

A Deep Network for Joint Registration and Parcellation of Cortical Surfaces

Abstract. Cortical surface registration and parcellation are two essential steps in neuroimaging analysis. Conventionally, they are performed independently as two tasks, ignoring the inherent connections of these two closely-related tasks. Essentially, both tasks rely on meaningful cortical feature representations, so they can be jointly optimized by learning shared useful cortical features. To this end, we propose a deep learning framework for joint cortical surface registration and parcellation. Specifically, our approach leverages the spherical topology of cortical surfaces and uses a spherical network as the shared encoder to first learn shared features for both tasks. Then we train two task-specific decoders for registration and parcellation, respectively. We further exploit the more explicit connection between them by incorporating the novel parcellation map similarity loss to enforce the region-of-interests' boundary consistency, thereby providing extra supervision for the registration task. Conversely, parcellation network training also benefits from the registration, which provides a large amount of augmented data by warping one surface with manual parcellation map to another surface, especially when only few manually-labeled surfaces are available. Experiments on a dataset with more than 600 cortical surfaces show that our approach achieves large improvements on both parcellation and registration accuracy (over separately trained networks) and enables training high-quality parcellation and registration models using much fewer labeled data.

Keywords: Surface Registration · Parcellation · Deep Neural Network.

1 Introduction

Cortical surface registration and parcellation are two essential tasks in surface-based neuroimaging analysis. Cortical surface registration estimates a deformation field to align cortical features from different scans and establishes vertex-wise cortical correspondences across individuals or time points [4,13,16], thus enabling subsequent analyses, e.g., group comparison or longitudinal studies. Cortical surface parcellation learns the mapping between the cortical features and the anatomically or functionally meaningful regions, thereby parcellating the cortex into different regions of interest (ROIs) [2,11,14]. Conventionally, these two tasks are generally performed independently [6], ignoring their inherent close relationship. However, they both rely on learning effective cortical feature representations, one for inferring the deformation field and the other for predicting

parcellation labels, which means that they explore cortical features in different views and thus conducting them together enables better regularization for more meaningful and robust feature representation, and therefore, in short, they can help each other. For example, for the parcellation task, with the vertex-wise correspondence established by cortical surface registration, the manual parcellation labels from one subject or an atlas can be easily propagated to a new subject, thus helping the parcellation task to better learn the mapping from features to parcellation labels. For the registration task, the parcellated ROI boundaries can be used as an extra guidance (in addition to cortical geometric or functional features) to better learn the mapping from features to deformation field for registering two surfaces.

Therefore, in this paper, for the *first* time, we explore the idea of joint registration and parcellation of cortical surfaces through a deep spherical neural network, leveraging the spherical topology of the cerebral cortex. To extract more useful cortical features, we design a shared encoder to first learn the common features shared by both tasks. Then we train two task-specific decoders for registration and parcellation, respectively. We further exploit the more explicit connection between the two tasks which is formulated by a novel parcellation map similarity loss. This loss forces the warped predicted parcellation map of the moving surface to match the manual parcellation map of the fixed surface, e.g., an atlas, thereby providing extra supervision of the ROI boundary consistency for the registration task. Conversely, this loss can be considered as a data augmentation method that generates quasi-ground-truth parcellations for the unlabeled surfaces to help the semi-supervised learning of the parcellation network [12]. Consequently, the two tasks can mutually guide each other’s training and boost each other’s performance compared with separately trained networks, especially with limited labeled data, which is of significant importance for practical use cases where only few manually labeled surfaces are available.

2 Method

Our goal is to train a joint registration and parcellation network (**JRP-Net**) that generates the parcellation map of a given cortical surface based on its cortical features and simultaneously yields the individual-to-atlas deformation field, more accurately with few manually labeled surfaces. To this end, we formulate our approach as in Fig. 1. Our JRP-Net consists of three modules: a shared encoder (SE) for extracting mutual high-level features, a registration decoder (RD) for cortical surface registration and a parcellation decoder (PD) for surface parcellation. Let M, F be the moving surface and fixed surface defined on the sphere S^2 discretized using icosahedron subdivisions [4]. The SE module learns the mapping f_{SE} between an input surface map and extracted features $Z : Z_F = f_{SE}(F), Z_M = f_{SE}(M)$. The RD module takes the extracted features from both F and M as input and outputs the spherical velocity field $u : u = f_{RD}(Z_F, Z_M)$ and further derives the diffeomorphic deformation field $\phi = \exp(u)$ using 6 “scaling and squaring” layers as in [13]. The PD module takes

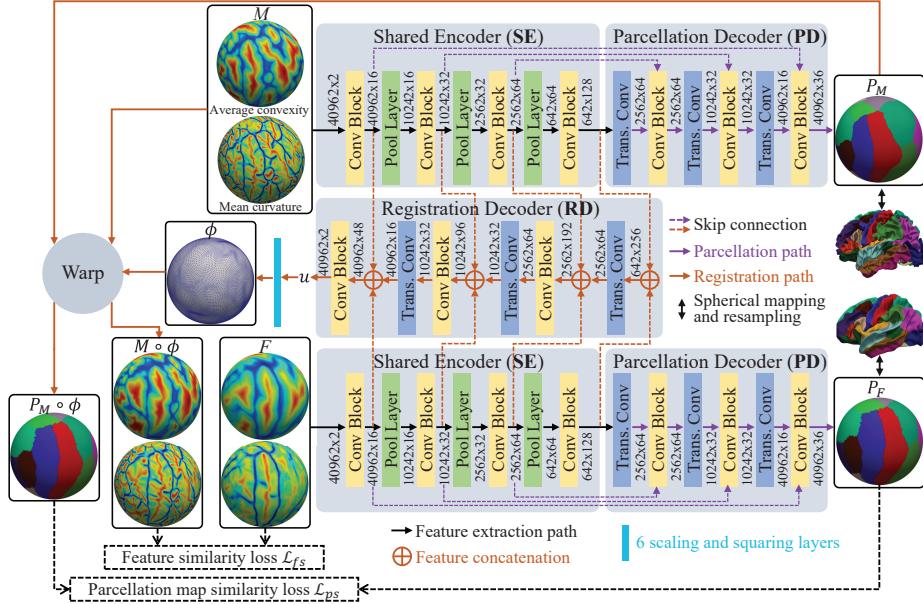


Fig. 1. Our JRP-Net for joint registration and parcellation of cortical surfaces. The input cortical surfaces are color-coded by two geometric features, i.e., average convexity (up) and mean curvature (down). The Conv Block contains repeated 1-ring-Conv+BN+ReLU. The Trans. Conv represents the spherical transposed convolution for upsampling the surface. The input and output sizes are denoted before and after each operation. See Sec. 2 for more interpretations of the math symbols.

the extracted features as input and outputs the parcellation map $P = f_{PD}(Z)$. We will detail the network architectures (Sec. 2.1), the specific losses (Sec. 2.2) and training strategy (Sec. 2.3) to train our JRP-Net effectively in the following.

2.1 Network Architecture

We construct our network based on the popular Spherical U-Net [14]. It leverages the spherical topology of cortical surfaces and extends convolution, pooling operations to the spherical space using 1-ring filter on regularly resampled spherical surfaces, and then constructs the network using corresponding spherical operations, achieving promising performance in various tasks, e.g., parcellation [14], registration [16] and harmonization [15].

Shared Encoder (SE). Our SE module shares a similar architecture with the encoder part of Spherical U-Net. It consists of 4 repeated 1-ring-Convolution +Batch Normalization (BN)+ReLU layers in four resolutions with three spherical mean pooling layers between them. This basic encoder architecture has demonstrated good representation ability in many tasks [10] and can be trained to extract shared features at various spatial scales for both registration and parcellation tasks.

Parcellation Decoder (PD). The PD module predicts the parcellation map based on the extracted feature maps Z using a similar decoder architecture as in Spherical U-Net, with modifications to the feature channels and resolutions.

Registration Decoder (RD). Our RD module is similar to a recent unsupervised learning framework [16]. It is used to fuse the separately extracted high-level features Z_F and Z_M at multiple scales and predict the corresponding deformation field. Compared to previous registration methods [16,1] that directly take the concatenation of F and M as input and output the deformation field using a single network, our novel registration network (SE+RD) has two main advantages. 1) Extracting features exclusively using SE and computing deformation field exclusively using RD makes the training objectives of the encoder and the decoder more explicit and distinct and thus easier to accomplish; 2) Our deep multi-scale feature concatenation in RD enables more effective learning of the differences of the high-level boundary features of F and M , which is important for computing the deformation field aligning F and M [7].

2.2 Loss Functions

Feature Similarity Loss (\mathcal{L}_{fs}). \mathcal{L}_{fs} is applied to the registration network to enforce the feature similarity between the warped moving surface (moved surface) and the atlas surface:

$$\mathcal{L}_{fs}(F, M, \phi) = \|F - M \circ \phi\|^2 - \lambda_{cc} \frac{\text{cov}(F, M \circ \phi)}{\sqrt{\sigma_F \cdot \sigma_{M \circ \phi}}}, \quad (1)$$

where ϕ is the learned deformation field and $M \circ \phi$ represents the moved surface maps, $\text{cov}(\cdot, \cdot)$ is the covariance, σ is the standard deviation and λ_{cc} is the weight for the correlation coefficient term.

Spherical Deformation Smoothness Loss (\mathcal{L}_s). We use the same operator ∇_s as in [16] on the spherical surface to approximate the spherical deformation's gradients on the sphere and accordingly penalize the gradients as:

$$\mathcal{L}_s(\phi) = \frac{1}{N} \sum_{n=1}^N \|\nabla_s R_{v_n}(u)\|^2, \quad (2)$$

where $R_{v_n}(u)$ denotes the local 1-ring velocity vectors of vertex v_n on a surface with N vertices. Hence, it can encourage the deformation field to be smooth.

Supervised Parcellation Loss (\mathcal{L}_{sp}). We use the weighted cross entropy loss to supervise the training of the parcellation network when manual parcellation maps are available:

$$\mathcal{L}_{sp}(P, P^*) = -\frac{1}{N} \sum_{n=1}^N W_c \log\left(\frac{\exp(P(v_n)[c])}{\sum_{j=1}^C \exp(P(v_n)[j])}\right), \quad (3)$$

where P and P^* are the predicted and manual parcellation maps respectively, W_c is the inverse proportions of the c -th ROI's area in P^* , $P(v_n)[j]$ represents the probability of vertex v_n predicted as label j and c is the manual label of

v_n . Considering the real applications of cortical surface analysis, we assume F (atlas) always has manual labels and thus $\mathcal{L}_{sp}(P, P^*) = \mathcal{L}_{sp}(P_F, P_F^*)$, when only F is labeled, otherwise $\mathcal{L}_{sp}(P, P^*) = \mathcal{L}_{sp}(P_F, P_F^*) + \mathcal{L}_{sp}(P_M, P_M^*)$ when both F and M are labeled.

Parcellation Map Similarity Loss (\mathcal{L}_{ps}). We use the multi-class Dice loss, which addresses imbalanced parcellation labels inherently [12]:

$$\mathcal{L}_{ps}(P_F, P_M, \phi) = -\frac{1}{K} \sum_{k=1}^K \text{Dice}(P_F^k, P_M^k \circ \phi) = -\frac{1}{K} \sum_{k=1}^K \frac{2 \sum_v P_F^k(v) \cdot (P_M^k \circ \phi)(v)}{\sum_v P_F^k(v) + \sum_v (P_M^k \circ \phi)(v)}, \quad (4)$$

where k indicates a ROI label (out of K ROIs) and v is the vertex location. The P_M in Eq. 4 is manually labeled parcellation map when available or predicted using f_{PD} otherwise. Hence it can provide extra ROI boundary consistency supervision for the registration task and augmented quasi-ground-truth parcellations of M for the parcellation task, when M is not labeled.

2.3 Training Strategy

To train the proposed JRP-Net, we design a progressive training strategy to learn the network parameters in an easy-to-hard manner with three steps. 1) We train the parcellation network (SE+PD) by minimizing \mathcal{L}_{sp} using the strong supervision from all available manually labeled surfaces for 2,000 iterations (i.e., 200 epochs for 10 labeled surfaces when batch size is 1, which empirically pre-trains the parcellation network sufficiently). 2) We jointly train registration and parcellation networks (SE+PD+RD) by optimizing $\mathcal{L}_{fs} + \lambda_s \mathcal{L}_s + \lambda_{sp} \mathcal{L}_{sp}$, where λ represents the weights, using all available surfaces for 10 epochs. 3) We incorporate \mathcal{L}_{ps} and optimize the full objective $\mathcal{L}_{fs} + \lambda_s \mathcal{L}_s + \lambda_{sp} \mathcal{L}_{sp} + \lambda_{ps} \mathcal{L}_{ps}$ for 100 epochs (or early stopped when obtaining stable results). Herein, \mathcal{L}_{ps} teaches PD to parcellate an unlabeled surface such that the predicted parcellation matches the inversely warped manual parcellation of the atlas via RD. Conversely, \mathcal{L}_{ps} enforces RD to correctly warp the predicted parcellation map via PD to match the atlas parcellation map, thus enabling mutual learning between the two tasks.

3 Experiments and Results

3.1 Experimental Setting

We used an infant dataset with 623 cortical surfaces, which were reconstructed via an in-house computational pipeline tailored for infants. We mapped each surface onto the sphere with 2 features at each vertex, i.e., ‘sulc’ (average convexity) and ‘curv’ (mean curvature) [3], and labeled it into 36 gyrus-based regions based on the parcellation protocol in [2]. We used the public UNC 4D infant cortical surface atlas [5] as the fixed surfaces. We randomly split the data into 60% for training, 10% for validation, and 30% for testing and made sure the longitudinal surfaces from the same subject are in the same set. All the results reported are then solely based on the hold-out test set.

Baselines. We used our SE+PD architecture in Fig. 1 (similar to Spherical U-Net [14]) as the baseline parcellation model (**BL-Parc**) trained via \mathcal{L}_{sp} , and SE+RD architecture as the baseline registration model (**BL-Reg**) trained via $\mathcal{L}_{fs} + \lambda_s \mathcal{L}_s$. Jointly training two independent BL-Parc and BL-Reg models (i.e., task-specific encoders for each task instead of the SE) is called **JRP-w/oSE-w/o \mathcal{L}_{ps}** . Similarly, **JRP-w/SE-w/o \mathcal{L}_{ps}** means jointly training with SE but without \mathcal{L}_{ps} . We also compared with 3 available registration methods, FreeSurfer [4], Spherical Demons [13], and an unsupervised learning approach [16].

Implementation Details. We implemented our method using PyTorch. We used Adam optimizer with a fixed learning rate 5e-4 for all deep learning experiments. Sulc and curv values were firstly normalized between [-1, 1]. The loss weights are $\lambda_{cc}=1.2$, $\lambda_s=10.0$, $\lambda_{sp}=2.0$ and $\lambda_{ps}=5.0$. Note that when computing \mathcal{L}_{fs} , we assign different weights for the similarities of sulc and curv maps, 0.75 for sulc and 0.25 for curv, because sulc is a more robust cortical folding measure, while curv is more variable and contains more noises for registration. These parameters are empirically determined. We used the official codes of Spherical Demons and FreeSurfer (7.1.0) for their experiments. We run all the experiments on a PC with an NVIDIA RTX2080 Ti GPU and an Intel Core i7-9700K CPU.

3.2 Results

We evaluate all the methods using the Dice overlap metric computed as in Eq. 4. For parcellation task, it is between predictions and the ground truth parcellations; for registration task, it is between moved manual parcellation maps and the fixed (atlas) parcellation map. We also provide the within-group correlation coefficient (CC) of moved cortical features as in [16,5,9] for additionally evaluating the within-group spatial normalization accuracy for registration task.

Comparison of Registration-Only Methods. As shown in Table 1, using our BL-Reg alone for cortical surface registration achieves comparable accuracy with available registration methods, but is 50+ times faster than [16], 500+ times faster than [13,4]. Note that conventional methods [13,4] do not have GPU implementation and we report the time on CPU for them in Table 1 (other methods are evaluated on GPU). We also run BL-Reg on the same CPU, it takes 0.4s and thus is still much faster than [13,4]. Compared to [16] that directly concatenates F and M as input in a coarse-to-fine manner, our SE+RD architecture fuses high-level features in deep feature space, thus avoiding the time-consuming re-interpolation of deformations in original spherical space while obtaining better results. This indicates the effective learning of high-level features in SE and the deformation computation based on boundary differences in RD. To additionally validate the topology-preserving registrations, we computed the folded triangles [8] in moved surfaces and found there is no folded triangles for all methods.

Ablation Study on Shared Encoder (SE). Comparing JRP-w/SE-w/o \mathcal{L}_{ps} with BL-Parc and BL-Reg in Table 1, we can see that jointly training SE improves Dice over separately training two networks by 0.86% and 0.84% for parcellation and registration tasks respectively, which indicates that SE learns and exploits the shared useful features between the two tasks successfully.

Table 1. Average (standard deviation) performance of cortical surface parcellation and registration on the hold-out test set using different models trained with all manual parcellations in the training set. Parc. means parcellation, Reg. means registration.

Models	Parcellation	Registration			Run time
		Dice (%)	Dice (%)	CC_sulc	
Parc. only (BL-Parc)	88.48(3.68)	-	-	-	0.01s
Reg. only (FreeSurfer [4])	-	76.69(5.52)	0.7788(0.0549)	0.2792(0.0813)	~30min
Reg. only (Spherical Demons [13])	-	76.58(5.83)	0.7825(0.0675)	0.2720(0.1237)	~1.5min
Reg. only (Zhao et al. [16])	-	77.03(5.66)	0.7859(0.0526)	0.2955(0.0810)	~10s
Reg. only (BL-Reg) (Ours)	-	77.34(3.82)	0.7946(0.0374)	0.2968(0.0796)	0.16s
JRP-w/oSE-w/ \mathcal{L}_{ps} (Ours)	88.48(3.68)	88.78(3.12)	0.8209(0.0339)	0.3313(0.0794)	0.01s+0.16s
JRP-w/SE-w/o \mathcal{L}_{ps} (Ours)	89.34(2.47)	78.18(4.39)	0.7996(0.0466)	0.3095(0.0953)	0.16s
JRP-w/SE-w/ \mathcal{L}_{ps} (Ours)	89.98(2.53)	90.23(2.56)	0.8288(0.0317)	0.3359(0.0676)	0.16s

Ablation Study on Parcellation Map Similarity Loss (\mathcal{L}_{ps}). Table 1 shows incorporating \mathcal{L}_{ps} substantially improves the performance, with >10% and 1.5% Dice improvement over separately trained networks for registration and parcellation, respectively. For JRP-w/oSE-w/ \mathcal{L}_{ps} , the \mathcal{L}_{ps} is computed between two manual parcellations according to Eq. 4, thus it does not provide any information for training the independent parcellation network but provides maximum extra supervision for the registration network and leads to huge improvement on registration accuracy. Further with SE (JRP-w/SE-w/ \mathcal{L}_{ps}), \mathcal{L}_{ps} back-propagates gradients from RD to SE and thus also enhance the performance of parcellation task by using more effective features in SE.

How Registration and Parcellation Help Each Other. We simulate the common practical use case where only few manually labeled surfaces are available. We randomly used N surfaces' manual parcellations to train BL-Parc, BL-Reg and JRP-Net (JRP-w/SE-w/ \mathcal{L}_{ps}) on the training set. To achieve more reliable results, we repeated the experiment 10 times for each model (each time with randomly selected manual parcellations). Fig. 2 shows the results on the test set averaged over 10 times experiments. We can see that in the extreme case where only atlases' parcellations are available (11 atlases in UNC 4D atlas [5]), our method still achieves good performance, with large improvements over independently trained models. This demonstrates that registration did help parcellation by providing estimated supervision from augmented parcellation maps on unlabeled surfaces. Conversely, the registration performance is also boosted by the parcellation task (when $N>3$) even using predicted parcellations to en-

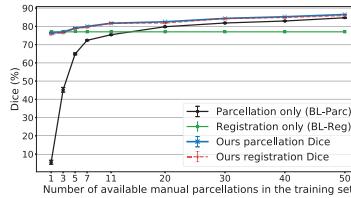


Fig. 2. Test Dice of the models trained with different numbers of manual parcellations.

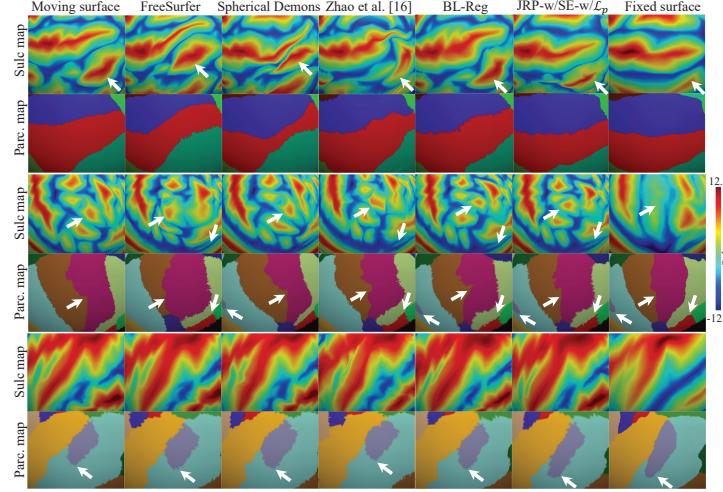


Fig. 3. For each case (left to right): the moving surface' sulc map (first row) and manual parcellation map (second row), corresponding moved maps by different methods, and fixed surface' maps. Note that all the methods fairly take the features (sulc and curv) as input and output the deformation field. The manual parcellation maps are only for validating the registration performance and were not used to drive the registration.

force the anatomical consistency. Also note that the lines of registration Dice and parcellation Dice of our JRP-Net in Fig. 2 are highly coincident. A paired t-test with a significance level 0.05 on the 10 times experiments shows there is no significant difference between them, which could be a strong evidence that the two tasks mutually guide each other's training to successfully learn the optimal features using SE and find the optimal solution for both tasks. Fig. 3 shows some practical registration cases from the test set that available registration methods all fail to align but are correctly aligned using our JRP-Net ($N=50$). The suboptimal results of available methods are an inherent problem because they only use feature similarity as the registration objective, while our JRP-Net incorporating the parcellation map similarity can effectively solve this problem.

4 Conclusion

In this paper, we propose the JRP-Net for joint registration and parcellation of cortical surfaces, which are connected by the shared encoder and parcellation map similarity loss. By leveraging the inherent relation between the two tasks, our shared encoder is effective in extracting more meaningful features shared by both tasks. The parcellation map similarity loss further enables effective mutual training between the two tasks. Both visual and quantitative results show large improvements of our method in both registration and parcellation over separately learned networks. In future, we will release our model for enhancing cortical surface registration and parcellation for the research community.

References

- Cheng, J., Dalca, A.V., Fischl, B., Zöllei, L., Initiative, A.D.N., et al.: Cortical surface registration using unsupervised learning. *NeuroImage* **221**, 117161 (2020)
- Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al.: An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage* **31**(3), 968–980 (2006)
- Fischl, B.: Freesurfer. *Neuroimage* **62**(2), 774–781 (2012)
- 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)
- Li, G., Wang, L., Shi, F., Gilmore, J.H., Lin, W., Shen, D.: Construction of 4d high-definition cortical surface atlases of infants: Methods and applications. *Medical image analysis* **25**(1), 22–36 (2015)
- Li, G., Wang, L., Yap, P.T., Wang, F., Wu, Z., Meng, Y., Dong, P., Kim, J., Shi, F., Rekik, I., et al.: Computational neuroanatomy of baby brains: A review. *NeuroImage* **185**, 906–925 (2019)
- Liu, L., Hu, X., Zhu, L., Heng, P.A.: Probabilistic multilayer regularization network for unsupervised 3d brain image registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 346–354. Springer (2019)
- Möller, T.: A fast triangle-triangle intersection test. *Journal of graphics tools* **2**(2), 25–30 (1997)
- 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)
- Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
- Wu, Z., Li, G., Wang, L., Shi, F., Lin, W., Gilmore, J.H., Shen, D.: Registration-free infant cortical surface parcellation using deep convolutional neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 672–680. Springer (2018)
- Xu, Z., Niethammer, M.: Deepatlas: Joint semi-supervised learning of image registration and segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 420–429. Springer (2019)
- 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)
- Zhao, F., Wu, Z., Wang, L., Lin, W., Gilmore, J.H., Xia, S., Shen, D., Li, G.: Spherical deformable u-net: Application to cortical surface parcellation and development prediction. *IEEE Transactions on Medical Imaging* (2021)
- Zhao, F., Wu, Z., Wang, L., Lin, W., Xia, S., Shen, D., Li, G., Consortium, U.B.C.P., et al.: Harmonization of infant cortical thickness using surface-to-surface cycle-consistent adversarial networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 475–483. Springer (2019)
- Zhao, F., Wu, Z., Wang, L., Lin, W., Xia, S., Shen, D., Li, G., Consortium, U.B.C.P., et al.: Unsupervised learning for spherical surface registration. In: International Workshop on Machine Learning in Medical Imaging. pp. 373–383. Springer (2020)