King's Research Portal

Document Version
Peer reviewed version

[Link to publication record in King's Research Portal](#)

A Deep-Discrete Learning Framework for Spherical Surface Registration

Mohamed A. Suliman, Logan Z. J. Williams, Abdulah Fawaz, and Emma C. Robinson

Department of Biomedical Engineering, King's College London.
{mohamed.suliman, logan.williams, abdulah.fawaz, emma.robinson}@kcl.ac.uk

Abstract. Cortical surface registration is a fundamental tool for neuroimaging analysis that has been shown to improve the alignment of functional regions relative to volumetric approaches. Classically, image registration is performed by optimizing a complex objective similarity function, leading to long run times. This contributes to a convention for aligning all data to a global average reference frame that poorly reflects the underlying cortical heterogeneity. In this paper, we propose a novel unsupervised learning-based framework that converts registration to a multi-label classification problem, where each point in a low-resolution control grid deforms to one of fixed, finite number of endpoints. This is learned using a spherical geometric deep learning architecture, in an end-to-end unsupervised way, with regularization imposed using a deep Conditional Random Field (CRF). Experiments show that our proposed framework performs competitively, in terms of similarity and areal distortion, relative to the most popular classical surface registration algorithms and generates smoother deformations than other learning-based surface registration methods, even in subjects with atypical cortical morphology.

Keywords: Deep learning · unsupervised learning · cortical surface registration · conditional random fields

1 Introduction

The human cerebral cortex is highly convoluted structure, with complex patterns of functional organisation that vary considerably across individuals [1,13]. Image registration is an important tool that supports comparison of brain images, through mapping of all data to a global average space. However, the degree of variation of cortical morphology and topography across individuals generates considerable uncertainty with regards to the optimal mapping between brains.

Recently, cortical surface registration algorithms [11,26,23,22] have led to improvements in the precision with which it is possible to compare features on the cortical surface through learning mappings that regularize deformations with respect to displacements along with the cortical sheet. Increasingly, these methods support multimodal registration [23,22,27], allowing improved evaluation of cortical functional areas [5,13]. However, increasing evidence suggests that there

is a limit to these improvements since cortical topographies vary in ways that break the diffeomorphic assumptions of classical registration algorithms [13].

To this end, learning-based registration algorithms [6,16] present an attractive framework for exploring these problems, on the grounds that they train fast, unsupervised deformation frameworks that can learn to adapt to sub-populations in the data [6]. Previous learning-based registration frameworks have predominately been generated for 2D or 3D Euclidean domains such as brain volumes [3,6,7,8], lung CT [15,16,12], and histology [4,24,20]. However, increasing efforts have been made to adapt convolutional networks to non-Euclidean domains [19,21,28], resulting in the development of tools for learning-based registration of surfaces and point clouds [2,25,28]. Most notably, the recent S3Reg framework [27] proposes the first learning-based registration framework for cortical surfaces, leveraging the Spherical U-Net algorithm [28] to learn multi-resolution hexagonal filters across regular subdivisions of an icosphere.

One limitation of the Spherical U-Net is that its hexagonal convolutions do not generate a rotationally equivariant solution since the lack of a global coordinate system on a sphere causes filter orientation to flip at the poles. When learning registrations for S3Reg, this generates distortions that must be corrected through averaging warps learned across multiple rotated orientations of the sphere. Recent work showed that contrary to Spherical U-Net, MoNet convolutions (learned from a mixture of Gaussian kernels) could indeed be trained to be rotationally equivariant [9]. Therefore, we develop a new framework for spherical cortical registration based on MoNet, which also takes inspiration from deep-discrete registration frameworks [15,16] designed to learn larger deformations than deep regression frameworks. We hypothesize that this could improve the generalization of our framework to brains with atypical topographies.

Contributions In this paper, we propose the first deep-discrete framework for cortical surface registration. We validate on the alignment of cortical folds and compare the proposed network against state-of-the-art classical [11,26,23,22] and learning-based [27] methods. Specific focus is placed on evaluating the smoothness of the generated warps and the capacity of the network to generalize to atypical cortical morphologies.

2   Method

The proposed method combines ideas from the discrete frameworks of [16,23,22] to propose a Deep-Discrete spherical Registration (DDR) network for alignment of cortical surface features. The full DDR architecture consists of three networks that compose global rigid rotations and non-linear warps (learned in a coarse-to-fine fashion). The objective is to learn a spatial transformation $\boldsymbol{\Phi} : M \to F$, that aligns cortical features on a moving mesh (M) to those of a fixed mesh (F) by optimizing a dissimilarity metric $\mathcal{L}$ of the form:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathcal{L}\left(\boldsymbol{\Phi}_{\boldsymbol{\theta}}; F, M\right) + \Sigma\left(\boldsymbol{\Phi}_{\boldsymbol{\theta}}\right), \tag{1}$$

where $\boldsymbol{\theta}$ are transformation parameters that parametrize $\boldsymbol{\Phi}$, while $\Sigma\left(\cdot\right)$ is a regularization function that imposes smoothness on $\boldsymbol{\Phi}$.

Let $f\left(\cdot\right)$ represents our deep learning network architecture, with $\boldsymbol{\eta}$ being a set of learnable parameters; then, our deep learning image registration problem may be represented as

$$\boldsymbol{\theta} = f_{\boldsymbol{\eta}}\left(F, M\right). \tag{2}$$

In all cases, data is presented to the network as concatenated cortical metric maps of F and M, defined on a sphere S^2 that is parametrized by a sixth-order icosphere (that has 40962 vertices). The base architecture of each network is a MoNet U-Net.

2.1 Rotation Architecture

Registration is learned in 3 stages, with the first network learning global rotations. To obtain the rotation matrix between M and F, we use a MoNet U-Net to estimate the 12 parameters of the rotation matrix. With conventional rotation estimation methods, such as quaternions and Euler angles, being shown to be discontinuous in the real Euclidean spaces of four or fewer dimensions [30], we apply the continuous 6D representation formula of the 3D rotations estimation proposed in [30]. The network optimization is driven using an unsupervised mean-squared-error loss. The rotationally aligned subjects are then passed to the next registration stage.

2.2 Deep-Discrete Networks

Deformable registration is subsequently learned using two deep-discrete networks, which learn optimal displacements as a classification problem, regularized by a deep CRF. Let $\{\mathbf{c}_i\}_{i=1}^{N_c} \in G \subset S^2$ be a set of N_c control points, on the input image sphere, generated from a low-resolution icosphere (bottom left of Fig. 1), and let $\{\mathbf{l}_i\}_{i=1}^{N_l} \in S^2$ represent a set of label points, defined around each control point $\mathbf{c}_i$, from a high-order icosphere (bottom left of Fig. 1). Then, the objective of DDR is to learn to predict the optimal label (displacement) for each control point to ensure features of the fixed and moving mesh optimally overlap. Importantly unlike classical discrete frameworks [23,22], for which run-time is linked to the label dimensionality, DDR is far less constrained by the extent of the label space.

Fig. 1 shows an overview of our proposed approach. The first part of the network takes M and F concatenated to a single input, then passes this through a U-Net followed by a feedforward spherical network (FSN) to output $Q =$ Softmax$\left(U\right) \in \mathbb{R}^{N_c \times N_l}$: softmax probabilities for each label. The second part of the network is a CFR-RNN network, which imposes smoothness by encouraging neighboring control points to take similar labels.
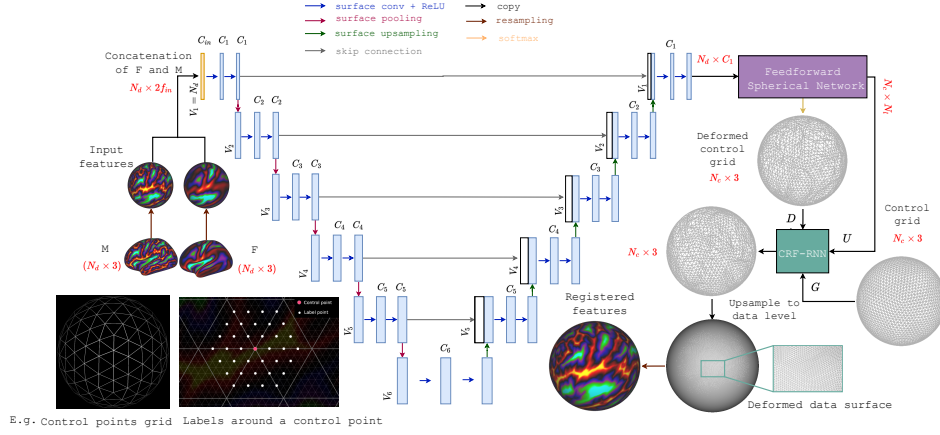
Fig. 1: DDR network architecture. The dimensions in red represent the input and the output dimensions at different levels on the network, while blue boxes represent features in the spherical space. N_d is the number of data vertices, N_c is the total number of control points, N_l is the number of labels around each control point, and f_{in} is the number of features in each subject.

Classifier Architecture The surface U-Net implements a symmetric encoder-decoder architecture with six resolution stages i, each defined at a different icosphere order. The number of icosphere vertices V_i at each stage is related to the previous one through $V_{i+1} = (V_i + 6)/4$. MoNet convolutional filters [19] with a kernel size of 3, spherical polar pseudo-coordinates, and mean aggregation operators are applied to learn C_i features at each level. In our network, we set $C_{i+1} = 2C_i$ for all i. Each convolution is followed by a LeakyReLU activation with parameter 0.2. Downsampling and upsampling are performed using the hexagonal mean pooling and transpose convolution operations proposed by [28].

The learned features at the final stage of the decoder, which is of dimensionality of $N_d \times C_1$, are then passed to a feedforward spherical network that decompresses channel dimensions over r convolutional layers, each learning $\bar{C}_i$ features, to return an output of resolution $N_d \times N_l$ (a label prediction for each location in the input mesh); then regularized by downsampling label predictions U to the desired control point resolution of $N_c \times N_l$. Once optimal labels are determined from the softmax operation Q, we deform our control grid accordingly using the spherical coordinates of the labels, i.e., $\mathbf{l}_i$.

CRF-RNN Network: On its own, the classifier is of limited use since cortical registration is an ill-posed problem with many possible solutions. Therefore, smoothness is imposed through a CRF-RNN, adapted from [29], that optimizes the following CRF energy cost function:

$$E = \sum_i Q_{(\mathbf{c}_i, \mathbf{l}_i)} + \sum_{i \neq j} \varphi \left(\mathbf{l}_{\mathbf{c}_i}, \mathbf{l}_{\mathbf{c}_j} \right). \tag{3}$$

Here, $Q_{(\mathbf{c}_i,\mathbf{l}_i)}$ represents the cost (likelihood) of deforming $\mathbf{c}_i$ to $\mathbf{l}_i$ while

$$\varphi\left(\mathbf{l}_{\mathbf{c}_i},\mathbf{l}_{\mathbf{c}_j}\right) = \mu\left(\mathbf{l}_i,\mathbf{l}_j\right) K_G\left(\mathbf{l}_{\mathbf{c}_i},\mathbf{l}_{\mathbf{c}_j}\right), \tag{4}$$

measures the pairwise cost of deforming $\mathbf{c}_i$ and $\mathbf{c}_j$ to the label points $\mathbf{l}_i$ and $\mathbf{l}_j$, respectively. Moreover, μ is a learnable label compatibility function that captures correspondences between different pairs of label points (i.e., penalizing the assignment of different labels to different control points with similar properties), and K_G is a Gaussian kernel of the form [18,29]

$$K_G\left(\mathbf{l}_{\mathbf{c}_i},\mathbf{l}_{\mathbf{c}_j}\right) = \omega\left(\mathbf{c}_i,\mathbf{c}_j\right)\exp\left(-\frac{1}{2\gamma^2}\left(\left(\mathbf{l}_{\mathbf{c}_i}-\mathbf{l}_{\mathbf{c}_j}\right)\boldsymbol{\Lambda}\left(\mathbf{l}_{\mathbf{c}_i}-\mathbf{l}_{\mathbf{c}_j}\right)\right)\right), \tag{5}$$

with $\mathbf{l}_{\mathbf{c}_i}$ being the spatial location of the deformed $\mathbf{c}_i$, ω being learnable filter weights, $\boldsymbol{\Lambda}$ being symmetric, positive-definite, kernel characterization matrix, and γ being a kernel parameter. Energy (3) is optimized using the Recurrent Neural Network (RNN) implementation of [29], which is based on learning multiple iterations of a mean-field CRF.

Final Deformed Grid and the Loss Function: The updated deformed control grid from the CRF-RNN network is upsampled to the input data icosphere order using barycentric interpolation. Then, the moving image features are resampled to the deformed data grid and compared to those of the fixed image. Optimization is driven using an unsupervised loss function (2) that is a sum of the mean-squared error and the cross-correlation. To allow for more user control over the balance between accurate alignment and smooth deformation, we add a diffusion regularization penalty on the spatial gradients of the $\boldsymbol{\Phi}$ in the form $\lambda\left(|\bigtriangledown\boldsymbol{\Phi}_{\mathbf{x}}| + |\bigtriangledown\boldsymbol{\Phi}_{\mathbf{y}}| + |\bigtriangledown\boldsymbol{\Phi}_{\mathbf{z}}|\right)$. The hexagonal filter is applied to compute $\bigtriangledown$.

Coarse to Fine Registration: As with surface registration methods, we perform multi-stage registration in the form of coarse-to-fine using two DDR networks. The first network is trained to align image features using a coarse grid of control points. The deformed grid is then upsampled to the higher level using the hexagonal upsampling method and passed to the second network that uses a higher resolution control grid.

3 Experiments

To validate our proposed framework, we conduct a series of experiments that register individual cortical surfaces to an atlas. In each case, the proposed framework was compared against classical surface registration methods: FreeSurfer [11], Spherical Demons [26], and Multimodal Surface Matching (MSM) [23,22], as well as the learning-based method S3Reg [27].

Datasets and Preprocessing: Experiments were performed using cortical surface data collected as a part of the adult Human Connectome Project (HCP) [14] resampled to a regular icosphere of order 6 using barycentric interpolation. A total number of 1110 brain MRI scans were used in the experiments with 888-111-111 split for train-validation-test and batch sizes of 1 for all. For simplicity, registration was driven solely using sulcal depth as a feature, i.e., $C_{in} = 2$; however, we point out that the network can be straightforwardly adapted to accept multi-channel features.

Implementation and Training: MoNet filters are implemented using the PyTorch Geometric library [10], while optimization is performed using ADAM [17] with a learning rate of 10^{-3}. During the coarse registration stage, we use $N_c = 162$ vertices formed from an icosphere of order 2 with a 26.7 mm distance between adjacent points. $N_l = 600$ labels are generated from an icosphere of order 5 with 3.7 mm neighboring distances. Finally, we set $C_1 = 32, \gamma = 0.7, \lambda = 1.5, r = 1$, and $\bar{C}_1 = 600$. For the fine registration, a total of $N_c = 2542$ points, formed using the vertices of an icosphere of order 4 with 6.9 mm neighboring distances, are applied. $N_l = 1000$ label vertices are generated from an icosphere of order 8 with 0.4 mm neighboring distances. Finally, we let $C_1 = 2, \gamma = 0.2, \lambda = 0.6, r = 5$, and set $[\bar{C}_i]_{i=1}^5 = [8, 16, 64, 128, 1000]$. All networks are trained on 100 epochs and we report the network performance with the best validation score.

Baseline Registration Methods: We validate against the official implementations of Spherical Demons (SD)[1], MSM Pair [2], MSM Strain[3], and the Spherical U-Net multi-stage registration algorithm (S3Reg)[4]. Freesurfer is run as part of the HCP Pipeline, so we obtain its result from there.

To achieve fair comparison across all methods, we ran parameter optimization and report performance across all runs. For SD, we run an additional 11 experiments by setting the number of smooth iterations to take $[1, 5, 10]$ and the smoothing variance σ_x to be $[1, 2, 6, 10]$. For MSM Pair, we run 22 experiments modifying the regularization penalty $\lambda \in [0.0001, 0.2]$, while for MSM Strain, we run 22 experiments with $\lambda \in [0.0001, 0.9]$. For S3Reg, we ran 7 more experiments by varying smoothness regularization at different registration levels to take the sets $[2, 5, 6, 8]$, $[2, 10, 12, 20]$, $[2, 10, 12, 14]$, $[2, 5, 12, 16]$, $[2, 10, 6, 8]$, $[2, 10, 12, 8]$, and $[2, 5, 6, 16]$; in each case training for 100 epochs and reporting the performance of the network with the best validation score. As FreeSurfer is not tunable by the user, we only report its results with default parameterization. Note that all these methods register two surfaces at 4 levels of icosphere subdivisions (coarse to fine), with S3Reg having an additional spherical transform network that ensures a diffeomorphic registration.

---

[1] https://github.com/ThomasYeoLab/CBIG

[2] Available through FSLv6.0

[3] https://github.com/ecr05/MSM_HOCR

[4] https://github.com/zhaofenqiang/SphericalUNetPackage

Evaluation Metrics: All methods are compared in terms of cross-correlation (CC) similarity, areal strain (J), shape distortion R, and run time. Here $J = \lambda_1/\lambda_2$ and $R = \lambda_1/\lambda_2$, where λ_1 and λ_2 represent the eigenvalues of the local deformation gradient F_{pqr} estimated from the deformation of each triangular face, defined by vertices: p, q, r. $\log_2 J$ is equivalent to areal distortion.

4   Results

Fig. 2 plots the similarity performances of the different runs of all methods versus the 95th percentile of the absolute value of J. At each similarity level, DDR returns distortions within the range of the best classical methods (SD and MSM Strain) and reduced extremes of strain J relative to that of S3Reg.

This trend is repeated across the full distributions of $|J|$ and $|R|$ for each comparable run. Fig. 3 plots the full histogram distributions for J and R across all runs that return a CC value of approximately 0.88. Most of the areal distortions of SD and DDR at this CC level are around zero. However, S3Reg, MSM Pair, and Freesurfer all lead to extreme distortions across subjects, with an average of 150 vertices at each subject returning $|J|$ above 2 in S3Reg compared to 0.567 for DDR.

Table. 1 further reports summary statistics for these histogram distributions in terms of the mean, max 95th, and 98th percentiles of $|J|$ and the mean CC across all runs for these parameters. Again, DDR performs comparably to SD and MSM Strain, whereas S3Reg and MSM Pair provide the worst performance. These values reflect the best performance of S3Reg (in terms of distortion) across all runs. In terms of average run time, for a PC with NVIDIA Titan RTX 24GB GPU and Intel Core i9-9820X 3.30 GHz CPU, DDR has the least GPU and CPU times across all methods.
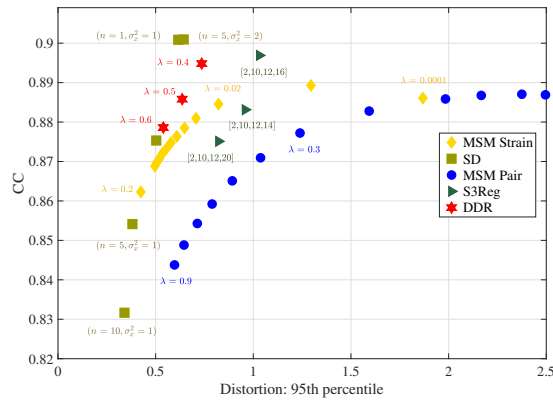

Fig. 2: Similarity performances of all methods vs. the 95th percentile of the areal distortion at multiple regularization levels across runs.
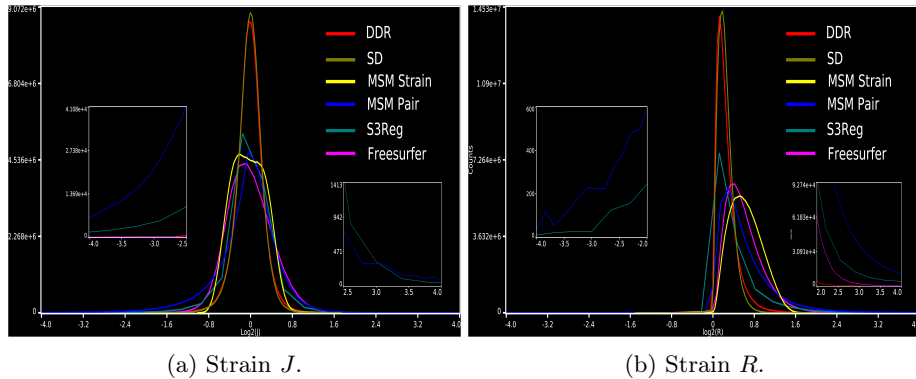
(a) Strain J.   (b) Strain R.

Fig. 3: Histogram plots comparing methods distortions J and R across all test subjects for runs with CC ~ 0.88.

Table 1: Distortions measures and average runtime for different methods obtained for CC ~ 0.88; top: classical methods, bottom: learning-based methods.

Methods	CC Similarity	Areal Distortion				Shape Distortion		Avg. Time	
		Mean	Max	95%	98%	Mean	Max	CPU	GPU
Freesurfer	0.75	0.34	11.73	0.82	1.00	0.63	6.77	30 min	-
MSM Pair	0.877	0.41	9.17	1.24	1.76	0.62	9.05	13 min	-
MSM Strain	0.880	0.27	1.06	0.70	0.66	0.64	1.93	1 hour	-
SD	0.875	0.18	2.00	0.50	0.65	0.24	1.98	1 min	-
S3Reg	0.875	0.266	22.22	0.82	1.16	0.51	21.65	8.8 sec	8.0 sec
DDR	0.878	**0.19**	**2.66**	**0.53**	**0.71**	**0.26**	**3.14**	**7.7 sec**	**2.3 sec**

Fig. 4 presents a qualitative comparison of the alignment quality for 2 subjects across all methods, together with the strain J metrics across the surface. Results show that DDR, SD, and MSM Strain generate comparable alignment for reduced distortion relative to S3Reg and MSM Pair that produce alignments that result in peaks of high distortions.

In Fig. 5, we provide qualitative comparisons of the methods registration performance with the template on subjects with atypical cortical folding patterns at CC of ~ 0.88 (except for FreeSurfer, which has 0.75 CC). Moreover, we also provide the associated strain J and strain R of these methods. The figure shows clearly that DDR, SD, and MSM Strain all produce good comparable alignments with reduced distortions, while MSM Pair and S3Reg provide alignments with peaks of high distortions.

Finally, we plot in Fig. 6 the histograms for J and R across all test subjects for methods that return a CC value of approximately 0.90 (i.e., SD = 0.901 CC, DDR = 0.895 CC, and S3Reg = 0.897 CC). Fig. 6 shows that most SD and DDR distortions at this CC level are around zero, while S3Reg leads to excess distortions.
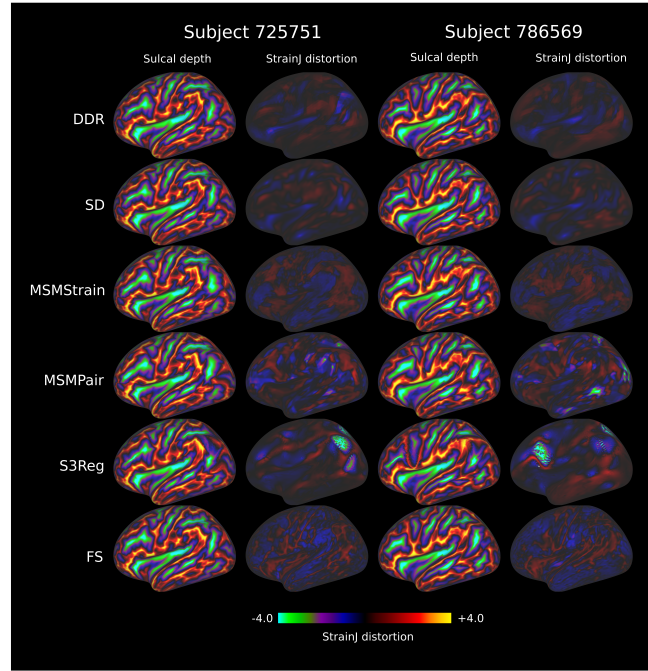
Fig. 4: Qualitative comparisons of sulcal depth alignment for 2 subjects. Results reflect distortion values presented in Fig. 3 and Table. 1. Columns 1 and 3 show registered sulcal depth maps; columns 2 and 4 reflect strain J across the surface. Cyan areas of the distortion maps highlight significant distortions.

5   Conclusions

In this work, we propose the first deep-discrete registration (DDR) framework for cortical surface alignment that aligns two surfaces by deforming a set of control points on the surface into a finite set of possible deformations. Our results show that DDR outperforms other deep learning-based cortical surface registration frameworks (S3Reg) in terms of similarity and distortions measures and provides a competitive performance to state-of-the-art conventional surface registration methods. Future work will extend DDR to multimodal alignment with high dimensional feature sets and address topographical variation.

(a)
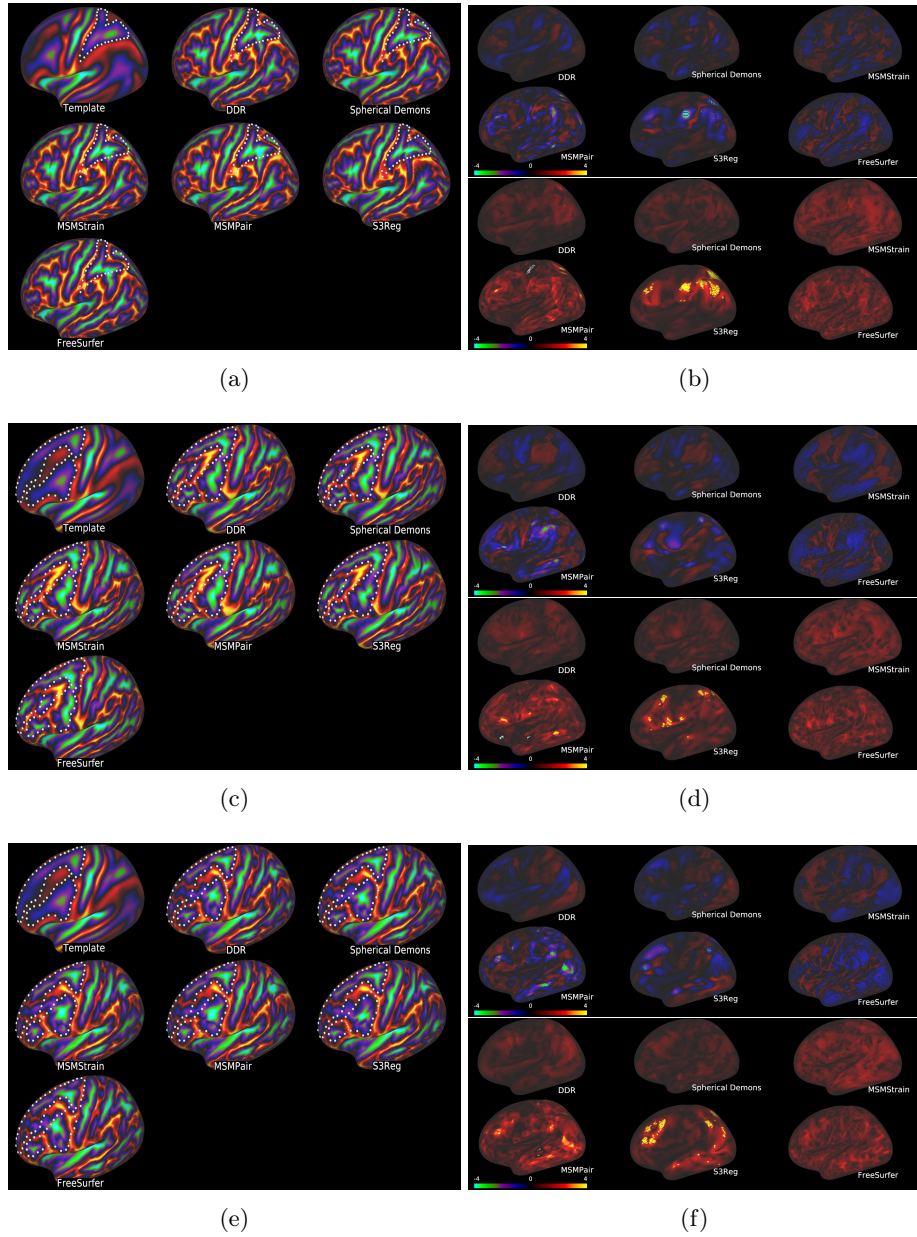

(b)

(c)

(d)

(e)

(f)

Fig. 5: (a), (c), (d) Qualitative comparisons of methods' performance on subjects with atypical cortical folding patterns at CC of ~ 0.88. (b), (d), (f) Associated Strain J (top) and Strain R (bottom).
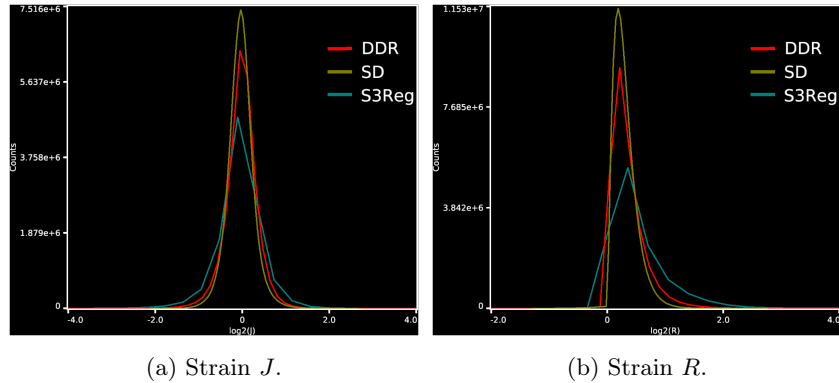
(a) Strain J.   (b) Strain R.

Fig. 6: Histogram plots comparing distortions across all test subjects for methods with CC ~ 0.90.

References

1. Amunts, K., Malikovic, A., Mohlberg, H., Schormann, T., Zilles, K.: Brodmann's areas 17 and 18 brought into stereotaxic space—where and how variable? Neuroimage **11**(1), 66–84 (2000)
2. Aoki, Y., Goforth, H., Srivatsan, R.A., Lucey, S.: Pointnetlk: Robust & efficient point cloud registration using pointnet. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7163–7172 (2019)
3. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. IEEE transactions on medical imaging **38**(8), 1788–1800 (2019)
4. Borovec, J., Kybic, J., Arganda-Carreras, I., Sorokin, D.V., Bueno, G., Khvostikov, A.V., Bakas, S., Eric, I., Chang, C., Heldmann, S., et al.: Anhir: automatic non-rigid histological image registration challenge. IEEE transactions on medical imaging **39**(10), 3042–3052 (2020)
5. Coalson, T.S., Van Essen, D.C., Glasser, M.F.: The impact of traditional neuroimaging methods on the spatial localization of cortical areas. Proceedings of the National Academy of Sciences **115**(27), E6356–E6365 (2018)
6. Dalca, A., Rakic, M., Guttag, J., Sabuncu, M.: Learning conditional deformable templates with convolutional networks. Advances in neural information processing systems **32** (2019)
7. De Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I.: A deep learning framework for unsupervised affine and deformable image registration. Medical image analysis **52**, 128–143 (2019)
8. Fan, J., Cao, X., Xue, Z., Yap, P.T., Shen, D.: Adversarial similarity network for evaluating image alignment in deep learning based registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 739–746. Springer (2018)
9. Fawaz, A., Williams, L.Z., Alansary, A., Bass, C., Gopinath, K., da Silva, M., Dahan, S., Adamson, C., Alexander, B., Thompson, D., et al.: Benchmarking geometric deep learning for cortical segmentation and neurodevelopmental phenotype prediction. bioRxiv (2021)

10. Fey, M., Lenssen, J.E.: Fast graph representation learning with PyTorch Geometric. In: ICLR Workshop on Representation Learning on Graphs and Manifolds (2019)
11. Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M.: High-resolution intersubject averaging and a coordinate system for the cortical surface. Human brain mapping **8**(4), 272–284 (1999)
12. Fu, Y., Lei, Y., Wang, T., Higgins, K., Bradley, J.D., Curran, W.J., Liu, T., Yang, X.: Lungregnet: An unsupervised deformable image registration method for 4d-ct lung. Medical physics **47**(4), 1763–1774 (2020)
13. Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., et al.: A multimodal parcellation of human cerebral cortex. Nature **536**(7615), 171–178 (2016)
14. Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., et al.: The minimal preprocessing pipelines for the human connectome project. Neuroimage **80**, 105–124 (2013)
15. Heinrich, M.P.: Closing the gap between deep and conventional image registration using probabilistic dense displacement networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 50–58. Springer (2019)
16. Heinrich, M.P., Hansen, L.: Highly accurate and memory efficient unsupervised learning-based discrete ct registration using 2.5 d displacement search. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 190–200. Springer (2020)
17. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
18. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected crfs with gaussian edge potentials. Advances in neural information processing systems **24** (2011)
19. Monti, F., Boscaini, D., Masci, J., Rodola, E., Svoboda, J., Bronstein, M.M.: Geometric deep learning on graphs and manifolds using mixture model cnns. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5115–5124 (2017)
20. Pielawski, N., Wetzer, E., Öfverstedt, J., Lu, J., Wählby, C., Lindblad, J., Sladoje, N.: Comir: Contrastive multimodal image representation for registration. Advances in neural information processing systems **33**, 18433–18444 (2020)
21. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 652–660 (2017)
22. Robinson, E.C., Garcia, K., Glasser, M.F., Chen, Z., Coalson, T.S., Makropoulos, A., Bozek, J., Wright, R., Schuh, A., Webster, M., et al.: Multimodal surface matching with higher-order smoothness constraints. Neuroimage **167**, 453–465 (2018)
23. 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.: Msm: a new flexible framework for multimodal surface matching. Neuroimage **100**, 414–426 (2014)
24. Shao, W., Banh, L., Kunder, C.A., Fan, R.E., Soerensen, S.J., Wang, J.B., Teslovich, N.C., Madhuripan, N., Jawahar, A., Ghanouni, P., et al.: Prosregnet: A deep learning framework for registration of mri and histopathology images of the prostate. Medical image analysis **68**, 101919 (2021)
25. Wang, Y., Solomon, J.M.: Deep closest point: Learning representations for point cloud registration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3523–3532 (2019)

26. Yeo, B.T., Sabuncu, M.R., Vercauteren, T., Ayache, N., Fischl, B., Golland, P.: Spherical demons: fast diffeomorphic landmark-free surface registration. IEEE transactions on medical imaging **29**(3), 650–668 (2009)

27. Zhao, F., Wu, Z., Wang, F., Lin, W., Xia, S., Shen, D., Wang, L., Li, G.: S3reg: superfast spherical surface registration based on deep learning. IEEE Transactions on Medical Imaging **40**(8), 1964–1976 (2021)

28. Zhao, F., Xia, S., Wu, Z., Duan, D., Wang, L., Lin, W., Gilmore, J.H., Shen, D., Li, G.: Spherical u-net on cortical surfaces: methods and applications. In: International Conference on Information Processing in Medical Imaging. pp. 855–866. Springer (2019)

29. Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H.: Conditional random fields as recurrent neural networks. In: Proceedings of the IEEE international conference on computer vision. pp. 1529–1537 (2015)

30. Zhou, Y., Barnes, C., Lu, J., Yang, J., Li, H.: On the continuity of rotation representations in neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5745–5753 (2019)