Reconstruction of Diffusion Anisotropies using 3D Deep Convolutional Neural Networks in Diffusion Imaging

Simon Koppers, Matthias Friedrichs and Dorit Merhof

Abstract The reconstruction of neural pathways is a challenging problem in case of crossing or kissing neuronal fibers. High angular resolution diffusion imaging models are required to identify multiple fiber orientations in a voxel. Disadvantage of those models is that they require a multitude of acquired gradient directions, otherwise these models become inaccurate. We present a new approach to derive the fiber orientation distribution function using a Deep Convolutional Neural Network, which remains stable, even if less gradient directions are acquired. In addition, the Convolutional Neural Network is able to improve the signal in a voxel by extracting useful information of surrounding neighboring voxels. Subsequently, the functionality of the network is evaluated using 100 different brain datasets from the Human Connectome Project.

1 Introduction

Diffusion-weighted MRI is able to provide subject specific information about the course and location of white matter tracts in the human brain. In cases of kissing or crossing fibers, which occur in 60% to 90% of human white matter [5], high angular resolution diffusion imaging (HARDI) is needed to better represent such complex structures. For this purpose many different gradient directions are typically

Simon Koppers
Institute of Imaging & Computer Vision, RWTH Aachen University
e-mail: Simon.Koppers@lfb.rwth-aachen.de

Matthias Friedrichs
Institute of Imaging & Computer Vision, RWTH Aachen University
e-mail: Matthias.Friedrichs@rwth-aachen.de

Dorit Merhof
Institute of Imaging & Computer Vision, RWTH Aachen University
e-mail: Dorit.Merhof@lfb.rwth-aachen.de
acquired, which linearly increases the acquisition time, resulting in an acquisition time of several minutes up to hours. However, for clinical application it is necessary to reduce the number of required gradient directions to keep the acquisition time in a feasible range. As a downside, this reduction will result in a blurred reconstruction, if standard reconstruction applications are applied.

While Neural Networks have been known for decades, Deep Neural Networks have gained interest recently in the field of Diffusion Imaging [4, 7, 6]. It was shown that Deep Learning is able to stabilize state-of-the-art approaches, if only a few gradient directions are acquired. Nevertheless, none of these approaches is so far able to directly reconstruct the whole fiber orientation distribution function (fODF). Furthermore, the Diffusion Signal was so far reconstructed in a voxel-wise manner, excluding additional neighboring information.

In this work, a novel way to reconstruct a non-quantized fODF by including neighboring information is presented. For this purpose, we decompose the spherical fODF utilizing well-known Spherical Harmonics (SH), which are able to represent spherical signals in a general and complete way. The resulting SH coefficients are fitted utilizing a Neural Network regression approach. In addition, previously excluded neighboring information are included using Deep Convolutional Neural Networks (CNN), which are able to extract local neighboring information.

2 Methods

In order to represent an fODF, a model-free and non-sparse representation of a spherical signal is required. For this purpose, SH are utilized to represent the spherical signal, which are fitted utilizing a Deep Regression CNN. In the first part of this section, the material, which is used in this work, is described. The second part describes the utilized CNN.

2.1 Training Data and Labels

In order to evaluate the performance of our algorithm, a subset of 100 uncorrelated healthy brain scans from the Human Connectome Project is utilized. All scans are acquired with a 3T Siemens Connectome Skyra MRI scanner, a resolution of 1.25 × 1.25 × 1.25 mm² and 288 diffusion gradients, comprising 18 diffusion gradients with $b = 0 \frac{mm^2}{s}$ and 3 shells comprising 90 diffusion gradients each, at $b = 1000 \frac{mm^2}{s}$, $b = 2000 \frac{mm^2}{s}$, $b = 3000 \frac{mm^2}{s}$, respectively. Because of its high resolution (145 × 174 × 145 with 288 gradients directions each), the data is further reduced by selecting 5000 non-isotropic voxel neighborhoods (3 × 3 × 3) from each subject, resulting in 500,000 unique sets. In addition, only the third shell ($b = 3000 \frac{mm^2}{s}$) is utilized.

The CNN is trained on 80 uncorrelated scans (400,000 samples), while validation is performed on the remaining 20 subjects (100,000 samples). To show the impact
of a reduced number of gradients, three additional datasets are generated. For this purpose, the diffusion signal is equidistantly resampled for 15, 30 and 45 gradient directions utilizing the well-known SH.

### 2.1.1 Neighborhood

Due to the fact that additional information can be gained utilizing the signals’ neighborhood [2], three different neighborhoods are compared. The first neighborhood consists out of 6 additional signals, which are marked in green in Figure 1. The second neighborhood adds the blue-marked voxels to the first neighborhood resulting in 18 additional voxels. The biggest neighborhood includes 26 additional signals, which is represented by the surrounding cube (see Fig. 1).

![3D voxel neighborhood](image)

Fig. 1: 3D voxel neighborhood, which is represented as a hierarchy of three types of neighborhoods (green; green + blue; green + blue + red).

### 2.1.2 Gold standard

In order to generate a valid label for training and comparison, the well-known and state-of-the-art constrained spherical deconvolution [12, 11] (CSD) is utilized. The idea of CSD is that the diffusion signal can be decomposed, if its underlying single-fiber response function is known. Based on this response function, the diffusion signal can be deconvoluted in order to generate the fODF.

The CSD is applied on the raw signal as well as on its resampled versions containing 45, 30 and 15 gradient directions, respectively. To be comparable and stable for each set of gradient directions, the reconstruction order is set to 4 to be stable. Its output is a sampled fODF, which is described utilizing SH of order 8, resulting in 45 coefficients, which are used as label for training.
Since CSD is unable to directly define the resulting fiber direction, the Ball-and-Stick [10] (BS) model is utilized to calculate a relative ground truth fiber direction. It is a well-known model and one of the most accurate approaches to identify the correct fiber direction [8], if the correct number of fibers is known.

2.2 Deep Regression Convolutional Neural Network

Assuming the fODF to be sufficiently represented by SH coefficients, a regression approach is required to fit these coefficients, which is based on Deep Learning. In this work a Deep Regression CNN is utilized to fit the SH coefficients. CNNs are a special case of neural networks, in which the weights between neurons are combined in a matrix. This matrix can be interpreted as a convolutional kernel, which can be applied to 1D, 2D and also higher dimensional signals. In this work, additional information about the fODF are gained by combining 3D neighboring signals, which is why a 3D CNN is utilized.

In addition, each neural network contains an activation function. We choose the Rectified Linear Unit (ReLU), because of its good convergence behavior, with

\[ Y = \max(0, x) \]  

which is applied after each convolutional layer, with \( x \) as input and \( Y \) as output signal.

A combination of these two layers is further called a functional unit (FU), which is further specified by its convolutional kernel size, the number of kernels and its number of input channels. Due to the fact that every kernel leads to a new dataset, the number of input channels of the actual layer needs to be equivalent to the number of kernels of the previous layer. Furthermore, zero-padding with size \( p \) may be considered during convolution.

In most CNNs, several layers are stacked together, followed by a fully-connected layer that maps different features to the output. In the end, a cost function calculates a loss, which is utilized to train the network. For our CNN, a standard least square loss function is utilized, which is in general a good loss function for a regression problem. Training is performed using a stochastic gradient descend algorithm with error backpropagation.

2.3 The Net

The neural network is composed out of four FUs, while the last FU is fully connected to the output. The whole network is presented in Fig. 2. In order to be invariant to different gradient directions, the input and output signal are represented utilizing SH. Due to this fact, the number of input channels in the first layer and the
number of kernels in the last layer are defined by the number of expected coefficients.

For this CNN, the input signal is represented using an SH order of 4, resulting in 15 input coefficients, in order to be more robust to noise. The output represents the fODF utilizing the SH coefficients corresponding to an order of 8.

Fig. 2: The 3D CNN, which is utilized to reconstruct the fODF based on a diffusion signal and its neighborhood.
3 Experimental Results

The CNN is implemented in Python based on the TensorFlow framework [1], while the CSD is implemented in Python using Dipy [3].

Training is performed using stochastic gradient descend with a learning rate of \( \eta = 0.01 \) and a batchsize of 100.

In order to validate the performance of the CNN, two different error quantities are evaluated. On the one hand, the surface deviation of the fODF is calculated, which is an important quantity for probabilistic tractography approaches and diffusion characteristics. It is evaluated using the normalized mean square error [6] with

\[
NMSE = \frac{\|S_{\text{true}} - S_{\text{rec}}\|_2^2}{\|S_{\text{true}}\|_2^2},
\]

(2)

which is common error metric for comparing surfaces. \( S_{\text{true}} \) defines the original signal vector for each gradient direction, while \( S_{\text{rec}} \) is the reconstructed signal vector.

Furthermore, the angular accuracy is evaluated, which is particularly important for generating a deterministic tractography. In order to evaluate the angular accuracy, 250 single-fiber validation voxels, 250 validation voxels with two fibers and 250 three fiber validation voxels are manually selected, based on their relative ground truth fODF, while BS defines the relative ground truth fiber direction.

The mean angular error (MAE) is calculated based on

\[
MAE = \frac{1}{k} \sum_{i=1}^{k} \alpha_i,
\]

(3)

where \( k \) defines the number of fibers in a voxel and \( \alpha_i \) the angular error of fiber \( i \).

3.1 fODF Reconstruction

In order to evaluate the reconstruction quality, each fODF is relatively compared to its gold standard fODF based on 90 gradient directions. The resulting qualities for the CSD and the CNN are presented in Table 1, which contains the resulting NMSE. In addition, it should be noted that the CNN is completely retrained for each gradient and neighborhood scenario.

The smallest error, about 1.65\%, occurs for a reconstruction based on 45 gradient directions utilizing the CSD, while the highest NMSE, about 32.81\%, is also observed for the CSD, when only 15 gradient directions are used for reconstruction. In any other case, the CSD gets outperformed by the CNN. Taking the neighborhood into account, it can be seen that there is a tendency that a neighborhood results in a lower NMSE, if less gradient directions are available. The lowest NMSE is achieved utilizing 18 neighbor voxels (green + blue).
Table 1: Impact of signal neighborhood on the surface deviation based on 15, 30 and 45 diffusion gradients with CNN as well as CSD.

<table>
<thead>
<tr>
<th>Method</th>
<th>fODF Deviation [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Neighborhood 15 Gradients</td>
</tr>
<tr>
<td>DL</td>
<td>green + blue + red</td>
</tr>
<tr>
<td></td>
<td>green + blue</td>
</tr>
<tr>
<td></td>
<td>green</td>
</tr>
<tr>
<td></td>
<td>none</td>
</tr>
<tr>
<td>CSD</td>
<td>-</td>
</tr>
</tbody>
</table>

In addition to the NMSE, Fig. 3 presents a randomly chosen region out of an axial slice. It contains the fODF for each voxel, with a big fiber bundle running from the upper left to the right. Moreover, there is a crossing of two fiber bundles within the box region. This region of interest is further evaluated in Fig. 4 for 15 and 90 gradient directions utilizing the CSD and the CNN for reconstruction. For evaluation purposes, no neighborhood is included for a fair comparison. Fig. 4b shows that the CNN is not able to identify weak fiber directions, which constitute only a small fraction to the remaining signal.

On the other hand, the fODF reconstructed by CSD gets worse as the number of gradient directions is reduced to 15. Due to this reduction of gradient directions,
3.2 Angular Accuracy

In order to evaluate the angular accuracy, each fODF is further processed using a local maximum finder to define a specific fiber direction. The resulting mean angular errors (MAE) of the regular CSD in comparison to the CNN with different neighborhoods and 15, 30 and 45 diffusion gradients are provided in Tab. 4, 3 and 2.

Each table includes the mean angular error for the single-fiber, two-fiber and three-fiber case. For 45 gradient directions (see Tab. 2) the smallest error is achieved utilizing CSD for single-fiber cases and three-fiber cases, while the CNN is slightly
Reconstruction of Diffusion Anisotropies

better for two-fiber cases. Considering Tab. 3 both algorithms achieve nearly the same results. If only 15 gradient directions are available, the CNN achieves the smallest angular error in each case, whereas the error of the CSD reconstruction is significantly increased.

In addition, taking the neighborhood into account, the same tendency as before can be observed. The angular error decreases for 15 and 30 gradient directions if a neighborhood is used to reconstruct the fODF, while it increases for 45 gradient directions. Once again, the 18-neighborhood (green + blue) achieves the best result.

Table 2: Impact of signal neighborhood on angular accuracy for 45 acquired diffusion gradients utilizing the CNN in comparison to CSD. The angular error represents the mean angular error for 1-Fiber, 2-Fiber and 3-Fiber voxels.

<table>
<thead>
<tr>
<th>Method</th>
<th>Neighborhood</th>
<th>Angular Error [°]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1 Fiber</td>
</tr>
<tr>
<td>DL</td>
<td>green + blue + red</td>
<td>1.86</td>
</tr>
<tr>
<td></td>
<td>green + blue</td>
<td>1.84</td>
</tr>
<tr>
<td></td>
<td>green</td>
<td>1.79</td>
</tr>
<tr>
<td></td>
<td>none</td>
<td>1.73</td>
</tr>
<tr>
<td>CSD</td>
<td>none</td>
<td>1.65</td>
</tr>
</tbody>
</table>

Table 3: Impact of signal neighborhood on angular accuracy for 30 acquired diffusion gradients utilizing the CNN in comparison to CSD. The angular error represents the mean angular error for 1-Fiber, 2-Fiber and 3-Fiber voxels.

<table>
<thead>
<tr>
<th>Method</th>
<th>Neighborhood</th>
<th>Angular Error [°]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1 Fiber</td>
</tr>
<tr>
<td>DL</td>
<td>green + blue + red</td>
<td>1.98</td>
</tr>
<tr>
<td></td>
<td>green + blue</td>
<td>1.96</td>
</tr>
<tr>
<td></td>
<td>green</td>
<td>2.08</td>
</tr>
<tr>
<td></td>
<td>none</td>
<td>1.91</td>
</tr>
<tr>
<td>CSD</td>
<td>none</td>
<td>1.90</td>
</tr>
</tbody>
</table>
Table 4: Impact of signal neighborhood on angular accuracy for 15 acquired diffusion gradients utilizing the CNN in comparison to CSD. The angular error represents the mean angular error for 1-Fiber, 2-Fiber and 3-Fiber voxels.

<table>
<thead>
<tr>
<th>Method</th>
<th>Neighborhood</th>
<th>Angular Error [◦]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 Fiber</td>
<td>2 Fibers</td>
</tr>
<tr>
<td>DL</td>
<td>green+blue+red</td>
<td>2.86</td>
</tr>
<tr>
<td></td>
<td>green+blue</td>
<td>2.87</td>
</tr>
<tr>
<td></td>
<td>green</td>
<td>2.91</td>
</tr>
<tr>
<td></td>
<td>none</td>
<td>3.43</td>
</tr>
<tr>
<td>CSD</td>
<td>none</td>
<td>3.63</td>
</tr>
</tbody>
</table>

4 Discussion

Considering the surface deviation, Tab. 1, Fig. 3 and Fig. 4 show that the CNN results in a higher NMSE for 45 gradient directions than the CSD, while it achieves a lower NMSE for 30 and 15 gradient directions. In addition, the performance of the CNN stays stable in comparison to the performance of the CSD, which decreases and nearly doubles the resulting NMSE in comparison to the CNN. Furthermore, this objective result is verified in Fig. 4d, which is visually very similar to the CSD label with 90 gradient directions in Fig. 4a. Moreover, Fig. 4c shows that the CSD approach is not able to correctly reconstruct crossings for 15 gradient directions.

Additionally, the performance of the CNN improves if neighboring voxels are included. Comparing the different neighborhoods shows that the second neighborhood (green+blue) achieves the best NMSE for 15 and 30 gradient directions, while any neighborhood results in a decreasing performance if 45 gradient directions are available. This coincides with the previously stated hypothesis, that additional information can be gained from neighboring voxels if only few measured gradient directions are available. In addition, it should be noted that the biggest neighborhood including the red voxels (see Fig. 1) decreases the performance in comparison to the second neighborhood (green + blue). This deterioration may be due to the fact that the distance between the main voxel and the included neighboring voxels increases, which results in a decreased correlation between the corresponding signals. In those cases, the CNN is not able to train the corresponding weights to zero, which results in a decreased performance because of included weighted noise. The same effect can be seen for more than 45 gradient directions.

The resulting mean angular error, presented in Tab. 2, 3 and 4, shows an effect similar to the surface deviation. While the angular error for the first and second fiber is on the same level for the CNN as well as for the CSD, the CNN is not able to identify the third fiber even for 45 gradient directions, resulting in a high angular error. The same effect can be seen in Fig. 4b, which shows that the CNN is mostly
unable to identify fibers with a small fraction compared to the remaining fibers in the voxel, even for 90 gradient directions. Nevertheless, the CNN outperforms the CSD for 30 and 15 gradient directions. Here, the CNN stays relatively stable, while the CSD becomes unstable, which may be due to the fact that it is only able to reconstruct the main fiber in a voxel.

As before for the surface deviation, the impact of neighboring signals is even more visible. The second neighborhood (green + blue) results in a lower angular error. Again, the gap between the CSD and the CNN increases as the number of gradients decreases. In addition, the performance of the biggest neighborhood still doesn’t result in an improvement.

The fact, that the CNN is in most cases not able to identify the correct third fiber direction can be explained as follows:

- First, the chosen SH order may be too low, which would result in an fODF blur, while fibers with small inter-fiber angles can’t be distinguished. On the other hand, an increasing SH order would result in a bigger output vector, which increases the complexity of the 3D CNN. Because of this, the 3D CNN may not be able to achieve a better performance than the empirically chosen 3D CNN.
- Secondly, the CNN is trained utilizing 80 different subjects for training. This results in a high variance, which results in a good generalization of the CNN, making it insensitive against head orientations and different diffusivities. However, this positive effect could turn into a disadvantage, because as the generalization of the CNN increases, it may get less sensitive to subtle changes of the diffusion signal. To face this issue, [9] showed that a synthetically generated dataset can be utilized for training in order to generate a unique CNN for a specific subject. This would result in a net, which is not blurred due to high variance.
- The last point is that the chosen label is based on the CSD applied on real human data. A disadvantage of this method is that noise, which is inevitably contained in each acquisition, leads to a noisy label. This noise would especially affect weak signals, which mostly occur for signals of the second or third fiber. Again, a synthetic dataset could be utilized for training, in order to obtain a noise free label.

Since not all MRI scanners provide the possibility to measure high b-values, the CNN is also evaluated using the same dataset and a b-value of $b = 1000 \text{ s mm}^2$. Utilizing this new dataset, the performance of the CNN as well as for the CSD decreases dramatically, making it unsuitable for clinical purpose. This may be, due to the reduced contrast of the diffusion signal within a voxel at a lower b-value.

### 5 Conclusion

The results of this work show that the proposed 3D CNN approach is able to reconstruct a quantization-free fODF and outperforms the CSD approach, if less than 45 gradient directions are available. In addition, we show that the CNN can be further
improved by including additional information from neighboring voxels. The best performance is reached by including the 18-neighborhood (green + blue in Fig. 1). For 15 gradient directions, this neighborhood improves the surface deviation from 18.89% to 13.12%, while the angular error improves from 3.43° to 2.87° for a single fiber case, from 10.87° to 6.80° for a two fiber case and from 18.06° to 16.01° in case of three crossing fibers in a voxel. On the other hand, we showed that including a neighborhood may lead to a decreased performance, if 45 and more gradient directions are available. In this case, the acquired signals contain enough information to reconstruct the fODF, while neighboring signals only add noise. Considering the applicability in a clinical scenario, it has to be taken into account that the trained CNN cannot be applied to data acquired with another scanner or scanner protocol, e.g. with different b-value.

In future work, we will investigate the behavior across different scanner types and acquisition protocols. Moreover, we will investigate a multi-shell approach, containing multiple b-values, in order to collect more information about the correct fODF from each voxel. In addition, we will investigate the impact of noise on the reconstruction.

Acknowledgements This work was supported by the International Research Training Group (IRTG 2150) of the German Research Foundation (DFG).

Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54 MH091657) 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.

References


