Group Convolutional Neural Networks for DWI Segmentation

Renfei Liu * RENFEI.LIU@DI.KU.DK and **François Lauze** FRANCOIS@DI.KU.DK
University of Copenhagen

Erik Bekkers† E.J.BEKKERS@UVA.NL
University of Amsterdam

Kenny Erleben KENNY@DI.KU.DK and **Sune Darkner** DARKNER@DI.KU.DK
University of Copenhagen

Abstract

We present a Group Convolutional Network for Segmentation of Diffusion Weighted Imaging data (DWI). The network incorporates group actions that are natural for this type of data, in the form of $SE(3)$ equivariant convolutions, i.e., roto-translation equivariant convolutions. The equivariance property provides an important inductive bias and may alleviate the need for data augmentation strategies. Instead of performing group equivariant convolutions via spectral (Fourier-based) approaches, as is common for $SE(3)$ equivariance, we implement direct and light-weight regular group convolutions. We study the effect of equivariance and weight sharing over $SE(3)$ on performances of the networks on DWI scans from the Human Connectome project. We show how that full $SE(3)$ equivariance improves segmentations, while limiting the number of learnable parameters.

Keywords: DWI, Group action, Homogeneous spaces G-CNN, Image Segmentation

1. Introduction

In this work, we propose a group convolutional neural network (G-CNN) for Diffusion Weighted Imaging (DWI) data. CNNs rely on assumed translational symmetries in data and have shown very robust performance in imaging tasks, especially medical imaging ones, and they are highly memory-efficient thanks to their weight-sharing property. When data offer more structure than translation, they can be used to build generalized CNNs. These Group and Geometric CNNs (GCNN) have been studied intensively and applied in many situations in the few past years, see e.g. (Masci et al., 2015; Cohen and Welling, 2016; Boscaini et al., 2016; Bekkers et al., 2018; Cohen et al., 2020). Structure beyond translational symmetries is particularly apparent in DWI data; A DWI scan can be modeled as a

* This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 801199. This paper only contains the author's views. The Research Executive Agency and the Commission are not responsible for any use that may be made of the information it contains. Data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. This project is also partially funded by 3Shape A/S.

† This project is also partially funded by the research programme VENI (grant number 17290), financed by the Dutch Research Council (NWO).

function $f : \mathbb{R}^3 \times \mathbb{S}^2 \to \mathbb{R}$ providing at each (x, v), with spatial position x and direction v, a response (Tuchs, 2004). A rigid transformation of a sample (i.e. by the action of the group $SE(3)$), should, up to the limitations of acquisition protocol, be reflected in the signal.

The space $\mathbb{R}^3 \times \mathbb{S}^2$ is a *homogeneous space* under the action of $SE(3)$: a point in $\mathbb{R}^3 \times \mathbb{S}^2$ can be transformed in any other point by a rigid transformation. This notion of homogeneous space is at the heart of the extension of CNNS to GCNNs (Cohen et al., 2020; Bekkers, 2019).

Our task at hand is the classification/segmentation of diffusion data. The inductive bias provided by the knowledge of these transformations may prove important for our task, especially when annotated data is limited. How to incorporate this knowledge? This is classically done by data augmentation, in the hope that the network will learn transformation-aware features during training. Incorporating, on the other hand, information about group actions on the data has shown to boost performances of these networks (Bekkers et al., 2018) whilst avoiding the need for augmentation, which, for DWI data, is non-trivial. To exploit the rigid motion action in the space of DWI - $\mathbb{R}^3 \times \mathbb{S}^2 \to \mathbb{R}$, we propose an $SE(3)$-GCNN.

Most CNNs approaches for processing of DWI signals discard its specific structure. For instance, Golkov et al. (Golkov et al., 2016) built multi-layer perceptrons in q-space for kurtosis and NODDI mappings. On the other hand, the importance of spherical equivariant or invariant structure has been acknowledged for some years now. The importance of the extraction of rotationally invariant features beyond Fractional Anisotropy (Basser et al., 1994) has been recognized in series of DWI works, for their usefulness in understanding microstructures, see for instance (Schwab et al., 2013; Caruyer and Verma, 2015; Novikov et al., 2018; Zucchelli et al., 2020). In (Sedlar et al., 2020), a spherical U-net based on (Driscoll and Healy, 1994) was used for neurite orientation. Cohen et al. (Cohen et al., 2018) lifted spherical functions to the 3D-rotation group $SO(3)$ using convolutions based on spherical/$SO(3)$-Fourier transforms. In (Sedlar et al., 2021), this idea was used for microstructure parameter estimation. In (Chakraborty et al., 2018; Banerjee et al., 2019), it was used for disease classification. In (Müller et al., 2021), a 6-D - 3D space and q-space - NNs with roto-translation was proposed. Several authors (Gens and Domingos, 2014; Cohen and Welling, 2016; Weiler et al., 2018; Worrall et al., 2017; Kondor and Trivedi, 2018; Bekkers et al., 2018; Andrearczyk et al., 2020; T.S. Cohen and M. Weiler and B. Kicanaoglu and M. Welling, 2019; Cohen et al., 2020; Aronsson, 2021) further explored the group convolution path for Lie groups and their homogeneous spaces.

The contributions of this paper are as follows:

1. Unlike existing works that use Fourier-based convolutions, we implement convolution in all our experiments in a direct way as it is usually done in the image analysis community, which is way more light-weight and is based on localized filters.

2. We show in our experiments that the equivariance we provide in our networks improves tasks performed on the DWI data.

3. We demonstrate in our results that for DWI data, exploiting both the spatial and the spherical structures of the signals is essential, as opposed to handling spatial translations or spherical rotations individually.

In the rest of this paper, we propose GCNNs with two types of group action, with equivariant layers for these actions and nonlinear ones. We show how incorporating these actions improve DWI segmentation performance compared to classical CNNs and CNNS with limited notion of symmetry (Liu et al., 2021), by evaluating them on scans from the Human Connectome Project (HCP) (Van Essen et al., 2013).

2. Method

The networks we present are built by extending standard CNNs to groups G and homogeneous spaces $\mathcal{M}$ on which they act by extending convolution operations to them. We do not follow, however, the common path of irreducible representations for implementing convolutions/correlations over $\mathbb{S}^2$ or $SO(3)$ (Cesa et al., 2021). Instead, we provide a direct convolution implementation that is much light-weight. For the general theory of convolutions on groups and *homogeneous spaces*, we refer the readers to (Cohen et al., 2020).

2.1. Standard convolution operations

Each group G we consider is endowed with a left-invariant Haar measure. Each homogeneous space we consider is endowed with a G-invariant measure. Functions are assumed to be square-integrable for these measures. The layers $\mathcal{L}$ we define are all equivariant $\mathcal{L}(L_g f) = L_g(\mathcal{L}f)$, w.r.t. the regular *left translation* representation L_g of G on the respective input/output spaces.

Lifting layer. A function $f : \mathcal{M} \to \mathbb{R}^N$ can be *lifted* to the group G via a kernel $\kappa : \mathcal{M} \to \mathbb{R}^N$ by

$$\kappa * f(g) = \sum_{i=1}^{N} \int_{\mathcal{M}} f_i(m)\kappa_i(g^{-1}m)\, dm \tag{1}$$

Group convolution layer. A feature function $F : G \to \mathbb{R}^N$ can be transformed by a convolution kernel $K : G \to \mathbb{R}^N$ by

$$K * F(g) = \sum_{i=1}^{N} \int_{G} F_i(h)K_i(h^{-1}g)dh. \tag{2}$$

Projection layer. If needed, feature map $F : G \to \mathbb{R}^n$ can be projected to a function $f : \mathcal{M} \to \mathbb{R}^n$ by summarizing on the fibres

$$\overline{F}(m) = \max_{h \in G_{m_0}} F(gh), \quad \text{for any } g \text{ with } g.m_0 = m, \tag{3}$$

where the max is computed component-wise.

Activation functions and separable kernels. A point-wise activation function α, such as ReLU, is trivially equivariant. On manifolds with a product structure, $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$, both for homogeneous spaces and groups, using separable kernels $\kappa = \kappa_{\mathcal{M}_1} \otimes \kappa_{\mathcal{M}_2}$, the layers can be split by sequential application of convolutions on these sub-domains as to obtain an implementation of separable group convolutions (Knigge et al., 2022).

Table 1: The groups and homogeneous spaces in this work. For each group and each homogeneous space, typical elements are provided, as well as the corresponding group actions.

G \ $\mathcal{M}$	$\mathbb{R}^3, x$	$\mathbb{S}^2, \boldsymbol{v}$	$\mathbb{R}^3 \times \mathbb{S}^2, (x, \boldsymbol{v})$
$\mathbb{T}^3, \boldsymbol{t}$	$x + \boldsymbol{t}$		
$SO(3), R$		$R\boldsymbol{v}$	
$SE(3), (R, \boldsymbol{t})$	$Rx + \boldsymbol{t}$	$R\boldsymbol{v}$	$(Rx + \boldsymbol{t}, R\boldsymbol{v})$

Spaces and groups. The spaces used in this work are $\mathbb{R}^3$, the sphere $\mathbb{S}^2$ and the product space $\mathbb{R}^3 \times \mathbb{S}^2$. The groups that we consider are: translations of $\mathbb{R}^3$ - $\mathbb{T}^3 \simeq \mathbb{R}^3$, 3D rotations - $SO(3)$, and the special Euclidean group $SE(3) = SO(3) \ltimes \mathbb{T}^3$. Table (1) shows the different combinations of spaces and groups. Entries left empty are not used or fail to be homogeneous spaces for standard group actions on them.

2.2. Discretization of spherical signals

The way spherical signals are numerically handled have major implications for our networks. A DWI signal is treated as a discretization of a signal $f : \mathbb{R}^3 \times \mathbb{S}^2 \to \mathbb{R}$. DWIs are acquired, for each voxel, at N fixed directions $p_1, \ldots, p_N$ on $\mathbb{S}^2$ (here $N = 90$). The directional measurements can be represented in two different ways.

- Type 1. Ignoring the spherical structure, at each voxel x, we get a measurement vector $I(x) = (I(x, p_1) \ldots, I(x, p_N)) \in \mathbb{R}^N$. Thus a DWI signal is a mapping $I : \mathbb{R}^3 \to \mathbb{R}^N$.
- Type 2. A DWI signal is a function $I : \mathbb{R}^3 \times \mathbb{S}^2 \to \mathbb{R}$ by interpolating the spherical signal $I(x, \boldsymbol{v})$ at each voxel x using a Watson kernel (Jupp and Mardia, 1989) s.t. that the anti-podal symmetry of the kernel resembles the same property of DWI signals.

3. Experiments and Results

We evaluate our method on the DWI brain scan dataset from the human connectome project (Van Essen et al., 2013). We classify the human brains into 4 regions - cerebrospinal fluid (CSF), subcortical, white matter, and grey matter. An illustration of the task can be found in fig. 2.

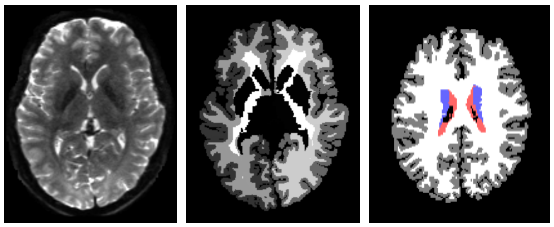

Figure 2: Left to right: original diffusion data, the ground-truth segmentation, and the processed ground-truth that we are going to learn from. The label colors for CSF, subcortical, white matter and grey matter are red, blue, white and grey respectively. The figures only illustrate the data, they are not necessarily from the same slice of the same scan.

We used the pre-processed DWI data (Van Essen et al., 2013) and normalized each DWI scan for the b-1000 images with the voxel-wise average of the b_0. The labels provided with the T1-image were transformed to the DWI using nearest neighbor interpolation (fig. 2). Since the 4 brain regions we are classifying have imbalanced numbers

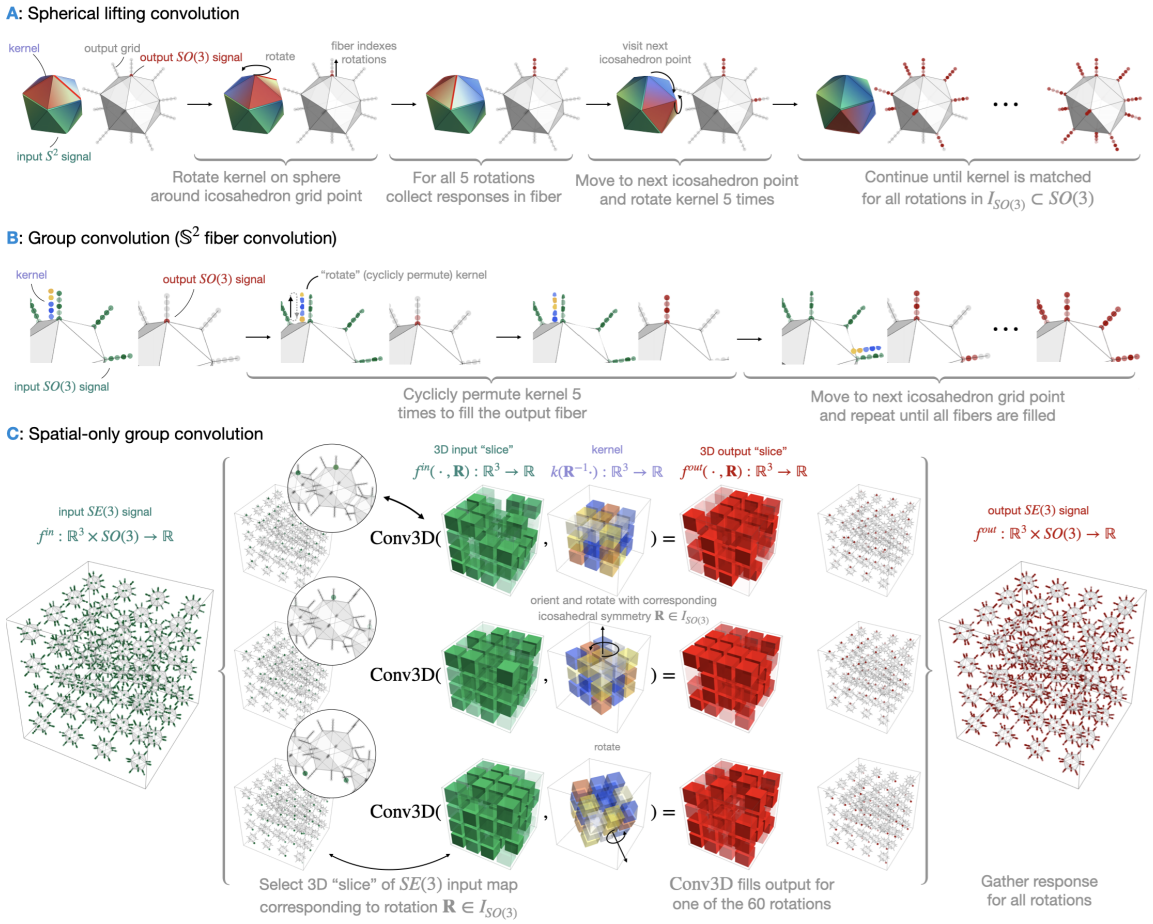

Figure 1: The three group convolution operators used in this paper. Fig. A shows the spherical part of the separable lifting convolution. The star-shaped kernel moves over all 12 icosahedron vertices like a spider crawling on a sphere. At each vertex, the kernel rotates 5 times, aligned with the edges of the icosahedron, and gets 5 rotational responses as such. There are in total 60 responses from all 12 vertices, coming from 60 rotations effectively applied to the local spherical kernel - the icosahedral discretization $I_{SO(3)} \subset SO(3)$. Fig. B shows the spherical part of the separable group convolution. The 1D kernel only lives on the fibers, and is rotated (permuted) again 5 times to get the responses of different orientations, as in the lifting convolution. Fig. C shows the spatial part of the separable convolution. The spatial kernel is a 3D grid and is rotated 60 times with the icosahedral symmetry rotations on which the input as sampled; each rotated kernel convolves the corresponding input "rotation channel".

of voxels, we use a Focal Loss (Lin et al., 2018) to counter the class imbalance. For Focal Loss, all experiments use $\alpha = (0.35, 0.35, 0.15, 0.15)$ for CSF, subcortical, WM, and GM respectively, and $\gamma = 2$. For Watson Kernel, all experiments that used this interpolation (Type 2 discretization) have $\kappa = 10$. Batch size for all experiments is 100.

3.1. Experiment setup

To reduce the computation burden, as inputting a full DWI volume is intractable, we use spatial windows of N^3 voxels, with $N = 1$ for $SO(3)$-action network, i.e. point-wise spherical CNNs, and $N = 7$ for the rest. In addition, due to the effect of convolutions without padding using kernels with a spatial extent, the 7^3 grid of voxels shrinks to 1^3. $\mathbb{S}^2$ is discretized by a regular icosahedron. $SO(3)$ is discretized as the icosahedral rotation group with 60 elements. Each vertex of the icosahedron is fixed by 5 rotations, isomorphic to the subgroup of $SO(2)$ consisting of rotations of angle $2k\pi/5$, $k = 0 \dots 4$, see also fig. 1.

3.2. Networks and architectures

3.2.1. $\mathbb{T}^3$: Classical CNN

For the $\mathbb{T}^3$ group action, the $\mathbb{S}^2$-structure of the signals is ignored, using Type 1 discretization. Each voxel becomes a flattened vector. Using the group $\mathbb{T}^3$, we just obtain a multi-channel standard CNN. We use a $ReLU(\mathbb{R}^3 conv) - ReLU(\mathbb{R}^3 conv) - ReLU(\mathbb{R}^3 conv) - FC$ architecture, with a small and a big network setup. The number of feature channels in the small network is $(90 - 5 - 5 - 5 - 4)$, which we label as Classical$^-$, and the big network with $(90 - 120 - 120 - 90 - 4)$ is labeled as Classical$^+$.

3.2.2. $SO(3)$: Baseline

This time the spatial structure is ignored, thus each voxel is a data sample. Type 2 discretization is used. The GCNN takes as input a spherical function, and classifies it by performing $SO(3)$-lifting, $SO(3)$-convolutions and summarization. The convolved function on $SO(3)$ is then projected back to $\mathbb{S}^2$ by this summarization. Convolutions over the $SO(2)$ fibers is illustrated in fig. 1 B. In the experiments, we use the $ReLU(lift) - ReLU(gconv) - project - FC$ architecture as was used in (Liu et al., 2021), but with true $SO(3)$-convolution. Here $lift$ corresponds to the spherical lifting convolution (fig. 1 A) and $gconv$ to the fiber convolution (fig. 1 C). The projection layer takes the maximum of the 5 rotations (fibers) to collapse the function back to the sphere. Two capacities are used $(1-5-4)$ and $(10-20-4)$, resp. named Baseline$^-$ and Baseline$^+$.

3.2.3. $SE(3)$: Ours

Type 2 discretization is used and the network uses the full interplay between spatial roto-translations and corresponding rotations of the spherical signal. It is separated into a spherical part and a spatial part as explained above. To perform the segmentation task, the projection layer collapses the function on $SE(3)$ back to $\mathbb{R}^3$ by summarizing over $SO(3)$. Both lifting and group convolution layers are separated into 2 parts - the spherical and the spatial part. Detailed illustrations can be found in fig. 1. We use the $ReLU(lift) - ReLU(gconv) - ReLU(gconv) - ReLU(gconv) - project - FC$ architecture. Here $gconv$ refers to a separable $SE(3)$ convolution with a fiber convolution (fig. 1 B) followed by a spatial group convolution (fig. 1 C). The separable convolution explained above is illustrated in fig. 3. We use $(5 - 5 - 5 - 5 - 5 - 5 - 5 - 4)$ for a small network (Ours$^-$) and $(10 - 20 - 20 - 40 - 40 - 20 - 10 - 4)$ for a big network (Ours$^+$).

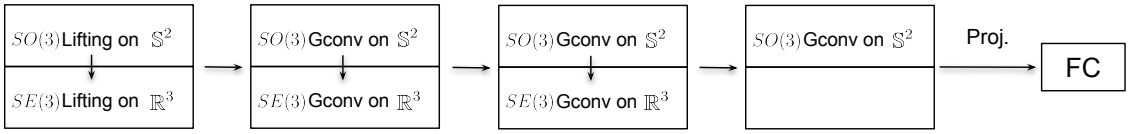

Figure 3: Architecture of our $SE(3)$ network. Each block is a convolutional layer split into 2 separable layers. The last block before the FC layer is equivalent to a single $\mathbb{S}^2$-layer as explained in section 3.1. The vertical arrows in each block shows the separable convolutions. First the spherical convolution is applied, followed by the spatial convolution.

3.3. Results

As was done in (Liu et al., 2021), we trained all networks using **1** scan, validated using **1** scan, and tested using **50** scans. Training with more scans has been tried, but it did not seem to improve the results significantly, thus we continued using only one scan in training.

We evaluate the accuracies and Dice scores and accuracies of the classification of the 4 regions respectively, and the overall classification accuracy across all test scans. Numbers reported in table 2 are the means and standard deviations of the respective criteria across 50 test scans. For each class, the accuracy is calculated by $\frac{\#ClassCorrect}{\#ClassSamples}$, and the Dice score by $\frac{2TP}{2TP+FP+FN}$ for the class. The overall accuracy is calculated by $\frac{\#Correct}{\#AllSamples}$.

We trained all models until they converge and start to overfit, thus models are stopped at different epochs based on evaluation on the validation set. Classical$^-$ and Classical$^+$ were trained for 34 and 19 epochs, Baseline$^-$ and Baseline$^+$ were both trained for 31 epochs, and Ours$^-$ and Ours$^+$ were trained for 41 and 15 epochs.

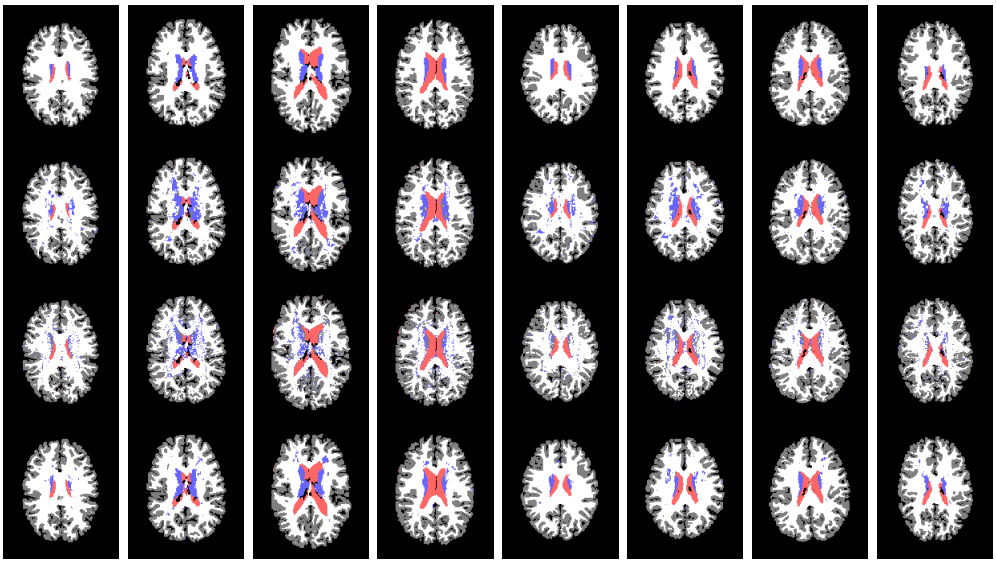

Figure 4: From top to bottom: ground-truth, predictions from Classical$^+$, Baseline$^+$, and Ours$^+$. The colors of CSF, subcortical, WM and GM resp. are red, blue, white, and grey.

The Dice scores and accuracies of all experiments can be found in table 2. It is easy to see that our method, with the smallest model capacity, has the best performance. The

Table 2: Statistics of Dice scores and Accuracies.

G	Model (#Param)	CSF	Subcortical	WM	GM	Overall
Dice scores						
$I : \mathbb{R}^3 \to \mathbb{R}^N$						
$\mathbb{T}^3$	Classical$^-$(13539)	0.756 ± 0.07	0.376 ± 0.043	0.834 ± 0.011	0.839 ± 0.02	
	Classical$^+$(972694)	$\mathbf{0.804} \pm 0.053$	0.583 ± 0.036	0.856 ± 0.011	0.893 ± 0.009	
$I : \mathbb{R}^3 \times \mathbb{S}^2 \to \mathbb{R}$						
$SO(3)$	Baseline$^-$(286)	0.75 ± 0.073	0.185 ± 0.04	0.801 ± 0.012	0.83 ± 0.011	
	Baseline$^+$(2104)	0.754 ± 0.069	0.334 ± 0.037	0.805 ± 0.013	0.841 ± 0.012	
$SE(3)$	Ours$^-$(2514)	0.769 ± 0.06	0.621 ± 0.038	0.854 ± 0.01	0.891 ± 0.008	
	Ours$^+$(59914)	0.788 ± 0.05	$\mathbf{0.746} \pm 0.034$	$\mathbf{0.877} \pm 0.008$	$\mathbf{0.909} \pm 0.006$	
Accuracies						
$I : \mathbb{R}^3 \to \mathbb{R}^N$						
$\mathbb{T}^3$	Classical$^-$(13539)	0.792 ± 0.08	0.415 ± 0.053	$\mathbf{0.879} \pm 0.024$	0.789 ± 0.034	0.806 ± 0.017
	Classical$^+$(972694)	0.815 ± 0.061	0.702 ± 0.026	0.834 ± 0.022	0.89 ± 0.011	0.854 ± 0.012
$I : \mathbb{R}^3 \times \mathbb{S}^2 \to \mathbb{R}$						
$SO(3)$	Baseline$^-$(286)	0.742 ± 0.082	0.145 ± 0.04	0.804 ± 0.024	0.85 ± 0.016	0.788 ± 0.011
	Baseline$^+$(2104)	0.778 ± 0.07	0.379 ± 0.065	0.784 ± 0.024	0.848 ± 0.02	0.792 ± 0.013
$SE(3)$	Ours$^-$(2514)	0.81 ± 0.065	0.692 ± 0.029	0.857 ± 0.022	0.874 ± 0.019	0.856 ± 0.01
	Ours$^+$(59914)	$\mathbf{0.896} \pm 0.042$	$\mathbf{0.826} \pm 0.023$	0.857 ± 0.017	$\mathbf{0.912} \pm 0.014$	$\mathbf{0.883} \pm 0.008$
Accuracies from rotated test set						
$\mathbb{T}^3$	Classical$^+$	0.632 ± 0.097	0.452 ± 0.02	0.434 ± 0.018	0.5 ± 0.03	0.471 ± 0.015
$SO(3)$	Baseline$^+$	0.769 ± 0.074	0.307 ± 0.059	0.782 ± 0.024	0.846 ± 0.02	0.786 ± 0.013
$SE(3)$	Ours$^+$	$\mathbf{0.88} \pm 0.048$	$\mathbf{0.659} \pm 0.028$	$\mathbf{0.83} \pm 0.019$	$\mathbf{0.868} \pm 0.018$	$\mathbf{0.84} \pm 0.009$

Classical$^+$ setup works well, but it is at the cost of much bigger model capacity. Additionally, Classical$^-$ is not significantly better than Baseline$^+$, even though it has a way larger capacity. Demonstrations of predictions from all models with high capacity can be found in fig. 4. Predictions from Ours$^+$ are much less noisy - especially for subcortical regions - than others.

In order to test the model resistance to variations, we rotated the test set using rotations randomly sampled from an octahedral rotation group; Accuracies from the rotated data with high capacity models are found at the bottom of table 2. Models with rotational group actions (Baseline$^+$ and Ours$^+$) are resistant to data variation, and Ours$^+$ remains the best in performance.

4. Conclusion

We proposed an $SE(3)$ GCNN for DWI scan segmentation by using a natural action of $SE(3)$ on the product space $\mathbb{R}^3 \times \mathbb{S}^2$, which models the space where DWI data is measured. As it is a homogeneous space for this action, we develop equivariant/invariant GCNNs for functions defined on it. This strategy keeps the required network capacity small, while mitigating the need for data augmentation, which is usually more expensive either in computation or in storage. Experiments show that our method is superior to ones that discard either the spatial symmetries on $\mathbb{R}^3$ or the rotational symmetries on $\mathbb{S}^2$. Additionally, tested with rotated data, models with rotational group actions demonstrate again the impact of equivariance, especially for our $SE(3)$-based model.

References

V. Andrearczyk, J. Fageot, and A. Depeursinge. Local Rotation Invariance in 3D CNNs. *Medical Image Analysis*, 65, 2020.

J. Aronsson. Homogeneous Vector Bundles and $\mathcal{G}$-Equivariant Convolutional Neural Networks, 2021.

M. Banerjee, R. Chakraborty, D. Archer, D. Vaillancourt, and B. C. Vemuri. DMR-CNN: A CNN Tailored for DMR Scans with Applications to PD Calssifiaction. In *Proceedings of International Symposium on Biomedical Imaging*, 2019.

P.J. Basser, J. Mattiello, and D. LeBihan. MR Diffusion Tensor Spectroscopy and Imaging. *Biophys. J.*, 66(1):259–267, 1994.

E. J. Bekkers. B-Spline CNNs on Lie groups. In *International Conference on Learning Representations*, 2019.

E. J. Bekkers, M. Lafarge, M. Veta, K.A.J. Eppenhof, J.P.W. Pluim, and R. Duits. Roto-Translation Covariant Convolutional Networks for Medical Image Analysis. In *Proc. MICCAI 2018*, pages 440–448, 2018.

D. Boscaini, J. Masci, E. Rodolà, and M. Bronstein. Learning Shape Correspondence With Anisotropic Convolutional Neural Networks. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29, 2016.

E. Caruyer and R. Verma. On Facilitating the Use of HARDI in Population Studies by Creating Rotation-Invariant Markers. *Medical Image Analysis*, 20(1):87–96, 2015.

G. Cesa, L. Lang, and M. Weiler. A Program to Build E (N)-Equivariant Steerable CNNs. In *International Conference on Learning Representations*, 2021.

R. Chakraborty, M. Banerjee, and B.C. Vemuri. A CNN for Homogeneous Riemannian Manifolds with Application to NeuroImaging, 2018.

T. Cohen, M. Geiger, and M. Weller. A General Theory of Equivariant CNNs on Homogeneous Spaces. *Advances in Neural Information Processing Systems (NeurIPS 2019)*, 32: 9142–9153, 2020.

T.S. Cohen and M. Welling. Group equivariant convolutional neural networks. In *Int. Conf. Machine Learning*, pages 2990–2999, 2016.

T.S. Cohen, M. Geiger, J. Köhler, and M. Welling. Spherical CNNs. In *International Conference on Learning Representations*, 2018.

J.R. Driscoll and D.M. Healy. Computing Fourier Transforms and Convolutions on the 2-Sphere. *Adv. Appl. Math*, 15(2), 1994.

R. Gens and P.M. Domingos. Deep Symmetry networks. In *NIPS*, pages 2537—2545, 2014.

V. Golkov, A. Dosovitskit, J I. Sperl, M. I. Menzel, M. Czisch, P. Särmann, T. Brox, and D. Cremers. q-Space Deep Learning: Twelve-Fold Shorter and Model-Free Diffusion MRI Scans. *IEEE Trans. Med. Im.*, 35(5):1344–1351, 2016.

P. E. Jupp and K. V. Mardia. A Unified View of the Theory of Directional Statistics, 1975-1988. *International Statistical Review / Revue Internationale de Statistique*, 57(3): 261–294, 1989.

D. M. Knigge, D. W. Romero, and E. J Bekkers. Exploiting redundancy: Separable Group Convolutional Networks on Lie Groups. In *International Conference on Machine Learning*, pages 11359–11386. PMLR, 2022.

R. Kondor and S. Trivedi. On the Generalization of Equivariance and Convolution in Neural Networks to the Action of Compact Groups. In *Proc. ICML*, pages 2747–2755, 2018.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal Loss for Dense Object Detection, 2018.

R. Liu, F. Lauze, K. Erleben, and S. Darkner. Bundle Geodesic Convolutional Neural Network for DWI Segmentation from Single Scan Learning. In Suheyla Cetin-Karayumak, Daan Christiaens, Matteo Figini, Pamela Guevara, Noemi Gyori, Vishwesh Nath, and Tomasz Pieciak, editors, *Computational Diffusion MRI*, pages 121–132, Cham, 2021. Springer International Publishing. ISBN 978-3-030-87615-9.

J. Masci, D. Boscaini, M. Bronstein, and P. Vandergheynst. Geodesic Convolutional Neural Networks on Riemannian Manifolds. In *Proceeding of 3dRR*, 2015.

P. Müller, V. Golkov, V. Tomassini, and D. Cremers. Rotation-Equivariant Deep Learning for Diffusion MRI, 2021.

D.S. Novikov, J. Veraart, I.O. Jelescu, and E. Fieremans. Rotationally-Invariant Mapping of Scalar and Orientational Metrics of Neuronal Microstructure with Diffusion MRI. *NeuroImage*, 174:518–538, 2018.

E. Schwab, H. E. Cetingül, B. Asfari, and E. Vidal. Rotational Invariant Features for HARDI. In *Proc. IPMI*, 2013.

S. Sedlar, T. Papadopoulo, R. Deriche, and S. Deslauriers-Gauthier. Diffusion MRI fiber Orientation Distribution Function Estimation using Voxel-wise Spherical U-net. In *International MICCAI Workshop 2020 - Computational Diffusion MRI*, Lima, Peru, October 2020. URL https://hal.archives-ouvertes.fr/hal-02946371.

S. Sedlar, T. Alimi, A.and Papadopoulo, R. Deriche, and S. Deslauriers-Gauthier. A Spherical Convolutional Neural Network for White Matter Structure Imaging via dMRI. In M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pages 529–539. Springer, 2021.

T.S. Cohen and M. Weiler and B. Kicanaoglu and M. Welling. Gauge Equivariant Convolutional Networks and the Icosahedral CNN. In *Proc. ICML*, pages 1321–1330, 2019.

D. S. Tuchs. Q-Ball Imaging. *Magnetic Resonance in Medicine*, 52:1358–1372, 2004.

David C. Van Essen, Stephen M. Smith, Deanna M. Barch, Timothy E.J. Behrens, Essa Yacoub, and Kamil Ugurbil. The WU-Minn Human Connectome Project: An Overview. *NeuroImage*, 80:62 – 79, 2013.

M. Weiler, M. Geiger, M. Welling, W. Boomsma, and T.S. Cohen. 3D Steerable CNNs: Learning Rotationally Equivariant Features in Volumetric Data. In *Proc. NIPS*, 2018.

D.E. Worrall, S.J. Garbin, D. Turmukhambetov, and G.J. Brostow. Harmonic Networks: Deep Translation and Rotation Equivariance, 2017.

Mauro Zucchelli, Samuel Deslauriers-Gauthier, and Rachid Deriche. A Computational Framework for Generating Rotation Invariant Features and its Application in Diffusion MRI. *Medical Image Analysis*, 60, 2020.