



Published in final edited form as:

Neuroimage. 2022 June ; 253: 119033. doi:10.1016/j.neuroimage.2022.119033.

SDnDTI: Self-supervised deep learning-based denoising for diffusion tensor MRI

Qiyuan Tian^{1,2,*}, Ziyu Li³, Qiuyun Fan^{1,2}, Jonathan R. Polimeni^{1,2,4}, Berkin Bilgic^{1,2,4}, David H. Salat^{1,2}, Susie Y. Huang^{1,2,4}

¹Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts, United States

²Department of Radiology, Harvard Medical School, Boston, Massachusetts, United States

³Department of Biomedical Engineering, Tsinghua University, Beijing, P. R. China

⁴Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States.

Abstract

Diffusion tensor magnetic resonance imaging (DTI) is a widely adopted neuroimaging method for the in vivo mapping of brain tissue microstructure and white matter tracts. Nonetheless, the noise in the diffusion-weighted images (DWIs) decreases the accuracy and precision of DTI derived microstructural parameters and leads to prolonged acquisition time for achieving improved signal-to-noise ratio (SNR). Deep learning-based image denoising using convolutional neural networks (CNNs) has superior performance but often requires additional high-SNR data for supervising the training of CNNs, which reduces the feasibility of supervised learning-based denoising in practice. In this work, we develop a self-supervised deep learning-based method entitled “SDnDTI” for denoising DTI data, which does not require additional high-SNR data for training. Specifically, SDnDTI divides multi-directional DTI data into many subsets of six DWI volumes and transforms DWIs from each subset to along the same diffusion-encoding directions through the diffusion tensor model, generating multiple repetitions of DWIs with identical image contrasts but different noise observations. SDnDTI removes noise by first denoising each repetition of DWIs using a deep 3-dimensional CNN with the average of all repetitions with higher SNR as the training target, following the same approach as normal supervised learning based denoising methods, and then averaging CNN-denoised images for achieving higher SNR. The denoising efficacy of SDnDTI is demonstrated in terms of the similarity of output images and resultant DTI metrics compared to the ground truth generated using substantially more DWI volumes on two datasets with different spatial resolution, b-values and number of input DWI volumes provided by the Human

*Correspondence to: Qiyuan Tian, Ph.D., Athinoula A. Martinos Center for Biomedical Imaging, 149 13th Street, Charlestown, MA, 02129, United States. qtian@mgh.harvard.edu.

Code availability

The source codes of BM4D implemented using MATLAB are publicly available (<https://www.cs.tut.fi/~foi/GCF-BM3D>). The MATLAB-based software of AONLM is publicly available (<https://sites.google.com/site/pierrickcoupe/software/denoising-for-medical-imaging/mri-denoising/mri-denoising-software>). The source codes of MPPCA implemented using MATLAB are publicly available (https://github.com/NYU-DiffusionMRI/mppca_denoise). The source codes of SDnDTI implemented using MATLAB and Keras application programming interface will be made publicly available (<https://github.com/qiyuantian/SDnDTI>).

Connectome Project (HCP) and the Lifespan HCP in Aging. The SDnDTI results preserve image sharpness and textural details and substantially improve upon those from the raw data. The results of SDnDTI are comparable to those from supervised learning-based denoising and outperform those from state-of-the-art conventional denoising algorithms including BM4D, AONLM and MPPCA. By leveraging domain knowledge of diffusion MRI physics, SDnDTI makes it easier to use CNN-based denoising methods in practice and has the potential to benefit a wider range of research and clinical applications that require accelerated DTI acquisition and high-quality DTI data for mapping of tissue microstructure, fiber tracts and structural connectivity in the living human brain.

Keywords

convolutional neural network; supervised learning; residual learning; image synthesis; diffusion tensor transformation; normal aging

Introduction

Diffusion tensor magnetic resonance imaging (DTI)^{1–5} is a neuroimaging method that noninvasively maps the brain tissue properties and white matter tracts in the in vivo human brain, which has a wide range of applications in research and clinical studies. DTI has proven to be a useful tool for characterizing and monitoring microstructural changes related to development^{6,7}, normal aging^{6,8}, neurodegeneration⁹, plasticity¹⁰ and a number of neurological^{11,12} and psychiatric^{13,14} disorders. In addition to investigating white matter, high-resolution DTI has also been used for mapping human cerebral cortical gray matter anisotropy and microstructure^{15–19}. For these reasons, DTI is an essential imaging modality adopted in many large-scale neuroimaging studies such as the Alzheimer’s Disease Neuroimaging Initiative^{20,21}, the Parkinson Progression Marker Initiative²², and the UK Biobank Imaging Study²³.

Nonetheless, relatively long acquisition times pose a barrier to performing high-quality DTI. Since diffusion-weighted MRI creates image contrast by attenuating the MR signal based on how readily water molecules diffuse, diffusion-weighted images (DWIs) are noisy, especially in acquisitions using strong diffusion encoding (i.e., high b-values) or high spatial resolution. Consequently, DTI often needs several times more measurements (e.g., 20 DWI volumes, 5–10 seconds per volume) than the theoretical minimum of 6 DWI volumes to derive the six unique elements of the tensor. For DTI performed at high (1.25 to 1 mm isotropic) or ultra-high (sub-millimeter isotropic) spatial resolution, the required number of measurements may be far more than 20, for example, in mapping cortical microstructure¹⁵.

Image denoising provides a feasible alternative to improve the quality of DTI from a shorter scan. Image denoising aims to recover a clean image with high signal-to-noise ratio (SNR) from noise-degraded observations, which is a highly ill-posed inverse problem. In the computer vision field, numerous denoising algorithms have been proposed to remove noise from natural and biomedical images such as non-local means (NLM) filtering²⁴, block-matching and 3-dimensional filtering (BM3D)²⁵ and K-singular value decomposition (K-SVD) denoising²⁶ and their 3-dimensional extensions for volumetric data^{27–31}. Many of

these algorithms are able to deal with spatially varying non-Gaussian noise and therefore can be readily applied to denoise diffusion MRI data. MRI reconstruction methods that often regularize the image formation process using prior knowledge, such as sparseness^{32–36} and low rank^{37–43}, can also achieve denoising effects. Many image restoration methods are also designed to exploit the additional redundant information originating from multiple diffusion encodings along various directions for increased denoising efficacy. A representative example is the widely adopted Marchenko–Pastur principal component analysis (MPPCA) algorithm^{44,45}, which isolates and suppresses the noise-only component of the spatial-diffusion signals in the eigenspectrum domain. Many variants of this algorithm based on similar principles have been proposed^{31,46,47}. Along this line of research, another novel method “Patch2Self”⁴⁸ approximates the signal intensity of each voxel in a DWI volume from a multi-b-value and multi-directional diffusion dataset by learning a locally linear relationship between it and the signal intensities of small spatial patches around this voxel from other DWI volumes of this dataset in a self-supervised manner. Since the random noise cannot be approximated, Patch2Self removes noise and has been demonstrated useful for a variety of empirical and simulated diffusion datasets. Moreover, multiple DWIs have also been jointly reconstructed to exploit their inter-image correlation for enhanced SNR^{38,49,50}. Another category of methods explicitly imposes a model of the signals in diffusion space to remove noise^{51,52}.

Emerging deep learning technologies, particularly convolutional neural networks (CNNs), offer another powerful tool set for image denoising. With supervision, CNNs can automatically learn to fully utilize the redundancy embedded in the data and effectively restore noise-free images from their noisy observations. It has been shown that the CNN for denoising (i.e., DnCNN⁵³) with a simple network architecture outperforms the state-of-the-art BM3D denoising method. CNN-based denoising has been widely adopted for fluorescence microscopy^{54,55}, optical coherence tomography⁵⁶, x-ray imaging⁵⁷, x-ray computed tomography⁵⁸, PET^{59–63} and MRI^{58,64–71}. For diffusion MRI, many studies have proven the superiority of CNNs in estimating high-quality scalar diffusion metrics from DTI^{72–77} and more advanced diffusion models^{73,76–78} as well as voxel-wise axonal orientations⁷⁹ from a small amount of input data for faster imaging. In parallel to these studies, the DeepDTI⁷⁰ method leverages CNNs to denoise six DWI volumes sampled along optimally selected diffusion-encoding directions with the target clean images synthesized from tensors fitted using more data, achieving approximately four-fold acceleration over non-denoised images.

Despite their superior performance, most deep learning-based denoising methods require additional high-SNR data for the supervised training of CNNs and are therefore more difficult to use in practice compared to conventional algorithms. On the one hand, the performance of a pre-trained CNN might be compromised when it is directly applied to a new dataset acquired with different hardware systems and sequences, which exhibit different image contrast, spatial resolution, SNR level, and so on. On the other hand, to create a custom CNN optimized for the data in a new application requires additional high-SNR image data as the training target from numerous subjects, even for fine-tuning parameters of pre-trained CNNs. These high-quality training targets are usually obtained from a large

number of measurements from longer scans, which might be challenging to acquire and are also unavailable for legacy data.

To address this challenge, we present a self-supervised deep learning framework for denoising DTI data entitled “SDnDTI” that does not require external high-SNR data for training. Specifically, SDnDTI generates multiple repetitions of DWI volumes exhibiting identical image contrasts but independent noise observations from multi-directional diffusion DTI data through the diffusion tensor model. SDnDTI removes noise by first denoising each repetition of DWIs using a deep 3-dimensional CNN with the average of all repetitions as the higher-SNR training target, following the same approach as normal supervised learning based denoising methods, and then averaging CNN-denoised images for achieving higher SNR. In this work, we lay out the framework for SDnDTI and then systematically quantify the similarity of denoised images and resultant DTI metrics from SDnDTI compared to the ground truth generated from substantially more DWI volumes using two separate datasets acquired with different spatial resolutions and b-values, namely, those from the Human Connectome Project (HCP) and Lifespan HCP in Aging. We show that SDnDTI results substantially improve upon those from the raw data, outperform those from state-of-the-art denoising algorithms, including BM4D^{25,30}, adaptive optimized NLM (AONLM)²⁹ and MPPCA, and are comparable to those from supervised deep learning-based denoising using external ground-truth data as the training target. Because of the superior performance and reduced requirement for training data, we anticipate easier deployment and wider use of SDnDTI in practice that might benefit a broader range of clinical and neuroscientific research.

METHODS

SDnDTI pipeline

SDnDTI employs the approach of first denoising, followed by averaging. Acquiring and averaging multiple repetitions of noisy images is a common practice to increase image SNR. Instead of directly averaging the acquired images, this approach first denoises each single noisy image using a CNN with the averaged image as the higher-SNR training target, following normal supervised learning based denoising methods. Since each denoised image has equivalent SNR compared to the averaged image due to the superior denoising performance of the CNN, the average of all denoised images recovers higher SNR than the averaged image and thus achieves denoising effects. Take n acquired noisy images for example, each CNN-denoised image is similar to the average of n noisy images and therefore averaging n CNN-denoised images is approximately equivalent to averaging n^2 noisy images, which achieves higher SNR than directly averaging n acquired noisy images (i.e., $n^2 - n$ difference). The denoising effect is stronger for larger n .

Implementing the approach of denoising first followed by averaging for multiple interspersed $b = 0$ image volumes of a DTI dataset is straightforward. If multiple repetitions of DWI volumes are acquired, this approach can be readily applied. However, DWI volumes in a DTI dataset are normally sampled along uniformly distributed directions exhibiting different image contrasts, and therefore the raw acquired data do not readily provide multiple repetitions of DWI volumes with identical image contrasts but independent noise

observations. SDnDTI leverages the diffusion tensor model to transform the image contrast of DWIs. Specifically, SDnDTI divides all DWI volumes into several subsets (e.g., m subsets), each with six DWI volumes along diffusion-encoding directions optimized for the tensor fitting, then estimates diffusion tensors from each subset of DWI volumes (along with the averaged $b = 0$ image volume), and finally synthesizes DWI volumes along a fixed set of six directions from the estimated diffusion tensors. Consequently, m repetitions of six DWI volumes with identical image contrasts but independent noise observations are generated. To avoid the loss of angular sampling coverage, SDnDTI instead synthesizes volumes along all acquired directions rather than six directions.

The SDnDTI pipeline for a DTI acquisition consisting of three $b = 0$ image volumes and 18 DWI volumes is demonstrated in Figure 1 and could be extended to any number of DWI volumes. The diffusion-encoding directions of the 18 DWI volumes need to be optimized such that they can be divided into three subsets of six directions which are optimal for the diffusion tensor fitting, and the 18 directions are also uniformly distributed on a sphere to ensure uniform angular coverage.

For each subset of six DWI volumes, tensor fitting is performed along with the averaged $b = 0$ image volume to estimate low-quality diffusion tensors, which are then used to synthesize DWI volumes sampled along the 18 acquired encoding directions. The synthesis of these DWI volumes serves to transform DWI volumes sampled along different encoding directions to the same directions while maintaining full angular coverage. A single $b = 0$ image volume and 18 synthesized DWI volumes serve as the inputs to the CNN. Specifically, for each voxel, the diffusion tensor $\mathbf{D}_6 = [D_{xx} \ D_{yy} \ D_{zz} \ D_{xy} \ D_{xz} \ D_{yz}]^T$ consisting of six unique elements is estimated using ordinary linear squares fit as:

$$\mathbf{D}_6 = \mathbf{A}_6^{-1} \mathbf{C}_6, \quad (1)$$

where $\mathbf{C}_6 = [c_1 \ c_2 \ c_3 \ c_4 \ c_5 \ c_6]^T$ represents the apparent diffusion coefficients (ADCs) along the 6 diffusion-encoding directions, with $c_i = -\ln(S_i/S_0)/b_i$ ($i = 1, 2, 3, 4, 5, 6$), S_0 as the non-diffusion-weighted signal intensity from the average of the three $b = 0$ image volumes, S_i as the diffusion-weighted signal intensity and b_i as the b-value, and $\mathbf{A}_6 = [\alpha_1 \ \alpha_2 \ \alpha_3 \ \alpha_4 \ \alpha_5 \ \alpha_6]^T$ represents the diffusion tensor transformation matrix, with $\alpha_i^T = [g_{ix}^2 \ g_{iy}^2 \ g_{iz}^2 \ 2g_{ix}g_{iy} \ 2g_{ix}g_{iz} \ 2g_{iy}g_{iz}]$ ($i = 1, 2, 3, 4, 5, 6$) solely depending on the diffusion-encoding directions $(g_{ix}, g_{iy}, g_{iz})^T$. In order to avoid large noise amplification when solving the tensor, the six diffusion-encoding directions need to be selected to minimize the condition number of \mathbf{A}_6 ⁸⁰. The image intensities of each voxel in the synthesized DWI volumes along the 18 acquired diffusion-encoding directions are calculated as:

$$\mathbf{S}_{18} = S_0 e^{-diag(\mathbf{b}) \cdot \mathbf{A}_{18} \mathbf{D}_6} = S_0 e^{-diag(\mathbf{b}) \cdot \mathbf{A}_{18} \mathbf{A}_6^{-1} \mathbf{C}_6} \quad (2)$$

, where \mathbf{A}_{18} is the diffusion tensor transformation matrix associated with the 18 acquired directions, and $diag$ stands for the diagonalization operation of a vector of the b-values \mathbf{b} of the 18 DWI volumes. The calculated image intensities of \mathbf{S}_{18} along the directions of

the six DWI volumes used for the tensor fitting are identical to the raw acquired image intensities since the diffusion tensor transformation is well conditioned and fully invertible (except for very few outlier voxels where some DWI intensities because of noise happen to be higher than the $b = 0$ image intensity such as on the edge of the brain where the $b = 0$ image intensities are very low). Only the remaining 12 computed image intensities of S_{18} are actually synthesized, which are different from the raw acquired image intensities and also exhibit different noise characteristics.

The tensor fitting is also performed on three $b = 0$ image volumes and all 18 DWI volumes to estimate a diffusion tensor with higher quality, which is then used to generate synthesized DWI volumes with higher SNR along the 18 acquired diffusion-encoding directions in a similar way as described above. The averaged $b = 0$ image volume and the 18 synthesized DWI volumes with higher SNR serve as the target of the CNN.

For each subject, three pairs of inputs and targets consisting of one $b = 0$ image volume and 18 DWI volumes are created. The $b = 0$ image volume and DWI volumes are jointly denoised to enhance data redundancy for boosted CNN performance. Any CNN for denoising can be trained in a supervised fashion to improve the SNR of each subset of input image volumes. The training can be performed using data from many subjects in a project which need to be denoised or from a single subject in a subject-specific fashion. Generally speaking, denoising the data of numerous subjects jointly is beneficial because the increased amount of training data can be used to train deeper CNNs with more parameters for boosted denoising performance. The denoised images of all three subsets from the CNN are averaged to obtain the final denoised results, which are then used for tensor fitting to derive scalar and orientation DTI metrics.

A simple approach can be used to select the optimal 18 diffusion-encoding directions. First, the six optimized diffusion-encoding directions from the DSM scheme⁸⁰ that minimize the condition number of the diffusion transformation matrix to 1.3228 while being as uniform as possible are chosen. Second, three sets of the six optimized directions are randomly rotated for many times and the 18 most uniform directions, i.e., with the minimal electrostatic potential energy⁸¹, are chosen.

Modified U-Net

A modified 10-layer 3-dimensional (3D) U-Net⁸² (MU-Net) was used to learn the mapping from the input noisy image volumes to the residuals between the input and output image volumes with higher SNR (i.e., residual learning), which were then added to the input image volumes to generate resultant denoised image volumes. Specifically, all max pooling and up-sampling layers of U-Net were removed to preserve the native spatial resolution at each layer, and the number of kernels at each layer was kept constant ($k = 192$). Compared to 2D convolution, 3D convolution ($d \times d \times d = 3 \times 3 \times 3$ kernel size, $1 \times 1 \times 1$ stride) increases the data redundancy from an additional spatial dimension for improved image synthesis performance and avoids boundary artifacts between 2D image slices. Essentially, MU-Net is composed of a sequence of paired convolution, batch normalization and rectified linear unit (ReLU) activation layers with several short paths from early layers to later layers. These skip connections serve to alleviate the vanishing-gradient problem and strengthen feature

propagation. MU-Net represents an intermediate network between a plain network (e.g., VDSR⁸³ and DnCNN⁵³) without any short paths and a densely connected network (e.g., DenseNet⁸⁴) that comprehensively connects each layer to every other layer.

Human Connectome Project data

Pre-processed diffusion MRI data of 20 unrelated young healthy subjects from the HCP WU-Minn-Ox Consortium (<https://www.humanconnectome.org>) were used for this study. The acquisition methods and parameter values have been described in detail previously^{85–87}, and those relevant to this study are briefly described below. Diffusion MRI data were acquired in the whole brain at 1.25 mm isotropic resolution using a two-dimensional diffusion-weighted pulsed-gradient spin-echo echo-planar imaging (DW-PGSE-EPI) sequence, with three b-values (1, 2, 3 ms/ μm^2) and two phase-encoding directions (left–right and right–left). For each b-value, 90 diffusion-encoding directions uniformly distributed on a sphere were acquired⁸⁸. The diffusion data were corrected for gradient nonlinearity, eddy current and susceptibility induced distortions and co-registered using the FMRIB Software Library (FSL) software^{89–92} (<https://fsl.fmrib.ox.ac.uk>). The image volumes acquired with the left–right and right–left phase-encoding directions were combined into a single image volume by FSL’s “eddy” function. Only the 18 interspersed combined $b = 0$ image volumes and 90 combined DWI volumes at $b = 1 \text{ ms}/\mu\text{m}^2$ were used in this study. In addition, the volumetric segmentation results from the FreeSurfer^{93,94} (<https://surfer.nmr.mgh.harvard.edu>) reconstruction on the T_1 -weighted data were also used in this study to derive brain tissue masks for results evaluation.

Human Connectome Project in Aging data

The diffusion and T_1 -weighted MRI data of 20 unrelated healthy adults (ages 36–93, mean age 65.8 ± 19.1 , 10 female) from the Lifespan HCP in Aging (HCP-A) study^{95,96} were used for this study. The data were acquired as part of the HCP-A at the Massachusetts General Hospital Martinos Center for Biomedical Imaging with approval from the institutional review board and written informed consent from all participants. The subjects were randomly selected with uniformly distributed ages. The data acquisition was performed using a 3-T MRI scanner (MAGNETOM Prisma; Siemens Healthcare, Erlangen, Germany) equipped with the Siemens 32-channel Prisma head coil for signal reception.

Whole-brain diffusion data were acquired at 1.5 mm isotropic resolution using a two-dimensional DW-PGSE-EPI sequence with the following parameters: repetition time = 3,230 ms, echo time = 89.2 ms, contiguous axial slices, simultaneous multi-slice factor = 4, without in-plane acceleration, with two b-values (1.5, 3 ms/ μm^2) and two phase-encoding directions (anterior–posterior and posterior–anterior). For $b = 1.5 \text{ ms}/\mu\text{m}^2$ and $b = 3 \text{ ms}/\mu\text{m}^2$, 93 and 92 diffusion-encoding directions uniformly distributed on a sphere were acquired. Only the 28 interspersed $b = 0$ image volumes and 186 DWI volumes at $b = 1.5 \text{ ms}/\mu\text{m}^2$ were used in this study.

Whole-brain T_1 -weighted data were acquired at 0.8-mm isotropic resolution using a 3-dimensional sagittal multi-echo magnetization-prepared rapid acquisition with gradient echo (ME-MPRAGE) sequence⁹⁷ with the following parameters: repetition time = 2,500 ms, echo

time = 1.8/3.6/5.4/7.2 ms, inversion time = 1000 ms, flip angle = 8°, partial Fourier factor = 6/8, generalized autocalibrating partial parallel acquisition (GRAPPA) factor = 2.

Human Connectome Project in Aging data Processing

The HCP-A diffusion data were corrected for eddy current and susceptibility induced distortions and co-registered using the “topup” and “eddy” functions from the FSL software. In order to compare to results from the MPPCA denoising method^{44,45}, which needs to be applied to the unprocessed raw data from the scanner, to other methods, the resultant warp field maps for correcting and aligning each image volume from the “eddy” function were saved out with the “--dfields” option. Each warp field map was applied to individual image volume from the MPPCA-denoised raw data using FSL’s “applywarp” function with the “spline” interpolation to obtain the distortion-free and co-registered MPPCA-denoised data. Each warp field map was also applied to individual image volume from the raw diffusion data using FSL’s “applywarp” function with the “spline” interpolation to obtain the distortion-free and co-registered raw data, which were used for the subsequent image processing and denoising. The “topup” and “eddy” processing were not performed on the raw data and MPPCA-denoised data separately because the resultant warp field maps might be different due to distinctive input images, which might confound the subsequent comparison of denoising methods. The image volumes acquired with the opposite phase-encoding directions were averaged and combined into a single image volume to account for the signal loss in the brain regions with large susceptibility induced distortions due to the absence of in-plane acceleration, resulting in 14 combined $b = 0$ image volumes and 93 combined DWI volumes at $b = 1.5 \text{ ms}/\mu\text{m}^2$. The corrected data directly from the “eddy” function were not used, because the “eddy” function internally replaces the detected outlier image slices with its own estimation, which introduces a confounding factor for the comparison of MPPCA and other denoising methods.

FreeSurfer reconstruction was performed on the T_1 -weighted data using the “recon-all” function.

Image processing

For each subject, the diffusion data were corrected for spatially varying intensity biases using the averaged $b = 0$ image volume with the unified segmentation routine implementation in the Statistical Parametric Mapping software (SPM, <https://www.fil.ion.ucl.ac.uk/spm>) with a full-width at half-maximum of 60 mm and a sampling distance of 2 mm.

The volumetric brain segmentation results (i.e., `aparc+aseg.mgz`) from FreeSurfer reconstruction on the T_1 -weighted data were re-sampled to the diffusion image space with an affine transformation using nearest neighbor interpolation. For the HCP data, the diffusion data and the T_1 -weighted were co-registered already and therefore the affine transformation was simply an identity matrix. For each HCP-A subject, the affine transformation was derived using the averaged $b = 0$ image volume with FreeSurfer’s “`bbregister`” function⁹⁸. Binary masks of brain tissue that excluded the cerebrospinal fluid

(CSF) were obtained using FreeSurfer's "mri_binarize" function with "--gm" and "--all-wm" options which were used for evaluating results.

Input data selection

Three $b = 0$ image volumes and 18 DWI volumes of each subject from HCP and two $b = 0$ image volumes and 12 DWI volumes of each subject from HCP-A were used to demonstrate the denoising efficacy of SDnDTI on input data with different spatial resolution, b -values and number of input DWI volumes as well as compare to other denoising methods. Because the diffusion-encoding directions of the pre-acquired HCP and HCP-A data were fixed, it was impossible to obtain the optimal 6 diffusion-encoding directions from the DSM scheme or their rotations that minimized the condition number of the diffusion transformation matrix to 1.3228. Therefore, all possible sets of six diffusion-encoding directions that associate with a diffusion tensor transformation matrix with a condition number lower than 1.6 were used (31 sets for HCP data and 46 sets for HCP-A data). These sets of diffusion-encoding directions were selected by randomly rotating the 6 optimal directions from the DSM scheme to six new directions and then keeping the set of the six nearest directions if their associated condition number was lower than 1.6. Then three out of 31 sets of six directions for the HCP data and two out of 46 sets of six directions for the HCP-A data were randomly picked many times, and the 18 or 12 selected directions with the lowest electrostatic potential energy⁹⁹ were chosen, which were uniformly distributed on a sphere (Figure 1, Supplementary Figure 1).

SDnDTI denoising implementation

The MU-Net of SDnDTI was implemented using the Keras application programming interface (<https://keras.io>) with a Tensorflow backend (<https://www.tensorflow.org>). The mean absolute error (MAE, i.e., L1 loss) was used to optimize the CNN parameters using the Adam optimizer¹⁰⁰ with default parameters (except for the learning rate). The learning rate was empirically set to 0.0001. Only the MAE within the brain mask was used.

To account for subject-to-subject variations in image intensity, the intensities of the input and target images of SDnDTI were standardized by subtracting the mean image intensity and dividing by the standard deviation of image intensities across all voxels within the brain mask from the input images. Input and target images were brain masked. The training data were flipped along the anatomical left-right direction to augment the training data. All 3D convolutional kernels were randomly initialized with a "He" initialization¹⁰¹. Blocks of $64 \times 64 \times 64$ voxel size were used for training (8 blocks from each subject) due to the limited memory of graphics processing unit (GPU).

The training and validation were performed on 20 subjects using a Tesla V100 GPU with 16 GB memory (NVIDIA, Santa Clara, CA). For each epoch, 80% randomly selected blocks were used for training and the remaining 20% were used for validation. The batch size was set to one, the largest size that can be accommodate by the GPU. The training and validation were performed for 40 epochs for the HCP data (~60 minutes per epoch) and 60 epochs for the HCP-A data (~30 minutes per epoch). During the training, the training error kept decreasing while the validation error decreased at first, reached the minimum

and then started increasing when the CNN parameters started to be overfitted. The MU-Net parameters with the minimum validation errors were used (from the 34th epoch for the HCP data and the 19th epoch for the HCP-A data).

The learned network parameters were applied to the whole brain volume of each subject. The standardized image intensities were transformed back to the normal range by multiplying with the standard deviation of image intensities across all voxels within the brain mask from the input images and then add the mean image intensity of brain voxels.

The network implementation and training parameters (e.g., learning rate, the way to select the suitable number of training epoch based on tracking the validation error) were kept the same in the following sections if not explicitly specified.

Effects of the amount of training data

Experiments were performed on the HCP data to assess the effects of the amount of training data on the SDnDTI performance. For the HCP data, the MU-Net of SDnDTI was trained on data from subsets of 10 subjects, 5 subjects, or 1 subject out of the 20 subjects in the similar way as described above, resulting in 2, 4, and 20 networks with optimized parameters respectively. Each optimized MU-Net was then applied to the data of subjects used for its training to denoise images.

For the training using 10 subjects, the training and validation were performed for 80 epochs (~60 minutes per epoch) and the networks from the 50th and 57th epoch with the lowest validation errors were used for denoising the data from the first 10 subjects and the last 10 subjects, respectively. For the training using 5 subjects, the training and validation were performed for 80 epochs (~15 minutes per epoch) and the networks from the 52nd, 47th, 52nd and 43rd epoch were used for denoising the data from each quarter of the 20 subjects, respectively. For the training using 1 subject, the training and validation were performed for 100 epochs (~3 minutes per epoch) and the networks from the 69th, 84th, 41st, 98th, 57th, 88th, 81st, 89th, 86th, 98th, 89th, 79th, 98th, 96th, 78th, 77th, 79th, 72nd, 95th, 79th epoch with the lowest validation errors were used for denoising the data from each of the 20 subjects.

Network generalization and fine tuning

In order to evaluate the generalization of the MU-Net of SDnDTI, two $b = 0$ image volumes and 12 DWI volumes of each of the 20 subjects from HCP (1.25 mm isotropic resolution, $b = 1 \text{ ms}/\mu\text{m}^2$) were selected and used to train a MU-Net in the same way as described above. In order to be directly applied to the HCP-A data, the input and target DWIs were synthesized along the encoding directions from the HCP-A input data from the tensor fitted using each subset of the HCP data and all selected data, respectively. The training and validation were performed for 40 epochs (~40 minutes per epoch) and the network from the 30th epoch with the lowest validation error was directly applied to denoise the data (1.5 mm isotropic resolution, $b = 1.5 \text{ ms}/\mu\text{m}^2$) from each of the HCP-A subjects.

The data from the HCP-A were used for fine-tuning parameters of the pre-trained MU-Net using the HCP data. Specifically, for each of the 20 subjects from the HCP-A, another

MU-Net, initialized with parameters of the MU-Net learned from the HCP data, were further fine-tuned on the data of this subject. The training and validation were performed for 40 epochs (~1 minute per epoch) and the networks from the 23rd, 12nd, 32nd, 18th, 23rd, 29th, 23rd, 23rd, 34th, 20th, 30th, 14th, 17th, 29th, 27th, 30th, 33rd, 19th, 21st, 8th epoch with the lowest validation errors were used for denoising the data from each of the 20 HCP-A subjects.

For comparison, an MU-Net was also trained from randomly initialized parameters on the data from each of the 20 HCP-A subjects. The training and validation were performed for 80 epochs (~1 minute per epoch) and the networks from the 40th, 30th, 41st, 36th, 48th, 37th, 38th, 46th, 65th, 44th, 30th, 33rd, 40th, 35th, 37th, 42nd, 43rd, 36th, 41st, 25th with the lowest validation errors were used to denoise the data from each of the 20 HCP-A subjects. Compared to the fine tuning, training from random initialization required more epochs to converge.

Denoising using other methods

For comparison, diffusion data were also denoised using three state-of-the-art traditional denoising methods, including BM4D, AONLM and MPPCA, as well as using supervised learning with external high-SNR data. BM4D and AONLM were applied to pre-processed HCP and HCP-A diffusion data. MPPCA should only be used for raw unprocessed diffusion data since the image resampling of the preprocessing steps changes the noise characteristics and makes the noise assumption of MPPCA invalid. Therefore, MPPCA was not evaluated on the preprocessed HCP data. MPPCA was only applied to the unprocessed HCP-A data. The MPPCA-denoised HCP-A data were then corrected for distortions and co-registered as described above.

BM4D denoising, an extension of the well-known BM3D algorithm for volumetric data, was set to estimate the unknown noise standard deviation of the Rician noise and perform collaborative Wiener filtering with “modified profile” option and default parameters using the publicly available MATLAB-based software (<https://www.cs.tut.fi/~foi/GCF-BM3D>). The AONLM was performed assuming Rician noise with $3 \times 3 \times 3$ block and $7 \times 7 \times 7$ search volume^{28,29} using the publicly available MATLAB-based software (<https://sites.google.com/site/pierrickcoupe/software/denoising-for-medical-imaging/mri-denoising/mri-denoising-software>). The MPPCA denoising was performed with $5 \times 5 \times 5$ kernel and “full” sampling using the publicly available MATLAB-based software (https://github.com/NYU-DiffusionMRI/mppca_denoise).

Supervised learning-based denoising using MU-Net with external high-SNR data as the training target was also performed for comparison. In this case, the input of the MU-Net is the raw acquired three $b = 0$ image volumes and 18 DWI volumes or two $b = 0$ image volumes and 12 DWI volumes with low SNR for the HCP and HCP-A data, respectively. The output of the MU-Net is three $b = 0$ image volumes and 18 DWI volumes or two $b = 0$ image volumes and 12 DWI volumes with high SNR for the HCP and HCP-A data, respectively. For each subject, the high-SNR $b = 0$ image volume was computed by averaging all available $b = 0$ image volumes (i.e., 18 or 14 volumes for each HCP or HCP-A subject, respectively). The high-SNR DWI volumes were synthesized from the

diffusion tensor generated using all diffusion data (i.e., 18 $b = 0$ image volumes and 90 DWI volume for each HCP subject, 14 $b = 0$ image volumes and 93 DWI volumes for each HCP-A subject) as described in Equation 2. For both the HCP and HCP-A data, training and validation were performed on the data from 20 subjects for 40 epochs (~21 minutes per epoch for the HCP data and ~14 minutes per epoch for the HCP-A data). The MU-Net from the 31st epoch and the 32nd epoch were used for denoising the HCP and HCP-A data, respectively.

Quantitative comparison

Resultant denoised images and the DTI metrics including the primary eigenvector (V1), fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), and radial diffusivity (RD) were compared to the ground truth for evaluating the denoising performance. The ground-truth $b = 0$ image volumes were computed by averaging all available $b = 0$ image volumes for each subject and the ground-truth DWI volumes were synthesized from the diffusion tensor generated using all available diffusion data. Diffusion tensor fitting was performed using ordinary linear squares fitting using FSL's "dtifit" function to derive the diffusion tensor, V1, FA, MD, AD and RD.

The MAE, Peak SNR (PSNR) and structural similarity index (SSIM)¹⁰² were used to quantify the similarity between the raw input images and denoised images using different methods compared to the ground-truth images. For these calculations, the image intensities of BM4D-, AONLM- and MPPCA-denoised image volumes were first standardized in the same way as for preparing the input and output data for SDnDTI. The standardized image intensities for denoised image volumes from all methods, within the range $[-3, 3]$, were transformed to the range of $[0, 1]$ by adding 3 and dividing by 6. PSNR and SSIM were computed using MATLAB's "psnr" and "ssim" function. The group mean and group standard deviation of the MAE, PSNR and SSIM across the 20 subjects from HCP and HCP-A were calculated, respectively.

The MAE of V1, FA, MD, AD, RD compared to the ground-truth DTI metrics within the brain (excluding the cerebrospinal fluid) were used to quantify the accuracy of the DTI results. The difference of V1 was computed as the angle (between 0 and 90°) between the two primary eigenvectors for comparison. The group mean and group standard deviation of the mean values of the MAE for different metrics across the 20 subjects from HCP and HCP-A were calculated, respectively.

Results

Figure 3 shows significantly improved SNR and quality for the denoised DWIs from SDnDTI (denoised $b = 0$ images are available in Supplementary Fig. S2). The DWI from subset 2 of SDnDTI input data (Fig. 3, a–c, iii), which was synthesized from the diffusion tensor generated using 6 DWIs, exhibited only slightly higher noise level than that of the raw acquired DWI from subset 1 of SDnDTI input data (Fig. 3, a–c, ii). The image similarity was comparable to the ground-truth DWI (0.033 vs. 0.031 MAE, 27.10 dB vs 27.82 dB PSNR, and 0.875 vs 0.891 SSIM) due to the use of optimized diffusion-encoding directions, which did not make the subsequent denoising task more challenging. Consequently, the

SDnDTI-denoised raw DWI (Fig. 3, d–f, i) and SDnDTI-denoised synthesized DWI (Fig. 3, d–f, ii) were indeed visually and quantitatively similar (0.017 vs. 0.017 MAE, 32.51 dB vs 32.32 dB, and 0.960 vs 0.961), as well as similar to the target DWI synthesized from the diffusion tensor generated using all 18 DWI volumes (Fig. 3, a–c, iv, 0.018 MAE, 32.04 dB PSNR, 0.956 SSIM).

The final result of SDnDTI, i.e., the average of the three denoised DWIs from all three subsets, achieved further improved image quality (Fig. 3, d–f, iii, 0.012 MAE, 34.85 dB PSNR, 0.979 SSIM). The SDnDTI-denoised DWI outperformed the training target of SDnDTI (Fig. 3, a–c, iv), as well as the raw acquired DWI, and the BM4D-denoised (Fig. 3, d–f, iv, 0.022 MAE, 30.72 dB PSNR, 0.942 SSIM) and AONLM-denoised raw DWI (Fig. 3, d–f, v, 0.023 MAE, 30.62 dB PSNR, 0.939 SSIM). The SDnDTI-denoised DWI also preserved more textural details around the internal capsule (Fig. 3, e, iii) compared to the BM4D (Fig. 3, e, iv) and AONLM (Fig. 3, e, v) results. The resultant DWI from supervised denoising (Fig. 3, a–c, v) with the ground-truth DWI as the training target achieved the highest image similarity compared to the ground-truth DWI (0.012 MAE, 36.43 dB PSNR, 0.982 SSIM) as expected. The residual maps between all denoised images and ground-truth images do not contain anatomical structure or biases reflecting the underlying anatomy (Fig. 3, rows c, f).

The group mean (\pm the group standard deviation) across the 20 HCP subjects of the MAE, PSNR and SSIM for the $b = 0$ images and the DWI shown in Figure 3 derived from the different methods are quantified in Table 1 (results for all DWIs available in Supplementary Tables 1–3). The MAE of SDnDTI-denoised DWI was approximately one third of that of the raw DWI (0.012 ± 0.00073 vs. 0.033 ± 0.0021), two thirds of that of the target DWI during the SDnDTI training (0.019 ± 0.0012), half of those of BM4D (0.023 ± 0.0012) and AONLM results (0.024 ± 0.0012), and equivalent to that of the supervised denoising results (0.012 ± 0.00072). The group mean (\pm the group standard deviation) of the PSNR of SDnDTI-denoised DWI was approximately 7 dB higher than that of the raw DWI (34.50 ± 1.32 dB vs. 27.13 ± 0.57 dB), 3 dB higher than that of the target DWI during the SDnDTI training (31.55 ± 0.92 dB), 4 dB higher than those of BM4D (30.23 ± 0.47 dB) and AONLM results (30.16 ± 0.46 dB), and 1.3 dB lower than that of the supervised denoising results (35.83 ± 0.54 dB). The group mean (\pm the group standard deviation) of the SSIM of SDnDTI results was about 0.1 higher than that of the raw DWI (0.98 ± 0.0024 vs. 0.88 ± 0.013), 0.03 higher than that of the target DWI during the SDnDTI training (0.95 ± 0.0054), 0.04 higher than those of BM4D (0.94 ± 0.0056) and AONLM results (0.94 ± 0.0058), and equivalent to that of the supervised denoising results (0.98 ± 0.0023).

Figure 4 demonstrates the capability of SDnDTI for estimating noise. The noise estimated by different methods (i.e., the residual maps between the acquired DWI and the denoised DWI) did not contain any noticeable anatomical structure or biases reflecting the underlying anatomy. The estimated noise maps of the synthesized DWI from the tensor fitted using 18 DWIs (Fig. 4b), the supervised denoising results (Fig. 4c) and the SDnDTI results (Fig. 4d) were visually more similar to the noise map from the ground-truth DWI (Fig. 4a) than those from the BM4D (Fig. 4e) and AONLM (Fig. 4f) results.

Figure 5 shows the ability of SDnDTI to recover detailed anatomical information from the noisy inputs as mapped in primary eigenvector V1-encoded FA maps. The FA maps from a single subset of SDnDTI-denoised data (Fig. 5d, e) were substantially less noisy compared to the map derived from the raw data (Fig. 5b). The FA map from the average of three subsets of SDnDTI-denoised data (Fig. 5f) further improved upon the maps from each single subset data (Fig. 5d, e) and was visually similar to the map from the supervised denoising (Fig. 5c). The SDnDTI map (Fig. 5f) was slightly blurred compared to the ground-truth map (Fig. 5a), but sharper than the map derived from BM4D (Fig. 5g) and AONLM (Fig. 5h), which was clearly depicted in the gray matter bridges that span the internal capsule with characteristic stripes (Fig. 5f–h, yellow boxes). These textures were buried in noise in the map derived from raw data (Fig. 5b, yellow box). Because of the blurring, the FA of the cortical gray matter in the BM4D- and AONLM-denoised results also significantly reduced (the green contour surrounding the gyrus in Fig. 5g, h, red boxes), which was preserved more in the SDnDTI-denoised results (Fig. 5f, red box).

The difference of five common DTI metrics, including V1, FA, MD, AD, RD between the results derived from different methods and ground-truth data is displayed for a representative subject (Fig. 6, Supplementary Fig. S3) and quantified for 20 HCP subjects (Table 2). The group means of the MAE derived from each subset of SDnDTI-denoised data (Fig. 6, iv, Table 2, f–h) remarkably improved upon the SDnDTI inputs (Table 2, b–d), which was similar to those from the raw data (Fig. 6, ii, Table 2, a). Because of the averaging, the group mean of the MAE derived from the final results of SDnDTI (Fig. 6, v, Table 2, i) substantially outperformed those from the raw data (Fig. 6, ii, Table 2, a) and were superior to those from BM4D-denoised (Fig. 6, vi, Table 2, j) and AONLM-denoised (Fig. 6, vii, Table 2, k) raw data. The group means of the MAE of BM4D- and AONLM-denoised raw data were in general very similar. The supervised-denoising (Fig. 6, iii, Table 2, e) achieved the lowest MAE for all DTI metrics, which were marginally lower than those from SDnDTI results (i.e., 0.025 ± 0.0014 vs. 0.026 ± 0.0015 for FA, $0.028 \pm 0.0023 \mu\text{m}^2/\text{ms}$ vs. $0.028 \pm 0.0023 \mu\text{m}^2/\text{ms}$ for MD, $0.040 \pm 0.0025 \mu\text{m}^2/\text{ms}$ vs. $0.041 \pm 0.0027 \mu\text{m}^2/\text{ms}$ for AD, and $0.030 \pm 0.0024 \mu\text{m}^2/\text{ms}$ vs. $0.030 \pm 0.0025 \mu\text{m}^2/\text{ms}$ for RD), except for the MAE for primary eigenvector (i.e., $10.52^\circ \pm 0.75^\circ$ vs. $11.20^\circ \pm 0.81^\circ$).

The denoising performance of SDnDTI using different numbers of training subjects is depicted in Figure 7. The primary eigenvector V1-encoded FA maps were visually very similar without noticeable differences. Even when the MU-Net of SDnDTI was trained on the data of each single subject, SDnDTI could still preserve image sharpness, such as the striated texture spanning the internal capsule (Fig. 7d), better than BM4D and AONLM (Fig. 5g, h). Quantitatively, the group means of the MAE in DTI metrics were the lowest if the data of the 20 HCP subjects were jointly denoised, in which case there was plenty of training data for optimizing the MU-Net of SDnDTI. When the number of training subjects was reduced from 20 to 10, the group means of the MAE in DTI metrics only marginally increased, presumably because the data from 10 subjects were still sufficient to train the MU-Net of SDnDTI. The denoising performance decreased more rapidly from 10 to 5 subjects, and 5 to 1 subject, especially for the primary eigenvector. Even when the MU-Net of SDnDTI was trained on the data of each single subject for denoising, the group means of the MAE in DTI metrics of SDnDTI were still slightly lower than those from BM4D and

AONLM (e.g., $12.44^{\circ} \pm 0.97^{\circ}$ vs. $12.46^{\circ} \pm 0.89^{\circ}$ and $12.64^{\circ} \pm 0.87^{\circ}$ for primary eigenvector, $0.031 \pm 0.0026 \mu\text{m}^2/\text{ms}$ vs. $0.036 \pm 0.0027 \mu\text{m}^2/\text{ms}$ and $0.033 \pm 0.0027 \mu\text{m}^2/\text{ms}$ for MD).

The difference of five common DTI metrics, including V1, FA, MD, AD, RD between the results derived from different approaches and ground-truth data is displayed for a representative HCP-A subject (Fig. 8 for V1, Fig. 9 for FA, Supplementary Figs. S4–S6 for MD, AD and RD) and quantified for the 20 HCP-A subjects (Table 3). SDnDTI substantially improved the SNR of the V1 and FA maps from the raw data (Figs. 8, 9, a, ii, vs. c, ii). The FA maps from supervised denoising (Fig. 9, a, iii), SDnDTI (Fig. 9, c, ii) and MPPCA (Fig. 9, c, i) appear sharper than those from BM4D (Fig. 9, a, iv) and AONLM (Fig. 9, a, v) and are visually more similar to the ground-truth map (Fig. 9, a, i).

Quantitatively, the group means of the MAE from SDnDTI (Table 3f) were substantially lower than those from the raw data (Table 3, a), as well as were superior to those from BM4D (Table 3, c), AONLM (Table 3, d) and MPPCA (Table 3, e), especially for MD, AD and RD for BM4D and AONLM and primary eigenvector, FA, AD and RD for MPPCA. As expected, the group means of the MAE of supervised denoising were the lowest (Table 3, a), which were slightly lower than those from SDnDTI for scalar metrics FA, MD, AD and RD while showing a greater advantage compared to those from SDnDTI for the primary eigenvector. The MU-Net of the SDnDTI generalized to different data reasonably well. When applying the MU-Net of SDnDTI trained on the data from HCP subjects directly to the data from HCP-A subjects acquired with very different hardware systems and protocols, the resultant maps were cleaner than those from the raw data (Figs. 8, 9, a, ii vs. c, iii), with the group means of the MAE (Table 3, g) lower than those from the raw data (Table 3, a) and MPPCA (Table 3, e), but higher than those from BM4D (Table 3, c), AONLM (Table 3, d) and SDnDTI trained on data from the 20 HCP-A subjects (Table 3, f). If the HCP MU-Net was further fine-tuned using the data of each single HCP-A subject, the denoising performance of the fine-tuned MU-Net was equivalent to the one optimized using much more training data from 20 HCP-A subjects (Figs. 8, 9, c, ii, vs. c, iv, Table 3, f vs. h). In contrast, the denoising performance of the MU-Nets trained from random initialization on each single HCP-A subject was slightly inferior to that of the fine-tuned ones (Table 3, h vs. i), which still outperformed the raw data (Table 3, a) and MPPCA (Table 3, e).

Discussion

In this study, we have developed a self-supervised deep learning approach called SDnDTI for denoising DTI data that does not require external high-SNR data for training. SDnDTI works by first denoising each single image volume of the multi-directional DTI data with the averaged image volume as the target using CNNs and then averaging multiple denoised results. The performance of SDnDTI is systematically evaluated in terms of the quality of output images and DTI metrics, as well as compared to supervised learning based denoising and conventional state-of-the-art denoising algorithms BM4D, AONLM and MPPCA on two different datasets provided by HCP and HCP-A. SDnDTI-denoised images preserve textural details and are sharper than those from BM4D and AONLM as well as similar to the ground truth with low MAEs of ~ 0.01 for $b = 0$ images and ~ 0.012 for DWIs, high PSNR of ~ 36 dB for $b = 0$ images and ~ 35 dB for DWIs, and high SSIM of ~ 0.99 for $b =$

0 images and ~ 0.98 for DWIs. SDnDTI derived images and DTI metrics are comparable to those from supervised denoising, substantially outperform those from the raw data, and are superior to those from BM4D, AONLM and MPPCA. SDnDTI is capable to generalize to different datasets and benefits from further fine tuning and more training data when the data of numerous subjects are jointly denoised.

“Patch2Self”⁴⁸ is another recently proposed self-supervised learning based method for removing noise from diffusion MRI data (publicly available as the “patch2self” function from the Dipy software¹⁰³, <https://dipy.org>), with a different denoising mechanism comparing to SDnDTI. Specifically, Patch2Self assumes a linear relationship between the signal intensity of a voxel in a DWI volume and the signal intensities from local spatial patches (e.g., $(2r+1) \times (2r+1) \times (2r+1)$ where r represents the patch radius) around this voxel from other DWI volumes (e.g., $n-1$) in a diffusion dataset. Patch2Self automatically learns the $(2r+1)^3 \times (n-1)$ to one mapping using linear regression (e.g., ordinary least squares regression, Ridge regression). Since the independent and random noise in each DWI volume cannot be approximated, the approximated DWI signals are free of noise, thereby achieving denoising. On the other hand, SDnDTI employs the approach of first denoising, followed by averaging. Specifically, when multiple repetitions of noisy images are available, instead of directly averaging these repetitions, each single noisy image is first denoised using the averaged image with higher SNR as the target for training the CNN, following the same approach as supervised learning based denoising methods. The CNN-denoised images are then averaged. The major difference between Patch2Self and SDnDTI lies in the training target. Patch2Self uses noisy DWI signals as the training target while SDnDTI uses higher-SNR DWI signals obtained from data averaging as the training target. The approach of first denoising, then averaging, rather than directly averaging, could also be applied to any applications in which multiple measurements are acquired, such as many repetitions of noisy T₁-weighted and T₂-weighted data at sub-millimeter isotropic spatial resolution. Patch2Self can be readily used for any type of diffusion data while SDnDTI is specifically designed for denoising DTI data with the potential to be extended to multi-shell data with higher b-values. In terms of denoising performance, Patch2Self-denoised images (“ols” option, $r=0, 1, \text{ and } 2$) are similar to those from BM4D and AONLM (Supplementary Figure S8, a–d, Supplementary Table S4a), quantified using data from the 20 HCP healthy subjects from our study. The primary eigenvectors and FA derived from Patch2Self are slightly less accurate compared to those from BM4D and AONLM, while the MD, AD and RD derived from Patch2Self are similar to those from BM4D and AONLM (Supplementary Figure S8e, f, Supplementary Table S4b).

The efficacy of the “first denoising then averaging” concept relies on the superior performance of deep learning-based denoising using CNNs, which are capable to map the noisier image data to the cleaner image data without compromising image quality (Figure 3, Supplementary Figure S2, Table 1, Supplementary Table 1–3). The superior performance is due to the use of residual learning and deep 3D CNNs. On the one hand, residual learning (i.e., learning the residuals between the input noisier image data and the target high-SNR image data) not only boosts the CNN performance^{53,83,104,105} since the CNN only needs to synthesize the high-spatial-frequency information, but also preserves image sharpness and textual details. On the other hand, deep 3D CNNs (i.e., 10-layer 3D MU-Net

in SDnDTI) can fully exploit the redundant information contained in the data. DTI is the imaging modality that benefits the most from the “first denoising then averaging” concept since numerous DWI volumes are acquired by nature.

Implementing the “first denoising then averaging” concept for denoising multi-directional DWI volumes leverages domain knowledge of diffusion MRI physics. The challenge lies in the fact that DWI volumes in DTI data are sampled along uniformly distributed directions thus exhibiting different image contrast, while this concept requires several repetitions of DWI volumes with identical image contrasts but different noise observations. SDnDTI addresses this challenge by transforming DWIs sampled along one set of directions to another set of directions through the diffusion tensor model (Equations 1, 2), i.e., first fitting a tensor model using DWIs along one set of directions and then synthesizing DWI volumes along another set of directions. In this way, the DWI volumes from all subsets can be transformed to along the same diffusion-encoding directions. To avoid the loss of angular sampling coverage, SDnDTI synthesizes DWIs along all acquired directions rather than fewer directions as in each subset of data. Directly averaging multiple repetitions of DWI volumes that are synthesized in SDnDTI does not improve the overall SNR of the original DTI dataset or the quality of derived DTI metrics (Supplementary Figure S7). The increased SNR originates from the supervised learning based denoising using the CNN. The minimum required number of DWI volumes for SDnDTI is 12, for which two subsets of six DWI volumes could be obtained for generating two repetitions of DWI volumes with identical contrast but different noise observations. SDnDTI could be extended for diffusion data with higher b-values and multiple shells using more sophisticated signal models, such as diffusion kurtosis^{106,107}, multi-compartment model¹⁰⁸ and spherical convolution model^{109,110}. For example, the DWI volumes of a diffusion dataset could be divided into subsets for diffusion kurtosis fitting, and the estimated kurtoses from each subset are then used for generating DWI volumes along the same diffusion-encoding directions for CNN based denoising as in SDnDTI. The denoised data could be used for advanced fiber orientation estimators (e.g., CSD^{109,110} and BEDPOSTX^{111–113}) and microstructural models (e.g., NODDI¹¹⁴ and WMTI^{115,116}) that require multiple and higher b-values. Since SDnDTI requires model fitting, it only works for pre-processed diffusion MRI data and therefore could be easily used as an additional post-processing step without intervening the standard diffusion MRI data processing pipeline.

The diffusion-encoding directions in each subset of DTI data must be carefully selected such that the noise is not amplified during the image transformation process. Otherwise, the subsequent denoising task becomes more difficult and the denoising results cannot match the target DWIs even using a CNN. SDnDTI adopts the uniform encoding directions from the DSM scheme that minimize the condition number of the diffusion tensor transformation matrix to 1.3228, which successfully suppress noise amplification during the transformation. The transformed images are only slightly noisier than the raw acquired images (Fig. 3, a–c, ii, Table 2, a–d). Since the HCP and HCP-A data are pre-required and thus the directions used in this study are only approximately as designed in the DSM scheme (i.e., the condition number of the diffusion tensor transformation matrix is ~ 1.6). We expect SDnDTI performance to be improved if the actual DSM directions could be used for the data acquisition. Moreover, SDnDTI uses 6 DWI volumes for each subset because the HCP

and HCP-A data are not very noisy. For extremely noisy DTI data such as those from sub-millimeter isotropic resolution, data along much more directions are often acquired (e.g., two repetitions of 256 directions at 1-mm isotropic spatial resolution¹⁵) and more DWI volumes (e.g., 20, 30 DSM directions) should be assigned for each subset for robust tensor fitting and image transformation. In these cases, advanced tensor fitting methods such as weighted least squares and RESTORE¹¹⁷ could also be utilized for improving the fitting performance.

It is beneficial to jointly denoise the DTI data of all subjects in a study using SDnDTI because the CNN performance improves as the amount of training data increases (Figure 7). It is not trivial to theoretically determine the required number of subjects for training, which depends on the number of parameters of the CNN (~12 million for the adopted 3D MU-Net) and the information contained in the image data of each subject (influenced by factors such as the brain size, image resolution and the number of DWI volumes in a DTI dataset). Empirically, the performance of SDnDTI trained and validated using data from 10 subjects from HCP is almost identical to that of using data from 20 subjects while decreases if using data from 5 subjects (Figure 7), suggesting that data from at least 10 subjects are required to optimize the MU-Net of SDnDTI given the HCP imaging parameters (e.g., 1.25 mm isotropic resolution) and DTI protocol (e.g., 3 b = 0 image volumes and 18 DWI volumes). On the other hand, a CNN with fewer parameters (e.g., shallower or with less kernels at each layer) can be adopted, which might be well trained using limited data and be more effective, but the trade-off between the CNN parameter number and the subject number needs to be determined empirically.

A preferable approach to account for the limited training data is to fine-tune parameters of the SDnDTI CNN pre-trained using big data provided by large-scale neuroimaging studies such as HCP. Our experiments show that SDnDTI results from the MU-Nets trained and validated using data from each single HCP-A subject cannot compare to those from BM4D and AONLM, even though they outperform those from the raw data, MPPCA and the MU-Nets trained and validated on the HCP data (Table 3). This is presumably due to insufficient training data, since SDnDTI results from the MU-Net trained and validated using data from 20 HCP-A subjects indeed improve and outperform those from BM4D and AONLM (Table 3). However, further adapting parameters of the MU-Net trained and validated on the HCP data using the data from each single HCP-A subject substantially increases the quality of denoising (Table 3), which then outperforms that from all tested conventional denoising algorithms and is essentially identical to that from the MU-Net trained and validated using 20 HCP-A subjects. In addition to reducing the requirement for training data, fine-tuning also helps accelerate training convergence and reduce training time. Fine-tuning the HCP MU-Net using data from each single HCP-A subject only requires 8 to 34 minutes of training and validation time, while training and validating the MU-Net with randomly initialized parameters on data from 20 HCP-A subjects take ~20 hours, even though they produce identical results. We will make our codes for SDnDTI publicly available (<https://github.com/qiyuantian/SDnDTI>) which can be used to pre-train the CNN of SDnDTI using HCP data for fine-tuning, a recommended approach in practice, or train the CNN for supervised learning using HCP data for fine-tuning if additional high-

quality data is available from a few subjects in some applications, after all the supervised denoising achieves the highest performance (Figs. 5, 6, 8, 9, Tables 2, 3).

There are several limitations of our work. First, it is worth noting that SDnDTI is a denoising algorithm specifically designed for DTI data while other denoising algorithms evaluated in our work are more general. MPPCA and Patch2Self can denoise any diffusion MRI data while BM4D and AONLM can denoise any images. Second, since SDnDTI requires specially designed uniform diffusion-encoding directions that can be divided into subsets of six directions optimal for diffusion tensor fitting, the performance of SDnDTI on data acquired with standard diffusion-encoding directions including a large amount of legacy data is unclear, which we will characterize in future work. Future work will also focus on reducing the requirement of SDnDTI for optimized diffusion-encoding directions and render SDnDTI generalizable to any DTI or diffusion MRI data. Finally, the comparisons and evaluations in our study were only performed on empirical data provided by HCP and HCP-A. Future work may focus on characterizing different denoising algorithms using simulated diffusion MRI data¹¹⁸.

Summary

This study presents a data-driven self-supervised deep learning-based denoising method entitled SDnDTI for DTI that does not require additional high-SNR data as the target for training. SDnDTI works by first denoising each single image volume of the multi-directional DTI data with the averaged image volume as the target using CNNs and then averaging multiple denoised results for recovering even higher SNR, a concept known as “first denoising then averaging”. SDnDTI-denoised DWIs preserve image sharpness and textural details and are similar to the ground-truth DWIs with low MAEs of ~ 0.012 , high PSNR of ~ 35 dB, and high SSIM of ~ 0.98 . SDnDTI-denoised images and derived DTI metrics are comparable to results from supervised learning-based denoising that use ground-truth images as the training target, and are superior to results from the raw data, and BM4D-, AONLM- and MPPCA-denoised data. SDnDTI generalizes well to different datasets and fine-tuning parameters of the pre-trained CNN of SDnDTI further improves denoising performance as well as shortens training time. By excluding the need for external high-SNR data and the generalization of CNNs, SDnDTI increases the feasibility of deep learning and CNN-based denoising methods in a wider range of clinical and neuroscientific studies that benefit from faster DTI acquisition and improved DTI data quality.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the NIH Grants P41-EB015896, P41-EB030006, U01-EB026996, U01-EB025162, U01-AG052564, K23-NS096056, R01-NS118187, R01-EB017337, R01-MH111419, R01-EB028797, R03-EB031175, K99-AG073506, and an MGH Claflin Distinguished Scholar Award. The diffusion and anatomical data of 20 young healthy adults were provided by the Human Connectome Project, WU-Minn-Ox Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; U54-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. B.B. has provided consulting services to Subtle Medical.

Data availability

The diffusion and T₁-weighted MRI data of 20 subjects from the Human Connectome Project WU-Minn-Ox Consortium and are publicly available (<https://www.humanconnectome.org>). The diffusion and T₁-weighted MRI data from the Lifespan Human Connectome Project in Aging are publicly available (<https://www.humanconnectome.org/study/hcp-lifespan-aging>)

References

1. Basser PJ, Mattiello J.& LeBihan D.MR diffusion tensor spectroscopy and imaging. *Biophysical journal* 66, 259–267 (1994). [PubMed: 8130344]
2. Basser PJ, Mattiello J.& LeBihan D.Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B* 103, 247–254 (1994). [PubMed: 8019776]
3. Pierpaoli C, Jezzard P, Basser PJ, Barnett A.& Di Chiro G.Diffusion tensor MR imaging of the human brain. *Radiology* 201, 637–648 (1996). [PubMed: 8939209]
4. Basser PJ & Pierpaoli C.Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of magnetic resonance. Series B* 111, 209–219 (1996). [PubMed: 8661285]
5. Pierpaoli C.& Basser PJ Toward a quantitative assessment of diffusion anisotropy. *Magnetic Resonance in Medicine* 36, 893–906 (1996). [PubMed: 8946355]
6. Yeatman JD, Wandell BA & Mezer AA Lifespan maturation and degeneration of human brain white matter. *Nature communications* 5 (2014).
7. Yeatman JD, Dougherty RF, Ben-Shachar M.& Wandell BA Development of white matter and reading skills. *Proceedings of the National Academy of Sciences* 109, E3045–E3053 (2012).
8. Salat D.et al. Age-related changes in prefrontal white matter measured by diffusion tensor imaging. *Annals of the New York Academy of Sciences* 1064, 37–49 (2005). [PubMed: 16394146]
9. Nir TM et al. Effectiveness of regional DTI measures in distinguishing Alzheimer’s disease, MCI, and normal aging. *NeuroImage: clinical* 3, 180–195 (2013). [PubMed: 24179862]
10. Huber E, Donnelly PM, Rokem A.& Yeatman JD Rapid and widespread white matter plasticity during an intensive reading intervention. *Nature communications* 9, 2260 (2018).
11. Rosendaal S.et al. Regional DTI differences in multiple sclerosis patients. *NeuroImage* 44, 1397–1403 (2009). [PubMed: 19027076]
12. Zheng Z.et al. DTI correlates of distinct cognitive impairments in Parkinson’s disease. *Human brain mapping* 35, 1325–1333 (2014). [PubMed: 23417856]
13. Kubicki M.et al. DTI and MTR abnormalities in schizophrenia: analysis of white matter integrity. *NeuroImage* 26, 1109–1118 (2005). [PubMed: 15878290]
14. Cullen KR et al. Altered white matter microstructure in adolescents with major depression: a preliminary study. *Journal of the American Academy of Child & Adolescent Psychiatry* 49, 173–183. e171 (2010). [PubMed: 20215939]
15. McNab JA et al. Surface based analysis of diffusion orientation for identifying architectonic domains in the in vivo human cortex. *NeuroImage* 69, 87–100 (2013). [PubMed: 23247190]
16. Wang F.et al. In vivo human whole-brain Connectom diffusion MRI dataset at 760 μm isotropic resolution. *Scientific Data* 8, 1–12 (2021). [PubMed: 33414438]
17. Liao C.et al. High-fidelity, high-isotropic-resolution diffusion imaging through gSlider acquisition with and T1 corrections and integrated B₀/R_x shim array. *Magnetic Resonance in Medicine* 83, 56–67 (2020). [PubMed: 31373048]
18. Liao C.et al. Phase-matched virtual coil reconstruction for highly accelerated diffusion echo-planar imaging. *NeuroImage* 194, 291–302 (2019). [PubMed: 30953837]
19. McNab JA et al. High resolution diffusion-weighted imaging in fixed human brain using diffusion-weighted steady state free precession. *NeuroImage* 46, 775–785 (2009). [PubMed: 19344686]

20. Jack CR Jr et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging* 27, 685–691 (2008). [PubMed: 18302232]
21. Mueller SG et al. The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics* 15, 869–877 (2005). [PubMed: 16443497]
22. Marek K. et al. The parkinson progression marker initiative (PPMI). *Progress in Neurobiology* 95, 629–635 (2011). [PubMed: 21930184]
23. Miller KL et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature neuroscience* 19, 1523 (2016). [PubMed: 27643430]
24. Buades A, Coll B. & Morel J-M A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation* 4, 490–530 (2005).
25. Dabov K, Foi A, Katkovnik V. & Egiazarian K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing* 16, 2080–2095 (2007). [PubMed: 17688213]
26. Aharon M, Elad M. & Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing* 54, 4311–4322 (2006).
27. Gerig G, Kubler O, Kikinis R. & Jolesz FA Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on medical imaging* 11, 221–232 (1992). [PubMed: 18218376]
28. Coupé P. et al. An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. *IEEE transactions on medical imaging* 27, 425–441 (2008). [PubMed: 18390341]
29. Manjón JV, Coupé P, Martí-Bonmatí L, Collins DL & Robles M. Adaptive non-local means denoising of MR images with spatially varying noise levels. *Journal of Magnetic Resonance Imaging* 31, 192–203 (2010). [PubMed: 20027588]
30. Maggioni M, Katkovnik V, Egiazarian K. & Foi A. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Transactions on Image Processing* 22, 119–133 (2012). [PubMed: 22868570]
31. Bazin P-L et al. Denoising High-field Multi-dimensional MRI with Local Complex PCA. *Frontiers in neuroscience* 13, 1066 (2019). [PubMed: 31649500]
32. Lustig M, Donoho D. & Pauly JM Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine* 58, 1182–1195 (2007). [PubMed: 17969013]
33. Block KT, Uecker M. & Frahm J. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 57, 1086–1098 (2007).
34. Liang D, Liu B, Wang J. & Ying L. Accelerating SENSE using compressed sensing. *Magnetic Resonance in Medicine* 62, 1574–1584 (2009). [PubMed: 19785017]
35. Murphy M. et al. Fast ℓ_1 -SPIRiT Compressed Sensing Parallel Imaging MRI: Scalable Parallel Implementation and Clinically Feasible Runtime. *IEEE transactions on medical imaging* 31, 1250–1262 (2012). [PubMed: 22345529]
36. Otazo R, Kim D, Axel L. & Sodickson DK Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI. *Magnetic resonance in medicine* 64, 767–776 (2010). [PubMed: 20535813]
37. Hu Y. et al. Motion-robust reconstruction of multishot diffusion-weighted images without phase estimation through locally low-rank regularization. *Magnetic resonance in medicine* 81, 1181–1190 (2019). [PubMed: 30346058]
38. Hu Y. et al. Multi-shot diffusion-weighted MRI reconstruction with magnitude-based spatial-angular locally low-rank regularization (SPA-LLR). *Magnetic Resonance in Medicine* 83, 1596–1607 (2020). [PubMed: 31593337]
39. Kim TH, Bilgic B, Polak D, Setsompop K. & Haldar JP Wave-LORAKS: Combining wave encoding with structured low-rank matrix modeling for more highly accelerated 3D imaging. *Magnetic resonance in medicine* 81, 1620–1633 (2019). [PubMed: 30252157]
40. Haldar JP Low-rank modeling of local k-space neighborhoods (LORAKS) for constrained MRI. *IEEE Transactions on Medical Imaging* 33, 668–681 (2013).
41. Shin PJ et al. Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion. *Magnetic resonance in medicine* 72, 959–970 (2014). [PubMed: 24248734]

42. Haldar JP & Zhuo J.P-LORAKS: low-rank modeling of local k-space neighborhoods with parallel imaging data. *Magnetic resonance in medicine* 75, 1499–1514 (2016). [PubMed: 25952136]
43. Bilgic B.et al. Highly accelerated multishot echo planar imaging through synergistic machine learning and joint reconstruction. *Magnetic Resonance in Medicine* 82, 1343–1358, doi:10.1002/mrm.27813 (2019). [PubMed: 31106902]
44. Veraart J, Fieremans E.& Novikov DS Diffusion MRI noise mapping using random matrix theory. *Magnetic resonance in medicine* 76, 1582–1593 (2016). [PubMed: 26599599]
45. Veraart J.et al. Denoising of diffusion MRI using random matrix theory. *NeuroImage* 142, 394 (2016). [PubMed: 27523449]
46. Manjón JV et al. Diffusion Weighted Image Denoising Using Overcomplete Local PCA. *PLoS One* 8, e73021 (2013).
47. St-Jean S, Coupé P.& Descoteaux M.Non Local Spatial and Angular Matching: Enabling higher spatial resolution diffusion MRI datasets through adaptive denoising. *Medical image analysis* 32, 115–130 (2016). [PubMed: 27082655]
48. Fadnavis S, Batson J.& Garyfallidis E.Patch2Self: Denoising Diffusion MRI with Self-Supervised Learning. *Advances in Neural Information Processing Systems* 33, 16293–16303 (2020).
49. Haldar JP et al. Improved diffusion imaging through SNR-enhancing joint reconstruction. *Magnetic resonance in medicine* 69, 277–289 (2013). [PubMed: 22392528]
50. Varadarajan D.& Haldar JP A majorize-minimize framework for Rician and non-central chi MR images. *IEEE transactions on medical imaging* 34, 2191–2202 (2015). [PubMed: 25935028]
51. Sperl JI et al. Model-based denoising in diffusion-weighted imaging using generalized spherical deconvolution. *Magnetic resonance in medicine* 78, 2428–2438 (2017). [PubMed: 28244188]
52. Dong Z.et al. Model-based reconstruction for simultaneous multislice and parallel imaging accelerated multishot diffusion tensor imaging. *Medical physics* 45, 3196–3204 (2018). [PubMed: 29758101]
53. Zhang K, Zuo W, Chen Y, Meng D.& Zhang L.Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing* 26, 3142–3155 (2017). [PubMed: 28166495]
54. Weigert M.et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature methods* 15, 1090–1097 (2018). [PubMed: 30478326]
55. Krull A, Buchholz T-O & Jug F.Noise2void-learning denoising from single noisy images. *The IEEE Conference on Computer Vision and Pattern Recognition*, 2129–2137 (2019).
56. Devalla SK et al. A deep learning approach to denoise optical coherence tomography images of the optic nerve head. *Scientific reports* 9, 1–13 (2019). [PubMed: 30626917]
57. Gondara L.Medical image denoising using convolutional denoising autoencoders. *The IEEE International Conference on Data Mining Workshops (ICDMW)*, 241–246 (2016).
58. Chang Y, Yan L, Chen M, Fang H.& Zhong S.Two-stage convolutional neural network for medical noise removal via image decomposition. *IEEE Transactions on Instrumentation and Measurement* 69, 2707–2721 (2019).
59. Chen KT et al. Ultra-Low-Dose 18F-Florbetaben Amyloid PET Imaging Using Deep Learning with Multi-Contrast MRI Inputs. *Radiology* 290, 649–656 (2018). [PubMed: 30526350]
60. Ouyang J, Chen KT, Gong E, Pauly J.& Zaharchuk G.Ultra-low-dose PET reconstruction using generative adversarial network with feature matching and task-specific perceptual loss. *Medical physics* 46, 3555–3564 (2019). [PubMed: 31131901]
61. Wang Y.et al. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. *NeuroImage* 174, 550–562 (2018). [PubMed: 29571715]
62. Serrano-Sosa M, Spuhler K, DeLorenzo C.& Huang C.PET Image Denoising Using Structural MRI with a Novel Dilated Convolutional Neural Network. *Journal of Nuclear Medicine* 61, 434–434 (2020).
63. Xu J.et al. Ultra-low-dose 18F-FDG brain PET/MR denoising using deep learning and multi-contrast information. *The SPIE Medical Imaging Conference* 11313, 113131P (2020).

64. Gong E, Pauly JM, Wintermark M.& Zaharchuk G. Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. *Journal of Magnetic Resonance Imaging* 48, 330–340, doi:10.1002/jmri.25970 (2018). [PubMed: 29437269]
65. Benou A, Veksler R, Friedman A.& Raviv TR Ensemble of expert deep neural networks for spatio-temporal denoising of contrast-enhanced MRI sequences. *Medical image analysis* 42, 145–159 (2017). [PubMed: 28802145]
66. Jiang D. et al. Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural network. *Japanese journal of radiology* 36, 566–574 (2018). [PubMed: 29982919]
67. Pierrick C. MRI denoising using Deep Learning and Non-local averaging. arXiv preprint arXiv:1911.04798 (2019).
68. Gong E. et al. Deep learning and multi-contrast based denoising for low-SNR Arterial Spin Labeling (ASL) MRI. *SPIE Medical Imaging Conference* 11313, 113130M (2020).
69. Kidoh M. et al. Deep learning based noise reduction for brain MR imaging: tests on phantoms and healthy volunteers. *Magnetic Resonance in Medical Sciences* 19, 195 (2019). [PubMed: 31484849]
70. Tian Q. et al. DeepDTI: High-fidelity six-direction diffusion tensor imaging using deep learning. *NeuroImage* 219, 117017 (2020).
71. Tian Q. et al. Improved cortical surface reconstruction using sub-millimeter resolution MPRAGE by image denoising. *NeuroImage* 233, 117946 (2021). [PubMed: 33711484]
72. Li H. et al. Deep learning diffusion tensor imaging with accelerated q-space acquisition. In *Proceedings The Machine Learning (Part II) Workshop of the International Society for Magnetic Resonance in Medicine*, Washington, District of Columbia, USA (2018).
73. Gong T. et al. Efficient reconstruction of diffusion kurtosis imaging based on a hierarchical convolutional neural network. In *Proceedings of the 26th Annual Meeting of the International Society for Magnetic Resonance in Medicine (ISMRM)*, Paris, France, 1653 (2018).
74. Aliotta E, Nourzadeh H, Sanders J, Muller D.& Ennis DB Highly accelerated, model-free diffusion tensor MRI reconstruction using neural networks. *Medical physics* 46, 1581–1591 (2019). [PubMed: 30677141]
75. Li Z. et al. Fast and Robust Diffusion Kurtosis Parametric Mapping Using a Three-dimensional Convolutional Neural Network. *IEEE Access* 7, 71398–71411 (2019).
76. Golkov V. et al. q-Space deep learning: twelve-fold shorter and model-free diffusion MRI scans. *IEEE transactions on medical imaging* 35, 1344–1351 (2016). [PubMed: 27071165]
77. Golkov V. et al. q-Space Deep Learning for Twelve-Fold Shorter and Model-Free Diffusion MRI Scans. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention* 9349, 37–44 (2015).
78. Gibbons EK et al. Simultaneous NODDI and GFA parameter map generation from subsampled q-space imaging using deep learning. *Magnetic Resonance in Medicine* 81, 2399–2411 (2018). [PubMed: 30426558]
79. Lin Z. et al. Fast learning of fiber orientation distribution function for MR tractography using convolutional neural network. *Medical physics* 46, 3101–3116. (2019). [PubMed: 31009085]
80. Skare S, Hedehus M, Moseley ME & Li T-Q Condition number as a measure of noise performance of diffusion tensor data acquisition schemes with MRI. *Journal of Magnetic Resonance* 147, 340–352 (2000). [PubMed: 11097823]
81. Jones D, Horsfield M.& Simmons A. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 42 (1999).
82. Falk T. et al. U-Net: deep learning for cell counting, detection, and morphometry. *Nature Methods* 16, 67–70, doi:10.1038/s41592-018-0261-2 (2019). [PubMed: 30559429]
83. Kim J, Kwon Lee J.& Mu Lee K. Accurate image super-resolution using very deep convolutional networks. *The IEEE Conference on Computer Vision and Pattern Recognition*, 1646–1654 (2016).
84. Huang G, Liu Z, Van Der Maaten L.& Weinberger KQ Densely connected convolutional networks. *IEEE Conference on Computer Vision and Pattern Recognition*, 4700–4708 (2017).

85. Glasser MF et al. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* 80, 105–124 (2013). [PubMed: 23668970]
86. Ugurbil K. et al. Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *NeuroImage* 80, 80–104 (2013). [PubMed: 23702417]
87. Sotiropoulos SN et al. Advances in diffusion MRI acquisition and processing in the Human Connectome Project. *NeuroImage* 80, 125–143 (2013). [PubMed: 23702418]
88. Caruyer E, Lenglet C, Sapiro G. & Deriche R. Design of multishell sampling schemes with uniform coverage in diffusion MRI. *Magnetic Resonance in Medicine* 69, 1534–1540 (2013). [PubMed: 23625329]
89. Smith SM et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23, S208–S219 (2004). [PubMed: 15501092]
90. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW & Smith SM FSL. *NeuroImage* 62, 782–790 (2012). [PubMed: 21979382]
91. Andersson JL, Skare S. & Ashburner J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage* 20, 870–888 (2003). [PubMed: 14568458]
92. Andersson JL & Sotiropoulos SN An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage* 125, 1063–1078 (2016). [PubMed: 26481672]
93. Fischl B. FreeSurfer. *NeuroImage* 62, 774–781 (2012). [PubMed: 22248573]
94. Dale AM, Fischl B. & Sereno MI Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage* 9, 179–194 (1999). [PubMed: 9931268]
95. Harms MP et al. Extending the Human Connectome Project across ages: Imaging protocols for the Lifespan Development and Aging projects. *NeuroImage* 183, 972–984 (2018). [PubMed: 30261308]
96. Bookheimer SY et al. The Lifespan Human Connectome Project in Aging: An overview. *NeuroImage* 185, 335–348 (2019). [PubMed: 30332613]
97. van der Kouwe AJ, Benner T, Salat DH & Fischl B. Brain morphometry with multiecho MPRAGE. *NeuroImage* 40, 559–569 (2008). [PubMed: 18242102]
98. Greve DN & Fischl B. Accurate and robust brain image alignment using boundary-based registration. *NeuroImage* 48, 63–72 (2009). [PubMed: 19573611]
99. Jones DK The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: a Monte Carlo study†. *Magnetic Resonance in Medicine* 51, 807–815 (2004). [PubMed: 15065255]
100. Kingma DP & Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
101. He K, Zhang X, Ren S. & Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *The IEEE International Conference on Computer Vision*, 1026–1034 (2015).
102. Wang Z, Bovik AC, Sheikh HR & Simoncelli EP Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 600–612 (2004). [PubMed: 15376593]
103. Garyfallidis E. et al. Dipy, a library for the analysis of diffusion MRI data. *Frontiers in neuroinformatics* 8 (2014).
104. He K, Zhang X, Ren S. & Sun J. Deep residual learning for image recognition. *The IEEE conference on computer vision and pattern recognition*, 770–778 (2016).
105. Pham C-H et al. Multiscale brain MRI super-resolution using deep 3D convolutional networks. *Computerized Medical Imaging and Graphics* 77, 101647 (2019).
106. Lu H, Jensen JH, Ramani A. & Helpert JA Three-dimensional characterization of non-gaussian water diffusion in humans using diffusion kurtosis imaging. *NMR in Biomedicine* 19, 236–247 (2006). [PubMed: 16521095]
107. Wu EX & Cheung MM MR diffusion kurtosis imaging for neural tissue characterization. *NMR in Biomedicine* 23, 836–848 (2010). [PubMed: 20623793]

108. Assaf Y.& Basser PJ Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. *NeuroImage* 27, 48–58 (2005). [PubMed: 15979342]
109. Tournier JD, Calamante F, Gadian DG & Connelly A. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage* 23, 1176–1185 (2004). [PubMed: 15528117]
110. Tournier JD, Calamante F.& Connelly A. Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage* 35, 1459–1472 (2007). [PubMed: 17379540]
111. Behrens T. et al. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic resonance in medicine* 50, 1077–1088 (2003). [PubMed: 14587019]
112. Behrens T. et al. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature neuroscience* 6, 750–757 (2003). [PubMed: 12808459]
113. Jbabdi S, Sotiropoulos SN, Savio AM, Graña M.& Behrens TE Model-based analysis of multishell diffusion MR data for tractography: How to get over fitting problems. *Magnetic Resonance in Medicine* 68, 1846–1855 (2012). [PubMed: 22334356]
114. Zhang H, Schneider T, Wheeler-Kingshott CA & Alexander DC NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *NeuroImage* 61, 1000–1016 (2012). [PubMed: 22484410]
115. Fieremans E, Jensen JH & Helpert JA White matter characterization with diffusional kurtosis imaging. *NeuroImage* 58, 177–188 (2011). [PubMed: 21699989]
116. Fieremans E. et al. Novel white matter tract integrity metrics sensitive to Alzheimer disease progression. *American Journal of Neuroradiology* 34, 2105–2112 (2013). [PubMed: 23764722]
117. Chang LC, Jones DK & Pierpaoli C. RESTORE: robust estimation of tensors by outlier rejection. *Magnetic Resonance in Medicine* 53, 1088–1095 (2005). [PubMed: 15844157]
118. Maier-Hein KH et al. The challenge of mapping the human connectome based on diffusion tractography. *Nature communications* 8, 1349 (2017).

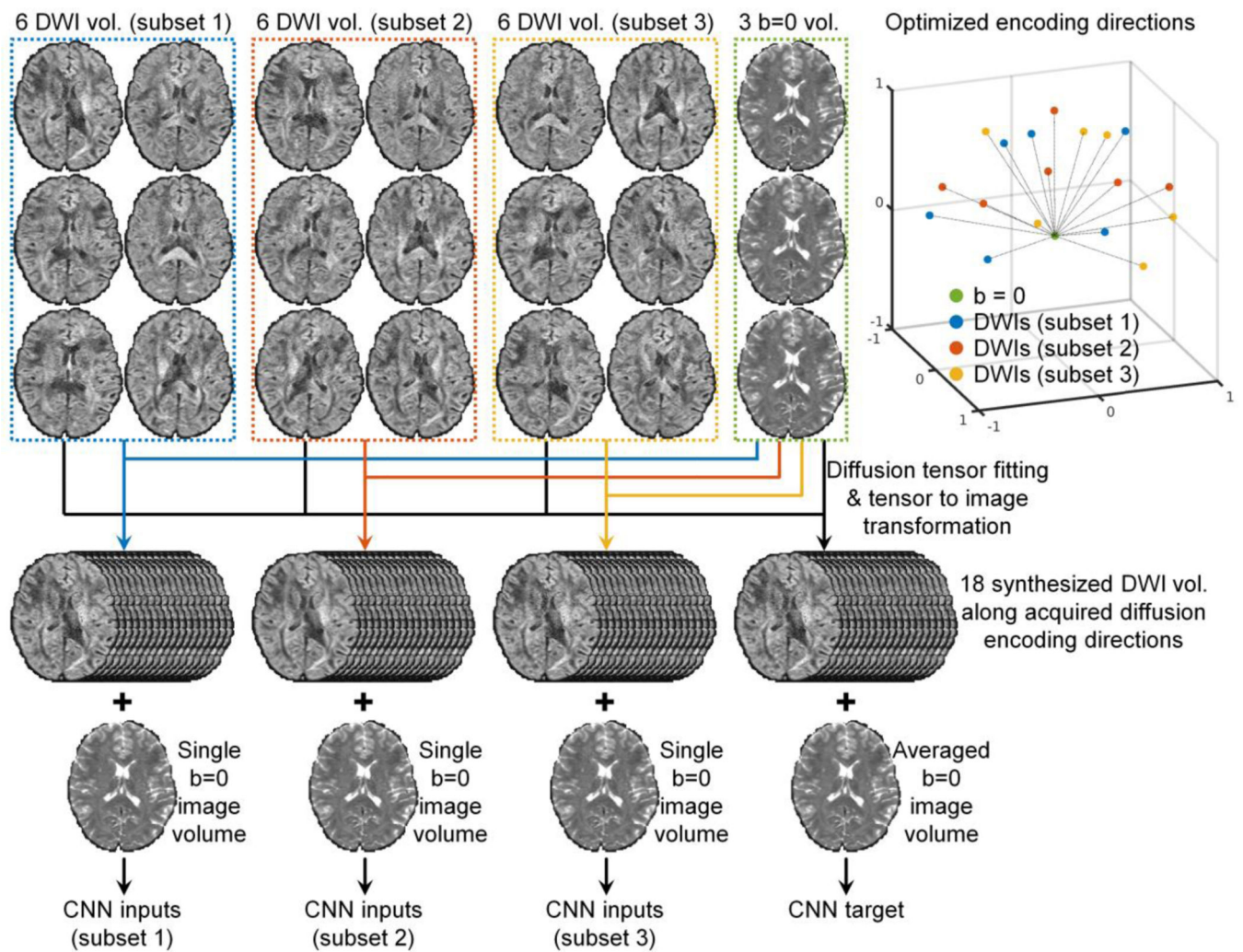


Figure 1. SDnDTI pipeline.

The SDnDTI pipeline for a DTI acquisition consisting of three $b = 0$ image volumes and 18 diffusion-weighted image volumes is demonstrated.

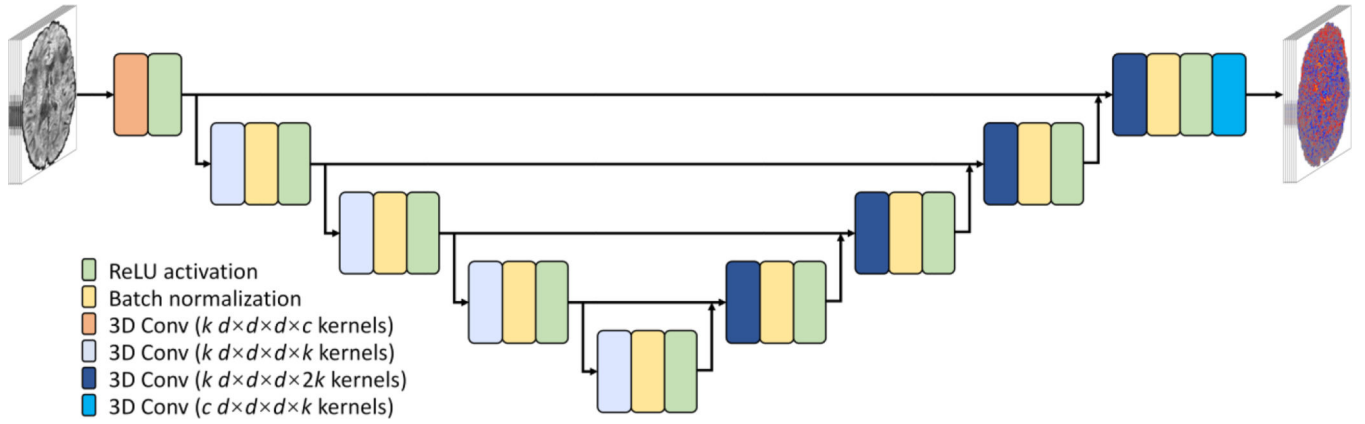


Figure 2. Modified U-Net (MU-Net) architecture.

MU-Net is modified from U-Net by removing all max pooling and up-sampling layers and keeping the number of kernels constant across all layers. The input is c noisy image volumes (one $b = 0$ image and $c - 1$ diffusion-weighted image volumes). The output is c residual images volumes between the input noisy image volumes and high-quality target image volumes. Network parameters of $k = 192$ and $d = 3$ were adopted in this study.

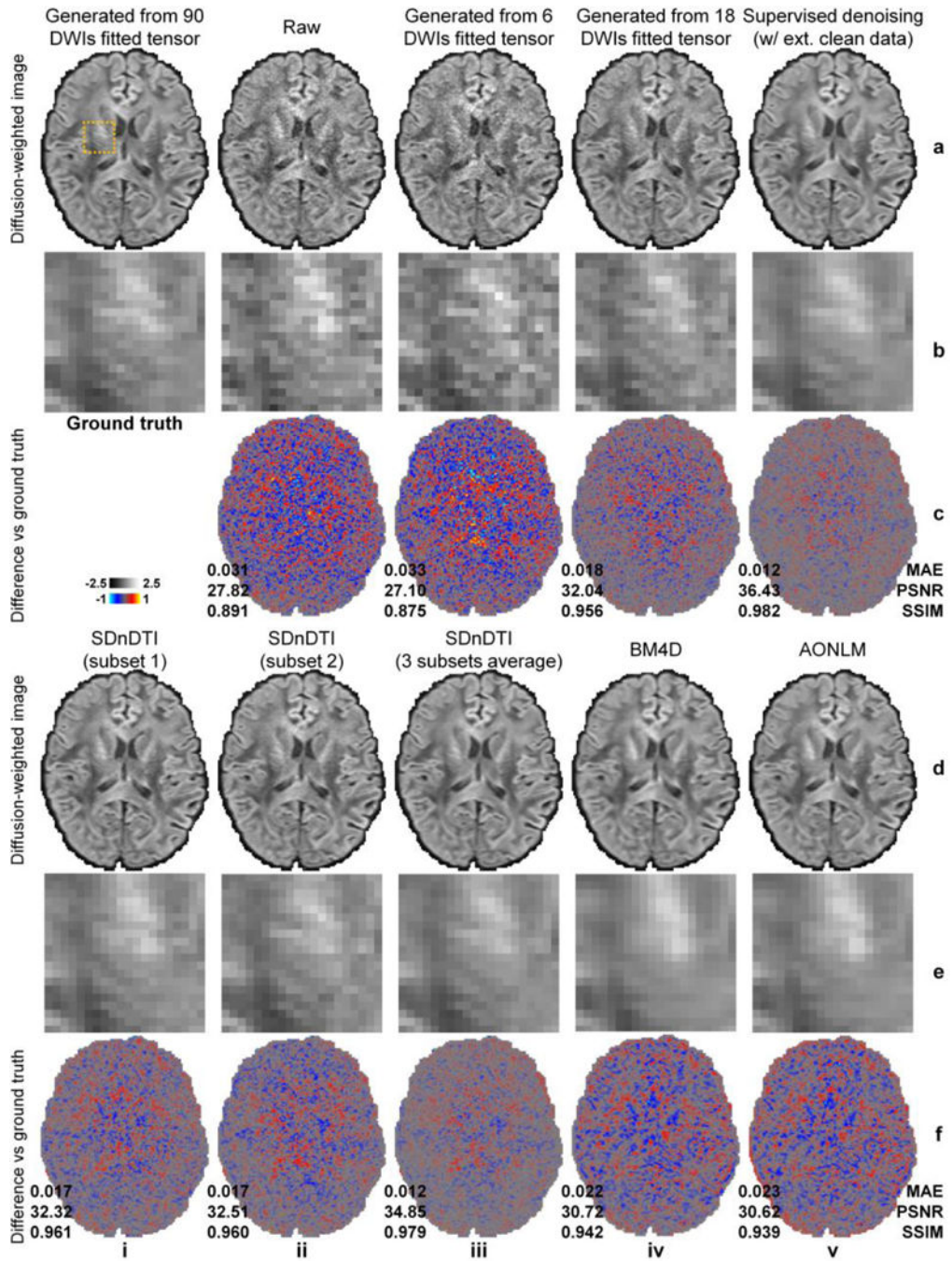


Figure 3. Denoised images.

Diffusion-weighted images (DWIs) sampled approximately along the superior-inferior direction (i.e., $[-0.18, 0.26, -0.95]$) of a representative HCP subject from the ground-truth data (i.e., synthesized from the diffusion tensor fitted using all 18 $b = 0$ images and 90 DWIs) (a, i), subset 1 of SDnDTI input data (i.e., raw acquired image) (a, ii), subset 2 of SDnDTI input data (i.e., synthesized image using the diffusion tensor fitted using three $b = 0$ images and six DWIs) (a, iii), synthesized data from the diffusion tensor fitted using three $b = 0$ images and 18 DWIs (a, iv), supervised learning denoised data using MU-Net

with the ground-truth data as the training target (a, v), subset 1 of SDnDTI-denoised data (d, i) (i.e., the raw DWI (a, ii) denoised by SDnDTI), subset 2 of SDnDTI-denoised data (d, ii) (i.e., the synthesized DWI (a, iii) denoised by SDnDTI), the average of all three subsets of SDnDTI-denoised data (d, iii), BM4D-denoised data (d, iv), and AONLM-denoised data (d, v), along with a region of interest in the deep white matter with fine textures (yellow box in a, i) displayed in enlarged views (rows b, e) and residual images compared to the ground-truth DWI (rows c, f). The mean absolute error (MAE), peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) of different images compared to the ground-truth DWI are used to quantify image similarity compared to the ground truth.

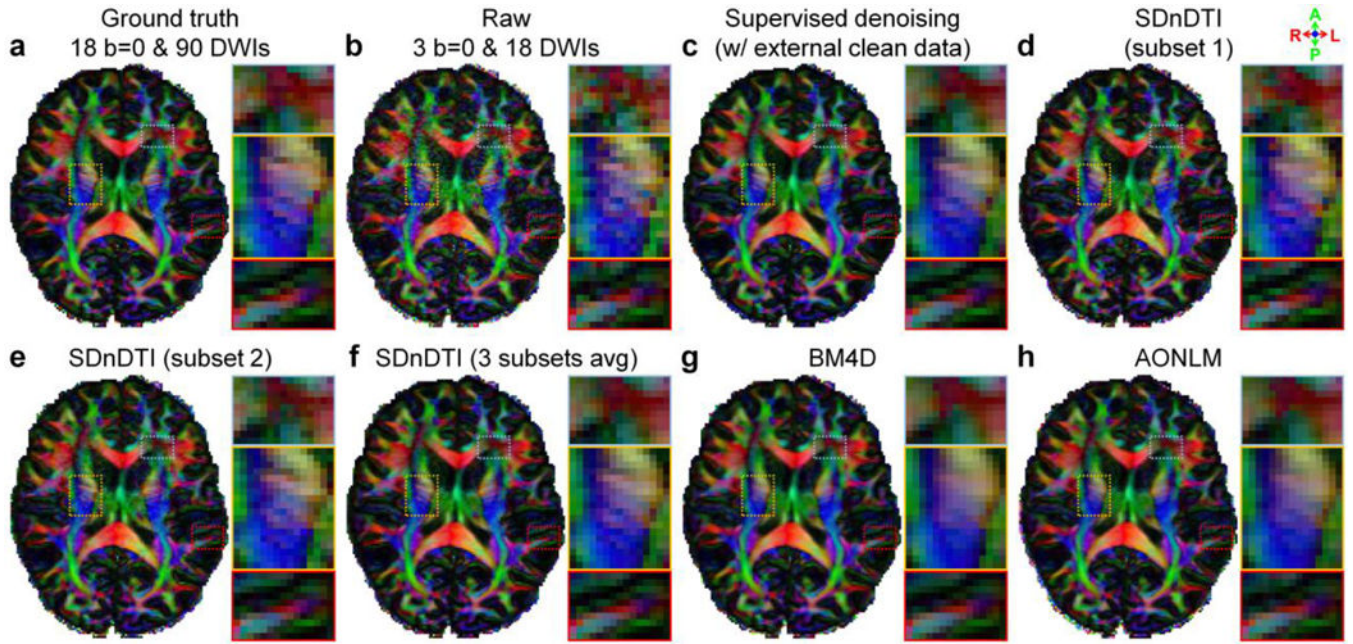


Figure 5. Structure mapping.

Fractional anisotropy (FA) maps color encoded by the primary eigenvector (red: left–right; green: anterior–posterior; blue: superior–inferior) derived from the diffusion tensors fitted using all 18 $b = 0$ and 90 diffusion-weighted images (DWIs) (ground truth, a), raw data consisting of three $b = 0$ and 18 DWIs (b), the raw data denoised by supervised learning with the ground-truth images as the training target (i.e., supervised denoising) (c), the subset 1 of SDnDTI-denoised data (d), the subset 2 of SDnDTI-denoised data (e), the average of all three subsets of SDnDTI-denoised data (f), and the raw data denoised by BM4D (g) and AONLM (h) from a representative HCP subject. Three regions of interest in the deep white matter (yellow boxes) and sub-cortical white matter surrounded by gray matter (red boxes) or with intersecting fiber tracts (blue boxes) are displayed in enlarged views.

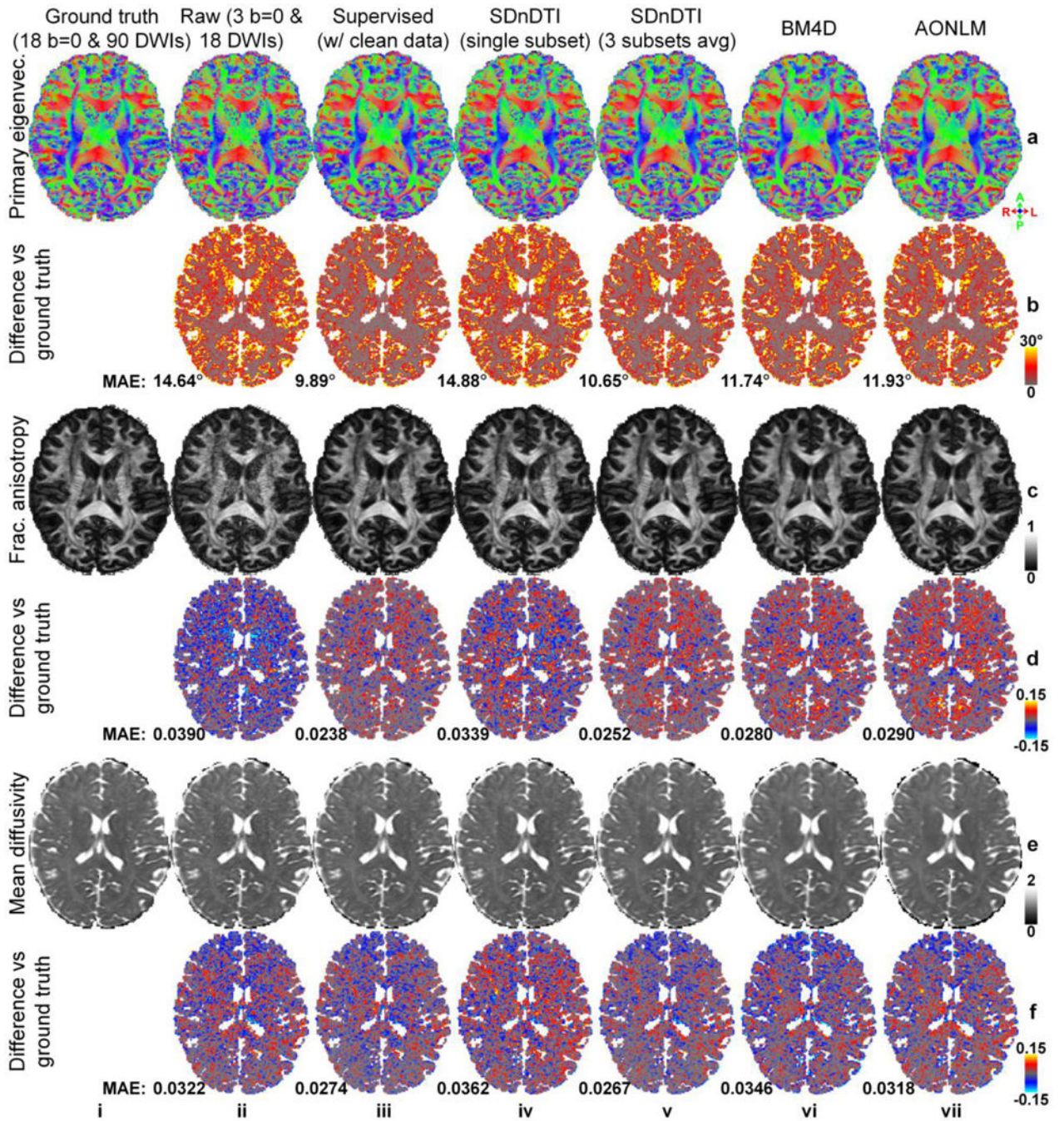


Figure 6. DTI metrics.

Maps of color-encoded primary eigenvector (red: left–right; green: anterior–posterior; blue: superior–inferior) (row a), fractional anisotropy (row c) and mean diffusivity (row e) derived from the diffusion tensors fitted using all 18 $b = 0$ and 90 diffusion-weighted images (DWIs) (ground truth, column i), raw data consisting of three $b = 0$ and 18 DWIs (column ii), the raw data denoised by supervised learning with the ground-truth images as the training target (i.e., supervised denoising) (column iii), the subset 2 of SDnDTI-denoised data (column iv), the average of all three subsets of SDnDTI-denoised data (column v), and the raw data

denoised by BM4D (column vi) and AONLM (column vi), and their residual maps (rows b, d, f) compared to the ground-truth maps from a representative HCP subject. The mean absolute error (MAE) of each map compared to the ground truth within the brain (excluding the cerebrospinal fluid) is displayed at the bottom of the residual map. The unit of the diffusivity is $\mu\text{m}^2/\text{ms}$.

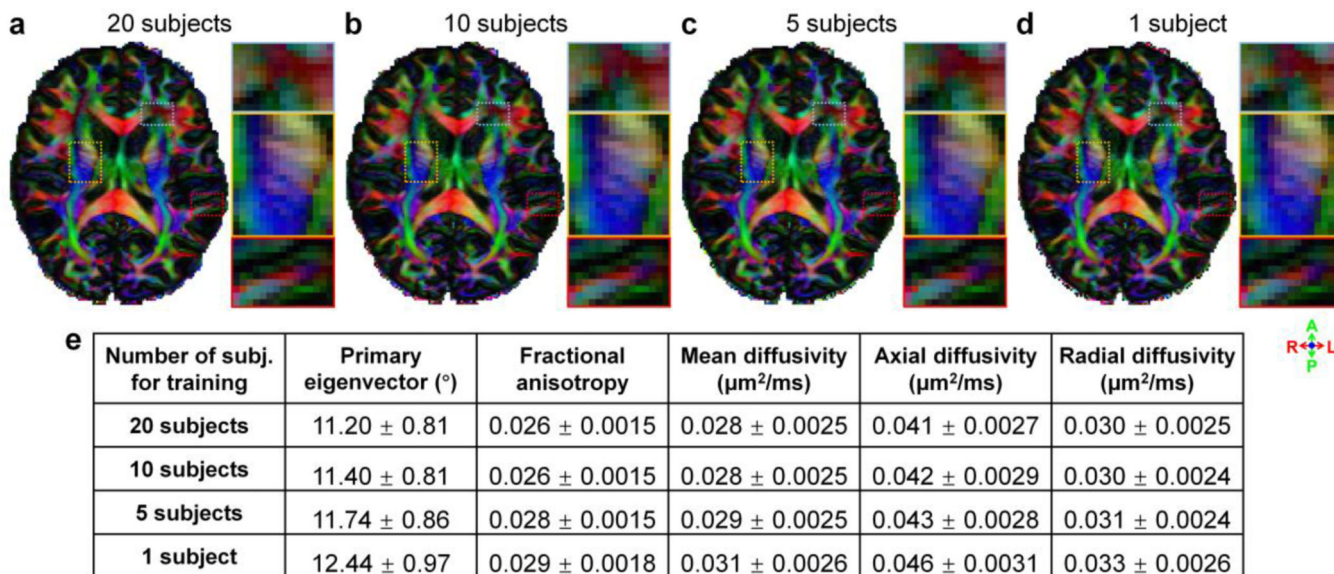


Figure 7. Effects of training data.

Fractional anisotropy maps color encoded by the primary eigenvector (red: left–right; green: anterior–posterior; blue: superior–inferior) derived from the raw data consisting of 3 $b = 0$ and 18 DWIs denoised by SDnDTI trained using data of 20 subjects (a), 10 subjects (b), 5 subjects (c) and 1 subject (d) from the HCP. Three regions of interest in the deep white matter (yellow boxes) and sub-cortical white matter surrounded by gray matter (red boxes) or with intersecting fiber tracts (blue boxes) are displayed in enlarged views. The table lists the group mean (\pm group standard deviation) across the 20 HCP subjects of the mean absolute error (MAE) of different DTI metrics compared to the ground-truth DTI metrics derived from 18 $b = 0$ and 90 DWIs.

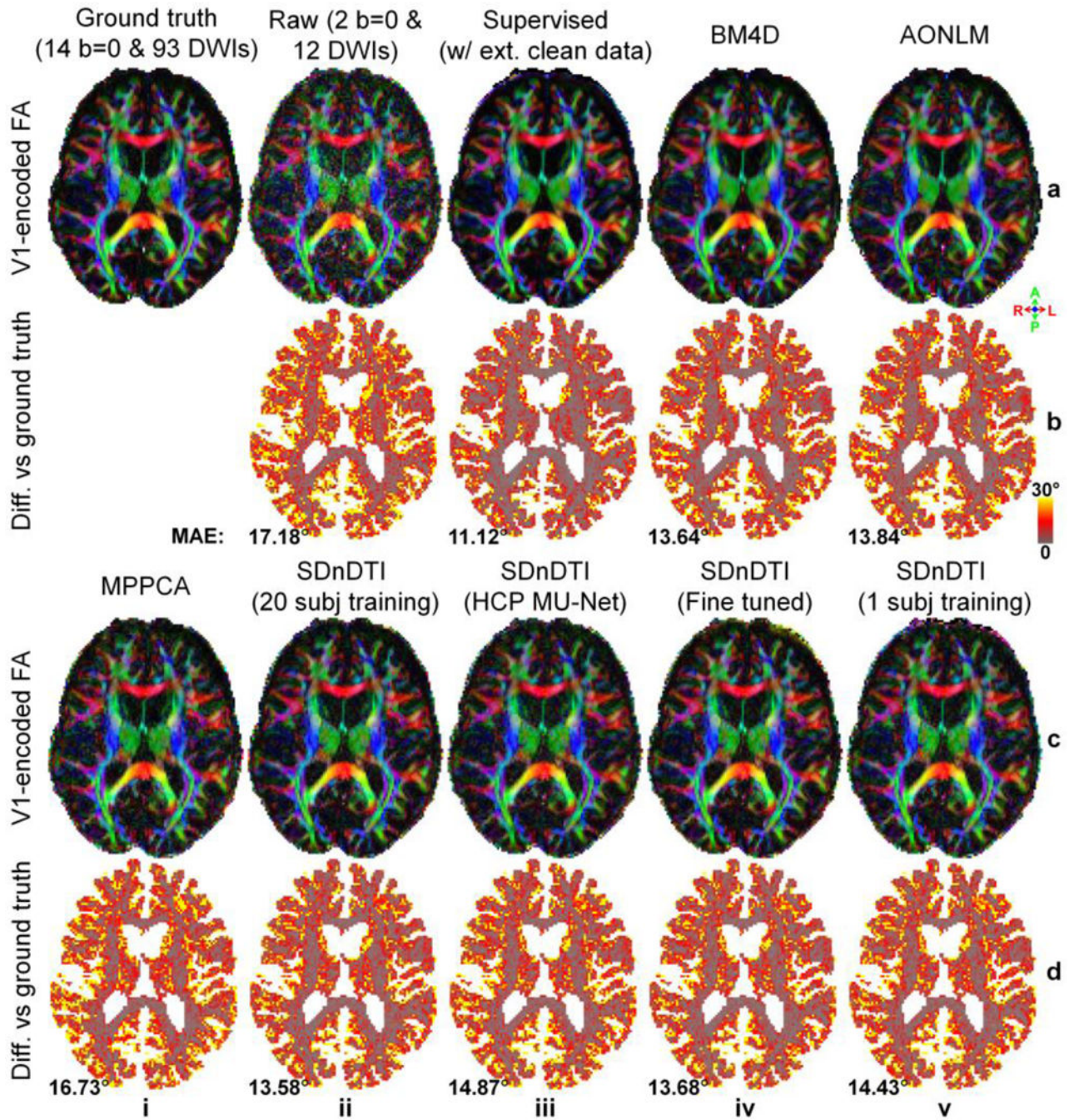


Figure 8. Primary eigenvector.

Maps of color-encoded primary eigenvector (red: left–right; green: anterior–posterior; blue: superior–inferior) modulated by the fractional anisotropy (row a, c) derived from the diffusion tensors fitted using all 14 $b = 0$ and 93 diffusion-weighted images (DWIs) (ground truth, a, i), raw data consisting of two $b = 0$ and 12 DWIs (a, ii), the raw data denoised by supervised learning with the ground-truth DWIs as the training target (i.e., supervised denoising) (a, iii), BM4D (a, iv), AONLM (a, v) and MPPCA (c, i), and SDnDTI (c, ii–v), and their residual maps (rows b, d) compared to the ground-truth map from a representative

HCP-A subject. SDnDTI results were generated by an MU-Net trained on the data from 20 HCP-A subjects (c, ii), an MU-Net trained on the data from 20 HCP subjects (c, iii), an MU-Net with parameters from the MU-Net trained on the data from 20 HCP subjects as initialization and further fine-tuned using the data of each HCP-A subject (c, iv), and an MU-Net trained on the data from the data of each HCP-A subject (c, v). The mean absolute error (MAE) of each map compared to the ground truth within the brain (excluding the cerebrospinal fluid) is displayed at the bottom of the residual map.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

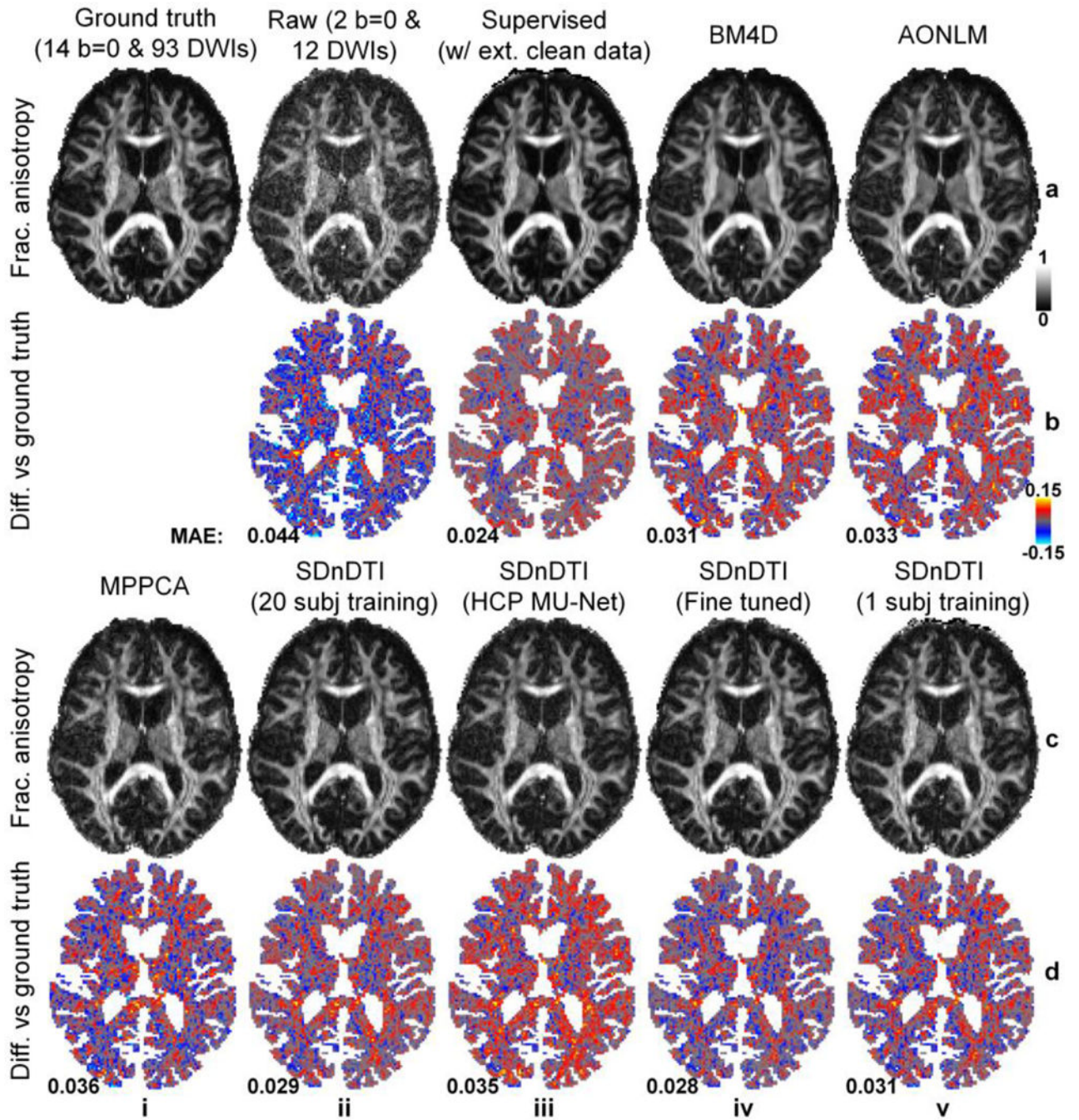


Figure 9. Fractional anisotropy.

Maps of fractional anisotropy (row a, c) derived from the diffusion tensors fitted using all 14 $b = 0$ and 93 diffusion-weighted images (DWIs) (ground truth, a, i), raw data consisting of two $b = 0$ and 12 DWIs (a, ii), the raw data denoised by supervised learning with the ground-truth DWIs as the training target (i.e., supervised denoising) (a, iii), BM4D (a, iv), AONLM (a, v) and MPPCA (c, i), and SDnDTI (c, ii–v), and their residual maps (rows b, d) compared to the ground-truth map from a representative HCP-A subject. SDnDTI results were generated by an MU-Net trained on the data from 20 HCP-A subjects (c, ii), an

MU-Net trained on the data from 20 HCP subjects (c, iii), an MU-Net with parameters from the MU-Net trained on the data from 20 HCP subjects as initialization and further fine-tuned using the data of each HCP-A subject (c, iv), and an MU-Net trained on the data from the data of each HCP-A subject (c, v). The mean absolute error (MAE) of each map compared to the ground truth within the brain (excluding the cerebrospinal fluid) is displayed at the bottom of the residual map.

Table 1.**Image similarity.**

The group mean (\pm group standard deviation) across the 20 HCP subjects of the mean absolute error (MAE), peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) of $b = 0$ images and the diffusion-weighted images (DWI, along the same diffusion-encoding direction as the DWI shown in Figure 3) from the raw data (a), each subset of SDnDTI input data (b–d) (a single raw $b = 0$ image or DWIs synthesized from the diffusion tensor fitted using three $b = 0$ images and six DWIs), the averaged $b = 0$ image or DWIs synthesized from the diffusion tensor fitted using three $b = 0$ images and 18 DWIs (f), the raw data denoised by supervised learning with the ground-truth images as the training target (i.e., supervised-denoising) (f), each subset of SDnDTI-denoised data (g–i), the average of all three subsets of SDnDTI-denoised data (j), and the raw data denoised by BM4D (k) and AONLM (l) compared to the ground-truth images generated from the tensor fitted using all 18 $b = 0$ and 90 DWIs.

		a	b	c	d	e	f	g	h	i	j	k	l
		Raw	Synth. (subset 1)	Synth. (subset 2)	Synth. (subset 3)	Synth. (18 DWIs)	Supervised	SDnDTI (subset 1)	SDnDTI (subset 2)	SDnDTI (subset 3)	SDnDTI (average)	BM4D	AONLWI
MAE	b = 0		0.017 \pm 0.00097	0.015 \pm 0.00098	0.015 \pm 0.00088	0.011 \pm 0.00087	0.0099 \pm 0.00073	0.014 \pm 0.00091	0.013 \pm 0.00087	0.013 \pm 0.00078	0.01 \pm 0.00082	0.014 \pm 0.00063	0.013 \pm 0.0007
	DWI	0.033 \pm 0.0021	0.034 \pm 0.0021	0.035 \pm 0.0021	0.035 \pm 0.0022	0.019 \pm 0.0012	0.012 \pm 0.00072	0.018 \pm 0.0011	0.018 \pm 0.00099	0.018 \pm 0.0011	0.012 \pm 0.00073	0.023 \pm 0.0012	0.024 \pm 0.0012
PSNR (dB)	b = 0		32.19 \pm 0.62	32.94 \pm 0.58	33.41 \pm 0.6	35.31 \pm 0.75	36.54 \pm 0.76	33.18 \pm 0.69	34.00 \pm 0.6	34.35 \pm 0.63	35.85 \pm 0.77	33.67 \pm 0.46	34.12 \pm 0.55
	DWI	27.13 \pm 0.57	26.89 \pm 0.61	26.34 \pm 0.61	26.45 \pm 0.71	31.55 \pm 0.8	35.83 \pm 0.54	31.87 \pm 0.86	31.68 \pm 0.93	31.70 \pm 1.09	34.50 \pm 1.32	30.23 \pm 0.47	30.16 \pm 0.46
SSIM	b = 0		0.96 \pm 0.0042	0.97 \pm 0.004	0.97 \pm 0.0039	0.99 \pm 0.0023	0.99 \pm 0.0019	0.98 \pm 0.0029	0.98 \pm 0.0025	0.98 \pm 0.0024	0.99 \pm 0.0018	0.98 \pm 0.002	0.98 \pm 0.0022
	DWI	0.88 \pm 0.013	0.88 \pm 0.013	0.86 \pm 0.013	0.87 \pm 0.013	0.95 \pm 0.0054	0.98 \pm 0.0023	0.96 \pm 0.0047	0.96 \pm 0.0047	0.96 \pm 0.0047	0.98 \pm 0.0024	0.94 \pm 0.0056	0.94 \pm 0.0058

Table 2.**Errors of DTI metrics.**

The group mean (\pm group standard deviation) across the 20 HCP subjects of the mean absolute error (MAE) of different DTI metrics derived from the raw data consisting of three $b = 0$ and 18 diffusion-weighted images (DWIs) (a), each subset of SDnDTI input data (b–d), the raw data denoised by supervised learning with the ground-truth images as the training target (i.e., supervised-denoising) (e), each subset of SDnDTI-denoised data (f–h), the average of all three subsets of SDnDTI-denoised data (i), and the raw data denoised by BM4D (j) and AONLM (k) compared to the ground-truth DTI metrics derived from 18 $b = 0$ and 90 DWIs.

HCP Data		Primary eigenvector ($^{\circ}$)	Fractional anisotropy	Mean diffusivity ($\mu\text{m}^2/\text{ms}$)	Axial diffusivity ($\mu\text{m}^2/\text{ms}$)	Radial diffusivity ($\mu\text{m}^2/\text{ms}$)
a	Raw	15.48 \pm 1.05	0.042 \pm 0.0031	0.033 \pm 0.0028	0.057 \pm 0.0041	0.037 \pm 0.003
b	Raw (subset 1)	24.90 \pm 1.25	0.086 \pm 0.0069	0.043 \pm 0.0030	0.10 \pm 0.0071	0.056 \pm 0.0044
c	Raw (subset 2)	24.92 \pm 1.32	0.086 \pm 0.0067	0.043 \pm 0.0032	0.10 \pm 0.0075	0.056 \pm 0.0043
d	Raw (subset 3)	24.86 \pm 1.34	0.085 \pm 0.0069	0.043 \pm 0.0033	0.10 \pm 0.0079	0.056 \pm 0.0043
e	Supervised	10.52 \pm 0.75	0.025 \pm 0.0014	0.028 \pm 0.0023	0.040 \pm 0.0025	0.030 \pm 0.0024
f	SDnDTI (subset 1)	15.73 \pm 1.05	0.036 \pm 0.0021	0.039 \pm 0.0029	0.058 \pm 0.0034	0.042 \pm 0.0030
g	SDnDTI (subset 2)	15.99 \pm 1.11	0.036 \pm 0.0021	0.037 \pm 0.0027	0.057 \pm 0.0034	0.041 \pm 0.0027
h	SDnDTI (subset 3)	15.80 \pm 1.09	0.035 \pm 0.0019	0.037 \pm 0.0027	0.056 \pm 0.0033	0.040 \pm 0.0027
i	SDnDTI (average)	11.20 \pm 0.81	0.026 \pm 0.0015	0.028 \pm 0.0025	0.041 \pm 0.0027	0.030 \pm 0.0025
j	BM4D	12.46 \pm 0.89	0.029 \pm 0.0016	0.036 \pm 0.0027	0.051 \pm 0.0031	0.037 \pm 0.0026
k	AONLM	12.64 \pm 0.87	0.030 \pm 0.0016	0.033 \pm 0.0027	0.050 \pm 0.0032	0.035 \pm 0.0026

Table 3.**Errors of DTI metrics.**

The group mean (\pm group standard deviation) across the 20 HCP-A subjects of the mean absolute error (MAE) of different DTI metrics derived from the raw data consisting of two $b = 0$ and 12 diffusion-weighted images (DWIs) (a), the raw data denoised by supervised learning with the ground-truth DWIs as the training target (i.e., supervised denoising) (b), BM4D (c), AONLM (d), MPPCA (e), and SDnDTI (f–i) compared to the ground-truth DTI metrics derived from 14 $b = 0$ and 93 DWIs. SDnDTI results were generated by an MU-Net trained on the data from 20 HCP-A subjects (f), an MU-Net trained on the data from 20 HCP subjects (g), an MU-Net with parameters from the MU-Net trained on the data from 20 HCP subjects as initialization and further fine-tuned using the data of each HCP-A subject (h), and an MU-Net trained on the data from the data of each HCP-A subject (i).

HCP-A data		Primary eigenvector (°)	Fractional anisotropy	Mean diffusivity ($\mu\text{m}^2/\text{ms}$)	Axial diffusivity ($\mu\text{m}^2/\text{ms}$)	Radial diffusivity ($\mu\text{m}^2/\text{ms}$)
a	Raw	16.17 \pm 0.72	0.043 \pm 0.0029	0.025 \pm 0.0023	0.054 \pm 0.0035	0.030 \pm 0.0024
b	Supervised	10.14 \pm 0.56	0.023 \pm 0.00092	0.020 \pm 0.0023	0.032 \pm 0.0021	0.023 \pm 0.0023
c	BM4D	12.78 \pm 0.58	0.031 \pm 0.0014	0.032 \pm 0.0040	0.049 \pm 0.0040	0.034 \pm 0.0034
d	AONLM	13.04 \pm 0.61	0.033 \pm 0.0016	0.027 \pm 0.0027	0.046 \pm 0.0028	0.030 \pm 0.0023
e	MPPCA	15.71 \pm 0.69	0.036 \pm 0.0019	0.025 \pm 0.0024	0.047 \pm 0.0028	0.028 \pm 0.0024
f	SDnDTI (20 subj)	12.63 \pm 0.61	0.028 \pm 0.0014	0.022 \pm 0.0022	0.039 \pm 0.0021	0.024 \pm 0.0022
g	SDnDTI (HCP MU-Net)	14.06 \pm 0.58	0.036 \pm 0.0014	0.023 \pm 0.0023	0.045 \pm 0.0019	0.027 \pm 0.0022
h	SDnDTI (fine tuned)	12.66 \pm 0.70	0.028 \pm 0.0016	0.022 \pm 0.0023	0.039 \pm 0.0025	0.025 \pm 0.0023
i	SDnDTI (1 subj)	13.65 \pm 0.65	0.031 \pm 0.0015	0.024 \pm 0.0024	0.043 \pm 0.0025	0.027 \pm 0.0023