

Rotation-Equivariant Deep Learning for Diffusion MRI

Philip Müller¹, Vladimir Golkov¹, Valentina Tomassini², and Daniel Cremers¹

¹Computer Vision Group, Technical University of Munich, Munich, Germany, ²D'Annunzio University, Chieti-Pescara, Italy

Synopsis

Convolutional networks are successful, but have recently been outperformed by new neural networks that are equivariant under rotations and translations. These new networks do not struggle with learning each possible orientation of each image feature separately. So far, they have been proposed for 2D and 3D data. Here we generalize them to 6D diffusion MRI data, ensuring joint equivariance under 3D roto-translations in image space and the matching rotations in q -space, as dictated by the image formation. We validate our method on multiple-sclerosis lesion segmentation. Our proposed neural networks yield better results and require less training data.

Introduction

Convolutional networks are very successful in medical imaging because they are translation-equivariant, i.e. they detect features well, regardless of their translation (location). Recently, neural networks that are equivariant under 3D rotations and translations (i.e. the group $SE(3)$) were proposed and proved even more successful^{1,2}. Learning the many possible orientations of a feature (e.g. a neural fiber) is more complicated than generalizing automatically to all orientations via $SE(3)$ -equivariance. In diffusion MRI (dMRI), deep learning is highly beneficial³ and $SE(3)$ -equivariance is appropriate because neural fibers can have a large variety of orientations. So far, $SE(3)$ -equivariant networks have been proposed only for 3D data. Here we propose $SE(3)$ -equivariant networks for 6D dMRI data and demonstrate their benefits.

Methods

An $SE(3)$ -equivariant network layer was proposed for 3D point clouds¹ and 3D images². Both approaches use $SE(3)$ -equivariant filters built as weighted sums from pre-defined basis filters that use spherical harmonics as their angular part, and a radial basis. Feature maps are so-called spherical-tensor fields of possibly different orders for each channel^{1,2}. These spherical tensors are combined with the filter using Clebsch–Gordan coefficients.

A roto-translation of an object in the scanner causes a roto-translation of the image in position space (p -space) and rotation in q -space. We propose a network layer that respects this dMRI-inherent equivariance. We generalize prior work^{1,2} from 3D data to the 6D space of dMRI scans.

As q -space is not translated, we multiply two radial bases: on the input and output q -vector lengths, respectively. For the angular part, we use the spherical harmonics twice: once applied to p - and once to q -space coordinate offsets. Both are combined into the angular part of the basis filters using Clebsch–Gordan coefficients. This leads to the following definition of our proposed layer \mathcal{L} :

$$\begin{aligned} \mathcal{L}_{m_{out}}^{(c_{out})}[\mathbf{I}](\mathbf{p}_{out}, \mathbf{q}_{out}) &= \sum_{c_{in}, l_{filter}, l_p, l_q, k_1, k_2, k_3} \mathbf{W}_{c_{in}, c_{out}, k_1, k_2, k_3}^{(l_{filter}, l_p, l_q)} \sum_{\substack{m_{filter} \in \{-l_{filter}, \dots, l_{filter}\}, \\ m_{in} \in \{-l_{in}, \dots, l_{in}\}}} \mathbf{C}_{(l_{filter}, m_{filter})(l_{in}, m_{in})}^{(l_{out}, m_{out})} \\ &\times \sum_{\substack{\mathbf{p}_{in} \in \mathbb{R}^3, \\ \mathbf{q}_{in} \in \mathbb{R}^3}} \varphi_1^{(k_1)}(\|\mathbf{p}_{out} - \mathbf{p}_{in}\|_2) \varphi_2^{(k_2)}(\|\mathbf{q}_{out}\|_2) \varphi_3^{(k_3)}(\|\mathbf{q}_{in}\|_2) \\ &\times \sum_{\substack{m_p \in \{-l_p, \dots, l_p\}, \\ m_q \in \{-l_q, \dots, l_q\}}} \mathbf{C}_{(l_p, m_p)(l_q, m_q)}^{(l_{filter}, m_{filter})} \mathbf{Y}_{m_p}^{(l_p)} \left(\frac{\mathbf{p}_{out} - \mathbf{p}_{in}}{\|\mathbf{p}_{out} - \mathbf{p}_{in}\|_2} \right) \mathbf{Y}_{m_q}^{(l_q)} \left(\frac{\mathbf{q}_{out} - \mathbf{q}_{in}}{\|\mathbf{q}_{out} - \mathbf{q}_{in}\|_2} \right) \mathbf{I}_{m_{in}}^{(c_{in})}(\mathbf{p}_{in}, \mathbf{q}_{in}), \end{aligned}$$

where \mathbf{I} denotes the input feature map, \mathbf{p}_{out} and \mathbf{q}_{out} are p - and q -space coordinates in the output and \mathbf{p}_{in} , \mathbf{q}_{in} in the input feature map, c_{out} is the index of one of the output channels, l_{out} is shorthand for $l_{out}(c_{out})$, i.e. the order of the output channel c_{out} , m_{out} indices the components of the output tensor of channel c_{out} , with $-l_{out} \leq m_{out} \leq l_{out}$, the index c_{in} goes over all input channels, l_{in} is shorthand for $l_{in}(c_{in})$, i.e. the order of the input channel c_{in} , with $-l_{in} \leq m_{in} \leq l_{in}$, l_{filter} is the filter order (frequency index) used to index the angular filter basis, with $|l_{out} - l_{in}| \leq l_{filter} \leq (l_{out} + l_{in})$, l_p and l_q are the orders of the p - and q -space parts of the filter, with $|l_p - l_q| \leq l_{filter} \leq (l_p + l_q)$, and $|l_{filter} - l_p| \leq 1$, k_1, k_2, k_3 are indices of the radial bases for p -space coordinate offsets and q -space output and input coordinates, \mathbf{W} are learned weights, \mathbf{C} are the (real) Clebsch–Gordan coefficients, $\varphi_1^{(k_1)} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$, $\varphi_2^{(k_2)} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$, $\varphi_3^{(k_3)} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ are sets of radial bases, \mathbf{Y} are the (real) spherical harmonics.

The derivation and proof of equivariance are a generalization of prior work^{1,2} to 6D dMRI data.

Note that the only learnable parameters are \mathbf{W} and optional parameters in $\varphi_1^{(k_1)}$, $\varphi_2^{(k_2)}$, and $\varphi_3^{(k_3)}$. The proposed layer is not the only option how a generalization of the equivariant 3D layer to dMRI is possible, we also investigated applying the spherical harmonics only once to the difference of coordinate offsets from p - and q -space but found it to be less effective.

The effectivity of the proposed layer is studied by doing segmentation of multiple sclerosis (MS) lesions using a dataset⁴ containing dMRI brain scans with ground-truth annotations of MS lesions. Our architecture uses five of the proposed equivariant layers. Various hyperparameters yielded similar results. Each of the layers is followed by a gated nonlinearity¹ and swish⁵ for scalar channels, except for the last layer, which uses sigmoid.

Results & Discussion

Fig. 1 shows the segmentation of six scans from the validation set (not used for training) with ground truth, predictions from our model, and a non-rotation-equivariant reference model that uses 3D convolutional layers, and shows their receiver operating characteristic (ROC) and precision-recall curves. Our model achieves 0.982 area under the ROC curve (AUC), 0.630 average precision (AvgPrec), and 0.592 Dice score. Figs. 2-3 show how the models are affected when the training dataset size is reduced.

While already the non-rotation-equivariant model roughly predicts the ground truth, our model predicts it more accurately. It can especially be seen that while our model predicts the MS lesions very confidently and reports very small values outside the areas around the lesions, the non-rotation-equivariant model is very uncertain at many positions. Our model outperforms the non-rotation-equivariant model by 3% ROC AUC, 39% AvgPrec, and 22% Dice score. Also, our model requires four times fewer training samples than the non-rotation-equivariant model to achieve comparable quality.

Conclusions

Our results show that the proposed layer can increase the performance through better generalization and decrease the required number of training samples on dMRI datasets. The equivariance of the layer allows the use of many parameters that can effectively capture the essence of the dataset so that the model does not underfit while still restricting it so that overfitting is effectively reduced without using additional regularization. Thus, the method likely improves results on other dMRI datasets.

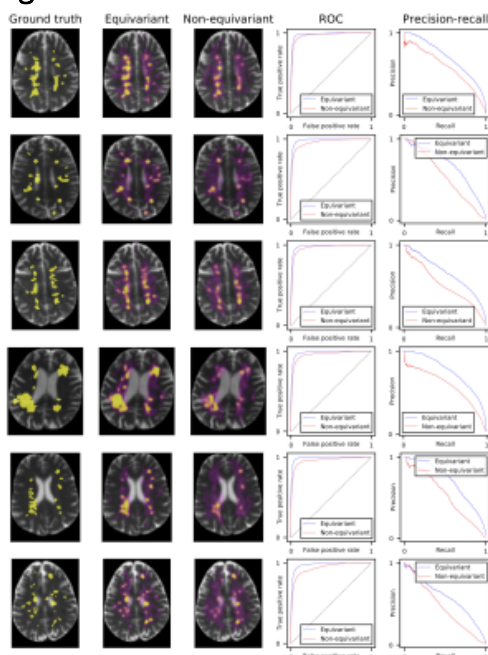
Acknowledgements

This work was supported by the Munich Center for Machine Learning (Grant No. 01IS18036B) and the BMBF project MLwin.

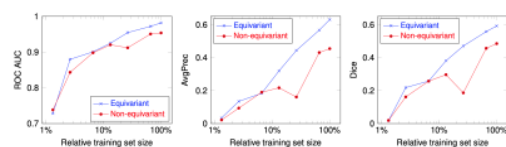
References

1. M. Weiler, M. Geiger, M. Welling, et al. "3D Steerable CNNs: Learning Rotationally Equivariant Features in Volumetric Data". Advances in Neural Information Processing Systems 31 (NIPS 2018). 2018;10381–10392.
2. N. Thomas, T. Smidt, S. M. Kearnes, et al. "Tensor Field Networks: Rotation- and Translation-Equivariant Neural Networks for 3D Point Clouds". 2018. arXiv:1802.08219.
3. V. Golkov, A. Dosovitskiy, J. I. Sperl, et al. "q-Space Deep Learning for Twelve-Fold Shorter and Model-Free Diffusion MRI Scans". MICCAI 2015, pp 37-44.
4. I. Lipp, C. Foster, R. Stickland, et al. "Predictors of training-related improvement in visuomotor performance in patients with multiple sclerosis: A behavioural and MRI study". Mult Scler. 2020;1352458520943788.
5. P. Ramachandran, B. Zoph, Q. V. Le. "Searching for activation functions". 2017. arXiv:1710.05941.

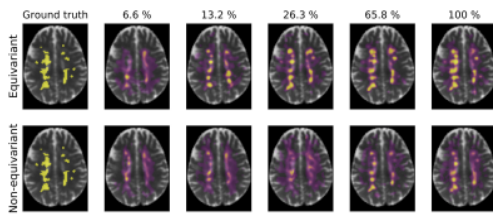
Figures



Segmentation of multiple-sclerosis lesions in six scans from the validation set. (a) Ground truth of one example slice per scan, (b) predictions for that slice using our equivariant model, (c) predictions for that slice using the non-rotation-equivariant reference model with 3D convolutional layers, (e) ROC curves of all models on the full scans, (f) precision-recall curves of all models on the full scans. While our equivariant model is very certain (yellow areas) at most positions, the non-rotation-equivariant model has large areas of high uncertainty (purple areas).



Comparison of our equivariant model (blue) with the non-rotation-equivariant model (red) both trained on reduced subsets of the training set. The plots show ROC AUC (left), AvgPrec (middle), and Dice scores (right). Our equivariant model trained on one quarter of the scans achieves similar results to the non-rotation-equivariant model trained on the full training dataset. When trained on two thirds of the training scans, our equivariant model outperforms the non-rotation-equivariant model by 1.9% in ROC AUC score, by 24.7% in AvgPrec score, and by 15.0% in Dice score.



Segmentation of multiple-sclerosis lesions in a scan from the validation set (not used for training) using our equivariant (top) and the non-rotation-equivariant (bottom) model, both trained on reduced subsets of the training set (from left to right) with the ground truth segmentation in the left column. While our equivariant model achieves quite accurate segmentations when trained on 26.3% of the training scans, the segmentations of the non-rotation-equivariant model only start getting accurate when trained on 65.8% of the set, indicating better generalization of our model.