Master's Thesis

SPHARM-Net: Spherical Harmonics-based Convolution for Cortical Parcellation

Seungbo Ha

Department of Computer Science and Engineering

Ulsan National Institute of Science and Technology

2022

SPHARM-Net: Spherical Harmonics-based Convolution for Cortical Parcellation

Seungbo Ha

Department of Computer Science and Engineering

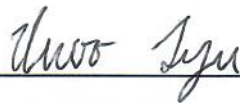Ulsan National Institute of Science and Technology

SPHARM-Net: Spherical Harmonics-based Convolution for Cortical Parcellation

A thesis/dissertation submitted to
Ulsan National Institute of Science and Technology
in partial fulfillment of the
requirements for the degree of
Master of Science

Seungbo Ha

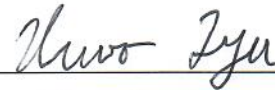06.02.2022 of submission

Approved by

_____

Advisor

Ilwoo Lyu

SPHARM-Net: Spherical Harmonics-based Convolution for Cortical Parcellation

Seungbo Ha

This certifies that the thesis/dissertation of Seungbo Ha is approved.

06.02.2022 of submission

_____

Advisor: Ilwoo Lyu

_____

Jimin Lee

_____

Jaejun Yoo

Abstract

We present a spherical harmonics-based convolutional neural network (CNN) for cortical parcellation, which we call SPHARM-Net. Recent advances in CNNs offer cortical parcellation on a fine-grained triangle mesh of the cortex. Yet, most CNNs designed for cortical parcellation employ spatial convolution that depends on extensive data augmentation and allows only predefined neighborhoods of specific spherical tessellation. On the other hand, a rotation-equivariant convolutional filter avoids data augmentation, and rotational equivariance can be achieved in spectral convolution independent of a neighborhood definition. Nevertheless, the limited resources of a modern machine enable only a finite set of spectral components that might lose geometric details. In this work, we propose (1) a constrained spherical convolutional filter that supports an infinite set of spectral components and (2) an end-to-end framework without data augmentation. The proposed filter encodes all the spectral components without the full expansion of spherical harmonics. We show that rotational equivariance drastically reduces the training time while achieving accurate cortical parcellation. Furthermore, the proposed convolution is fully composed of matrix transformations, which offers efficient and fast spectral processing. In the experiments, we validate SPHARM-Net on two public datasets with manual labels: Mindboggle-101 (N=101) and NAMIC (N=39). The experimental results show that the proposed method outperforms the state-of-the-art methods on both datasets even with fewer learnable parameters without rigid alignment and data augmentation. Our code is publicly available at https://github.com/Shape-Lab/SPHARM-Net.

Contents

List of Figures

I Introduction

Cortical parcellation is a key tool that supplies regions of interests (ROIs) in cortical morphological studies. Regarding anatomical/functional associations, the cortex can be subdivided into non-overlapping sub-regions by tracing regional boundaries defined by parcellation protocols [1–3]. Nevertheless, earlier cortical parcellation was a tedious and error-prone task as often handcrafted by experts, which requires tremendous human resources. This necessitates an automatic parcellation technique to reduce such extensive manual effort significantly. An initial attempt was to employ label fusion, in which the reference labels are aggregated on a target after surface registration [4]. This approach yet requires non-rigid warping [5–7] of all references to a target, and the label fusion needs to be carefully designed to incorporate anatomical variability across the references.

Recently, convolutional neural networks (CNNs) have shown remarkable success for their flexibility and scalability in semantic segmentation with the continuing evolution of computing power. In neuroimaging studies, CNNs become more popular for tissue segmentation of volumetric data [8–12] than the conventional approaches that require either handcrafted features or extensive non-rigid registration. In particular, an encoder-decoder architecture such as U-Net [13] can jointly learn both local and global geometric details without handcrafted feature definitions [8–12]. An important note is that most CNNs are generally designed for structured data (i.e., uniform grid) that simplify a semantic segmentation task. Yet, the extension to cortical parcellation is rather non-trivial because surface data form an irregular grid in general.

Later, CNNs have been evolved to support semantic segmentation on an unstructured grid (mesh) of surface data [14–19]. A graph representation is used to define convolutional filters on an unstructured grid using either node-wise message passing [14–17] or spectral decomposition [18, 19] regardless of graph regularity. As a cortical surface is already of a graph form, this approach can be straightforwardly extended to accomplish a cortical parcellation task in cortical sub-regions [19] or whole cortex [16]. These studies reported that the graph CNNs achieve higher parcellation accuracy than non-CNN-based cortical parcellation.

Another popular approach is spherical CNNs that can support a spherical form of data with rotational convolution. Spherical CNNs define their convolutional filters via either known spatial tessellation [20–24] or spherical harmonic decomposition [25,26]. In neuroimaging studies, a cortical surface is enforced to be genus-zero in modern cortical surface reconstruction pipelines such as FreeSurfer [6]. They offer spherical parametrization of the cortical surfaces, which may enable spherical CNNs to accomplish a cortical parcellation task. To date, nevertheless, only a few spherical CNNs are available for cortical parcellation yet [22, 27, 28] perhaps due to (1) fine data resolution and (2) lack of manual annotation, which is seldom discussed in most spherical CNNs.

Several spherical CNNs [20–24] define convolutional filters on a known spherical grid such as icosahedral [29] or HEALPix [30] tessellation. They extended convolutional filters of Euclidean CNNs, by which spherical convolution can be applied over spherical data in a similar way to conventional Euclidean CNNs. Yet, such spherical convolution is feasible under a certain neighborhood definition.

1
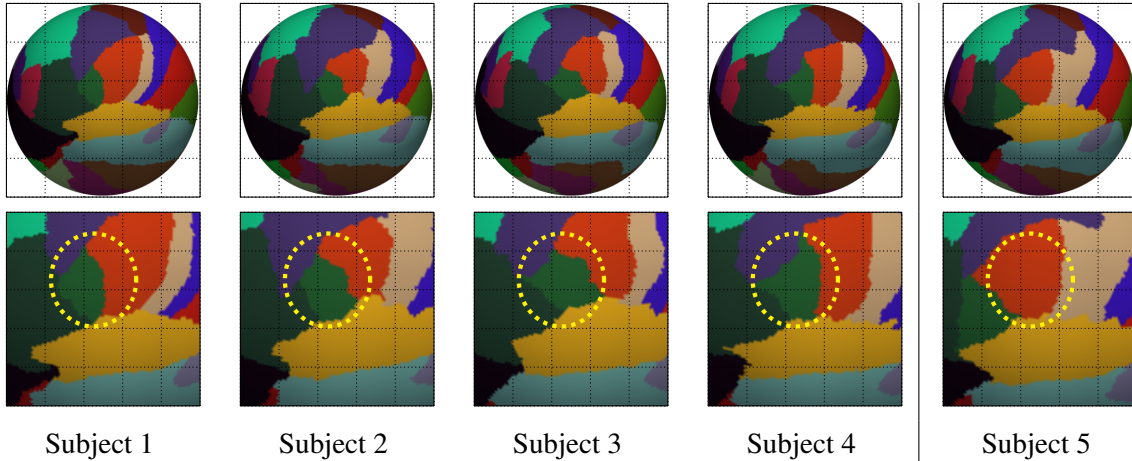
Figure 1: Anatomical variability. *Top*: overall structures on the unit sphere are well aligned after rigid alignment. *Bottom*: in *pars orbitalis* colored by *dark green* (*yellow* dotted ellipse), an instance to be inferred (Subject 5) can be spatially far off from the majority (Subjects 1-4) due to high anatomical variability, which may be difficult to annotate even after flawless rigid alignment unless rotational equivariance is considered.

Moreover, one of the main challenges in cortical parcellation is anatomical variability. Even with flawless rigid alignment, some ROIs having similar patterns can be spatially far off from the majority as shown in Fig. 1. Recent cortical parcellation approaches [22, 27, 28] attempted to relieve such variability by employing an extra preprocessing step such as random rotation. However, the preprocessing may not fully cover the entire configuration space of spherical rotation, and the increased sample size takes longer training time. This suggests that we should consider rotational equivariance to learn anatomical variability. Some spherical CNNs [23, 24] can support rotation-equivariant convolution with a specific neighborhood definition, which might be promising to incorporate anatomical variability.

Spherical harmonics can transform any spherical function (signal) into a series of spectral components akin to Fourier transform. The convolution theorem [31] states that spherical convolution is equivalent to point-wise multiplication of harmonic coefficients and can be independent of a neighborhood definition. Here, spherical convolution can achieve rotational equivariance by integrating over $SO(3)$ [25] or imposing a rotational symmetry constraint [26, 32], yielding zonal convolutional filters [26]. Although promising, the extension to semantic segmentation especially in neuroimaging studies is not popular yet perhaps due to limited machine capacity. On a modern machine, the expansion of spherical harmonics is limited up to a certain bandwidth, by which geometric details encoded in high-frequency components can be smoothed out in the existing spectral convolution.

In this work[1], we present a novel spherical harmonics-based spherical CNN (SPHARM-Net) for cortical parcellation. SPHARM-Net seeks the following key features: (1) rotational equivariance, (2) constrained full-bandwidth convolution, and (3) an end-to-end framework for semantic segmentation of

---

[1]The work is based on the previously published paper, Ha and Lyu [33], and partially adapts text descriptions and figures from the published paper.

cortical surface data. To this end, we follow the foundations of spectral convolution proposed by [26]. In particular, we adapt a standard U-Net architecture [13] for cortical parcellation that extends [26] originally designed for object recognition tasks. Furthermore, we propose a constrained spherical convolutional filter that supports an infinite set of spectral components. The resulting filter avoids the harmonic truncation in the original spectral convolution [26] that might lose geometric details. This can thus bridge a gap between theoretical consideration and its practical realization. Thanks to rotational equivariance, SPHARM-Net handles anatomical variability by extracting vertex-wise features without extensive data augmentation and rigid alignment required in conventional cortical parcellation. Moreover, the proposed convolution including harmonic decomposition is fully composed of matrix transformations, which is well suitable for parallel computing for efficient data processing. Fig. 2 illustrates a schematic overview of the proposed cortical parcellation.
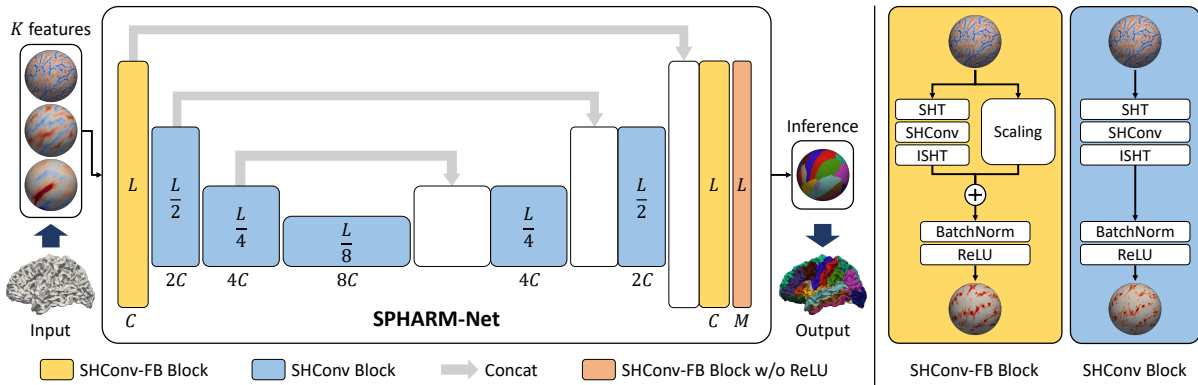
Figure 2: A schematic overview of the proposed cortical parcellation. *Left*: From an input cortical surface, K extracted geometric features are mapped onto the unit sphere to serve as the input channels of SPHARM-Net. In the encoding (decoding) phase, each block halves (doubles) the harmonic bandwidth L while doubling (halving) the number of channels C. The final block (*orange*) aggregates the learned information to infer M parcellation labels that are then mapped back to the original cortical surface. *Right*: the proposed blocks perform SHT on a spherical signal followed by rotation-equivariant spectral convolution and ISHT. The SHConv-FB block (*yellow*) further adds a scaled spherical signal to the SHConv block (*blue*) to support constrained full-bandwidth convolution (see the technical details in this work). In the encoding and decoding phases, every block is composed of batch normalization and nonlinear activation layers following the proposed convolution.

3

II Background

2.1 Spherical Harmonic Transform

We briefly review concepts of the spherical harmonic transform and its properties as a core of our methodology. More technical details are discussed in the review literature [31, 34, 35]. Despite various definitions of spherical harmonics, they can be defined in a complex domain as follows:

$$Y_l^m(\theta, \varphi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) e^{im\varphi}, \tag{1}$$

$$Y_l^{-m}(\theta, \varphi) = (-1)^m \bar{Y}_l^m(\theta, \varphi), \tag{2}$$

where P_l^m is the associated Legendre polynomial and $\bar{Y}_l^m$ is the conjugate of Y_l^m. In our problem setting, real harmonic basis functions are more intuitive to handle spherical signals from the cortex since the real-valued signals are often provided by cortical geometry (e.g., mean curvature or cortical thickness). Formally, we denote a real harmonic function of degree l and order m ($|m| \leq l$) by Y_{lm}. At a spherical location $(\theta, \varphi) \in [0, \pi] \times [-\pi, \pi]$, the standard irreducible real harmonic basis is written by

$$Y_{lm} = \begin{cases} \frac{i}{\sqrt{2}} \left(Y_l^m - (-1)^m Y_l^{-m} \right), & m < 0 \\ Y_l^0, & m = 0 \\ \frac{1}{\sqrt{2}} \left(Y_l^{-m} + (-1)^m Y_l^m \right), & m > 0. \end{cases} \tag{3}$$

Given a real-valued function $f : S^2 \to \mathbb{R}$, its spherical harmonic expansion is obtained by

$$f(\theta, \varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \hat{f}_{lm} Y_{lm}(\theta, \varphi), \tag{4}$$

where $\hat{f}_{lm}$ is the associated coefficient of Y_{lm} as follows:

$$\hat{f}_{lm} = \int_{S^2} f(\theta, \varphi) Y_{lm}(\theta, \varphi) \, d\theta \, d\varphi. \tag{5}$$

We refer Eqs. (5) and (4) to the spherical harmonic transform (SHT) of f and its inverse spherical harmonic transform (ISHT), respectively. A key property of SHT is that the spectral energy distribution of f is independent of a rotation of f. SHT can thus offer a rotation-invariant representation of f in the spectral domain. More formally, consider the l_2 norm of coefficients calculated over the orders at a given degree l:

$$\|\hat{f}_l\| = \sqrt{\sum_{m=-l}^{l} (\hat{f}_{lm})^2}. \tag{6}$$

The norm remains unchanged regardless of the orientation of f [34]. Hence, a collection of the l_2 norm across harmonic degrees encodes a rotation-invariant descriptor of f:

$$\left(\|\hat{f}_0\|, \|\hat{f}_1\|, \cdots, \|\hat{f}_l\|, \cdots \right). \tag{7}$$

Interested readers are referred to [34] for in-depth proof.

2.2 Rotation-Equivariant Convolution

Consider two spherical functions $f, h : S^2 \to \mathbb{R}$, and we let h be a convolutional filter (kernel). Their convolution is then defined as an integral over the 3D rotation group $SO(3)$:

$$(f * h)(x) = \int_{g \in \mathrm{SO}(3)} f(g\eta)h(g^{-1}x)\,\mathrm{d}g, \tag{8}$$

where η is the north pole and $\|x\|^2 = 1$. Unfortunately, the solution is not of a closed form in general. This hence necessitates some special manipulation to make spherical convolution feasible: (1) spatial sampling along a spherical grid or (2) spectral transformation. A popular approach is to re-tessellate the unit sphere along a semi-uniform grid via well-known tessellation techniques such as icosahedral subdivision [29], in which the integral can be approximated as a sum over the entire spherical grid. Also, 1-ring or 2-ring neighborhoods are usually considered to simplify spherical convolution for computational efficiency. Despite such a simple operation, the solution heavily depends on the quality of spatial sampling (graph regularity, edge length, etc.) due to the absence of uniform spacing on a spherical grid. Moreover, the approximated integral is incapable of looping through the entire $SO(3)$ because any spherical grid is finite, on which the resulting spherical convolution (i.e., numerical accuracy) is always sensitive to a choice of tessellation techniques. Furthermore, these reasons make rotation-invariant convolution non-trivial to be maintained in the grid-based approach.

Alternatively, spherical convolution is performed by point-wise multiplication of harmonic coefficients followed by ISHT that converts the harmonics back into the spherical domain [25, 26]. In particular, our interest is to define spherical convolution that holds rotational equivariance. In this context, spherical convolution can be achieved by transforming either a function of $SO(3)$ [25] or a spherical function [26] via SHT. The former offers a flexible filter by lifting spherical functions onto $SO(3)$ to enable rotation-equivariant convolution. On the other hand, the entire processing in the latter lies in the spherical domain without the costly lifting despite its relatively strict constraints on a convolutional filter. In this work, we employ the latter approach for computational efficiency and more intuitive interpretation.

More formally, we denote the SHT of f and h at degree l and order m by $\hat{f}_{lm}$ and $\hat{h}_{lm}$, respectively. As stated in [31], there exists a convolutional filter h supporting a rotation-equivariant spherical convolution if h is indirectional and axis-symmetric (also known as a spherical zonal/isotropic function), i.e., $\hat{h}_{lm} = 0$ ($m \neq 0$). This consequently reduces the spherical harmonics to the ordinary Legendre polynomials. To this end, we have the following operation for asymmetry (or left-) convolution in the spectral domain [31]:

$$(\widehat{f * h})_{lm} = C(l)\hat{f}_{lm}\hat{h}_{l0}, \tag{9}$$

where $C(l) = 2\pi\sqrt{4\pi/(2l+1)}$. To verify that the resulting convolution holds rotational equivariance,

we use the properties of the rotation-invariant descriptor in Eq. (7).

$$
\begin{aligned}
\|(\widehat{f * h})_l\| &= C(l)\|\hat{f}_{l,-l}\hat{h}_{l0} + \cdots + \hat{f}_{ll}\hat{h}_{l0}\| \\
&= C(l)\|\hat{h}_{l0} \cdot (\hat{f}_{l,-l} + \cdots + \hat{f}_{ll})\| \\
&= C(l)|\hat{h}_{l0}|\|\hat{f}_{l,-l} + \cdots + \hat{f}_{ll}\| \\
&= C(l)|\hat{h}_{l0}|\|\hat{f}_l\|.
\end{aligned}
\tag{10}
$$

This suggests that the resulting convolution is rotation-equivariant regardless of the orientation of f for a given zonal function h. Finally, the spatial convolution can be recovered by a linear combination of spherical harmonic basis functions with their associated harmonic coefficients (see Eq. (4)).

$$
(f * h)(\theta, \varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (\widehat{f * h})_{lm} Y_{lm}(\theta, \varphi)
\tag{11}
$$

$$
= \sum_{l=0}^{\infty} \sum_{m=-l}^{l} C(l)\hat{f}_{lm}\hat{h}_{l0} Y_{lm}(\theta, \varphi).
\tag{12}
$$

On a modern machine, the harmonic expansion is band-limited yet because of limited machine capacity. This hinders the realization of such well-defined harmonic behaviors by the harmonic truncation, yielding overly smoothed effects. We will address this challenge under a certain condition and discuss detailed discretization techniques in the next section.

III Methods

3.1 Problem Statement

Our goal is to subdivide a cortical surface into M sub-regions. Suppose that a cortical surface consists of K distinct geometric features such as mean curvature. By valid spherical parametrization, the i-th geometric feature can be mapped onto the unit sphere as a spherical function $f_i : S^2 \to \mathbb{R}$, $i = 1, \cdots, K$. Here, cortical parcellation can be achieved by estimating discriminant functions given the cortical geometry. More formally, we denote a collection of spherical locations by $\mathscr{V} = \{v \in S^2 : v = (\theta, \varphi) \in [0, \pi] \times [-\pi, \pi]\}$. At a spherical location v, we infer its label $z \in \{1, 2, \cdots, M\}$. Consider a discriminant that v belongs to z. In cortical parcellation, we have M discriminant functions $\mathscr{F} : \mathbb{R}^K \to \mathbb{R}$:

$$\mathscr{F}(f_1(v), \cdots, f_K(v); z). \tag{13}$$

Once $\mathscr{F}$ is optimized, the highest discriminant is calculated to assign cortical labels to unseen instances. In supervised learning, $\mathscr{F}$ is a model that can be learned from observations of f and their labels. In the remainder of this work, we will describe the SPHARM-Net architecture to model $\mathscr{F}$.

3.2 Constrained Convolution with Full Bandwidth

As stated, expanding an infinite set of spherical harmonics is practically infeasible because of limited machine capacity. Existing approaches thus employ the band-limited spherical harmonic expansion up to a certain bandwidth L [26, 35]. Despite their computational benefits, such harmonic truncation discards geometric details; overly smoothed signals might discourage local feature detection. To support the full harmonic expansion, we propose a convolutional (zonal) filter, where the coefficients of the high-frequency components are constrained. Formally, we decompose Eq. (12) into two terms as follows:

$$(f * h)(\theta, \varphi) = \sum_{l=0}^{L} \sum_{m=-l}^{l} C(l) \hat{f}_{lm} \hat{h}_{l0} Y_{lm}(\theta, \varphi) + \sum_{l=L+1}^{\infty} \sum_{m=-l}^{l} C(l) \hat{f}_{lm} \hat{h}_{l0} Y_{lm}(\theta, \varphi), \tag{14}$$

where $\{\hat{h}_{l0}\}_{l=0}^{\infty}$ are learnable parameters. As learning an infinite set of $\hat{h}_{l0}$ is infeasible, the second term is discarded in the conventional band-limited convolution. Our goal is to design a specialized convolutional filter h that can maintain high-frequency components. Here, we constrain $\hat{h}_{l0} = \alpha/C(l)$, $\forall l > L$ for a single learnable parameter $\alpha \in \mathbb{R}$. This yields

$$(f * h)(\theta, \varphi) = \sum_{l=0}^{L} \sum_{m=-l}^{l} C(l) \hat{f}_{lm} \hat{h}_{l0} Y_{lm}(\theta, \varphi) + \sum_{l=L+1}^{\infty} \sum_{m=-l}^{l} C(l) \hat{f}_{lm} \frac{\alpha}{C(l)} Y_{lm}(\theta, \varphi). \tag{15}$$

By arithmetic manipulation, we rewrite Eq. (15) as follows:

$$
\begin{aligned}
&(f * h)(\theta, \varphi) \\
&= \sum_{l=0}^{L} \sum_{m=-l}^{l} [C(l) \hat{f}_{lm} \hat{h}_{l0} Y_{lm}(\theta, \varphi) - C(l) \hat{f}_{lm} \frac{\alpha}{C(l)} Y_{lm}(\theta, \varphi)] + \sum_{l=0}^{\infty} \sum_{m=-l}^{l} C(l) \hat{f}_{lm} \frac{\alpha}{C(l)} Y_{lm}(\theta, \varphi)
\end{aligned}
\tag{16}
$$

$$= \sum_{l=0}^{L} \sum_{m=-l}^{l} C(l) \hat{f}_{lm} \left(\hat{h}_{l0} - \frac{\alpha}{C(l)} \right) Y_{lm}(\theta, \varphi) + \sum_{l=0}^{\infty} \sum_{m=-l}^{l} C(l) \hat{f}_{lm} \frac{\alpha}{C(l)} Y_{lm}(\theta, \varphi). \tag{17}$$

7

This indicates that the proposed convolutional filter can be decomposed into band-limited and full-bandwidth components. To investigate the proposed filter, we consider its spatial form:

$$h(\theta, \varphi) = \sum_{l=0}^{L} \left(\hat{h}_{l0} - \frac{\alpha}{C(l)} \right) Y_{l0}(\theta, \varphi) + \sum_{l=0}^{\infty} \frac{\alpha}{C(l)} Y_{l0}(\theta, \varphi). \tag{18}$$

The first term is a band-limited zonal filter. We will show that the second term h_{FB} has a closed form. Since h_{FB} is a zonal filter and Y_{l0} is involved in the ordinary Legendre polynomial, we have

$$
\begin{aligned}
h_{\mathrm{FB}}(\theta, \varphi) &= \sum_{l=0}^{\infty} \frac{\alpha}{C(l)} Y_{l0}(\theta, \varphi) \\
&= \sum_{l=0}^{\infty} \frac{\alpha}{C(l)} \cdot \sqrt{\frac{2l+1}{4\pi}} P_l^0(\cos\theta) \\
&= \sum_{l=0}^{\infty} \alpha \frac{2l+1}{8\pi^2} P_l^0(\cos\theta) \\
&= \frac{\alpha}{4\pi^2} \sum_{l=0}^{\infty} \frac{2l+1}{2} P_l^0(\cos\theta).
\end{aligned}
\tag{19}
$$

Since the Dirac delta function δ has the following series of the Legendre polynomials for $x, y \in [-1, 1]$:

$$\delta(x-y) = \sum_{l=0}^{\infty} \frac{2l+1}{2} P_l^0(x) P_l^0(y), \tag{20}$$

and $P_l^0(1) = 1$ for all l, we can simplify Eq. (19):

$$h_{\mathrm{FB}}(\theta, \varphi) = \frac{\alpha}{4\pi^2} \delta(\cos\theta - 1). \tag{21}$$

We now have a closed form of h_{FB} as an impulse response at the north pole ($\theta = 0$) and zero, otherwise. Hence, the spatial form of h is composed of a band-limited zonal filter and an impulse response at the north pole, which can maintain the high-frequency components of f as shown in Fig. 3. From Eq. (4), we can show that the second term of Eq. (17) becomes a scaled ISHT of $\hat{f}$. By letting $\hat{h}'_{l0} = \hat{h}_{l0} - \alpha/C(l)$, we further simplify Eq. (17):

$$(f * h)(\theta, \varphi) = \sum_{l=0}^{L} \sum_{m=-l}^{l} C(l) \hat{f}_{lm} \hat{h}'_{l0} Y_{lm}(\theta, \varphi) + \alpha f(\theta, \varphi), \tag{22}$$

where $\{\hat{h}'_{l0}\}_{l=0}^{L}$ and α are learnable parameters. Consequently, the proposed full-bandwidth convolution is feasible without the infinite harmonic expansion of both f and h. We can also verify that rotational equivariance holds in the convolution by Eq. (10). It is noteworthy that the conventional band-limited convolution is a special case of the proposed approach, where no impulse response is learnable (i.e., $\alpha = 0$).

One may observe that the formula of the proposed convolution is similar to that of the residual networks proposed in [36]. However, we emphasize that our approach stems from a different perspective of how high-frequency components are efficiently maintained associated with geometric details in spectral convolution. This has not been studied well in existing spherical harmonics-based CNNs. Moreover, the proposed full-bandwidth convolution can be further simplified by the spectral pooling operation [26] in our framework unlike [36], which will be described in the next section.

8
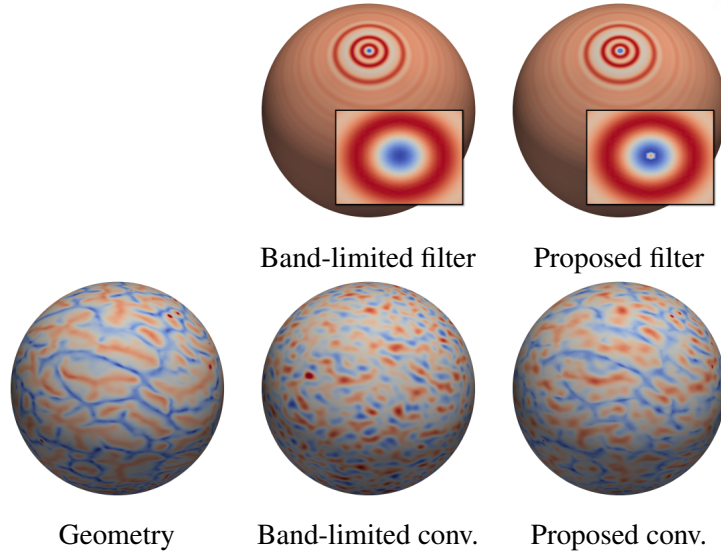
Figure 3: An example learned filter. *Top*: unlike a conventional filter (*left*), the proposed filter (*right*) has a learnable impulse response at the north pole. *Bottom*: after spectral convolution of cortical geometry (*left*) with the filters above, the band-limited convolution (*middle*) loses geometric details due to the truncation of the high-frequency components. On the other hand, the proposed convolution (*right*) maintains geometric details without the infinite expansion of spherical harmonics.

3.3 Spectral Pooling

In CNNs, a pooling operation is widely used on a uniform grid, which resamples input data by extracting a maximal or average activation within a sub-region of the input data. In spherical CNNs, the pooling operation on a uniform grid can be easily generalized over a spherical grid, on which the neighborhoods are used to define a spherical sub-region for local data aggregation [20–22]. As expected, their performance relies on the quality of ruling a spherical grid; unfortunately, uniform spatial tessellation of the unit sphere is topologically implausible, which can likely over-/under-estimate local activation. Moreover, the spatial pooling typically tends to suffer from stairstep artifacts as the reduction of spatial resolution is enforced [37]. This motivates a spectral pooling technique [26, 37] that limits the harmonic expansion to address the tessellation issue posed in the spatial pooling. The spectral pooling naturally allows low-pass filtering of the input data to smoothly adjust spatial changes while maintaining the original spatial resolution.

In particular, the spectral pooling proposed in [26] has the following advantages over the conventional spatial pooling: fast operation in the spectral domain, independence of spherical tessellation, and rotational equivariance. To employ the spectral pooling, we simply limit the bandwidth of the proposed convolution by $L' \leq L$ without adjusting the spatial resolution. Formally, the spectral pooling is applied to the convoluted signals from Eq. (22) as follows:

$$\sum_{l=0}^{L'} \sum_{m=-l}^{l} C(l) \hat{f}_{lm} \hat{h}_{l0} Y_{lm}(\theta, \varphi). \tag{23}$$

Hence, geometric details can be easily controlled by tuning L'. Specifically, we decrease L' as the

encoding level increases akin to [26]. Similarly, we refine geometric details by increasing L' as the decoding level increases.

3.4 Numerical Implementation

Spherical Harmonic Basis

In our approach, spherical harmonics need to be expanded to encode spherical signals sufficiently. Although the analytic expansion is readily available, numerical stability should be considered due to machine precision in practice. Consider the following Schmidt semi-normalized associated polynomial

$$\bar{P}_l^m(\cos\theta) = \sqrt{2 - \delta_{m0}}\sqrt{\frac{(l - |m|)!}{(l + |m|)!}} P_l^{|m|}(\cos\theta). \tag{24}$$

where δ_{m0} is the Kronecker delta function. By plugging Eq. (24) into Eq. (3), we can rewrite a real harmonic basis without introducing complex terms.

$$Y_{lm}(\theta, \varphi) = \begin{cases} (-1)^m \sqrt{\frac{2l+1}{4\pi}} \bar{P}_l^m(\cos\theta)\sin(-m\varphi), & m < 0 \\ \sqrt{\frac{2l+1}{4\pi}} \bar{P}_l^0(\cos\theta), & m = 0 \\ (-1)^m \sqrt{\frac{2l+1}{4\pi}} \bar{P}_l^m(\cos\theta)\cos(m\varphi), & m > 0. \end{cases} \tag{25}$$

From a numerical perspective, evaluating Eq. (25) may suffer from numerical instability as l becomes larger due to the factorial expansion in Eq. (24) [38]. However, Eq. (24) can be efficiently computed by a standard three-term recursion without the explicit factorial expansion [38, 39]. The approach is known to guarantee numerical stability up to a harmonic degree of 2,700 on a modern machine [38]. As the harmonic expansion at each spherical location is computationally independent, the expansion over the entire sphere can be further accelerated by parallel processing for computational efficiency. Since we assume that spherical data share common tessellation in this work, the harmonic bases need to be computed once.

SHT and ISHT

In our approach, spatial signals need to be recovered via ISHT to enable nonlinear activation of the networks. Then, SHT converts the spatial activation back into the spectral domain, which requires proper tessellation of the unit sphere. Conventionally, a linear least squares is widely used to estimate spherical harmonic coefficients in SHT by collecting spherical signals over a discrete grid [35]. However, this approach is computationally expensive due to the pseudo-inverse of discretized harmonic bases. For computational efficiency, we rather utilize icosahedral subdivision [29] to employ a nearly uniform grid on the unit sphere. Motivated by [40], we numerically approximate the integral of Eq. (5) by a finite sum over the icosahedral tessellation without solving the least squares. This consequently simplifies both SHT and ISHT into a form of matrix multiplication. More formally, we denote the sampling points of the icosahedral tessellation by $v_i \in S^2$, $i = 1, \cdots, N$. Given a proper resolution of the tessellation, we

have the following approximation of SHT from Eq. (5):

$$\hat{f}_{lm} \approx \sum_{i=1}^{N} f(v_i) Y_{lm}(v_i) \Delta(v_i), \tag{26}$$

where $\Delta(v_i)$ is the area at v_i calculated across the adjacent triangles of v_i. As reported in [40], the approximation error becomes negligible for $N \geq 40,962$. As the triangles of the icosahedral tessellation are nearly regular, the area at v_i can be estimated by

$$\Delta(v_i) = \frac{1}{3} \sum_{j=1}^{k_i} T_j(v_i), \tag{27}$$

where k_i is the number of the adjacent triangles of v_i and $T_j(v_i)$ is the associated area [40]. From Eqs. (26) and (27), SHT can be thus rewritten as a form of matrix multiplication. We denote the respective collections of spherical signals and harmonic coefficients by

$$f = [f(v_1), \cdots, f(v_N)]^\top \text{ and } \hat{f}_l = \left[\hat{f}_{l,-l}, \cdots, \hat{f}_{ll}\right]^\top, \tag{28}$$

and the discretized harmonic basis at degree of l by

$$y_{lm} = [Y_{lm}(v_1), \cdots, Y_{lm}(v_N)]^\top \text{ and } \mathbf{Y}_l = \left[y_{l,-l}, \cdots, y_{ll}\right]. \tag{29}$$

Given a bandwidth L, we have

$$\mathbf{Y} = [\mathbf{Y}_0, \cdots, \mathbf{Y}_L] \text{ and } \hat{f} = \begin{bmatrix} \hat{f}_0 \\ \vdots \\ \hat{f}_L \end{bmatrix}. \tag{30}$$

Hence, the harmonic coefficients of y_{lm} can be obtained by

$$\hat{f}_{lm} = y_{lm}^\top \text{diag}(a) f, \tag{31}$$

where $a = [\Delta(v_1), \cdots, \Delta(v_N)]^\top$. This yields the following form of matrix multiplication:

$$\hat{f} = \mathbf{Y}^\top \text{diag}(a) f. \tag{32}$$

For ISHT, we can easily show

$$f = \mathbf{Y}\hat{f}. \tag{33}$$

This suggests that both SHT and ISHT can be efficiently solved by parallel processing while reducing the order of computation compared to the least squares approach.

3.5 SPHARM-Net

We adapt a standard U-Net architecture [13] that consists of encoder (contracting networks) and symmetric decoder (expanding networks) layers. Specifically, SPHARM-Net is composed of successive convolutional blocks that include multiple components in order: SHT (Eq. (32)), a convolutional layer

11

(either Eqs. (22) or (23)), ISHT (Eq. (33)), a batch normalization layer [41], and a rectified linear unit (ReLU) [42]. Like [13], SPHARM-Net contracts the convolutional networks in the encoding phase while analogously expanding the convolutional networks in the decoding phase (see Fig. 2). As stated, we can control the contraction/expansion of the convolutional networks by harmonic bandwidth L' rather than spatial resolution. We set $L' = L/2$ for the entry/final pooling operation. We then halve (double) L' and double (halve) channels in the encoding (decoding) phase like a standard U-Net [13]. As described already, our convolutional layers support two types of operations: spectral convolution with full bandwidth (SHConv-FB, Eq. (22)) and with the spectral pooling (SHConv, Eq. (23)). The proposed convolutional block uses SHConv except for the entry and final layers that maintain fine geometric details. This follows the original concept of fine feature representations in a standard U-Net. Furthermore, we parameterize the spectrum by every i-th degree (we empirically set $i = 5$) and linearly interpolate the harmonic coefficients to encourage spatial localization of the convolutional filter and to achieve computational efficiency [26, 43]. Once again, the proposed convolution including SHT and ISHT is fully composed of matrix transformations, which offers fast learning and inference.

IV   Results

Table 1: ROIs defined by the Desikan-Killiany Protocol. *: unavailable in Mindboggle-101. *: merged in NAMIC.

*0. Unknown	*1. Bank of Superior Temporal	2. Caudal Anterior Cingulate	3. Caudal Middle Frontal
*4. Corpus Callosum	5. Cuneus	6. Entorhinal	7. Fusiform
8. Inferior Parietal	9. Inferior Temporal	10. Isthmus Cingulate	11. Lateral Occipital
12. Lateral Orbitofrontal	13. Lingual	14. Medial Orbitofrontal	15. Middle Temporal
16. Parahippocampal	17. Paracentral	18. Pars Opercularis	19. Pars Orbitalis
20. Pars Triangularis	21. Pericalcarine	22. Postcentral	23. Posterior Cingulate
24. Precentral	25. Precuneus	26. Rostral Anterior Cingulate	27. Rostral Middle Frontal
28. Superior Frontal	29. Superior Parietal	30. Superior Temporal	31. Supramarginal
*32. Frontal Pole	*33. Temporal Pole	34. Transverse Temporal	*35. Insula
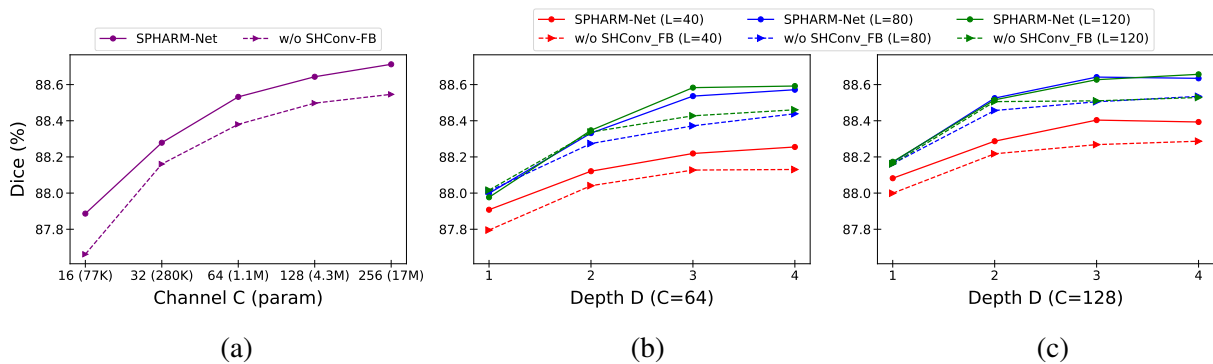
Figure 4: Ablation study. Starting from a base configuration ($C = 128$, $L = 80$, and $D = 3$), the ablation study is performed by controlling each of them while fixing others. (a) at $L = 80$ and $D = 3$, a constant performance gap is observed across C between with and without SHConv-FB. (b) at $C = 64$, a performance gap is relatively constant across D at $L = 40$, while it becomes significantly wider from $D = 3$ at $L = 80$ and 120. (c) similar trends to (b) are observed at $C = 128$. The performance becomes plateau both with and without SHConv-FB from $D = 3$. (b-c) it is noteworthy that SHConv-FB at $C = 64$ outperforms the ablated SPHARM-Net at $C \geq 64$.

4.1   Experimental Setup

Imaging Data

We validated SPHARM-Net on two public datasets: Mindboggle-101 (101 participants) [3] and NAMIC (39 participants) [1]. In both datasets, the cortical surfaces were reconstructed with their invertible spherical mappings by a standard FreeSurfer pipeline [6] and then manually annotated with 35 subregions ($M = 35$) via the Desikan-Killiany (DK) protocol [1] (see Table 1). In [3], they released the Mindboggle-101 dataset with the revision of the DK protocol to avoid spatial ambiguity: the *insula* was additionally defined from the unknown region (0 in Table 1), while 4 regions (1, 4, 32, and 33 in Table 1)

13

were merged into their neighboring ROIs. This results in a total of 32 regions ($M = 32$). Following [27, 44], we used three geometric features ($K = 3$) to represent multi-resolution of cortical geometry: mean curvature of both inflated and smoothed gray/white surfaces as well as average convexity [6]. All the individual spheres were re-tessellated via icosahedral subdivision of 40,962 vertices [29]. Finally, we only used the left hemisphere for training due to similar anatomical patterns observed in the right hemisphere.

Training and Evaluation

To train our model, we used the Dice loss proposed in [8] and the Adam optimizer [45] at an initial learning rate of 0.01 with decay by a factor of 0.1 if no improvement is made in two consecutive epochs. We employed 5-fold cross-validation, in which the original datasets are split into 5 partitions with 60% (3 partitions) for training, 20% (1 partition) each for validation and test. For each fold, we then circulated the partitions to achieve full validation. In the test phase, we used the learned model parameters at the peak accuracy in the validation set. The vertex-wise discriminant in the test set was then transferred to the original surface tessellation to determine final labels [44]. Furthermore, we shuffled the datasets prior to the cross-validation by setting 5 different random seeds to reduce a potential data bias introduced in the training phase (i.e., a total of five 5-fold cross-validation procedures). To evaluate performance, we computed the mean Dice score for each ROI across the random seeds. We used the gray/white surfaces to compute Dice overlap as they are fed to the spherical mapping [6], where sphere-based parcellation can be evaluated with minimal area distortion. All the experiments were conducted on an NVIDIA GeForce RTX 3090 with 24 GB memory.

4.2 Baseline Methods

We compared SPHARM-Net to the state-of-the-art cortical parcellation methods: FreeSurfer [4] and spherical CNNs [21, 22, 27]. For fair comparisons, we aligned the sphere of each cortex using its cortical geometry to reduce potential misalignment errors that might affect the performance of the baseline methods. It is noteworthy that our approach is independent of initial alignments because of rotational invariance, which we report in the experimental results. We also followed the same training configurations of the baseline methods as their original work [22, 27], which are optimally tweaked for a cortical parcellation task (data augmentation strategy, hyperparameter settings, etc.). For [4], we co-registered the individual spheres for each dataset in an atlas building scheme [5] using [6] as surface registration is required in [4]. We applied the same evaluation criteria as used in SPHARM-Net to the baseline methods. We evaluated the performance of SPHARM-Net over the baseline methods in terms of Dice overlap (whole cortex and individual ROIs) by conducting paired *t*-tests.

4.3 Ablation Study

We evaluated the impact of SHConv-FB on the Mindboggle-101 dataset. To this end, we replaced SHConv-FB with SHConv by fixing $\alpha = 0$. For three tunable hyperparameters in SPHARM-Net (chan-
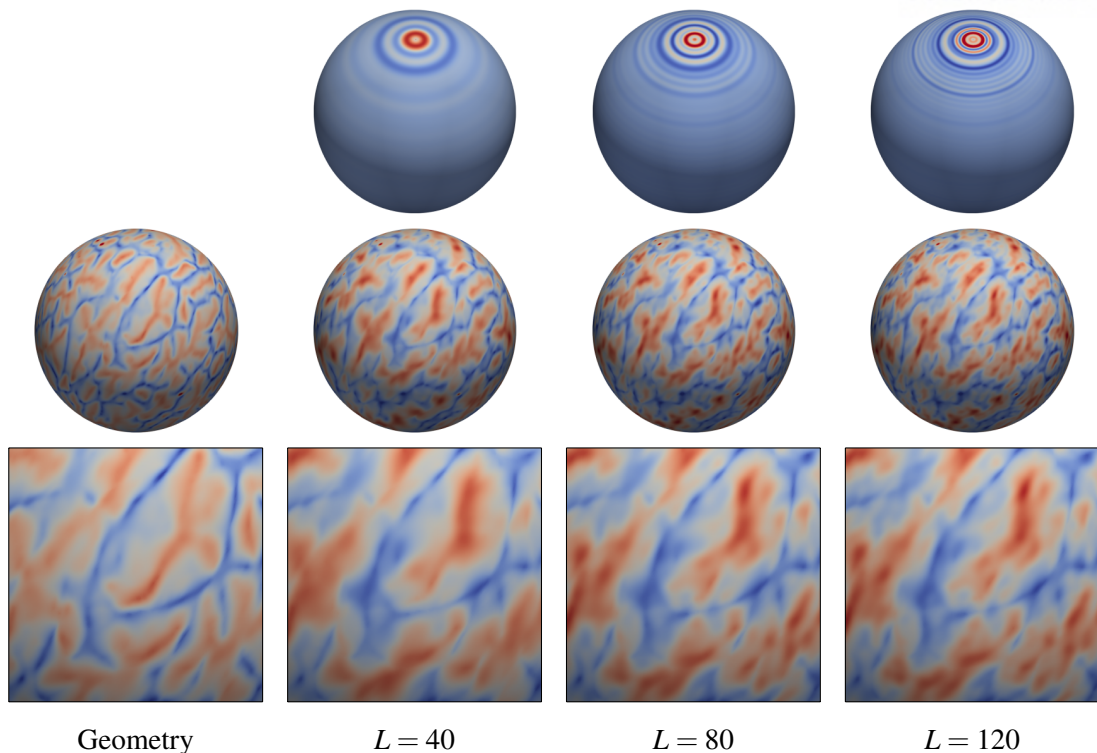
Figure 5: Learned feature maps with SHConv-FB. *Top*: learned filters at different bandwidth *L*. More global patterns are observed at lower bandwidth while local details are sharpened as *L* increases. *Middle*: filtered signals. Overall patterns become closer to the original geometry at lower bandwidth while more local responses are added as *L* increases. *Bottom*: zoomed-in signals. No huge visual difference between $L = 80$ and 120 may indicate similar parcellation performance (see Fig. 4(b-c)).

nel size *C*, encoder-decoder depth *D*, and harmonic bandwidth *L*), we varied each of them by fixing others in this work. To determine the reference degree *L*, we followed the harmonic reconstruction error in [35]. As the encoding block halves *L*, we chose $L = 80$ for divisibility by 2 up to $D = 4$ and for the spectral parametrization at every 5th degree. We also probed the importance of the spectral pooling.

Channel Size

Due to the GPU memory limit, we validated SPHARM-Net up to a channel size of 256 at $D = 3$ and $L = 80$. Overall, we observed the improved performance as *C* increases in SPHARM-Net both with and without SHConv-FB as shown in Fig. 4(a). This is likely because the increased channel size offers more flexibility in the spherical convolution. Nevertheless, a constant performance gap is observed across all the channel sizes, in which we found statistically significant difference for the respective channel sizes ($p < 1.0 \times 10^{-4}$). It is noteworthy that SHConv-FB at $C = 64$ outperforms the ablated SPHARM-Net at $C \geq 64$.

Depth and Bandwidth

We further validated SPHARM-Net by controlling D and L at $C = 64$ (Fig. 4(b)) and 128 (Fig. 4(c)), respectively. In both scenarios, we observed similar trends at $L = 40$, where a performance gap is relatively constant across D ($p < 0.05$). At $L = 80$ and 120, a performance gap is marginal between with and without SHConv-FB up to $D = 2$ while it becomes significantly wider from $D = 3$ ($p < 5.0 \times 10^{-5}$ at $C = 64$ and $p < 1.0 \times 10^{-4}$ at $C = 128$). At $C = 128$, the performance then becomes plateau both with and without SHConv-FB from $D = 3$ ($p > 0.12$). Like the ablation study of C, it can be observed that SHConv-FB at $D = 3$ outperforms the ablated SPHARM-Net for all the depths at $L = 80$ and 120. This reduces the computing burden while improving parcellation performance. Therefore, we could conclude from this experiment that SHConv-FB can consistently offer more accurate inference by considering high-frequency components than the ablated SPHARM-Net in terms of Dice overlap under the controlled configurations of the three hyperparameters. Throughout the ablation study, we suggest $C = 128$, $D = 3$, and $L = 80$, which can fit a GPU device with 8 GB memory, and we used this configuration throughout the remaining experiments. Fig. 5 shows an example of the learned feature maps after the proposed convolution varying L with SHConv-FB.

Spectral Pooling

We investigated the impact of the spectral pooling on SPHARM-Net. Since the original pooling [26] was proposed for object recognition tasks, its impact on cortical parcellation needs to be justified. For each block of SPHARM-Net, we replaced SHConv with SHConv-FB, and L was fixed at 80. Due to the GPU memory limit, we used $C = 64$ in this evaluation. SPHARM-Net without the spectral pooling achieves $88.51 \pm 1.87\%$ in Dice overlap on the Mindboggle-101 dataset, which is comparable to that with the spectral pooling ($88.49 \pm 1.83\%$, $p = 0.32$). Nevertheless, it is noteworthy that 3.26 times more learnable parameters (3.58M) are required without the spectral pooling. This suggests that the spectral pooling can efficiently capture global geometric features while reducing the number of learnable parameters.

4.4 Comparisons to Baseline Methods

We compared SPHARM-Net to the baseline methods and investigated the importance of rotational equivariance in spherical convolution. In addition to the suggested configuration ($C = 128$, $L = 80$, and $D = 3$, SPHARM-Net (128)), we further adjusted the hyperparameters to match the number of learnable parameters to that of the baseline methods for fair comparisons (see Tables 2 and 3). We then reported Dice scores for all the methods and their p values against SPHARM-Net (128) after paired t-tests. As SPHARM-Net is independent of rigid alignment, we reported the performance before and after rigid alignment, and no augmented samples were used.

Table 2: Mean Dice score on Mindboggle-101. *Param*: # of learnable parameters. *Rot*: rigid alignment. *Aug*: data augmentation.

Method	Param	Dice (%)	p value	Rot	Aug
SPHARM-Net (128)	4.31M	88.64 ± 1.83	-	Yes	No
SPHARM-Net (64)	1.10M	88.49 ± 1.83	2.3×10^{-11}	Yes	No
SPHARM-Net (16)	77K	87.86 ± 1.84	2.3×10^{-40}	Yes	No
FreeSurfer [4]	-	87.41 ± 1.82	5.6×10^{-21}	Yes	No
SU-Net [22]	1.67M	87.04 ± 1.90	4.2×10^{-31}	Yes	No
SU-Net+Aug [22]	1.67M	88.33 ± 1.57	2.1×10^{-5}	Yes	Yes
UGSCNN [21]	5.18M	84.16 ± 2.85	6.9×10^{-33}	Yes	No
UGSCNN+Aug [27]	5.18M	86.28 ± 2.54	3.5×10^{-18}	Yes	Yes
SPHARM-Net (128)	4.31M	88.64 ± 1.83	0.66	No	No
SPHARM-Net (64)	1.10M	88.49 ± 1.87	8.6×10^{-8}	No	No
SPHARM-Net (16)	77K	87.86 ± 1.85	9.0×10^{-41}	No	No

Mindboggle-101 Dataset

Statistical tests reveal that SPHARM-Net at $C = 64$ (SPHARM-Net (64)) and 128 outperforms the baseline methods as summarized in Table 2. The minimal SPHARM-Net ($C = 16$, SPHARM-Net (16)) achieves even better performance than the baseline methods except for SU-Net with data augmentation. However, it should be noted in this case that SPHARM-Net (16) utilizes only 77K learnable parameters compared to 1.67M of SU-Net (21 times more than ours). As observed in Table 2, data augmentation in the baseline spherical CNNs is key to improving their performance. This is mainly due to the absence of rotational equivariance in their spherical convolution, which needs rotational data augmentation to cover the configuration space of spherical rotation. By contrast, rigid alignment and rotational data augmentation statistically have no effects on SPHARM-Net for all the channels ($p = 0.96$ at $C = 16$ and $p = 0.80$ at $C = 64$) because the proposed rotation-equivariant convolution already encodes SO(3). In ROI-wise evaluation, we found that 21, 32, and 15 regions are significantly improved by SPHARM-Net (128) against FreeSurfer, UGSCNN with data augmentation, and SU-Net with data augmentation, respectively. No regions show worse performance than the baseline spherical CNNs but the *insula* than FreeSurfer after multi-comparison correction using false discovery rate (FDR) [46] at $q = 0.05$ (see Fig. 6). We further investigated the performance of SPHARM-Net in a worst-case scenario. Fig. 8 illustrates an example subject with the lowest Dice score of the whole cortex in SPHARM-Net. Despite comparable performance to the baseline method even in this case, SPHARM-Net still improves small regions in visual inspection.

NAMIC Dataset

To investigate the effectiveness of rotational equivariance, we further evaluated SPHARM-Net on the NAMIC dataset that includes a smaller number of participants but more parcellation labels than the Mindboggle-101 dataset. Since the number of ROIs is a bit larger than that of the Mindboggle-101 dataset, the learnable parameters increase in all the methods. As expected, the overall performance across all the methods becomes worse than that on the Mindboggle-101 dataset. Nevertheless, we observed similar trends to the Mindboggle-101 dataset and even SPHARM-Net (16) outperforms all the baseline spherical CNNs ($p < 4.2 \times 10^{-3}$). It is noteworthy that the importance of data augmentation in this dataset is much stressed in the baseline spherical CNNs. A huge performance gap is clearly observed between before and after introducing data augmentation as shown in Table 3. On the other hand, SPHARM-Net achieves comparable performance regardless of rigid alignment ($p = 0.75$ at $C = 16$ and $p = 0.98$ at $C = 64$). Once again, this suggests that rotational equivariance in the proposed convolution avoids extensive data augmentation without losing performance. Meanwhile, no significant difference in overall Dice overlap is revealed between FreeSurfer and SPHARM-Net (128). In ROI-wise evaluation, we found that 1, 35, and 14 regions are significantly improved by SPHARM-Net (128) against FreeSurfer, UGSCNN with data augmentation, and SU-Net with data augmentation, respectively. No regions show worse performance than the baseline methods after FDR at $q = 0.05$ as shown in Fig. 7. Fig. 9 illustrates an example of the improved Dice overlap in a small region by SPHARM-Net.

Table 3: Mean Dice score on NAMIC. *Param*: # of learnable parameters. *Rot*: rigid alignment. *Aug*: data augmentation.

Method	Param	Dice (%)	p value	Rot	Aug
SPHARM-Net (128)	4.32M	85.48 ± 2.52	-	Yes	No
SPHARM-Net (64)	1.10M	85.33 ± 2.45	2.0×10^{-4}	Yes	No
SPHARM-Net (16)	78K	84.88 ± 2.61	1.2×10^{-14}	Yes	No
FreeSurfer [4]	-	85.51 ± 2.53	0.86	Yes	No
SU-Net [22]	1.67M	76.79 ± 2.73	4.5×10^{-26}	Yes	No
SU-Net+Aug [22]	1.67M	84.50 ± 2.14	1.4×10^{-10}	Yes	Yes
UGSCNN [21]	5.18M	76.74 ± 5.63	7.3×10^{-15}	Yes	No
UGSCNN+Aug [27]	5.18M	81.53 ± 4.36	2.3×10^{-8}	Yes	Yes
SPHARM-Net (128)	4.32M	85.45 ± 2.46	0.16	No	No
SPHARM-Net (64)	1.10M	85.33 ± 2.46	4.1×10^{-4}	No	No
SPHARM-Net (16)	78K	84.86 ± 2.56	2.9×10^{-17}	No	No

4.5 Further Performance Evaluation

We further investigated the performance of SPHARM-Net in terms of spherical tessellation, runtime benchmark, and rotation-equivariant inference on the right hemisphere from the left hemisphere. We
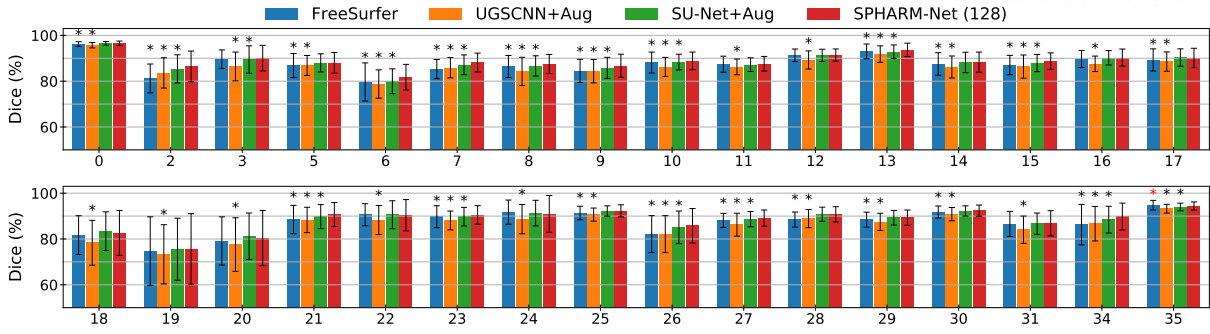
Figure 6: ROI-wise Dice score on Mindboggle-101. The statistical significance is reported after multi-comparison correction among the 32 ROIs (FDR at $q = 0.05$). SPHARM-Net shows higher Dice scores than FreeSurfer, UGSCNN+Aug, and SU-Net+Aug in 21, 32, and 15 ROIs, respectively. No ROI is observed with worse performance in SPHARM-Net against the baseline CNNs, whereas the *insula* (35) achieves worse performance than FreeSurfer. *: statistical significance with better performance. *: statistical significance with worse performance. *Label number*: see Table 1.
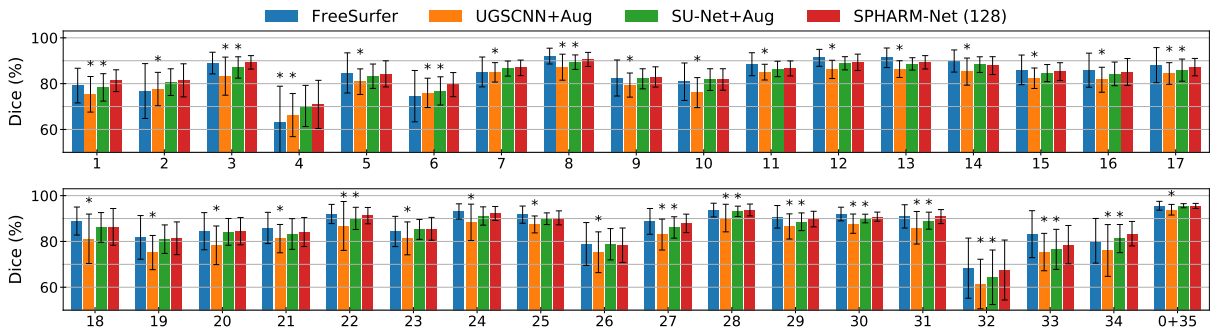
Figure 7: ROI-wise Dice score on NAMIC. The statistical significance is reported after multi-comparison correction among the 35 ROIs (FDR at $q = 0.05$). SPHARM-Net shows Dice scores than FreeSurfer, UGSCNN+Aug, and SU-Net+Aug in 1, 35, and 14 ROIs, respectively. No ROI is observed with worse performance in SPHARM-Net against the baseline methods. *: statistical significance. *Label number*: see Table 1.

used the Mindboggle-101 dataset for the evaluation, and the same baseline spherical CNNs as in the previous section were employed to evaluate runtime benchmark and rotational equivariance.

Spherical Tessellation

The proposed filter is independent of a neighborhood definition as it is a collection of spherical harmonic coefficients. Meanwhile, SHT approximates the integral of Eq. (5) by Eq. (26) in SPHARM-Net and can be thus affected by sampling uniformity. To investigate this, we evaluated SPHARM-Net on an equiangular grid along θ and φ with an interval of $\pi/144$ (40,898 vertices) as well as a lower resolution of icosahedral subdivision (10,242 vertices). Table 4 summarizes Dice scores with respect to spherical tessellation. A performance gap is observed between with and without rigid alignment in the equiangular sampling ($p = 2.4 \times 10^{-4}$) as denser sampling points exist around the poles than the equator in a unit

	Manual annotation Dice (%)	FreeSurfer 82.07	UGSCNN+Aug 83.43
		SU-Net+Aug 85.07	SPHARM-Net (128) 84.52
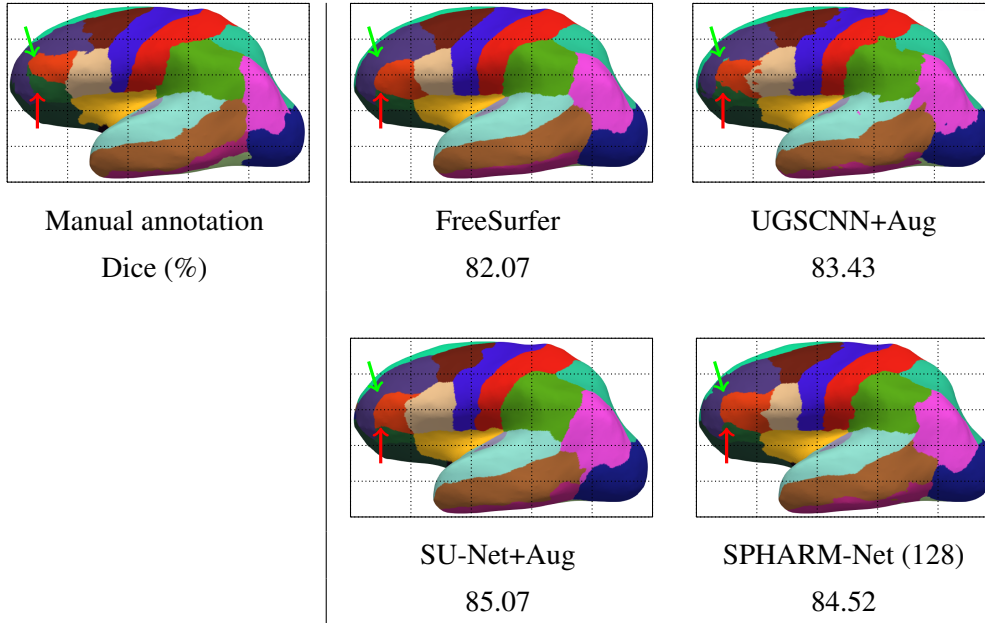
Figure 8: Visual inspection of worst-case scenario (lowest mean Dice) in SPHARM-Net within Mindboggle-101 and its comparison to the baseline methods. Overall, all the methods achieve comparable Dice scores in this example while a subtle visual difference is observed in small regions. On the boundary of *Triangularis* (*orange*), the upper (*green* arrow) and lower (*red* arrow) boundaries are underestimated by the baseline methods, whereas SPHARM-Net offers improved inference on the upper boundary. The inflated surface is used for better visualization.

Dice (%)	81.00	67.60	72.64	85.63	87.73
Manual annotation	FreeSurfer	UGSCNN+Aug	SU-Net+Aug	w/o SHConv-FB	SPHARM-Net (128)
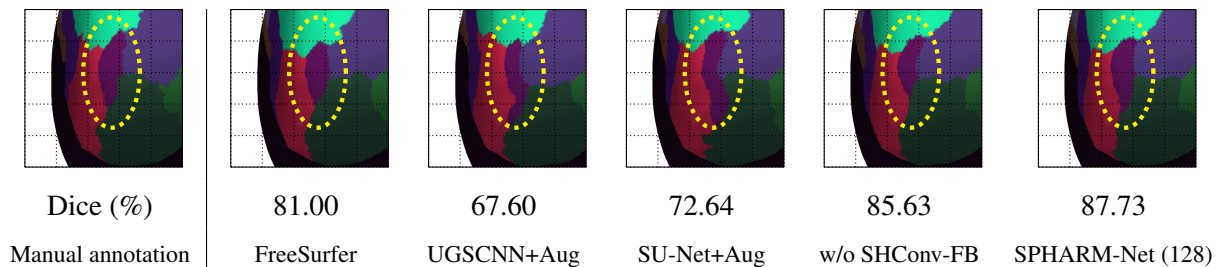
Figure 9: Visual inspection of improved parcellation accuracy in a small region. The zoomed-in *frontal pole* colored by *violet* (*yellow* dotted ellipse) is illustrated from a coronal view. The overall size in FreeSurfer is smaller than the manual annotation, whereas longer tails are annotated by the baseline spherical CNNs. SPHARM-Net without SHConv-FB improves the accuracy but shrinks the overall size, which is likely because geometric details are overly smoothed out by the harmonic truncation. This limitation is further overcome by the use of SHConv-FB that supports full-bandwidth convolution. The inflated surface is used for better visualization.

area, compared to the icosahedral subdivision. In the icosahedral subdivision, consistent performance is observed regardless of rigid alignment ($p = 0.39$) while Eq. (26) becomes less accurate than the higher resolution, yielding lower accuracy.

Table 4: SPHARM-Net (128) with different spherical tessellation on Mindboggle-101. *Vertices*: # of vertices. *Rot*: rigid alignment.

Method	Vertices	Dice (%)	Rot
Icosahedron	40,962	88.64 ± 1.83	Yes
Icosahedron	10,242	88.57 ± 1.85	Yes
Equiangle	40,898	88.24 ± 1.83	Yes
Icosahedron	40,962	88.64 ± 1.83	No
Icosahedron	10,242	88.55 ± 1.82	No
Equiangle	40,898	88.17 ± 1.82	No

Runtime Benchmark

Table 5 summarizes runtime comparisons to the baseline spherical CNNs. As SPHARM-Net reserves spherical harmonic bases (1,001 MB GPU memory on 40,962 vertices in single precision), it tends to utilize more memory than the baseline spherical CNNs. Meanwhile, SPHARM-Net is free from rotational data augmentation, by which it converges in 20 epochs, whereas [22] and [27] converge in 15 and 30 epochs with a large number of samples. As SPHARM-Net (128) has a higher iteration-per-second rate than [22] and [27], the training time of SPHARM-Net (128) is 38.85 and 19.21 times faster than [22] and [27], respectively, as summarized in Table 5. Although [22] achieves the fastest inference time, all the methods take less than a second. Note that the data preprocessing is excluded in measuring training and inference time. In this experiment, we observed that SPHARM-Net can achieve comparable performance to that of the baseline methods with a small sample size but while offering fast training.
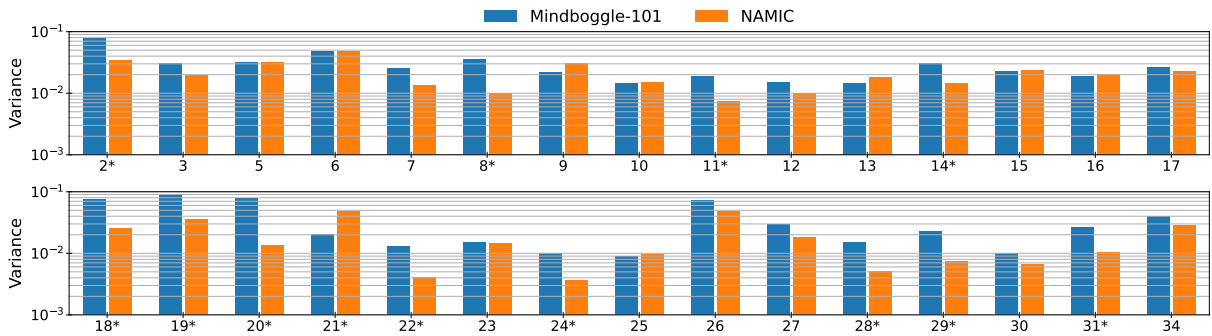
Figure 10: Area variance of common ROIs between Mindboggle-101 and NAMIC. To test equal variance, ROIs are mapped onto the unit sphere. Because some ROIs are merged into their neighbors in Mindboggle-101, computing variance can have a scaling issue. Hence, the mapped area of each ROI is divided by the mean area of that ROI in each dataset. F-tests reveal that 13 regions have different variance between the two datasets after FDR at $q = 0.05$, potentially indicating a performance gap between the two datasets. *: statistical significance. *Label number*: see Table 1.

Table 5: Benchmark on Mindboggle-101 at batch size of 1. The GPU memory consumption may vary depending on CUDA kernels. *Samples*: # of training samples. *Ips*: iteration-per-second rate. *Epoch*: maximum epochs. *Param*: # of learnable parameters.

Method	Memory (MB)	Samples	Ips	Epoch	Training (min)	Inference (msec)	Param	Dice (%)
SPHARM-Net (128)	8,081	60	9.19	20	2.18	44.14	4.31M	88.64 ± 1.83
SPHARM-Net (64)	4,297	60	15.17	20	1.32	25.54	1.10M	88.49 ± 1.83
SPHARM-Net (16)	2,817	60	16.55	20	1.21	17.60	77K	87.86 ± 1.84
SU-Net+Aug [22]	1,665	3,000	8.87	15	84.55	6.44	1.67M	88.33 ± 1.57
UGSCNN+Aug [27]	2,331	660	7.89	30	41.82	261.25	5.18M	86.28 ± 2.54

Rotational Equivariance

We further investigated rotational equivariance. We used the right hemisphere as an external validation set for the trained models in Table 2. Since the right hemisphere is reflected and potentially rotated, we built the two respective test sets with and without reflection and rigid alignment of the right hemisphere. As shown in Table 6, huge performance degeneration is observed without reflection and rigid alignment in the baseline spherical CNNs, whereas SPHARM-Net consistently subdivides the right hemisphere regardless of the data preprocessing ($p = 0.75$). This is because the data orientation is learned by rotational equivariance and the proposed zonal (isotropic) filters generate consistent responses independent of the data reflection.

Non-rigid Data Augmentation

To understand the impact of non-rigid data augmentation, we investigated the performance of SPHARM-Net on the augmented training samples generated by [27]. As shown in Table 7, no significant difference was revealed between with and without the augmentation technique ($p = 0.68$). This consistency may occur because the proposed filter is isotropic, implying limited flexibility in fully capturing non-rigid local changes. Further investigation into the filter representation could improve SPHARM-Net.

Table 6: Inference on the right hemisphere using the models (Table 2) trained from the left hemisphere in Mindboggle-101. *Ref+Rot*: reflection and rigid alignment of test data.

Method	Dice (%)	Ref+Rot
SPHARM-Net (128)	86.46 ± 1.84	Yes
SU-Net+Aug [22]	85.83 ± 1.88	Yes
UGSCNN+Aug [27]	84.01 ± 3.74	Yes
SPHARM-Net (128)	86.45 ± 1.84	No
SU-Net+Aug [22]	14.91 ± 2.11	No
UGSCNN+Aug [27]	15.83 ± 2.04	No

Table 7: SPHARM-Net (128) with non-rigid augmentation on Mindboggle-101. *Rot*: rigid alignment. *Aug*: data augmentation.

Method	Dice (%)	*p* value	Rot	Aug	Samples
SPHARM-Net (128)	88.64 ± 1.83	-	Yes	No	60
SPHARM-Net (128)+Aug	88.62 ± 1.86	0.68	Yes	Yes	660

V   Conclusion

In this work, we proposed SPHARM-Net, a novel spherical harmonics-based CNN built on the convolution theorem. The main motivation of using spherical harmonics is to model full-bandwidth convolution and to preserve rotational equivariance while avoiding a specific neighborhood definition for spherical convolution. To overcome limited machine capacity, we introduced a learnable impulse response at the north pole, where a single learnable parameter covers full bandwidth. This improves the flexibility of the proposed filter in spherical convolution by allowing high-frequency components. The experimental results show that the enhanced descriptive features can be extracted from spherical signals and rotational equivariance can capture spatially varying anatomical sub-regions. We observed that rotational equivariance can drastically reduce the training time while improving parcellation performance. Furthermore, the proposed convolution is fully composed of matrix transformations, which offers efficient and fast training and inference. On public datasets, we showed that SPHARM-Net outperforms the state-of-the-art methods with fewer learnable parameters without extensive data augmentation.

References

[1] R. S. Desikan *et al.*, "An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest," *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006.

[2] B. Thomas Yeo *et al.*, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *Journal of neurophysiology*, vol. 106, no. 3, pp. 1125–1165, 2011.

[3] A. Klein and J. Tourville, "101 labeled brain images and a consistent human cortical labeling protocol," *Frontiers in neuroscience*, vol. 6, p. 171, 2012.

[4] B. Fischl, A. Van Der Kouwe, C. Destrieux, E. Halgren, F. Ségonne, D. H. Salat, E. Busa, L. J. Seidman, J. Goldstein, D. Kennedy *et al.*, "Automatically parcellating the human cerebral cortex," *Cerebral cortex*, vol. 14, no. 1, pp. 11–22, 2004.

[5] B. T. Yeo, M. R. Sabuncu, T. Vercauteren, N. Ayache, B. Fischl, and P. Golland, "Spherical demons: fast diffeomorphic landmark-free surface registration," *IEEE transactions on medical imaging*, vol. 29, no. 3, pp. 650–668, 2009.

[6] B. Fischl, "Freesurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.

[7] I. Lyu, H. Kang, N. D. Woodward, M. A. Styner, and B. A. Landman, "Hierarchical spherical deformation for cortical surface registration," *Medical image analysis*, vol. 57, pp. 72–88, 2019.

[8] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*. IEEE, 2016, pp. 565–571.

[9] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. N. L. Benders, and I. Išgum, "Automatic segmentation of mr brain images with a convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1252–1261, 2016.

[10] H. Chen, Q. Dou, L. Yu, J. Qin, and P.-A. Heng, "Voxresnet: Deep voxelwise residual networks for brain segmentation from 3d mr images," *NeuroImage*, vol. 170, pp. 446–455, 2018.

[11] A. G. Roy, S. Conjeti, N. Navab, and C. Wachinger, "Quicknat: A fully convolutional network for quick and accurate segmentation of neuroanatomy," *NeuroImage*, vol. 186, pp. 713–727, 2019.

[12] Y. Huo *et al.*, "3d whole brain segmentation using spatially localized atlas network tiles," *NeuroImage*, vol. 194, pp. 105–119, 2019.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[14] J. Masci, D. Boscaini, M. Bronstein, and P. Vandergheynst, "Geodesic convolutional neural networks on riemannian manifolds," in *Proceedings of the IEEE international conference on computer vision workshops*, 2015, pp. 37–45.

[15] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 1025–1035.

[16] K. Gopinath, C. Desrosiers, and H. Lombaert, "Graph convolutions on spectral embeddings for cortical surface parcellation," *Medical image analysis*, vol. 54, pp. 297–305, 2019.

[17] S. Gong, L. Chen, M. Bronstein, and S. Zafeiriou, "Spiralnet++: A fast and highly efficient mesh convolution operator," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019.

[18] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *International Conference on Learning Representations (ICLR)*, 2017.

[19] G. Cucurull *et al.*, "Convolutional neural networks for mesh-based parcellation of the cerebral cortex," in *International Conference on Medical Imaging with Deep Learning (MIDL)*, 2018.

[20] S.-B. Seong, C. Pae, and H.-J. Park, "Geometric convolutional neural network for analyzing surface-based neuroimaging data," *Frontiers in neuroinformatics*, vol. 12, p. 42, 2018.

[21] C. M. Jiang, J. Huang, K. Kashinath, Prabhat, P. Marcus, and M. Niessner, "Spherical CNNs on unstructured grids," in *International Conference on Learning Representations (ICLR)*, 2019.

[22] F. Zhao *et al.*, "Spherical u-net on cortical surfaces: methods and applications," in *International Conference on Information Processing in Medical Imaging*. Springer, 2019, pp. 855–866.

[23] N. Perraudin, M. Defferrard, T. Kacprzak, and R. Sgier, "Deepsphere: Efficient spherical convolutional neural network with healpix sampling for cosmological applications," *Astronomy and Computing*, vol. 27, pp. 130–146, 2019.

[24] P. De Haan, M. Weiler, T. Cohen, and M. Welling, "Gauge equivariant mesh cnns: Anisotropic convolutions on geometric graphs," in *International Conference on Learning Representations (ICLR)*, 2021.

[25] T. S. Cohen, M. Geiger, J. Köhler, and M. Welling, "Spherical cnns," in *International Conference on Learning Representations (ICLR)*, 2018.

[26] C. Esteves, C. Allen-Blanchette, A. Makadia, and K. Daniilidis, "Learning so (3) equivariant representations with spherical cnns," *International Journal of Computer Vision*, vol. 128, no. 3, pp. 588–600, 2020.

[27] P. Parvathaneni *et al.*, "Cortical surface parcellation using spherical convolutional neural networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 501–509.

[28] F. Zhao *et al.*, "Spherical deformable u-net: Application to cortical surface parcellation and development prediction," *IEEE transactions on medical imaging*, vol. 40, no. 4, pp. 1217–1228, 2021.

[29] J. R. Baumgardner and P. O. Frederickson, "Icosahedral discretization of the two-sphere," *SIAM Journal on Numerical Analysis*, vol. 22, no. 6, pp. 1107–1115, 1985.

[30] K. M. Gorski *et al.*, "Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere," *The Astrophysical Journal*, vol. 622, no. 2, p. 759, 2005.

[31] J. R. Driscoll and D. M. Healy, "Computing fourier transforms and convolutions on the 2-sphere," *Advances in applied mathematics*, vol. 15, no. 2, pp. 202–250, 1994.

[32] E. J. Bekkers, "B-spline cnns on lie groups," in *International Conference on Learning Representations (ICLR)*, 2020.

[33] S. Ha and I. Lyu, "Spharm-net: Spherical harmonics-based convolution for cortical parcellation," *IEEE Transactions on Medical Imaging*, 2022.

[34] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3D shape descriptors," in *Symposium on Geometry Processing*, Jun. 2003.

[35] M. K. Chung, K. M. Dalton, L. Shen, A. C. Evans, and R. J. Davidson, "Weighted fourier series representation and its application to quantifying the amount of gray matter," *IEEE transactions on medical imaging*, vol. 26, no. 4, pp. 566–581, 2007.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[37] O. Rippel, J. Snoek, and R. P. Adams, "Spectral representations for convolutional neural networks," *arXiv preprint arXiv:1506.03767*, 2015.

[38] S. A. Holmes and W. E. Featherstone, "A unified approach to the clenshaw summation and the recursive computation of very high degree and order normalised associated legendre functions," *Journal of Geodesy*, vol. 76, no. 5, pp. 279–299, 2002.

[39] J. A. Jacobs, *Geomagnetism: Volume 4*. Elsevier, 2016, vol. 4.

[40] M. Chung, "Heat kernel smoothing on unit sphere," in *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.*, 2006, pp. 992–995.

[41] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, 2015, pp. 448–456.

[42] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *International Conference on Machine Learning*, 2010, p. 807–814.

[43] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," in *International Conference on Learning Representations (ICLR)*, 2014.

[44] I. Lyu *et al.*, "Labeling lateral prefrontal sulci using spherical data augmentation and context-aware training," *Neuroimage*, vol. 229, p. 117758, 2021.

[45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, 2015.

[46] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal statistical society: series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

Acknowledgements

This project would not have been possible without the support of my adviser, Ilwoo Lyu. His insight and patience encouraged me in all the time of my academic research. With his theoretical and technical advice, I could finally experience how mathematics can be used practically in computer science. Moreover, I could see that I can utilize knowledge that I have at the moment. I also would like to thank Ulsan National Institute of Science and Technology for giving me an opportunity to do this work and study with great professors.

From the bottom of my heart, I would like to thank my dear friends and family who have endured the whole journey with me. Academic discussion and daily conversation with you have made me grow. I wish you all better than your hearts desire.