Learning joint surface reconstruction and segmentation, from brain images to cortical surface parcellation
Karthik Gopinath, Christian Desrosiers, Herve Lombaert
ETS Montreal, Canada

A R T I C L E   I N F O
Keywords:
Surface reconstruction
Cortical parcellation

A B S T R A C T
Reconstructing and segmenting cortical surfaces from MRI is essential to a wide range of brain analyses. However, most approaches follow a multi-step slow process, such as a sequential spherical inflation and registration, which requires considerable computation times. To overcome the limitations arising from these multi-steps, we propose SegRecon, an integrated end-to-end deep learning method to jointly reconstruct and segment cortical surfaces directly from an MRI volume in one single step. We train a volume-based neural network to predict, for each voxel, the signed distances to multiple nested surfaces and their corresponding spherical representation in atlas space. This is, for instance, useful for jointly reconstructing and segmenting the white-to-gray-matter interface and the gray-matter-to-CSF (pial) surface. We evaluate the performance of our reconstruction and segmentation method with a comprehensive set of experiments on the MindBoggle, ABIDE and OASIS datasets. Our reconstruction error is found to be less than 0.52 mm and 0.97 mm in terms of average Hausdorff distance to the FreeSurfer generated surfaces. Likewise, the parcellation results show over 4% improvements in average Dice with respect to FreeSurfer, in addition to an observed drastic speed-up from hours to seconds of computation on a standard desktop station.

1. Introduction
Brain surface analysis requires the accurate reconstruction and segmentation of cortical surfaces from MRI volumes (Querbes et al., 2009; Glasser et al., 2016; Billot et al., 2023). Standard surface processing pipelines for reconstructing cortical surfaces (Fischl et al., 2004; Dahnke et al., 2013; Kim et al., 2005; Kriegeskorte and Goebel, 2001; Shattuck and Leahy, 2002) and hippocampus (Styner et al., 2006; Puonti et al., 2016) follow a sequence of costly operations that often include: white matter segmentation, surface mesh generation from the segmentation masks, mesh smoothing and projection to a sphere, topological correction of the projected mesh, and fine-tuning of re-projected mesh on the segmented volume. The cortical surfaces are segmented into neuroanatomical parcels in a subsequent and highly-expensive step. Such segmentation can take several hours to finish, which involves the re-projection of each surface to a sphere via a metric-preserving inflation process, registration to a spherical atlas (Fischl et al., 1999; Klein and Tourville, 2012) and cortical parcellation using atlas labels (Desikan et al., 2006).

Recently, Henschel et al. (2020) developed a framework called FastSurfer using deep learning that accelerates the processing times for brain segmentation and spectral embedding for registration to a spherical atlas. Despite reducing computation times considerably, this pipeline still performs reconstruction and segmentation in two consecutive steps. To overcome this limitation, Cruz et al. (2021), Ma et al. (2021), Hoopes et al. (2022) and Bongratz et al. (2022) proposed a deep learning based models for cortical surface reconstruction. This method draws inspiration from Park et al. (2019), and samples points on a reference grid of arbitrary resolution to reconstruct a surface without the need for an explicit segmentation step. However, this process is highly expensive in terms of both computation and memory for detailed surfaces with hundreds of thousands of points. Additionally, DeepCSR only performs surface reconstruction, and cannot be used for parcellation which is one of the most time-costly operations in conventional neuroimaging pipelines (Fischl et al., 2004). Similarly, (Ma et al., 2021) proposed a 3D deep learning framework for pial surface reconstruction. Approaches that directly operate (Wu et al., 2019; López-López et al., 2020) or learn on surface data (Gopinath et al., 2019; Lombaert et al., 2015; Gopinath et al., 2020) have been used for cortical parcellation, but are designed to process single surfaces separately for each subjects. Spectral embeddings of surface meshes in a low-dimensional space can be exploited to predict cortical parcellation labels (Lombaert et al., 2015; Germanaud et al., 2012). However, a major limitation of these early learning approaches is that mesh nodes are considered separately instead of jointly. Recent work has proposed using graph convolutional
networks (GCN) (Gopinath et al., 2019; Wu et al., 2019; He et al., 2020; Gopinath et al., 2020) to exploit the connectivity information of a mesh graph. Similarly, DBPN (Zhang and Wang, 2019) introduced a two-stage spatial graph convolution network and ASEGAT (Li et al., 2022) proposed to use an anatomically constrained squeeze-and-excitation self attention graph network to perform a cortical parcellation in the original mesh space. While such strategy provides an accurate and faster parcellation of the cortical surface, it is highly sensitive to the quality of the surface reconstruction step. For instance, small errors or holes in the reconstructed cortical mesh may causing graph convolution methods to fail at parcellation tasks. In a similar manner, SPHARM-Net (Ha and Lyu, 2022a) uses a spherical harmonics-based convolutional neural network for vertex-wise cortical parcellation. Furthermore, recent parcellation methods such as (Zhao et al., 2019, 2021; Parvathaneni et al., 2019; Ha and Lyu, 2022b) rely on inflated spheres of the cortical surface to perform spherical convolution-based parcellation. Extracting these cortical meshes, inflating them and registering them to spheres are typically computationally expensive steps, often ignored by these methods.

We propose SegRecon, a novel deep learning model for the joint reconstruction and parcellation of cortical surfaces. Our end-to-end model works directly on MRI volumes and predicts a dense set of surface points along with their corresponding parcellation labels (see Fig. 1). A 3D CNN based UNet (Çiçek et al., 2016) predicts for each input voxel of the volume, the brain hemisphere, its signed distance to the nested surfaces (white matter and pial surfaces) of that hemisphere, and the spherical coordinates in the registered atlas space. We use this predicted nested signed distance surfaces for surface reconstruction and the spherical coordinates for surface parcellation. By learning to solve this multi-task problem, the network can be used to reconstruct and segment brain surfaces efficiently and in a topologically-accurate manner (Bazin and Pham, 2007).

The main contributions of our work are the following:

- To our knowledge, we propose the first deep learning model for the joint reconstruction of multiple nested surfaces and their segmentation, with an application on brain surfaces. This contrasts with existing approaches, which either perform surface reconstruction and segmentation in separate steps (Henschel et al., 2020), are limited to reconstruction (Cruz et al., 2021), or require a pre-generated mesh as input (Gopinath et al., 2019; Wu et al., 2019; Lombaert et al., 2015);
- Compared to the current surface reconstruction learning approaches (such as DeepCSR), the proposed method implements a fully-convolutional architecture that densely predicts the location of all input voxels relative to cortical surfaces, in a single feed-forward pass. Our method also leverages a novel surface reconstruction loss that controls the minimum and maximum distance between white matter and pial surfaces (i.e., cortical thickness), thereby ensuring that these surfaces never cross;
- We present a comprehensive set of experiments involving three publicly-available datasets, i.e., MindBoggle (Klein et al., 2017), OASIS (Marcus et al., 2007) and ABIDE-I (Di Martino et al., 2014), that compare the surface reconstruction and segmentation accuracy of our method against several baselines. Our results demonstrate the major advantages of our method over standard brain surface analysis pipelines are recent approaches for cortical parcellation. With respect to the widely-used FreeSurfer software, our method generates surfaces with an average Hausdorff distance less than 0.52 mm, while boosting the parcellation Dice by 4.3% and being several orders of magnitude faster.

The proposed work is a substantial extension of our MICCAI article (Gopinath et al., 2021). The major extensions are: firstly, the reconstruction of multiple nested surfaces extracts both white matter and pial surfaces; secondly, a novel surface reconstruction loss ensures that the white and pial surfaces never cross; finally, the evaluation of the surface reconstruction is made on two independent test sets with varying factors, such as image acquisition, processing parameters, age, and cortical surface alterations.

In the next section, we present our proposed joint reconstruction and segmentation approach, describing in detail the network architecture, training losses and inference steps. The performance of our
method is then evaluated on the MindBoggle (Klein et al., 2017), OASIS (Marcus et al., 2007) and ABIDE-I (Di Martino et al., 2014) datasets. The ablation study and comparison to the state-of-art in these experiments demonstrate important benefits of our method.

2. Method

An overview of SegRecon is shown in Fig. 2 with the end-to-end surface reconstruction and segmentation steps illustrated. Let \( D = \{(X_i, S^W, S^P, Y_i)\}_{i=1}^{n} \) be a training set where each example is composed of: a 3D volume \( X_i \in \mathbb{R}^{3} \) with voxel set \( \Omega \subset \mathbb{Z}^3 \), a white matter surface \( S^W \in \mathbb{R}^{m \times 3} \) defined by \( m \) points, a pial surface \( S^P \in \mathbb{R}^{n \times 3} \) defined by \( n \) points, and a one-hot encoded segmentation \( Y_i \in \{0, 1\}^{m \times oc} \) of the white matter surface, where \( c \) is the number of segmentation classes.

The goal is to train a model \( f \) parameterized by \( \theta \) which maps an input 3D volume \( X \) to a white matter surface \( S^W \) with corresponding parcel labels \( Y \), and a pial surface \( S^P \).

One of the main challenges in this task comes from the disparity between the well-defined grid space of images \( X \) and the domain of surfaces \( S^W \) and \( S^P \), where the number of points can vary from one surface to another and points can lie anywhere in 3D space. In Cruz et al. (2021), this problem is solved by giving as input to model \( f \) both the image \( X \) and a query point \( p \in \mathbb{R}^3 \) in the template space. The model then predicts if \( p \) belongs to the surface in \( X \) or, alternatively, its distance to this surface. To reconstruct a surface at inference time, the model is queried over a fixed reference grid. While this strategy allows reconstructing a surface at arbitrary resolution, it suffers from two important drawbacks. First, since the template points which can be in the hundreds of thousands are queried independently, reconstructing a surface requires significant time and computation. Moreover, unlike dense prediction approaches, this strategy does not exploit the spatial relationship between points. Last, because feature maps need to be computed for the whole 3D volume \( X \), it also needs a large amount of memory.

To overcome these drawbacks, we instead learn a model that densely projects voxels of the input volume \( X \) to a spherical atlas space. Specifically, \( f \) maps each voxel \( v \in \Omega \) to a vector

\[
f_v(X) = [d_w^v, d_p^v, \phi_v, \gamma_v, h_{l\ell}^w, h_{l\ell}^p].
\]

where \( d_w^v \) is the signed distance from \( v \) to its nearest surface point, such that \( d_w^v \leq 0 \) if \( v \) is inside the surface else \( d_w^v > 0 \). Similarly, \( d_p^v \) is the signed distance from \( v \) to its nearest pial surface. \( \phi_v, \gamma_v \) are the polar angle and azimuthal angle of \( v \in \Omega \) defining its position in the spherical atlas, and \( h_{l\ell}^w, h_{l\ell}^p \in [0, 1] \) are the probabilities that \( v \) is in the left hemisphere, right hemisphere and background, respectively. Here, polar and azimuthal angles are normalized so to lie in the \([-1, 1] \) range. A further topological correction step (Bazin and Pham, 2007) over the predicted surface points prevents the extraction of critical points yielding topological defects. The resulting white and pial surfaces are defined implicitly as the 0-levelset of their respective distance map and can be efficiently reconstructed using an iso-surface extraction algorithm such as the Marching Cubes (Lorensen and Cline, 1987).

2.1. Training the model

Denote \( \hat{f} = f(X_i) \) as the predicted vector for an image \( X_i \) and let \( f_i \) be the corresponding ground-truth. To train the model, we use the following loss function

\[
L(\theta; D) = \sum_{i=1}^{n} \epsilon_{\text{warp}}(\hat{f}_i, f_i) + \epsilon_{\text{surf}}(\hat{f}_i, f_i) + \lambda_1 \epsilon_{\text{hemi}}(\hat{f}_i, f_i) + \lambda_2 \epsilon_{\text{thick}}(\hat{f}_i).
\]

The first loss term, \( \epsilon_{\text{warp}} \), ensures that the signed distance of voxels to the white matter surface, as well as their position in the spherical atlas space, are well predicted. Dropping index \( i \) for simplicity, it is defined as

\[
\epsilon_{\text{warp}}(\hat{f}, f) = \sum_{v \in \Omega} \mathbb{I}_{P} \left[ (d_w^v - d_w^v)^2 + \min(\phi_v - \phi_v')^2, (1 + \gamma_v - \gamma_v')^2 \right].
\]

where \( \mathbb{I}_{P} \) is the indicator function, equal to 1 if predicate \( P \) is true else, 0 otherwise. We only consider voxels within a distance of \( \epsilon \) to the nearest white matter surface point in order to focus learning on relevant points close to our surface. This is achieved with function \( \mathbb{I}_{|d_w^v| < \epsilon} \) in Eq. (3). Additionally, we consider the non-uniqueness of spherical coordinates (e.g., \( -\pi \equiv \pi \)) by computing, for each angle, the minimum \( L_\ell \) distance from the predicted angle or this angle plus \( 1 \) to the ground-truth. Using the minimum function of the two cyclic angles aids in optimizing the training of both polar and azimuth angles.

The distance \( d_w^v \) is, therefore, defined between the center of the voxel \( v \) in image space and the nearest point on white matter surface \( S^W \). In this work, we use the white matter surface mesh generated by FreeSurfer for training. The sign of \( d_w^v \) is determined using the white-matter segmentation mask, with voxels inside the white matter having
a negative distance. Likewise, the ground-truth spherical coordinates \( \phi \) and \( \gamma \) are obtained using FreeSurfer (Fischl et al., 2004) with the Desikan–Killiany–Tourville (DKT) atlas (Klein and Tourville, 2012).

The second loss term, \( \ell_{\text{vol}} \), ensures that the signed distance of voxels to the nearest pial surface is predicted accurately. We define it as

\[
\ell_{\text{vol}}(\tilde{d}, \tilde{t}) = \sum_{v \in \Omega} \left[ \max(\tilde{d}_v - \tilde{d}_v^w, 0)^2 + \max(\tilde{d}_v^w - \tilde{d}_v - \tilde{t}_v^w, 0)^2 \right].
\]

(6)

where \( \tilde{d}_v \) is the distance defined between the center of the voxel \( v \) and its closest point on pial surface \( S^p \) obtained by FreeSurfer pial meshes. The sign of the distance \( \tilde{d}_v \) is estimated using the brain segmentation mask with voxels inside the brain mask having negative distance. Similar to \( \ell_{\text{surf}} \) in Eq. (3), \( 1_p \) is the indicator function used to restrict the training to the useful voxels within a distance of \( \epsilon \) to the closest pial surface.

The third term, \( \ell_{\text{hem}} \), enables the network to predict if a voxel \( v \) lies in the left hemisphere (\( lh \)), in the right hemisphere (\( rh \)) or is outside both (bg). This selection is necessary since the surface atlas is defined separately for each hemisphere. Here, we use cross-entropy as loss function:

\[
\ell_{\text{hem}}(\tilde{d}, \tilde{t}) = -\sum_{v \in \Omega} \sum_{c \in \{lh, rh, bg\}} h_c^v \log \tilde{h}_c^v.
\]

(7)

The ground-truth hemisphere masks are once again obtained from FreeSurfer.

Since the white matter and pial surfaces are reconstructed from two separate predictions, it may happen that predicted surfaces are near to the ground-truth while still violating anatomical constraints. For example, in very thin regions of the cortex, the reconstructed surfaces may overlap or even cross each other. To avoid this problem, we add a last term to the loss function, \( \ell_{\text{thick}} \), which controls the minimum and maximum distance between the surfaces:

\[
\ell_{\text{thick}}(\tilde{d}, \tilde{t}) = \sum_{v \in \Omega} \left[ \max(\tilde{d}_v - \tilde{d}_v^{w}, \tilde{t}_v - \tilde{t}_v^{w}, 0) + \max(\tilde{d}_v^{w} - \tilde{d}_v - \tilde{t}_v, 0) \right].
\]

(8)

where \( t_{\text{min}} \) and \( t_{\text{max}} \) are the minimum and maximum allowed inter-surface distances (cortical thickness). These hyperparameters can be set based on the dataset ground-truth or some clinical reference. For instance, a cortical thickness range from 1 to 4.5 mm is reported in Fischl and Dale (2000). We use similar values in this work: \( t_{\text{min}} = 0 \) and \( t_{\text{max}} = 5 \) mm. Effectively, this prevents surfaces from crossing each other or separating beyond 5 mm. As defined in Eq. (6), this penalty is only calculated for voxels inside the pial surface, i.e., voxels \( v \) such that \( \tilde{d}_v^w \leq 0 \).

2.2. Surface reconstruction and segmentation

Once the network is trained, it can be used to reconstruct and segment surfaces directly from a test volume \( X \). First, we feed the volume to the network to obtain a prediction vector for all voxels. Since the network is fully-convolutional, this can be done efficiently in a single feed-forward pass. Next, we apply a small-width Gaussian filter on the predicted 3D white matter surface distance map \( \hat{d}_v^w \) using a single convolution operation and employ a topological correction step (Bazin and Pham, 2007) to overcome any defects in the surface. The same steps are followed to extract the 3D pial surface using distance map \( \hat{d}_v^p \). The surface is reconstructed using the Marching Cubes algorithm (Lorensen and Cline, 1987) on the 0-level set of its predicted signed distance map, smoothed with a Gaussian kernel and topologically corrected with the method of Bazin and Pham (2007).

To segment the surface, we first compute the near-surface voxels in each hemisphere as follows:

\[
S^v = \left\{ v \in \Omega \mid |d_v| \leq \epsilon \land c = \arg \max \tilde{h}_c^v \right\}, \quad c \in \{lh, rh\}.
\]

(9)

Next, we find the nearest-neighbor to a given reference atlas \( R^c \) for all the near-surface voxels \( v \in S^v \) using their predicted angles \( \phi_v \) and \( \gamma_v \). The segmentation labels from this reference atlas \( R^c \) are later projected back to the near-surface voxels \( S^v \). A majority voting across multiple atlases is eventually applied to obtain the final parcellation labels.

2.3. Implementation details

The overall architecture of StoReCON is shown in Fig. 2. As an input, we provide the skull-stripped, intensity normalized 3D T1-MRI volume. We use a 3D-UNet architecture similar to Çiçek et al. (2016) in order to map the input voxel to a point in the spherical atlas space.

We apply a softmax activation in the first three output channels to predict the probability of a voxel belonging to the background, left hemisphere, or right hemisphere. The polar and azimuthal angles, \( \phi_v \) and \( \gamma_v \), are predicted with a tanh activation. The FreeSurfer generated cortical meshes and the corresponding registered spheres are used to pre-compute the surface distances and interpolate the ground-truth spherical coordinates for each voxels respectively. The last two output
channels produce the signed distance map \( \hat{d}^v \) and \( \hat{d}^w \) for each voxel \( v \). The \( \lambda_1 \) and \( \lambda_2 \) are set to 0.05 and 1 respectively. The network parameters, \( \theta \), are optimized using a stochastic gradient descent with the Adam optimizer (Kingma and Ba, 2014). During training, we pick the maximum distance of surface voxels in Eq. (3) to be \( \epsilon = 2 \, \text{mm} \), which corresponds to the overall average thickness reported in Fischl and Dale (2000). We employ the (Lewiner et al., 2003) implementation of the marching cubes in scikit-image on the 0-levelset of its predicted signed distance map, smoothed with a Gaussian kernel of \( \sigma = 0.5 \, \text{mm} \) to ensure that the reconstructed surfaces are one-connected meshes with no cracks or tears. We use an Intel I7 desktop machine with 16Gb RAM and Nvidia RTX 2080 GPU for our work.

### 3. Experiments and results

To benchmark the performance of our method, we use one of the largest publicly-available dataset containing manual surface parcellation, MindBoggle (Klein et al., 2017). This dataset contains 101 subjects with MRI volumes, FreeSurfer processed meshes, and 32 manually-labeled cortical parcels. We split the dataset randomly into training, validation, and testing using a ratio of 70-10%-20%. We also use the ABIDE-I (Di Martino et al., 2014) and OASIS (Marcus et al., 2007) databases as independent test sets to measure the surface reconstruction error of our method with FreeSurfer-generated cortical and pial surfaces. The ABIDE dataset contains brain surfaces for 1035 subjects with 530 healthy and 505 autism spectrum disorder (ASD) subjects. Likewise, the OASIS dataset comprises a total of 226 brain surfaces from 93 healthy subjects and 133 subjects with Alzheimer’s disease (AD). These two datasets are used to validate the robustness of the method to various factors, including image acquisition, pre-processing parameters, and dataset quality.

As found in Table 1, our method obtained a mean AAD below 0.35 mm and mean HD less than 0.93 mm for the white matter surface, in both datasets. Similarly, reconstructed pial surfaces in both datasets have a mean AAD no greater than 0.58 mm and mean HD less than 1.53 mm for pial surfaces in both datasets. These results, obtained for subjects of very different ages and with cortical alterations, are comparable to those obtained for the MindBoggle test set. The qualitative results in Fig. 4 validate the visual similarity in surface reconstruction of our method, across datasets. Furthermore, we present a visualization of the surface reconstruction geometry overlaid with the curvatures of the white surface on inflated surfaces (Fig. 5).

In a first experiment, we validate the benefit of using a signed distance (SD) map, when reconstructing the white and pial cortical surfaces, by comparing it against using a binary mask (BW). To predict the binary mask, we use an architecture similar to the one in Fig. 2 where the last two output layers (corresponding to white matter and pial surfaces) are generated with sigmoid activations. As reported in Table 1, an improvement in CD from 5.0 mm to 1.7 mm is obtained when a signed distance map is used for white matter surface reconstruction. A similar improvement over the binary mask approach is also observed in terms of AAD and HD. Qualitative results, presented in Fig. 3, show that the meshes reconstructed using signed distance maps are more regular and closer to FreeSurfer-generated meshes, compared to those obtained with binary masks.

Our surface reconstruction method was also tested on the OASIS and ABIDE datasets, not used for training, to evaluate its robustness. As found in Table 1, our method obtained a mean AAD below 0.35 mm and mean HD less than 0.93 mm for the white matter surface, in both datasets.

### 3.2. Effect of reference atlas on parcellation

Instead of predicting class probabilities for each voxel, as in standard 3D segmentation networks, the proposed network predicts spherical atlas coordinates (i.e., angles \( \phi \) and \( \gamma \)). This has two important advantages: (i) considerably reducing the number of outputs for the

**Table 1**

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Left White Matter</th>
<th>Right White Matter</th>
<th>Left Pial</th>
<th>Right Pial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CD</td>
<td>AAD</td>
<td>HD</td>
<td>CD</td>
</tr>
<tr>
<td>SD map</td>
<td>1.760(0.23)</td>
<td>0.337(0.02)</td>
<td>0.850(0.06)</td>
<td>1.772(0.20)</td>
</tr>
<tr>
<td>BW map</td>
<td>5.019(0.35)</td>
<td>0.630(0.04)</td>
<td>3.104(0.60)</td>
<td>4.956(0.27)</td>
</tr>
<tr>
<td>OASIS CN</td>
<td>1.820(0.68)</td>
<td>0.360(0.05)</td>
<td>0.805(0.12)</td>
<td>1.726(0.45)</td>
</tr>
<tr>
<td>OASIS AD</td>
<td>2.003(0.75)</td>
<td>0.396(0.05)</td>
<td>0.843(0.11)</td>
<td>1.973(0.70)</td>
</tr>
<tr>
<td>ABIDE CN</td>
<td>1.995(0.80)</td>
<td>0.357(0.05)</td>
<td>0.934(0.17)</td>
<td>1.920(0.52)</td>
</tr>
<tr>
<td>ABIDE AD</td>
<td>1.946(0.70)</td>
<td>0.353(0.04)</td>
<td>0.936(0.16)</td>
<td>1.931(0.68)</td>
</tr>
</tbody>
</table>
number of classes to only two, and (ii) providing information on the precise location of a voxel inside a parcel instead of simply measuring if a voxel is inside a parcel or not. As we will show in the next section, this continuous prediction strategy leads to a higher accuracy compared to a standard segmentation approach. However, the final predicted labels depend on the reference atlas.

For assessing the impact of the reference atlas on segmentation performance, we randomly select five subjects from the training set and use the spherical coordinates and parcellation labels of their surface mesh nodes as different atlases Ref_1, …, Ref_5. Table 2 reports the mean Dice score obtained for test subjects using each of the five atlases. While a high accuracy is obtained in all cases, the performance also varies significantly from 84.60% to 87.33%.

To provide greater robustness to the choice of an atlas, we apply a simple multi-atlas strategy in which a separate prediction is obtained for each atlas, and individual predictions are combined using majority voting. The number of atlas or references picked for combining labels can impact the average Dice overlap of parcellation. Table 3 reports majority voting computed across multiple atlases. The subscripts indicate the number of atlases used for majority voting. The accuracy obtained by combining a different number of atlases referenced with subscripts in Table 3 shows an increase in Dice overlap from 85.29% to 89.40%. However, the computation time also increases from 1.7 s to 89.9 s, with an increase in the number of atlases referenced for majority voting.

Having a trade-off between performance and computation time, Table 2 (last column) shows the majority voting strategy with 5 reference atlas obtaining a Dice score of 88.69%.

3.3. Comparison with the state-of-the-art

We now compare our joint reconstruction and parcellation method SegRecon against several baselines and recent approaches for these tasks. Table 4 reports the performance of tested methods in terms of average Dice scores over both hemispheres, mean Hausdorff distances computed on the surface manifold over both hemispheres, as well as runtimes. To evaluate the benefit of predicting cortical parcels using spherical atlas coordinates, we first train a 3D-UNet to predict the parcellation label probabilities directly at the voxel level as in standard 3D segmentation networks. This baseline, called DirectSeg in Table 4, gives a low Dice score of 79.95%. As mentioned above, this is due to the greater number of network outputs (i.e., one output per class) compared to simply predicting the two spherical atlas coordinates.

We also evaluate the FreeSurfer parcellation against the manual labels provided in the MindBoggle dataset. FreeSurfer considerably improves parcellation accuracy compared to DirectSeg with a Dice score of 84.39%. However, this comes at the price of a significant increase in computation times, from 300 ms per volume for DirectSeg to a few hours for FreeSurfer.
A maximum of 9.91%. However, these approaches require generating FS + GCN, FS + ASEGAT, and FS + SPHARM-Net, instead of individually. As a result, SPHARM-Net (Ha and Lyu, 2022b), in our training setting. As a result, a spectral graph convolutional network (GCN) (Gopinath et al., 2019) is evaluated. Finally, we exploit the surface parcellation next spectral method, the connectivity of nodes in the mesh graph is exploited in the prediction using a spectral graph convolutional network.

The labels of embedded nodes are then predicted in the FreeSurfer mesh graph using the main eigen-components of its Laplace matrix. The standard deviation is reported inside the parenthesis. Column 1-5: The average Dice overlap (in %) obtained after using five different references as an atlas for label propagation. The standard deviation is reported inside the parenthesis. Column 1–6: The average Dice overlap (in %) obtained after using majority voting with 1, 3, 5, 35, 70 different references as an atlas for label propagation. The standard deviation is reported inside the parenthesis.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Dice overlap (%)</th>
<th>Hausdorff (mm)</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>DirectSeg</td>
<td>79.95 ± 2.58</td>
<td>300 ms</td>
<td></td>
</tr>
<tr>
<td>FreeSurfer</td>
<td>84.39 ± 0.91</td>
<td>2.11 ± 0.29</td>
<td>4 h</td>
</tr>
<tr>
<td>FS + SRF</td>
<td>79.89 ± 2.62</td>
<td>1.97 ± 0.40</td>
<td>2 h + 18 s</td>
</tr>
<tr>
<td>FS + DBPN</td>
<td>84.60 ± 3.40</td>
<td>2 h + 1 s</td>
<td></td>
</tr>
<tr>
<td>FS + GCN</td>
<td>86.61 ± 2.45</td>
<td>1.66 ± 0.44</td>
<td>2 h + 3 s</td>
</tr>
<tr>
<td>FS + ASEGAT</td>
<td>89.00 ± 2.29</td>
<td>2 h + 1 s</td>
<td></td>
</tr>
<tr>
<td>FS + SPHARM-Net</td>
<td>88.48 ± 1.88</td>
<td>1.36 ± 0.27</td>
<td>2 h + 1 s</td>
</tr>
<tr>
<td>Recon+Seg</td>
<td>44.50 ± 0</td>
<td>4 s</td>
<td></td>
</tr>
<tr>
<td>w/o hemisphere</td>
<td>59.58 ± 12.20</td>
<td>3.94 ± 3.14</td>
<td>8 s</td>
</tr>
<tr>
<td>SemoRecon (Ours)</td>
<td>88.69 ± 1.84</td>
<td>1.20 ± 1.36</td>
<td>8 s</td>
</tr>
</tbody>
</table>

The first row shows the performance of a DirectSeg a 3D-CNN network on surface parcellation. The second row illustrates the results of the traditional FreeSurfer algorithm for parcellation. In the third and fourth row, we show the ability of a Spectral Random Forest (SRF), DBPN, graph convolutional network (GCN) and ASEGAT learning based approach to segment the cortical surface. The eighth row shows the importance of learning hemisphere segmentation in our work. Finally, in the last row, we show the performance of our proposed model. The Hausdorff distance is computed directly on the surface mesh.

Third, we show the advantage of predicting cortical surfaces directly from 3D images, as in our method, compared to working with surface meshes computed previously. Toward this goal, we compare our method with four mesh-based surface parcellation methods, named FS + SRF, FS + DBPN (Zhang and Wang, 2019), FS + GCN (Gopinath et al., 2018), and FS + ASEGAT (Li et al., 2022). The results reported for these methods are taken from their original manuscripts since their code is not publicly available. The first one, Spectral Random Forest (SRF), DBPN, graph convolutional network (GCN) and ASEGAT learning based method (Bazin and Pham, 2007), a mesh is generated for each surface voxel compared with the hemisphere prediction. In this way, our model predicts the iso-surface for surface reconstruction. The accurate prediction of polar and azimuth angles (ϕ̂, and ̂ ̂) for obtaining parcel labels from the atlas yields an average Dice score of 88.69%. Similar improvements of our method compared to other approaches are also found for the Hausdorff distance metric. Qualitative results obtained by our surface segmentation method are shown in Fig. 6, where we illustrate the differences between the predicted and manual label boundaries for four different parcels or regions.

### 4. Discussion and conclusion

We presented SemoRecon, a novel deep learning end-to-end model for the joint reconstruction and segmentation of nested surfaces, directly from MRI volumes. Our model learns multiple signed distance functions that represent surfaces implicitly as iso-levels. An inter-surface distance loss, computed from the distance maps during training, ensures that surfaces do not cross and that the predicted cortical thickness is anatomically possible. After applying a topological correction method (Bazin and Pham, 2007), a mesh is generated for each surface from their signed distance map using the (Lewiner et al., 2003) implementation of Marching Cubes algorithm (Lorensen and Cline, 1987). Jointly, the model also learns to predict the spherical coordinates of each voxel in a registered atlas space. The propagation of labels from the atlas space effectively segments the cortical white matter surface.

Our experiments used the largest publicly available dataset of manually-labeled brain surfaces (Klein et al., 2017), as well as the ABIDE-I (Di Martino et al., 2014) and OASIS (Marcus et al., 2007) datasets, to evaluate the surface reconstruction and segmentation accuracy of our method. We first showed the advantage of employing a signed distance map over a binary surface mask for reconstructing cortical surfaces. When comparing surfaces reconstructed by our method to those produced by FreeSurfer, using a continuous signed distance map significantly reduces the Hausdorff distance from 2.8 mm to 1 mm. Fig. 3 shows the irregularities and artifacts in the reconstructed surface.
due to the use of binary map. We then validated the robustness of our reconstruction method on the ABIDE and OASIS datasets which were not used in training. The method yields a Hausdorff distance of less than 1.5 mm on samples from these datasets, obtained with varying acquisition protocols and corresponding to subjects with very different ages and cortical alterations. The surfaces reconstructed with our method, presented in Fig. 3, 4, are visually similar to FreeSurfer meshes which require few hours of runtime to generate. Similarly, the curvature of the white surfaces, as seen in Fig. 5, highlights the quality of our surface reconstructions.

We analyzed the impact on performance of the reference atlas selected for transferring cortical parcellation labels to the surface. While Dice scores ranging from 84.60% to 87.33% were obtained with 5 different atlases, an improved Dice of 88.69% was achieved via a multi-atlas strategy combining the predictions for different atlases with majority voting. We also compared our method against several baselines and state-of-the-art approaches for cortical parcellation. Our approach has higher Dice score than directly predicting cortical labels with 3D-UNet (79.9%) which, unlike our method, cannot be used to reconstruct cortical surfaces. Moreover, it achieved a significantly higher mean Dice score than FreeSurfer (84.3%) with substantially reduced computation times over compared to this method (hours vs. seconds). Likewise, it improved by over 2% Dice a state-of-art parcellation method based on GCN that requires pre-computed surfaces as input.

While the potential of our method has been demonstrated in our results, our method has also limitations in both reconstruction and parcellation, identified next.

Reconstruction: The reconstruction of surfaces from explicit distance functions creates a dependency on the input resolution of the MRI. Our results have evaluated the robustness of our method on a 1 mm^3 MRI volume. Our approach could however potentially fail when a surface reconstruction is made on coarser MRI resolutions. Furthermore, our proposed surface reconstruction has only been evaluated on structural T1 MRI. While an extension to other modalities and resolutions should follow the same methodology with a re-training, fine-tuning and additional data augmentation, its validation remains to be performed. On an additional note, our method adds a penalty term on thickness during training to ensure non-overlapping white and pial surfaces implicitly, while also using an external topological correction technique. To this matter, despite the added dependency on these external techniques, our results indicate that the reconstructed cortical surfaces have no measured and observed topological errors. Assessing alternative topological correction methods remains a potential for future research. As an extension of our work, we further anticipate that adding constraints to the CNN network that guarantee a topological correctness could potentially improve the accuracy of the surface reconstruction. Likewise, the presence of abnormalities such as tumors and lesions in the brain could hinder the estimation of the surface distance functions and, hence, their reconstructions.

Segmentation/Parcellation: Our approach consists of performing a cortical surface parcellation by predicting the spherical coordinates in a registered atlas space. Predicting such spherical coordinates also enables the use of different atlases at inference time with no retraining. As shown in Table 2, the results demonstrate a sensitivity to the choice of the reference atlas picked for label propagation. To overcome this limitation, we have proposed to aggregate the labels from multiple reference atlases in a majority voting scheme. The computation time for our parcellation is, therefore, dependent on the number of references picked for a label look-up. Furthermore, our method also assumes a necessary registration between the reference atlases. Our majority voting could therefore potentially fail if these reference atlases are not adequately aligned to the same template space.

Future work: Our method has currently been evaluated on cerebral cortices, while it could also be in practice applied to various other surface data, such as cardio-vascular surfaces. Moreover, although our model includes a loss to control the distance between reconstructed surfaces and prevent them from crossing one another, incorporating more powerful topological constraints during training could possibly remove the need for our current use of external topological correction.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data we used for this manuscript is publically available.

Acknowledgments

This work was supported financially by the Canada Research Chair on Shape Analysis in Medical Imaging, the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Fonds de recherche du Québec – Nature et technologies (FRQNT), Canada. We also gratefully acknowledge the support of the NVIDIA Corporation with the donation of a GPU used for this research.

References


