}_k\circ\boldsymbol{I}_0\right\|^2 \\ & + \lambda_0\left\|\nabla\boldsymbol{u}_0\right\|^2 + \sum_{k=1}^{N}\lambda_k\left\|\nabla\boldsymbol{u}_k\right\|^2 + \sum_{k=1}^{N}\alpha_k\left\|\bar{\boldsymbol{u}}_k\right\|^2 + \text{const.}\end{aligned} \tag{6}$$

To avoid atlas drift (Aganj et al., 2017) during construction, instead of minimizing the first two data fidelity terms in the mid-space ($\boldsymbol{I}_0$), we evaluate losses in the atlas ($\boldsymbol{A}$) and subject ($\boldsymbol{I}_k$) spaces using the invertibility property of diffeomorphic deformation fields.

3.4. Neural network approach and semi-supervision with task fMRI data

Fig. 2 shows the diagram of the proposed network architecture. The cortical surface of each subject is first 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, generating images as inputs to the network (Cheng et al., 2020).

To solve the maximum likelihood function in Eq. (4), we design a network that takes a parameterized image and the atlas as inputs and outputs *three* deformation fields. The joint deformation, which models the relatively large inter-subject variance, is shared among and composed with individual deformations. Using spatial transform layers, we can efficiently warp geometric and functional images based on the generated deformation fields.

In this study, we use task fMRI data to quantify the accuracy of functional alignment. Critically, we use the 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 U-net. Rather they are used only for functional loss evaluation. In this way, we remove the requirement of having task fMRI data at test time, as the deformation fields can be inferred only using geometric features.

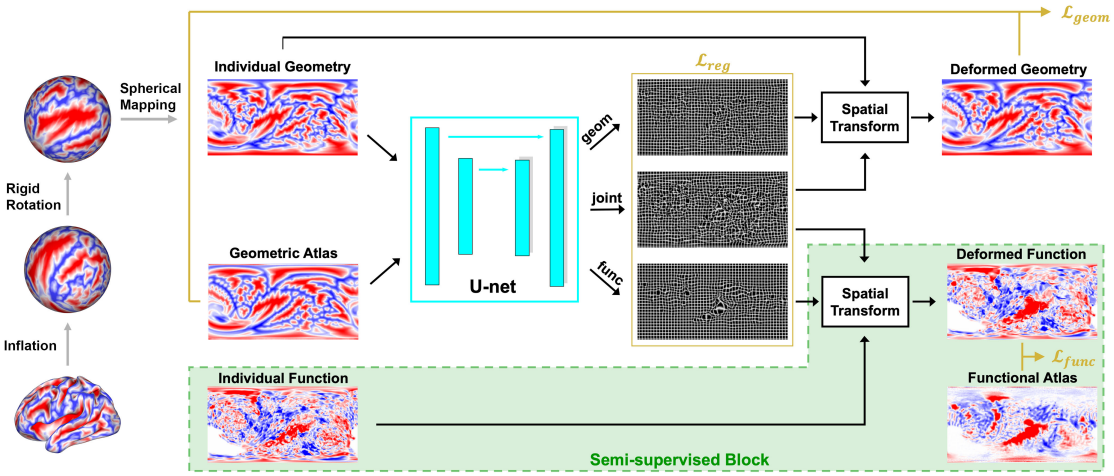


Figure 2: Overview of the network architecture and preprocessing pipeline. Preprocessing (left) includes inflation of the subject surface into a sphere, rigid registration to the average space, and spherical parameterization onto a 2D grid. The network (right) takes the geometric features from the subject as well as the atlas as inputs and outputs a large joint and two small deformation fields for registration of folding patterns and functions separately. Task fMRI data were used for evaluating the functional loss only (the green block) in a semi-supervised manner.

3.5. Implementation

We implemented the network shown in Fig. 2 based on the core architecture in VoxelMorph (`https://github.com/voxelmorph/voxelmorph`) (Balakrishnan et al., 2019; Dalca et al., 2019). We used a 5-layer encoder with the number of filters of [128, 256, 384, 512, 640] and a symmetric decoder for the U-Net followed by 2 more convolutional layers with the number of filters of [64, 32]. Each layer involves convolution, max-pooling/up-sampling, and LeakyReLU activation. Prior and distortion corrections were performed identically to that described in (Cheng et al., 2020). During training, we randomly sampled the training data into mini-batches with a batch size of 8. For each batch, we augmented the data with random deformations (max $\sigma = 4$ and distortion corrected) and Gaussian noise (max $\sigma = 6$). We used the LAMB optimizer (You et al., 2020) 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 features and geometric features were set to 0.7:0.3 to give the network stronger motivation for functional registration, as functional features do not go through the networks. To encourage the smoothness of the deformation fields, we set the regularization hyperparameter λ_0 in Eq. (6) for the joint (large) deformation to be 0.1 and λ_k for the individual (small) deformations to be 0.2, all relative to the data fidelity term, which has a weight of 1. We used TensorFlow (Abadi et al., 2016) with Keras front-end (Chollet, 2018) and the Neurite package (Dalca et al., 2018b) for implementations and all experiments were conducted in the Dell Workstation with dual Intel Xeon Silver 6226R CPUs and an Nvidia RTX6000 GPU.

4. Experiments

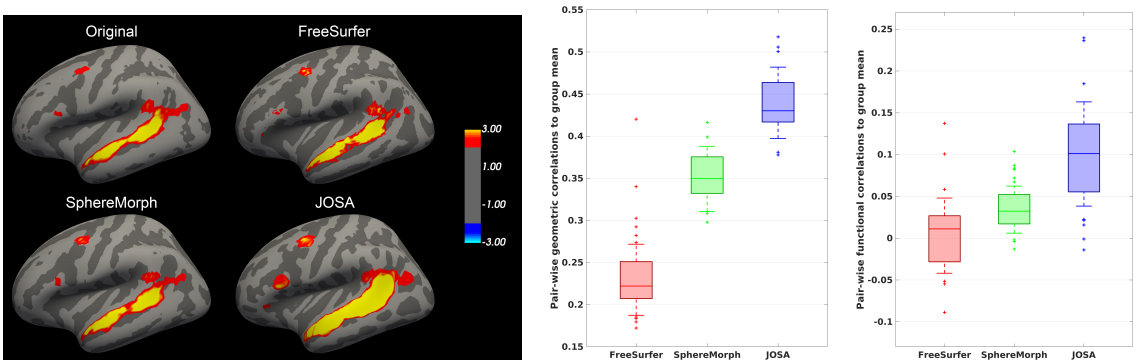
Language task fMRI data. We used a subset of 150 subjects from a large-scale language mapping study (Lipkin et al., 2022). The task used to localize the language network in these subjects was a reading one contrasting sentences and lists of nonwords strings in a standard blocked design with a counterbalanced condition order across runs. Each stimulus consisted of 12 words/nonwords, and stimuli were presented one word/nonword at a time at the rate of 450 ms per word/nonword. 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 100ms, for a total trial duration of 6s. Experimental blocks lasted 18s (with 3 trials per block), and fixation blocks lasted 14s. Each run (consisting of 5 fixation blocks and 16 experimental blocks) lasted 358s. 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. T1-weighted images were collected in 176 sagittal slices with 1 mm isotropic resolution. Functional data (BOLD) were acquired using an EPI sequence with 4 mm thick near-axial slices, 2.1 mm $\times$ 2.1 mm (G2) in-plane resolution, TR = 2,000 ms, and TE = 30 ms. Preprocessing and analysis of the data was performed using FreeSurfer v6.0.0 as described in (Lipkin et al., 2022) with the exception of that the subjects' surfaces were reconstructed from the T1 images (using default *recon-all* parameters) and data were analyzed on the subjects' native ("self") surface instead of the fsaverage. Moreover, data were not spatially smoothed. A "sentence vs. nonword" contrast t-map was generated for each subject using first-level GLM analysis based on the blocked design. We randomly split the data into a training set with 110 subjects and a validation set with the remaining 40 subjects.

Baselines. We compared JOSA with FreeSurfer registration (Fischl et al., 1999), and SphereMorph (Cheng et al., 2020). For FreeSurfer registration, we ran `mris_register` for each individual subject in the validation set to register them to the FreeSurfer average space. For SphereMorph, we trained the network described in (Cheng et al., 2020) with a single deformation field, and we used the average feature maps as the fixed atlases. At test time, we used the deformation field generated based on the subject's geometric features to warp the subject's functional data to the atlas space.

Evaluation. Qualitatively, we computed the group mean images of both the geometric and the functional data in the validation set after registration. We then visualized them by superimposing the functional group mean map with the curvature group mean map using Freeview (Fischl, 2012). Quantitatively, we measured the registration accuracy as the correlations between the registered individual data and the group mean. We show the pair-wise correlation improvement for each method (i.e., the correlation difference for the same subject after and before registration).

5. Results

Fig. 3(a) shows the group mean maps after registration, where the curvature map is shown in dark gray and the thresholded language task activation map is shown using a heat map. We assume a better alignment leads to a cleaner group mean image with higher peaks and sharper transitions from task active to non-active regions. FreeSurfer and SphereMorph



(a) (b) (c)

Figure 3: Comparison of registration result. (a) Average language activation map across test subjects superimposed on average curvature map; (b) box-plot of pair-wise correlation of individual curvature map to the group mean; (c) Counterpart of (b) but for function.

significantly improved the alignment of folding patterns with marginal improvement in functional alignment, which is as expected because their registrations are purely driven by geometric features. In contrast, JOSA achieved a better alignment in both folding patterns and function. In particular, the predominant language region in the superior temporal gyrus shows a substantially stronger response and clearer functional boundaries, and additional active language-responsive regions were discovered around inferior frontal regions near Broca's area, indicating an improved structural and functional alignment after the joint registration.

Quantitatively, Fig. 3(b) and 3(c) show the pair-wise registration *improvement* in correlation for each method. As shown in Fig. 3(b), all three methods show substantial improvement in aligning geometry, while JOSA yielded the highest correlation increase. Fig. 3(c) confirmed our observation in qualitative visualization that FreeSurfer and SphereMorph marginally improved functional registration, whereas JOSA achieved a substantial improvement due to its separate modeling of geometry and function.

6. Conclusion

We developed JOSA, a joint geometric and functional registration framework with atlas building. JOSA shows superior performance in registering folding patterns and task-active regions in comparison to other traditional or learning-based methods. Using a semi-supervised training strategy, JOSA lifts the burden of acquiring functional data at test time, which may greatly benefit its future translation to clinical applications. The current approach is potentially limited by the size of the dataset as well as the single-task contrast used in the study. We plan to expand our framework to a broad range of functional data with a larger number of subjects to better explore the relationship between geometry and function.

Acknowledgments

Acknowledgments withheld.

References

Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mane, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viegas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, 2016.

Iman Aganj, Juan Eugenio Iglesias, Martin Reuter, Mert Rory Sabuncu, and Bruce Fischl. Mid-space-independent deformable image registration. *NeuroImage*, 152:158–170, 2017.

Vincent Arsigny, Olivier Commowick, Xavier Pennec, and Nicholas Ayache. A log-euclidean framework for statistics on diffeomorphisms. In *Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science, pages 924–931, Berlin, Heidelberg, 2006.

John Ashburner. A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1): 95–113, 2007.

B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, 2008.

Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. VoxelMorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38(8):1788–1800, 2019.

M. Faisal Beg, Michael I. Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61(2):139–157, 2005.

Xiaohuan Cao, Jianhua Yang, Jun Zhang, Dong Nie, Minjeong Kim, Qian Wang, and Dinggang Shen. Deformable image registration based on similarity-steered CNN regression. In *Medical Image Computing and Computer Assisted Intervention*, Lecture Notes in Computer Science, pages 300–308, Cham, 2017.

Jieyu Cheng, Adrian V. Dalca, Bruce Fischl, and Lilla Zöllei. Cortical surface registration using unsupervised learning. *NeuroImage*, 221:117161, 2020.

Francois Chollet. Keras: the Python deep learning library. *Astrophysics Source Code Library*, page ascl:1806.022, 2018.

Bryan R. Conroy, Benjamin D. Singer, J. Swaroop Guntupalli, Peter J. Ramadge, and James V. Haxby. Inter-subject alignment of human cortical anatomy using functional connectivity. *NeuroImage*, 81:400–411, 2013.

Adrian Dalca, Marianne Rakic, John Guttag, and Mert Sabuncu. Learning conditional deformable templates with convolutional networks. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Adrian V. Dalca, Guha Balakrishnan, John Guttag, and Mert R. Sabuncu. Unsupervised learning for fast probabilistic diffeomorphic registration. In *Medical Image Computing and Computer Assisted Intervention*, Lecture Notes in Computer Science, pages 729–738, Cham, 2018a.

Adrian V. Dalca, John Guttag, and Mert R. Sabuncu. Anatomical priors in convolutional networks for unsupervised biomedical segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9290–9299, 2018b.

C. Davatzikos and N. Bryan. Using a deformable surface model to obtain a shape representation of the cortex. *IEEE Transactions on Medical Imaging*, 15(6):785–795, 1996.

Bob D. de Vos, Floris F. Berendsen, Max A. Viergever, Hessam Sokooti, Marius Staring, and Ivana Isgum. A deep learning framework for unsupervised affine and deformable image registration. *Medical Image Analysis*, 52:128–143, 2019.

B. Fischl, N. Rajendran, E. Busa, J. Augustinack, O. Hinds, B. T. T. Yeo, H. Mohlberg, K. Amunts, and K. Zilles. Cortical folding patterns and predicting cytoarchitecture. *Cerebral Cortex*, 18(8):1973–1980, 2008.

Bruce Fischl. FreeSurfer. *NeuroImage*, 62(2):774–781, 2012.

Bruce Fischl, Martin I. Sereno, Roger B. H. Tootell, and Anders M. Dale. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, 8(4):272–284, 1999.

Martin A. Frost and Rainer Goebel. Measuring structural–functional correspondence: Spatial variability of specialised brain regions after macro-anatomical alignment. *NeuroImage*, 59(2):1369–1381, 2012.

J. Swaroop Guntupalli, Michael Hanke, Yaroslav O. Halchenko, Andrew C. Connolly, Peter J. Ramadge, and James V. Haxby. A model of representational spaces in human cortex. *Cerebral Cortex*, 26(6):2919–2934, 2016.

J. Swaroop Guntupalli, Ma Feilong, and James V. Haxby. A computational model of shared fine-scale structure in the human connectome. *PLOS Computational Biology*, 14 (4):e1006120, 2018.

Max Jaderberg, Karen Simonyan, Andrew Zisserman, and koray kavukcuoglu. Spatial transformer networks. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Julian Krebs, Tommaso Mansi, Hervé Delingette, Li Zhang, Florin C. Ghesu, Shun Miao, Andreas K. Maier, Nicholas Ayache, Rui Liao, and Ali Kamen. Robust non-rigid registration through agent-based action learning. In *Medical Image Computing and Computer Assisted Intervention*, Lecture Notes in Computer Science, pages 344–352, Cham, 2017.

Julian Krebs, Hervé Delingette, Boris Mailhé, Nicholas Ayache, and Tommaso Mansi. Learning a probabilistic model for diffeomorphic registration. *IEEE Transactions on Medical Imaging*, 38(9):2165–2176, 2019.

Benjamin Lipkin, Greta Tuckute, Josef Affourtit, Hannah Small, Zachary Mineroff, Hope Kean, Olessia Jouravlev, Lara Rakocevic, Brianna Pritchett, Matthew Siegelman, Caitlyn Hoeflin, Alvince Pongos, Idan A. Blank, Melissa Kline Struhl, Anna Ivanova, Steven Shannon, Aalok Sathe, Malte Hoffmann, Alfonso Nieto-Castanon, and Evelina Fedorenko. Probabilistic atlas for the language network based on precision fMRI data from >800 individuals. *Scientific Data*, 9(1):529, 2022.

J. B. Antoine Maintz and Max A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

Karl-Heinz Nenning, Hesheng Liu, Satrajit S. Ghosh, Mert R. Sabuncu, Ernst Schwartz, and Georg Langs. Diffeomorphic functional brain surface alignment: Functional demons. *NeuroImage*, 156:456–465, 2017.

Marc Niethammer, Roland Kwitt, and Francois-Xavier Vialard. Metric learning for image registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8463–8472, 2019.

Emma C. Robinson, Saad Jbabdi, Matthew F. Glasser, Jesper Andersson, Gregory C. Burgess, Michael P. Harms, Stephen M. Smith, David C. Van Essen, and Mark Jenkinson. MSM: A new flexible framework for Multimodal Surface Matching. *NeuroImage*, 100:414–426, 2014.

Emma C. Robinson, Kara Garcia, Matthew F. Glasser, Zhengdao Chen, Timothy S. Coalson, Antonios Makropoulos, Jelena Bozek, Robert Wright, Andreas Schuh, Matthew Webster, Jana Hutter, Anthony Price, Lucilio Cordero Grande, Emer Hughes, Nora Tusor, Philip V. Bayly, David C. Van Essen, Stephen M. Smith, A. David Edwards, Joseph Hajnal, Mark Jenkinson, Ben Glocker, and Daniel Rueckert. Multimodal surface matching with higher-order smoothness constraints. *NeuroImage*, 167:453–465, 2018.

Mert R. Sabuncu, Benjamin D. Singer, Bryan Conroy, Ronald E. Bryan, Peter J. Ramadge, and James V. Haxby. Function-based intersubject alignment of human cortical anatomy. *Cerebral Cortex*, 20(1):130–140, 2010.

Hessam Sokooti, Bob de Vos, Floris Berendsen, Boudewijn P. F. Lelieveldt, Ivana Isgum, and Marius Staring. Nonrigid image registration using multi-scale 3D convolutional neural networks. In *Medical Image Computing and Computer Assisted Intervention*, Lecture Notes in Computer Science, pages 232–239, Cham, 2017.

Helmuth Steinmetz and Rüdiger J. Seitz. Functional anatomy of language processing: Neuroimaging and the problem of individual variability. *Neuropsychologia*, 29(12):1149–1161, 1991.

Nienke van Atteveldt, Elia Formisano, Rainer Goebel, and Leo Blomert. Integration of letters and speech sounds in the human brain. *Neuron*, 43(2):271–282, 2004.

Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72, 2009.

Xiao Yang, Roland Kwitt, and Marc Niethammer. Fast predictive image registration. In *Deep Learning and Data Labeling for Medical Applications*, Lecture Notes in Computer Science, pages 48–57, Cham, 2016.

B.T.T. Yeo, M.R. Sabuncu, T. Vercauteren, N. Ayache, B. Fischl, and P. Golland. Spherical demons: Fast diffeomorphic landmark-free surface registration. *IEEE Transactions on Medical Imaging*, 29(3):650–668, 2010.

Yang You, Jing Li, Sashank Reddi, Jonathan Hseu, Sanjiv Kumar, Srinadh Bhojanapalli, Xiaodan Song, James Demmel, Kurt Keutzer, and Cho-Jui Hsieh. Large batch optimization for deep learning: Training BERT in 76 minutes. In *International Conference on Learning Representations*, 2020.

Fenqiang Zhao, Zhengwang Wu, Fan Wang, Weili Lin, Shunren Xia, Dinggang Shen, Li Wang, and Gang Li. S3Reg: Superfast spherical surface registration based on deep learning. *IEEE Transactions on Medical Imaging*, 40(8):1964–1976, 2021.