Autoencoder-based quality assessment for synthetic diffusion-MRI data

Leon Weninger[1], Maxim Drobjazko[1], Chuh-Hyoun Na[2], Kerstin Jütten[2], Dorit Merhof[1]

[1]Imaging and Computer Vision, RWTH Aachen University
[2]Department of Neurosurgery, University Hospital RWTH Aachen
leon.weninger@lfb.rwth-aachen.de

Abstract. Diffusion MRI makes it possible to assess brain microstructure in-vivo. Recently, a variety of deep learning methods have been proposed that enhance the quality and utility of these acquisitions. For deep learning methods, a large amount of training data is necessary, but difficult to obtain. As a solution, different approaches to synthetic data creation have been published, but it is unclear which approach produces data that best matches the in-vivo characteristics. Here, a methodology to assess the quality of synthetic diffusion data which is based on denoising autoencoders is proposed. For this, the reconstruction errors of autoencoders trained only on synthetic data were evaluated. The more the synthetic data resembles the real data, the lower the reconstruction error. Using this method, we evaluated which of four different synthetic data simulation techniques produced data that best resembled the in-vivo data. We find that modeling diffusion MRI data with patient- and scanner specific values leads to significantly better reconstruction results than using default diffusivity values, suggesting possible benefits of precision medicine approaches in diffusion MRI analysis.

1 Introduction

MRI scans are used as a primary method in detecting diseases such as traumas, brain aneurysms, strokes, and tumors. Currently, medical professionals analyze these scans visually. However, deep learning could assist, e.g. by automatic detection of diseases, as deep learning approaches can notice subtle differences and patterns in data. In clinical research, deep learning methods have been established in a variety of diffusion MRI applications [1].

Deep learning approaches need a high amount of data to train on. Since the number of available MRI scans is always limited and groundtruth data is scarce, synthetic data can help to alleviate this problem. Software-based diffusion phantoms have been in use since several years [2]. However, if deep learning approaches are to be trained on synthetic data, the synthetic data needs to match the characteristics of in-vivo data as closely as possible. If the generating

model of synthetic data only matches a subset of the characteristics of the in-vivo data, the trained algorithm will not be able to generalize to the complete range of in-vivo data. On the other hand, if the generating model induces too much variety, we hypothesize that a trained deep learning algorithm will perform suboptimally, as the neural network will lose specificity.

Different approaches for creating synthetic data exist, and it is not clear which one is optimally suited for deep learning models. The generating models need to be tested and compared in order to differentiate model qualities. We propose to assess the quality of synthetic diffusion data with an autoencoder, a network architecture originally designed for denoising [3]. Fitted on synthetic data and evaluated on real data, it can be used as an indicator for quality assessment. Better suited synthetic data will lead to better reconstruction performance on in-vivo data.

Our proposed autoencoder is self-adaptive to the shape of the input, i.e., to the number of diffusion directions, and was trained and evaluated on single-voxel data. Comparing the reconstruction performance on a local study dataset as well as on data from the Human Connectome Project (HCP), an open access dataset containing high quality diffusion MRI scans [4], the quality of 4 different synthetic diffusion data models was assessed.

2 Materials and Methods

2.1 Data

Two different datasets were used: The freely available HCP dataset (isotropic voxel size of 1.25mm, three-shell b=1000,2000,3000 s/mm^2, 90 gradient directions per shell), as well as a dataset containing diffusion MRI scans (isotropic voxel size of 2.4mm, single-shell b=1000 s/mm^2, 64 gradient directions) of 28 healthy subjects acquired at the University Hospital Aachen. From the HCP dataset, all subjects of the "100 unrelated subjects" dataset were selected, and only the b=1000 s/mm^2 shell data was utilized. For the locally acquired data, all subjects have given written informed consent. The scans were approved by the ethics committee of the Medical Faculty of Aachen University (EK 294/15), and acquired according to the standards of Good Clinical Practice and the Declaration of Helsinki. The diffusion data of the local study were corrected for susceptibility induced and eddy current distortions with tools from FSL [5]. On the accompanying T1 image, the tissue was segmented into white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) with FSL Fast, and the segmentation map was transformed in diffusion space using an affine registration of the T1 and diffusion image. The HCP data was already preprocessed, and tissue segmentation maps were readily available.

2.2 Synthetic Data Creation

Four different synthetic data generating methods [6,7,8,9], which were all previously used in deep learning settings by various groups, were evaluated. For every subject, individual synthetic data was created. Three of the data generating

techniques were multi-tensor simulations, for which only the settings for possible tensor eigenvalues differed. These multi-tensor simulations were made from up to three different single-fiber WM compartments with random main directions and random volume percentage, as well as possible GM and CSF compartments with a random compartment contribution from 0 to 100% each.

Syn The first method, fully synthetic, further referred to as "Syn", is based on plausible pre-determined values for all compartments. The default values of the Dipy [6] diffusion simulation toolbox were used for the WM tensor eigenvalues, i.e. an axial diffusivity of $0.0015\,\mathrm{s/mm^2}$, and a radial diffusivity of $0.0003\,\mathrm{s/mm^2}$. The diffusivity was set to $0.0005\,\mathrm{s/mm^2}$ for GM and to $0.003\,\mathrm{s/mm^2}$ for CSF.

Sampled Second, exemplary diffusion tensor eigenvalues were sampled from the subject in question using eroded tissue segmentation maps for all three tissues, which was similarly proposed in [7]. For WM, only single-fiber voxels, identified by an FA value larger than 0.7 were retained. Thus, the tensor eigenvalues follow the individual characteristics, and a variety of eigenvalues were employed. This method is further labelled "Sampled".

SampleAndMean Third, eigenvalues were sampled for WM, but for CSF and GM the mean diffusivity values obtained from the subject were used. This method is thus referred to as "SampleAndMean", and was originally proposed in [8]. Taking the mean diffusivity instead of sampling GM and CSF was proposed in order to reduce noise in the synthetic data, as the diffusion signal can be especially noisy in CSF due to low signal levels.

RandomWM Finally, [9] proposed to simulate the diffusivity of free water with $0.003\,\mathrm{s/mm^2}$, and the diffusion attenuation of other microstructure compartments with a random variable per gradient following a uniform distribution $\mathrm{U}(0, 1)$. This synthetic diffusion data was originally employed to train a model that predicts the water compartment in a voxel, it was thus not intended to be a holistic simulation of diffusion signals. Nevertheless, it could be that such a model is also useful for other tasks, e.g., for neural network pretraining, and is thus also compared as "RandomWM" in the experiments.

2.3 Denoising Autoencoder

With the denoising autoencoder, the diffusion data was reduced to a minimum size in the hidden layers, from which the original signal was reconstructed with the decoder part of the autoencoder. As single voxels were used as input, and a fully connected architecture was chosen, an automatically adjusting input size for different MRI scans was necessary. Two reducing and two increasing steps were chosen, and the reduction and increase of size was defined relatively to the previous layer size. The optimal reduction size was set to 1/8 of the input

size, as experimentally determined by training and evaluating on real data. The intermediate layer size was set to 4/6, creating an hourglass-like shape. A batch size of 100, ReLU activation functions, a mean square error and an Adam optimizer with a learning rate of 0.0005 were chosen. 300.000 voxels were created for training, split into 250.000 voxels for training and 50.000 for validation. The neural network had to be trained for 50 epochs in order to reach convergence.

3 Results

The reconstruction performance of the autoencoders trained on the different synthetic datasets was compared using the raw diffusion attenuated signal, as well as metrics derived from fitted diffusion tensors.

First, to assess the reconstruction performance, the diffusion attenuated signal was used, i.e., the diffusion signal divided by the b0 measurement. The mean absolute deviation of the reconstructed signal on the brain MRI scans is shown in Fig. 1. Both, in the HCP as well as in the study data, the reconstruction performance for Syn, SampleAndMean and Sample was nearly identical with no statistically significant differences, while the RandomWM mode performed significantly worse. Meanwhile, the reconstruction performance between the HCP acquisitions and the study data was different. This effect can be explained by the acquisition settings: The MRI scanners used in the HCP and local study were similar, but the HCP used voxel sizes of 1.25mm, while 2.4mm were used in the local study, i.e., one voxel in the local data corresponds to $(\frac{2.4}{1.25})^3 \approx 7$ voxels in the HCP data. Thus, ignoring differences due to other acquisition settings and the preprocessing pipeline, a difference in signal-to-noise ratio (SNR) of $\sqrt{(\frac{2.4}{1.25})^3} \approx 2.66$ was expected between the two datasets. This falls exactly in line with the difference in raw reconstruction performance: The mean absolute error for the three multi-tensor simulations was 0.014 on the local study data, and 0.037 for the HCP data.

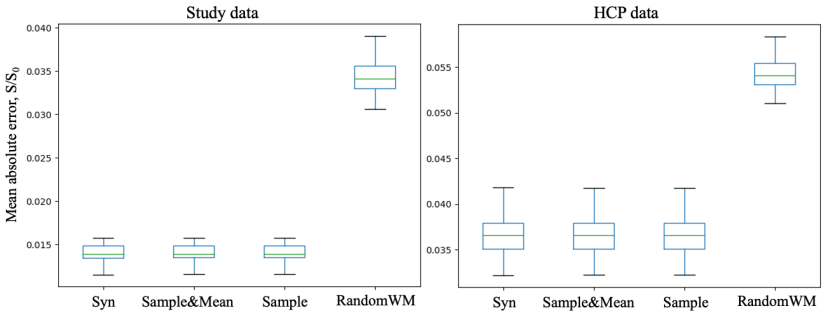

Fig. 1. Mean absolute error of the diffusion signal attenuation. Note the different y-axis between the two different datasets.

Second, the reconstruction performance on metrics derived from fitted diffusion tensors was assessed. The difference in fractional anisotropy (FA) for the

three multi-tensor modes Syn, SampleAndMean and Sample was negligible with a mean absolute error of 0.0075 for the three settings in local study data, and 0.0104 in the HCP data, while the RandomWM model was again significantly worse, with a mean absolute deviation of 0.17 on the local study data, and 0.22 on the HCP data.

However, significant differences between the multi-tensor modes could be observed for the reconstruction of the main fiber direction (Fig. 2). Especially, a strong difference between Syn and the WM sampling models could be observed, while the differences between the mean or sampling of GM and CSF diffusion attenuation did not affect the results as strongly. The performance of RandomWM (local study: $42.4 \pm 3.2°$, HCP: $42.7 \pm 2.8°$) was considerably worse than for the other techniques, it is not displayed for better visualization.

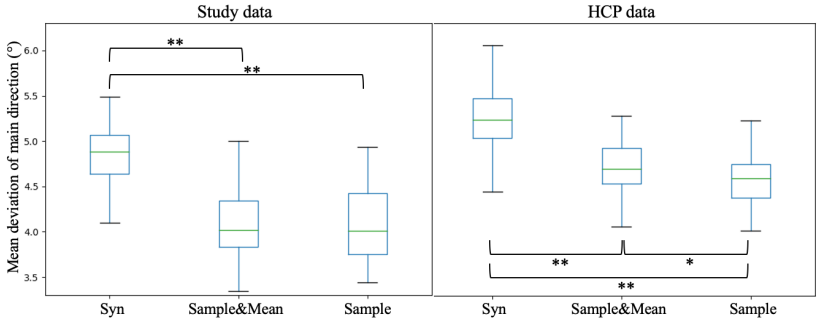

Fig. 2. Mean deviation of the main direction of the diffusion tensor in degree for all voxels with an fractional anisotropy (FA) value between 0.5 and 0.9. * statistically significant with $p<0.05$, ** significant with $p<10^{-8}$.

4 Discussion

In the analysis of raw signal reconstruction performance, the three different multi-tensor models had negligible difference, while the WM simulation with random values lead to worse results. This random value approach covers the largest variety in diffusion signals, but the autoencoder network, which needs to compress the data, is not able to distill the necessary information into the smaller latent space. Regardless of the task, modern deep learning approaches need to distill information at some point, which does not seem to be possible with a too unspecific data generation approach such as the method labeled RandomWM. Similar effects should occur when neural networks are trained on this synthetic data for other tasks, e.g. microstructure prediction or tractography.

Multi-tensor models seem to be a good choice: The matching of SNR difference between the two datasets and reconstruction performance is an indicator for close-to-optimal reconstruction performance of the employed denoising autoencoder. How the diffusion tensor eigenvalues should be chosen depends on the

application case: The simulation of multi-tensor data with pre-set tensor eigenvalues using openly available software is simple and fast, and the reconstruction performance on the raw signals and FA values is already good. Therefore, it is feasible to use such simulated data for quick and scanner-independent training or pretraining of neural networks, where the directionality of fibers is not relevant. If the application is dependent not only on raw signal values or FA, but instead on the direction of the WM fibers, the individual characteristics, depending on scanner and subject, should not be ignored. For example, in the application of deep learning to tractography [10], minor errors in fiber direction integrate over the whole fiber, possibly leading to major errors in the end region of the fiber. In such cases, a precision medicine approach is necessary: The diffusivity of WM and other compartments should be set to patient-specific values.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – grant 269953372/GRK2150 and grant ME 3737/19-1.

References

1. Ravi D, Ghavami N, Alexander DC, et al. Current Applications and Future Promises of Machine Learning in Diffusion MRI. MICCAI Workshop on Computational Diffusion MRI (CDMRI). 2019; p. 105–121.
2. Neher PF, Laun FB, Stieltjes B, et al. Fiberfox: Facilitating the creation of realistic white matter software phantoms. Magn Reson Med. 2014;72(5):1460–1470.
3. Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016.
4. Essen DCV, Smith SM, Barch DM, et al. The WU-Minn Human Connectome Project: An overview. NeuroImage. 2013;80:62–79. Mapping the Connectome.
5. Smith SM, Jenkinson M, Woolrich MW, et al. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage. 2004;23 Suppl. 1:S208–S219.
6. Garyfallidis E, Brett M, Amirbekian B, et al. Dipy, a library for the analysis of diffusion MRI data. Front Neuroinform. 2014;8:8.
7. Schultz T. Learning a Reliable Estimate of the Number of Fiber Directions in Diffusion MRI. Medical Image Computing and Computer-Assisted Intervention – MICCAI. 2012; p. 493–500.
8. Weninger L, Koppers S, Na CH, et al. Free-Water Correction in Diffusion MRI: A Reliable and Robust Learning Approach. MICCAI Workshop on Computational Diffusion MRI (CDMRI). 2019; p. 91–99.
9. Molina-Romero M, Wiestler B, Gómez P, et al. Deep Learning with Synthetic Diffusion MRI Data for Free-Water Elimination in Glioblastoma Cases. Medical Image Computing and Computer Assisted Intervention – MICCAI. 2018; p. 98–106.
10. Poulin P, Cote MA, Houde JC, et al. Learn to Track: Deep Learning for Tractography. Medical Image Computing and Computer Assisted Intervention - MICCAI. 2017; p. 540–547.