Can deep learning based super-resolution enhance the upsampling of diffusion MRI for brain fiber tractography?

Author	*Supervisor*	*Examiner*	*Second reviewer*
S.D. Bishas	A. De Luca	A. Leemans	M. Froeling
Utrecht University	UMC Utrecht	UMC Utrecht	UMC Utrecht
MSc Medical Imaging	Imaging Sciences Institute	Imaging Sciences Institute	Imaging Sciences Institute

***Abstract*—Diffusion MRI fiber tractography is the only technique that allows to investigate the orientation of white matter fibers non-invasively. The quality of fiber tractography critically depends on the underlying diffusion MRI quality, including imaging resolution. However, obtaining high-resolution diffusion MRI data is technically challenging, and results in longer scan times, lower signal-to-noise ratios, and limited spatial coverage. Super-resolution techniques have emerged as a promising method for improving image resolution by enhancing low-resolution images to high-resolution images after acquired dMRI data. In conventional super resolution, already acquired data is enhanced in resolution using, for examples, spatial filters. In recent years, deep learning techniques have also been suggested promising for super-resolution models that enhance the quality of diffusion MRI data. In this project, we investigated the combination of conventional upsampling techniques with a specific deep learning architecture, convolutional neural networks, to perform deep super resolution (DSR) of diffusion MRI, and evaluated its impact on subsequent fiber tractography. We considered 28 datasets from the Human Connectome Project acquired at 1.25mm^3 isotropic. Data was downsampled by k-space truncation to 2.5mm^3 isotropic, then re-upsampled to the original resolution using nearest neighbours, cubic b-spline and zero padding in k-space, respectively. The results of the upsampling were given as input to a DSR network with 10 layers. Different types of DSR inputs were considered, namely 2D patches, 2D whole slices, and 3D volumes. The combination of zero-padding and the DSR with a 2D patch approach produced the most optimal overall results with mean squared error, peak signal-to-noise ratio, and structural similarity index of 36.109 ± 9.24, 29.418 ± 0.723, and 0.929 ± 0.007 as compared to ZP only, respectively. This approach has been further investigated for fiber tractography to assess the reconstruction of white matter tracts. The results revealed minimal visible differences between fiber tractography with the original data, after downsampling and after DSR implementation. Overall, these results suggest that the tested DSR networks are promising to improve the spatial detail of diffusion MRI data, but are still insufficient to improve the quality of fiber tractography.**

***Index Terms*—Super-resolution, diffusion MRI, deep learning, convolutional neural network, diffusion tensor imaging, residual learning, fiber tractography.**

1. Introduction

The utility of magnetic resonance imaging (MRI) in medical imaging has gained widespread attention, particularly in brain analysis, due to its superior soft tissue contrast and the ability to acquire high spatial resolution with minimal health risks. Quantitative analysis of brain MRI has been used to detect brain disorders such as Alzheimer's, epilepsy, schizophrenia, multiple sclerosis, malignancy, and degenerative diseases [1].

Recently, the application of dMRI have shown significant advancements in this field [2]. dMRI measures the displacement of water molecules, which allows to infer the local fiber orientations of specific tissues and provide insights into the connectomics of the brain [3]. Fiber tractography, a process that visualizes the long-range connectivity throughout the brain, utilizes the fiber orientation information obtained from dMRI data to map the main white matter bundles [4]. This information is crucial in pre-operative neurosurgical treatment planning, as it evaluates brain functionality [2].

The quality of fiber tractography reconstructions in the brain has been shown to strongly depend on the imaging resolution of the underlying dMRI [5]. More generally, high-resolution MRI images are critical for precise identification of abnormal tissue structures and accurate quantitative image analysis [6]. However, obtaining high resolution dMRI data is challenging, as it requires longer scan times, lower signal-to-noise ratios, and limited spatial coverage [7]. While advances in MRI technology, such as increased field strength, have improved image quality and resolution in structural and functional imaging, such improvements have not yet systematically impacted the quality of these images.

Super-resolution (SR) techniques offer an alternative method for improving image resolution by enhancing low resolution (LR) images to high-resolution (HR) images [8] after their acquisition. The overall framework is visualized in Fig. 1. The initial step involves downsampling of the original HR data, which will subsequently undergo an upsampling process. Most common upsampling methods are cubic spline (CS) interpolation, nearest neighbor (NN) interpolation, and zero-padding (ZP) in k-space. These LR interpolated data

serves as a basis for reconstructing the HR image. SR approaches have the potential to increase the signal-to-noise ratio and spatial details of images while not affecting the scan time and do not require special adaptations. The challenge of image resolution enhancement is not new, and a number of traditional interpolation techniques can be used to increase the spatial coverage of LR images. However, these methods often fail to recover high-frequency details, such as edges. Deep learning based SR techniques have been suggested promising to overcome this limitation by learning specific image features on high quality datasets, which can be used to enhance conventionally acquired LR images [9], [10].

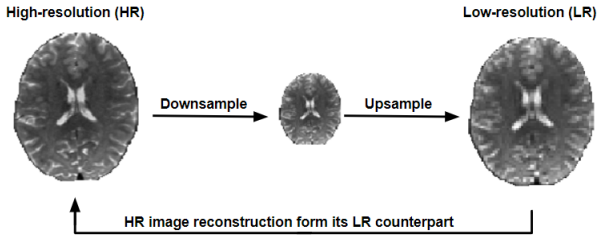

Fig. 1: Overall framework of super-resolution

.

The implementation of SR techniques poses significant challenges. The challenge of developing a mapping between the LR and HR image spaces in dMRI is compounded by the lack of a clear definition of this mapping. The complexity of this high-dimensional mapping is further addressed by the large amounts of raw data involved.

The aim of this study is to evaluate whether deep learning based SR (DSR) in combination with conventional upsampling approaches can improve the quality of dMRI data for fiber tractography. There is a large variety of promising deep learning approaches out there, but a sparse quantity in the context of dMRI. After a literature search, one network architecture was chosen as a starting point. The performance of this network was tested in combination with different interpolation methods as input data. Furthermore, the most promising approach was used for diffusion tensor imaging and fiber tractography analysis.

2. Theory

2.1. Fundamentals of diffusion MRI

To investigate complex fiber connections of the human brain, high-resolution dMRI becomes essential [3]. The dis placement of water molecules along different axes will be measured to reveal the fine fiber structures of the brain. Brain fibers (axons) are impermeable, which constrains the water molecules to move in their main direction.

In the field of brain dMRI, high resolution data is used for accurate analysis and fiber tracking, a technique that leverages the ability to construct fiber pathways [3]. Despite not being representative of actual pathways, these pathways offer insights into the directional relationships between fibers. The challenge of converting low resolution data into high resolution data in order to uncover previously unseen fiber pathways has led to studies exploring the use of deep learning techniques to supplement missing details [11], [12].

The principle of tensor based fiber tractography is based on the assumption of one dominant fiber orientation per voxel and its global fiber trajectories follows these local fiber orientations. This is illustrated in Fig. 2. Each local fiber orientation can be compromised as 3D vectors. By assuming one dominant orientation, the information is reduced to the first eigenvector. The accuracy and length of the global fiber trajectories depend on three criteria: the fractional anisotropy (FA), angle deviation as a predefined threshold and the step size. By altering the predefined thresholds, a trade off has to be made between the amount of false negative and false positive pathway propagation segments. Moreover, the computation time and data storage increases with smaller step sizes.

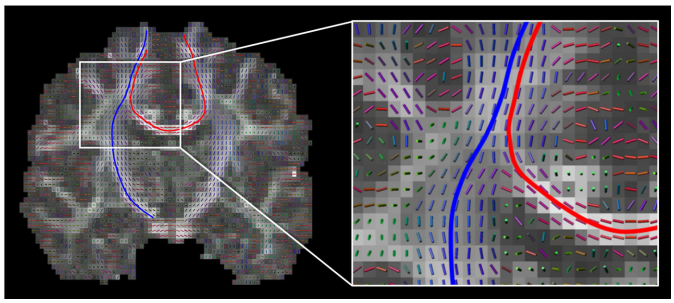

Fig. 2: Fibers constructed by fiber tractography. The concept of fiber tractography relies on the assumption that each voxel contains a single predominant fiber orientation and that the overall fiber trajectories follow these local fiber orientations.

.

In order to investigate the effects of imaging resolution on fiber tractography, Tian et al. [11] conducted a study which is presented in Figure 3. The study employed the raw low resolution dMRI data, which was obtained by adding noise. Due to the fluctuations in intensities within brain voxels, fewer fiber tracts could be generated by following the criteria when compared to the original ground-truth data.

2.2. A brief review of deep learning based super-resolution

In recent years, deep learning techniques have undergone rapid development. As a result, models based on deep learning for SR have been extensively studied and have shown remark able success [13]. The availability of large datasets has enabled these networks to attain high accuracy. Extracting specific features from raw images has been the fundamental approach of deep learning neural networks [14]. This type of machine learning uses its self-learning capability to extract a hierarchy of low-dimensional features [1]. Moreover, deep learning algorithms are able to generalize for large amounts of data. Different deep learning methods have been used to SR utilities, with the earliest being the Convolutional Neural Networks (CNN) based method, such as SRCNN [10]. SR algorithms using deep learning techniques typically differ in some aspects,

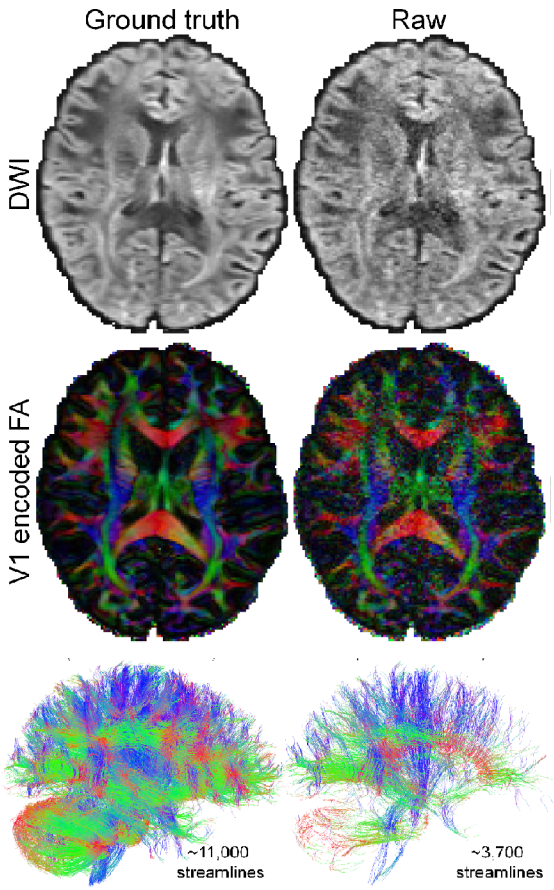

Fig. 3: Comparison between the raw low-resolution and ground-truth dMRI data. The raw dataset was intentionally corrupted with noise. This enables differences in the FA maps and number of streamlines generated through the use of fiber tractography.

such as network architectures, loss functions, learning principles and strategies. The optimization of networks involves adjusting hyper-parameters, including the number of layers, kernel sizes, epochs, early stopping and learning rate [15].

In 1989, CNNs were first introduced and are currently the most used deep learning architecture for image analysis. The complexity of CNNs remains in the amount of layers. Some approaches include more than 100 layers resulting in millions of weights and connections up to billions. The structure of the CNNs consists of convolutions layers, non-linearity layers, pooling (sub-sampling) layers and classification.

- Convolutional layer: extracts features to learn the relationship between pixels of the input images, by applying convolutional filters. At the first layer, simple features, such as edges or lines will be extracted. More complex features could be extracted further down the layers. The feature information will be stored in feature maps.
- Non-linearity layer: after each convolutional layer non linearity operations are applied, which operate as activation functions. It uses the feature maps for a sparse representation, in the sake of computational efficiency. Rectified linear units (ReLUs) are commonly used, being computational cheap, but lacking in its gradient discontinuity.
- Pooling (sub-sampling) layer: pooling layers modifies the sparse representation in such a way that it summarizes the statistics of non-overlapping neighborhoods to reduce parameters and risk of over-fitting.
- Loss function: the learning of the model consists of using applicable loss functions. These functions are used to enhance the models ability to discriminate between intra class similarity and inter-class separability.

Another promising network, which functions as a two player mini game, is the GAN [16]. The model contains a generator G and a discriminator D. The goal of G is to generate output data, which looks like it came from the training data. While, D has to decide if the output data is generated or picked from the training data.

Like the proposed SRGAN by Ledig et al. [17], G tries to generate HR images from LR input images. The generator itself is a CNN, which is trained to generate HR images and the discriminator is another neural network trying to differentiate the super-resoluted images from the HR images. It is crucial for the training procedure to have a balance between both players, so neither player can outperform one another. This will result in a stable network which is determined to acquire high quality images.

Tian et al. [11], [18] has recently proposed a deep learning based SR method, the SRDTI, to synthesize high-resolution diffusion MRI data for diffusion tensor imaging. It utilizes deep learning to derive the six unknowns in a diffusion tensor using a reduced amount of data. The framework maps non diffusion-weighted images, six diffusion-weighted images as input of the 3D CNN. The model utilizes the learning process of the residuals between the input data and ground-truth data. This approach enables residual learning to enhance CNN performance by adding the predicted residuals to the original input data. Thereby, acquiring high-quality data for DTI metrics and fiber tractography. The deep CNN employed takes advantage of the redundancy in spatial information and diffusion-encoding directions within the data to improve performance.

In the study by Luo et al. [19], a novel approach to dMRI SR was presented. The proposed method involves the utilization of a 3D convolution kernel for reconstruction in both the space and angle domains. Additionally, the study introduces an adversarial learning mechanism and an attention mechanism to address limitations of traditional loss functions in fully quantifying the differences between high-dimensional data and not considering important feature maps. The study demonstrates the implementation of a deep learning based dMRI reconstruction and the utilization of adversarial learning in combination with a 3D convolution kernel.

While network depth grows by increasing the amount of layers, feature information of various layers could get lost in the process [20]. This issue can be characterized by the vanishing gradient problem. Here, a significant amount of

layers uses activation functions, of which the gradients will approach zero and the training procedure will be hampered. Skip connections can be used connecting any two layers to preserve information. To exploit this concept, dense connections are introduced to connect layers to all forward layers. This solves the gradients vanishing problem and improves the propagation of feature information. Dense networks use a gate unit to learn block weights. Another approach is to concatenate feature maps of previous layers, hereby increasing the parameters exponentially.

The objective of this study is to assess the effectiveness of deep learning based methods in improving the quality of upsampling beyond that achieved by conventional methods. To achieve this, we decided to employ the relatively straighforward CNN architecture from Tian et al. [18] and analyze its impact on various interpolation techniques.

3. Methods

3.1. Data acquisition

Diffusion MRI data was provided by the Human Connectome Project (HCP). This dataset contains pre-processed T1-weighted scans of 28 volunteers, from which the training, validation and testing sets are separated by 18, 5, 5 subjects, respectively. The subject data is organized into 18 b=0 s/mm^2 images, 90 b=1000 s/mm^2 images, 90 b=2000 s/mm^2 images, and 90 b=3000 s/mm^2 images, with an isotropic voxelsize of 1.25 mm.

3.2. Data pre-processing

The data processing was executed using Matlab version 9.12.0. To remove outliers, data points belonging to the last percentile were eliminated from the raw data acquired from the HCP. Inclusion of outliers in the dataset can cause substantial shifts in mean and standard deviation, resulting in intensity values that are less comparable across datasets. This step was essential to normalize the data post-sampling for the DSR network. Normalization was performed by subtracting the mean and dividing the standard deviation of brain voxels in the input images. Brain masks were extracted to exclude background noise from the training process. The threshold for each HCP dataset was determined by calculating the mean value of the first b=0 s/mm^2 volume. This was followed by performing a single morphological opening and closing operation using a structural element in the shape of a disk with a radius of 10 pixels.

3.2.1. Downsampling: The data underwent k-space downsampling using Fourier transformation to simulate a realistic acquisition of low-resolution images in MRI scans. The HCP dataset has a 1.25 mm isotropic configuration, and scaling the data by a factor of 2 resulted in the cropping of each dMRI data size from 145x174x145 to 72x86x72 in k-space. By leveraging the linear relationship between k-space and image space resolution, upsampling to the eigenvector image space yielded data with 2.50 mm isotropic voxels.

3.2.2. Upsampling: We evaluated three conventional methods for upsampling, namely cubic spline interpolation (CS), nearest neighbor interpolation (NN), and zero-padding (ZP). The CS and NN methods were applied immediately following the application of inverse Fourier transform. ZP was implemented in k-space by adding rows and columns containing zeros, followed by the application of inverse Fourier transform. Fig. 4 visualizes the difference in image details for each upsampling technique. Being the simplest interpolation approach, NN was included as a baseline reference.

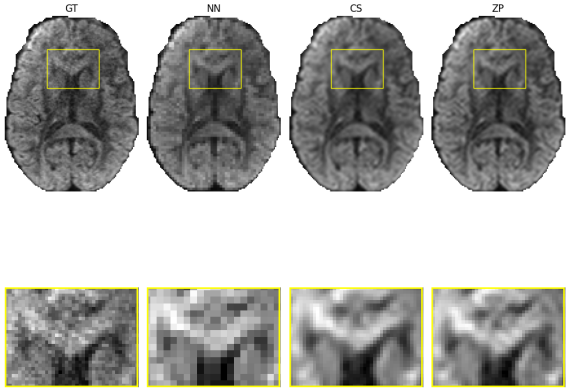

Fig. 4: An example image of the ground-truth (GT), nearest neighbor (NN), cubic spline (CS) and the zero-padding (ZP) used for the upsampling process.

3.3. Network implementation

In this study, the proposed DSR network is based on the DeepDTI method by Tian et al. [11]. The 2D and 3D model of the DSR network architecture are visualized by Fig. 5. The main difference between the models are the difference in input data dimensions and the convolutional kernels used. The proposed 2D and 3D networks utilize LR data with its corresponding residuals to the GT data and mask as input. The mask is used to train on the residuals within this region of interest. A deep convolutional neural network (CNN), consisting of stacked convolutional filters paired with ReLU activation functions and batch normalization in the middle layers. By using ReLU, networks can learn piece-wise linear mappings between LR and HR images, which results in faster training convergence and higher reconstruction quality, compared to networks using other nonlinear functions such as a sigmoid [21]. Furthermore, in order to accelerate and stabilize training of deep CNNs, Sergey et al. [22] states that batch normalization reduces internal covariate shift of networks. The DSR is employed to map the input image volumes to the residuals between the input and output image volumes. The output is a HR image generated by adding the predicted residuals to the input LR image.

While the 2D network is computationally less demanding, the full image slices in addition to patches were analyzed with the same model. The slices have dimensions of 145x174, while the patches have dimensions of 64x64 and 64x64x64 in 3D. In order to consider both image and angle spaces, this

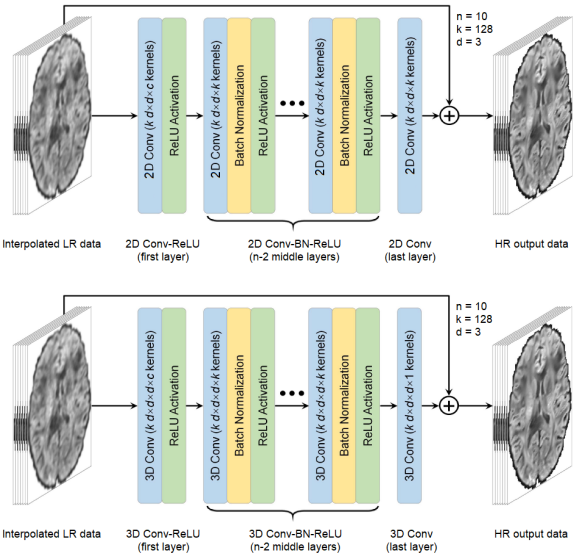

Fig. 5: The proposed 2D and 3D network utilizes low-resolution data with its corresponding ground-truth data and mask as input. A deep 3-dimensional convolutional neural network, consisting of stacked convolutional filters paired with ReLU activation functions (n = 10 layers, k = 128 filters, d = 3 kernel size, c = 1, 18 or 90 channels), is employed to map the input image volumes to the residuals between the input and output image volumes. The output is a high-resolution image generated by adding the predicted residuals to the input low-resolution image.

study was implemented with a multi-volume (MV) approach by an additional channel feature. This channel incorporates all available dMRI volumes from each b-value. Hereby, 18 b0 volumes and 90 volumes of each higher order b-value were predicted at once to account for the angular resolution.

The DSR was implemented using the Keras (version 2.9.0) and Tensorflow (version 19.11) frameworks. For the implementation of the DSR, the empty slices and patches were excluded, so the model would not predict empty images, while the majority of some slices and patches are black. The optimization of CNN parameters was achieved using the Adam optimizer with a learning rate of 7.3e-3 and the mean-square-error (L2) loss function, which was compared to the ground-truth images. The utilization of L2 loss leads to an increase in the distance between larger and smaller errors. [23] Specifically, the L2 loss uses a more severe penalty on larger errors, while penalizing less towards smaller errors. The L2 loss was calculated only within the predefined brain mask.

3.4. Statistical analysis

Following the optimization of the network and the training process, an independent testing set was used for statistical analysis. Five subjects were predicted to evaluate the network performance using the mean squared error (MSE), peak signal to noise ratio (PSNR), and structural similarity metrics (SSIM). The residuals between the predicted data and the ground-truth eigenvector data were analyzed to determine areas in which the model demonstrated the greatest capability to supplement missing details.

MSE is a metric that combines the variance and bias of an estimator [24]. In the case of an unbiased estimator, the MSE is equal to the variance of the estimator. Like variance, the MSE has units of measurement that are the square of the quantity being estimated.

$$MSE = \frac{1}{MN} \sum_{n=0}^{M-1} \sum_{m=1}^{N} [\hat{g}(n,m) - g(n,m)]^2 \qquad (1)$$

In Eq. 1, the terms $g(n,m)$ and $\hat{g}(n,m)$ represent the original image and the reconstructed image, respectively. The terms M and N represent the dimensions of the image.

PSNR is a measure that compares the power of the original signal to the power of the noise that is affecting its quality [25]. The comparison is expressed in decibels (dB). Since signals often have a wide dynamic range, the logarithm of the ratio is used to compute the PSNR. The dynamic range is the difference between the largest and smallest possible values of the signal, which can vary depending on the signal's quality. The PSNR is expressed by Eq. 2, where the peakval is the maximum possible pixel value of the image.

$$PSNR = 10 \log_{10} \left(\frac{peakval^2}{MSE} \right) \qquad (2)$$

SSIM is a perception-based model that considers image degradation as a change in the perception of structural information [26]. It also takes into account other important perception-based factors, such as luminance masking and contrast masking. Structural information refers to strongly interdependent or spatially close pixels that provide important information about visual objects in the image domain. Luminance masking refers to the phenomenon where the distorted part of an image is less visible in the edges, while contrast masking refers to the distortion being less visible in the texture of the image. SSIM is used to estimate the perceived quality of images and videos by measuring the similarity between the original and the recovered images.

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \qquad (3)$$

The SSIM can be expressed as Eq. 3, where x and y are the two images being compared. The variables l(x,y), c(x,y), and s(x,y) represent the luminance, contrast, and structure of the images, respectively. The constants α, β, and γ are weighting parameters that determine the relative importance of each factor in the calculation of SSIM. The equation expresses the similarity between the two images based on their luminance, contrast, and structure.

Based on these evaluation metrics, the best combination of the model and resampling method was picked for further investigation. The FA images of all 28 subjects were aligned to MNI space using ANTS with the PNL TBSS PIPELINE [27]. Subsequently, residual images were co-registered to

MNI space for group analysis. The mean residuals and mean absolute residuals were then computed.

Diffusion tensor imaging (DTI) was applied to the dMRI data produced by the most optimal combination. Due to data complexity, only the b0s and b1000s image volumes are used for further investigation at this stage. The fractional anisotropy (FA) and mean diffusivity (MD) maps obtained with LR and DSR were compared voxelwise.

At last, whole brain FT, left arcuate FT and the corresponding dice scores were analysed on all 5 test subjects. The whole brain FT will illustrate the ability to reproduce a fair amount of fibers for analyzing specific white matter pathways. The arcuate fasciculus is a crucial white matter pathway connecting the frontal language production regions with the superior temporal gyrus, also known as the Wernicke's area, which is responsible for auditory perception in the brain [28]. For each subject, the left arcuate will be manually delineated individually. Although the precise extraction of a specific tract can be a challenging task, the ROI will remain consistent across the high-resolution GT, LR and DSR predictions. The dice score is employed to evaluate the overlapping regions between GT and both LR and DSR.

FT parameters:

- 'SeedPointsRes' = [2 2 2] mm
- 'StepSize' = 0.6 mm
- 'FAThresh' = 0.2
- 'AngleThresh' = 30°
- 'FiberLengthRange' = [30 500] mm

4. Results

The performance of the DSR network using different up sampling methods, namely nearest neighbor (NN), cubic spline (CS), and zero-padding (ZP) in combination with 2D and 3D models was evaluated. The mean squared error (MSE), peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) were calculated for five independent test subjects, and the results were presented in Table I and the box-plots of Fig. 6. The NN method exhibited the highest overall MSE and the lowest PSNR and SSIM. In contrast, the ZP method had the lowest MSE and the highest PSNR and SSIM. The CS method performed better than NN, but it lacked the ability to reconstruct a higher resolution compared to the ZP method.

To gain a deeper understanding of the results, an overview of one example slice is provided by Fig. 7 for visualization purposes. The models attempted to correct for the loss of image details, which are perceivable within the central portion of the scoped areas. The largest differences between the modalities can be observed by examining the NN approach, where each network model attempts to introduce different image details to the LR input data. The LR NN data initially possesses fewer details, because of its utilization of intensities for neighboring interpolation. Its blocky appearance persists when implementing the DSR models, albeit smoother in appearance. Furthermore, the CS and ZP methods contain

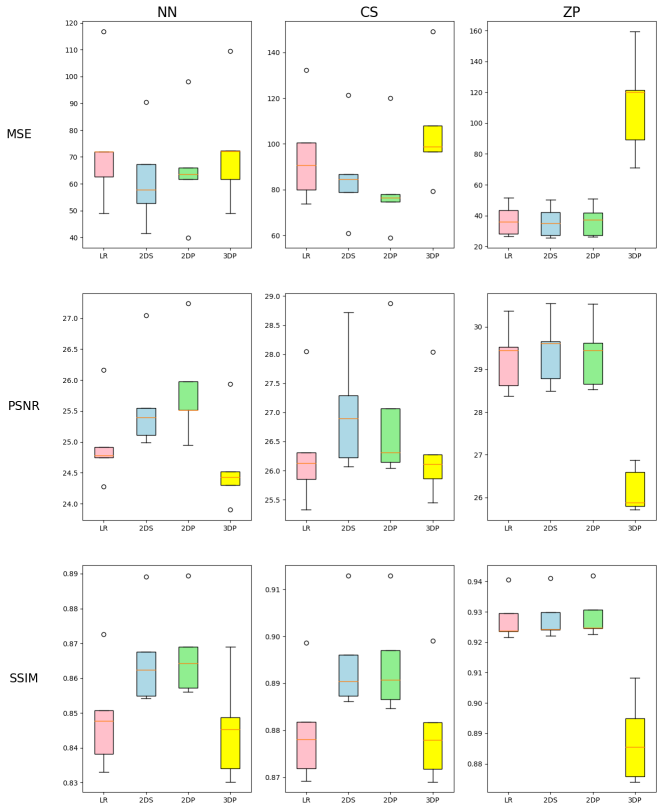

Fig. 6: Results of the DSR network performances. The NN, CS and ZP have been implemented in combination with the 2D slices (2DS), 2D patches (2DP) and 3D patches (3DP) approach.

image properties similar to a denoised smoother representation of the GT data and maintain this characteristic when executing various models. The center of these LR images exhibits greater intensities in comparison to the GT. While the 2D models of the CS and ZP techniques reduced the intensities in these regions, the 3D model conserved a larger proportion of the image details from the LR input.

The addition of 2D DSR models demonstrated a notable improvement in lowering the MSE by about 9.85%, while increasing the PSNR (3.41%) and SSIM (1.16%). In contrast, the 3D DSR model slightly enhanced the LR data and often resulted in a reduction in image resolution. The improvement of the neural networks was the highest in combination with NN, albeit the final statistics were still lower compared to ZP and CS. The input data format, whether as patches or full image slices, showed minor differences in performance, as both 2D slice and 2D patch methods displayed similar results throughout the process. Finally, the combination of ZP and a neural network with 2D patch approach demonstrated the most optimal overall results, with a MSE, PSNR, and SSIM of 36.109 ± 9.24, 29.418 ± 0.723, and 0.929 ± 0.007, respectively. The complete performances for each b-value are presented in Appendix I.

TABLE I: DSR network performances

		LR	2D slices	2D patches	3D patches
NN	MSE	95.421 ± 20.61	86.411 ± 19.657	81.62 ± 20.385	106.348 ± 23.36
	PSNR	24.978 ± 0.63	25.617 ± 0.742	25.839 ± 0.772	24.616 ± 0.691
	SSIM	0.848 ± 0.014	0.866 ± 0.013	0.867 ± 0.012	0.845 ± 0.014
CS	MSE	74.48 ± 22.799	65.866 ± 18.653	61.954 ± 16.519	72.908 ± 20.207
	PSNR	26.333 ± 0.918	26.888 ± 1.055	27.039 ± 0.95	26.348 ± 0.891
	SSIM	0.88 ± 0.01	0.894 ± 0.01	0.895 ± 0.01	0.88 ± 0.011
ZP	MSE	37.118 ± 9.416	36.661 ± 9.286	**36.109 ± 9.24**	112.282 ± 30.288
	PSNR	29.265 ± 0.71	29.36 ± 0.728	**29.418 ± 0.723**	26.17 ± 0.472
	SSIM	0.928 ± 0.007	0.928 ± 0.007	**0.929 ± 0.007**	0.888 ± 0.013

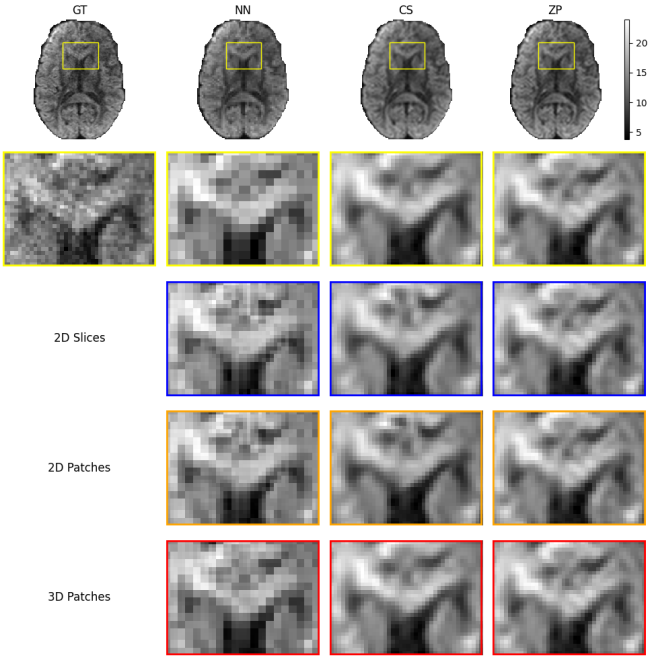

Fig. 7: Example slice to visualize the performances of the DSR network. The NN, CS and ZP have been implemented in combination with the 2D slices, 2D patches and 3D patches approach. The region highlighted by the yellow rectangle has been dissected and scoped for all approaches.

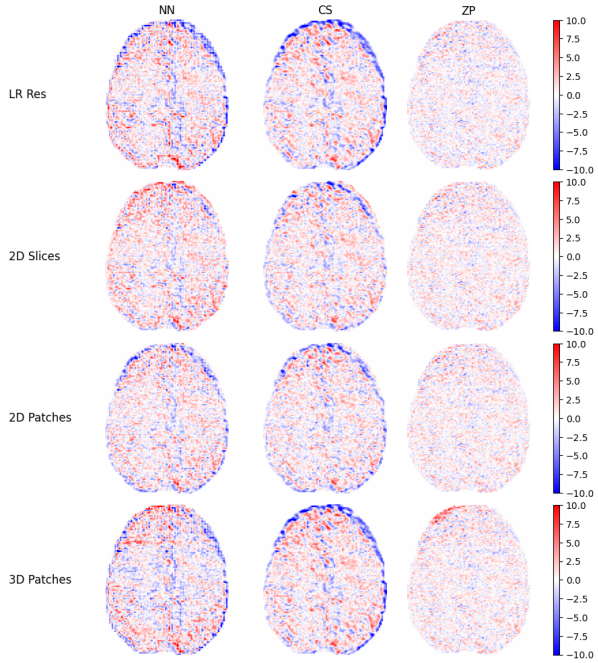

Fig. 8: The residuals of the DSR algorithm. A sample slice was selected from one of the test subjects and visualized to assess the variations in upsampling techniques with respect to the network models. The color coding of the residuals in the image visualizes that light colors correspond to lower residuals, while darker colors indicate higher residuals. Lower residuals indicate a better correspondence between the generated image and ground truth data.

The residuals of one test subject's b=1000s/mm^2 image are illustrated in Fig. 8. The residual analyses support the results in Table I and Fig. 6. The NN and CS methods initially exhibit a larger variance of residual amplitudes compared to the ZP method. The network tends to overcompensate for brain edges towards the skull when using the 2D slice method on the NN and CS data. In contrast, ZP demonstrates minor variances in residual amplitude among different models. The 3D model, on the other hand, overcompensates only specific areas towards the outer edges of the brain. Although, there is limited potential for improvement with ZP compared to NN and CS, the network consistently produced minor adjustments that enhanced the images with the 2D implementations.

For the quantitative analysis of FA maps, the residuals compared to the GT were calculated and the results are visualized in Fig. 9. The figure displays the mean absolute error (mae) and the mean of the residuals for the LR and DSR methods, with the mean mae above and the mean residuals at the bottom. The mean mae indicates the absolute values of the residuals, whereas the mean contains negative values. The findings indicate a minor variation in the mean mae residuals. DSR prediction displays fewer high residuals as compared to the LR map. Additionally, the mean mae for the LR is slightly higher than that of the DSR when examining the largest residuals in the image. A majority of these larger residuals are located at the boundaries of white and gray matter. In general, the FA maps demonstrate differences between the LR and DSR techniques. The LR method exhibits more error margins within the white matter and edges compared to the DSR maps. This corresponds to a higher degree of similarity between the DSR and GT compared to the LR and GT data.

Fig. 10 visualizes the absolute FA and MD maps of a test subject with the corresponding difference compared to the GT. There is no apparent improvement when comparing FA

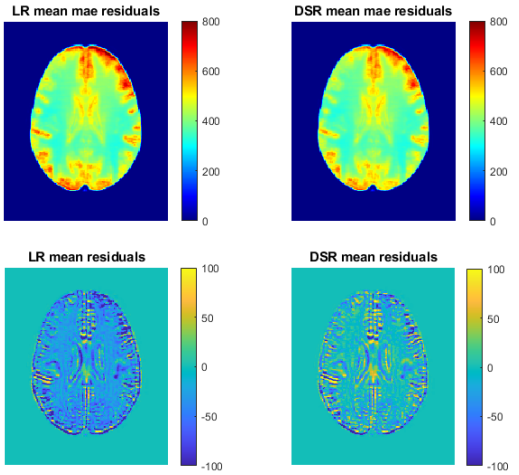

Fig. 9: The quantitative representation of the mean absolute error (mae) residuals above and the mean residuals of the FA maps below. These maps were constructed by aligning and compressing all available data.

and MD obtained with the LR and the DSR. The mean FA of the GT, LR and DSR are 0.451±0.138, 0.448±0.142, and 0.449±0.142, respectively.

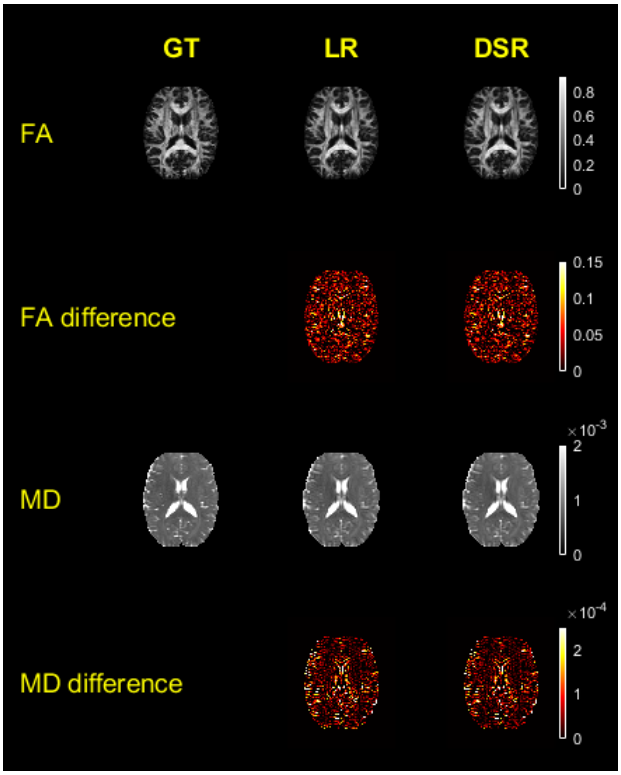

Fig. 10: The DTI based results of the absolute fractional anisotropy and the mean diffusivity (MD). The differences of the low-resolution (LR) and predicted high-resolution (DSR) with respect to the ground-truth (GT) are visualized with heatmaps. The mean FA of the GT, LR and DSR are 0.451±0.138, 0.448±0.142, and 0.449±0.142, respectively.

The whole brain tractography of one subject from the test set is visualized in Fig. 11. GT contains the highest number of fibers, a total of 70349 fibers. Conversely, the LR and DSR contain 68732 and 68999 fibers, respectively. The LR exhibits the lowest fiber count, whereas the DSR was able to reconstruct more fiber pathways. The visualization highlights that the LR is less dense in comparison to the GT, while the DSR prediction appears to be denser than the GT. The DSR generated more similar fiber pathways around the fiber tract of the LR data, even though the DSR fiber tract still contains considerably fewer fibers than the GT.

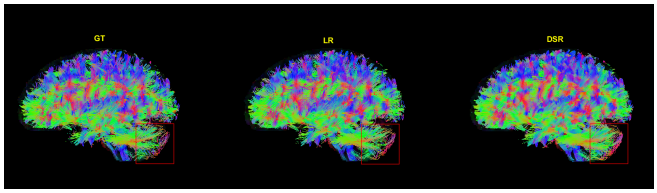

Fig. 11: Whole brain tractography visualizing the GT, LR and DSR prediction of one test subject. The GT, LR and DSR contains 70349, 68732, and 68999, respectively.

Fig.13 visualizes the outcomes of the whole brain FT, left arcuate FT, and the corresponding dice scores. In general, GT of the whole brain FT comprises a greater number of fibers than the LR and DSR techniques. Furthermore, the DSR method is capable of generating marginally more fiber tracts than the LR approach. However, there are no notable differences among the number of fibers related to GT, LR, and DSR in the fiber tracts of the left arcuate. The dice scores exhibit minor differences of overlapping regions between the LR and DSR data with the GT, but remain in the range of 0.82 and 0.89.

The overlap among the GT, LR and DSR fiber tracts for subject number 1 are illustrated in Figure 12. The fiber tracts of all other tested subjects are presented in Appendix II. The majority of the fibers show comparable orientation, and there are not many abnormal tracts visible. Although there is a slight variability in the number of fibers, with the GT containing 324, LR 298, and DSR 297 fibers. The dice scores for LR and DSR are 0.85 and 0.86, respectively, with the DSR marginally improving the dice score.

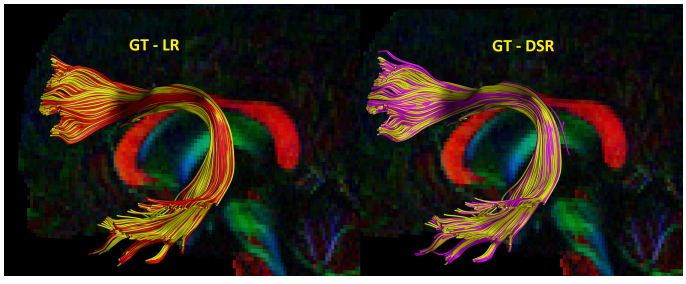

Fig. 12: The fiber tracts of the left arcuate from the GT (red), LR (purple) and DSR (blue) overlapped from volunteer number 1.

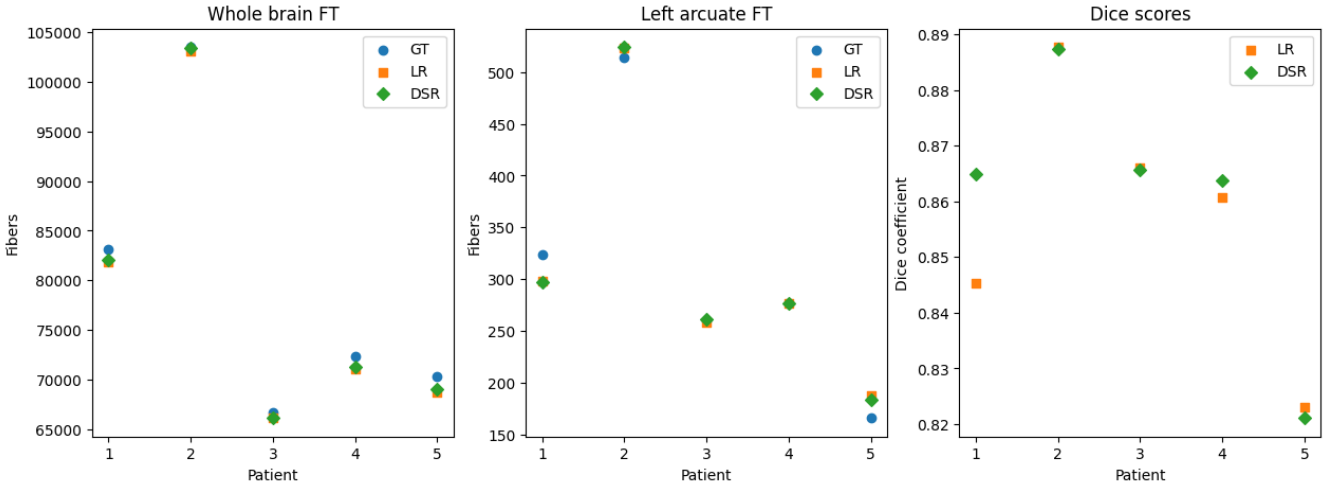

Fig. 13: Resulting plots of the whole brain fiber tractography, left arcuate fiber tractography and the corresponding dice scores. All 5 test subjects are included and the GT, LR and DSR are visualized.

5. Discussion

In this work, DSR networks in combination with conventional upsampling approaches were evaluated to improve the quality of dMRI data for fiber tractography.

Different types of DSR inputs were considered for enhancing diffusion MRI data, including 2D patches, 2D whole slices, and 3D volumes. The combination of ZP and the DSR method using 2D patches resulted in the most optimal overall results, as measured by MSE, PSNR, and SSIM. The respective values were 36.109 ± 9.24, 29.418 ± 0.723, and 0.929 ± 0.007, as compared to ZP only.

Subsequently, the performance of this approach was evaluated on fiber tractography to assess the reconstruction of white matter pathways. The results revealed minimal visible differences between fiber tractography using the original data, downsampling, and DSR implementation. These findings suggest that the tested DSR networks are promising for improving the spatial detail of diffusion MRI data but are still insufficient to enhance the quality of fiber tractography.

The performance of three different image reconstruction methods, namely, the NN, CS and ZP techniques were used for evaluation. Our results demonstrated that the LR NN data initially started with the lowest resolution, as expected, due to its reliance on pixel intensities from neighboring pixels while interpolating. Although the NN method initially consisted of its blocky appearance, the DSR models were able to add spatial image details, resulting in larger improvements compared to CS and ZP. Although the NN method was expected to have limited potential for improvement, it demonstrated some enhancements, likely due to the larger room for optimization. The CS method performed better than NN, providing a smoother and denoised representation of the GT data, although it was unable to achieve a higher resolution than the ZP method. We selected the ZP method as the representative method for image reconstruction, as it mimics the MRI acquisition process in k-space. While high-frequency image details are discarded with ZP, contrast-related image properties were largely preserved. The findings demonstrate the preservation of the data, as evidenced by MSE of the LR ZP technique being approximately half that of LR using CS and almost a third of the LR NN MSE, as measured by 37.118 ± 9.416, 74.48 ± 22.799, and 95.421 ± 20.61, respectively. In addition to the PSNR and SSIM metrics indicating higher values for the LR ZP approach compared to the LR NN and CS data.

Furthermore, this study investigated the performance of 2D and 3D models in enhancing image resolution. The 2D models showed some improvements in terms of reducing MSE and increasing PSNR and SSIM. On the other hand, the 3D model lacked in the ability enhance low-resolution data and often resulted in a decrease in image resolution. This was however in contrast with our expectations, while the 3D CNN of the study by Tian et al. [18] achieved promising results, namely, PSNR values of approximately 31 dB and SSIM values of around 0.98. The key differences between this study and ours is the use of six optimized directions on a dataset including only b=0 s/mm^2 and b=1000 s/mm^2 images, along with their corresponding T1-weighted anatomical images. As a result, the performance of the DSR model might be influenced by the input data used by Tian et al., which incorporates the knowledge of its underlying dMRI physics. This could potentially explain why the current 3D DSR model, despite having an additional dimension compared to the 2D models, may be insufficient in capturing tissue properties.

Our experiments also involved the use of patches and whole image slices as input data, which revealed that both 2D slice and 2D patch methods demonstrated similar performance. This observation was unexpected. While the use of whole image slices corresponds with many slices into the training process, excluding only the empty ones, patches contain smaller sizes, leading to the exclusion of more patches due to the removal of empty patches. Consequently, the model was expected to

obtain more information regarding internal brain structures. However, our results suggest that this process did not majorly alter the predictive capacity of the model.

We evaluated the performance of DSR models in restoring these high-frequency image properties. Unfortunately, our results revealed minimal improvements on the ZP datasets. The network was unable to restore much of the image details, such as edges. While the 2D models did not degrade the resolution of their input data, they could not find the optimal configurations to enhance the already higher resolution data. Our findings suggest that while DSR enables greater improvements in lower resolution images, it lacks the ability to enhance higher resolution images. This limitation may be due to factors such as large network depth or less optimal hyper-parameters, as previously discussed.

The residual analysis revealed large variances throughout the brain, primarily located at the edges of different tissue properties, between gray and white matter regions. Although the residual analysis was limited to one slice of one dataset, it provided insights into the adjustments made by different models to conventional upsampling methods.

For the FA and MD analysis using diffusion tensor imaging data, there was minor improvement achievable when comparing FA and MD obtained with the LR and the DSR. However, the DSR method was able to generate slightly fewer residuals of higher amplitude than the LR approach. To evaluate the , all available data was first aligned and combined into a single volume and compared the residuals to the FA maps of the GT. The results concerning the quantitative analysis of FA showed that there was minor differences between the mean absolute error (mae) of the LR and DSR methods. These variances were more prominently visible at the edges of the brain. The LR mean mae was slightly higher than the DSR when examining the largest residuals in these regions. Subsequently, the LR technique contained more error margins within the white matter than the DSR, when visualizing the mean residual maps, which indicates an improvement on spatial image details by using the DSR network.

We also evaluated the performance of the GT, LR, and DSR methods in generating whole brain and left arcuate fiber tracts. The whole brain fiber tract generated using the GT method contained a higher number of fibers overall, as compared to the LR and DSR methods. The DSR method produced slightly more fiber tracts than the LR method. As expected, the LR method resulted in a less dense fiber tract compared to the GT method, reflecting the fewer fibers constructed according to the same fiber tractography criteria. However, the DSR method exhibited a denser representation of the GT data. While the expectation was to reconstruct more fibers as compared to the LR data, the DSR method may also potentially discover new fiber pathways by denoising the GT data. Conversely, there were no major differences observed in the number of fibers among the GT, LR, and DSR methods for the left arcuate fiber tract analysis.

The dice scores showed minor differences of overlapping regions between the LR and DSR data with the GT. Therefor, it is difficult to determine the performance of the model based on the dice scores. This could also discuss the fact of connecting different parts of the brain, which were not applicable when evaluating the GT data. The DSR methods could potentially implement the denoising on the GT data, which could explain the differences in fiber reconstruction and regions of overlap.

According to the results, the DSR network showed promise in improving dMRI data through the multi-volume (MV) approach, but this approach is unfortunately limited to a specific input data structure from the HCP dataset. This implies that the DSR model can only be implemented with a particular data size. The MV approach requires a substantial amount of data input, which can be challenging to optimize with numerous parameters. Bao et al. [20] demonstrated that increasing the depth of the network can result in the loss of feature information from various layers due to the vanishing gradient problem. This is characterized by the activation functions of many layers, which cause the gradients to approach zero and hinder the training procedure. This issue could potentially explain the minimal improvements of the DSR method when applied to conventional upsampling techniques.

Furthermore, hyper-parameter optimization was additionally a challenge due to the many degrees of freedom involved. Most of the hyper-parameters were optimized manually, like the amount of layers, number of filters, epochs, early stopping and learning rate. However, this process was time-consuming and yielded minimal effects. As the depth of the neural network increases with the number of layers and filters, the epochs and early stopping are crucial criteria for preventing overfitting [15]. Early stopping refers to the number of epochs that the model can go without improving, beyond which the current model is presented as the best possible configuration. However, this may also indicate that the model has become trapped in a local minimum and has not yet achieved the optimal model, which may require more epochs. This could well be the case with the MV approach, while the training procedure was relatively fast compared to the SV. Without the use of early stopping, the model allows large computational costs, even after finding the most optimal model, which is computationally inefficient.

However, it should be noted that the orientation of fibers and endpoints of fibers exhibit variations, when visualizing all the tracts overlapped on top of each other. This could be attributed to the connection of various brain regions, or due to the inaccuracy of the ROI placements. All ROIs are manually placed, so it is compelling that some tracks outside of the arcuate were also included.

6. Conclusion

In conclusion, the aim of this study was to evaluate whether the DSR network in combination with conventional upsampling approaches can improve the quality of dMRI data for fiber tractography. DSR networks were evaluated for enhancing diffusion MRI data, including 2D patches, 2D whole slices, and 3D volumes. The combination of ZP and the 2D patches DSR method yielded the most optimal overall results based

on measurements of MSE, PSNR, and SSIM, with respective values of 36.109 ± 9.24, 29.418 ± 0.723, and 0.929 ± 0.007, when compared to ZP alone.

Subsequently, the performance of this approach was assessed on fiber tractography to evaluate the reconstruction of white matter tracts. The results demonstrated minimal visual differences between fiber tractography using the original data, downsampling, and DSR implementation. These findings suggest that the tested DSR networks hold promise for improving the spatial detail of diffusion MRI data, but they still fall short in enhancing the quality of fiber tractography.

It would be beneficial for future studies to focus on hyperparameter optimization, given the large degree of freedom involved. The large input data size may not be advantageous for training, so reducing the amount of data using spherical harmonics (SH) to compress it to the desired number of angles could be a viable solution. Using the compressed SH data as input volumes for the DSR model could potentially improve performance while reducing the number of parameters involved in the process. Additionally, exploring more complex models such as the SRGAN could improve network performance beyond the current CNN model.

Overall, these results suggest that the tested DSR networks are promising to improve the spatial detail of diffusion MRI data, but seem to minimally impact the reconstruct brain connectivity by fiber tractography, suggesting the need for additional research in this direction.

REFERENCES

[1] J. Bernal, K. Kushibar, D.S. Asfaw, S. Valverde, A. Oliver, R. Marti, and X. Llado. Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. *Artificial Intelligence in Medicine*, 95:64–81, 2019.

[2] Y. Assaf, H. Johansen-Berg, and T. de Schotten. The role of diffusion mri in neuroscience. *NMR in biomedicine*, 32, 2017.

[3] E. Dai, S. Liu, and H. Guo. High-resolution whole-brain diffusion mri at 3t using simultaneous multi-slab (smslab) acquisition. *NeuroImage*, 237, 2021.

[4] J.D. Tournier. Diffusion mri in the brain–theory and concepts. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 112:1–16, 2019.

[5] R. Jones, G. Grisot, J. Augustinack, C. Magnain, D. A. Boas, B. Fischl, ..., and A. Yendiki. Insight into the fundamental trade-offs of diffusion mri from polarization-sensitive optical coherence tomography in ex vivo human brain. *Neuroimage*, 214:116704, 2020.

[6] Y. Chen, Y. Xie, Z. Zhou, F. Shi, A.G. Christodoulou, and D. Li. Brain mri super resolution using 3d deep densely connected neural networks. *IEEE 15th international symposium on biomedical imaging*, pages 739–742, 2018.

[7] S.O. Dumoulin, A. Fracasso, W. van der Zwaag, J.C.W. Siero, and N. Petridou. Ultra-high field mri: Advancing systems neuroscience towards mesoscopic human brain function. *NeuroImage*, 168:345–357, 2018.

[8] W. Yang, X. Zhang, Y. Tian, W. Wang, J. H. Xue, and Q. Liao. Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, 21:3106–3121, 2019.

[9] C. Dong, C. Loy, and X. Tang. Accelerating the super-resolution convolutional neural network. *in Proceedings of the European Conference on Computer Vision*, page 391–407, 2016.

[10] C. Dong, C.C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. *Proc. Eur. Conf. Comput. Vis.*, page 184–199, 2014.

[11] Q. Tian, B. Bilgic, Q. Fan, C. Liao, c. Ngamsombat, Y. Hu, T. Witzel, K. Setsompop, Polimeni J.R., and Huang S.Y. Deepdti: High-fidelity six-direction diffusion tensor imaging using deep learning. *NeuroImage*, 219, 2020.

[12] H. Li, Z. Liang, C. Zhang, R. Liu, J. Li, W. Zhang, ..., and L. Ying. Superdti: Ultrafast dti and fiber tractography with deep learning. *Magnetic resonance in medicine*, 86(6):3334–3347, 2021.

[13] Z. Wang, J. Chen, and S. C. Hoi. Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3365–3387, 2020.

[14] Z. Akkus, A. Galimzianova, A. Hoogi, D.L. Rubin, and B.J. Erickson. Deep learning for brain mri segmentation: State of the art and future directions. *Journal of Digital Imaging*, 30:449–459, 2017.

[15] L. Rice, E. Wong, and Z. Kolter. Overfitting in adversarially robust deep learning. *In International Conference on Machine Learning*, pages 8093–8104, 2020.

[16] I. Sánchez and V. Vilaplana. Brain mri super-resolution using 3d generative adversarial networks. *1st Conference on Medical Imaging with Deep Learning (MIDL 2018)*, 2018.

[17] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[18] Q. Tian, Z. Li, Q. Fan, C. Ngamsombat, Y. Hu, C. Liao, ..., and S. Y. Huang. Srdti: Deep learning-based super-resolution for diffusion tensor mri. *NeuroImage*, 2021.

[19] S. Luo, J. Zhou, Z. Yang, H. Wei, and Y. Fu. Diffusion mri super-resolution reconstruction via sub-pixel convolution generative adversarial network. *Magnetic Resonance Imaging*, 88:101–107, 2022.

[20] L Bao, F. Ye, C. Cai, K. Zeng, P.C.M van Zijl, and Z. Chen. Undersampled mr image reconstruction using an enhanced recursive residual network. *Journal of Magnetic Resonance*, 305:232–246, 2019.

[21] J.S. Choi and M. Kim. A deep convolutional neural network with selection units for super-resolution. *In Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 154–160, 2017.

[22] I. Sergey and S. Christian. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn.*, page 448–456, 2015.

[23] Z. Pan, W. Yu, B. Wang, H. Xie, V. S. Sheng, J. Lei, and S. Kwong. Loss functions of generative adversarial networks (gans): opportunities and challenges. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 4(4):500–522, 2020.

[24] J. Søgaard, L. Krasula, M. Shahid, D. Temel, K. Brunnstrom, and M. Razaak. Applicability of existing objective metrics of perceptual quality for adaptive video streaming. *Society for Imaging Science and Technology IST International Symposium on Electronic Imaging*, 2016.

[25] R.G. Deshpande, L.L. Ragha, and S.K. Sharma. Video quality assessment through psnr estimation for different compression standards. *Indonesian Journal of Electrical Engineering and Computer Science*, 11:918–924, 2018.

[26] A.C. Brooks. Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions. *IEEE Transactions on Image Processing*, 17:1261–1273, 2008.

[27] T Billah, S Bouix, and O. Pasternak. Generalized tract based spatial statistics (tbss) pipeline. *https://github.com/pnlbwh/tbss*, 2019.

[28] D. Giampiccolo and H. Duffau. Controversy over the temporal cortical terminations of the left arcuate fasciculus: a reappraisal. *Brain*, 145(4):1242–1256, 2022.

7. APPENDIX

7.1. Appendix I

TABLE II: DSR network performances on NN

Network performance					
		NN		NN+Network	
b-value			2D slices	2D patches	3D patches
b0	MSE	46.118 ± 3.171	29.644 ± 2.482	23.479 ± 1.582	38.18 ± 2.055
	PSNR	28.427 ± 3.169	30.362 ± 3.423	31.331 ± 3.218	29.019 ± 3.08
	SSIM	0.87 ± 0.068	0.909 ± 0.052	0.91 ± 0.052	0.874 ± 0.065
b1000	MSE	132.821 ± 11.243	132.007 ± 12.21	119.224 ± 11.65	160.744 ± 13.655
	PSNR	23.893 ± 3.734	24.113 ± 4.105	24.608 ± 4.112	23.009 ± 3.768
	SSIM	0.853 ± 0.077	0.876 ± 0.067	0.878 ± 0.067	0.849 ± 0.078
b2000	MSE	99.029 ± 6.374	81.238 ± 5.627	77.776 ± 5.5	98.744 ± 6.029
	PSNR	24.597 ± 2.987	25.497 ± 3.084	25.656 ± 2.967	24.598 ± 2.967
	SSIM	0.849 ± 0.081	0.864 ± 0.075	0.867 ± 0.074	0.849 ± 0.081
b3000	MSE	64.794 ± 4.22	58.411 ± 3.595	60.15 ± 3.649	74.297 ± 5.637
	PSNR	25.75 ± 2.594	26.279 ± 2.723	26.147 ± 2.725	25.345 ± 2.751
	SSIM	0.839 ± 0.086	0.848 ± 0.083	0.849 ± 0.082	0.833 ± 0.091

TABLE III: DSR network performances on CS

Network performance					
		CS		CS+Network	
b-value			2D slices	2D patches	3D patches
b0	MSE	27.434 ± 2.201	17.95 ± 1.543	20.418 ± 1.5	25.091 ± 1.18
	PSNR	30.815 ± 3.248	32.682 ± 3.585	32.224 ± 3.728	31.05 ± 3.173
	SSIM	0.91 ± 0.046	0.936 ± 0.033	0.934 ± 0.034	0.913 ± 0.044
b1000	MSE	111.886 ± 11.135	88.059 ± 8.497	94.196 ± 10.0	105.933 ± 10.561
	PSNR	25.086 ± 4.328	25.951 ± 4.158	25.943 ± 4.531	25.224 ± 4.228
	SSIM	0.887 ± 0.058	0.909 ± 0.047	0.909 ± 0.047	0.888 ± 0.058
b2000	MSE	73.153 ± 5.178	60.276 ± 4.425	67.036 ± 4.456	73.106 ± 5.091
	PSNR	25.981 ± 3.187	26.764 ± 3.083	26.403 ± 3.263	25.981 ± 3.179
	SSIM	0.88 ± 0.064	0.894 ± 0.058	0.893 ± 0.058	0.88 ± 0.064
b3000	MSE	48.461 ± 3.697	46.398 ± 3.691	45.994 ± 2.09	49.709 ± 3.323
	PSNR	27.036 ± 2.702	27.301 ± 2.812	27.257 ± 2.712	26.901 ± 2.663
	SSIM	0.867 ± 0.072	0.873 ± 0.07	0.873 ± 0.07	0.866 ± 0.073

TABLE IV: DSR network performances on ZP

Network performance					
		ZP		ZP+Network	
b-value			2D slices	2D patches	3D patches
b0	MSE	10.438 ± 0.951	9.045 ± 0.945	8.571 ± 0.858	10.704 ± 0.948
	PSNR	34.907 ± 3.647	35.564 ± 3.796	35.761 ± 3.743	34.602 ± 3.561
	SSIM	0.954 ± 0.024	0.958 ± 0.022	0.959 ± 0.022	0.951 ± 0.025
b1000	MSE	46.556 ± 4.906	45.206 ± 4.836	45.318 ± 4.867	65.055 ± 5.053
	PSNR	28.82 ± 4.133	28.969 ± 4.149	28.953 ± 4.144	27.148 ± 3.881
	SSIM	0.941 ± 0.031	0.941 ± 0.031	0.941 ± 0.031	0.926 ± 0.04
b2000	MSE	39.462 ± 2.042	38.632 ± 2.472	40.463 ± 2.809	258.518 ± 1.904
	PSNR	28.719 ± 3.178	28.815 ± 3.181	28.594 ± 3.146	20.51 ± 3.137
	SSIM	0.928 ± 0.04	0.928 ± 0.04	0.93 ± 0.039	0.816 ± 0.094
b3000	MSE	30.37 ± 2.319	29.604 ± 2.746	29.502 ± 2.677	30.236 ± 2.295
	PSNR	29.166 ± 2.868	29.279 ± 2.872	29.295 ± 2.874	29.193 ± 2.87
	SSIM	0.91 ± 0.051	0.91 ± 0.05	0.91 ± 0.05	0.909 ± 0.051

7.2. Appendix II

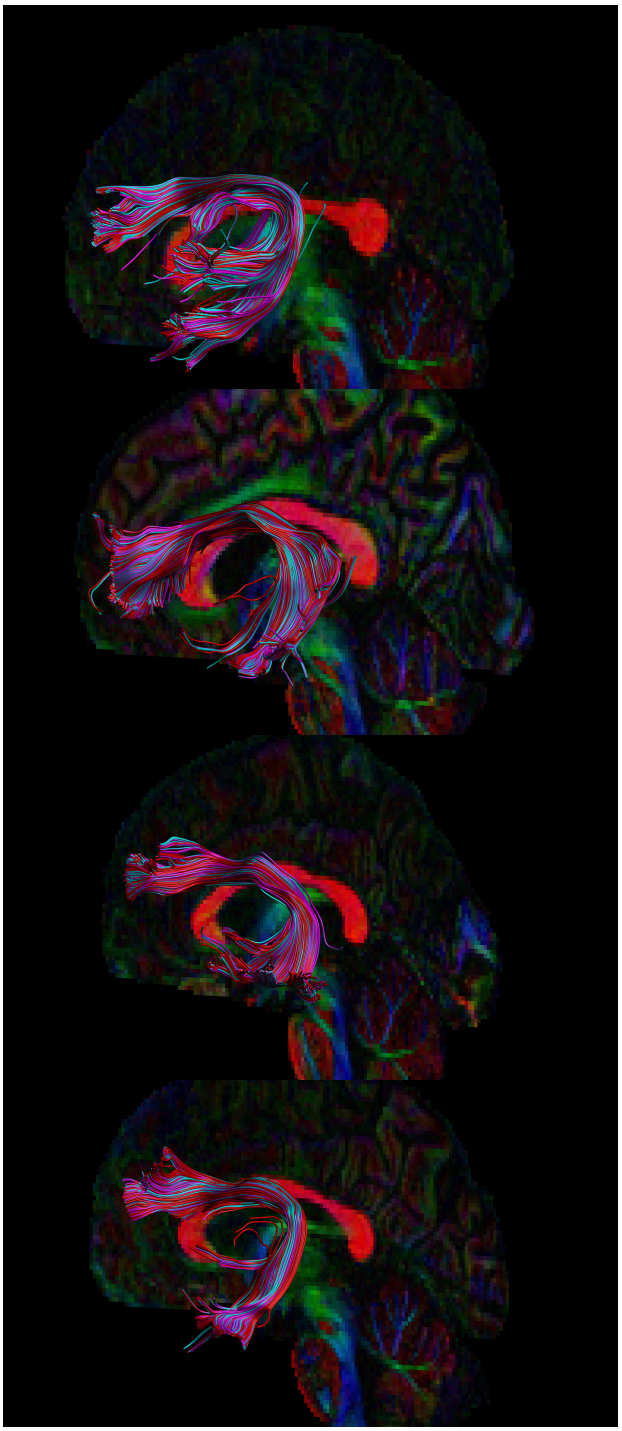

Fig. 14: The fiber tracts of the left arcuate from the GT (red), LR (purple) and DSR (blue) overlapped from volunteer number 2, 3, 4 and 5 in chronological order.